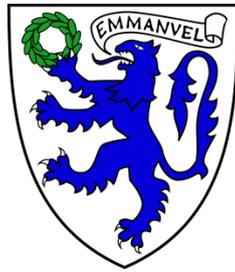


Physiological and Molecular Consequences of Large Y Chromosome Long Arm Deletions in Mice

Emma Elizabeth Philippa Johnson MRes



Emmanuel College



Department of Pathology
University of Cambridge

September 2017

This dissertation is submitted for the degree of Doctor of Philosophy. This dissertation is submitted to the University of Cambridge in accordance with requirements of the degree of Doctor of Philosophy in the School of the Biological Sciences.

Physiological and Molecular Consequences of Large Y Chromosome Long Arm Deletions in Mice

Emma Elizabeth Philippa Johnson

Summary

The mammalian Y chromosome contains genes important for male sexual maturity and reproduction. The mouse Y chromosome long arm harbours a number of multi-copy genes whose absence or reduced representation has been linked to sperm defects and offspring sex ratio distortion in favour of females. Understanding the biological mechanisms of sex ratio distortion and related sperm aberrations could not only result in benefits for fertility research, but also in the development of methods for large scale animal breeding pre-implantation sex selection. The distortion has been linked to an intragenomic conflict between the X and Y chromosomes that impacts spermiogenesis. Since the proportion of X- and Y-bearing sperm does not differ in affected animals, and there is no selective destruction of male embryos post-fertilisation, a functional difference must exist between the X- and Y-bearing sperm. This thesis describes the investigation into the physiological and molecular mechanisms of a large Y-chromosome long arm deletion in the mouse model MF1XY^{RIII}qdel.

The examination of physiological characteristics revealed a distinct sperm morphology within the deletion model. Detailed characterisation of sperm shape demonstrated that aberrations consistently occur within specific regions of the sperm head, linking the distorted morphology to particular maturation stages in the sperm cycle. Using sperm fluorescence *in situ* hybridisation, a novel and detailed comparison of X- and Y- bearing sperm has shown that a subtle distinction in shape also exists between the X- and Y- bearing sperm in the deletion model. Breeding data were examined and showed a skew towards female offspring and a slightly reduced litter size. Sperm enzyme activity assays did not reveal altered hyaluronidase activity in MF1XY^{RIII}qdel sperm.

Physiological differences between X- and Y- bearing sperm must result from differential gene expression, complicated by the syncytial nature of sperm development. To explore this, a detailed molecular characterisation of the MF1XY^{RIII}qdel phenotype in developing haploid spermatids was performed. Cellular elutriation and fractionation techniques were employed to separate spermatids at different stages of maturation and isolate different subcellular compartments. Differences in the transcriptional profile between these populations were analysed by microarray and RNA sequencing analysis of total and micro RNA. This work yielded a collection of coding and non-coding transcripts which show distinctive expression and compartmentalisation differences between the deletion model and its wild-type counterpart across several sperm maturation stages.

Combining these strategies has led to the identification of several gene products potentially implicated in observed physiological differences and the offspring sex ratio skew, providing candidate genes for further research.

Author's Declaration

This dissertation is the result of my own work conducted between 2013 and 2017, and includes nothing which is the outcome of work done in collaboration except as specified in the text. I further state that no substantial part of my dissertation has been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. This thesis does not exceed the prescribed word limit of 60,000 words set by the School of Biological Sciences, University of Cambridge.

Emma Elizabeth Philippa Johnson MRes

This thesis is dedicated to my family and friends. Thank you for everything.

Acknowledgements

This thesis, and the experience I have had conducting the work summarised in it, have only come about because of the involvement of several people. I want to start by thanking my supervisor, Professor Nabeel Affara; a man whom, like a great work of art, excels at comforting the disturbed and disturbing the comfortable. Nabeel has made me feel welcome in the lab from the moment I arrived and has made it harder for me to say goodbye. The best of what I have learnt from Nabeel has been all that we've spoken about away from the bench. I only hope to find another mentor from whom I can gain as much.

Thank you to my second supervisor, Dr Peter Ellis, who is now based at the University of Kent. Peter is an exceptionally pleasant man to know, whose detailed knowledge of this (and nearly every) subject has entertained me from the beginning. I am pleased to say that he only appears to use his powers for good. Thank you, Peter, for always being on the other end of an email. My third supervisor, Dr Alan Mileham, has supported me in things I've undertaken and made sure that opportunities I have had have been the best they can be. He has provided excellent conversation on the occasions that we have had the pleasure of meeting and has been a great source of advice. My final lab mentor is Dr Benjamin Skinner. Although not an official supervisor, Ben is the person I have worked most closely with over the past three years. Ben is unflagging in his enthusiasm for science. Thank you for the constant stream of insults and for taking mine well in return. You have been a great source of help and guidance over the past few years and I am grateful for it. I will not miss your puns.

I would like to thank Jo Bacon and Kim Lachani for their kind tuition, especially at the beginning of my time in the lab, and for their efforts and assistance with several experiments that did not quite make it into the thesis. Thank you also to Jo, and to Kerry Harvey for the kind words they provided when I needed them most. Thank you Dr Carole Sargeant for sharing some of the fascinating things she knows over the last few years and for thoroughly enriching my time in the lab. Thank you to fellow PhD student Courtney Landers for being a provider of much needed lunch time walks and musings. Thank you to Claire for always a cheery 'hello' and to all members of the lab for making the last four years what they have been. A final thank you to our wonderful part II students Myrto Vlazaki and Becki Nunn, who are both inspiring and lovely people.

From the Cambridge Genomics Services team, I'd first like to thank Dr Ed Farnell for answering all (or most of) my long list of questions with enthusiasm and for his questionable sense of humour. Thank you especially to Julien Bauer and Maria Gomez for being the foundation of my understanding of bioinformatics. I am very grateful for all the time they spent with me answering my many questions and helping me when I needed it. Thank you also to Dr Alex Covney, Dr Emily Clemente, Phil Howden, Chris Reitter and Marta Andrada for all their help and delightful conversation; and to Phil and Chris for answering my early morning phone calls to the lab! Thanks to Carl Wignall for

responding humorously to my IT woes and providing thoughtful “constructive” criticism of my handling of technical issues.

My sincerest thanks go to the kind members of the Griffin lab, and in particular to Dr Claudia Rathje for not only allowing me to use their lab space, but setting me up and booking time slots for me on their excellent microscope. Thank you for being my second lab-home. Additional thanks must go to Claudia and Bill (and the parrots) for housing me and entertaining me during a number of my visits to Canterbury.

Colleges can be a big part of student life at Cambridge, and Emmanuel College has been an important part of mine. I would like to thank the Emmanuel porters in particular. Their help and kindness has been indispensable, and they have always made me feel welcome whenever I visit the college, no matter how long I’ve been away.

From further afield, I must mention Jessie Martin and Laramie Pence. During my time at the University of Maryland in the summer of 2016, both Jessie and Laramie did everything they could to share their skills and knowledge with me and, most importantly, they made my time there thoroughly memorable. My gratitude extends to all members of the Telugu lab, who were hugely accommodating, facilitated my first experience of microinjection and CRISPR based experiments, and helped expand my understanding of the field.

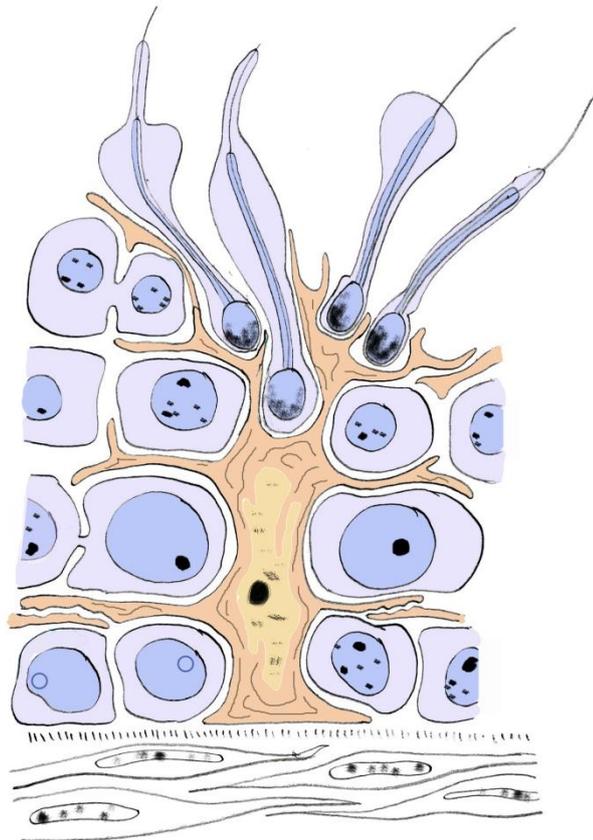
There are many more people I would like to thank, and without taking up several more pages, I will not be able to do so in full. So, I’d instead like to briefly thank my mentor Dr Paolo D’Avino, Dr Debbie Drage, Dr Nigel Slater, the Pathology graduates, and my friends at Emmanuel College and beyond. I’d also like to thank my sponsors and funders, the BBSRC, Genus Plc, and KTN for all the experiences and opportunities they have enabled and supported. It would not have been possible without them.

I have been extremely lucky to have enjoyed so much of what I have done over the years and this all begins with having remarkable parents. Thank you for everything you have done and still do for me. I couldn’t have asked for two more special and interesting people in my life. Thank you especially for being there for me over the phone, by text, and with emails over the last few years. I’ve missed you (and James!) hugely but never really felt we were that far apart. Thank you for being who you are; strong, loving, silly, and delightfully odd.

Finally, I want to thank my incredible partner, (Dr) Dave Williams. Without Dave, I don’t know where I would have found the mental fortitude to overcome the various challenges of the last four years. His intelligent and calm advice has been both indispensable and entertaining, and the great patience and care he has given me have made me feel ten times what I am. He is resilient, clever,

talented, hugely considerate, and one of the most caring people I have ever met. All the washing up, making of dinners, and doing of laundry may or may not be repaid.

I wish everyone the best and thank you all for your part in making the last four years what they have been. The writing of this thesis has been made notably easier by contributions from the British summer, primarily in the form of thick clouds and rain.



This thesis was prepared using Microsoft Word 2007. Figures were produced in one of the following: RStudio (v. 3.3.0), Adobe Photoshop CS, Inkscape v.0.9.

Table of Contents

1. Introduction	24
1.1. Sex Determination and the Meiotic Sexes	24
1.1.1. Sexual reproduction	25
1.2. The Sex Chromosomes	26
1.2.1. Evolution of the sex chromosomes	26
1.2.2. Anatomy of the mouse Y chromosome	30
1.2.3. The X chromosome	32
1.3. Spermatogenesis, Sperm Structure and Function	33
1.3.1. Sperm head morphology and indicators of sperm fitness	40
1.3.2. Gene expression throughout the course of spermatogenesis	44
1.3.3. Cytoplasmic sharing	45
1.4. Transmission Ratio Distortion, Sex Chromosome Meiotic Drive and Sex Ratio Distortion	47
1.4.1. Mendelian ratios and transmission ratio distortion	47
1.4.2. Transmission ratio distortion and sex chromosome meiotic drive in mouse	48
1.4.3. Mammalian sex ratios, intragenomic conflict and sex ratio distortion across taxa	49
1.4.4. Intragenomic conflict in the mouse	50
1.5. Deletions of the Male-Specific Region of the Y Chromosome Long Arm	52
1.5.1. Mouse MSYq deletion models	53
1.5.2. MSYq deletions and offspring sex ratio distortion	57
1.5.3. Offspring sex ratio distortion and physiological characteristics of the MF1XY ^{R111} qdel mouse	58
1.6. Transcriptomics and the Study of Gene Expression - Investigating the Transcriptome	60
1.6.1. RNA sequencing	61
1.6.2. Microarray	61
1.7. Rational for Investigatory Work	64

1.7.1.	Understanding mechanisms of meiotic drive: fertility research and applications in agriculture	64
1.8.	Candidate Mechanisms and Genes.....	65
1.9.	Structure of the Thesis.....	68
2.	Materials and Methods	70
2.1.	Animal Care and Preparation of Tissues, Cells and Fractions.....	70
2.1.1.	Extraction of mature sperm cells	70
2.1.2.	Extraction of testis cells and preparation of cellular suspension for elutriation	71
2.1.3.	Counting of cells using a haemocytometer and trypan blue staining.....	71
2.2.	Morphology and Microscopy Methods	72
2.2.1.	Silver nitrate staining of sperm	73
2.2.2.	Categorisation of sperm heads by shape	73
2.2.3.	Quantitative sperm head morphology methods.....	75
2.2.4.	Sperm fluorescence <i>in situ</i> hybridisation (sperm FISH)	75
2.2.5.	Fluorescence <i>in situ</i> hybridisation (FISH) of metaphase cells	77
2.3.	Other Physiological Assessment Methods.....	78
2.3.1.	Assessment of sperm hyaluronidase activity	78
2.4.	Preparation of Cells and Subcellular Fractions for RNA Sequencing and Microarray Gene Expression Analysis	81
2.4.1.	Separation of testis cell types by centrifugal elutriation	82
2.4.2.	Subcellular fractionation of cells	83
2.4.3.	DAPI and lectin staining for examination of cellular fractions	86
2.4.4.	Cell counting from microscopy images using ImageJ	87
2.5.	RNA Extraction and Sample Preparation	87
2.5.1.	RNA extraction	87
2.5.2.	Quality control of RNA samples.....	88
2.5.3.	Increasing concentration of purified RNA samples using Concentrator Plus.....	89
2.6.	RNA Sequencing Library Preparation	89
2.6.1.	Total RNA Illumina library preparation	89
2.6.2.	Small RNA Illumina library preparation.....	90

2.7.	Microarray Sample and Library Preparation	91
2.7.1.	Affymetrix total RNA sample preparation and processing	91
2.7.2.	Toray miRNA sample preparation and processing	92
3.	Physiological Characterisation of the MF1XY^{RIII}qdel Model	93
3.1.	Chapter Introduction.....	93
3.1.1.	Chapter objectives.....	94
3.2.	Offspring Sex Ratio and Litter Size in the MF1XY^{RIII}qdel Colony.....	94
3.2.1.	Introduction to offspring sex ratio and litter size assessment	94
3.2.2.	Results of offspring sex ratio and litter size assessment	95
3.2.3.	Discussion of findings of offspring sex ratio and litter size assessment	98
3.3.	Assessment of Sperm Head Morphology in MF1XY^{RIII}qdel mice.....	98
3.3.1.	Introduction to assessing morphology	98
3.3.2.	Results of qualitative sperm head morphology assessment	99
3.3.3.	Discussion of findings of qualitative sperm head morphology assessment	101
3.4.	Quantitative Assessment of Sperm Head Morphology in MF1XY^{RIII}qdel Mice	102
3.4.1.	Introduction to quantitative sperm head morphology assessment	103
3.4.2.	Structure and implementation of sperm head morphology analysis	104
3.4.3.	Image analysis.....	104
3.4.4.	Results of quantitative sperm head morphology assessment	109
3.4.5.	Discussion of findings of quantitative sperm head morphology assessment within and between strains	122
3.5.	Assessment of Sperm Hyaluronidase Activity in the Mature Sperm of MF1XY^{RIII}qdel Mice	124
3.5.1.	Introduction to assessing hyaluronidase activity.....	124
3.5.2.	Structure and execution of hyaluronic acid activity assessment	125
3.5.3.	Results of sperm head hyaluronidase assay	129
3.5.4.	Discussion of findings of sperm head hyaluronidase assay	131
3.6.	Chapter Summary	132
4.	RNA Sequencing Study of Gene Expression in Spermatids.....	134
4.1.	Chapter Introduction.....	134

4.1.1.	Investigating differential expression during spermiogenesis	135
4.1.2.	Investigation of transcript abundance in differing subcellular regions of spermatids in mice	136
4.1.3.	Messenger RNA transcript localisation and sharing	139
4.1.4.	Application of RNA sequencing	140
4.1.5.	Chapter objectives.....	141
4.1.6.	Notes for chapter.....	141
4.2.	Sample Preparation	142
4.2.1.	Cellular elutriation: examination of retrieved cell composition.....	143
4.2.2.	Further sample processing and subcellular fractionation	144
4.2.3.	RNA sample quality control	145
4.2.4.	RNA sequencing	145
4.3.	Data Processing and Quality Control.....	145
4.3.1.	Examination of technical and quality controls before trimming and mapping	146
4.3.2.	Trimming, mapping and quality control of reads.....	151
4.3.3.	Single replicate data and generation of differential expression log fold change values	154
4.3.4.	Overall trends in RNA sequencing data.....	155
4.4.	Results of Differential Expression Analysis of RNA Sequencing Data	158
4.4.1.	Increased expression of sex chromatin in MF1XY ^{RIII} qdel.....	158
4.4.2.	Expression patterns of control genes across the comparison data sets	161
4.4.3.	Investigation of differentially expressed genes: identification of candidate genes in offspring sex ratio distortion in MF1XY ^{RIII} qdel.....	165
4.4.4.	Investigation of differential expression of protein-coding and non-coding genes across whole cell samples	171
4.4.5.	Functional annotation analysis of shortlisted protein-coding genes by STRING (search tool for the retrieval of interacting genes/proteins)	176
4.4.6.	Examination of genes of interest in RNA sequencing dataset	178
4.5.	Chapter Discussion	179
4.6.	Chapter Summary	183

5. Microarray Study of Gene Expression in Spermatids	184
5.1. Chapter Introduction.....	184
5.1.1. Hypothesis and experimental design.....	184
5.1.2. Chapter objectives.....	186
5.1.3. Notes for chapter.....	187
5.2. Sample Preparation	188
5.2.1. Results of cellular population enrichment.....	189
5.2.2. Subcellular fractionation of whole cell populations.....	193
5.2.3. RNA sample quality control	193
5.3. Normalisation and Data Treatment	194
5.3.1. Data preparation.....	194
5.3.2. Examination of technical controls.....	195
5.3.3. Normalisation of microarray data: raw data (pre-normalisation) and post-normalisation quality control	196
5.4. Overall Trends in Microarray Data	198
5.4.1. Principal components analysis (PCA) – clustering analysis	198
5.4.2. Upregulation of sex-linked genes	202
5.5. Results of Differential Expression Analysis of MF1XY^{RIII} and MF1XY^{RIII}qdel Spermatid RNA Samples	204
5.5.1. Examination of control transcripts.....	206
5.5.2. Results of matched comparisons between strains – MF1XY ^{RIII} (WT) vs MF1XY ^{RIII} qdel (YQ)	208
5.5.3. STRING (search tool for the retrieval of interacting genes/proteins) analysis of differentially expressed protein-coding genes.....	213
5.5.4. Predicting sharing behaviour of differentially expressed genes.....	216
5.5.5. Genes of interest from the literature and previous work.....	218
5.6. Chapter Discussion	219
5.6.1. Comparison of results from RNA sequencing (Chapter 4) and microarray data (Chapter 5)	224
5.7. Chapter Summary	225

6. Micro RNA Expression Study of Spermatids and Mature Sperm.....	227
6.1. Chapter Introduction.....	228
6.1.1. Hypothesis and experimental design.....	233
6.1.2. Chapter objectives.....	234
6.1.3. Notes for chapter.....	234
6.2. Sample Preparation: Library Preparation, RNA Sample and RNA Library QC, Sequencing and Array Hybridisation of Samples.....	235
6.3. Preparation of Small RNA Sequencing Data.....	236
6.3.1. Raw data quality control	238
6.3.2. Trimming of low quality bases and adapters	240
6.3.3. Mapping, counting, normalisation and differential expression analysis	241
6.4. Results of Small RNA Sequencing Analysis	244
6.5. Preparation of miRNA Microarray Data	245
6.5.1. Micro RNA microarray raw data quality control.....	246
6.5.2. Normalisation and post-normalisation quality control.....	248
6.6. Results of miRNA Microarray Expression Analysis.....	251
6.7. Further Examination of Differentially Expressed miRNAs.....	252
6.8. Chapter Discussion	254
6.9. Chapter Summary	256
7. Discussion and Further Work.....	257
7.1. Further Work.....	261
8. Conclusion.....	263
Appendix.....	265
Supplementary Material	265

Table of Figures

Figure 1.1. Sketch of the predicted ‘1’ enfant’	25
Figure 1.2. Fusion of autosomal <i>Sox3</i>	27
Figure 1.3. Overview of the evolution of sex chromosomes	28
Figure 1.4. Simplified representation of the mouse (<i>Mus musculus</i>) Y chromosome	29
Figure 1.5. Comparative morphology of mouse and human sperm.....	34
Figure 1.6. Testis section from a MF1XY ^{RIII} mouse	35
Figure 1.7. Diagram of the murine testis	36
Figure 1.8. Overview of cellular progression during mammalian spermatogenesis	37
Figure 1.9. Structure of the mouse oocyte	39
Figure 1.10. A selection of sperm head morphologies from across metazoa.....	42
Figure 1.11. Progression of spermatogenesis in the mouse testis	45
Figure 1.12. RNA <i>in situ</i> of digoxigenin-labelled <i>Smok2b</i> transcripts	47
Figure 1.13. The ‘presumed mode of origin’ of the deletion Y chromosome.....	53
Figure 1.14. Examples of sperm taken from XY ^{RIII} qdel males and sperm from wild-type XY ^{RIII}	59
Figure 1.15. Metaphase spreads from MF1XY ^{RIII} and MF1XY ^{RIII} qdel	60
Figure 2.1. Flow chart overview of sample preparation for both RNA seq and microarray.....	81
Figure 2.2. Simplified diagram of Beckmann Coulter J26-XP Centrifugal Elutriator set up	82
Figure 2.3. Flow diagram of sample preparation resulting in collection of subcellular fractions.....	85
Figure 2.4. Cytospin chamber set up.....	86
Figure 3.1. Boxplot of litter sizes across strains	97
Figure 3.2. Individual litter sizes recorded by sire.....	97
Figure 3.3. Silver nitrate stained sperm heads provided as examples of morphology	100
Figure 3.4. The proportion of sperm categorised under each morphological type	100
Figure 3.5. Examples of dominant sperm head phenotypes seen in the strains compared	102
Figure 3.6. Simplified workflow of quantitative sperm head morphology assessment	105
Figure 3.7. Screenshot of detected nuclei during the manual curation stage	106
Figure 3.8. Example images of DAPI-stained sperm taken pre-FISH and post-FISH.....	107
Figure 3.9. Examples of post-FISH mapping to pre-FISH images	108
Figure 3.10. Overlay of the consensus nuclear shape from each of WT and YQ populations	110
Figure 3.11. Histogram showing the distribution of data for two parameters used in morphology comparisons	110
Figure 3.12. Scatterplot showing distributions of individual sperm heads from WT and YQ mice across two sperm head parameters.....	112
Figure 3.13. Segmented sperm head angle profile per population.....	112
Figure 3.14. Diagram of descriptive sperm regions.....	113

Figure 3.15. Comparison of segmented sperm head regions between WT and YQ populations.....	113
Figure 3.16. Graphical overlay of average sperm head shape from each X- and Y-bearing sperm population	114
Figure 3.17. Probability density functions of two sperm head parameters	114
Figure 3.18. Violin plot comparison of WT and YQ sub-populations by area in square microns.....	116
Figure 3.19. Comparison of segmented sperm head regions between WTX, WTY, YQX and YQY sperm populations	117
Figure 3.20. Proportion of sperm populations represented in each clustered (phenotype) dataset.....	119
Figure 3.21. Comparison of three major sperm head parameters and sperm DNA content	121
Figure 3.22. Comparison of primary incubation times used in enzyme-only experiments.....	126
Figure 3.23. Comparison of buffer solutions PBS and HTF	126
Figure 3.24. Enzyme-only concentration versus transmittance	128
Figure 3.25. Equivalent hyaluronidase concentration of sperm samples from individual MF1XY ^{RIII} (WT) and MF1XY ^{RIII} qdel (YQ) mice	130
Figure 3.26. Equivalent hyaluronidase concentration of sperm samples	131
Figure 4.1. Demonstration of potential distorter / responder product interaction.....	137
Figure 4.2. Demonstration of potential distorter / responder mechanisms implicated in offspring sex ratio distortion in MSYqdel mice.....	138
Figure 4.3. Overview of sample preparation conducted for analysis of RNA sequencing data	143
Figure 4.4. Overview of RNA data processing pipeline following RNA sequencing	146
Figure 4.5. Reads retrieved from each sample.....	147
Figure 4.6. Boxplots of quality values distributed across each read of the sample examined.....	148
Figure 4.7. Per sequence mean quality scores	150
Figure 4.8. Per sample read distribution by read categories before and after rRNA filtering	153
Figure 4.9. PCA plot of all RNA libraries post-data processing.....	156
Figure 4.10. Heatmap comparing all RNA samples for similarity of RNA composition.....	158
Figure 4.11. Comparison of overall expression of sex-linked and autosomal genes in early stage spermatids	159
Figure 4.12. Comparison of overall expression of sex-linked and autosomal genes in late stage spermatids	160
Figure 4.13. Heatmaps of log fold change (logFC) values	164
Figure 4.14. Overview of selection criteria for short-listing of genes	166
Figure 4.15. Heatmaps of log fold change (logFC) values	169
Figure 4.16. Graphical output of STRING protein functional association analysis at medium confidence.....	177
Figure 4.17. Remaining functional association networks at high confidence.....	178
Figure 5.1. Overview of total RNA sample preparation for microarray	186

Figure 5.2. Description of cell type attributions used for examination of cell composition	190
Figure 5.3. Examples of images used to determine the composition of cell populations	191
Figure 5.4. Array quality control workflow	195
Figure 5.5. Boxplots of probe fluorescence intensity pre- and post-normalisation	197
Figure 5.6. PCA plots containing all samples, pre- and post-normalisation.....	199
Figure 5.7. PCA plots of samples pre-normalisation	200
Figure 5.8. PCA plots of samples post-normalisation.....	201
Figure 5.9. Scatterplots showing up-regulation of X-linked genes and remaining Y-linked genes....	203
Figure 5.10. Heatmaps of log fold change (logFC) values	207
Figure 5.11. Output of STRING protein functional association analysis at medium confidence	214
Figure 6.1. Major enzymatic steps for the maturation of primary transcripts into mature miRNAs ..	229
Figure 6.2. The conserved maturation pathway of miRNAs and mechanisms of action.....	230
Figure 6.3. Changing spatiotemporal distribution of small non-coding RNAs in spermatogenesis ...	232
Figure 6.4. Overview of small RNA sequencing data preparation and analysis pipeline.....	237
Figure 6.5. Reads retrieved from each sample.....	238
Figure 6.6. Boxplots of quality (Q) values distributed across each read of the sample examined	239
Figure 6.7. Mean quality score across all small RNA samples.....	240
Figure 6.8. Mean quality score across all small RNA samples post-trimming.....	241
Figure 6.9. PCA plot of all samples post-normalisation and transformation in DESeq2.....	242
Figure 6.10. Heatmap comparing read composition for similarity	243
Figure 6.11. Simplified overview of miRNA data processing pipeline	246
Figure 6.12. Scatterplots of raw detected miRNA content	247
Figure 6.13. Simple representation of the effect of quantile normalisation across arrays	248
Figure 6.14. PCA plots of all miRNA microarray samples post-normalisation	250

List of Tables

Table 1.1. Ancestral genes	31
Table 1.2. Acquired genes	32
Table 1.3. Summary of MSYq deletion mouse models	54
Table 1.4. Traits found to be associated with the Yq deletion in B10.BR-Ydel mice	56
Table 1.5. Comparison of RNA sequencing and microarray expression analysis platforms	63
Table 2.1. Composition of mixture used for cell counting in a haemocytometer	72
Table 2.2. Categories used to define individual sperm head shapes within sperm populations	74
Table 2.3. Selection of enzyme concentrations used to generate an enzyme activity curve	80
Table 2.4. Composition of lysis buffer 1	83
Table 2.5. Composition of lysis buffer 2	84
Table 3.1. Offspring sex ratio comparison between sires across all offspring produced.....	96
Table 3.2. Standard deviation and error associated with sires in each strain	96
Table 3.3. Observed sperm head morphology by classification	101
Table 3.4. Number of successfully detected sperm heads used in comparative analysis.....	109
Table 3.5. Recorded difference between WT and YQ sperm populations by measured parameters ..	111
Table 3.6. Percentage differences between sub-populations by measured parameters.....	115
Table 3.7. Proportion of sperm populations represented by each clustered dataset.....	118
Table 3.8. Difference in DNA content between individual sperm.....	120
Table 3.9. Percentage difference in sperm head area.....	120
Table 3.10. Values of transmittance as calculated across enzyme-only experiments	128
Table 4.1. Approximation of subcellular fractions retrieved	139
Table 4.2. Explanation of sample names and abbreviations used in RNA sequencing chapter.....	142
Table 4.3. Table of CPM values of genes of known transcript origin (cell type)	161
Table 4.4. Comparison of CPM values in late spermatid and early spermatid populations	162
Table 4.5. Table of genes of known shared or non-shared behaviour	163
Table 4.6. List of protein-coding genes shortlisted by LogFC	170
Table 4.7. Genes shortlisted as potentially cis-limited	171
Table 4.8. Table containing differentially expressed genes in YQ early stage spermatids.....	173
Table 4.9. Table containing differentially expressed genes in YQ late stage spermatids.....	174
Table 4.10. Table containing differentially expressed genes in YQ mature sperm.	176
Table 4.11. Summary of transcript distribution of candidate genes identified from the literature	179
Table 5.1. Abbreviations used to describe RNA sample replicate sets.....	187
Table 5.2. Abbreviations used to describe RNA samples	188
Table 5.3. Cell type attribution per population for microarray cell samples	192
Table 5.4. Differential expression between WT and YQ samples at high (75%) stringency	205

Table 5.5. Differential expression between fractions & within strains at standard (50%) stringency	205
Table 5.6. Table of genes of known shared or non-shared behaviour	206
Table 5.7. Differentially expressed genes by sample comparison	208
Table 5.8. Minimal gene annotation of differentially expressed genes	216
Table 5.9. LogFC values from comparisons of control genes	217
Table 5.10. LogFC values from comparisons of candidate genes	218
Table 5.11. Behaviour of genes of interest identified from the literature	219
Table 5.12. Genes shortlisted in differential expression analysis from both major techniques	225
Table 6.1. Table of sample name abbreviations to be used throughout the chapter	235
Table 6.2. The number of genes showing logFC values of given value	244
Table 6.3. Details of miRNAs shortlisted for subsequent analysis.....	245
Table 6.4. A list of aberrant samples and the chips that they were hybridised to	249
Table 6.5. A list of miRNAs showing significant differential expression	251
Table 6.6. Differentially expressed miRNAs returning experimentally-validated interactions with shortlisted differentially expressed mRNAs	252
Table 6.7. Spermatid-expressed mRNAs & interaction with differentially expressed miRNAs	253

Chapter 1

1. Introduction

This thesis outlines the investigative work undertaken and insights gained in the study of the partial Y-chromosome deletion model, MFXY^{RIII}qdel. The MF1XY^{RIII}qdel model possesses a Y-chromosome long arm of one third the expected length for *Mus musculus* MF1 strain, resulting in a reduced copy number of multicopy genes that occupy the Y-chromosome long arm. Deletions within this region have been shown to result in defects in spermiogenesis and an altered sex ratio in offspring of the affected male. The extent of the sex ratio skew and whether there are associated fertility defects is correlated with the extent of the long arm deletion. It has been shown that the offspring sex ratio skew in the MF1XY^{RIII}qdel model does not come about as a result of reduced Y-bearing sperm production or as a result of male-biased embryo failure during gestation. Therefore, functional inequality between X- and Y-bearing sperm themselves is responsible for the distortion.

A number of spermatid-specific genes and genes important in spermatogenesis exist on the sex chromosomes. The sharing of gene products between developing spermatids means that genetically distinct haploid spermatids can develop into functionally equivalent haploid sperm. In order for functional inequality to exist between spermatids, one or more postmeiotic gene products that are not shared across the cytoplasmic bridges must be involved. In order to investigate the causes and consequences of altered sex chromatin expression in the MF1XY^{RIII}qdel model, several parallel investigations are pursued. These include the characterisation of physiological abnormalities and inequalities in sperm from both wild-type and deletion model mice, and an examination of transcriptional differences in post-meiotic spermatids between the two strains. This extends to an evaluation of subcellular distribution of transcripts, transcript sharing potential and identification of candidate genes for involvement in the resulting offspring sex ratio distortion and observed physiological consequences of Yq deletion.

1.1. Sex Determination and the Meiotic Sexes

Understanding the existence of sexes and their significance is a human cultural preoccupation. The extent and significance of the contribution of both sexes in reproduction is a central question within this. Human understanding of this topic has come some way since the ‘small man’ concept of preformationism in which a fully developed human was thought to reside in miniature form inside a single sperm (*Figure 1.1*). Or Pythagoras and others’ ‘spermist’ view in which males contribute the essential characteristics to the offspring, and females contribute only the ‘substrate’. Anthony Van

Leeuwenhoek's investigation of his own 'residue' revealed the existence of thousands of eel-like 'spermatic animalcules' which he suggested provided the substance to resulting offspring, whilst the female counterpart provided the nourishment. His student, Nicolas Hartsoeker, wrote of 'l'enfant' (the child) or 'le petit animal' (the small animal) when describing the fully formed person or animal curled inside a spermatozoan; cells which Leeuwenhoek referred to as 'animalcules' (Pinto-Correia 2015, Beukeboom and Perrin 2014, Hartsoeker 1694). We now understand the contribution of both the sperm and egg to be vitally important in the development of the individual and to be more complex than a simple 'substance' and 'nourishment' divide.

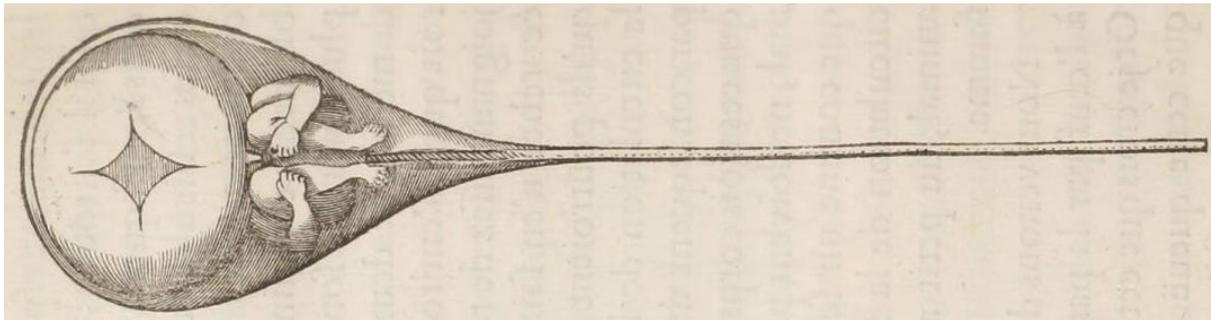


Figure 1.1. Sketch of the predicted 'l'enfant', a fully formed person residing inside the sperm, a reproductive theory supported by the preformationists. Sketch by Nicolas Hartsoeker from page 230 of his *Essay de Dioptrique* (1694). The image was cropped from a digital copy of the original work, provided by Linda Hall Library of Science, Engineering & Technology; LHL Digital (2017).

1.1.1. Sexual reproduction

Sexual reproduction is a form of reproduction involving the fusion of two distinct reproductive cells (gametes) generally borne by two individuals belonging to different sexes within the same species. Sexual reproduction is remarkably widespread amongst living organisms (Maynard-Smith 1978). In mammals, gametes are haploid cells - they contain only half the full complement of DNA required for the formation of a new individual. The fusion of gametes results in the formation of a zygote, a cell containing a full DNA complement following fertilisation of the female haploid cell that is able to develop through subsequent cell divisions into an independent organism.

Meiosis is a process of cell division whereby haploid cells form from full-complement diploid cells. It is essential for the production of sperm and oocytes (the gametes). Meiosis occurs across a broad range of organisms, is highly complex and highly conserved (Halary et al. 2011, Egel and Penny 2007), and is purported to have its origins in early asexual eukaryotes, evolving in part as defence against aberrant recombination (Wilkins and Holliday, 2009). The combination of meiosis and syngamy (fertilisation) is required for sexual reproduction as we know it. The origins of sexual reproduction can be traced back to about two billion years ago, alongside the first multicellular organisms (Zimmer 2009, Malik et al. 2008). The majority of the eukaryota, organisms which possess membrane-bound organelles and the taxon to which we and all other animals belong, reproduce

sexually. There are some exceptions, reoccurrences of asexual reproduction throughout the eukaryote tree (including many plants and fungi), and numerous phyla that are capable of reproducing either sexually (normally under times of stress) or asexually (Lee et al. 2010). Goodenough and Heitman (2014) state, 'That the LECA [last eukaryotic common ancestor] was sexual is no longer a matter of speculation/debate, as evidence of sex, and of genes exclusively involved in meiosis has been found in all the major eukaryotic radiations', citing several papers which confirm the existence of these genes in a multitude of diverse phyla.

The major benefits of sexual reproduction are thought to come from the ability to avoid the rapid deterioration of one genetic line of a population following the accumulation of deleterious mutations or vulnerabilities exposed by limited genetic diversity (Hollister et al. 2015, Henry et al. 2012, Muller 1932). However, sexual reproduction has a high immediate cost when compared to asexual reproduction or parthenogenesis (Gibson et al. 2017, Maynard-Smith 1971). The need to find, attract and mate with another individual alongside the potential disruption of adaptive genotypes and the time taken to achieve successful meiosis versus mitosis (Lehtonen et al. 2012), are amongst some of the many costs imposed by sexual reproduction as an adaptation. Sexual reproduction also induces a number of forms of competition: between the sexes in the passage of genetic information to the next generation; between individuals of the same sex in the form of mate competition, and between male gametes in the race to fertilise the oocyte (Lehtonen and Parker 2014, Parker 2014, Parker 1970). This competition has shaped and is shaped by many aspects of sex chromosome evolution. The chromosomes that, amongst other things, determine to which sex the offspring belong and influence many aspects of sexual reproduction and sexual competition.

1.2. The Sex Chromosomes

Mice, like most other mammals, only reproduce sexually and possess an XX/XY chromosome sex determining system. The sex chromosomes are the highly differentiated non-pairing chromosomes that determine the biological sex of the individual carrying them. In mammals, males are the heterogametic sex (XY) and females the homogametic sex (XX).

1.2.1. Evolution of the sex chromosomes

All sex chromosomes originated as an autosomal pair, gaining, at some point in their history, a major sex-determining factor (Beukeboom and Perrin 2014). The sex-determining locus in mammals is found on the Y chromosome, having first appeared in a mammalian ancestor common to placental mammals, marsupials and monotremes around 170 million years ago (Graves 2016, Beukeboom and Perrin 2014, Sinclair et al. 1990). The *Sry* (sex determining region of the Y) gene derived from an autosomal ancestor of X-linked *Sox3* (Sutton et al. 2011, Foster and Graves 1994) (see *Figure 1.2*)

and has a fundamental role in sex determination across all placental mammals barring a few exceptions in the rodent family that have secondarily lost the Y chromosome. It alone is both essential and sufficient to bring about male sex organ development in mice, as determined through induced *Sry* expression in XX mice (Koopman et al. 1991).

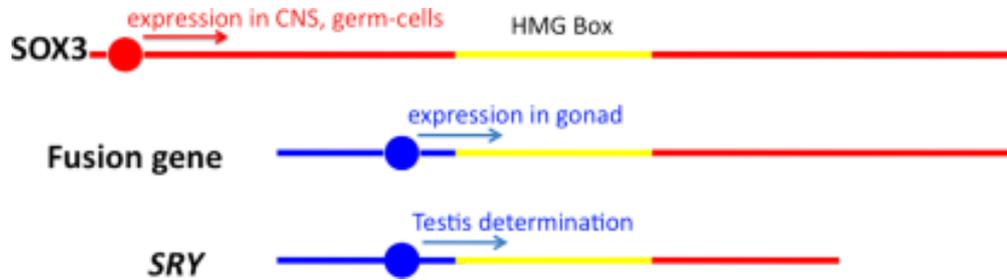


Figure 1.2. Fusion of autosomal *Sox3* with a gene accompanied by a promoter driving expression in the genital ridge. HMG = high mobility gene. Image taken from Graves (2013).

The introduction of a sex determining factor is thought to be the initiating factor in the evolution of the sex chromosomes as we know them. The induction of a sex-determining locus to one chromosome of a pair results in the rapid differentiation of the chromosome from its partner as the sex determining locus accumulates genes advantageous to males only. The suppression of recombination between the two chromosomes follows due to the accumulation of large-scale inversions on the Y, preventing disruption of the gene cluster that is advantageously inherited together. Recombination suppression expands as the proto-Y chromosome accumulates further sexually antagonistic genes, and undergoes sequence amplification of male-benefit genes. As the suppression of recombination expands, the chromosomes differentiate further, resulting in the accumulation of deleterious mutations and the decay of ancestral genes on the male-specific region of the Y (everything excluding the still recombining pseudo-autosomal region (PAR)). This results in purging of decayed Y-linked genes and dosage compensation of the X-chromosome to compensate for the unequal expression of X-linked genes between the sexes (XX versus XY). Dosage compensation is achieved through X-inactivation (of one X chromosome) in XX mammals. An overview of sex chromosome evolution is given in *Figure 1.3*. In mammals, the Y chromosome is often diminished in size compared to the much larger X chromosome. This reflects a differential loss of ancestral autosomal genes in chromosomes under differing selective pressures. However, the extent of Y chromosome degeneration and male-benefit sequence amplification varies across taxa, with even closely related mammals possessing Y chromosomes of dramatically different sizes. The euchromatic region of the male-specific region of the Y (MSY) in humans is 23Mb long, whereas in mouse this totals 89.5Mb, 98% of which is ampliconic (Soh et al. 2014).

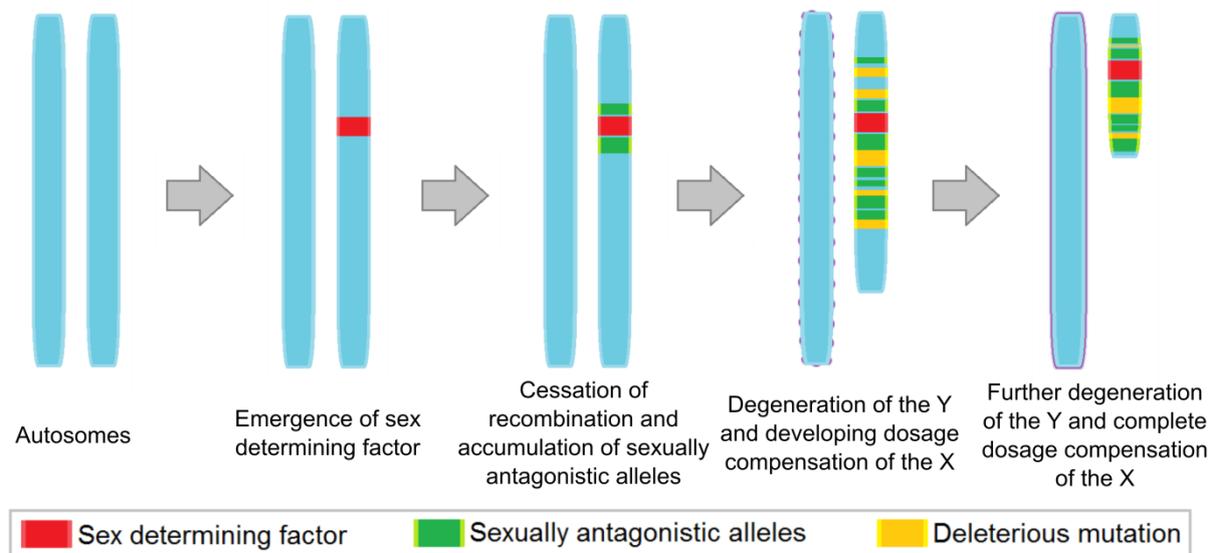


Figure 1.3. Overview of the evolution of sex chromosomes, as demonstrated in a male heterogametic system. Figure inspired by Abbott et al. (2017), and Wei and Barbash (2015).

Gonosomal male-specific gene amplification

The sex chromosomes are rich in genes involved in spermatogenesis. The lack of recombination across most of the Y chromosome leads to the gradual accumulation and fixation of male benefit genes. In addition to this, Turner (2007) and others have noted that the effects of any X-linked gene providing benefits to male individuals, whether dominant or not, will immediately appear in the male line due to the presence of a single X chromosome (the gene is hemizygous) and a lack of directly competing alleles (Turner 2007). This suggests, and is supported by evidence, that genes, especially recessive genes, beneficial only to males (spermatogenesis genes) will disproportionately accumulate on the X chromosome (Turner 2007, Khil and Camerini-Otero 2005, Lercher et al. 2003, Rice 1984, Fisher 1931). Sex chromosomes play a crucial role in the development of sexually dimorphic traits.

Insects have a diverse range of sex determination systems, but a sex-biased gene accumulation effect can be seen in some species with an XY sex-determining system. Species such as the stalk-eyed flies (Baker et al. 2016) and the model Pea aphid, *Acyrtosiphon pisum* (Jaquiéry et al. 2013), have a highly masculinised X chromosome enriched for testis-specific genes. In contrast, a paucity of male-specific genes has been noted in several members of the Hemiptera family of true bugs, some species of *Drosophila*, and others (Parisi et al. 2003, Pal and Vicoso 2015) where selection pressures differ. Though divergent from mammalian systems, this diversity demonstrates the plasticity of the sex chromosomes and highlights their importance and variability as carriers of sex-specific genes highly influenced by sexual selection and the inherent conflict of inheritance in organisms possessing differentiated sex chromosomes.

1.2.2. Anatomy of the mouse Y chromosome

The mouse Y chromosome can be divided into three distinctive regions: the short arm (Yp), the long arm (Yq) and the pseudo-autosomal region (PAR). All regions are known to contribute to successful reproduction. Unlike all other chromosomes of the mouse genome, the Y chromosome is considered acrocentric, possessing a small arm (Yp), a noticeable diversion from the otherwise telocentric collection of chromosomes of the mouse genome (Pertile et al. 2009). This small arm contains the two genes, *Sry* and *Eif2s3y*, essential for ‘maleness’ and spermatogenic proliferation, respectively (Yamauchi et al. 2016, Mazeyrat et al. 2001), and several other genes implicated in spermatogenesis (see Figure 1.4).

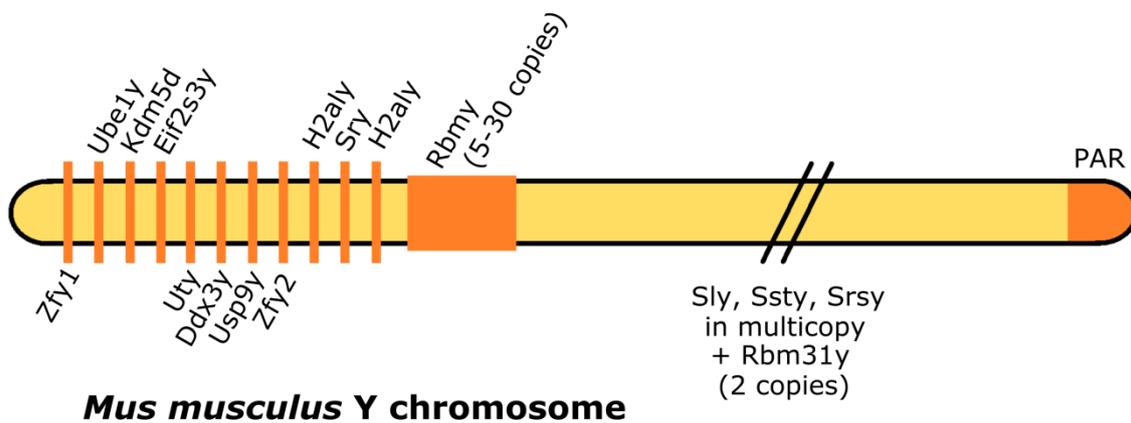


Figure 1.4. Simplified representation of the mouse (*Mus musculus*) Y chromosome. Bottom/top placement of gene names on the Y short-arm (Yp) is indicative of presence on the positive or negative strand, respectively. PAR = pseudo-autosomal region.

Given that *Sry* alone confers ‘maleness’, karyotypically female (XX or XO) animals bearing a transposed copy of *Sry* on the X chromosome exhibit sex reversal. In $X^{Srx^a}O$ males, almost the entire content of the Yp is transposed onto the distal region of an unpartnered X chromosome; the transposed material is termed the sex reversed region (Sxr^a). In 1992, Burgoyne et al. (1992) demonstrated that XO male mice that possess the Sxr^a attached to the normal X chromosome are phenotypically male but undergo spermatogenic arrest at the meiotic metaphase stages due to the lack of a binding partner for the X chromosome and consequent lack of synapsis at the first meiotic metaphase. However, by providing a diminished partner chromosome composed of most of the Y-chromosome PAR and fused to an X chromosome centromere (termed the Y^{*X}), they were able to restore full meiotic progression to the round spermatid stage (Burgoyne et al. 1992). These $XSrxY^{*X}$ males were therefore able to successfully carry out meiotic divisions, but were recorded as being sterile and exhibiting major sperm morphology consequences due to the absence of remaining Y-

linked genes. This demonstrated that the material on the Yq is essential for normal fertility and, in particular, for sperm morphology.

Study of the Y chromosome long arm by traditional methods is difficult due to the presence of genes such as *Sly* and *Ssty* in massive multicopy along the arm, with both *Sly* and *Ssty* being present in more than 80 copies. The difficulties in sequencing highly repetitive regions have meant that it was only in 2014 that Soh et al. published the complete sequence of the C57BL/6J mouse Y-chromosome in Nature (Soh et al. 2014), concluding an endeavour that had begun twelve years previous.

The male specific region of the Y chromosome long arm (MSYq)

Unlike in many other mammals, the mouse Y chromosome has a massively amplified long arm. The long arm of the Y-chromosome, excluding the pseudo-autosomal region (PAR) is referred to as the ‘male specific region of the Y chromosome long arm’ or ‘MSYq’, and is known to contain four massively amplified genes (*Sly*, *Srsy*, *Ssty1*, *Ssty2*) and two copies of *Rbm31y*. Contrary to previous assumptions, the male specific region of the Y chromosome (MSY) in mice contains a high proportion of actively transcribed genes, it is 99.9% euchromatic. Unlike the X chromosome, little of the autosomal ancestry of the Y chromosome remains (around 2%) (Soh et al. 2014). A recent comprehensive build of the ampliconic arm has provided highly detailed information on the structure of the MSY (Soh et al. 2014). The copy number of genes within the repetitive region of the Y long arm is highly variable, whilst variability in non-repetitive regions of the MSY is low even compared to autosomes (Morgan and Villena 2017, Case et al. 2015).

In 1969, Krzanowska showed that genes affecting sperm morphology were located on the Y-chromosome through the cross-breeding of inbred mouse strains (Krzanowska 1969). This was confirmed in subsequent experiments including examinations of $X^{Sxra}Y^{*X}$ and MSYq deletion mice (section 1.5). In 2009, it was shown that *Sly* and the *Ssty* genes on the Yq are essential contributors to a spermatid specific genomic conflict between the X and Y chromosomes (see section 1.4.4). The Yq contains three major ampliconic genes; these are described below. An overview of the composition of the MSY can be seen in both *Figure 1.4*, and in *Table 1.1* and *Table 1.2*, where ancestral and acquired genes are reported.

***Sly* – Sycp3-like Y-linked**

Sly is known to be the major Y-linked contributor to the post-meiotic intragenomic conflict between the sex chromosomes (see section 1.4.4) and, much like other Yq genes, it exists in multiple copies. Like its homologue, *Slx*, SLY protein contains a single functional Cor1 domain which is thought to mediate interaction with the sex chromatin (Comptour et al. 2014, Ellis et al. 2005). *Sly* likely originated following chimerism between the 5’ region of *Slx* and the 3’ region of *Xlr* (X-linked

lymphocyte regulated), with duplication of *Slx* exons 3-4 becoming exons 5-6 of *Sly* (Ellis et al. 2007). *Sly* is first transcribed in secondary spermatocytes and transcription continues in the round spermatid stage (Reynard et al. 2009). Major phenotypic disruptions in MSYq deletion mice have been linked primarily to downstream consequences of *Sly* depletion (Case et al. 2015).

***Ssty1/2* – Spermiogenesis specific transcript Y-linked 1/2**

It seems likely that the *Ssty* gene family, comprising *Ssty1* and *Ssty2*, originated from the autosomal gene *Spin* as a retroposed transcript (Touré et al. 2004a, Touré et al. 2004b, Oh et al. 1998). The two subfamilies share 84% similarity at the nucleotide level and, much like *Sly*, *Ssty* are transcribed specifically in testis. The *Ssty* family have also been implicated as a secondary mediator in the intragenomic conflict, potentially acting as a transporter or shepherd protein to SLX and SLY (Comptour et al. 2014).

***Srsy* – Serine-rich secreted Y-linked**

Another multicopy gene, now known as *Srsy*, is poorly understood and has not been shown conclusively to produce a protein product. It shows sequence similarity to genes like those found in the human MAGEA family (Morgan and Villena 2017).

These three gene families are contained within the ‘huge repeat array’ of the Y long arm. All genes within the array have functions in chromatin condensation during sperm maturation and are expressed almost exclusively in post-meiotic spermatids (Morgan and Villena 2017).

Y gene	Location	Copy number	X-homologue	Copy number
<i>Zfy1, Zfy2</i>	p	2	<i>Zfx</i>	1
<i>Ube1y1</i>	p	1	<i>Ube1x</i>	1
<i>Kdm5d</i>	p	1	<i>Kdm5c</i>	1
<i>Eif2s3y</i>	p	1	<i>Eif2s3x</i>	1
<i>Uty</i>	p	1	<i>Utx</i>	1
<i>Ddx3y</i>	p	1	<i>Ddx3x</i>	1
<i>Usp9y</i>	p	1	<i>Usp9x</i>	1
<i>Sry</i>	p	1	<i>Sox3</i>	1
<i>Rbmy</i>	p	30 ^a	<i>Rbmx</i>	1

Table 1.2. Acquired genes				
Y gene	Location	Copy number	X-homologue	Copy number
<i>H2a12y</i>	p	2	<i>H2a11</i>	14 ^a
<i>Rbm31y</i>	q	2	<i>Rbm31x</i>	1
<i>Sly</i>	q	126	<i>Slx</i>	25 ^a
			<i>Slx11</i>	14 ^a
<i>Ssty1</i>	q	85	<i>Sstx</i>	11 ^a
<i>Ssty2</i>	q	221		
<i>Srsy</i>	q	197	<i>Srsx</i>	14 ^a
<i>Prssly</i>	p	1 ^a		
<i>Teyorf1</i>	p	1 ^a		

Table 1.1 and Table 1.2. List of ancestral (top) and acquired (bottom) genes present on the Y, described alongside their location on the Y-chromosome, copy number and X homologue where appropriate. Location on the Y-chromosome is stated as either ‘P’ – short arm or ‘Q’ – long arm. ^a estimated (Soh et al. 2014). This table has been adapted from Soh et al. (2014).

The Y chromosome pseudo-autosomal region (PAR)

As in the X chromosome, the Y chromosome pseudo-autosomal region (PAR) is a distinct segment of the Y chromosome required for essential pairing and recombination between the X and Y chromosomes. It is a short (700Kb) region of homology between the sex chromosomes (Perry et al. 2001), without which spermatogenesis is seen to arrest at meiotic metaphase (Burgoyne et al. 1992). Genetic loci within the PAR are capable of exchange between the two chromosomes, unlike the rest of the X and Y. And, due to the presence of two copies of PAR genes during meiosis, all genes thus far identified in this region are known to escape X-inactivation (Perry et al. 2001).

1.2.3. The X chromosome

The mouse X chromosome, like many mammalian X chromosomes, is generally highly conserved and is enriched for essential brain development and reproductive genes. However, the generally constrained X contains regions of highly specialised gene content and rapid evolution that are evident across the chromosome. The regulatory constraints imposed by essential epigenetic X-chromosome dosage compensation mechanisms that result in X-chromosome inactivation require strict conservation of X-linked genes (Larson et al. 2016). However, as previously stated, it is expected that male-specific genes located on the X undergo more rapid evolution due to the hemizygous nature of males because selection will act more efficiently.

In 2008, Mueller et al. showed that, contrary to standard thought, the mouse X chromosome was enriched for multicopy genes showing specific post-meiotic expression in testis; with around 18% of all X chromosome genes expressed in postmeiotic spermatids. This gave strong evidence for the incomplete continuation of meiotic X-chromosome repression post meiosis, and unveiled possibilities of the interaction of the X and Y chromosomes after meiosis. These ampliconic gene families include *Slx*, *Gmcll1*, *Ott*, *Srsx* and many more. They determined around 12% of the X chromosome to be ampliconic and exposed a number of multicopy genes in non-ampliconic regions that also showed testis-biased expression (Mueller et al. 2008). *Slx* and *Slx11*, *Sstx*, and *Srsx* are genes that exist within ampliconic regions and are known to be X-linked homologues of *Sly*, *Ssty1* and *Ssty2*, and *Srsy*. Like their homologues, they share sequence similarity with autosomal *Sycp3*, *Spin* and melanoma/testis antigens respectively (Morgan and Villena 2017, Ellis et al. 2011a, Ellis et al. 2007, Oh et al. 1998). *Slx* and *Slx11* have been shown to be involved in the intragenomic conflict subsequently discovered in mice; acting antagonistically to their Y-linked homologue *Sly* (Cocquet et al. 2012).

1.3.Spermatogenesis, Sperm Structure and Function

Gametes are highly specialised haploid cells. Male and female gametes have developed under distinct evolutionary pressures and are highly differentiated from each other and from somatic cells. Male gametes (sperm) are extremely morphologically diverse, and their development has produced diverse solutions to overcome a range of selective pressures.

Mammalian sperm are highly motile, relatively short-lived, and competitive haploid cells produced by the male individuals of a species. Their primary purpose is to find and fertilise the oocyte (the female gamete), a large sedentary cell possessing multiple thick surrounding layers to its nuclear core. Sperm are a common feature across sexual species, not all are referred to as sperm and not all are motile (Birkhead et al. 2008), but they perform much the same function – delivering a consignment of DNA to the female gamete to begin the process of embryogenesis. In mammals, which (for the most part) utilise a XY sex determination system in which the male is the heterogametic sex (XY), the sperm is the sex-determining cell during reproduction. Rodent spermatozoa exhibit a wide range of sperm sizes and head shapes across species (Varea Sánchez et al. 2013), but many functional features of the sperm head and tail are conserved across mammalian species (Jonge and Barratt 2006). *Figure 1.5* shows the comparative morphology of mature mouse and human spermatozoa.

Mouse sperm Human sperm

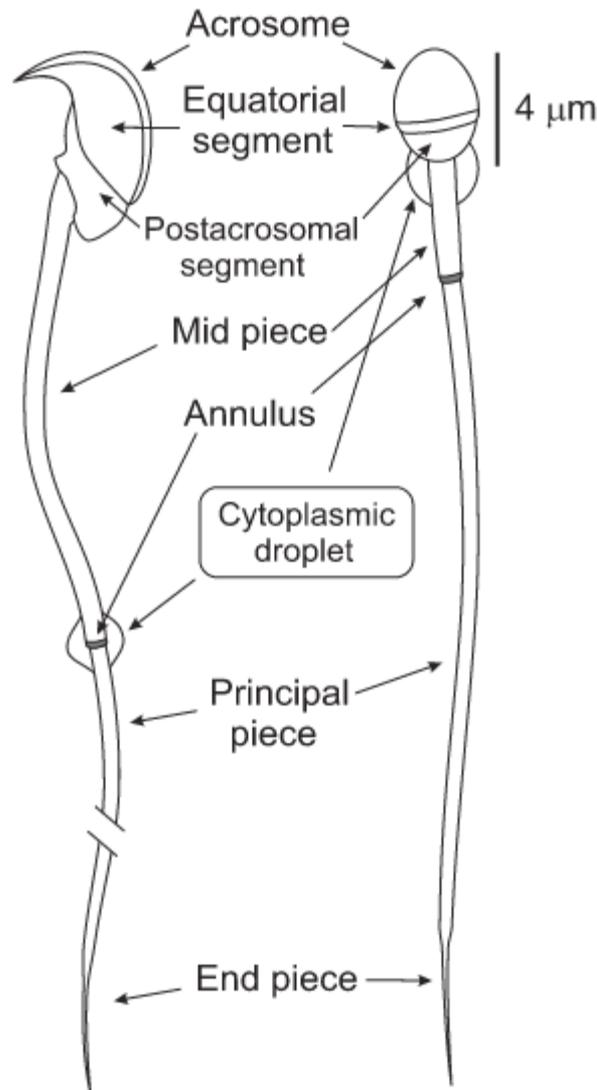


Figure 1.5. Comparative morphology of mouse and human sperm. Mouse sperm possess a longer mid-piece and tail region (shortened in diagram) - human sperm are around 55μm in length whilst mouse sperm are around 120μm long. All major regions of sperm are shared characteristics between the two and many other mammalian species. Image from Darszon et al. (2011).

Spermatogenesis is the process by which male gametes develop from spermatogonial stem cells. Spermatogonia undergo meiotic divisions and recombination resulting in reduction of the genome to half its former complement, becoming a haploid cell that contains only one copy of each chromosome and one of either the X or the Y sex chromosomes. In mouse, spermatogenesis takes place in the testes from 14 days post-partum. A full cycle, from stem cell to mature sperm, proceeds over 35 days, moving through mitotic, meiotic and post-meiotic phases (Eddy 2002). As sperm cells develop, they progress from the exterior edge of the testicular tubule (the basal compartment) to the interior lumen

(see *Figure 1.6*), where meiotic and post-meiotic phases of development take place. Post-meiotic spermatids originating from the same stem cell remain connected by a network of cytoplasmic bridges until late in the elongating/condensing stages of spermatid maturation (Greenbaum et al. 2011). When a testicular cross-section is taken, multiple stages of spermatogenesis can be seen in one tubule, this is due to the accumulation of overlapping waves of spermatogenesis. Mature sperm present in the lumen of the testicular tubules progress along the efferent ducts and through subsequent regions of the epididymis toward the vas deferens; *Figure 1.7* shows the simplified anatomy of the murine testis.

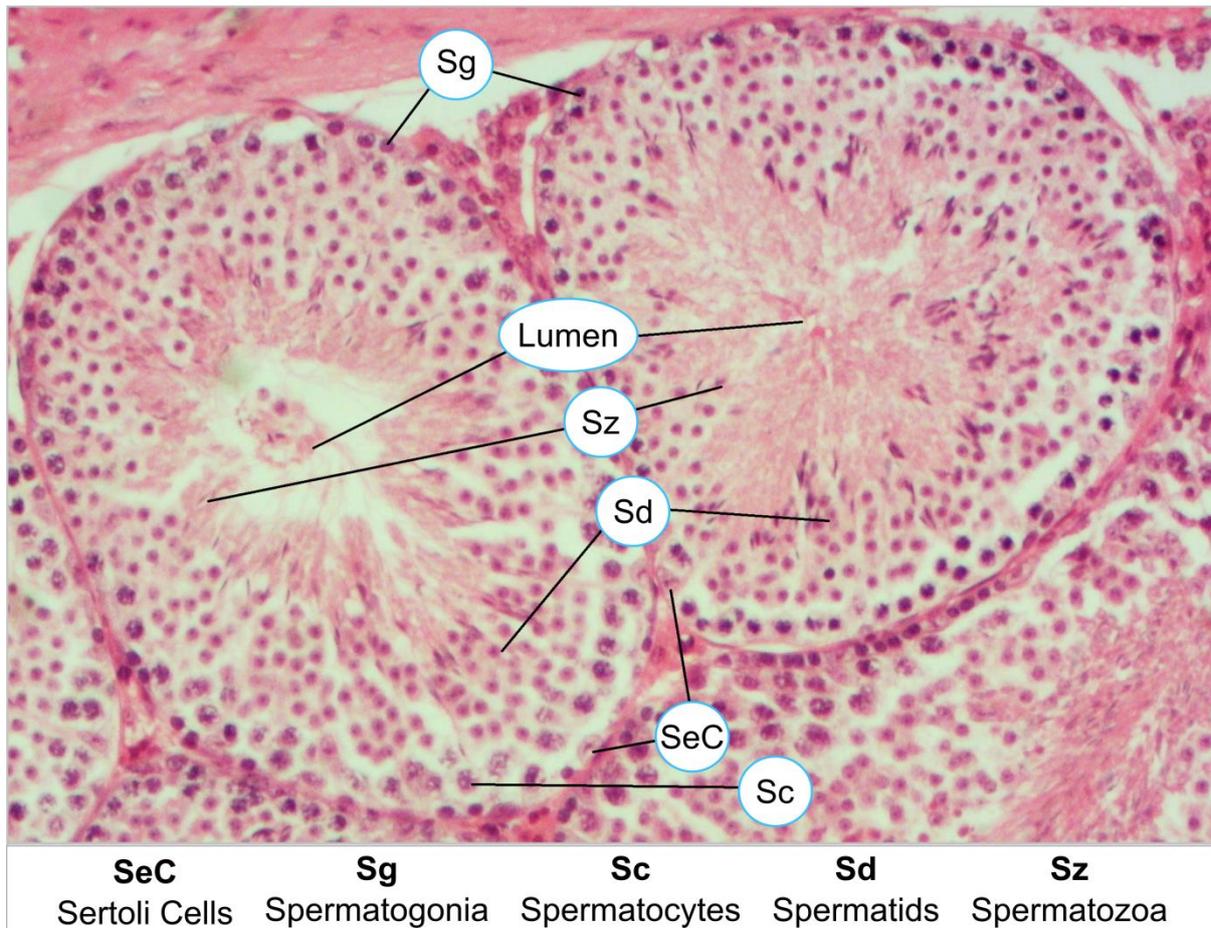


Figure 1.6. Testis section from a MF1XY^{R111} mouse, stained with haematoxylin and eosin (HE stain). Haematoxylin stains nucleic acids blue, and eosin stains proteinaceous material pink. Cropped image of two tubules at differing stages. Testicular cells are labelled as per the key at the base of the image. As sperm proceed through spermatogenesis from spermatogonial stem cell through to mature spermatozoa, they travel towards the inner region of the tubule, the lumen.

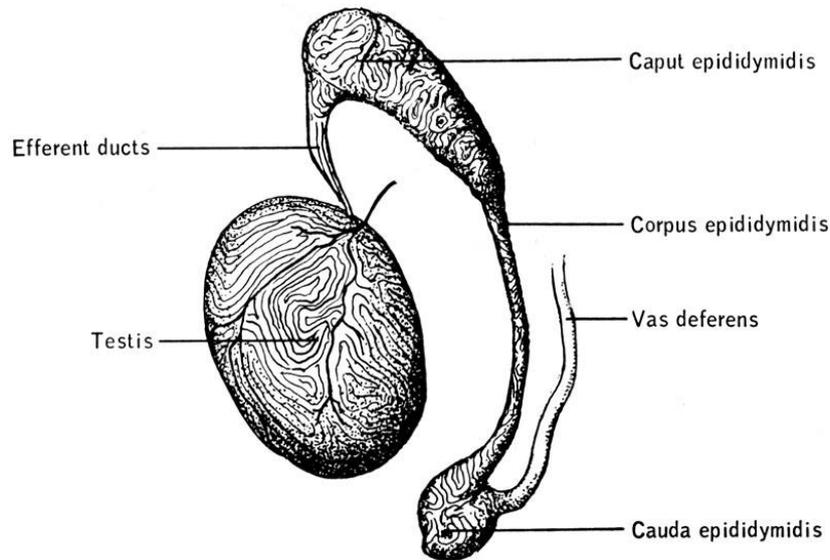


Figure 1.7. Diagram of the murine testis. The epididymis links the efferent ducts to the vas deferens and consists of the caput (head), corpus (body) and cauda (tail). Inside lies the epididymal duct. Mature sperm released into the lumen travel along the duct towards the vas deferens. Image: *Biology of the Laboratory Mouse*, by The Staff of The Jackson Laboratory, Earl. L.Green (editor), Chapter 13 – anatomy, Figure 13-46. ‘Drawing of the testis, efferent ducts, epididymus, and vas deferens (x 5.4)’¹

Spermatogenesis can be broadly split into two parts: spermatocytogenesis and spermiogenesis. Initial clonal expansion of spermatogonia and subsequent meiotic divisions occur within spermatocytogenesis; whereas the complex maturation of sperm cells from haploid round spermatids is termed spermiogenesis. This section focuses primarily on spermiogenesis but a brief description of spermatogonial proliferation and formation of haploid cells is given below. An overview of cellular progression during spermatogenesis can be seen in *Figure 1.8*.

Spermatocytogenesis begins with undifferentiated spermatogonial stem cells and ends with post-meiotic haploid spermatids

In reproductively mature male mice, testicular stem cell activity resides with a population of undifferentiated spermatogonia. This base population self-renews through mitotic proliferation alongside generating populations of differentiating spermatogonia. All spermatogonia are localised to the basement membrane of the testis with subsequently produced cell types progressing with maturation towards the lumen of the tubules (Yoshida et al.2006). Differentiating spermatogonia form primary spermatocytes through mitosis; primary spermatocytes are diploid cells that can undergo meiosis, first producing two haploid secondary spermatocytes per one primary spermatocyte following meiosis I and finally four haploid spermatids as a result of meiosis II. These cells are known as round spermatids.

¹ © 1968 The Jackson Laboratory – used with permission

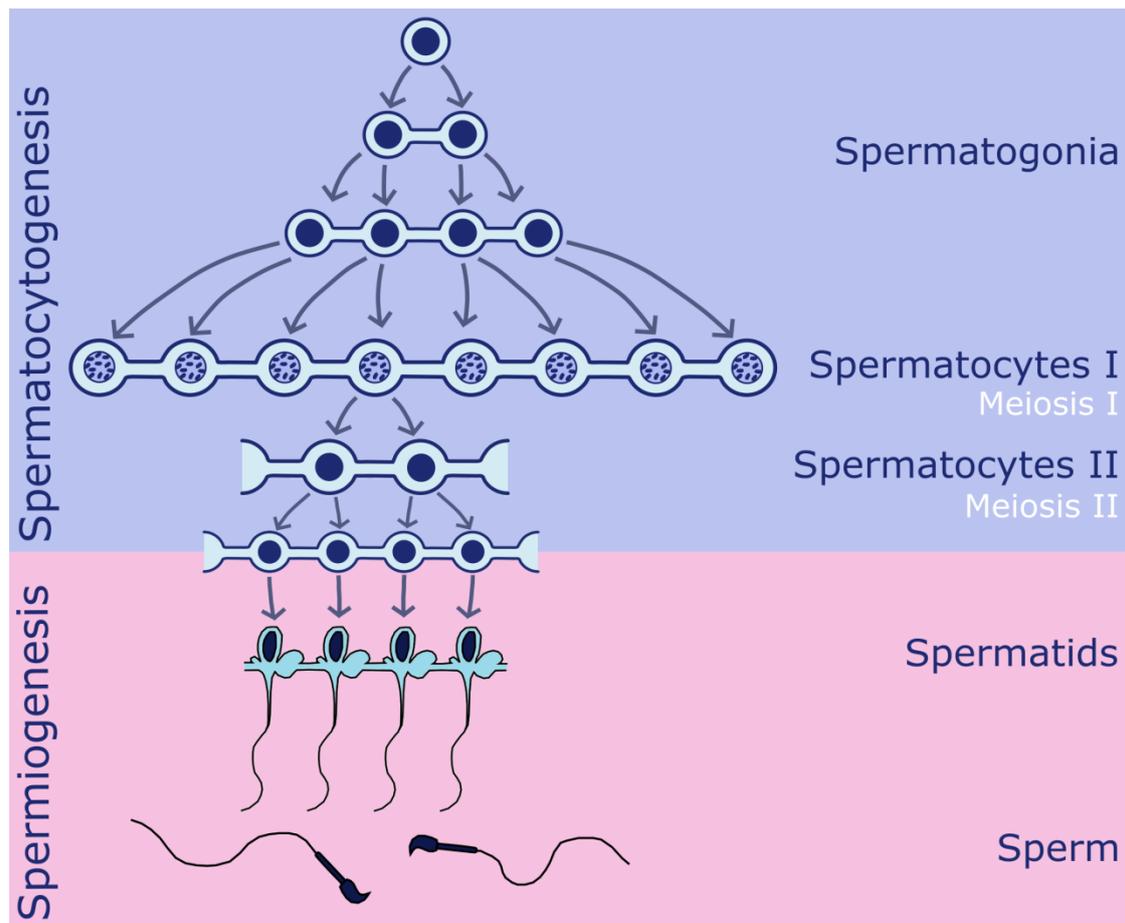


Figure 1.8. Overview of cellular progression during mammalian spermatogenesis. Spermatogenesis can be broadly divided into stages; spermatocytogenesis and subsequent spermioogenesis. Inspired by image from www.embryology.ch.

Extensive reshaping of the sperm head and shedding of cytoplasm during spermioogenesis

Following meiosis, cells progress from haploid round spermatids to highly condensed, flagellate and dorso-ventrally flattened mature spermatozoa is an extensive yet intricate process known as spermioogenesis. For many years, this process has been divided into descriptive stages, defined by Eugene Oakberg in 1956 (Oakberg 1956). Mature sperm possesses a number of distinctive and highly specialised features; formation of which takes place during the period of spermioogenesis. Round spermatids are the entry cell type; they are the outcome of male meiosis and are haploid, yet incapable of fertilisation *in vivo*. Not long after their formation, an acrosomal granule forms in each round cell and extends along one pole of the nucleus. Numerous pro-acrosomal granules form from trans-Golgi stacks, and later fuse together to form a single large acrosome that associates with the nuclear envelope (Yao et al. 2002, Abou-Haila and Tulsiani 2000). Concurrent to the start of acrosomal biogenesis is the movement of the nucleus towards the plasma membrane and an increase in the presence of microtubules surrounding it; prior to the formation of the manchette (Sperry 2012). The manchette is composed of self-assembling microtubules made up of an alpha-beta tubulin ($\alpha\beta$ tubulin)

heterodimer to which microtubule-associated proteins and motor proteins can bind (Mochida et al. 1999). It is expected to play a key part in the reshaping and streamlining of the nucleus, and redistribution of cytoplasmic contents of the cell. The manchette is a temporary structure, which undergoes complete disassembly prior to full sperm maturation. The axoneme is another tubulin-containing structure that develops consecutively with the manchette during spermiogenesis, but unlike the manchette it is not transient and plays an important part in sperm motility. The axoneme is the inner cytoskeletal core of the sperm flagella, the mechanism of propulsion in the spermatozoan, and acts as a scaffold constructed of nine outer microtubule pairs arranged in a ring and one inner pair in the centre (Lodish et al. 2000). Tubulins that are the primary component of both the manchette and the axoneme undergo extensive post-translational modifications, likely to affect microtubule function (Kierszenbaum 2002, Mochida et al. 1999). During the final stages of spermiogenesis, the majority of remaining cellular cytoplasm is depleted and forms a cytoplasmic drop which is subsequently discarded from the maturing cell and phagocytosed by the Sertoli cells (Cooper 2011, Huszar et al. 1998, Merton 1940). Throughout these late stages, the structure of sperm chromatin changes considerably, undergoing significant condensation and deactivation through the binding of specific nuclear proteins. Transition proteins (TP1 and TP2) first replace histones at the end of the elongating spermatid stage in the initiation of condensation (Meistrich 1989). Finally, a more extensive change in DNA condensation occurs when these transition proteins are replaced by protamines (P1 and P2); arginine-rich proteins that bind DNA in a non-specific manner to form toroidal coil subunits containing around 50kb of DNA (Brewer et al. 2002, Brewer et al. 1999).

As spermatozoa migrate along the epididymis from the testis towards the vas deferens, they undergo a maturation process that endows them with both motility and fertilisation capabilities. Through a series of genetically mediated time-dependent cellular modifications, spermatozoa gain the ability to move, navigate through the reproductive tract, undergo capacitation and bind to the oocyte (Dacheux and Dacheux 2014). Discrete segments of the epididymis contribute towards maturation of the spermatozoa in distinct and varied ways. This occurs through segment-specific expression of secretory and cellular proteins (Cornwall and Horsten 2007, Jonge and Barratt 2006). Although lacking both motility and the ability to fertilise an egg, it was previously thought that spermatozoa exiting the testis into the epididymis, across most mammalian species, were morphologically complete. However, novel structures are still being identified. Investigations by Lin et al. (2013) have demonstrated the acquisition of a 'hook rim (HR)' onto the apical hook of mouse sperm during transition through the proximal two-thirds of the caput epididymis.

The sperm acrosome

The mouse sperm acrosome is a thin, cap-like structure that covers a substantial proportion of the anterior part of the sperm head and contains digestive enzymes such as various hyaluronidases and

acrosin (Abou-Haila and Tulsani 2000). These enzymes function to diffuse the thick cumulus layer surrounding the oocyte and allow penetration of the zona pellucida for the purposes of fertilisation. The acrosome of sperm forms post-meiosis, during late spermiogenesis (Yao et al. 2002, Abou-Haila and Tulsiani 2000). Any deformation of the region may impede the acrosomal reaction, the fusion of the outer acrosomal membrane with the sperm plasma membrane, required for fertilisation, or the binding of the reacted acrosomal region with the oocyte zona pellucida.

The Oocyte

Mature oocytes are surrounded by an inner vitelline layer and the zona pellucida (a ‘thick vitelline layer’) that is further surrounded by cumulus cells embedded in a non-cellular matrix of hyaluronic acid polymers, proteins and proteoglycans (the cumulus oophorus), see *Figure 1.9*. Together, these form the cumulus oocyte complex (Yanagimachi 2011). In order to achieve successful fertilisation, the sperm must penetrate the cumulus oophorus, bind with and penetrate the zona pellucida, and finally fuse with the oocyte vitelline layer. This fusion allows for the combination of sperm-oocyte DNA to form a diploid cell.

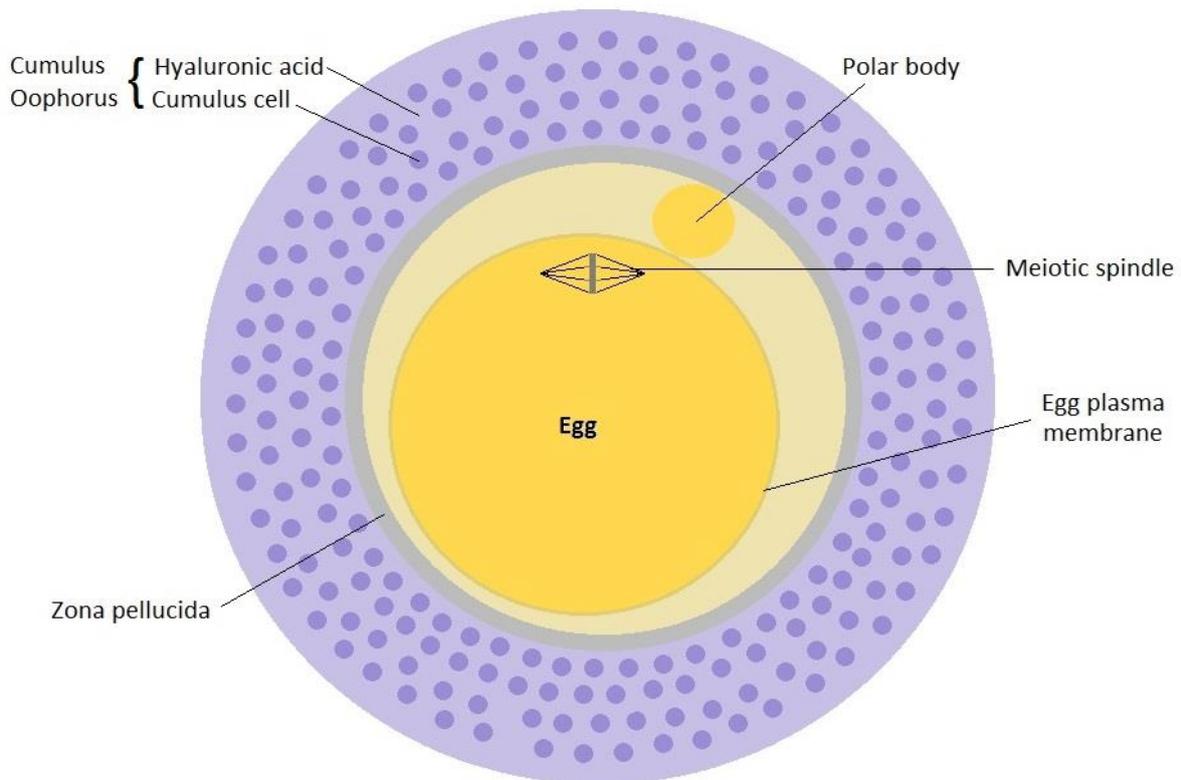


Figure 1.9. Structure of the mouse oocyte. Cumulus cells are embedded in a hyaluronic acid matrix which surrounds the zona pellucida.

1.3.1. Sperm head morphology and indicators of sperm fitness

Given the complex nature of sperm structure, formation and the processes required for a sperm to become fully capable of fertilisation, it is not surprising that a number of measures of sperm fitness exist relating to several important sperm features. Here, some important aspects of sperm fitness are discussed.

Sperm head morphology

Sperm morphology is an important indicator of sperm health and fertility and is frequently used as a key indicator for male fertility in clinical settings (see The WHO laboratory manual for the examination and processing of human semen. 2010). In general, sperm morphology assessment in scientific studies is carried out using qualitative methods such as often simply categorising individual sperm heads as ‘normal’ or ‘abnormal’ (Palmer et al. 2012, Styrna et al. 2002, Fabricant and Parkening 1982). Some efforts have been made to produce more objective and widely-applicable methods for discerning finer level differences in sperm head morphology (Amann and Waberski 2014, Varea-Sánchez et al. 2013), although often these do not provide much information beyond distance measurements.

Many methods focus on the examination of human spermatozoa for assessing either individual sperm or sperm populations for suitability of use in fertility treatments, such as *in vitro* fertilisation (IVF) and more especially intracytoplasmic sperm injection (ICSI). Techniques that allow for the ultra-high resolution (x6600 to x13000 magnification) examination of individual motile spermatozoa and subsequent injection of selected sperm have attracted studies that reported improved pregnancy rates following application of the technique (Setti et al. 2013, Bartoov et al. 2002).

Clinical attribution of human sperm quality value relies on determining the percentage of individual sperm within a single population (ejaculate) that meet a predefined morphological ‘norm’. For humans, the lower limit value for sperm quality is 4% of the sperm population being considered morphologically normal, revised down from 14% in 2010 (The WHO laboratory manual for the examination and processing of human semen. 2010). Morphological normalcy has been defined based on studies of sperm retrieved from the female reproductive tract and the surface of the zona pellucida, and, used in combination with other parameters, is a useful indicator of reproductive potential (Menkveld et al. 2011, Menkveld 2010, Guzick et al. 2001, Huszar and Vigue 1993).

As evolutionary forces have shaped sperm into competitive vehicles for the delivery of genetic information to the oocyte, sperm competition has effected change in sperm morphology. The shape of sperm varies greatly, even between closely related species (see *Figure 1.10*). This demonstrates a phenotypic flexibility, shaped by differing competitive pressures. Mice (and many murines) produce sperm with a noticeably falciform sperm head. Although this observation has been well documented

and broadly studied, the purpose of the hook is still uncertain, though there are a number of credible suggestions.

Sperm head shape diversity

Whilst a majority of known mammalian sperm head shapes do not differ dramatically from the standard ‘tadpole’ or spatulate shape (*Figure 1.10*, row a), there are many notable exceptions and sperm morphology across mammals varies remarkably. Much variability can be accounted for by post-copulatory sexual selection and differences in mating systems, especially dependent upon levels of promiscuity and therefore direct inter-male sperm competition within the female reproductive tract. A lack of inter-male competition, such as in the naked mole rat where a single dominant male-female pair reproduce, has led to poor fertilisation rates and an inconsistent morphology even within the same ejaculate (van der Horst 2011). *Figure 1.10* from Skinner and Johnson (2017) demonstrates some of the sperm head shape diversity seen within metazoa.

Mus musculus have a distinct sperm morphology shared by some, but not all, members of the family rodentia. There are a number of suggestions as to the purpose of the distinct falciform / hook shape sperm head in fertilisation, such as the capacity for sperm to ‘cling’ to the oviduct epithelium (Firman and Simmons 2009). Amongst the better supported theories is the improved ability of spermatozoa to form ‘sperm trains’ and improve sperm swimming performance (Fisher et al. 2014, Immler et al. 2007). This provides competitive benefits to closely related spermatozoa which is especially important in promiscuous species. The sperm of species such as the deer mouse (*Peromyscus*) have been shown to preferentially produce cooperative sperm groups with more closely related sperm (Fisher and Hoekstra 2010). Sperm head hook parameters have frequently been linked to species promiscuity (Tourmente et al. 2016, Immler et al. 2007, Moore et al. 2002), with some studies showing increased hook length positively correlated with increased swimming velocity (Varea-Sánchez et al. 2016). However, a recent study of 25 murid species found no relationship between the presence and extent of the hook-shaped sperm head and a predisposition to sperm train formation itself (Tormente et al. 2016). Firman et al. (2013) found no evidence of sperm conjugate formation in sperm of a murine species bearing three apical hooks. The relative rarity of the sperm-train strategy requires that more species be examined before the primary purpose of the apical hook is understood. There is currently no evidence of sperm train formation in *Mus musculus* species. In the meantime, the distinct apical hook shape of murid spermatozoa provides both challenges and opportunities in morphological assessment.

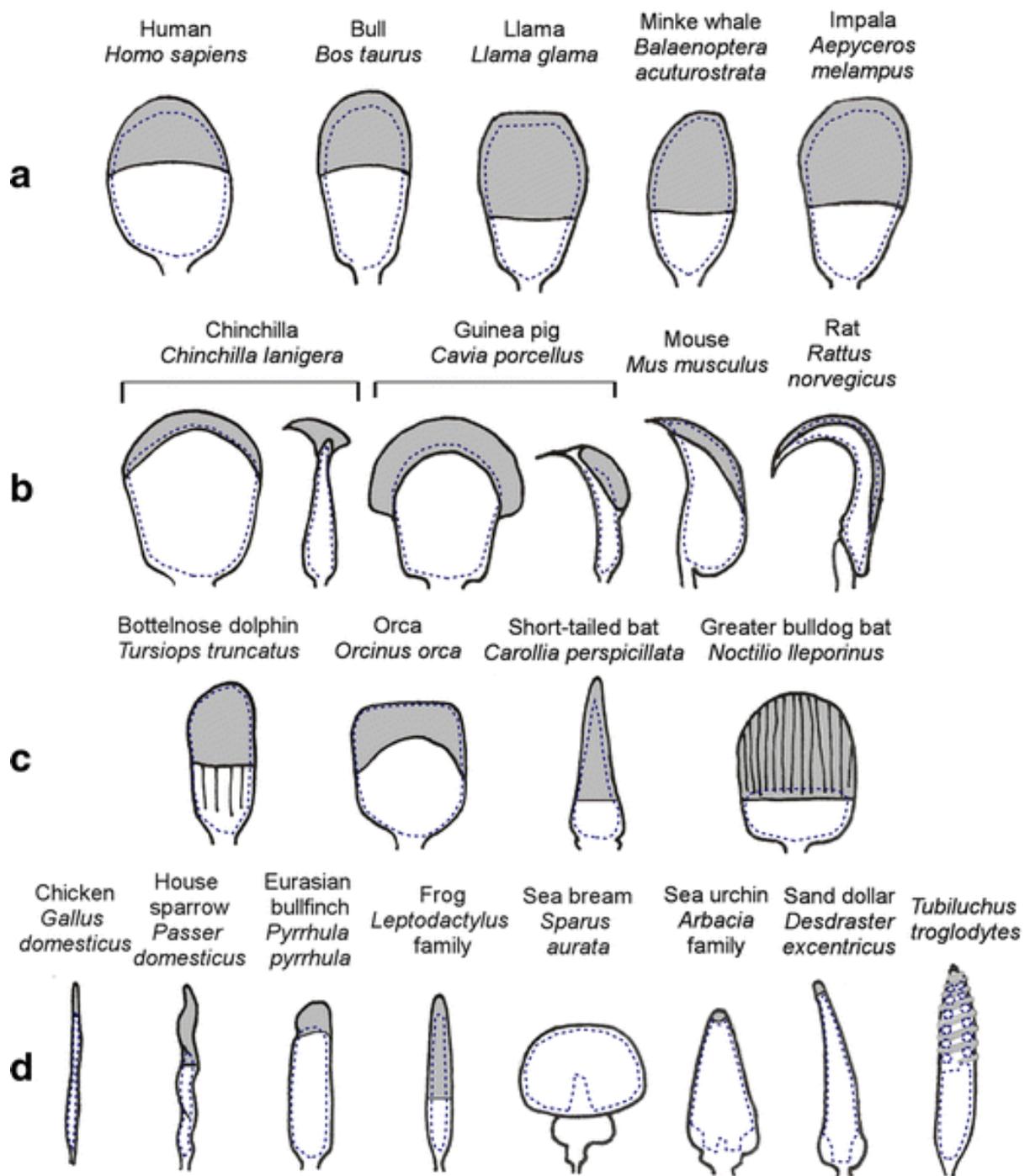


Figure 1.10. A selection of sperm head morphologies from across metazoa; acrosomal regions are shaded in grey and nucleus cross-sections denoted by a dashed outline. a) The typical ovate or paddle head shape seen in many mammals. b) Examples of giant acrosomes (including sagittal cross-sections) and falciform hooks seen in rodents. c) Atypical mammalian head shapes. d) Examples of morphologies from outside mammalia, including the anomalous sperm head of the Eurasian bullfinch (*Pyrrhula pyrrhula*), the rounded acrosome-less sperm head of the sea bream (*Sparus aurata*) and the spiralling acrosome sperm head of the ‘living fossil’, *Tubiluchus troglodytes*—the nucleus of which also forms a remarkable double spiral in the anterior portion of the sperm head, seen here in cross-section. Image and caption taken from Skinner and Johnson (2017).

The acrosome reaction

Successful fertilisation in the mouse and many other mammals involves the proper execution of two major events prior to sperm-egg fusion; execution of the sperm acrosome reaction, and binding to and penetration of the oocyte zona pellucida.

Spermatozoa must undergo the acrosome reaction (AR) before being able to fertilise the oocyte. This involves “multiple fusions of the outer acrosomal membrane with the overlying sperm membrane”, allowing contents of the acrosome to migrate to the external cell membrane (Yanagimachi 2011). In 2011, Jin et al. demonstrated through the video capture of GFP-tagged acrosomes in mouse spermatozoa that, contrary to popular belief, most mouse spermatozoa were already acrosome-reacted when first observed in the cumulus layer. This, alongside the observation that sperm are able to bind with and fertilise decumulated oocytes (Kang et al. 2010), suggests that alterations to sperm surface protein availability, whether through changes in physical accessibility or through co-operative protein modification, may occur in the luminal fluid of the female reproductive tract or at some other point prior to contact with the oocyte. Evidence for this has been seen in the case of a number of cell-surface proteins, including, for example, sperm adhesion molecule 1, *Spam1* (Griffiths et al. 2008, Chen et al. 2006, Deng et al. 1999).

Hyaluronidase activity

Hyaluronidases are enzymes that degrade hyaluronic acid (also called hyaluronan), and mammalian sperm acrosomes are known to contain a number of hyaluronidases. The female gamete, the oocyte, is surrounded by an extracellular layer called the ‘cumulus oophorus’ that is composed of cumulus cells suspended in a matrix of proteins and hyaluronic acid (*Figure 1.9*). Hyaluronidases perform a key role in fertilisation by catalysing the digestion of hyaluronic acid. By encouraging the dispersal of the cumulus layer, sperm can gain access to the oocyte to bind and penetrate the zona pellucida - another function of some hyaluronidase proteins. Each sperm carries its own hyaluronidase proteins which are translated and relocate to the acrosome during spermiogenesis.

The sperm acrosome has been shown to contain a number of hyaluronidases which undergo cellular exocytosis or migration to the external cell membrane prior to the sperm reaching the oocyte (Jin et al. 2011, Thaler and Cardullo 1995). A number of sperm head hyaluronidases, including *Spam1* and *Hyal5* have been identified in mouse (Kim et al. 2005, Lin et al. 1994); however, the full importance of their presence for successful fertilisation has been brought into question (Kang et al. 2010, Baba et al. 2002). Hyaluronidase activity and gene function has been linked to cases of transmission ratio distortion in mouse (Zheng and Martin-DeLeon 1999), and there are suggestions of reduced hyaluronidase activity in other MSYq deletion models (Styrna and Krzanowska 1995).

1.3.2. Gene expression throughout the course of spermatogenesis

During spermatogenesis, the germline undergoes both meiotic sex chromosome inactivation (MSCI) and postmeiotic sex chromatin repression (PSCR). Prior to meiosis, a large number of sex-linked genes are specifically expressed in spermatogonia, progenitors of spermatids (Wang et al. 2001). MSCI begins at the pachytene spermatocyte stage (stage 5 in *Figure 1.11*) facilitated by histone modifications as cells undergo meiosis. MSCI is a widely conserved phenomenon of male germ cell production and is a specialised form of the more general mechanism of meiotic silencing of unsynapsed chromatin (MSUC) (Turner 2007). It results from remodelling of the X and Y chromosomes and the formation of the ‘sex body’ or ‘XY-body’, a compacted structure composed of the sex chromosomes that is sequestered away into a peripheral nuclear subdomain (Turner 2007, Turner et al. 2000, Solari 1969). Failure to correctly undergo MSCI is a known factor in meiotic-based sterility. MSCI is thought to extend to silencing all mRNA X-linked genes during meiotic prophase in male gametogenesis, but it has been suggested that a substantial proportion of known X-linked miRNAs escape silencing during MSCI (Yan and McCarrey 2009). This conclusion is disputed and further evidence of this escape has not been forthcoming, with the opposite being confirmed in work by Royo et al. (2015). PSCR occurs in postmeiotic cells, though appears less pervasive than MSCI. Although haploid spermatids undergo PSCR there is still notable sex chromosome gene expression, especially in haploid round spermatids. Spermatid specific genes such as *Sly* (*Sycp3*-like on the Y), *Ssty1* and *Ssty2* (Spermatid specific transcript on the Y 1 and 2) reside on the long arm of the Y chromosome and their expression is required for proper formation of the mature sperm and full fertility. RNA sequencing and microarray analyses of postmeiotic germ cells have shown that sex chromosome gene expression is abundant in round spermatids, and that the composition of repressive epigenetic marks to the sex chromosomes is distinct to that seen on the autosomes (Moretti et al. 2016). The existence of partially penetrant PSCR is likely to have contributed substantially to selective pressure for the enrichment of spermatid-specific genes on the X and Y chromosomes.

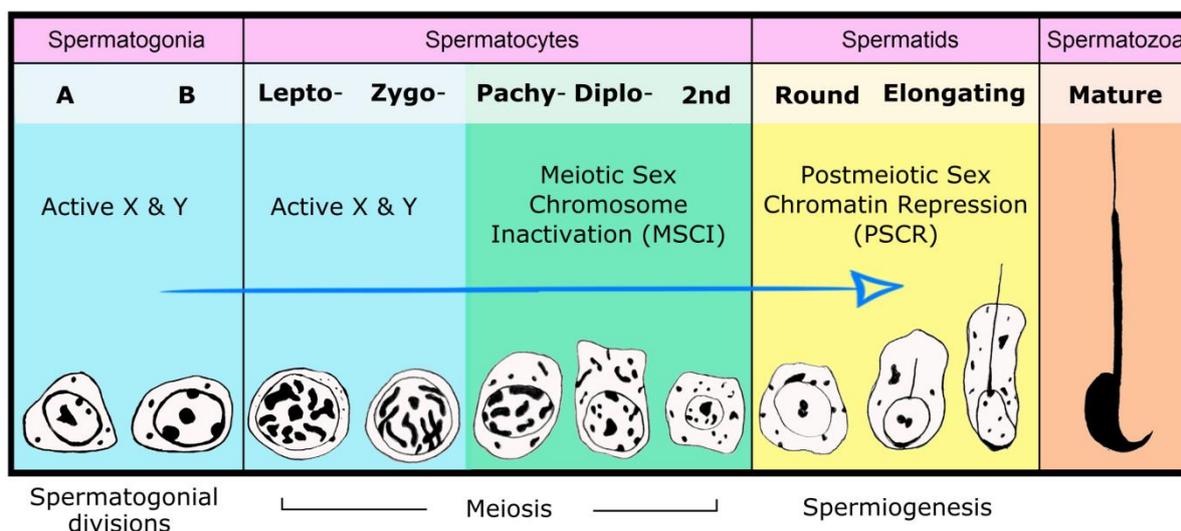


Figure 1.11. Progression of spermatogenesis in the mouse testis, from spermatogonial stem cells to mature spermatozoa. The diagram was created using personal observations and figures from Russell (1990) and Turner (2007) as references.

1.3.3. Cytoplasmic sharing

During Spermatogenesis, sperm cells develop as a syncytium, connected to one another by cytoplasmic bridges that allow for the sharing of genetic information between developing sister spermatids. This sharing allows for genetically distinct X- or Y-bearing haploid cells to maintain relative functional equivalence through the passage of products from sex-linked genes.

Cytoplasmic bridges are large (up to 3µm) intercellular bridges that exist between developing germ cells during gametogenesis, and join the cytoplasm of neighbouring cells to form a syncytium. Their existence has been known for more than a century, and they are highly conserved, being exhibited across multicellular organisms ranging from *Drosophila* to humans. Despite their prevalence, the molecular basis of their formation, maintenance, and function are not fully understood (Greenbaum et al. 2011).

Confirmation of cytoplasmic bridges facilitating sharing of gene products between developing spermatids has been demonstrated through examination of transcript and protein localisation to the bridges (Morales et al. 2002). Specifically, transcript sharing between developing spermatids has been confirmed using gene-specific autosomal knockout models and examination of sex-linked gene product sharing (Nipper et al. 2005; Caldwell and Handel 1991). Cytoplasmic bridges have also been shown to facilitate the movement of chromatoid bodies (mRNA containing ribonucleoprotein bodies) whose existence is specific to the round spermatid stage of spermatogenesis (Ventelä et al. 2003). Intercellular bridges are not limited to the reproductive system, for example in epiblast cells of the zebra fish embryo where they allow the transfer of proteins between distant but connected cells and facilitate communication between cells during gastrulation (Caneparo et al. 2011).

Sharing of cytoplasmic constituents across cytoplasmic bridges permits the development of ‘de facto diploid’ gametes of relative functional equivalence (Braun et al. 1989). Some genes that exist on the X-chromosome and are transcribed at the haploid spermatid stage, such as *Akap4*, are essential for sperm formation (Nipper et al. 2005). If gene product sharing were not possible or were restricted during spermatogenesis, the absence of such products in Y-bearing spermatids could lead to distinct functional differences between sperm carrying opposing sex chromosomes. Whilst there is evidence of transcript sharing for genes such as *Akap4* (Nipper et al. 2005) and *Prm1* (Caldwell and Handel 1991, Braun et al.1989), we also know that some RNA transcripts are not shared across cytoplasmic bridges.

Transmission ratio distortion (TRD) is the term used to describe the non-mendelian transfer of genetic information to the next generation, and in males, this may occur if one allele impairs the development or function of the sperm; see section 1.4 for more information. That individual sperm developed in syncytia can show distinct functional capability suggests that sperm do not develop under perfectly equal conditions. In order for functional differences to manifest between sperm that develop in syncytia, there must be a failure to share specific gene products between spermatids. This non-sharing must extend to both RNA transcript and protein products of a gene.

The *Smok* (sperm motility kinase) gene family shows restricted sharing in mouse spermatid syncytia. *Smok1* is an autosomal gene expressed in post-meiotic spermatids and the non-sharing of gene products from it and its mutant allele *Tcr*, result in segregated development of sperm populations in affected mice (Véron et al. 2009). Véron et al. (2009) confirmed the lack of sharing of *Tcr* and *Smok1* transcripts by producing a Myc-tagged *Tcr* transgene construct whose RNA product subsequently only appeared in a subset (around 50%) of the spermatid population. A similar experiment was conducted by the Affara lab (unpublished) has shown lack of sharing of *Smok2b* transcripts in a transgenic mouse line carrying a KillerRed-*Smok2b* fusion gene. *Figure 1.12* shows restricted distribution of *Smok2b* to approximately 50% of the spermatid population. This work was conducted by Kim Lachani and Dr Peter Ellis.

Several examples of transmission ratio distortion exist that are caused by reduced fertilising ability of a subset of the sperm population. A number of these examples include skewed inheritance of either one of the sex chromosomes resulting from functional inequivalence between X- and Y-bearing sperm. Transmission and sex ratio distortion are discussed in more detail in the next section, alongside additional examples of non-sharing of gene products.

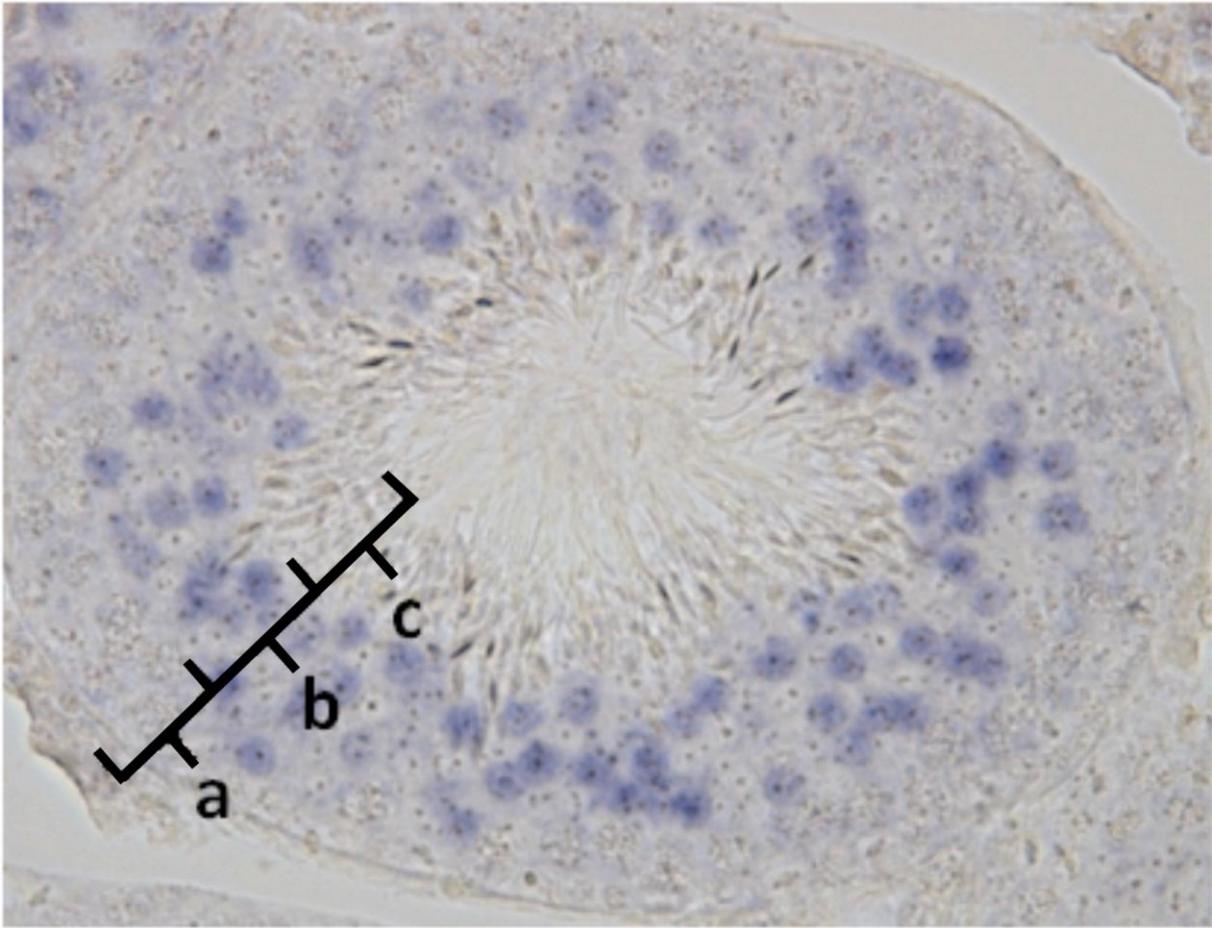


Figure 1.12. RNA *in situ* hybridisation of digoxigenin-labelled *Smok2b* transcripts in transgenic mouse testis from a line carrying a KillerRed-*Smok2b* fusion gene. This shows uneven distribution of transcripts between round spermatids. a) spermatocytes b) round spermatids c) late elongating spermatids and mature sperm. Image credit to Kim Lachani and Dr Peter Ellis.

1.4. Transmission Ratio Distortion, Sex Chromosome Meiotic Drive and Sex Ratio Distortion

1.4.1. Mendelian ratios and transmission ratio distortion

According to Mendel's law of random segregation, both copies of a gene have an equal chance of being passed on to the subsequent generation. Transmission ratio distortion (TRD), also known as meiotic drive or segregation distortion, involves 'selfish' genetic elements that are able to transmit at a higher than expected (>50%) frequency, above the Mendelian inheritance ratio. This observation can have many causes, but the term 'transmission ratio distortion' simply refers to instances in which "the two alleles carried by a heterozygote are transmitted unequally to the zygote at the time of fertilisation" (Lyon 2003). In males, this may occur if one allele impairs the development or function of the sperm. Distorters can emerge on any chromosome, and when these complexes occur on sex chromosomes this can result in preferential transmission of that chromosome to the next generation,

called sex ratio (or sex chromosome) meiotic drive (SRMD). Instances of SRMD and genomic conflict related sex ratio distortion are known to exist across several taxa, and a number of these distorters have been studied in *Drosophila*.

1.4.2. Transmission ratio distortion and sex chromosome meiotic drive in mouse

The most well-studied example of a transmission distorter in mammals is the *t*-complex; a highly varied region of chromosome 17 that has been found to exist in wild mice (Huang et al. 2013). When transmitted in a heterozygous form, the distorter genes of the *t*-complex act on the non-*t*-form of the responder gene, rendering any sperm carrying this form severely impaired or completely non-functional. The *t*-forms of the distorter loci act additively, and leads to a higher transmission of the *t*-form of the responder gene (between 50-99%) (Lyon 2003, Lyon 1986). Males homozygous for the *t*-haplotype are sterile due to major defects in sperm motility. Transmission distorters have been seen to accumulate in low recombination regions, where cooperation between distorter and responder genes is able to evolve through tighter linkage (Lyon 1984, Coquet et al. 2012).

Previous studies have shown that Chromosome 6 deficiencies are associated with impaired sperm function (Zackowski and Martin-DeLeon 1989). A case of transmission ratio distortion and associated offspring sex ratio distortion was noticed by the DeLeon lab (Department of Biological Sciences, Delaware) in mice heterozygous for a 6.16 Robertsonian fusion. An impaired fertilising ability was noted in mice carrying the fusion which was potentiated in X-bearing sperm. They found that in male mice heterozygous for the Robertsonian translocation Rb(6.15) sperm exhibited impaired fertilizing ability compared to normal sperm, and that the functional inequality bears out in X-bearing sperm leading to a skew in the offspring sex ratio (around 36 X : 64 Y) in resultant offspring (Zheng and Martin-DeLeon 1999, Aranha and Martin-DeLeon 1995, Aranha and Martin-DeLeon 1991). The reduced sperm fitness was linked to a damaged copy of *Spam1*, a prominent sperm head hyaluronidase (see section 1.3.1) present on mouse chromosome 6, where a number of other hyaluronidase genes also reside (Zheng et al. 2001a), suggesting an interaction with indirect sex-linked mediators of hyaluronidase activity. In work using two different translocation models with transmission ratio skewing, the group showed that *Spam1* maps to a location close to the centromere of chromosome 6, and at the junction of the Rb(6.16) and Rb(6.15) translocations. It is believed that the dysfunction of sperm bearing a damaged copy of Chromosome 6 may be as a result of damage caused to the *Spam1* gene (Deng et al. 1997). Study of *Spam1* has indicated a lack of sharing of both its mRNA and protein in developing spermatids (Zheng et al. 2001b).

1.4.3. Mammalian sex ratios, intragenomic conflict and sex ratio distortion across taxa

Both Fisher and Düsing argued that a 50:50 offspring sex ratio was required in most dioecious species on the basis of parental expenditure and optimal genetic representation in the subsequent generation (Fisher 1930, Düsing 1884 via Edwards 2000). However, there exist a number of instances in nature where the standard sex ratio is maintained away from 50% female.

As explained by Hamilton in 1967, the existence of sex ratio distorters provides one exception to Fishers' rule that offspring sex ratios are maintained at 50:50 on the basis of parental expenditure (Hamilton 1967, Fisher 1930, Düsing 1884 via Edwards 2000). Because one chromosome of the heterogametic sex is transmitted only through those offspring, there exists a strong selective advantage for the production of distorters for their transmission on that chromosome. A distorter on either sex chromosome can exert its effects naturally in a population as long as any deleterious effects on population fertility are outweighed by chromosomal benefits of the skewed inheritance. This often ultimately leads to the development of suppressor genes located on the opposite sex chromosome or autosomes (Cocquet et al. 2012, Atlan et al. 2003, Merçot et al 1995). Evidence of offspring sex ratio distortion exists in many animal models, both mammalian and non-mammalian. The phenomenon has been extensively studied in *Drosophila*, and some of these examples are discussed below.

An X-linked sex ratio distorter referred to as *Dox*, causes a deficiency of Y-bearing sperm and a resulting sex ratio skew towards female offspring in *Drosophila* (Branco et al. 2013). And both Y-linked and autosomal (Tao et al. 2007b) suppressor of SRMD have been reported. *Dox* is a well known X-linked facilitator of SRMD in *Drosophila*. Unlike the mouse Y chromosome that contains more than 700 protein coding genes (Soh et al. 2014), the *Drosophila* Y contains just 22 such genes (Ensembl BDGP6, INSDC Assembly GCA_000001215.4, Jul 2014). However, Y-linked suppressors are also known to exist amongst a number of species within this genus, and Y chromosomes are categorised as being 'suppressor' and 'non-suppressor' [of sex ratio distortion] (Montchamp-Moreau et al. 2001, Carvahlo et al. 1997). In *Drosophila*, suppressors of sex ratio distortion appear to act through epigenetic alterations, by RNA interference (Branco et al. 2013, Tao et al. 2007a, Tao et al. 2007b). Branco et al. (2013) reported the upregulation of a subset of post-meiotic genes and the down-regulation of a specific set of mitochondrial and gland-specific genes in cells of *Drosophila* with suppressor Y-chromosomes compared to those with a non-suppressor Y. Their observation that this differential expression was also evident in XXY females where expression of Y-linked protein coding genes was repressed, lead them to conclude that at least part of the suppressor complex acts through epigenetic mechanisms. In *Drosophila pseudoobscura*, the Y-sperm killing X-linked meiotic driver know as 'sex-ratio' also reduces competitiveness in X-bearing sperm, meaning affected males

are less effective in the presence of sperm competition (Price et al. 2008a). SRMDs and TRDs have been associated with adaptations of reproductive tactics in both the mouse and housefly and are suggested to lead to higher rates of polyandry as a result of reduced competitiveness of affected males (Sutter and Lindholm 2015, Manser et al. 2011, Price et al. 2008b).

In some cases, offspring sex ratio distortion is modulated by other factors. In *Drosophila*, the pro-female sex ratio distortion facilitated by *sex-ratio* (a distorter on the X) is greatly enhanced at lower temperatures (Tao et al. 2007a). Environmental and dietary factors also influence sex ratios in mammals; for example, the proportion of calories from fat in maternal diet was shown to influence sex ratio in mice (Rosenfeld et al. 2003). Rational evolutionary bases for adaptive sex ratios in response to environmental conditions have been suggested by Trivers and Willard (1973), though mechanisms of action remain mostly out of reach. Even the influence of separate organisms has been shown to alter sex ratios, such as in cases of infection by the widely prevalent bacteria *Wolbachia* in several species of arthropoda (Gunnarsson et al. 2009, Moreau and Rigaud 2000, Rigaud et al. 1999). The bacterium is only transmitted through the female line and propagates by the selective killing or feminisation of male embryos. In a number of cases, long-term infection by this bacteria has led to the adaptation of sex-determining systems in the host species (review by Charlat et al. 2003).

1.4.4. Intragenomic conflict in the mouse

In mouse, both the X and Y chromosomes contain *Sycp3*-like ampliconic genes. The X chromosome possesses both *Slx* and *Slx11* in multicopy; the homologue of these is *Sly*, found in multicopy on the long arm of the Y-chromosome; this is covered in more detail in sections 1.2.2 and 1.2.3. These genes are now known to act in opposition upon postmeiotic sex chromatin and take part in an intragenomic conflict (Morgan and Villena 2017, Cocquet et al. 2012).

Genes of the Y-chromosome long arm have been shown to recruit repressive epigenetic marks to sex chromatin during spermatogenesis (Cocquet et al. 2012, Cocquet et al. 2009, Reynard and Turner 2009). Through the production of *Sly* knockout mice using small interfering RNAs (siRNAs), it has been shown that *Sly* has a crucial role in postmeiotic sex chromatin repression (PSCR) via the recruitment of epigenetic markers following meiosis, when sex chromosomes are partially reactivated post-meiotic sex chromosome inactivation (MSCI) (Cocquet et al. 2012, Burgoyne et al. 2009, Cocquet et al. 2009). *Sly* represses X- and Y-linked gene expression, including the X-linked homologs of *Sly*; *Slx* and *Slx11*; with *Sly* deficiency resulting in sub-fertility and cell differentiation defects (Cocquet et al. 2012). SLY proteins localise to the sex chromatin (and a region of chromosome 5 corresponding to the *Speer* cluster) in post-meiotic spermatids and the loss of SLY results in the reduction of H3K9 (histone H3 Lysine 9) methylation and reduced expression of CBX1, a heterochromatin protein (Cocquet et al. 2009).

Slx and *Slx1l* have the opposite effect to *Sly*, stimulating XY gene expression via metabolic processes in spermatid cytoplasm and are essential for normal sperm differentiation and male fertility (Kotarska et al. 2014, Cocquet et al. 2010). In the absence of SLY, SLX and SLXL1 colocalise with the sex chromatin (and *Speer* cluster). In addition, *Slx* and *Slx1l* deficiency results in a delay in spermatid elongation and sperm release (Cocquet et al. 2010) and the overproduction of male offspring, away from the expected 50:50 ratio (Cocquet et al. 2012). The opposing functions of *Sly* and the two X-homologues, and the resulting intragenomic conflict, are likely to be the basis for the massive multicopy expansion of these regions of the sex chromosomes in mice through selection to maintain gene balance. Whilst deficiency in either gene leads to notable consequences to fertility; the absence or reduction of *Sly*, *Slx* and *Slx1l* genes simultaneously appears to partially rescue the phenotype (Cocquet et al. 2012). *Ssty* has also been implicated in the control of XY gene expression in post-meiotic spermatids. SSTY proteins have been shown to interact with both SLY and SLX/SLXL1 and to co-localise with sex chromatin, suggesting a potential shuttling or localisation role for this protein (Comptour et al. 2014).

Deficiency of *Sly* and *Ssty* in MSYq deletion mice results in global upregulation of X-linked genes and remaining Y-linked genes in post-meiotic spermatids (Ellis et al. 2005). Due to the existence of pro-female offspring sex ratio skews in MSYq deficient males, the distortion is expected to be mediated by an X-linked complex, for which the *Slx/Slx1l* and *Sly* regions may be competing regulators; *Sly* acting as repressor and *Slx/Slx1l* as stimulators. The MF1XY^{RIII}qdel mouse, which lacks 2/3 of the normal Yq, is both fertile and exhibits a sustained sex ratio skew towards female offspring of around 10% (section 1.5.2 and 1.5.3), providing an ideal opportunity to examine the physiological and molecular consequences of a substantial genomic deletion.

Although copy number variation (CNV) of *Sly* and *Ssty* is known to correlate with the extent of abnormal fertility phenotypes (Case et al. 2015, Comptour et al. 2014, Cocquet et al. 2012), Case et al. (2015) highlighted the additional potential contribution of other MSY genes, including *Srsy*, *Rbmy* and *Rbm31y*, to the conflict. Using multiple regression analysis to examine the relationship between a variety of MSYq deletion phenotypes (including offspring sex ratio distortion) and CNV of individual MSY-linked genes in consomic strains of B6-ChrY inheriting a *M. m. domesticus* Y chromosome, they highlighted the existence of significant relationships between the CNV of the above genes and offspring sex ratio distortion. Although intriguing, the analysis does not rule out the impact of association and reports only where relationships exist by correlation. No mechanistic means of involvement in the intragenomic conflict has been reported for these additional genes, and the observed impacts of these additional genes are not independent to known sex ratio regulators *Sly* and *Ssty* in this study.

1.5. Deletions of the Male-Specific Region of the Y Chromosome Long Arm

A number of naturally occurring deletion models exist with losses of part of or the entire long arm of the Y-chromosome (Yq) in mouse. The extent of the deletion varies, with male mice completely lacking the Yq being infertile (Grzmil et al. 2007). In some models, a symptom of such partial deletions (where males remain fertile) is a distortion of the sex ratio evident in the offspring of affected males (Ellis et al. 2005, Conway et al. 1994, Moriwaki et al. 1988)

The MF1XY^{RIII}qdel was first reported in 1994 (Conway et al. 1994), when a 2/3 deletion of the male specific region of the Y-chromosome (MSY) long arm was indentified alongside a 13% offspring sex ratio skew from wild-type in favour of females (the control strain produced 51% male offspring, the deletion strain produced 38% male offspring). This data was generated over 10 generations of 409 mice, with no reduction in overall fertility. The original documentation also mentions a 3.4x increase in the frequency of abnormal sperm heads compared to the wild-type. The substantial deletion first appeared in XYS^{r^a} males, which carried the RIII strain Y-chromosome. Both the full complement RIII Y and the 2/3 deletion RIII Y were then backcrossed onto an MF1 background (and later, separately, onto a C57BL/10 background) to improve fecundity and litter size, resulting in the MF1XY^{RIII} and MF1XY^{RIII}qdel models. The C57BL/10^{RIII}qdel males also demonstrated an offspring sex ratio skew in favour of females, though slightly less severe (C57BL/10^{RIII} = 57% male, C57BL/10^{RIII}qdel = 48% male) (Conway et al. 1994). The Y353/B gene referred to in the paper's title, is now known to be a member of the *Ssty1/2* gene family (discussed further in section 1.2.2).

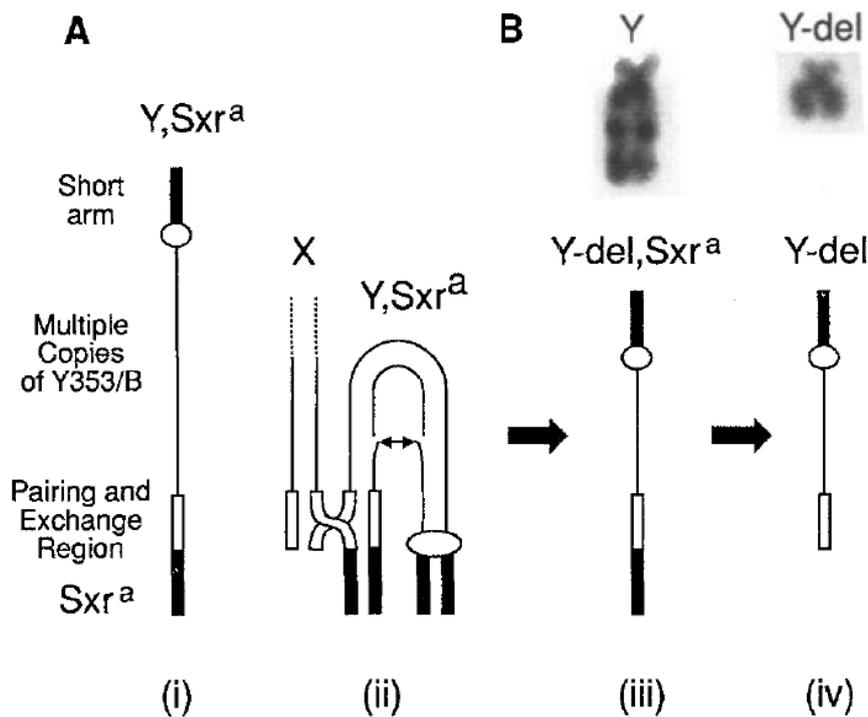


Figure 1.13. Original image from Conway et al. (1994) showing the ‘presumed mode of origin’ of the deletion Y chromosome (A). (i) The Sxr^a males carry an extra short arm region of the Y chromosome, called the $Sxra$, that is attached to the distal end of the Y-chromosome PAR. (ii) Demonstrates the self-synapsed structure between the short arm and the $Sxra$, and the suspected intrachromatid exchange between the repetitive sequences of the Y-chromosome long arm. (iii) The resulting $Yqdel Sxra$, before the Sxr^a region was lost through normal X-Y recombination to leave the $Yqdel$ (iv). Subset image (B) shows the Y chromosome (and substantial Y deletion of more than half the length of the chromosome) from bone marrow metaphases of both the RIII Y-chromosome (wild-type) and RIII- $Yqdel$. Image taken from Conway et al. (1994).

1.5.1. Mouse MSYq deletion models

Large scale chromosomal deletions occur naturally within mammalian populations. Within the research environment a number of deletion models have been maintained for study. There are several examples of Yq deletion mouse models, both previously and currently maintained that provide insights into the effects of small and large scale sex chromosome deletions in the context of intragenomic conflict (see *Table 1.3*). The mouse is an ideal model for studying mammalian genomics; being both easy to house and with a relatively short reproductive cycle, it is possible to work across generations of data within a few years. The massive amplification of genes on the sex chromosomes and their broadly understood genetic make-up also make them intriguing and useful models for research.

Previously and currently maintained mouse (<i>Mus musculus</i>) MSYq deletion models						
Deletion model name	Control model name	Approximate extent of MSYq deletion	Original reference	*% sperm with head abnormalities in Ydel	Litter size of Ydel/control	% male offspring in Ydel/control
B10.BR-Y ^{del}	B10.BR/SgSn	75%	Moriwaki et al. (1988)	70 – 88 ^a	6.89/6.59	41 / - ^b
X ^{Sxr} aY ^X	N/A	100%	Burgoyne et al. (1992)	100	sterile	N/A
MF1XY ^{RIII} qdel	MF1XY ^{RIII}	66%	Conway et al. (1994)	6.9 ^c	-/- ^d	38 / 51
C57BL/10 ^{RIII} qdel	C57BL/10 ^{RIII}	66%	Conway et al. (1994)	57.2 ^c	-/- ^d	46 / 57
CBA-Y ^{del}	CBA	66%	Styrna et al. (2003)	92	No difference	-/- ^d
XY ^{Tdym1} qdelSry	XY ^{Tdym1} Sry	90%	Touré et al. (2004a)	100	sterile	N/A
C57BL/6JBomTac	C57BL/6J	40%	Fischer et al. (2017)	35 ^e	No difference ^e	41 / 51 - 2016 data ^e

Table 1.3. Summary of MSYq deletion mouse models with approximate extent of the MSYq deletion listed alongside full Y-chromosome complement counterpart and reproductive phenotypic data for the Ydel strain.

*Sperm head abnormalities were categorised independently for each publication without full consensus between descriptions of ‘normal’ and ‘abnormal’.

^a70% abnormality as reported by Moriwaki et al. in 1988, and 88% abnormality as reported by Styrna et al. in 2002.

^b Male offspring ratio was not confirmed for the control strain, however it is implied as being neutral i.e. c.50%.

^c The definition of ‘normal’ included sperm with flattened acrosomes in this paper (Conway et al. 1994), ‘abnormal’ was reserved for the most severely deformed sperm.

^d Information not found.

^e from MacBride et al. 2017

Major spermatogenic defects were originally attributed to loss of the MSYq based on findings from the $XSxr^aY^{*X}$ line, originally an XO male line with a transposed RIII-derived Sxr^a (Y-chromosome short arm lacking some copies of *Rbmy*) and later with an added, diminished Y (see section 1.2.2). This mouse line is now known to lack the majority of copies of the Yp gene *Rbmy*; a gene previously implicated in maintaining sperm quality (Touré et al. 2004a, Mahadevaiah et al. 1998, Elliott et al. 1996). Given that *Rbmy* is known to be transcribed though not translated in spermatids, its impact on sperm morphology in a post-meiotic context is unclear (Szot et al. 2003). Mice used in inferring its importance possess a chimeric Y chromosome with the Yp and the Yq originating from distinct strains; meaning that the defects in sperm quality cannot necessarily be directly attributed to the loss of *Rbmy*. The contribution of *Rbmy* and genes such as *Rbm31y* to the maintenance of sperm integrity is uncertain. The recently discovered C57BL/6JBomTac is reported as lacking both *Rbm31y* copies, normally present on the MSYq (MacBride et al. 2017).

J. Styrna and colleagues have published several reports of their examination of a mouse model with a partial Y chromosome deletion similar to that of the MF1XY^{RIII}qdel. The B10.BR-Y^{del} and its congenic ‘wild-type’ strain B10.BR/SgSn first appeared in Japan in 1977 (Pietraszek and Styrna 1995) and were transported to Poland in 1981 (Styrna et al. 2003), where they were maintained by backcrossing -Y^{del} or /SgSn males to B10.BR females. The B10.BR-Y^{del} deletion is thought to extend to 3/4 of the standard B10.BR MSYq (the long arm of the male specific region of the Y chromosome) (Styrna et al. 1991b). Over the course of a number of years, B10.BR-Y^{del} males were found to have several differing traits when compared to B10.BR/SgSn males that possess the full Y-chromosome complement; see *Table 1.4*.

The authors also noted that sperm from B10.BR-Y^{del} with mildly deformed heads (as defined in Styrna et al. 2002) were not reduced in number in the oviduct versus the uterus, unlike the most severely abnormal sperm which showed a notable reduction. However, these findings were not compared with control B10.BR/SgSn sperm so it is difficult to conclude whether this finding relates to reduced fertilising ability in the Y^{del} or suggests that severe morphological abnormalities are associated with other lower functionality traits across both strains.

Trait	First Reported
Higher frequency of degenerated tubules	Styrna et al. (2002)
Higher level of X-Y dissociation at meiosis	Styrna et al. (2002)
Higher percentage of sperm with abnormal heads	Styrna et al. (1991a)
More caudal sperm with cytoplasmic droplet attached	Styrna et al. (2002)
Flat acrosomes and deficiency in proteolytic enzymes	Styrna et al. (2002)
Low numbers of sperm retrieved from the uterus	Styrna et al. (2002)
Higher proportion of dead sperm in vas deferens	Styrna and Kraznowska (1995)
More sterile copulations	Styrna et al. (2002)
Higher proportion of sperm lacking acrosine	Styrna et al. (1991b)
Impaired capacitation	Xian et al. (1992)
Lower scores in hyaluronic acid penetration tests	Styrna and Kraznowska (1995)

Table 1.4. Traits found to be associated with the Yq deletion in B10.BR-Y^{del} mice as compared to B10.BR/SgSn (full Y-chromosome complement) mice. Papers reporting the discovery of the trait are listed in the second column.

The B10.BR-Y^{del} males were also used to develop a CBA-Y^{del} model, through backcrossing of CBA females with B10.BR-Y^{del} males for the examination of genetic background effects on Yq deletion associated fertility symptoms (Styrna et al. 2003). Many of the traits observed in B10.BR-Y^{del} males were investigated in CBA-Y^{del} counterparts. Results showed that whilst the background genetic composition had some effect on the efficiency of spermatogenesis (minimal improved sperm morphology in CBA versus B10.BR males) and testis weight, spermatogenic defects could be widely attributed to the Y-chromosome deletion (Styrna et al. 2003). Observable deformations appeared in CBA-Y^{del} spermatogenesis at the round spermatid stage, where acrosomal abnormalities (including abnormal acrosomal vesicle shape and displacement of the developing acrosomal granule) were noted and progressed into the elongating spermatid stage. Though these abnormalities were noted in the CBA background mice, these traits are not expected to be limited to Y^{del} mice on a CBA background (Styrna et al. 2003).

Later work with the B10.BR-Y^{del} model has revealed inherited epigenetic effects of the Yq deletion. Kotarska et al. (2014) reported that female offspring of B10.BR-Y^{del} males produce oocytes with excessively thick and firm cumulus matrices and that these complexes release increased amounts of progesterone (a sperm chemo-attractant), possibly as a compensatory mechanism (Kotarska et al.

2014). The females are genetically identical to their B10.BR/SgSn-fathered counterparts, and display no divergent phenotypes except the reproductive traits mentioned above. The composition of the paternal Y-chromosome influences the heritable epigenetic marks that are established in X-bearing spermatozoa, and results in subtle phenotypic changes in the female offspring (Kotarska et al. 2013, Kotarska et al. 2014).

1.5.2. MSYq deletions and offspring sex ratio distortion

It has been shown that the Yq multicopy gene *Sly* works in opposition to the X chromosome homologues *Slx* and *Slx11* (Cocquet et al. 2012, Cocquet et al. 2010) to modulate the expression of sex chromatin. Genetic background influences the Yq deletion phenotype through differences in *Slx* and *Slx11* copy number. A dosage compensation effect can clearly be seen in closely related strains with differing Y chromosome composition. This has also been observed where crosses between different mouse strains or species have been made (Case et al. 2015).

It is in later studies of the B10.BR-Y^{del} model that an offspring sex ratio skew is referred to (Kotarska et al. 2014). The sex ratio skew in B10.BR-Y^{del} was first reported by Moriwaki et al. in 1988, at 0.41 male to female. This number was generated over 15 years, from 1300 mice in 190 litters (Moriwaki et al. 1988). Though no more recent published updates to the sex ratio skew in the B10.BR-Y^{del} are provided, the authors of recent publications point to the original 1988 publication and support this with reference to unpublished lab records.

A model with one of the largest MSYq deletions, the XY^{Tdym1}qdel*Sry*, was derived from the female XXY^{Tdym1} mouse line. The deletion was originally detected in male progeny of XXY^{Tdym1} females (a *Sry* negative line) crossed with males carrying an autosomal *Sry*. These offspring were sterile, owing to the XXY complement; consequently the female parent (carrying the deletion Y) was used to establish the line (Touré et al. 2004a). The extent of Yq deletions is strongly correlated with the extent of the sex ratio skew, spermatogenic defects and sperm head morphology abnormalities (Case et al. 2015, Conway et al. 2004, Touré et al. 2004). In 2015, Case et al. published a report assessing the individual impact of CNV of MSYq-linked genes through the use of consomic mice with Y chromosomes originating from different strains. Their results confirmed that *Sly* is the primary mediator of sex ratio distortion in MSYq deletion, however they also reported that *Ssty1*, *Srsy*, *Rbmy*, and *Rbm31y* CNV showed significant relationships with disruption of the offspring sex ratio.

Offspring sex ratio skewing has been shown to occur in many of the MSYq deletion mouse models discussed earlier in this chapter (section 1.5.1), and where it has not been shown, it has not been investigated. Case et al.'s investigation examined fifteen C57BL/6J consomic strains with Y-chromosomes differing in the composition of the male-specific region of the Y chromosome (MSY). Having examined CNV across seven important genes (five MSYq genes, *Slx* and *Rmby*) from all

consomic strains, they were able to produce a map of significance for chr Y CNVs and identified reproductive traits using multiple regression analysis. The research focussed predominantly on the impact of chr Y CNV on the extent and prevalence of central nervous system (CNS) autoimmune disease in female offspring of affected males, but examinations of sex ratio distortion and sperm head morphology were also made. As previously reported, *Sly* CNV was seen to have the greatest impact on both sex ratio and sperm head abnormalities, when compared with other MSY genes *Ssty1*, *Ssty2*, *Srsy*, *Rbmy* and *Rbm31y*. A significant ($p < 0.005$) relationship was identified between *Sly*, *Ssty1*, *Srsy*, and *Rbmy* CNV and both offspring sex ratio and sperm head abnormalities (Case et al. 2015). These findings echo the reports of Cocquet et al. (2012) that identified *Sly* and its homologues *Slx* and *Slxll* as mediators of the intragenomic conflict between X and Y, and of Comptour et al. (2014) that identified *Ssty* proteins as interacting partners in this process (see section 1.4.4). *Srsy* is a poorly characterised gene that exists in multicopy on the Yq (Soh et al. 2014), its function remains unknown and evidence of a protein counterpart has yet to come to fruition.

1.5.3. Offspring sex ratio distortion and physiological characteristics of the MF1XY^{RIII}qdel mouse

In 2006, Monika Ward and Paul Burgoyne published the findings of an Intracytoplasmic sperm injection (ICSI) study conducted in two models, the XY^{Tdym1}qdel and the XY^{RIII}qdel. The XY^{Tdym1}qdel (mentioned previously in section 1.5.1) exhibits almost complete abolition of *Sly* and *Ssty* copies. The XY^{RIII}qdel is an ancestor of our MF1XY^{RIII}qdel model and possesses the same Y chromosome. ICSI was conducted using both models; it involves the direct injection of the mature sperm (sometimes spermatid in other cases) into an oocyte for the purposes of fertilisation. ICSI removes any competitive selection related to sperm function prior to fertilisation that could be present *in vivo* or *in vitro*. Through their work Ward and Burgoyne concluded that:

“1) Yq deletions in mice do not bias the primary sex ratio and 2) Y^{RIII}qdel spermatozoa have poorer fertilizing ability than their X-bearing counterparts.”

(Ward and Burgoyne 2006)

The primary sex ratio refers to the relative number of X- and Y-bearing sperm produced by each model; they showed that an equal number of X- and Y-bearing sperm were produced by mice with Yq deletions. They recorded a reduced fecundity in XY^{RIII}qdel mice *in vitro*, although males appeared of normal fertility under laboratory breeding conditions; and, importantly, a significant difference in the proportion of female offspring between ICSI litters and natural mating litters. An offspring skew towards female of 40M:60F was noted in natural matings, which was abolished when fertilisation occurred by ICSI. No change was noted in XY^{RIII} litters. This leads to the conclusion that differences

in the offspring sex ratio occur as a result of differences in X- and Y-bearing sperm function prior to fertilisation.

In addition, Ward and Burgoyne examined morphological differences in all four models. *Figure 1.14* shows images of sperm taken from $XY^{RIII}qdel$ mice and the full Y chromosome control XY^{RIII} . As noted in other models of MSYq deletion (section 1.5.1), $XY^{RIII}qdel$ sperm produced sperm of an abnormal morphology; primarily with flattened sperm heads and distorted acrosomal hooks.

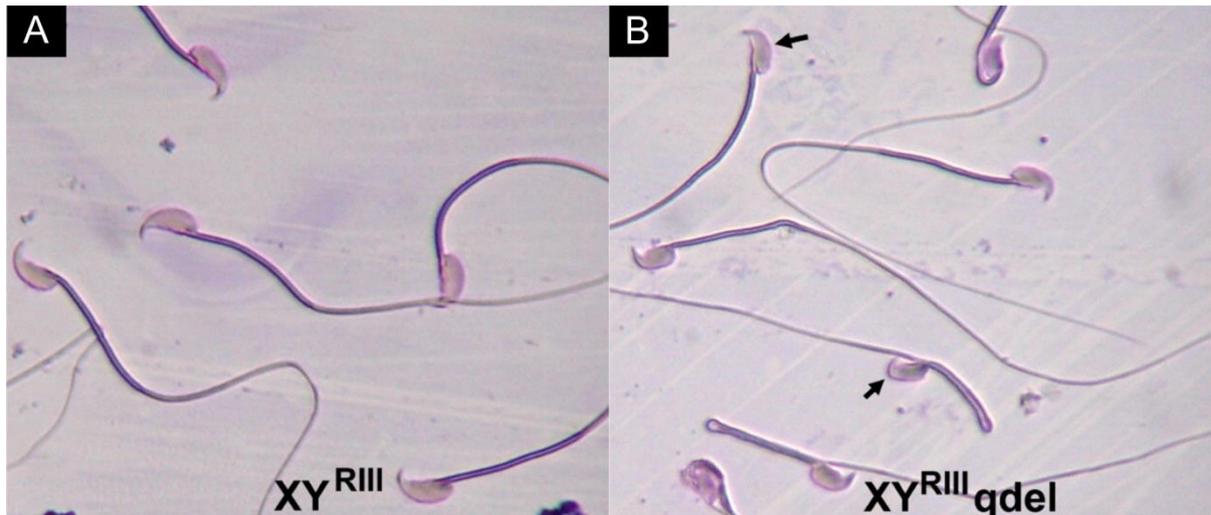


Figure 1.14. Examples of sperm taken from $XY^{RIII}qdel$ males alongside sperm from wild-type XY^{RIII} published in Ward and Burgoyne (2006). Sperm from $XY^{RIII}qdel$ mice show predominantly flattened acrosomes and other slight sperm head abnormalities (arrows). Image taken from Ward and Burgoyne (2006).

To date, our understanding of the physiological consequences of MSYq in the $XY^{RIII}del$ lineage come primarily from the Ward and Burgoyne (2006) and Conway et al. (1994) papers. The extent of the deletion has previously been shown in metaphase spreads of chromosomes (see *Figure 1.13*) and assessed by Southern blotting (Toure et al. 2005). *Figure 1.15* shows metaphase spreads produced for this investigation and the most recent comparison of Y chromosome length in the MF1 $XY^{RIII}qdel$ and conspecific full Y chromosome wild-type.

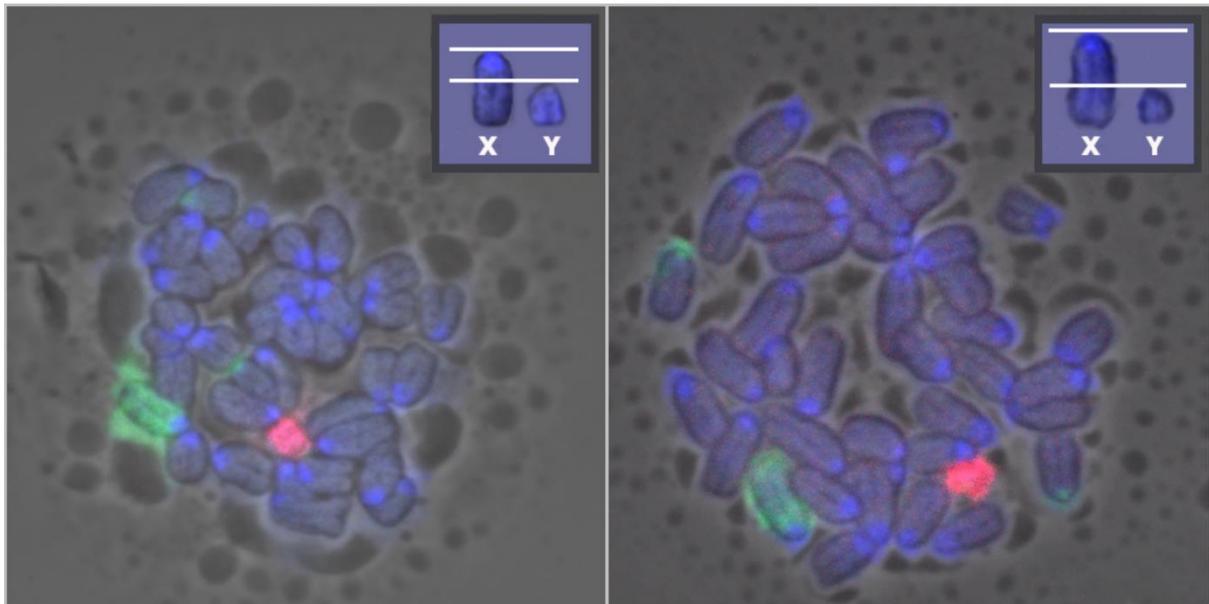


Figure 1.15. Metaphase spreads from MF1XY^{RIII} (left) and MF1XY^{RIII}qdel (right) derived from spleen cell culture. A fluorescent probe highlights the X chromosome (and a small autosomal region) in green, whilst another highlights the Y chromosome in red. Differences in comparative chromosome sizes are demonstrated inset. See section 2.2.5 for metaphase FISH methodology. Metaphase cells provided by Dr Benjamin Skinner.

1.6. Transcriptomics and the Study of Gene Expression – Investigating the Transcriptome

Gene expression is the mechanism by which genetic information contained in DNA is transcribed into mRNA. The transcriptome is the collection of all RNAs present in a sample, whether that be a collection of cells, a tissue sample, or a whole organism, at a given moment in time. Understanding the transcriptome, and therefore gene activity, has brought diverse benefits to biological research. It has brought insight to mechanisms of gene regulation, enabled the discovery of functions of previously unannotated or unknown genes, and elucidated interactions between different genetic elements and functional cascades. The molecular and analytical techniques for transcriptome analysis have advanced with some vigour over nearly two decades; from the use of real-time quantitative PCR (qPCR) and northern blotting for targeted examination of gene expression, through low-throughput Sanger sequencing, microarray, RNA sequencing and a range of novel burgeoning techniques for single cell transcriptome analyses. The two most common techniques used in contemporary biological research for whole transcriptome analyses are microarray - a probe-based sequence specific hybridisation platform, and RNA sequencing (RNA seq) – a next-generation sequencing technology allowing for the identification of novel transcripts. Each of these techniques has their own advantages and flaws, and both remain useful in considered and appropriate contexts. Transcriptome analyses are not limited to the examination of messenger RNA (mRNA) species; RNA sequencing and specifically

designed non-coding RNA (ncRNA) microarrays can detect and allow for the quantification of non-coding RNAs such as micro RNAs (miRNAs), piwi RNAs (piRNAs), long non-coding RNAs (lncRNAs) and many more. These species are now known to have essential roles in gene activation, regulation and epigenetic modification; and are a critical aspect of our understanding of biology, including functional genetics and disease mechanisms.

1.6.1. RNA Sequencing

RNA sequencing is a high-throughput sequence-based technology used for examining the transcriptome. It refers to the sequencing of RNA transcript cDNAs produced during preparation of samples specifically for this technique. Sequencing based approaches directly determine the cDNA sequence and allow for the mapping and quantification of sample transcripts to a specified reference genome. Differential expression analysis can be performed by measuring relative abundance of vast numbers of transcripts (counts) between experimental conditions. RNA sequencing use spread more rapidly following the refinement of sequencing technologies (Solexa and Illumina) in 2008 to allow for $>10^9$ transcript reads to be recorded (Lowe et al. 2017). Illumina are now the primary supplier of RNA sequencing platforms and reagents across the globe, and RNA sequencing is now the primary technology for the high-throughput examination of the transcriptome, overtaking microarray technologies in 2015 (Lowe et al. 2017).

During RNA sequencing, a purified RNA sample is converted into a cDNA library using specifically designed library preparation kits that also anneal adapters to one or both ends of each transcript. These samples can also be tagged with origin identifiers so that different samples can be pooled and sequenced together. The tagged sequences are then read by the sequencer in a high-throughput manner. Each library is amplified onto a solid surface with covalently attached linkers that hybridise to the library adapters. This creates cDNA clusters originating from a single library fragment. The sequences are then read per cluster and the number of transcripts present in each cluster is recorded. Single-end sequencing involves the reading of transcripts from one end, whilst paired-end sequencing reads the transcript from both ends. Whilst paired-end sequencing can be more laborious and costly, it provides greater sequencing accuracy and therefore is more likely to result in reads aligning to the reference genome. It also permits the discovery of gene fusions and splice isoforms (Wang et al. 2009).

1.6.2. Microarray

For many years, microarray has been the central technique of transcriptomics. It is used for the quantification of a broad range of predetermined sequences. Microarray analyses became common in the 1990s and 2000s, seeing important improvements to specificity, probe density and fluorescence detection over the years (Lowe et al. 2017, Hoheisel 2006, Heller 2002, Nelson 2001). The microarray

chip, which may now contain many hundreds of thousands of nucleic acid probes that specifically bind to a RNA sequence, defines the genes to be examined through the composition of the chip's probes. Modern total RNA and miRNA chips exist for a range of species commonly used in research and can be used to examine the expression of tens of thousands of individual genes. Microarray technology can only be as good as the associated bioinformatic data for the species, which is why commercial microarrays tend to be limited to model and economically important organisms including human, mouse, pig and rat. The Affymetrix GeneChip Mouse Gene 2.0 ST Array (used in experimental work described in Chapter 5), contains >698,000 probes for the detection of mRNA and long intergenic non-coding RNA (lincRNA) transcripts derived from >33,000 genes (Affymetrix, Mouse Gene Array Datasheet). Specialised microarray systems also exist for miRNA analyses, such as the Toray miRNA Oligo Mouse Chip v.2.1 (used in experimental work described in Chapter 6), which provides coverage of 1,900 miRNA species (Toray, miRNA Oligo Chip Datasheet).

Microarray platforms utilise a pre-defined set of nucleic acid probes (single stranded DNA sequences) that are integrated onto a glass 'chip' in a grid-like pattern, either through the addition of in-tact sequences onto the glass or through direct synthesis of sequences into the array, for the quantification of RNA sequences in the sample. RNAs in the experimental sample are pre-labelled with fluorescent dye prior to exposure to the array. These fluorescently labelled RNA sequences in the experimental sample then hybridise to the complementary probes on the array and their abundance can then be measured through fluorescence that occurs under specific, high-stringency conditions. Multiple probes detecting the same transcript are contained within a single 'spot' on the microarray; each feature can then be specifically identified by its location on the microarray chip and the expression level derived from the intensity of fluorescence at that location.

There are a number of manufacturers of array technologies who provide arrays of varying cost and for different ranges of organisms, and provide manufacturer specific microarray scanners and reagents; amongst these are Affymetrix, Agilent, Illumina, Qiagen, and Toray (3D-Gene). The main advantages of microarray technologies over techniques such as Illumina RNA sequencing are reduced cost, ease of preparation and analysis of retrieved data and, in some circumstances, higher reproducibility and quantitation accuracy (Lowe et al. 2017, Mantione et al. 2014). A key factor for consideration when designing an experiment is the number of biological replications required in order to recover meaningful and informative data. Cost is a major variable in transcriptome analytics, the reduced cost of microarray often offers a sensible alternative to RNA sequencing wherein more biological replicates can be performed, generating a more robust dataset.

Microarray technologies do have a number of limitations including the lack of potential for novel transcript discovery, that the array itself can only ever be as good as the associated bioinformatic data,

the potential for topping-out (overloading the array’s probes), and the limited depth of coverage compared to other techniques.

RNA sequencing versus microarray

When designing experimental work, investigators are faced with the decision between a variety of transcriptome analysis platforms. As previously mentioned, each platform provides advantages and disadvantages for transcriptome analysis. The chosen platform should be suited to the technology and the context of the experimental work for which there are many considerations. Below is a selection of the advantages and disadvantages of the two major transcriptome analysis technologies currently in use, RNA sequencing and microarray (*Table 1.5*). Factors such as reproducibility were not included as a numerical figure as they rely heavily upon the appropriate application of the technique and on subsequent analysis, it is suggested that the reproducibility for both technologies *when applied optimally* is high, even between research groups (Lowe et al. 2017, Łabaj and Kreil 2016, ‘t Hoen et al. 2013).

	Sequencing technologies	
Factor	RNA sequencing	Microarray
RNA input amount	Low (around 1ng total RNA)	Higher (around 1µg total RNA)
Sample preparation	More complex (RNA extraction, library preparation)	Simple (RNA extraction)
Complexity of analysis	More complex	More simple
Novel transcript discovery	Yes	Not possible
Background signal	Very low	Can be high
Species range	No limitations	Model species only – arrays are single- or multi-species specific
Non-coding RNA	Yes – separate small RNA library preparation recommended	Yes - dedicated small non-coding RNA arrays
Time	More time required (additional sample preparation, sequencing run, more complex data analysis)	Quick
Cost	More per sample	Less per sample

Table 1.5. Comparison of RNA sequencing and microarray expression analysis platforms. This table includes some of the comparative advantages and disadvantages of two of the major gene expression technologies. This table was constructed based on information primarily from Lowe et al. 2017 and Wang et al. 2009.

In this thesis, the following transcriptome analysis platforms were used and are discussed in more detail in the corresponding chapters: Illumina Total RNA sequencing (Chapter 4) and Affymetrix GeneChip Mouse Gene 2.1 ST Array (Chapter 5) for the examination of total RNA, and Illumina Small RNA sequencing (Chapter 6) and Toray miRNA Oligo Mouse Chip v.2.1 (Chapter 6) for the examination of miRNAs.

1.7.Rational for Investigatory Work

1.7.1. Understanding mechanisms of meiotic drive: Fertility research and applications in agriculture

UN projections suggest that the global human population will have reached 9.7 billion by the year 2050, with much of that population growth in Africa, and more than half of the predicted population living in countries where food shortage is already an issue (The World Population Prospects, 2015). As the global human population continues to increase and the effects of global warming challenge future markets, it is of great economic and ethical importance that food pressures as well as the well-being of livestock continue to be met and improved in addition to ensuring agricultural sustainability. Consumed in moderation, animal products provide a wealth of benefits to the end user. In countries where much dietary energy is provided by cereals or roots crops, meats and other animal products provide a valuable source of dietary fats, vitamins, proteins and compounds for growth and long-term health, and will likely remain important in the future diet (Bohrer 2017, McColl et al. 2013, Pereira and Vicente 2013, Fahrenkrug et al. 2010, Rosegrant et al. 2009). In 1995, annual global meat consumption was 55,551K tonnes; in 2015, this reached 67,451K tonnes. As populations increase and diets in developing countries change, this is projected to reach 77,460K tonnes by the year 2025 (OECD-FAO Agricultural Outlook, 2016). Genetically engineered (GE) animals are likely to play an important part in the future of sustainable food production. Disease-resistance has been the main focus of GE livestock research but opportunities for improving efficiencies elsewhere within animal product production are clear.

Sex selection in livestock

The ability to select for X- or Y-bearing sperm would have a profound effect upon the efficiency of livestock production for human consumption. As an example, such capacities could allow for the preferential production of female livestock for the dairy industry and selection of male stock in beef cattle lines. Current technologies utilise the inherent and significant difference in length, and therefore DNA quantity, between the X- and Y-chromosomes to allow for the sex-selection of sperm using flow cytometry (Rath et al. 2013, Sharpe and Evans, 2009). At present, this is the only commercially viable method of pre-implantation sex selection available, resulting in a 'sort-purity' of >90% (Rath et al.

2013). However, there are a number of limitations to this method, many of which can be attributed to its requirement to examine the DNA content of each spermatozoon, restricting the number of spermatozoa that can be sorted to a maximum of 8000 per second under ideal conditions (equivalent to 2 conventional bovine artificial insemination (AI) doses per hour) (Rath et al. 2013, Sharpe and Evans, 2009).

In addition, this method is only feasible for bovine livestock; porcine livestock require between 1.5×10^9 and 5.0×10^9 sperm per intra-cervical AI dose (Alm et al. 2006) and >50 million sperm per dose for deep intra-uterine AI (Rath et al. 2013, Grossfeld et al. 2005), making flow cytometry sorting of sperm commercially non-viable. There are also various stresses placed on the sperm before, during and after the sorting process that may impact upon their physical and functional integrity. Furthermore, the fate of Hoescht dye used in the process is poorly understood (Garner 2009).

Gaining an understanding of the biological mechanisms of sex ratio variation is particularly relevant to pig and cattle farming where one sex is considered substantially more economically valuable than the other. The mouse provides a well understood and easy-to-study mammalian species for study of transmission ratio distortion. A number of models exhibiting variation in sex chromosome content and offspring sex ratios have been produced and maintained, and have contributed to our understanding of these sex-influencing phenomena. Current methods of sex selection in livestock rely on staining sperm DNA with potentially harmful intercalating dyes and flow-sorting under UV laser illumination, this method is only applicable to cattle and there is currently no viable method for offspring sex ratio control in the pig. If a native, non-shared, sex chromosome specific marker were identified, it could open the door to the development and refinement of effective sex selective techniques that rely on integral offspring sex-specific antigens (OSSAs) rather than the addition of poorly understood DNA dyes into the food chain. The development of improved methods of sex ratio control would contribute significantly to the advancement of global food security and could likely present major benefits for the UK and global economy.

1.8.Candidate Mechanisms and Genes

A review of the literature provides a number of candidate genes suitable for further investigation for their potential role in transmission ratio distortion (TRD) in the $MF1XY^{RIII}qdel$ model. Candidate genes are required to evade sharing across cytoplasmic bridges at both RNA transcript and protein levels, and result in a phenotypic asymmetry between sperm from the same ejaculate, either through their own actions or through regulatory actions of other genes. It is unknown how many genes might be involved in the hypothesised distorter-effector complex, however, all genes must meet the above criteria, and it is hypothesised that at least one component of this complex is X-linked, given the resulting sex ratio skew and known effects of $MSYq$ deletion on sex chromatin expression. These

include genes mentioned as potential instigators of transmission ratio distortion in other models and genes implicated in similar physiological aberrations in MSYq deletion strains. Selected candidates and genes of interest are discussed in detail below, other candidates are summarised in the final subsection.

Hyaluronidases and *Spam1* (Sperm adhesion molecule 1)

Spam1 (sperm adhesion molecule 1) is a single-chain glycosyl phosphatidylinositol- (GPI) linked sperm surface protein. It is one of six known members of the hyaluronidase family and is the major mammalian testicular hyaluronidase, expressed at high levels in sperm cells (Dunn and Mager 2005). It has roles in both the penetration of the cumulus layer of the oocyte (through hyaluronidase activity) and in secondary binding, the binding of acrosome-reacted sperm to the zona pellucida (Cherr et al. 2001, Myles and Primakoff 1997, Lin et al. 1994). In mouse, *Spam1* is located on chr.6 and its deficiency has been linked to impaired sperm function and a notable sex ratio skew in Robertsonian translocation models (Zheng and Martin-DeLeon 1999) (see section 1.4.2).

Zheng et al (2001a) were able to demonstrate that *Spam1* is haploid expressed and showed that both the mRNA and protein are compartmentalised. They implied that *Spam1* transcripts are not shared between spermatids, making it a suitable candidate mechanism for their model of transmission ratio distortion (TRD) (Zheng et al. 2001b). Subsequent papers have shown *Spam1* mRNA to localise to the endoplasmic reticulum in developing spermatids (Morales et al. 2004). In 2002, Baba et al. showed that *Spam1* null mice are fertile; the sperm are able to bind with and pass through the zona pellucida of the oocyte (Baba et al. 2002). It also showed that when in the presence of competition, sperm completely lacking or with reduced *Spam1* expression lose out to sperm possessing fully functional forms. In a competitive fertilisation assay, 75% of cumulus intact oocytes were fertilised by wild-type sperm over *Spam1*^{-/-} sperm. This demonstrated that although *Spam1* null mice are competent to fertilise, they perform poorly in competition with *Spam1* competent sperm. This is amplified in cases of additional *Acr* (Acrosin) and/or *Prss21* (Protease, serine 21) loss (Zhou et al. 2012). The presence of SPAM1 protein in the female reproductive tract and subsequent uptake by sperm has also complicated its examination and role in fertility (Martin-DeLeon 2006, Zhang and Martin-DeLeon 2003).

Production and acquisition of SPAM1 protein is not exclusively intrinsic to the individual sperm - sperm are able to acquire external SPAM1 proteins during transit through the epididymis and through the estrous uterus and oviduct. Secretory SPAM1 can be found in all three main regions of the epididymis, and it has been shown that this bestows even *Spam1* homozygous knockout mouse (-/-) sperm with SPAM1-mediated hyaluronidase activity (Chen et al. 2006). In a report by Griffiths et al. (2008) they demonstrate the ability of individual sperm to bind secreted SPAM1 proteins found in the estrous uterus and oviduct to their surface during sperm capacitation. A reduction in *Spam1* activity,

potentially mediated by an X-linked distorter, and limited by non-sharing of transcripts to a sub-population of cells, could provide a functional disadvantage to Y-linked sperm in the MF1XY^{RMII}qdel model in the form of reduced hyaluronic acid binding ability and decreased rate of cumulus dispersal. To investigate this possibility, hyaluronidase activity assessments and examination of transcript distribution in sequencing experiments were conducted.

Progesterone Receptors - *Pgrmc1* (progesterone receptor membrane component 1)

Work by Kotarska et al. (2014 and 2013) has revealed zona-thickening in B10.BR(Ydel) females sired by B10.BR-Ydel (2/3 Yq deletion) males to be caused by increased amounts of progesterone synthesized by the oocyte cumulus. Progesterone in the female genital tract stimulates capacitation, motility, the acrosome reaction and acts as a sperm chemoattractant. Variation in female reproductive features between the B10.BR and B10.BR(Ydel) mice can be explained only “by indirect influence of paternal Y chromosomes that participate in the establishment of heritable epigenetic markers in X-bearing sperm” (Kotarska et al. 2014), as the two strains are otherwise genetically identical. These features resulted in impaired fertilisation ability of B10.BR(Ydel) oocytes by B10.BR-Ydel sperm but not B10.BR controls, shown to be due to decreased functional progesterone receptors present on the surface of B10.BR-Ydel sperm. Whole gene and mRNA expression analysis revealed significant increases in the level of *Hsd3b1* gene and *Cyp19a1* mRNA expression in B10.BR(Ydel) females in comparison to control females. These two genes are involved in two separate stages of cumulus progesterone production.

The origins of suggested epigenetic influences from the Y-chromosome are yet to be fully understood. It appears the effects of zona-thickening do not result in an exaggerated sex-ratio skew, but does reveal some of the symptomatic links to partial Yq deletion. This presents a further avenue for investigation of the interactions between the Yq deletion and resulting phenotypic changes.

Other genes of interest from the literature

Tcp11x2 (T-complex 11 X-linked protein 2) is a gene mentioned in Cocquet et al. 2012 as potentially involved in reproductive aberrations resulting from loss of *Sly*. *Tcp11x2* is the homologue to the human T-complex 11 X-linked protein 2, also known as NXF2 – nuclear RNA transport factor 2. The human homologue binds RNA, localises to nuclear envelope and exhibits RNA export activity. This protein could be implicated in altered localisation of RNA transcripts in YQ deletions. In addition, *Tcp11*, homologue of *Tcp11x2*, is an autosomal gene implicated in transmission ratio distortion seen in the T-complex (Safronova et al. 2002, Fraser and Dudley 1999). TCP-11 is found on the surface of mature sperm and is thought to play a role in capacitation and inhibition of spontaneous acrosome reaction (Fraser et al. 1997).

SLY protein has been shown to interact with DKKL1, KAT5 and APPBP2 through coimmunoprecipitation (Reynard et al. 2009). Although reduced presence of SLY in Yq deletions is unlikely to lead to directly altered transcription of these genes, these genes are involved in vital sperm development pathways and the reduced presence of *Sly* may result in subsequent alteration of these pathways. APPBP2 in particular is involved in the intracellular trafficking of proteins. DKKL1 gene family members are antagonists of the Wnt signalling pathway, which in various incarnations is responsible for signal transduction in regulation of the cytoskeleton and therefore cell morphology, and intracellular calcium regulation (review by Lai et al. 2009).

In addition to localisation with the PMSC, SLY1 protein has also been found to co-localise with the *Speer* (Spermatogenesis-associated glutamate (E)-rich protein) gene cluster on chromosome 5 in post-meiotic spermatids (Cocquet et al. 2012). Much like with PMSC, in the absence of SLY, SLX/SLXL1 proteins were also shown to localise with *Speer* (Cocquet et al. 2012). In mice lacking the full complement of *Sly* gene copies, *Speer* expression is increased. The *Speer* cluster represents one of the largest rodent-specific autosomal expansions; this and SLY/SLX localisation indicates its expansion may be heavily influenced by the intragenomic conflict between *Slx* and *Sly* and may represent a mediator or gene of important reproductive value in MSYq deletions (Cocquet et al. 2012).

The *Alkbh* family contains highly conserved members identified throughout the animal kingdom that are involved in DNA repair. *Alkbh1* has been implicated in sex ratio distortion (in favour of males) in mice alongside transmission ratio distortion of offspring carrying the hemi- or homozygous deletion for the gene; suggesting an associated inheritance or expression of the X-chromosome and this gene (Nordstrand et al. 2010). Although it is suggested that the deletion primarily affects development during the pachytene spermatocyte stage, there may be post-meiotic effects. The *Alkbh* family contains multiple members that may exhibit similar characteristics.

A vast number of other genes are associated with traits such as abnormal sperm head formation that have been recorded in MSYq mice. Any number of these genes may have roles to play in the altered reproductive traits seen, however there are too many to sensibly address individually. Genes shortlisted in expression analysis studies were instead evaluated for links to such traits and commented on appropriately in those chapters.

1.9. Structure of the Thesis

The thesis outlines discoveries made and steps taken in the investigation of the offspring sex ratio skew and associated physiological and molecular characteristics relating to a two-thirds MSYq deletion in the MFIXY^{RIII}qdel mouse model. The work described in this thesis includes physiological assays and sequencing techniques to identify and further investigate potential gene candidates for their

involvement in the offspring sex ratio skew and associated phenotypes. All techniques and materials used are described in Chapter 2 – Materials and Methods.

Chapter 3 examines multiple physiological effects of the MSYq deletion. This chapter thoroughly examines the offspring sex ratios seen in the MF1XY^{RIII}qdel and conspecific MF1XY^{RIII} models over the four year period of the investigation. The chapter employs traditional techniques for the qualitative examination of sperm morphology and a novel technique for the quantitative examination of sperm morphology and associated sex chromosome content. This novel technique ties the examination of sex chromosome content of individual sperm to their morphology for the first time. Sperm hyaluronidase activity, an important sperm function parameter, is also examined. This chapter directs further investigation by defining physiological effects of the MSYq deletion.

Chapters 4, 5 and 6 employ RNA samples generated from testicular extracts of both MF1XY^{RIII} and MF1XY^{RIII}qdel mice to perform a detailed examination of spermatid gene expression. Whole testis samples were elutriated, to enrich specific cell populations, and subsequently fractionated into subcellular compartments using graded cellular lysis.

Chapter 4 utilises RNA sequencing of total RNA to examine RNA transcript distribution amongst and within cellular populations and subcellular fractions. This broad sweep of the transcriptome allows for comparison within and between these populations and the MF1XY^{RIII} and MF1XY^{RIII}qdel strains to identify potentially differentially expressed or highly compartmentalised transcripts that may be involved in causing or facilitating the offspring sex ratio skew. This chapter also discusses the production and initial use of candidate gene identification and development of data analysis techniques.

Chapter 5 discusses a microarray experiment designed to examine the transcriptome with multiple replicates. The different technologies (RNA sequencing and microarray) are briefly compared and differential expression results discussed. A separate but complementary candidate gene list is generated using this analysis. Following this, a comparison of the results of the two expression analysis chapters allowed for the refinement and further exploration of selected candidates.

Chapter 6 utilises both small RNA sequencing and miRNA microarray technologies to identify interacting miRNA partners of previously short-listed protein-coding genes and examine the potential influence of epigenetic mechanisms on the physiological outcomes seen in the MF1XY^{RIII}qdel model.

Chapter 7 summarises the results of the work presented in chapters 1 through 6 and discusses the potential implications of the findings alongside suggestions and direction for future work.

Chapter 2

2. Materials and Methods

This chapter covers the general methods used in experimental work throughout the duration of the project. Where individual conditions or procedures vary from the below, this is noted in the corresponding experimental chapter.

2.1. Animal Care and Preparation of Tissues, Cells and Fractions

All animals used in investigatory work outlined in this thesis were housed and cared for at either: The Department of Physiology, Development and Neuroscience, Cambridge; or, The Biological Sciences Unit, Department of Pathology, University of Cambridge; or, Central Biomedical Services, Addenbrooke's Hospital Site, Hills Rd, Cambridge.

All animal procedures were in accordance with the United Kingdom Animal Scientific Procedures Act 1986 and were subject to local ethical review. The work was performed under Home Office Project licenses 80/2451 and 70/8925 (PPL holder: Dr Peter Ellis). Animals were culled using a Schedule 1 approved method (cervical dislocation), and all cells and tissues were prepared in the manner described in the below sections unless otherwise stated in the associated experimental subsection.

2.1.1. Extraction of mature sperm cells

The vas deferens and cauda were extracted from adult male mice by dissection. Following extraction of the vas deferens and cauda, mature sperm were collected into a petri dish (Thermo Fisher Scientific, USA) by immersing both regions in 1ml of Dulbecco's phosphate buffered saline (PBS) (Sigma, USA) warmed to 35°C on a heat block. Where the caudal sperm were required, two large cuts were made across the cauda using a razor blade and sperm were allowed to swim out for five minutes, after which the tissue was moved up and down in the PBS and gentle pressure applied to encourage sperm to exit the tissue. Where the caudal sperm were not required, the cauda was removed and the sperm extracted from the vas deferens by securing one end in place using tweezers and, using a pair of precision angle tweezers, applying gentle progressing pressure along the length of the duct. The debris (evacuated vas deferens) was then removed from the collected sample and disposed of. The mature sperm sample is then placed in a 1.5ml eppendorf (Starlab, UK). The sample was spun down in a microcentrifuge at 700g for 3 minutes, the supernatant removed and the sperm resuspended in fresh, warmed (35°C) PBS. This PBS wash is repeated, following which sperm are ready to use.

2.1.2. Extraction of testis cells and preparation of cellular suspension for elutriation

Centrifugal elutriation provides a means of separating live testis cells into differing populations by size. Testes were first extracted from adult male mice by dissection, and placed in 1ml of Dulbecco's modified eagle medium (DMEM) (Sigma, USA). The tunica albuginea is removed from the testes by, first, careful piercing of the layer using micro dissecting scissors. The layer is then held firmly at one end of the testis and an even, firm pressure is applied down the length of the testis using precision angle tweezers to extract the testis contents (seminiferous tubules). The tunica albuginea is then disposed of. The testes extract is then finely minced using a disposable razor blade.

DMEM 0.5% bovine serum albumin (BSA) (Sigma-aldrich, USA) was prepared on the day prior to actual elutriation of cells by adding 2.5g of BSA to 500ml of DMEM and shaking vigorously for 30 seconds. The mix was then left overnight at 4°C to allow the BSA to fully dissolve. Following extraction as described above, the minced testes extract was then resuspended in 18ml of fresh DMEM in a 50ml falcon tube (Greiner). Trypsin-EDTA (Sigma, USA) is added at 2.5mg/ml to facilitate cell dissociation, alongside RNase-free DNase I (Thermo Fisher Scientific, USA) at 50µg/ml to remove DNA. Trypsin functions to dissociate cell clusters (syncytia) formed between synchronously developing male germ cells connected by intercellular bridges. It also reduces the incidence of multinucleated cells (symplasts), which might otherwise result from the opening up of these bridges (Romrell et al. 1976). The suspension was then incubated in a water bath at 31°C for thirty minutes with stirring or manual agitation every two minutes, to increase the activity of both trypsin-EDTA and DNase. After thirty minutes, trypsin activity was quenched using foetal bovine serum (Thermo Fisher Scientific, USA) at a final concentration of 8%. The cell suspension was then passed through a 70µm nylon cell strainer (Falcon, USA) to remove larger debris and cellular clumps that have failed to dissociate. After being passed through the cell strainer, the suspension was centrifuged at 500g for fifteen minutes at room temperature to form a cell pellet. After centrifugation, the supernatant was removed and the cells were gently resuspended in 20ml of DMEM 0.5% BSA + 50µg/ml DNase I. The suspension was kept on ice from this point onwards. A small aliquot of pre-elutriation cell suspension was also kept for subsequent examination under the microscope.

2.1.3. Counting of cells using a haemocytometer and trypan blue staining

Total cell counts and cell concentrations can be estimated using a haemocytometer and an appropriate dye. In this instance, trypan blue (Sigma), for the staining of dead cells, was used as follows.

Cells were collected by either dissection (mature sperm, whole testis extract) or elutriation (spermatids). When using live cells, the cells were first combined with 4% paraformaldehyde (PFA)

in a 1:1 ratio (10µl cell suspension + 10µl 4% PFA) to kill cells for better absorption of the trypan blue stain and to immobilise spermatozoa. Mature sperm extracts were further diluted by adding 80µl of PBS, as the direct extract concentrations resulted in overcrowding of the haemocytometer. This is accounted for using a dilution factor (df) in the final calculation.

Cells	Fixation and dilution	Trypan blue	Dilution factor (df)
Mature sperm	10µl cell suspension + 10µl 4% PFA + 80µl PBS	100µl trypan blue	10
Other cells	10µl cell suspension + 10µl 4% PFA	20µl trypan blue	2

Table 2.1. Composition of mixture used for cell counting in a haemocytometer.

The cell/PFA mixture was then combined with trypan blue in a 1:1 ratio and left for five minutes at room temperature. Dampening the edge of the haemocytometer with a minute drop of water and placing a 22mm x 22mm coverslip directly over the counting chamber allowed for surface tension to develop and hold the coverslip in place. Newton’s rings are visible when the cover slip is properly secured. When placed at the edge of the chamber using a pipette, 20µl of the cell sample were immediately drawn into the 0.1mm deep chamber that was created by the cover slip. This was then allowed to settle at room temperature for two minutes. The cells were then visualised using a light microscope and all cells within the gridded counting chamber counted towards the total. The grid is divided into twenty five small squares, which make up one large square. A minimum of one hundred cells were counted to produce an estimation of original concentration using the calculations below.

n = cell count

s = number of squares counted

df = dilution factor

Addition of trypan is accounted for by doubling the result (included in calculation)

Using large squares: $2 \times n/s \times 10,000 \times df = \text{cells/ml}$

Using small squares: $2 \times n/s \times 250,000 \times df = \text{cells/ml}$

This gives a final cell concentration in cells per ml.

2.2.Morphology and Microscopy Methods

Qualitative sperm morphology assessment was preformed through visual examination and ranking of individual sperm heads into head morphology categories (see *Table 2.2*). Mature sperm were collected

independently from age-matched pairs of MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ) mice as described in section 2.1.1. Sperm suspensions were then dropped onto slides and stained using silver nitrate (see below). A ranked count was then compared between age-matched pairs and between genotypes as a whole.

2.2.1. Silver nitrate staining of sperm

Whilst numerous different staining techniques exist, silver nitrate staining of sperm has been shown to provide high resolution viewing and easy identification of sperm structures (Banaszewska et al. 2015, Andraszek and Smalec 2011, Bongso 1983), and can be easily examined using a simple light microscope. Examination of sperm heads was performed by one individual only.

The following solutions are required for silver staining of sperm:

Solution 1 (can be stored at room temperature)

2% gelatine (dissolved at 60°C with stirring) + 1% formic acid

Solution 2 (must be made fresh)

50% silver nitrate (0.5g in 1ml water)

Mature sperm were extracted as described in section 2.1.1. 10µl of sperm extract was placed onto a superfrost microscope slide (ThermoFisher) and smeared across the length using the edge of another slide. To fix the cells, a few drops of Methanol/acetic acid fix were placed on the slide and allowed to air dry for two minutes. The slides were then rinsed with 0.4% PhotoFlo (Kodak, UK) to clean the slide; this was then allowed to air dry. The slides may be stored at this point for later use.

Solution 1 and 2 were mixed on a 50 x 22mm cover slip, and the fixed sperm slide placed gently on top, avoiding trapping any air bubbles. The slides were then placed on a heat block set to 60°C until a golden brown colour developed (after about 30 to 45 seconds). The slides were finally rinsed under running cold water and placed on a slide rack to air dry.

2.2.2. Categorisation of sperm heads by shape

Following silver nitrate staining of sperm as previously described, sperm were visualised under a light microscope and individual sperm heads placed into one of five distinct morphological categories. These categories are described in *Table 2.2*, below.

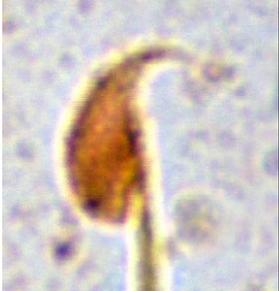
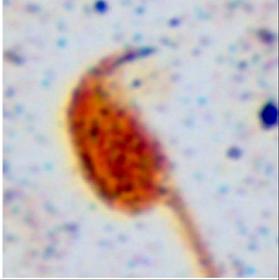
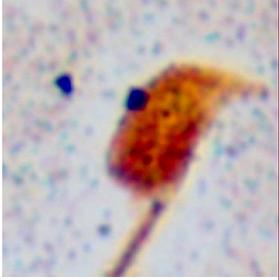
Head morphology	Description	Example
1 - Normal	Displays a smooth curved edge ending in a prominent curved hook. The postacrosomal segment appears to run from the equatorial segment to the body of the sperm in a smooth arc.	
2 - Flattened	Appears shortened in comparison to a normal sperm head, with the postacrosomal segment appearing distinctly flattened. Some straightening and/or shortening of the apical hook.	
3 - Severely flattened	Increased severity of characteristics seen in flattened head alongside flattening of the acrosomal cap, further reduced length and increasing perpendicularity of the apical hook to the sperm head. Sperm head appears more box-like than curved.	
4 - Brick-like	Sperm head appears notably brick-like. Severe reduction in length of the apical hook compared to normal morphology.	
5 - Other (extreme abnormality)	All other massive deformities	

Table 2.2. Categories used to define individual sperm head shapes within sperm populations. The name of the category (reflected in the results chapter, Chapter 3), a brief description of the sperm head shape and an example sperm head are given for each category. These categories were defined by the author.

2.2.3. Quantitative sperm head morphology methods

Sperm head shape was assessed using a new quantitative method. By combining techniques for imaging for morphology, the recording of individual cell locations, sperm fluorescence *in situ* hybridisation (sperm FISH), and nuclear morphology analysis software (Nuclear Morphology Analysis Software 1.13.6), it has been possible to discern changes in nuclear morphology of individual sperm heads and between populations, including differences between cells with differing sex chromosome content.

Sperm samples were extracted and prepared as described in section 2.1.1. Sperm samples were then fixed drop-wise in 3:1 methanol/acetic acid, up to a volume of 1ml. After thorough resuspension of the samples, sperm suspensions were then dropped by pipette (Gilson) onto polylysine slides. Initially, 10µl of methanol / acetic acid (Me:A) fixed sperm suspension was used at approximately 1.4×10^7 cells/ml concentration; these were allowed to air dry. Slides were then examined under phase to assess number and spread of cells. If required, either more sperm suspension (to increase sperm cell count) or Me:A (to increase spread) was added to achieve optimal distribution of cells for morphological assessment. Once the slides were dry, 2x 8µl DAPI (Vectashield) was added to either end of the slides, directly on top of the dropped sperm area. A 55 x 20 mm coverslip was then carefully laid on top, and slides were placed upside down on absorbent paper to remove excess DAPI. Slides were then placed at 4°C for a minimum of 10 minutes to allow the DAPI to fully penetrate the nucleus. After 10 minutes, the slides were ready for imaging.

Sperm heads were visualised at 100x under DAPI using an Olympus BX61 epi-fluorescence microscope equipped with a cooled CCD and appropriate filters (Griffin Laboratory, University of Kent). The system's motorised stage facilitates capture and re-capture of individual sperm heads pre- and post-FISH (fluorescence *in situ* hybridisation). Capture of images was conducted using SmartCapture 3 for Apple Mac (Digital Scientific, UK).

2.2.4. Sperm fluorescence *in situ* hybridisation (sperm FISH)

Fluorescence *in situ* hybridisation (FISH) was used to determine the sex chromosome content in mature sperm from both MF1XY^{RIII} and MF1XY^{RIII}qdel mice. The following solutions are required for successful sperm FISH:

Swelling solution:

Tris-HCl pH 7 0.1M

DTT 10mM

Permeabilisation solution:

Igepal 0.5%

Triton-X-100 0.5%

ddH₂O 99%

Pepsin solution:

HCl 0.01N

Pepsin 0.1mg/ml

Pepsin should be added once the solution has warmed, and immediately prior to use.

Probe mix (per slide):

3µl Murine whole chromosome X-paint - green (Cytocell, UK)

2µl Murine whole chromosome Y-paint - red (Cytocell, UK)

This mix should be kept in the dark as the paints are light-sensitive.

The first section of this protocol was adapted and optimised over many years from sperm FISH in Finch et al. 2008.

DAPI pre-staining for morphological examination of sperm heads was conducted as detailed in section 2.2.3. After initial image capture and recording of slide positions, slides were submerged in 2x saline sodium citrate buffer (SSC) to allow coverslips to float off. The slides were then washed in a further two rounds of 2x SSC for five minutes each time. The slides were placed through an ethanol series, two minutes in each of 70%, 80% and 100% ethanol. After air drying, slides were aged on a hotplate at 70°C for one hour. To swell the sperm heads and allow the probes to penetrate the highly condensed nucleus, slides were immersed in swelling solution in a fume hood for thirty minutes in the dark and at room temperature. Once swelled, slides are placed in 70% formamide heated to 75°C in a water bath; this allows for the denaturation of the sperm cell DNA in order that probes may bind successfully. Slides were then rinsed in 2x SSC, before being dehydrated in an ethanol series as previously described. Pepsin solution was warmed to 37°C and slides were immersed for twenty minutes to digest proteins, and consequently cytoplasm, and to improve probe penetration to the nucleus. After rinsing slides in 2x SSC, they were then placed in permeabilisation solution (on ice) for thirty minutes at 4°C. Slides were again rinsed in 2x SSC and dehydrated in an ethanol series at room temperature. To denature the slides in preparation for the addition of probes, both slides and probe mixture (wrapped in foil) were placed on a heat block or ThermoBrite (Abbott Molecular, USA) at 70°C for five minutes before slides alone were submerged in 70% formamide (in a fume hood) at 75°C for five minutes only. Slides were then immediately transferred to ice-cold 70% ethanol (in a fume hood) for two minutes, and the ethanol series completed as before with two minutes in each of 80% and 100% ethanol at room temperature. Once the slides were dry, 5µl of probe mixture was added to each slide (directly onto the region in which cells were most concentrated). Small 13mm diameter circular coverslips were added to the probe area and the area was carefully sealed with

rubber cement. Due to the addition of probes, slides were protected from light from this point onwards. After incubation at 75°C for two minutes, slides were placed in a moisture chamber overnight at 37°C to allow probes to hybridise.

The second day (post-hybridisation) protocol was adapted from Cytocell Murine Whole Chromosome Painting Probes, Instructions for Use (Cytocell, UK), 'Post-Hybridisation Washes'. After 24 hours incubation, slides were recovered from the moisture chamber and the rubber cement carefully removed using precision tweezers. Coverslips were gently removed and slides submerged in 0.4x SSC at 72°C for two minutes undisturbed. Slides were then moved to 2x SSC with 0.05% Tween-20 (Sigma, USA) for thirty seconds to remove any remaining probe mix. DAPI (Vectashield) was then applied as a counterstain; DAPI emits blue range fluorescence upon binding with AT regions of DNA, and so is extremely useful in highlighting the nuclear area of the cell. 8µl is added to opposite ends of the slide and a 55 x 22 mm coverslip placed carefully on top. Slides are then inverted on a piece of absorbent towel to remove any excess DAPI. If prolonged storage is required after visualisation, clear nail varnish can be added around the edge of the cover slip to prevent the slide drying out. After being stored at 4°C in the dark for ten minutes, slides can be viewed using fluorescence microscopy.

Images were captured on an Olympus BX61 epi-fluorescence microscope equipped with a cooled CCD and TexR, Fitc and DAPI optics, using SmartCapture 3 for Apple Mac (Digital Scientific, UK). Slide positions that were previously stored following DAPI image capture were loaded for accurate recapture of the same individual sperm heads post-FISH. Images were then processed as described in Chapter 3 section 3.4.3.

2.2.5. Fluorescence *in situ* hybridisation (FISH) of metaphase cells

This protocol was adapted from the Cytocell (Cambridge, UK) FISH protocol², and utilises murine X and Y whole chromosome paints. Metaphase cell suspensions produced from spleen cell culture were fixed in methanol/acetic acid (3:1) (Me:A). Five microlitres of suspension was dropped on a polylysine slide. A clean drop of Me:A was added to encourage dispersal and slower drying. Slides were checked under phase microscopy to assess the number of available metaphases. Slides were dehydrated in an ethanol series (70%, 85%, and 100%), each for two minutes at room temperature and allowed to dry. Slides were then aged overnight at 37°C. Following aging of slides, probe solution was prepared using 3µl X-paint and 2µl Y-paint per slide. The probe mixture and sample slides were then warmed on a hot plate at 37°C for five minutes. The required amount of probe mixture was placed onto the slide and a circular coverslip carefully applied to the area of interest. This was then sealed with rubber solution glue and allowed to dry completely in the dark, whilst still on the hotplate. Samples and probes were denatured simultaneously by heating the slide on a hotplate at 75°C for 2

² Cytocell FISH protocol - http://www.cytocell.com/assets/000/001/165/AMP_R-G_v002.00_original.pdf?1501237370

minutes in the dark. The slides were then placed in a humid, lightproof container at 37°C for between 4 and 24 hours. Following incubation, all traces of glue were carefully removed and slides were immersed in 0.4x SSC (pH 7.0) at 72°C for 2 minutes without agitation. After which slides were drained and submerged in 2x SSC with 0.05% tween-20 at room temperature (pH 7.0) for 30 seconds without agitation. Following which, 16µl DAPI (Vectashield) was applied to the slide (8µl at opposing sides of the test area). Slides were then sealed with a coverslip and clear nail varnish applied around the edge. The slides were stored at 4°C in the dark for a minimum of 10 minutes prior to viewing. Results of metaphase FISH can be seen in section 1.5.3.

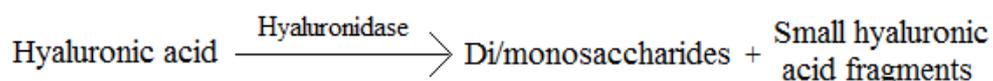
2.3. Other Physiological Assessment Methods

2.3.1. Assessment of sperm hyaluronidase activity

In order to assess levels of active hyaluronidase enzyme present in the sperm/on the sperm surface, a turbidimetric assay was devised, utilising spectrophotometry to determine levels of precipitated hyaluronic acid remaining post-enzymatic action. The protocol was adapted from methodologies from Deng *et al* (1999), Thaler & Cardullo (1995), and Sigma Aldrich “Enzymatic Assay of Hyaluronidase (3.2.1.35)”.

Sperm (or pure hyaluronidase enzyme) were first incubated with hyaluronic acid. This is the reaction to be indirectly measured. A second incubation of the resulting mixture with bovine serum albumin (BSA) allows measureable turbidity to develop. The resulting level of turbidity is dependent upon the amount of uncatalysed hyaluronic acid remaining in the solution. This turbidity is measured by a spectrophotometer that gives a measure of absorbance; this is converted to a value of transmittance for the production of an enzyme action curve. Enzyme activity in sperm can be estimated by using transmittance values from sperm and HA reactions to determine equivalent enzyme action values using the previously produced enzyme action curve. This is explained in more detail below.

Under the correct conditions and in the presence of active hyaluronidase, the following reaction takes place:



In the presence of hyaluronic acid (HA), bovine serum albumin (BSA) forms protein aggregates, which result in greater turbidity of the solution. This does not occur when BSA comes into contact with HA derived substrates formed following the enzymatic action of hyaluronidases on HA. Turbidity can be measured using a spectrophotometer to give an indirect measure of enzyme activity.

The following media/buffers were prepared for the purposes of the hyaluronidase assay:

Enzyme-only hyaluronidase preparation:

Hyaluronidase standard (Sigma, H4272-30mg) at 10mg/ml in QH₂O

Hyaluronic acid (HA) substrate solution:

Hyaluronic acid (Sigma, H5388-100mg) at 0.3mg/ml in 0.3M Sodium phosphate buffer (pH 5.35)

HA polymer may take time to dissolve, so it is best prepared the day prior to the assay, or it may be warmed in a 90°C water bath (not boiled) to encourage dissolution, and allowed to cool before use.

Acidified bovine serum albumin (BSA) solution:

Bovine serum albumin (BSA) at 1mg/ml in 0.5M Sodium acetate buffer (pH 3.75)

Protocol A - Creating an enzyme activity curve:

The initial objective was to create an enzyme curve of varied and known concentrations with which to compare subsequent absorbance/transmittance results from assays using sperm. In later experiments, sperm were assayed following extraction from age-matched mice of either genotype.

Commercially available hyaluronidase (Sigma H4272) was combined with ultrapure water to form a 1280u per 100µl standard as per instructions. This was then diluted in human tubal fluid (HTF) as required to solutions ranging from 2.0u/100µl to 0.125u/100µl (a negative control was also included – see table below), creating a set of enzyme standard curve samples. 125µl of each enzyme standard curve sample was added, separately, to 125µl of hyaluronic acid substrate solution and incubated in a water bath at 37°C for one hour. Primary incubation facilitates the breakdown of the hyaluronic acid (HA at 0.3mg/ml in sodium phosphate buffer) by any hyaluronidase present in solution by incubating the HA and sperm or enzyme (at known concentrations) for a fixed period of time. Subsequently the samples were removed from incubation and spun down to remove cell debris, after which 200µl of supernatant was placed in a clean 1.5ml eppendorf and 800µl of warmed (37.5°C) BSA added. The remaining hyaluronic acid in solution is co-precipitated with acidified BSA (800µl at 1mg/ml in 0.5M sodium acetate, pH 3.5) at a temperature of 37.5°C for 30 minutes for the secondary incubation period, allowing measurable turbidity to develop. Optimisation of primary and secondary incubation times is discussed in section 3.5.2. The solutions were placed into individual cuvettes and absorbance measured at 600nm in a spectrophotometer. Appropriate blank measurements were used. Each enzyme-only experiment was duplicated from the point of generating different dilutions for enzyme standard curve samples onwards, in order to produce a set of two replicate enzyme curves for each experiment.

Enzyme curve sample number	Enzyme concentration U/100µl
1	2.000
2	1.500
3	1.000
4	0.750
5	0.500
6	0.250
7	0.125
8	0.000

Table 2.3. Selection of enzyme concentrations used to generate an enzyme activity curve

Protocol B – Assessing hyaluronidase activity in sperm:

One age-matched pair of mice, of one animal from each examined strain (MF1XY^{RIII} and MF1XY^{RIII}qdel), were sacrificed as described in section 2.1. Sperm from the cauda and vas deferens was then extracted into 1ml warmed (35°C) PBS as described in section 2.1.1. The extract was then centrifuged at 500g for two minutes to pellet any remaining tissue fragments and debris. The supernatant (containing sperm) was retained and 10µl taken for counting in a haemocytometer as described in section 2.1.3. The sperm suspension was then diluted in warmed PBS into three populations at the following concentrations (see table below), and all were kept at 31°C until needed. This was repeated once to produce two replicates for each sperm sample. Chosen cell concentrations were optimised to provide a range of transmittance values that lay within the linear part of the enzyme curve (see section 3.5.2 for further detail). Sperm suspensions were then treated in much the same way as the hyaluronidase standard, beginning with primary incubation. Enzyme-only and sperm assays were run concurrently.

Concentration abbreviation	Cell concentration, cells/100µl
High	1 x 10 ⁶
Medium	5 x 10 ⁵
Low	2.5 x 10 ⁵

Transmittance values for each sample were then calculated as follow:

$$\text{Transmittance} = 10^{-\text{absorbance}}$$

An enzyme curve was produced, plotting enzyme concentration (x axis) against transmittance value (y-axis) to produce an enzyme activity curve. Using the rearranged linear equation derived from the linear part of the enzyme curve (enzyme concentrations from 0.125u to 1.0 u/100µl), equivalent enzyme concentration values were determined for sperm-assay-derived transmittance values. Giving a range of enzyme activity values for each sperm sample assessed.

2.4.Preparation of Cells and Subcellular Fractions for RNA Sequencing and Microarray Gene Expression Analysis

In order to obtain RNA from various subcellular fractions of different spermatid populations, testicular extracts were processed using the techniques outlined in this section. A diagrammatic overview is given in *Figure 2.1*. After processing of testicular tissue as outlined in section 2.1.2, spermatid populations were enriched by size (section 2.4.1), then fractionated into subcellular compartments (section 2.4.2), after which RNA was extracted (section 2.5.1) and RNA libraries prepared as required.

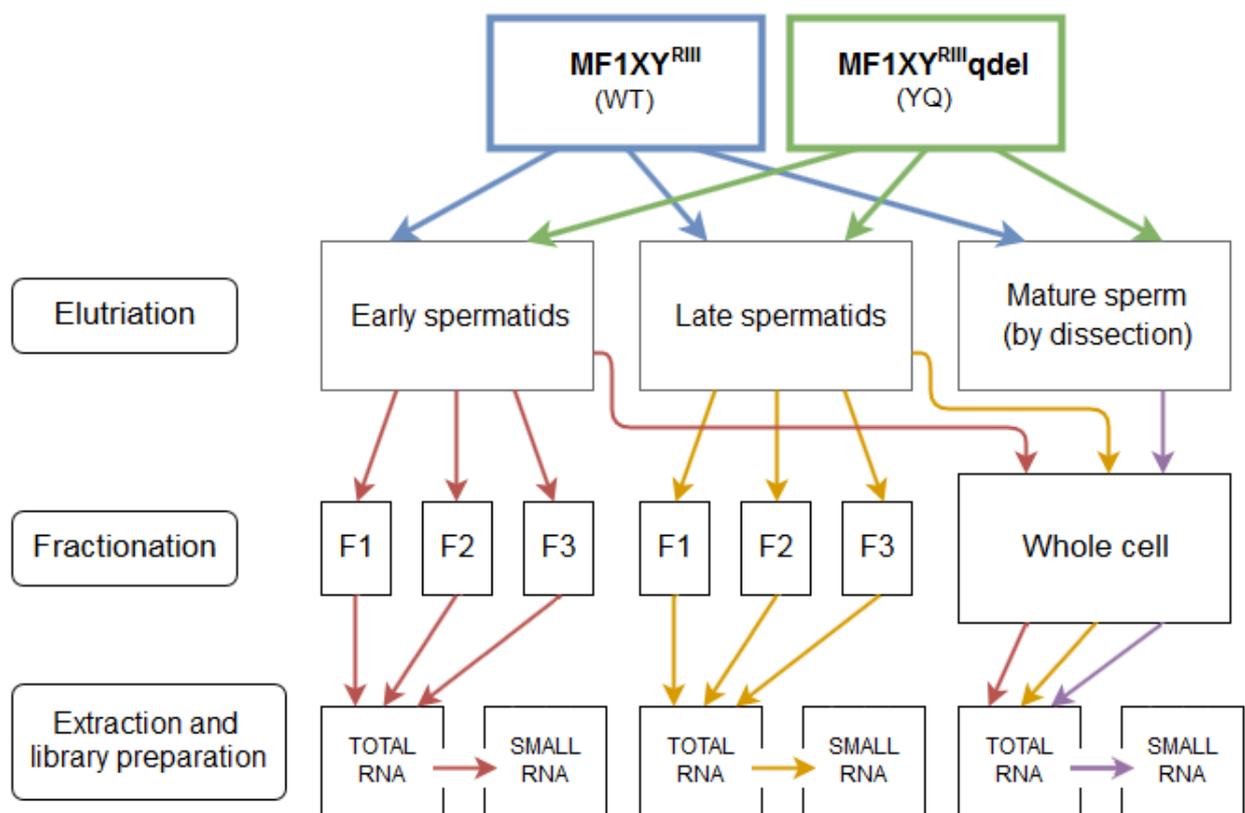


Figure 2.1. Flow chart overview of sample preparation for both RNA seq and microarray. Whole testes were taken from MF1XY^{RIII} and MF1XY^{RIII}qdel mice, cells were then elutriated into different cellular populations, fractionated into subcellular regions, and finally the RNA extracted and libraries prepared. Mature sperm cells were collected by dissection.

2.4.1. Separation of testis cell types by centrifugal elutriation

The Beckmann Coulter J26-XP Centrifugal Elutriator with the JE-5.0 Elutriator Rotor was used for the separation of living cells from a suspension by size. The centrifugal elutriator functions by passing the cell suspension through channels leading to a rapidly rotating elutriation chamber, where cells gather until the combined flow rate and centrifugal force reaches a point at which they will be ejected. This point differs for different sizes of cells. The set up includes a pressure gauge, to ensure that pressure in chamber (which is sealed except for one input and one output point) does not rise above the acceptable level. The flow rate is controlled through the MasterFlex LS Pump Drive. The pump and elutriator are calibrated before every run, by timing the collection of set volumes of water run through the centrifugal elutriator.

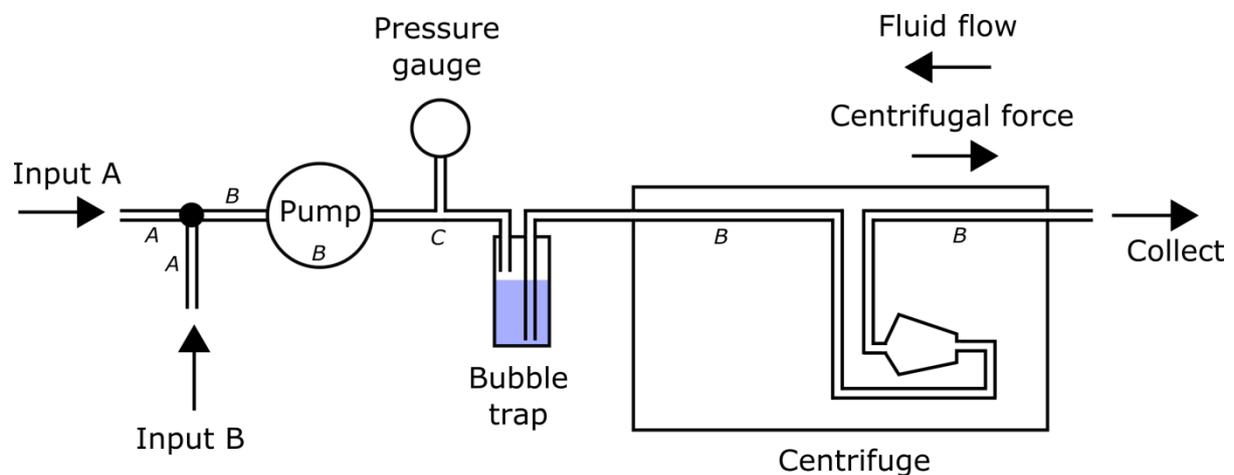


Figure 2.2. Simplified diagram of Beckmann Coulter J26-XP Centrifugal Elutriator set up. Letters in italic indicate the channel type used: *A* = Tygen 3350 1/8 x 1/4; *B* = MasterFlex 96400-16; *C* = Tygen 3-44-4X meets NSF-51 STD (stiff). Image created by Dr Ben Skinner (University of Cambridge).

A large (2 litre) container of DMEM 0.5% BSA was prepared and kept on ice during this process. The DMEM 0.5% BSA is used as an elutriation media. Separate suspensions of MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ) testes cells were generated as per section 2.1.2. The two suspensions were processed separately to generate strain specific cellular populations. The elutriator was first cleaned with 80% ethanol, cooled to 4°C, calibrated, and tested at different running speeds using ultrapure water to check for leaks and spikes in pressure. Once the set up has been satisfactorily tested, DMEM 0.5% BSA is loaded at 10ml/minute, at 3000rpm.

The input channel is transferred from the DMEM 0.5% BSA into the first cell suspension (20ml) being careful not to induce any air bubbles into the channel. This was avoided by using two separate channels that could be switched from active to passive interchangeably. The suspension was loaded at 10ml/minute and cells were allowed to accumulate in the chamber for fifteen minutes. In order to collect the late spermatid fraction (cells smaller in size than in the early spermatid fraction), the flow

rate was increased to 22.5ml/minute. The first 50ml of cells were collected into a 50ml falcon tube (Greiner) and stored on ice. An additional 100ml was collected and discarded to remove all remaining cells before the early spermatid fraction was collected. Immediately after the 100ml discard, the flow rate was increased to 40ml/minute to release the early spermatid fraction. As before, 50ml was collected and a subsequent 100ml discarded. All cell populations were kept on ice after collection. The elutriator was then shut down and rinsed through with pure DMEM 0.5% BSA to release all cells and prepare for the next elutriation. The process was then repeated for the second cell suspension, after which all equipment was disassembled and cleaned. Cells were then ready for use in generating subcellular fractions; this is covered in the next section. Where subcellular fractionation was not required, cells were spun down at 500g for fifteen minutes at 4°C, the supernatant aspirated and the cells resuspended in 10ml PBS.

2.4.2. Subcellular fractionation of cells

This section covers the methods used for retrieval of subcellular fractions from elutriated populations of testicular cell suspensions (*Figure 2.3*). The below protocol follows on from cellular population collection outlined in section 2.4.1.

Fractionation is facilitated by the application of a series of buffers with an increasing concentration of salts. The series of buffers used is an adaptation of the Hesketh series (Hesketh et al. 1991). Hesketh series buffers 1 and 2 were produced as described below. The third buffer described in Hesketh et al. 1991, was substituted for QIAzol Lysis Reagent (Qiagen, UK). Buffers were prepared the day prior to elutriation and fractionation, and were autoclaved and allowed to cool prior to use. RNase inhibitor Ribolock (Thermo Fisher) was then added at 135 units/ml.

Lysis buffer 1:

Reagent	Amount	Final concentration
1M Tris pH7.6	1ml	10 mM
Sucrose	8.56g	250 mM
KCl	186.3mg	25 mM
MgCl ₂	101.6mg	5 mM
CaCl ₂	7.3mg	0.5mM
Igepal	50µl	0.05%
ddH ₂ O	To 100ml	

Table 2.4. Composition of lysis buffer 1

Lysis buffer 2:

Reagent	Amount	Final concentration
1M Tris pH7.6	1ml	10 mM
Sucrose	8.56g	250 mM
KCl	969.1mg	130 mM
MgCl ₂	101.6mg	5 mM
CaCl ₂	7.3mg	0.5mM
Igepal	50µl	0.05%
ddH ₂ O	To 100ml	

Table 2.5. Composition of lysis buffer 2.

Continuing from section 2.4.1:

Cellular populations were centrifuged at 500g for fifteen minutes at 4°C to allow a visible cell pellet to form. The supernatant was then removed using a pipette and the cell pellet resuspended in 10ml sterile PBS.

At this point, small aliquots of each fraction were taken for cell counting (see section 2.1.3), and for production of slides for later staining and imaging (see section 2.4.3). For the purposes of RNA sequencing (Chapter 4), microarray (Chapter 5), and micro RNA (Chapter 6) investigations, a ‘whole cell’ aliquot of minimum 1×10^7 cells was taken for examination without subcellular fractionation for each cellular population collected (i.e. early spermatid and late spermatid). All cell suspensions were centrifuged, the supernatant removed and the cells resuspended in 10ml fresh PBS; this was repeated once.

Whole cell sample preparation

The ‘whole cell’ preparations were then centrifuged a final time, a small amount (150µl) of PBS was retained and the pellet gently resuspended in this volume. In a fume hood, 850µl of QIAzol was added to the resuspended cells. The mixture was vortexed and the cells were thoroughly homogenized by moving the mixture through a 19G needle (Terumo) attached to a 1ml syringe (BD) a minimum of ten times. This homogenised mixture was immediately placed on dry ice.

Generating subcellular fractions

The remaining cell suspensions were centrifuged a final time, the supernatant removed and the cells resuspended in 1.5ml of lysis buffer 1. Following resuspension, the sample was placed on ice for 5

minutes to allow sufficient time for lysis. The mixture was then immediately centrifuged at 1000g for five minutes at room temperature and the supernatant (Fraction 1) collected into a 15ml falcon tube (Greiner) . In the hood, QIAzol is added to the mixture in a 5:1 ratio (7.5ml QIAzol added). The mixture was then briefly vortexed to ensure thorough mixing of the sample. The pellet retrieved following lysis 1 was retained and washed again in lysis buffer 1 (1000g for five minutes) and the supernatant disposed of. Pellets were then resuspended in 1.5ml of lysis buffer 2 and incubated at 4°C for ten minutes. Centrifuging the mixture at 2000g for five minutes allows for separation of the final hard pellet and the supernatant. As before, this suspernatant was then mixed with 7.5ml QIAzol in a falcon tube, this time labelled ‘Fraction 2’. The remaining pellet was washed (without forceful resuspension) in lysis buffer 2 a final time, centrifuged and the supernatant disposed of. The final hard pellet is difficult to resuspend, so once 1ml of QIAzol had been added directly to the pellet and the eppendorf well sealed, the mixture was vortexed extensively to achieve as much break-up of the pellet as possible, after which the mixture was homogenised using a needle and syringe – the mixture was moved through the needle a minimum of twenty times. This final sample, labelled ‘Fraction 3’ was then stored on dry ice with the other fractionated samples.

It is possible to store the samples for a number of months at -80°C once suspended in QIAzol. Once ready, RNA extraction can proceed from this point.

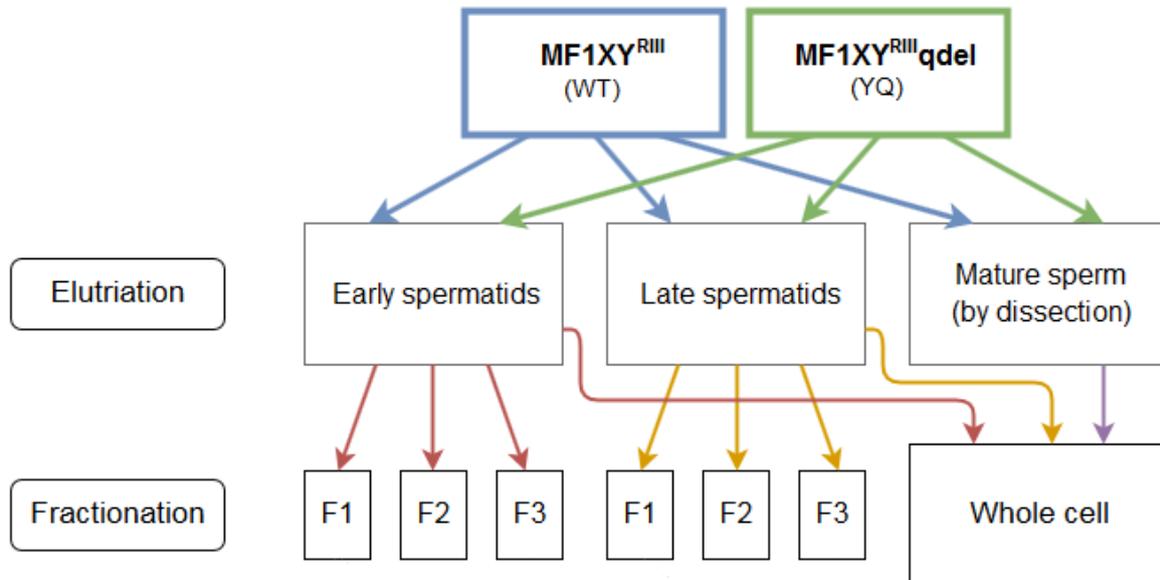


Figure 2.3. Flow diagram of sample preparation procedures resulting in the collection of subcellular fractions. Following centrifugal elutriation of spermatid populations, populations were collected into subcellular fractions through the application of lysis buffers to produce three separate fractions, entitled ‘F1’, ‘F2’ and ‘F3’. A subset of non-fractionated cells was retained from each enriched spermatid population for the examination of RNA in whole cells.

2.4.3. DAPI and lectin staining for examination of cellular fractions

Cells were collected as described in section 2.4.1. Approximately 4×10^4 cells were then resuspended in 200 μ l of PBS in preparation for cytospinning. The Cytospin 2 (Shandon, UK) was used to concentrate cells in a central area on the slide. Polylysine slides (Thermo Fisher) were used for greater adherence of cells. Cytospin chambers were set up as follows (Figure 2.4):

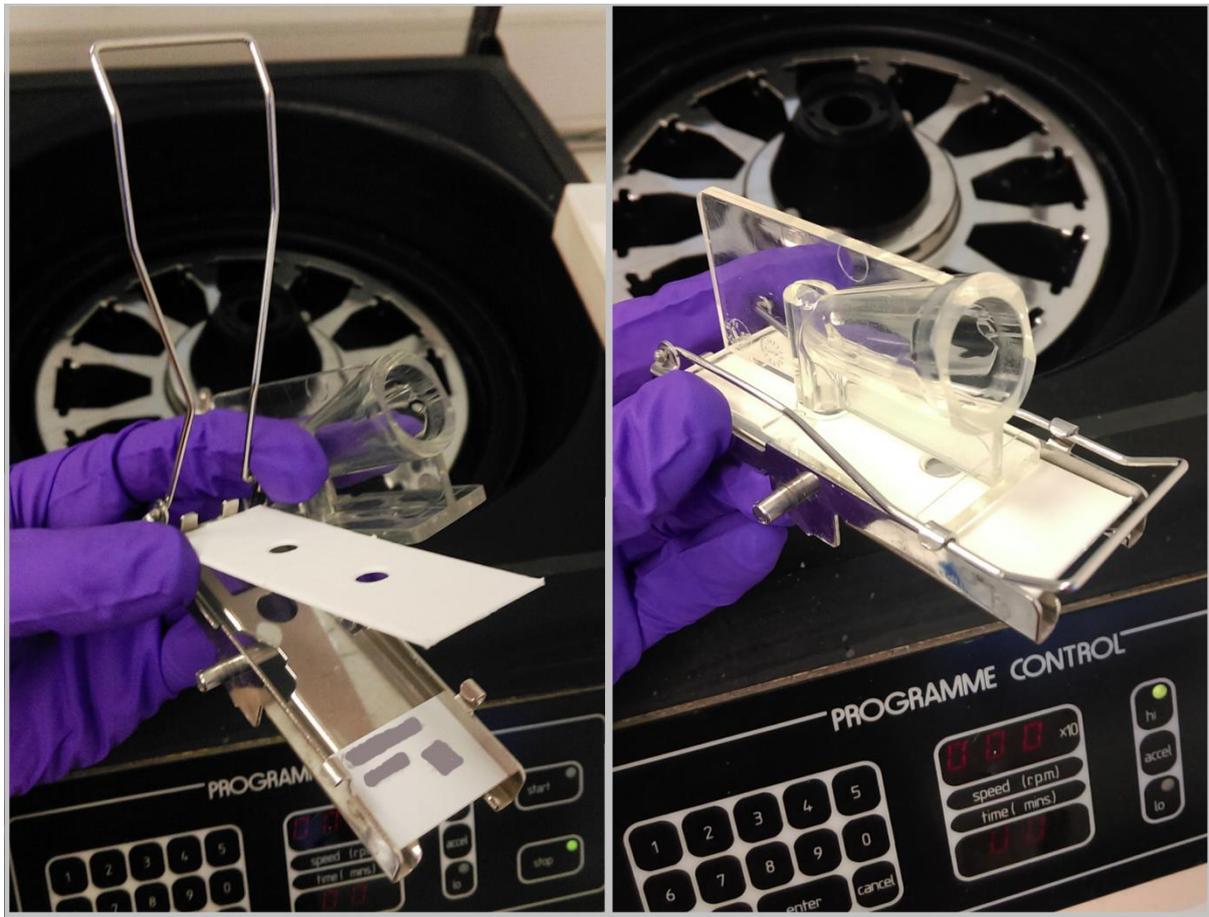


Figure 2.4. Cytospin chamber set up. In the left hand image the chamber can be seen in its constituent parts: clamp; glass slide; filter slip; plastic cytospin funnel. The image on the right shows the complete chamber ready for use in the cytospin.

Two hundred micro litres of cell suspension was dropped into each individual cytospin chamber. After two minutes at 800g, slides were released from their chambers and were ready for DAPI and lectin staining. To fix the cells, 500 μ l of 4% PFA was added to each slide and incubated for four minutes at room temperature. The slides were then rinsed (submerged) in three separate PBS solutions to remove excess PFA and then air dried. Slides were then stored at -80°C for later use as required. For acrosome visualisation, Alexa-coupled lectin-PNA staining is required. 100 μ l of a 1/500 dilution of Alexa-coupled PNA-lectin was added to each slide and the slides then incubated in a moisture chamber (in the dark) at room temperature for a minimum of one hour. Care was taken to minimise

exposure of the fluorophore to light. After one hour, the slides were washed in PBS for ten minutes and allowed to air dry (in the dark). DAPI was used for nuclear staining; one 8µl drop of DAPI was added at either end of the slide and a 50 x 22mm slide carefully lowered on top. The slides were inverted onto an absorbent towel to remove excess DAPI before clear nail varnish was applied to the edge of the coverslip to seal the slide. Slides were then ready for visualisation by fluorescent microscopy.

2.4.4. Cell counting from microscopy images using ImageJ

Following preparation of slides as described in sections 2.2.3 and 2.4.3, slides were imaged using a Nikon Mircophot SA fluorescent microscope with appropriate filters. Images were collected from several regions of the slide. Images were subsequently uploaded and split into separate colour channels using ImageJ (1.50b) standard package. Images were then imported to Microsoft Office Picture Manager (2007) where brightness and contrast were adjusted equally across all images to make the images clearer and cell recognition easier. Cells were manually identified and counted with the aid of ImageJ package Cell Counter (De Vos 2001³). Several images were collected of each slide and therefore each sample. A minimum of 500 cells per sample were counted and allocated to one of five categories. See chapter 5 section 5.2.1 for a description of the categories and results of centrifugal elutriation enrichment.

2.5. RNA Extraction and Sample Preparation

Following preparation of cells and fractions in QIAzol as described in sections 2.4.1 and 2.4.2, samples were taken forward for RNA extraction and subsequent library preparation for either Illumina RNA sequencing (section 2.6), Affymetrix microarray (section 2.7.1) or Toray microarray (section 2.7.2).

2.5.1. RNA extraction

In order to retain a broad range of RNAs, varying in size from micro RNAs (miRNAs) at around 20b in length in eukaryotic systems to messenger RNAs (mRNAs) which can reach more than 100Kb in length, a suitable RNA extraction kit was required. The miRNeasy mini kit (Qiagen, UK) allows for the purification of total RNA, including miRNA, from animal cells and tissues and was used in the preparation of RNAs from suspensions in QIAzol. Quality control and sample quantification of purified RNAs and RNA libraries was performed on three systems: Nanodrop 1000 (Thermo Scientific, US), Agilent 2100 Bioanalyzer (Agilent Technologies, UK), and Qubit 3.0 Fluorometer (Thermo Fisher Scientific, US). Where required, extracted RNA samples were concentrated using the

³ Cell Counter Image J - <https://imagej.nih.gov/ij/plugins/cell-counter.html>

Concentrator Plus (Eppendorf, Germany). Library preparation was completed using either the TruSeq RNA Library Preparation Kit v2 (Illumina, USA) or, for small RNA library preparation, the TruSeq Small RNA Library Prep Kit (Illumina, USA).

Purification of total RNA with miRNA

Purification of total RNA from mouse testis cell samples suspended in QIAzol was performed using the miRNeasy mini kit (Qiagen, UK), following the manufacturer's protocol, 'Purification of Total RNA, Including Small RNAs, from Animal Cells' (miRNeasy Mini Handbook - QIAGEN). An additional on-column DNase digestion (Appendix B) was applied for whole cell samples and subcellular fraction F3, where large amounts of nuclear material were expected to reside. Purified RNAs were elutriated into 30µl of RNase free water in two separate 1.5ml DNA/RNA LoBind microcentrifuge tubes (Eppendorf, Germany) per sample. Where initial sample volume exceeded the maximum recommended volume of the kit's RNA columns, columns were reused up to a maximum of 2ml sample. Any remaining unprocessed sample was refrozen and stored at -80°C.

2.5.2. Quality Control of RNA samples

Three separate measures of RNA quality were used to determine purified RNA quality and concentration. A separate 6µl aliquot of purified RNA was set aside from each main sample for the purposes of quality control, to avoid repeated freeze-thaw of the main samples.

A concentration reading and two separate quality readings were retrieved using the Nanodrop 1000 (Thermo Scientific, US), a spectrophotometer utilising surface tension to remove the need for sample containment units such as cuvettes, and allows for RNA/DNA/protein quality assessment using very small amounts (1µl) of sample. A concentration reading is given in ng/µl based on absorbance at 260nm. Separate quality readings for the detection of protein and phenol contaminants or other co-purified contaminants are given by a 260/280nm ratio and 260/230nm ratio respectively. A blank measurement was taken using RNase free water. The Nanodrop 1000 was used as per the manufacturer's instructions.

The Agilent 2100 Bioanalyzer (Agilent Technologies, UK) is commonly used to assess RNA quality prior to RNA sequencing by automated electrophoresis. RNA concentration estimation, RNA integrity number (RIN) and an electropherogram displaying distribution of RNA species by size are generated. The RIN value provides an indication of overall RNA integrity. Where sample concentration exceeded 50ng/µl (as estimated by testing on the Nanodrop 1000), the RNA 6000 Nano Kit (Agilent Technologies, UK) was used. RNA samples at lower estimated concentrations required the use of the RNA 6000 Pico Kit (Agilent Technologies, UK), which provides greater sensitivity. The Agilent 2100 Bioanalyzer was used as per the manufacturer's instructions.

The Qubit Fluorometric Quantitation system (Thermo Fisher Scientific, UK) was used as a third assessment of RNA quantity in purified RNA samples prior to library preparation for RNA sequencing. The Qubit 3.0 works by measuring the amount of signal emitted by fluorophores that activate only when bound to the sample molecule in question.

All three methods listed above require only 1µl of the sample for assessment, so are even suitable for low yield samples. All three methods were taken into account when adjusting concentration of samples for input into library preparation.

2.5.3. Increasing concentration of purified RNA samples using Concentrator Plus

Purified RNA extracts must meet minimum concentration requirements prior to library preparation and RNA sequencing. Illumina total RNA preparation requires a minimum input of 0.1µg RNA per 10µl, whilst small RNA library preparation requires a minimum input concentration of 1µg RNA per 5µl (Illumina Truseq Stranded Total RNA Sample Preparation Guide 2013). Samples that did not meet this minimum were concentrated further using the Concentrator Plus (Eppendorf, Germany), performed following the manufacturer's instructions.

2.6. RNA Sequencing Library Preparation

2.6.1. Total RNA Illumina library preparation

Library preparation is essential for converting purified total RNA into template molecules whose origin can be identified thanks to the addition of molecular indexes. This makes the strands suitable for the generation of clusters, multiplexing and subsequent DNA sequencing.

Library preparation of Total RNA was carried out using the Truseq Stranded Total RNA LT Prep Kit (Illumina, USA), following the Low Sample (LS) protocol (recommended for twenty four samples or fewer) provided by the manufacturer. Whole transcriptome sequencing of Total RNA, captures both coding and multiple non-coding sequences allowing for the detection and quantification of known and novel transcripts in a given sample. Input of RNA was normalised at 20ng per microlitre.

Library preparation begins with depletion of ribosomal RNA (rRNA); this is important as rRNA is hugely abundant in most tissue types, and its removal allows for greater depth of coverage of sequences of interest. The Ribo-Zero rRNA Removal Kit Human/Mouse/Rat (Illumina, USA) was used for this purpose prior to cDNA synthesis. First strand synthesis involves the reverse transcription of cleaved RNA fragments that have previously been primed with random hexamers (oligonucleotides covering multiple sequence variations of that size), required for priming single-stranded sequences (the cleaved RNA) for extension. Synthesis of the second strand removes the original RNA template,

replacing dTTP with dUTP and completes the synthesis of blunt-ended double-stranded cDNA. To prevent blunt-end to blunt-end cDNA ligation, a single adenine nucleotide is added at the 'adenylate 3' end' phase to the 3' end and allows for adapter-fragment ligation. The addition of adapter sequences allows for indexing of sequences and subsequent multiplex sequencing. Each sample is given a unique adapter sequence, so the fragment may hybridise successfully to the flow cell, and the origin of the sequence can be identified following DNA sequencing. The final enrichment step utilises a polymerase chain reaction (PCR) step for amplification of all successfully adapter-ligated cDNA fragments. PCR cycles are kept to a minimum to avoid over-representation of some transcripts (skewing) in the newly generated library (Illumina Truseq Stranded Total RNA Sample Preparation Guide, 2013). Library preparation was conducted over three weeks, samples being processed two at a time.

Total RNA Library validation

Total RNA libraries were examined for concentration and integrity using the Agilent 2100 Bioanalyzer (Agilent Technologies, UK) and a DNA specific Bioanalyzer chip. Concentration of libraries was first estimated using the Nanodrop 1000 (Thermo Scientific, US) to determine which Bioanalyzer chip should be used for a more thorough assessment. The DNA-1000 kit (Agilent, UK) was used where sample concentration was determined to be greater than 0.1ng/µl, and the High Sensitivity DNA Analysis Kit (Agilent, UK) used where concentration was likely to be lower.

Total RNA sequencing

Sequencing of total RNA libraries was performed in three separate runs of six samples (eighteen in total) on the NextSeq 500 (Illumina, Chesterford UK). Processing, quality control and analysis of RNA sequencing data is covered in Chapter 4.

2.6.2. Small RNA Illumina library preparation

The TruSeq Small RNA Sample Prep Kit (Illumina, USA) was used for preparation of small RNA libraries. The Illumina adapter sequences specific to this kit ligate directly to the common natural structure of micro RNAs (miRNA), allowing for specific amplification of this RNA species. The single purified total RNA samples generated in section 2.5 are still used as the RNA extraction technique allowed for small RNA fragments to be retained.

1µg of RNA in 5µl of ultrapure water was used for Small RNA library preparation. A DNase step was carried out as recommended by the manufacturer. Small RNA library prep begins with the ligation of the adapter sequences at both 3' and 5' ends of the RNA fragments. These specially designed adapters attach exclusively to mature miRNAs and serve to purify the RNA whilst uniquely tagging each sample so that the origin of each read can be determined post-sequencing. cDNA constructs are

generated using a single step PCR step that selectively amplifies RNA fragments that have successfully ligated 3' and 5' adapter sequences. A second short PCR step (eleven cycles) is performed for amplification of purified transcripts during which a sample index, unique to each sample, is attached to the transcripts. Following amplification, samples are run in parallel through a gel purification stage, where size selection takes place. The gels were visualised using SybrSafe (Thermo Fisher), a safer (and confirmed suitable) alternative to the ethidium bromide stain recommended in the manufacturer's protocol, under a Dark Reader Transilluminator and desired bands retrieved using a razor blade as per the manufacturer's instructions. Gel breaker tubes were obtained from IST Engineering Incorporated. Ethanol precipitation was not performed, so as to avoid reducing yield. The DNA was eluted by rotating on a powered eppendorf rotation rack for between two and twenty-four hours. The DNA is finally resuspended in Tris-HCl pH 8.5 and is ready for library validation.

Small RNA Library validation

Small RNA libraries were assessed for concentration and quality using the same methods as described for total RNA library validation (see section 2.6.1)

Small RNA sequencing

Sequencing of small RNA libraries was performed in two separate runs of nine samples (eighteen in total) on the NextSeq 500 (Illumina). Processing, quality control and analysis of RNA sequencing data is covered in Chapter 6.

2.7. Microarray Sample and Library Preparation

2.7.1. Affymetrix Total RNA sample preparation and processing

RNA samples were prepared as described in section 2.5. Library preparation is not required on the Affymetrix total RNA platform.

RNA was assessed for concentration and quality using a SpectroStar (BMG Labtech, Aylesbury, UK) and Bioanalyser (Agilent Technologies, Cheshire, UK). Microarray processing and quality control were performed by Cambridge Genomic Services (Department of Pathology, University of Cambridge) using a species specific Gene 2.1 ST Array Plate (Affymetrix, Wooburn Green, UK) in combination with WT PLUS amplification kit (Affymetrix) according to the manufacturer's instructions.

100ng Total RNA was amplified along with inline PolyA spike in control RNA, using the WT PLUS amplification kit (Affymetrix). Successfully amplified samples were labelled using the GeneChip WT

terminal labelling kit (Affymetrix) using the in line hybridization controls. Plate arrays were processed on the GeneTitan instrument (Affymetrix) using the GeneTitan Hybridization, Wash and Stain kit (Affymetrix). Samples were hybridized to the array, washed, stained and scanned using the array specific parameters provided by Affymetrix. Finally basic visual quality control was performed using Command Console Viewer (Affymetrix) prior to bioinformatic quality control.

Processing, quality control and analysis of microarray data is covered in Chapter 5.

2.7.2. Toray miRNA sample preparation and processing

RNA samples were prepared as previously described in Section 2.5. A miRNA labelling reaction using the 3D-Gene miRNA labelling kit is carried out immediately prior to RNA-microarray chip hybridisation. Micro RNA labelling and hybridisation were carried out by Cambridge Genomics Services (Department of Pathology, University of Cambridge) and utilised the Toray version 21 miRNA microarray chip. Hybridisation and capture was performed on the Toray platform. Processing, quality control and analysis of miRNA microarray data is covered in Chapter 6.

Chapter 3

3. Physiological Characterisation of the MF1XY^{RIII}qdel Model

This chapter examines reproductive and sperm phenotype characteristics of the MF1XY^{RIII}qdel model to determine whether they differ from the wild-type MF1XY^{RIII}. A number of physiological aberrations have previously been reported to be associated with MSYq deletions in mouse, including unusual sperm morphology and offspring sex ratio skewing towards females. We know from previous work that a functional inequality exists between X- and Y-bearing sperm in the deletion model, facilitating the offspring sex ratio skew; in this chapter a number of potentially causative characteristics are examined. Understanding what form these functional differences take is valuable for investigating the underlying mechanism of sex ratio distortion in this model.

3.1. Chapter Introduction

A number of abnormal physiological phenotypes have been observed in mice with partial deletions of the MSYq. It has been determined that the offspring sex ratio skew observed in MSYq deletion models, and more specifically in the XY^{RIII}qdel, is caused not by differences in X- and Y-bearing sperm number or increased post-fertilisation death of male embryos, but is due to differences in sperm function between X- and Y-bearing sperm (Ward and Burgoyne 2006). Understanding the physiological characteristics of the MF1XY^{RIII}qdel (YQ) model as compared to the control MF1XY^{RIII} (WT), provides information on potential fitness outcomes and reveals where functional differences may be present between the two strains and between X- and Y- bearing sperm within each strain. Physiological characteristics provide information to guide molecular analyses, where specific regions of the genome are known to contribute to the functional competency of sperm.

Sex ratio skewing has been observed and reported in a number of mice with large MSYq deletions (section 1.5.2). Over a near five year period, the YQ offspring skew towards females is seen to be 6% divergent from that of offspring produced by WT sires. A reduction in litter size is also noted for YQ males compared to WT.

Abnormal sperm morphology is a well described characteristic of MSYq deletion mice (section 1.5), however there has to date been no investigation of morphological differences between X- and Y-bearing sperm in these models. Any differences between X- and Y- bearing sperm could be linked to reduced fertilising capacity in the Y-bearing sperm. In this chapter the results of both qualitative and quantitative morphological assessments are reported. This reveals distinctive sperm head morphology present across both X- and Y-bearing sperm in the MF1XY^{RIII}qdel, associated with specific regions of

the sperm head. The analysis also reveals a greater morphological distinction between X- and Y-bearing sperm in YQ populations than observed in WT.

Hyaluronidase activity of sperm allows for the more rapid penetration of sperm into the inner layers of the oocyte. The cumulus oophorous is the external-most layer of the oocyte complex and efficient diffusion of this layer can provide a competitive edge to competing sperm. Differences in hyaluronidase activity have been touted as a source of functional inequality between individual sperm from the same ejaculate in a mouse model of transmission ratio distortion of the X-chromosome (section 1.4.2). Evidence of potential non-sharing of hyaluronidase RNA transcripts and proteins between developing spermatids made hyaluronidase activity an intriguing candidate as a source of functional inequality between haploid sperm in our model. However, hyaluronidase activity was compared between pools of MF1XY^{RIII} and MF1XY^{RIII}qdel sperm and revealed no discernible difference between the two strains in this instance.

Separate introductions are given for the investigation of each physiological characteristic as listed above.

3.1.1. Chapter objectives

The specific aims of this chapter are:

- O1. To define the extent of the offspring sex ratio skew, and consistency of the skew in our MF1XY^{RIII}qdel colony, with comparison to the wild-type MF1XY^{RIII} mouse strain.
- O2. To define observed morphological differences between the MF1XY^{RIII} and MF1XY^{RIII}qdel strains at a greater resolution than is currently provided for other models with a similar size deletion.
- O3. To determine whether morphological differences exist between X- and Y-bearing sperm of either strain, and to quantify these differences where they do exist.
- O4. To examine hyaluronidase activity as a potential factor of functional inequality in populations of MF1XY^{RIII}qdel sperm.

3.2. Offspring Sex Ratio and Litter Size in the MF1XY^{RIII}qdel Colony

3.2.1. Introduction to offspring sex ratio and litter size assessment

Across all known MSYq deletion mouse strains within which males retain fertility, an offspring sex ratio skew towards females is recorded (where examined) compared to full genomic complement mice from the congenic strain (section 1.5); the extent of the deletion correlates with the extent of the

offspring sex ratio skew when other factors are controlled for (Case et al. 2015, section 1.5.1). It has been shown that this relates primarily to the copy number of *Sly* genes on the MSYq. Given this information, we expect to observe a sex ratio skew towards females similar to that observed in other MSYq deletion models lacking around 2/3 of the MSYq; in the region of 9-13%.

Some conflicting information exists regarding the viability of males with MSYq deletions. Whilst strains lacking extensive portions of the MSYq of >90%, such as the $XSx7^aY^{*X}$ (Burgoyne et al. 1992) and $XY^{Tdyml}qdelSry$ (Touré et al. 2004a) are known to be incapable of producing offspring through natural mating, reports of strains with lesser deletions (c.66%) have produced mixed results (section 1.5.1). Here, as is the case in a number of reports, litter size is used as a proxy measurement of fecundity and is assessed across all litters produced in each strain.

Given the information we have regarding sex ratios and litter sizes, here the following hypotheses are tested:

Hypothesis (1): Sires from the MF1XY^{RIII}qdel line produce a different proportion of male offspring than sires belonging to the MF1XY^{RIII} line.

Hypothesis (2): Litter size means differ significantly between the MF1XY^{RIII} and MF1XY^{RIII}qdel strains.

3.2.2. Results of offspring sex ratio and litter size assessment

Rederived MF1XY^{RIII} and MF1XY^{RIII}qdel embryos were originally obtained from the Burgoyne lab in 2012. These mice formed the basis of breeding colonies for the two strains studied in the work described in this thesis. Both sets of males were bred with MF1 females already established at Cambridge to produce both MF1XY^{RIII} and MF1XY^{RIII}qdel colonies. Records of litter size and sex of offspring were collected over a four and a half year period.

Offspring sex ratio data

From the information given in *Table 3.1*, a difference in the percentage of male offspring of 6% can be seen between MF1XY^{RIII} and MF1XY^{RIII}qdel colonies. This shows all offspring produced over a nearly five year period, from establishment of the colony to the end of the period of study. MF1XY^{RIII} sires produced 2.03% fewer males than females over five years than females, whilst MF1XY^{RIII}qdel sires produced 22.93% fewer males than females over the same period. The hypothesis was tested using the chi-squared test for goodness-of-fit and the information in *Table 3.1*.

Across all offspring				
Paternal strain	Total offspring	Male	Female	Ratio male
MF1XY ^{RIII}	683	338	345	0.495
MF1XY ^{RIII} qdel	919	400	519	0.435

Table 3.1. Offspring sex ratio comparison between MF1XY^{RIII} and MF1XY^{RIII}qdel sires across all offspring produced.

The chi-squared test for goodness-of-fit was implemented in R and retrieved a p-value of 0.00029. The null hypothesis was rejected. *Table 3.2* shows a low standard deviation between sires within the two strains.

Sire specific information				
	No. of sires	Offspring male ratio (mean)	St.dev	St.err
MF1XY ^{RIII}	10	0.495	0.06	0.02
MF1XY ^{RIII} qdel	16	0.435	0.04	0.01

Table 3.2. Standard deviation and error associated with sires in each strain.

Litter size data

Litter size was examined across the two strains. A total of 54 MF1XY^{RIII} litters and 83 MF1XY^{RIII}qdel litters were examined. Although a reduced litter size was observed in MF1XY^{RIII}qdel, this was shown to overlap with distributions from MF1XY^{RIII} (*Figure 3.1*). A mean litter size of 12.4 was recorded for MF1XY^{RIII} and 11.1 for MF1XY^{RIII}qdel, with a variance of 17.4 and 19.5 respectively.

Student's unpaired t-test was used to test the null hypothesis that litter size means do not significantly differ between the two strains. The t-test produced a two-tailed p.value of 0.037, meaning we should reject the null hypothesis at p<0.05.

Litter size is expected to vary within and between individual sires, but without producing a greater number of mice simply to test this observation, statistical testing of these variables is of little value. Instead, a graphical examination of litter size between males was produced (*Figure 3.2*). *Figure 3.2* and *Figure 3.1* show there to be few outliers across both strains. From *Figure 3.2* it seems that distinct distributions exist between the strains, but with high variance.

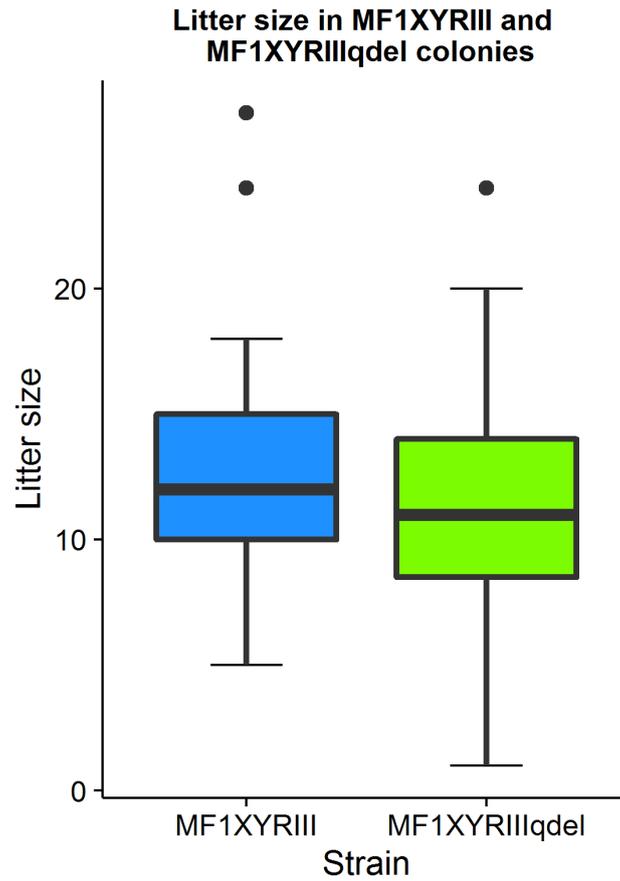


Figure 3.1. Boxplot of litter sizes across strains. Outliers are represented as black dots. Interquartile ranges and the median value are represented by the box plot.

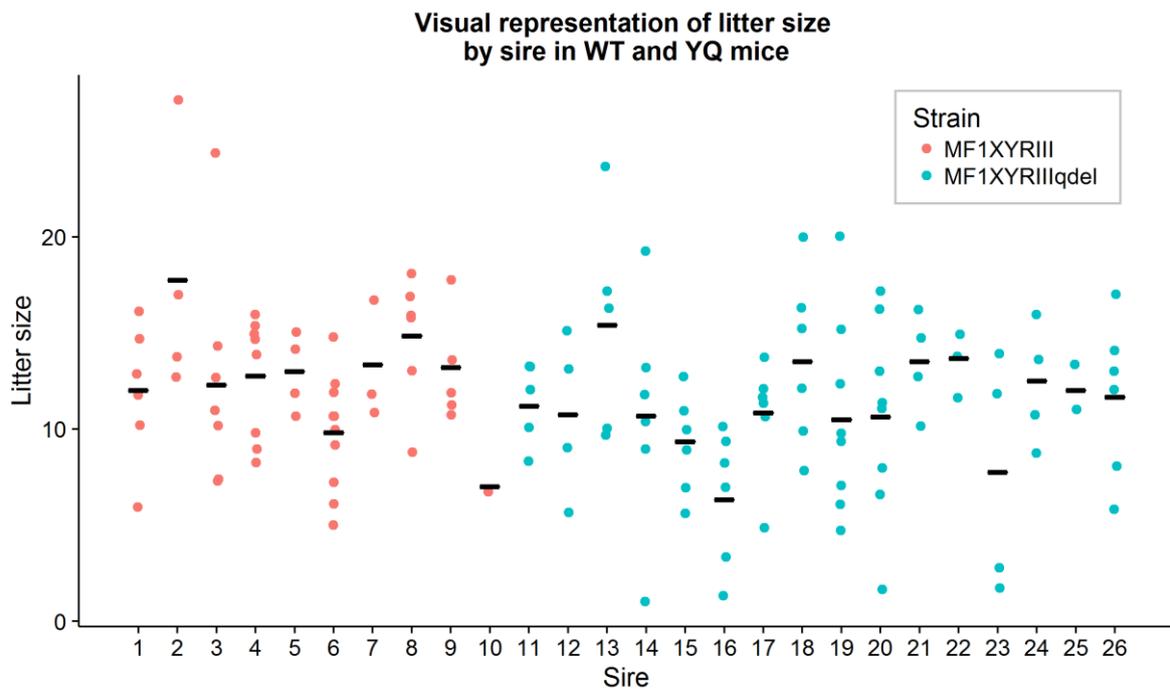


Figure 3.2. Individual litter sizes recorded by sire. Mean average values per sire are represented by a solid black line.

3.2.3. Discussion of findings of offspring sex ratio and litter size assessment

An offspring sex ratio skew towards females of 6% is seen in MF1XY^{RIII}qdel compared to MF1XY^{RIII}. The extent of the sex ratio skew has been shown to be strongly linked to the remaining copy number of certain genes on the MSYq (Case et al. 2015, Conway et al. 2004, Touré et al. 2004), and therefore the extent of the Yq deletion. The extent of the skew observed in our model fits with that observed in other models with similar deletions (section 1.5), though is perhaps slightly less than expected. A reduction in the ratio of male offspring per litter in both strains was noted over time, which appeared to affect WT litters more. As a result, the extent of the skew appears to have changed over time, from an initial skew of 11.5% in 2012 to 6% when considered across all years. This appears primarily due to a reduction in the percentage of male offspring born to MF1XY^{RIII} sires (55.94% in the 2012 to 49.5% over all years). Clearly, these comparisons are drawn across different sample sizes and it would not make sense to test the change statistically, however it is interesting as an observation. A small change in the ratios may be explained by subtle differences in copy number of X-linked genes involved in the intragenomic conflict between the original MF1 background of the founder individuals and the MF1 background of the resulting colonies as the lines were integrated over time.

Questions persist as to the relative fertility of mice with MSYq deletions. The results reported here show an overall reduction in litter size, which in this case was used as a proxy for fecundity, in MF1XY^{RIII}qdel compared to MF1XY^{RIII}. Direct competitive assays of MF1XY^{RIII} and MF1XY^{RIII}qdel mature sperm have not been conducted; it could be the case that the functional inequality that makes Y-bearing sperm less competitive than X-bearing sperm also affects X-bearing sperm but to a lesser extent, making YQ derived sperm less competitive overall. This might account for some reduction in litter size seen from YQ sires, if YQ sperm are less efficient at completing fertilisation.

3.3. Assessment of Sperm Head Morphology in MF1XY^{RIII}qdel mice

3.3.1. Introduction to assessing morphology

Qualitative assessment of sperm morphology by ranking of sperm heads into distinct morphological categories is common practise even in clinical settings (World Health Organisation (WHO) 2010). Qualitative assessment provides a relatively easy (though laborious) method for the identification of gross sperm abnormalities in different sperm populations, requiring only an appropriate stain, a means of visualisation and an operator. There are limitations to the technique; more subtle variation, particularly where strains are examined separately, may be lost. But the method still provides a valid basis for global examination of the frequency of cell morphology aberrations in sperm samples.

The sperm head morphology of strains MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ), was examined using silver staining and light microscopy. More than one thousand sperm heads were categorised by a single operator and the following results were generated. Reports across all models of MSYq deletion show a substantial proportion of sperm with aberrant head morphology, and it is expected that YQ mice will display a similar level of morphological abnormalities to those observed in mice with similar size deletions (section 1.5). This section aims to test the following hypothesis:

Hypothesis (1): The attribution of sperm head morphologies in sperm collected from MF1XY^{RIII}qdel mice will differ from that of sperm collected from MF1XY^{RIII} mice. Sperm head morphology in a sample sperm population is associated with the mouse strain from which the sperm population originates.

A detailed methodology can be found in section 2.2.1 and 2.2.2.

3.3.2. Results of qualitative sperm head morphology assessment

Qualitative assessment of sperm head morphology in accordance with categorisation outlined in section 2.2.2 revealed a notable departure from control strain (MF1XY^{RIII}) morphology in MF1XY^{RIII}qdel. Percentage morphology scores for strains (MF1XY^{RIII} and MF1XY^{RIII}qdel) were generated from combined percentage scores from individual mice: 574 individual sperm heads were assessed from 9 x MF1XY^{RIII} mice; and 448 sperm heads from 7 x MF1XY^{RIII}qdel mice. A minimum of fifty and a maximum of one hundred sperm heads were imaged and scored for each individual (s.d.=10.94).

MF1XY^{RIII} (WT) control mice showed 90.97% normal sperm head shape with 4.33% showing severe deformities (categorised as ‘other’) and 4.7% exhibiting some degree of sperm head flattening. In contrast, 84.18% of MF1XY^{RIII}qdel (YQ) sperm heads showed some degree of flattening with ‘flattened’ being the predominant sperm head shape; 46.88% of total. Only 10.19% of sperm were categorised as having normal head shape. Example morphologies for each category are shown in *Figure 3.3*. *Figure 3.4* shows the distribution of sperm head morphologies across all classifications in both strains. *Table 3.3* provides the percentage of sperm heads placed under each classification and the standard error between individual mice.

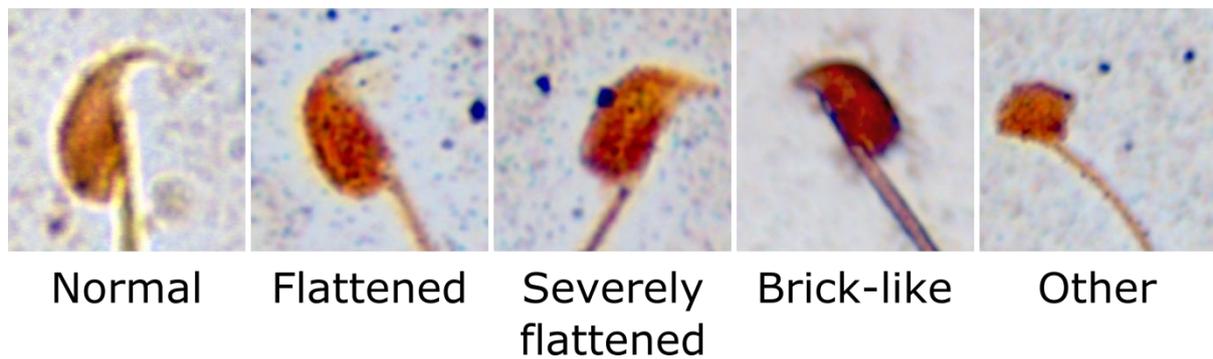


Figure 3.3. Silver nitrate stained sperm heads provided as examples of morphology for each designated category

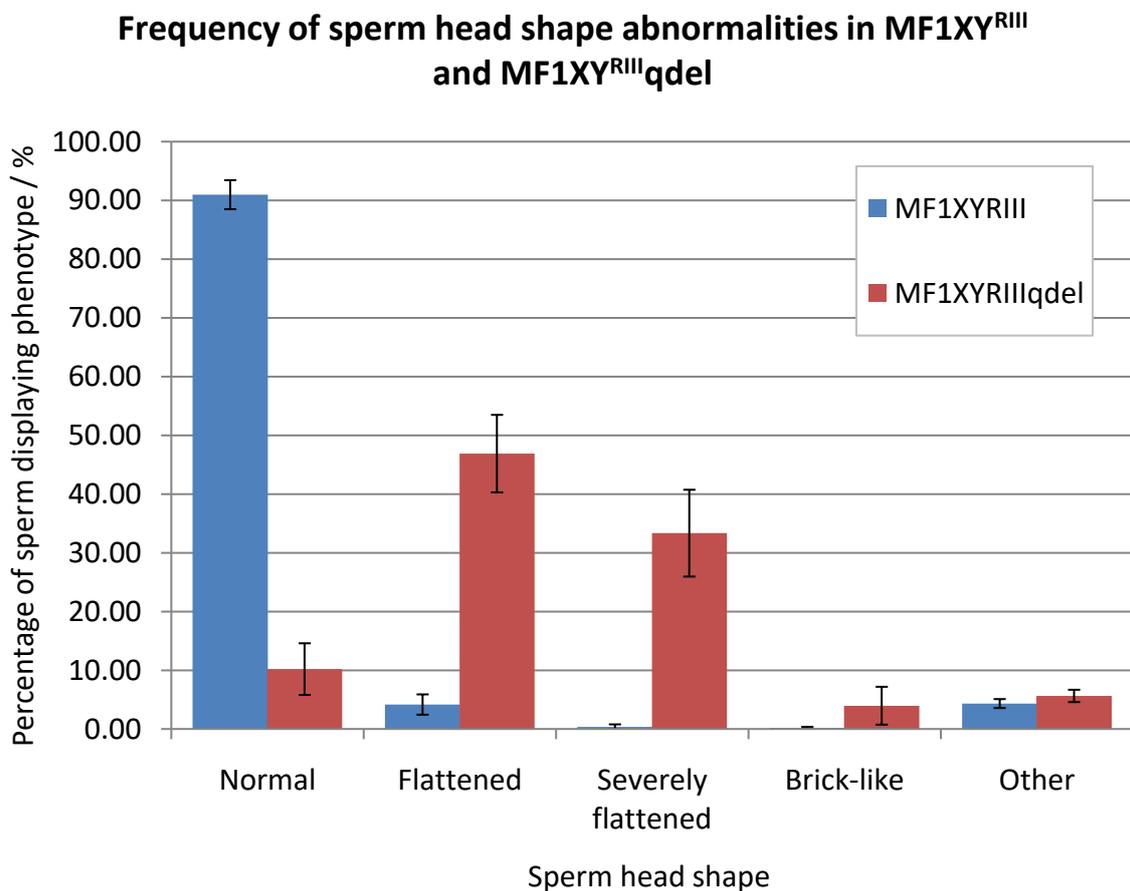


Figure 3.4. The proportion of sperm categorised under each morphological type; comparison of MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ). Values are given as a percentage of total sperm population assessed. Error bars represent standard error (WT n = 9, YQ n = 7) for each morphological category between individual mice.

Strain		Sperm head morphology classification / % attributed				
		Normal	Flattened	Severely flattened	Brick-like	Other
MF1XY ^{RIII}	Percentage of population / %	90.97	4.16	0.37	0.17	4.33
	Standard error	2.47	1.73	0.42	0.19	0.76
MF1XY ^{RIII} qdel	Percentage of population / %	10.19	46.88	33.34	3.96	5.63
	Standard error	4.40	6.60	7.39	3.22	1.05

Table 3.3. Observed sperm head morphology by classification represented as percentages of the assessed sperm population per strain. Standard error for each morphological category between individual mice is included (WT n = 9, YQ n = 7).

A chi-squared test for independence gives a p-value of <0.001 between the two strains. This implies a significant difference between the two strains in the proportion of sperm adhering to each morphological category, and leads to rejection of the null hypothesis.

The dominant phenotype observed in the wild-type MF1XY^{RIII} is ‘normal’ (90.97%). The dominant phenotype of the Yq deletion model, MF1XY^{RIII}qdel, is ‘flattened’ (46.88%) with an additional 33.34% appearing ‘severely flattened’; examples of which are seen above (*Figure 3.3*). The ‘flattened’ phenotype describes a sperm head that is mostly normal in appearance, except for a noticeably flattened acrosomal region, and somewhat flattened tail attachment region.

3.3.3. Discussion of findings of qualitative sperm head morphology assessment

In the B10.BR^{-Ydel} mouse model, another mouse strain with a >50% deletion of the male specific region of the Y chromosome long arm (MSYq), Styrna et al. (2002) reported 88% abnormal sperm head morphology and 31% in their control strain (B10.BR/SgSn). This compares to this author’s findings of 89.81% abnormal sperm head morphology in the MF1XY^{RIII}qdel deletion model and 9.03% abnormal sperm heads in the MF1XY^{RIII} control. These qualitative rankings were devised by the individual authors, and are likely to differ somewhat in how they are categorised. Recordings may also differ through observer error, but the descriptions and images given as representations of the normal and abnormal phenotypes are largely similar. Styrna et al. (2002) describe a ‘flattened’ acrosome phenotype that arises exclusively in their deletion model, and not in the control, occurring in 27.4% of the deletion model sperm population. From images provided in their publication, this phenotype seems to match well with the ‘severely flattened’ phenotype described in this experiment above, which was observed to occur in 33.34% of the deletion model population and 0.37% in the control population. Both Styrna et al. (2002) and I used sperm from the cauda epididymis and vas

deferens for examination. This appears to be a consistent phenotype across the deletion strains. Ward and Burgoyne (2006) provided much of the prior knowledge of the XY^{RIII}qdel phenotype from which our strain was derived. They provide a comparative image of XY^{RIII} and XY^{RIII}qdel sperm morphology in their publication, an altered version of which can be seen in section 1.5.3. This phenotype can be seen to match well with the results of this examination, showing consistency of the phenotype over time, and a distinct morphology associated with the deletion model (*Figure 3.5*).

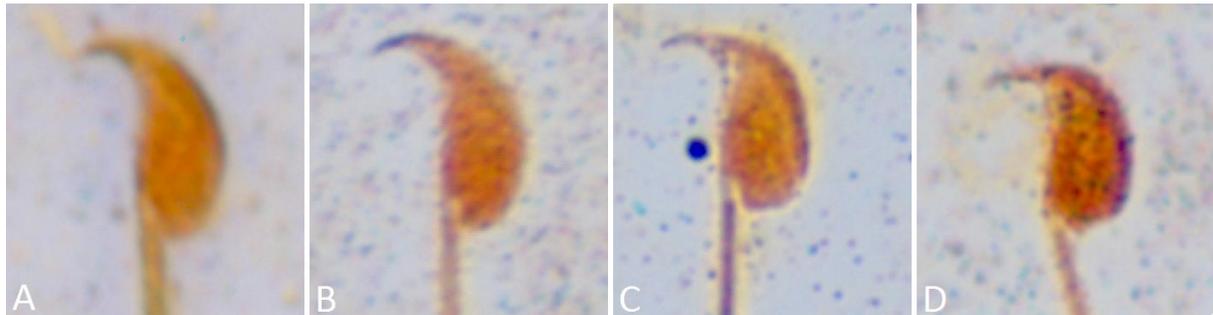


Figure 3.5. Examples of dominant sperm head phenotypes seen in the strains compared; MF1XY^{RIII} (A and B) and MF1XY^{RIII}qdel (C and D). A, B) Morphologically normal; C) ‘flattened’ phenotype; D) ‘severely flattened’ phenotype. Sperm heads in these images have been stained with silver nitrate and were viewed by light microscopy.

3.4. Quantitative Assessment of Sperm Head Morphology in MF1XY^{RIII}qdel Mice

By combining an objective technique for assessment of sperm head morphology with sperm fluorescence *in situ* hybridisation (FISH), it has been possible to generate a dataset and analysis pipeline that allows for the examination of chromosome content in relation to sperm head morphology. Here I report results of morphological comparisons between the wild-type (MF1XY^{RIII}) and MSYq deletion (MF1XY^{RIII}qdel) strains, and comparisons between X- and Y-bearing sperm within both strains. Sperm head measurements and comparisons were conducted using Nuclear Morphology Analysis software (version 1.13.7)⁴ created by Dr Benjamin Skinner (Department of Pathology, University of Cambridge).

Images and analysis in this section have been generated in R (R Core Team 2016) unless otherwise stated. Where images have been exported directly from the Nuclear Morphology Analysis software, this has been stated.

⁴ https://bitbucket.org/bmskinner/nuclear_morphology/wiki/Home

3.4.1. Introduction to quantitative sperm head morphology assessment

Mice with MSYq deletions exhibit abnormal sperm morphology (teratozoospermia) and an offspring sex ratio skew towards females. If the morphological abnormalities were a direct cause of reduced fertilising ability in Y-bearing sperm, we would expect to observe differences in morphology between X- and Y-bearing sperm within the MF1XY^{R^{III}}qdel phenotype, affecting at least some of either sperm population. Sperm morphology, in particular acrosomal shape and formation, has been linked to altered fertilising capability or shown to be an indicator of other aberrations in a number of species (de Boer et al. 2015, Fujihara et al. 2012, van der Horst et al. 2011, Chemes and Rawe 2003, section 1.3.1). It is often linked to chromosomal abnormalities and defects in chromatin condensation. In mouse, the relationship has been examined in several models; however comparative examinations of X- and Y-bearing sperm morphologies have not been published.

Sperm DNA and the head surrounding it are highly condensed and compacted, with protamines replacing histones as sperm mature and reduce in size, meaning the sperm head primarily represents a condensed nucleus. This makes examination of shape through nuclear staining a feasible prospect. Using the Nuclear Morphology Analysis software (version 1.13.7), it is possible to examine many facets of nuclear morphology, and quantify differences between a number of parameters in curated cell populations.

Fluorescence *in situ* hybridisation (FISH) is a commonly used molecular cytogenetic technique for the investigation of chromosome/gene content of cells. It utilises fluorescent probes that specifically bind to chromosomal regions based on sequence complementarity dependent upon the design of the probe. Sperm are haploid cells, containing either an X or Y chromosome. The addition of this technique to sperm morphology assessment allows for the separate examination of cells with distinct chromosome content.

In this investigation, a number of variables were examined. These included, but were not limited to: total sperm head area, sperm head perimeter, overall sperm head shape, and sex chromosome content. Overall sperm head shape was defined by an angle profile using a modified Zahn-Roskies transformation (Zahn and Roskies 1972). In its simplest form, starting from a single point, Zahn-Roskies involves the measurement of angular turns required to meet the subsequent point in a collection of (preferably) equally spaced x/y coordinates on the perimeter of any 2D object or curve. In the modification, distances between points are proportional to overall object size rather than being of a set value, meaning object shape can be defined independent of size. The stated parameters were assessed both between and within strains. The within strain investigation consisted of an assessment of morphology between X- and Y-bearing sperm (as determined by sex chromosome FISH). In previous work, an acrosomal flattening has been observed in models of MSYq deletion (Ward and Burgoyne 2006, Styrna et al. 2002). This was confirmed in section 3.3. It is therefore expected that

greater resolution of sperm head phenotypes will reveal the acrosomal region to be the major region of distinction between WT and YQ strains, and will allow for quantification of the differences in this region within and between strains.

Hypothesis (1): Head shape of sperm from MF1XY^{RIII}qdel mice is distinct from that of MF1XY^{RIII}.

Hypothesis (2): Sperm head shape of Y-bearing sperm is distinct from that of X-bearing sperm in MF1XY^{RIII}qdel mice, but not in MF1XY^{RIII} mice.

Hypothesis (3): The acrosomal region of the sperm head is the region of greatest dissimilarity between MF1XY^{RIII} and MF1XY^{RIII}qdel populations.

3.4.2. Structure and implementation of sperm head morphology analysis

To assess sex chromosome content of sperm and its relationship to sperm head morphology, two phases of image capture were completed: Phase I (morphology capture) and Phase II (FISH capture). As sperm chromatin is highly condensed, the cellular DNA is inaccessible to fluorescent probes. In order to overcome this, sperm were swelled using a swelling solution containing dithiothreitol (DTT). Swelling of cells ablates cellular morphology, so a capture and recapture technique made possible by the use of a motorised microscope stage was employed to examine sperm morphology pre-FISH and relocate the exact sperm heads for recapture and examination of sex chromosome content post-FISH.

Sperm cell suspensions from 3 x MF1XY^{RIII} (WT) and 3 x MF1XY^{RIII}qdel (YQ) age-matched mice were prepared as described in section 2.1.1. Initial image capture and sperm FISH were performed as described in section 2.2.3; following which, slide positions of individual sperm heads were loaded for secondary post-FISH recapture.

3.4.3. Image analysis

Figure 3.6 provides a simplified overview of the process of image analysis. Pre-FISH images were first imported for automated detection of nuclei. Manual curation of images was then performed to remove any nuclei where detection failed or was incomplete (*Figure 3.7*). Post-FISH images were mapped to previously imported pre-FISH morphology images. *Figure 3.8* gives examples of successful hybridisations; X-chromatin can be seen highlighted in green, whilst Y-chromatin is shown highlighted by a contrasting red fluorescent probe. Examples of successful and unsuccessful post-FISH mapping can be seen in *Figure 3.9*.

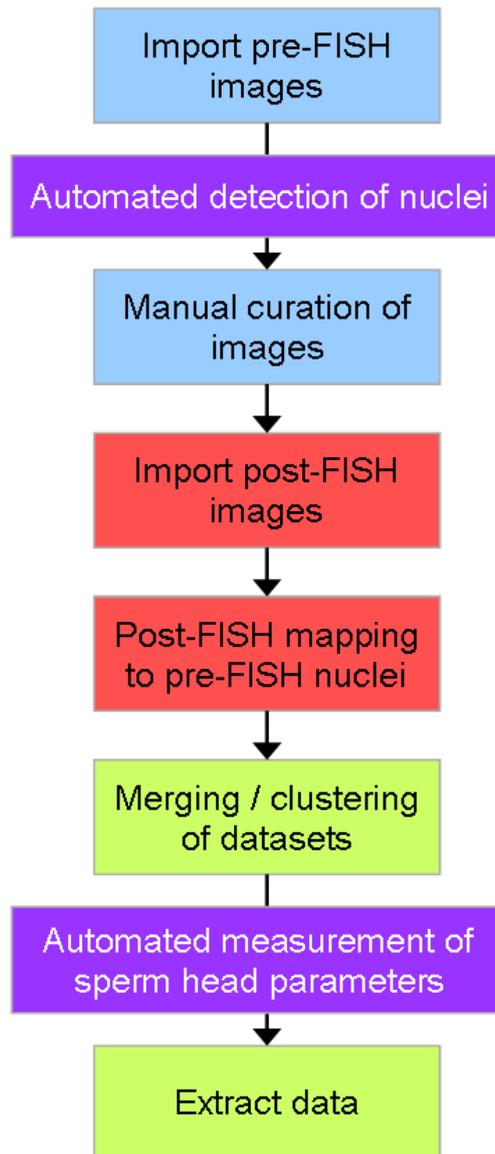


Figure 3.6. Simplified workflow of sperm head morphology assessment performed using Nuclear Morphology Analysis software.

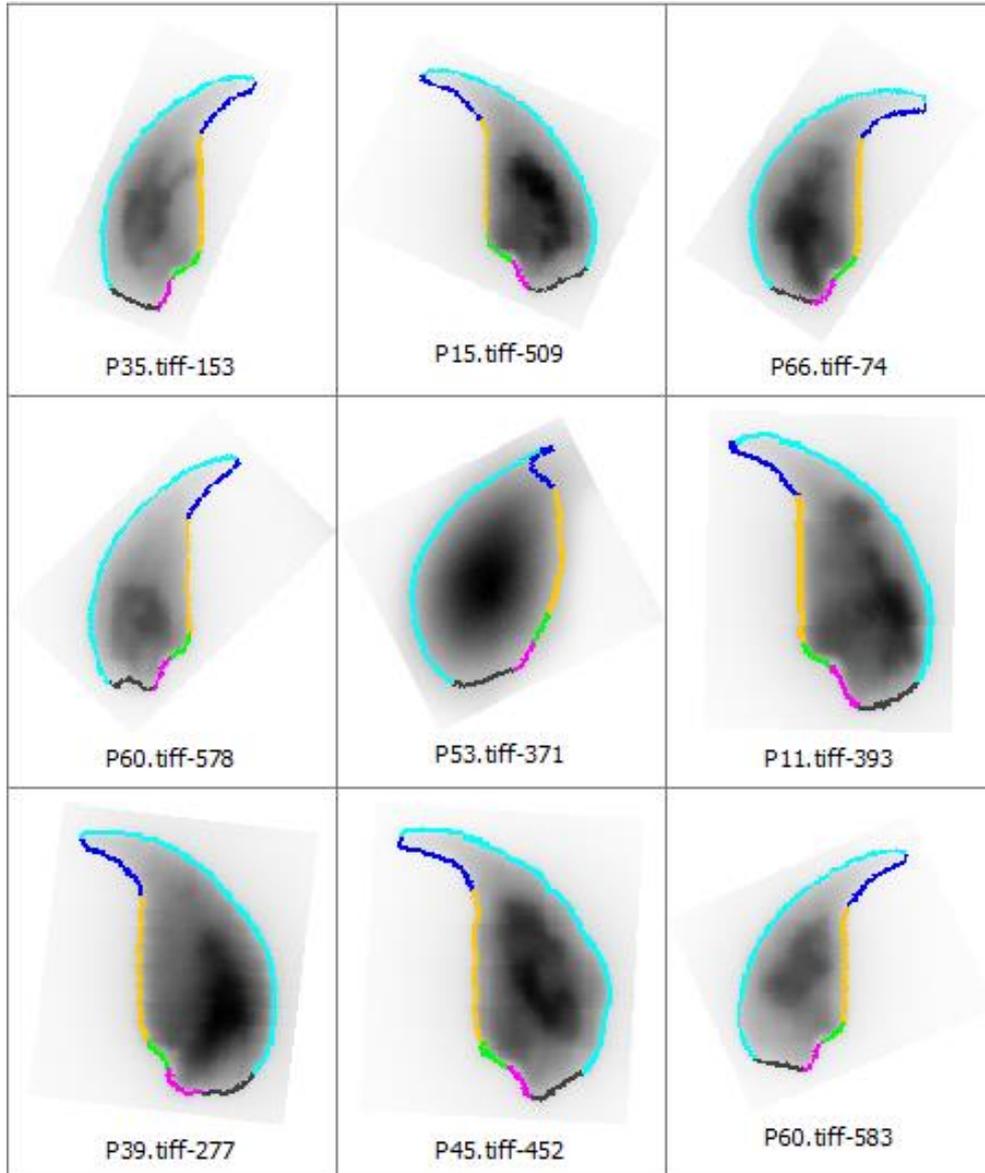


Figure 3.7. Screenshot of detected MF1XY^{R111} nuclei during the manual curation stage in Nuclear Morphology Analysis software. The central nucleus is an example of an out-of-focus capture that was discarded at this stage.

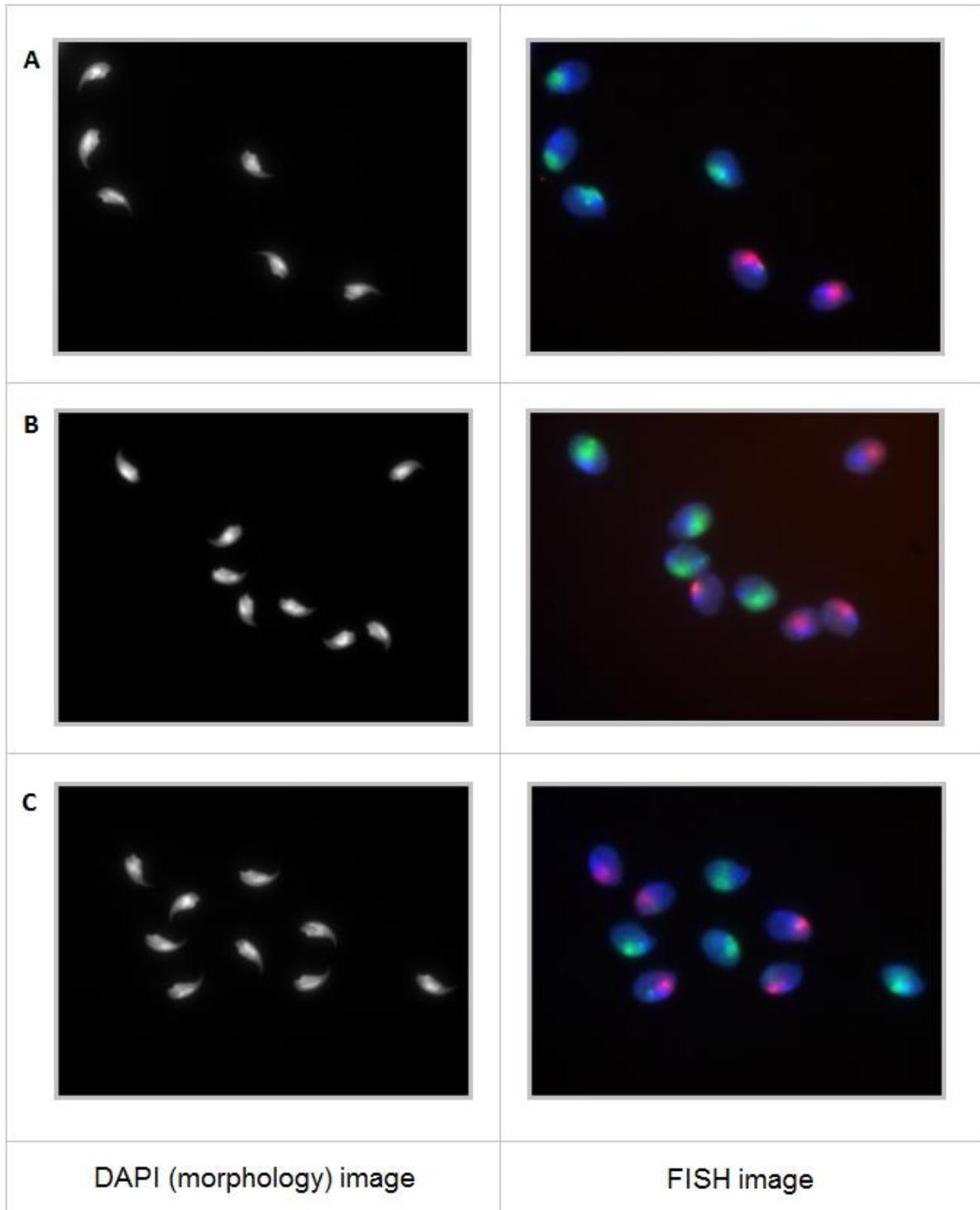


Figure 3.8. Example images of DAPI-stained sperm taken pre-FISH (left) and post-FISH (right). Sperm FISH requires swelling of the sperm heads, so images for morphological assessment must be taken prior to sperm-FISH and individual sperm heads relocated post-FISH for assessment of sex-chromosome content. The X chromosome was tagged with a green fluorophore, the Y chromosome with red.

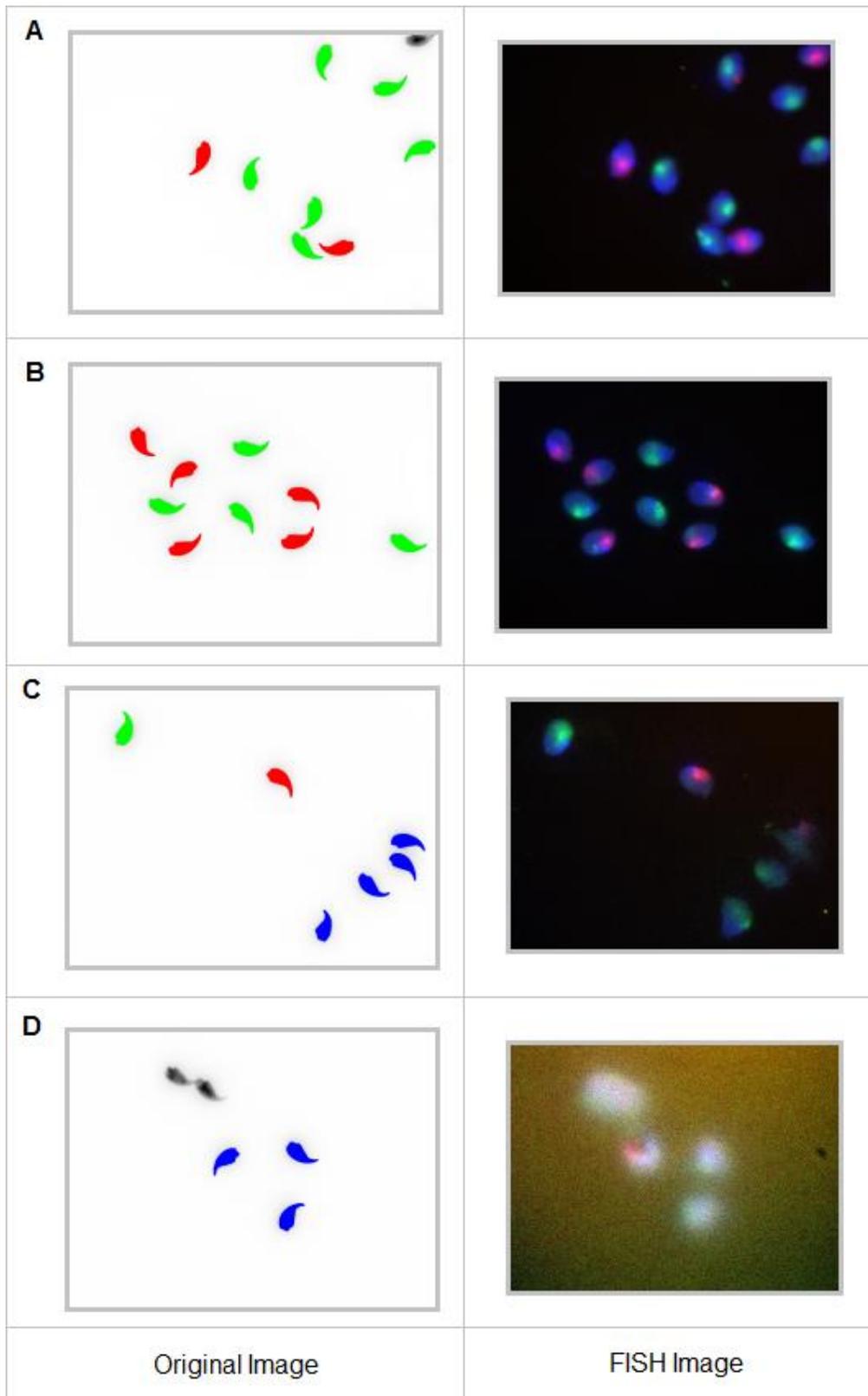


Figure 3.9. Examples of post-FISH mapping to pre-FISH images. Screenshots taken from the post-FISH mapping stage of quantitative morphological analysis in Nuclear Morphology Analysis software. In examples A and B, it was possible to assign a definitive signal category (X- or Y-bearing) to all successfully detected nuclei. Images C and D show some individual nuclei left unassigned where the nuclear signal appears ambiguous in the corresponding FISH image.

Following quality assessment of uploaded images and detected nuclei, a total of 2704 distinct nuclei-like shapes were detected. Of these, 2148 nuclei passed manual curation and could be conclusively determined as being either X- or Y-bearing. These were curated as described in *Table 3.4*. A total of 507 sperm could be confidently classified as X-bearing and 542 as Y-bearing sperm from YQ sperm populations and 555 and 544 respectively from WT populations; these were used in sex-chromosome content related morphology analysis and in comparative morphological assessments between WT and YQ population. Variation between animals was also assessed.

Strain	Number of sperm nuclei used in comparative analysis		
	X-bearing	Y-bearing	Total
WT	555	544	1099
YQ	507	542	1049
Total	1062	1086	2148

Table 3.4. Number of successfully detected sperm heads used in comparative analysis.

3.4.4. Results of quantitative sperm head morphology assessment

Results are broken into four main sections; a between strains analysis, a combined morphology and sex chromosome content analysis, a clustering analysis performed across all datasets, and a DNA content and sperm head size comparison.

Gross morphological differences between WT and YQ

Initial examination between strains shows divergent nuclear angle profiles between the populations. *Figure 3.10* shows consensus nuclear shape for each population. The consensus nucleus is generated by sampling the coordinates of each vertically-oriented nucleus at equivalent proportions around the perimeter and taking the median at each as the consensus point. The YQ sperm population shows a reduced overall size, distinctly reduced hook length, more flattened acrosomal region and reduced maximum height. Histograms of population distributions across selected parameters, such as shown in *Figure 3.11*, were used to visually determine that the data per population across all measured parameters appears normally distributed. This suggests that mean values provide an appropriate representation of the population. Further scrutiny of mean and median values showed them to be highly similar. *Table 3.5* shows major distinctions between WT and YQ sperm using mean values across a selection of parameters. An unpaired t-test was used to test for significance of difference between the population means for each measured parameter.

Where multiple hypotheses are tested in large datasets, type 1 error rate increases. To adjust for this, a Bonferroni correction was applied to retrieve an adjusted p.value significance level. In the below table, 7 different hypotheses were tested. This adjustment is calculated as follows:

$$p = \alpha / n \quad \text{where } \alpha \text{ equals the desired alpha (significance level) value and } n \text{ equals the number of hypotheses}$$

$$p^* = 0.05/7 = 0.0071$$

$$p^{**} = 0.01/7 = 0.0014$$

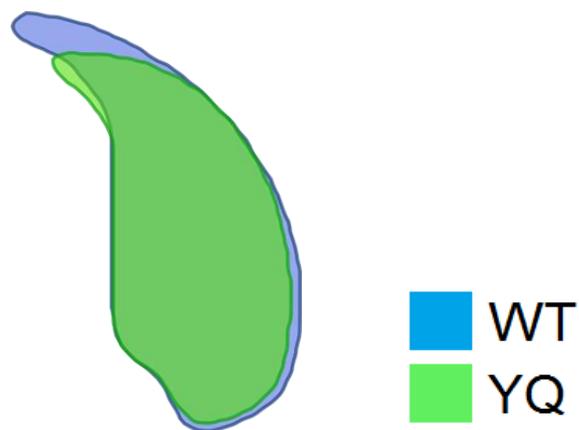


Figure 3.10. Overlay of the consensus nuclear shape from each of WT and YQ populations.

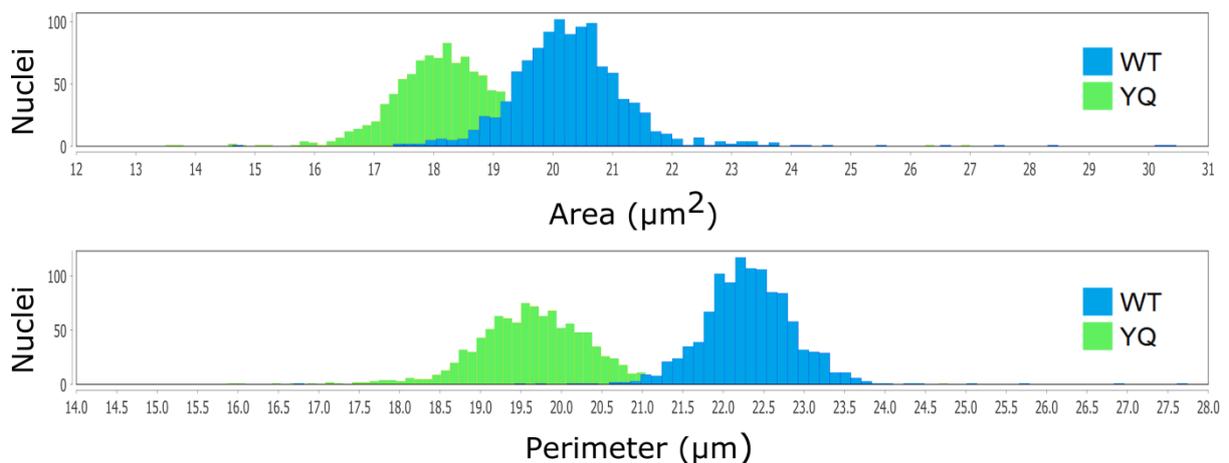


Figure 3.11. Histogram showing the distribution of data for two parameters used in morphology comparisons as an example of observed normal distribution.

Comparison of WT and YQ sperm heads using selected parameters				
Parameters	WT	YQ	% difference	p-value
Area (μm^2)	20.31	18.24	-10.19	**
Perimeter (μm)	22.32	19.65	-11.96	**
Max feret (μm)	8.74	7.63	-12.70	**
Circularity	0.51	0.60	17.65	**
Variability	5.62	7.19	27.94	**
Length of hook (μm)	1.91	1.20	-37.17	**
Width of body (μm)	3.51	3.29	-6.27	**

Table 3.5. Recorded values of and percentage difference between WT and YQ sperm populations by measured parameters. An unpaired t-test was used to test for significance of difference between the population means. ‘***’ indicates p.value at <0.0014.

Table 3.5 shows a significant difference between WT and YQ across several parameters. A difference of more than 10% between total area, perimeter, and maximum calliper diameter (recorded as max feret) of average sperm heads exists between the two populations. These are different measures of size that are intrinsically linked. The greatest difference can be seen in the mean length of the YQ sperm hook, which is less than two thirds of the WT hook. Both circularity and variability are increased in the YQ population. In the case of mouse sperm, max calliper diameter (max feret) represents the distance between the hook tip and the tail attachment region.

Area and perimeter are intrinsically linked parameters, representing different aspects of sperm head size. Figure 3.12 is a scatterplot of the two measures and presents a comparison between WT and YQ. It is easy to see two divergent populations using these determinants of size alone. To determine which regions of the sperm heads show greatest divergence between the two populations, sperm head angle profiles (as shown in Figure 3.13) were used to perform segmented profile comparisons. Segmented regions are defined by the greatest changes in the sperm angle profile; the local minima and maxima above and below 180 degrees in the median profile. A map of sperm head segments and region names is shown in Figure 3.14. An automated measure of each region was generated and could then be directly compared as below in Figure 3.15. As expected, all segments were reduced in size in YQ sperm. The regions that showed the greatest reduction were the under-hook region (segment 1) and the post-tail attachment region (segment 5). The extended acrosomal curve (segment 6) was reduced by 10.44%.

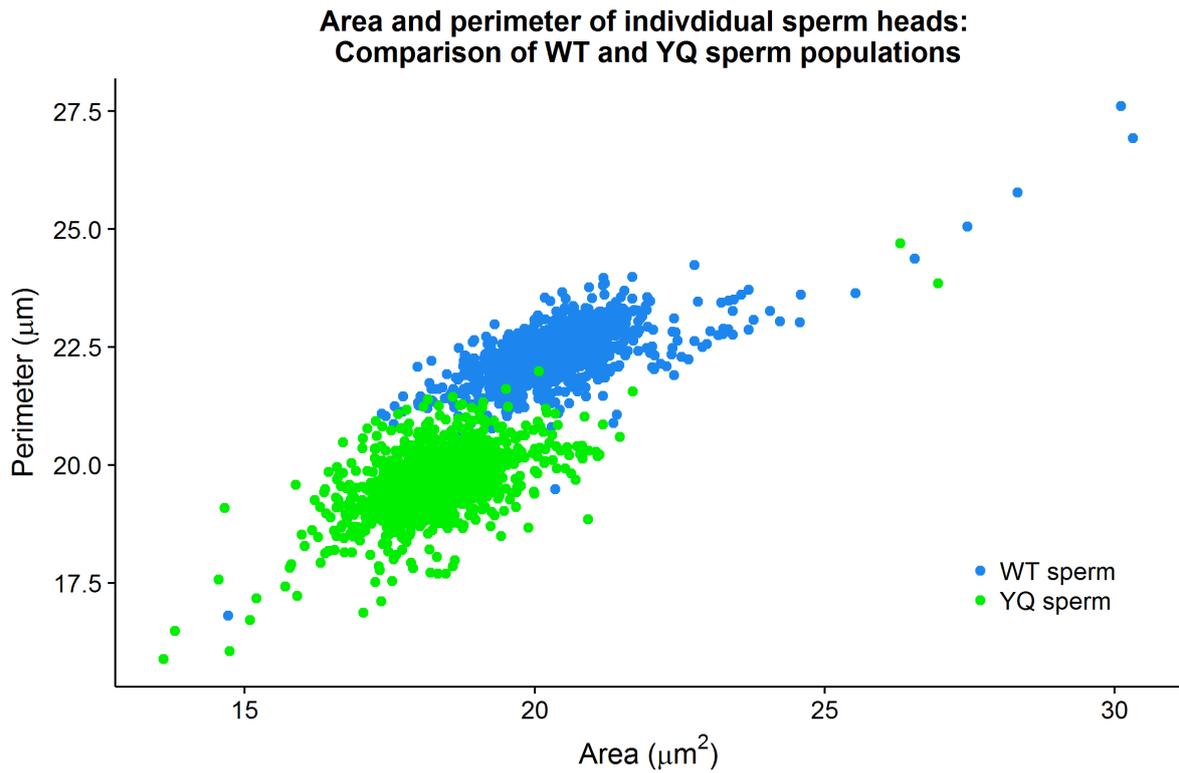


Figure 3.12. Scatterplot showing distributions of individual sperm heads from WT and YQ mice across two sperm head parameters; area (x-axis) and perimeter (y-axis).

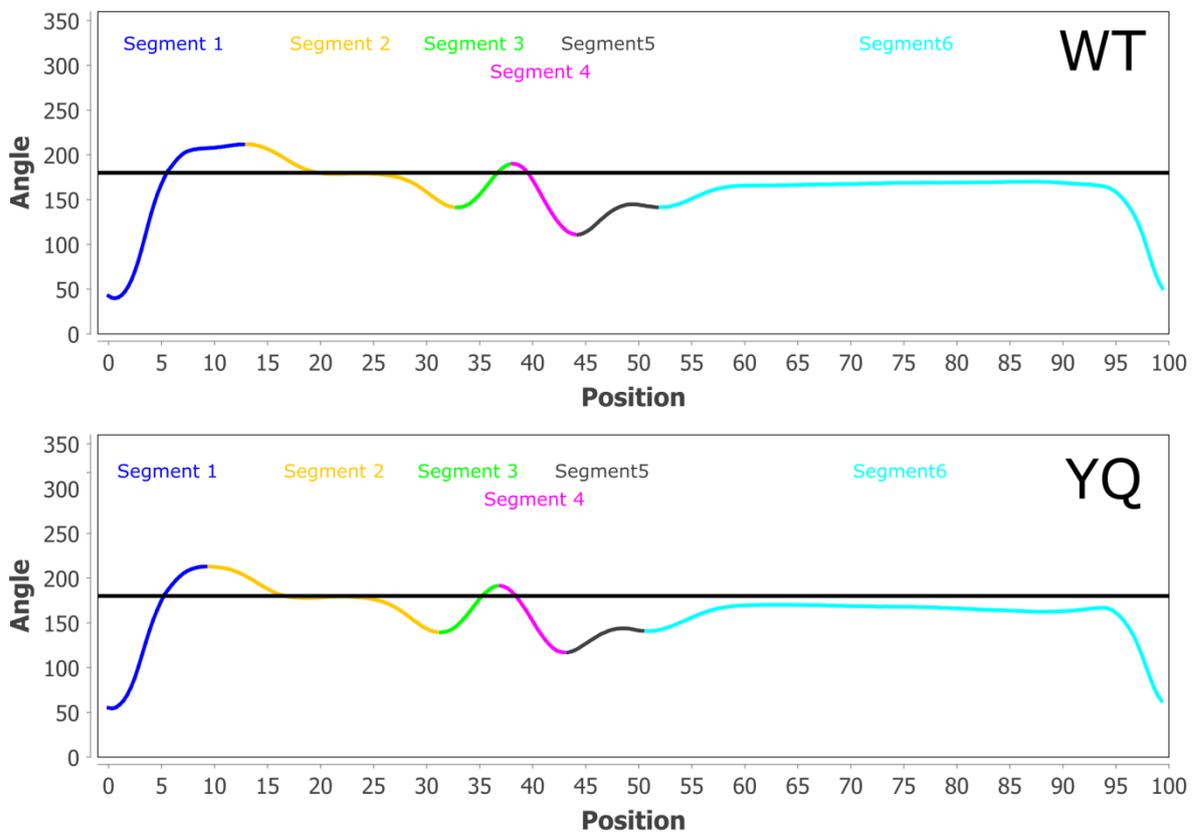


Figure 3.13. Segmented sperm head angle profile per population; WT (top) and YQ (bottom). Each coloured section denotes a separate segmented region of the sperm head.

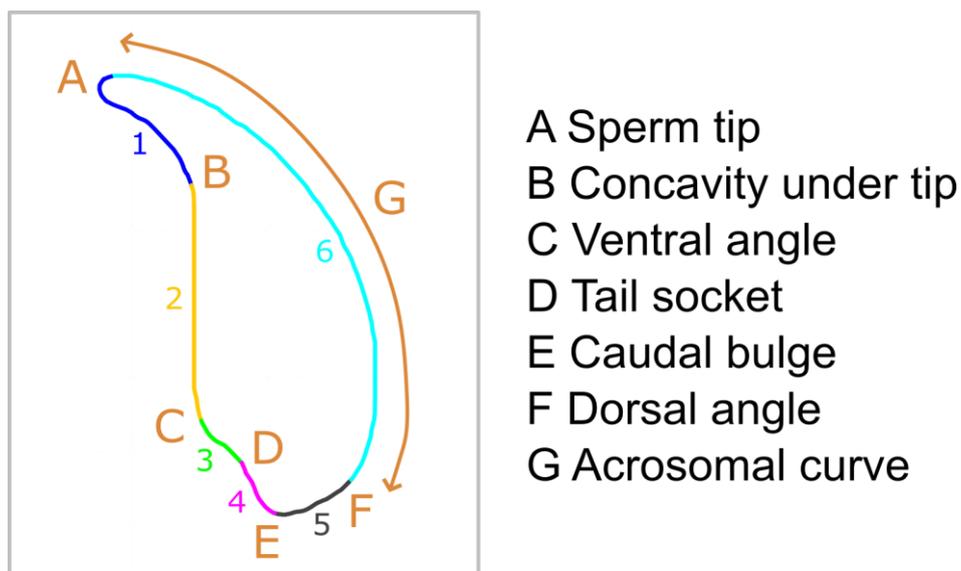
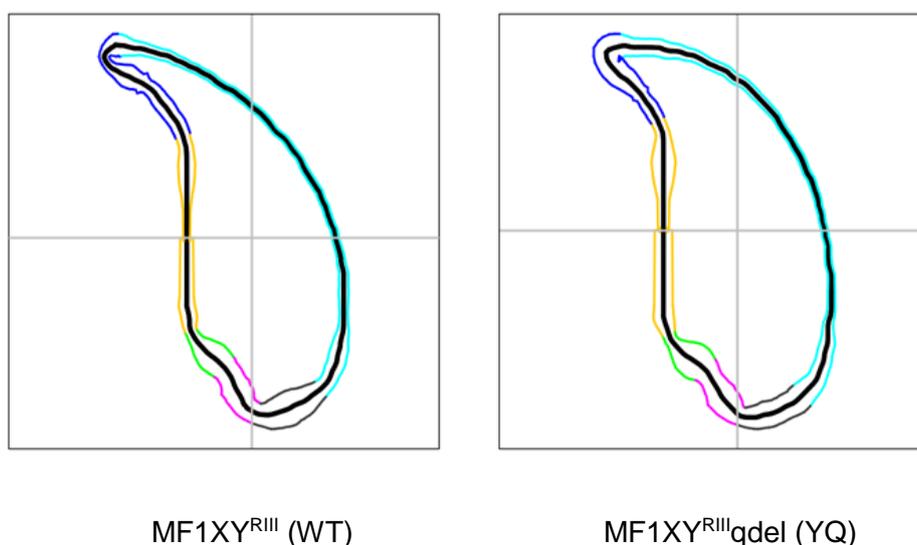


Figure 3.14. Diagram of descriptive sperm regions. Coloured and numbered segments relate to those seen in segment analysis and are defined further in *Figure 3.13* and *Figure 3.15*.



Strain	Length of different sperm head perimeter segments / μm					
	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
MF1XY ^{R^{III}} (WT)	2.77 \pm 0.01	4.58 \pm 0.02	1.11 \pm 0.01	1.32 \pm 0.01	1.71 \pm 0.01	10.82 \pm 0.02
MF1XY ^{R^{III}} qdel (YQ)	1.85 \pm 0.01	4.35 \pm 0.02	1.06 \pm 0.01	1.21 \pm 0.01	1.49 \pm 0.01	9.69 \pm 0.03
% difference	-33.21	-5.02	-4.50	-8.33	-12.87	-10.44

Figure 3.15. Comparison of segmented sperm head regions between WT and YQ sperm populations. The sperm head shape outlines (top) represent the average sperm head shape across all sperm from the given strain (WT on the left, YQ on the right). These are shown not to the same scale. The embedded table shows the lengths of defined sperm head perimeter segments in microns (rounded to two decimal places). The segment colours are reflected in the sperm head shape outlines. Values of standard deviation are given alongside perimeter length values. Where segmented colours are shown as an outline away from the consensus, this represents the interquartile range of the angle profile from that point of the consensus within the population.

Sperm head morphology and sex chromosome content

Comparisons were made between several aspects of sperm head morphology in the four subpopulations of WT and YQ sperm; WT X-bearing (WTX), WT Y-bearing (WTY), YQ X-bearing (YQX), and YQ Y-bearing (YQY). As shown in *Figure 3.16*, no differences were seen between WTX and WTY in visual examinations of a graphical overlay of consensus nuclei outline per population. However, a small difference, primarily in the hook tip and slight difference in overall head length were noted between the YQX and YQY populations. A scatterplot of area vs. perimeter was produced as in the previous comparison; however a notable overlap between X- and Y-bearing populations made distinguishing between the two difficult. Instead a comparison of probability density functions between the two parameters is shown in *Figure 3.17*. *Figure 3.17* shows a greater separation of within strain populations in YQ, with YQY representing the smaller population.

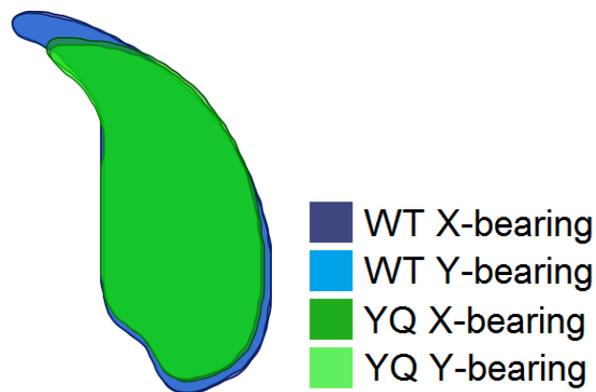


Figure 3.16. Graphical overlay of average sperm head shape from each X- and Y-bearing sperm population.

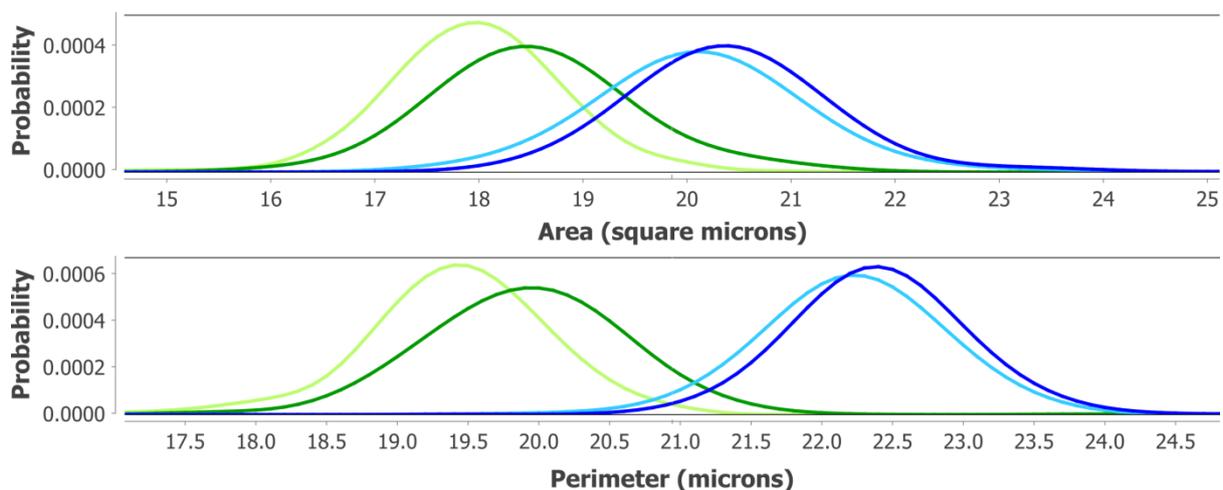


Figure 3.17. Probability density functions of two sperm head parameters; area (top) and perimeter (bottom). X- and Y-bearing populations can be seen to overlap within populations. YQ subpopulations appear more distinct than WT subpopulations. This image was adapted from output of Morphology Analysis Software. The X-axes have been shortened to focus on the area of interest.

A comparison of measured parameters (*Table 3.6*) confirmed greater differences in sperm morphology measures across the board, including measures of size, between YQX and YQY than between WTX and WTY. Length of hook showed the greatest proportional difference.

Adjusted p values were calculated as follows:

$$P^* = 0.05/14 = 0.0036$$

$$p^{**} = 0.01/14 = 0.0007$$

Parameters	Comparison of X- and Y-bearing sperm heads within strains							
	WT				YQ			
	WT X	WT Y	% difference	Sig test	YQ X	YQ Y	% difference	Sig test
Area (μm^2)	20.47	20.15	-1.56	**	18.56	17.95	-3.29	**
Perimeter (μm)	22.41	22.22	-0.85	**	19.91	19.40	-2.56	**
Max feret (μm)	8.77	8.70	-0.80	**	7.75	7.52	-2.97	**
Circularity	0.51	0.51	0.00	-	0.59	0.60	1.69	**
Variability	5.93	5.95	0.34	-	7.63	7.71	1.05	-
Length of hook (μm)	1.90	1.91	0.53	-	1.23	1.15	-6.50	*
Width of body (μm)	3.52	3.50	-0.57	-	3.31	3.28	-0.91	-

Table 3.6. Percentage differences between strain sperm sub-populations by measured parameters. ‘*’ indicates p.value at <0.0036, ‘**’ indicates p.value at <0.0007, ‘-’ shows the value indicates no significance. In YQ, the greatest differences are seen between area and length of hook. In WT, area and perimeter show the greatest divergence between subpopulations. Values in microns have been rounded to two decimal places for the purposes of displaying the data.

The measured difference between X- and Y-bearing sperm within populations was substantially less than that seen between strains. However, a significant difference between both WT sub-populations and between YQ sub-populations was seen; though with differing magnitude. The two parameters that showed significant difference in YQ but not in WT are circularity and length of hook. *Figure 3.18* shows the distribution of area values within each sub-population, where the magnitude of differences between populations can be clearly seen.

Comparison of WT and YQ sperm head area and variability within each population and by sex chromosome content

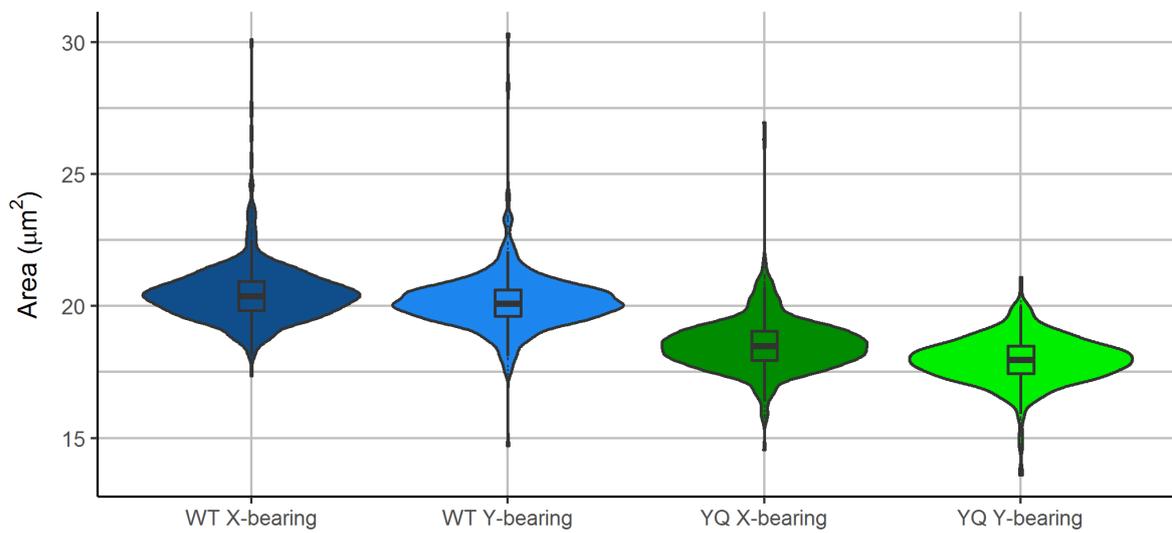
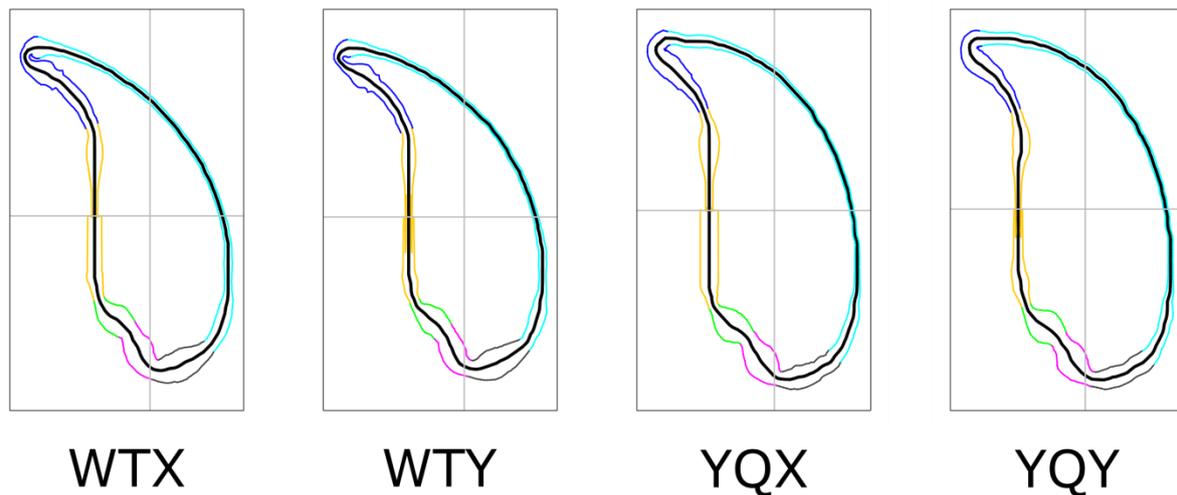


Figure 3.18. Violin plot comparison of WT and YQ sub-populations by area in square microns. This shows a greater distinction between YQX and YQY than is observed between WTX and WTY.

Analysis of sperm head regions following unbiased segmentation

Figure 3.19 gives results of segmented region measurements. Each measured parameter shows greater distinction between YQX and YQY than between WT subpopulations, except for segments 3 and 4 which form the tail-attachment region. A -4.04% difference in length is observed between WT subpopulations across this region compared to -1.25% in YQ.



	Length of different sperm head perimeter segments / μm					
Strain	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
WTX	3.05 ± 0.02	4.56 ± 0.03	1.20 ± 0.02	1.22 ± 0.03	1.77 ± 0.03	10.60 ± 0.03
WTY	3.04 ± 0.02	4.50 ± 0.03	1.19 ± 0.02	1.13 ± 0.02	1.75 ± 0.03	10.60 ± 0.03
% diff in WT	-0.54	-1.28	-0.87	-7.16	-0.67	0.03
YQX	2.11 ± 0.02	4.36 ± 0.03	1.17 ± 0.02	1.04 ± 0.02	1.58 ± 0.03	9.64 ± 0.04
YQY	2.04 ± 0.02	4.23 ± 0.03	1.17 ± 0.02	1.02 ± 0.02	1.53 ± 0.03	9.41 ± 0.04
% diff in YQ	-3.27	-3.09	-0.50	-2.09	-3.16	-2.40

Figure 3.19. Comparison of segmented sperm head regions between WTX, WTY, YQX and YQY sperm populations. The sperm head shape outlines (top) represent the average sperm head shape across all sperm from the given population. These outlines are shown not to the same scale. The embedded table shows the lengths of defined sperm head perimeter segments in microns. The segment colours are reflected in the sperm head shape outlines. Values of standard deviation are given alongside perimeter length values. Percentage difference in measurements from Y-bearing populations compared to their X-bearing counterparts within strain are given underneath the subpopulation measurements.

Separation of morphological groups by automated clustering

Using expectation maximisation clustering across all samples we were able to select out the most severely deformed sperm heads as determined by angle profile. The expectation maximisation algorithm is an iterative method of parameter maximum likelihood finding for unobserved variables (Dempster et al. 1977). Angle profile was used in isolation from other parameters to prevent clustering on already known size differences. Automated clustering retrieved 18 separate clusters of sperm heads. The most severely deformed sperm across all samples were identified by average nuclear shape and adherence to the cluster's angle profile; this is determined by visually examining tightness of fit of individual sperm angle profiles to that of the group. These sperm made up the two smallest clusters, totalling 34 sperm heads. An additional distinct cluster of sperm contained sperm displaying 'secondary' abnormalities; these included long and thin heads, bent hooks and enlarged heads. Finally, clusters of normal sperm and sperm with flattened acrosomes were seen. The origin dataset (WT X- or Y-bearing, YQ X- or Y-bearing) of each of these sperm heads was determined and the proportion of sperm belonging to each cluster is given in *Table 3.7*. This is displayed graphically in *Figure 3.20*.

Clustered dataset	% of sperm population represented by each clustered dataset			
	WT X-bearing (n=555)	WT Y-bearing (n=544)	YQ X-bearing (n=507)	YQ Y-bearing (n=542)
Severely deformed (n=34)	0.54	0.55	1.38	3.87
Secondary deformations (n=51)	1.80	1.10	3.55	3.14
Flattened acrosome (n=952)	0.36	0.55	90.34	90.22
Normal sperm (n=1111)	97.30	97.79	4.73	2.77

Table 3.7. Proportion of sperm populations represented by each clustered dataset

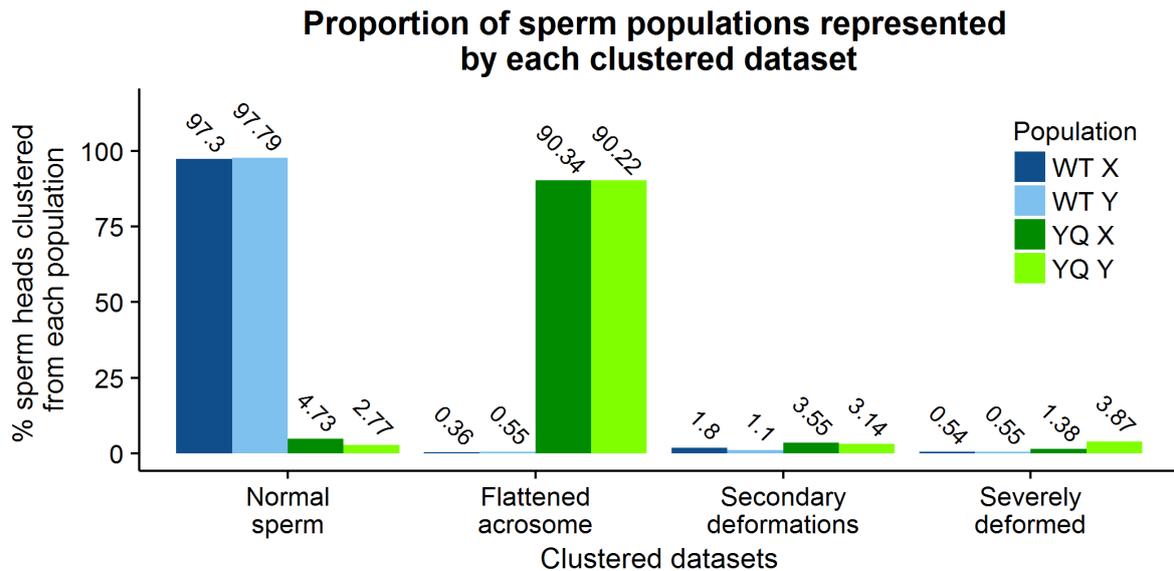


Figure 3.20. Proportion of sperm populations represented in each clustered (phenotype) dataset. Percentage values are given above bars.

A greater proportion of sperm were clustered into a severely abnormal category in YQ samples than in WT, and more in YQY than in YQX; although both strains possess severely deformed sperm in very low numbers. Sperm from the two strains show very little overlap in clustered phenotypes.

Differences in DNA content and comparisons with area measurements

Across sperm populations of both mouse strains, Y-bearing sperm appeared reduced in size, measured in 2D area, compared to X-bearing counterparts (-1.56% in WT, -3.29% in YQ). And YQ sperm reduced by -10.12% compared to WT sperm. We know that the total DNA content of these sperm populations will differ due to the haploid nature of sperm and the diminished Y chromosome of the YQ. In order to discern whether differences in size might be accounted for by differences in total DNA content, comparisons of overall DNA length (in bases) as listed in *Table 3.8*, and observed sperm size differences (using 2D area) as listed in *Table 3.9*, were made. DNA length values were collected from Ensembl Genome Browser, mouse genome GRCm38.p5⁵. YQ Y chromosome volume was calculated by subtracting 66% length of the mouse MSYq reported as 86.0Mb in length (Soh et al. 2014) from the standard Y chromosome length.

Differences in WT and YQ area cannot be accounted for by differences in DNA content (-1.07% DNA and -10.12% area). However, differences between X- and Y-bearing sperm in both strains show the opposite; difference in DNA content is greater than measured difference in area. WTY show -2.92% DNA difference and -1.56% difference in area compared to WTX; YQY show -5.03% DNA and -3.29% area compared to YQX. *Figure 3.21* provides a comparison of three major sperm head parameters and DNA content between all populations.

⁵ Ensembl Genome Browser - http://www.ensembl.org/Mus_musculus/Info/Index

Expected % difference in DNA content (bases) between sperm populations				
	Sperm population B			
Sperm population A	WT X	WT Y	YQ X	YQ Y
WT X	0.00	-2.92	0.00	-5.03
WT Y	3.01	0.00	3.01	-2.17
YQ X	0.00	-2.92	0.00	-5.03
YQ Y	5.29	2.22	5.29	0.00

Table 3.8. Difference in DNA content between individual sperm belonging to different sperm populations. Percentage values represent percentage difference of population B from population A. Percentage values represent differences in DNA content as calculated using Ensembl genome browser and Soh et al. 2014. Blue highlights show comparisons of greatest interest.

% difference in area (μm^2) between sperm populations				
	Sperm population B			
Sperm population A	WT X	WT Y	YQ X	YQ Y
WT X	0.00	-1.56	-9.33	-12.31
WT Y	1.59	0.00	-7.89	-10.92
YQ X	10.29	8.57	0.00	-3.29
YQ Y	14.04	12.26	3.40	0.00

Table 3.9. Percentage difference in sperm head area originally measured in square microns between sperm belonging to different sperm populations. Percentage values represent percentage difference of population B from population A. Blue highlights show comparisons of greatest interest.

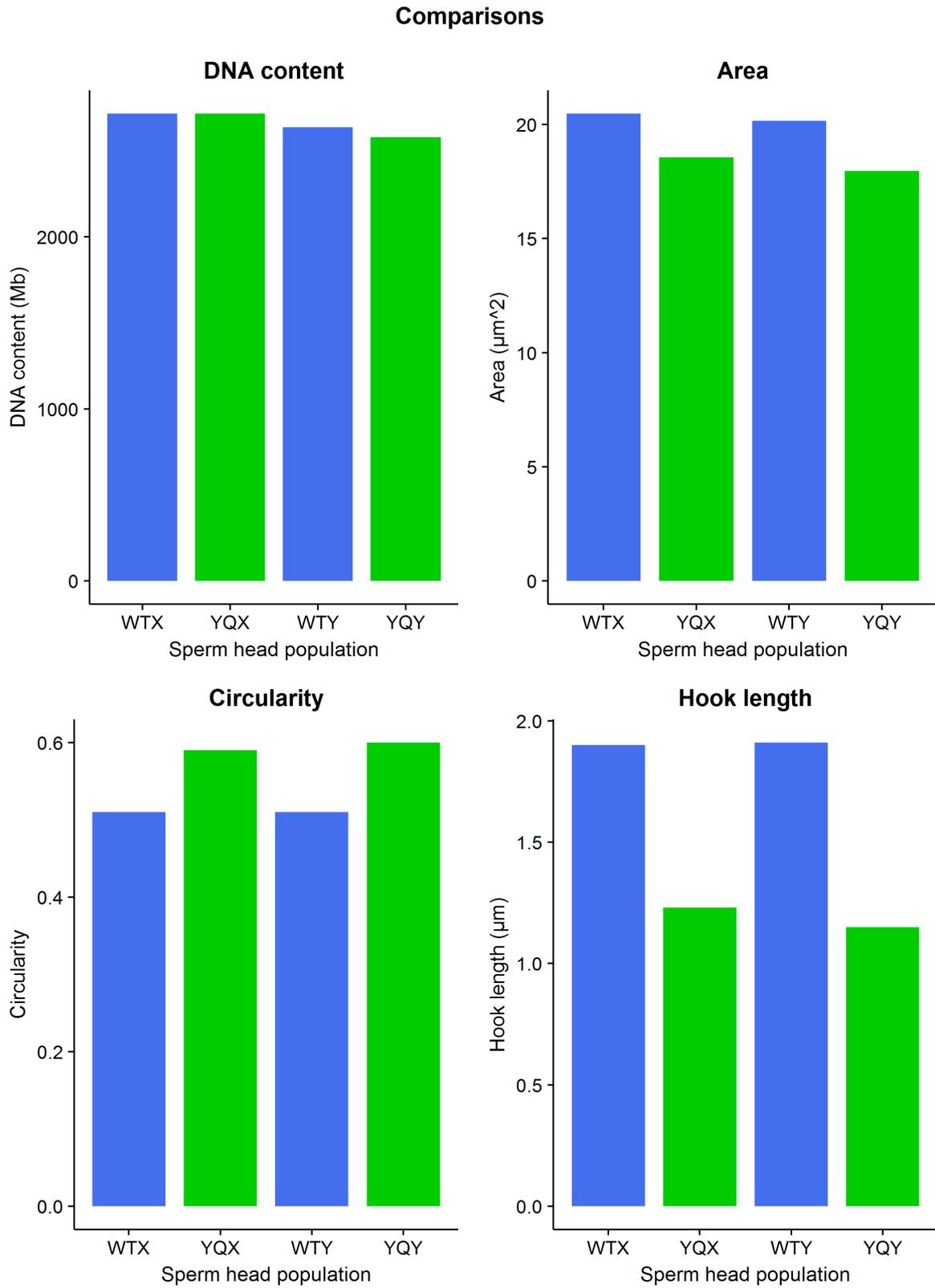


Figure 3.21. Comparison of three major sperm head parameters and sperm DNA content.

3.4.5. Discussion of findings of quantitative sperm head morphology assessment within and between strains

The above data show that sperm from YQ mice are smaller than WT sperm, display a distinct morphology (flattened acrosome) with differences appearing exacerbated in YQY. This difference is greater than that observed between WTY and WTX. Originating mouse strain is the dominant factor in sperm head differences in these populations. Differences in DNA content do not account for recorded differences in 2D sperm head size (area) between WT sperm and YQ sperm. In combination, these examinations reveal that whilst differences in DNA condensation from WT seem apparent across both subpopulations of YQ, the difference in DNA content between YQX and YQY more than accounts for the difference in overall area.

Sperm show differences in 2D size and not just shape between strains; a difference also exists between X- and Y-bearing populations within strains

Unbiased quantitative assessment has allowed us to determine that sperm nuclei differ in 2D size and not just shape between strains, not accounted for by difference in DNA content. The combination of decreased 2D area, perimeter and max feret measurements of sperm size with increased circularity and variability suggest aberrant DNA condensation may have a role to play in these differences across all YQ sperm. Altered condensation of the chromatin may give rise to the reduction in 2D nuclear size observed. Given the information, aberrant DNA condensation does not appear to be a characteristic particular or exacerbated in YQY and is unlikely to be a source of functional inequivalence between YQX and YQY. This is because the differences in area seen when compared to DNA content echo those seen between WTX and WTY. Defective chromatin condensation has also been associated with an increased variability in sperm head shape (Chemes and Sedo 2012, Beletti et al. 2005), a characteristic seen when comparing between strains but not between populations within strains in our data. The process of DNA condensation during spermiogenesis is a complex and highly regulated process (section 1.3), and aberrant chromatin condensation is a characteristic previously noted in sperm from males with large (>66%) MSYq deletions, though with a limited sample size and variability within models with smaller deletions (Yamauchi et al. 2010). A measurement of the Z-axis, providing insights into 3D shape, is required before conclusions can be drawn. A failure of appropriate flattening of the sperm head could result in comparatively poor motility of affected populations, providing a functional difference between strains and between YQ subpopulations. In future investigations, Z-axis sperm head measurements could be used to determine whether failure for DNA to fully condense is a likely cause of the abnormal morphology seen in YQ. This is further explored in the Further Work section in Chapter 7.

The existence of differences in 2D size between WTX and WTY populations not equivalent to differences in DNA content suggest that any subsequent effects of this difference are compensated for by one of the populations, or that the difference is so minute as not to be linked to differences in functional capacity between X- and Y-bearing sperm.

Measurement of specific regions of the sperm head and parameters of sperm shape and size reveal differences in X- and Y-bearing populations that are not consistent between strains

Sperm populations show significant differences across all seven measured parameters between strains. However, when examining populations of X- and Y- bearing sperm, some features only appear significantly different between YQ populations. Hook length and circularity show greater differences between YQ sub-populations than between WT sub-populations. This matches with the results of subsequent examinations of segmented sperm head profiles, where under hook length shows the greatest proportional difference between YQX and YQY sperm.

Decreased overall sperm head size is reflected in measurements of specific sperm head regions. Nuclear hook length shows a disproportionate reduction in YQ sperm compared to WT. Formation of the sperm hook (measured by resulting hook length) is strongly affected in both subpopulations of YQ sperm. WT hooks are, on average, 38.4% longer than YQ hooks. An additional distinction exists between YQ X- and Y-bearing sperm, whilst no difference exists between WT subpopulations. The contribution of each of the nucleus and the acrosomal granule to the hook structure and size is not well defined (Tourmente et al. 2016, Breed et al. 2014), much like the hook's competitive consequences. Nevertheless, several genes have been linked to hook development, and provide new candidates for investigation in subsequent work.

Although the extended acrosomal region showed a 10.44% reduced length in YQ sperm, a more noticeable difference can be seen in the curvature when segmented profiles are compared alongside one another (*Figure 3.15*). The change in size and curvature of this region might be explained in part by the marked reduction in under hook length (-33.21% in YQ).

Whilst cumulative difference in sperm head perimeter is greater between YQ X- and Y-bearing sperm than in WT (WTX vs WTY -0.85%; YQX vs YQY -2.56%), a proportionally larger difference was noted between WT subpopulations in segmented regions 3 and 4 which make up the tail attachment region. Why this might be the case and what consequences it might have are uncertain, however it does not appear to be associated with differences in functional capacity between the WTX and WTY, as reinforced by examination of offspring sex ratios in section 3.2.

Quantitative techniques reveal differences previously unseen using traditional qualitative methods

Whilst qualitative examinations led to the identification of a separate ‘severely flattened’ sperm head phenotype, this distinction was not noted in quantitative analysis. Within the quantitative data ‘flattened acrosome’ cluster, there were sperm that showed slightly more flattening; however this was not readily distinguishable as in the qualitative assessment. This discrepancy may result from inefficiencies in the qualitative technique (the perception of categorical differences where there are none), however the distinction between ‘flattened’ and ‘severely flattened’ in qualitative analysis may also arise from morphological distinction in the outer proteinaceous acrosome itself, and would therefore not be observed when DAPI staining the nucleus.

A large number of genes, including several sex-linked genes, are known to be involved in sperm head chromatin condensation and sperm hook formation. Particular attention will be paid to genes with these functions in subsequent chapters.

A technical note on probe efficiency is provided in the Supplementary Material.

3.5. Assessment of Sperm Hyaluronidase Activity in the Mature Sperm of MF1XY^{RIII}qdel Mice

3.5.1. Introduction to assessing hyaluronidase activity

Hyaluronidases perform a key role in fertilisation by catalysing the digestion of hyaluronic acid which is a major component of the cumulus oophorous that surround the oocyte. Each sperm carries its own hyaluronidase proteins which are translated and relocate to the acrosome during spermiogenesis. Hyaluronidase activity and gene function has been linked to cases of transmission ratio distortion in mouse (Zheng et al 2001a) and restricted sharing of some hyaluronidase gene products has been shown (Zheng et al. 2001b). Hyaluronidase activity represents a potential exponent of functional inequivalence between X- and Y-bearing sperm in the MF1XY^{RIII}qdel deletion mouse.

The sorting of sperm by sex chromosome content for examination of hyaluronidase activity has not been possible for this investigation. It is also not fully understood how flow sorting of mature sperm might affect hyaluronidase activity or hyaluronidase sperm surface protein retention. If a proportion of either sperm population (WT or YQ) demonstrate deficiencies in hyaluronidase activity, this may be detectable as a reduction across the whole population. It may also be the case that, as with previous investigations, effects are seen across the entirety of the YQ population and are exacerbated in YQ Y-bearing (YQY) individuals. Therefore, hyaluronidase activity was assessed across sperm populations pooled by originating strain and examined accordingly.

Hypothesis: Sperm from MF1XY^{RIII}qdel mice differ in the level of sperm surface hyaluronidase activity from that of wild-type MF1XY^{RIII} sperm.

3.5.2. Structure and execution of hyaluronic acid activity assessment

Hyaluronidase activity can be measured indirectly by determining the level of by-products that are present following the catalysis of hyaluronic acid by hyaluronidase, or by the subsequent absence of hyaluronic acid itself. For the full methodology, see section 2.3.1.

Calibration using enzyme curves

In order to assess differences between mouse models, first a standardised enzyme-only transmittance curve was established. Preliminary experiments resulted in adjusted timings, concentrations, reagents and methodology to produce the protocol detailed in section 2.3.1. Initial experiments focused on establishing a suitable range of enzyme concentrations and discerning optimal length incubation periods. MF1XY^{RIII} sperm was often tested alongside enzyme-only samples to discern any immediate, sperm-specific issues that might cause complications in later experiments.

Several optimisation experiments were conducted to determine an appropriate range of enzyme concentrations.

Incubation periods

Using enzyme-only data, maximum absorbance was reached at a concentration of 2u/100µl after 1 hour primary incubation. Further incubation, after maximum turbidity was reached, showed no additional change in the sample (*Figure 3.22*).

Sperm and enzyme diluent media selection

It was observed that when sperm were extracted into PBS and incubated with HA, a number of sperm subsequently displayed detached heads, and reduced motility was observed across both strains. The high osmolarity of PBS, relative to sperm, may have caused damage to the cells over extended *ex vivo* periods. Human tubal fluid (HTF) is a media commonly used for sustaining sperm post-extraction from the testes in *in vitro* fertilisation (IVF) experiments. It was subsequently found that human tubal fluid (HTF) better sustained the spermatozoa, reduced incubation times and inhibited enzyme activity less than PBS, producing a linear trendline closer to a concentration of zero (*Figure 3.23*). PBS was substituted for by HTF for subsequent hyaluronidase activity experiments.

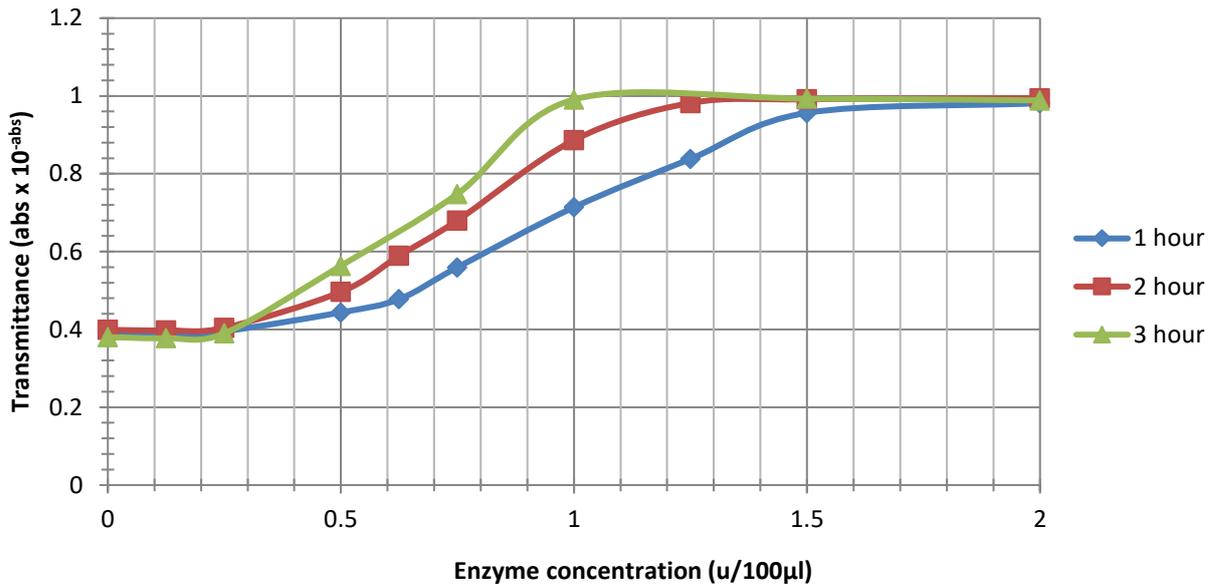


Figure 3.22. Comparison of primary incubation times used in enzyme-only experiments; secondary incubation time remained constant at 30 minutes. Absorbance readings taken from multiple experiments and expressed as mean for each incubation length. Phosphate buffered saline (PBS) was used as the enzyme diluent.

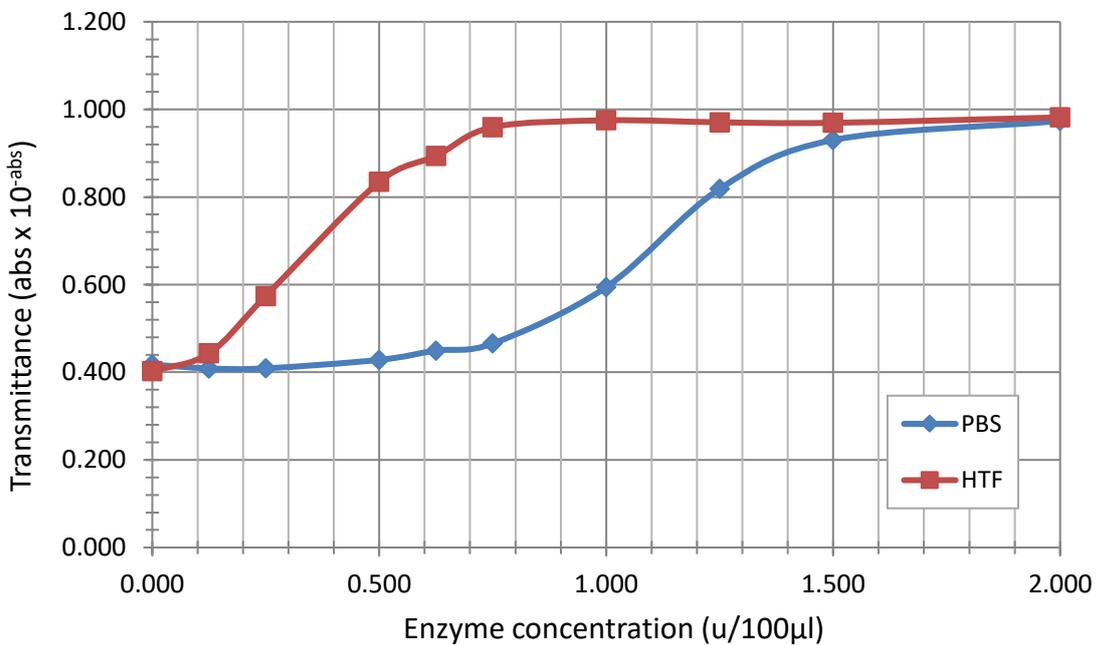


Figure 3.23. Comparison of buffer solutions phosphate buffered saline (PBS) and human tubal fluid (HTF) used in enzyme-only hyaluronidase activity turbidimetric experiments, after 1.5 hours primary incubation and 30 minutes secondary incubation. Data was obtained from four separate comparison tests and presented as a mean average for each data point.

Determining an appropriate sperm concentration range

Experiments using live sperm began with the primary intention of defining suitable concentrations of sperm cells for use in the assay such that resulting transmittance could be easily plotted on the linear part of the enzyme-only curve and hyaluronidase activity estimated per number of sperm. A concentration range of 2.5×10^5 to 1.0×10^6 cells/100 μ l was found to be well suited. Hyaluronidase activity was measured at three different sperm concentrations: High = 1×10^6 cells/100 μ l; Medium = 5×10^5 cells/100 μ l; Low = 2.5×10^5 cells/100 μ l. Sperm concentration was calculated using a haemocytometer and trypan blue staining of fixed sperm (see section 2.1.3).

Examination of hyaluronidase activity across two sperm populations

Experiments compared sperm from age-matched mice of both MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ) strains at three sperm concentrations (1×10^6 , 5×10^5 and 2.5×10^5 cells/100 μ l). Two replicate assays were performed for each mouse, producing a mean enzyme activity value for each individual sperm sample. Sperm from eight mice (four mice from each strain), examined in WT/YQ pairs over four separate experimental days, were used to generate the data. Enzyme-only experiments were run in parallel on each day to produce enzyme-only activity curves required for calculating equivalent sperm hyaluronidase activity values. Enzyme-only data were combined across all experimental days to produce a linear model of enzyme-only concentration and transmittance ($R^2 = 0.9974$), resulting in one linear equation being used to determine sperm enzyme activity across all samples (*Figure 3.24*).

It was determined that splining of enzyme-only data to produce a linear equation describing the relationship between enzyme concentration and transmittance (using concentration values 0.125u/100 μ l to 1.0u/100 μ l) could be used as an appropriate determinant of sperm enzyme concentration for measured transmittance values between 0.414 and 0.918.

Linear regression of sperm assay transmittance values were used to determine the concentration of hyaluronidase in sperm samples across all measured sperm concentrations.

$$Y = 0.5788x + 0.3445$$

Rearranged as follows:

$$X = (Y - 0.3445) / 0.5788$$

Enzyme-only fixed concentration transmittance

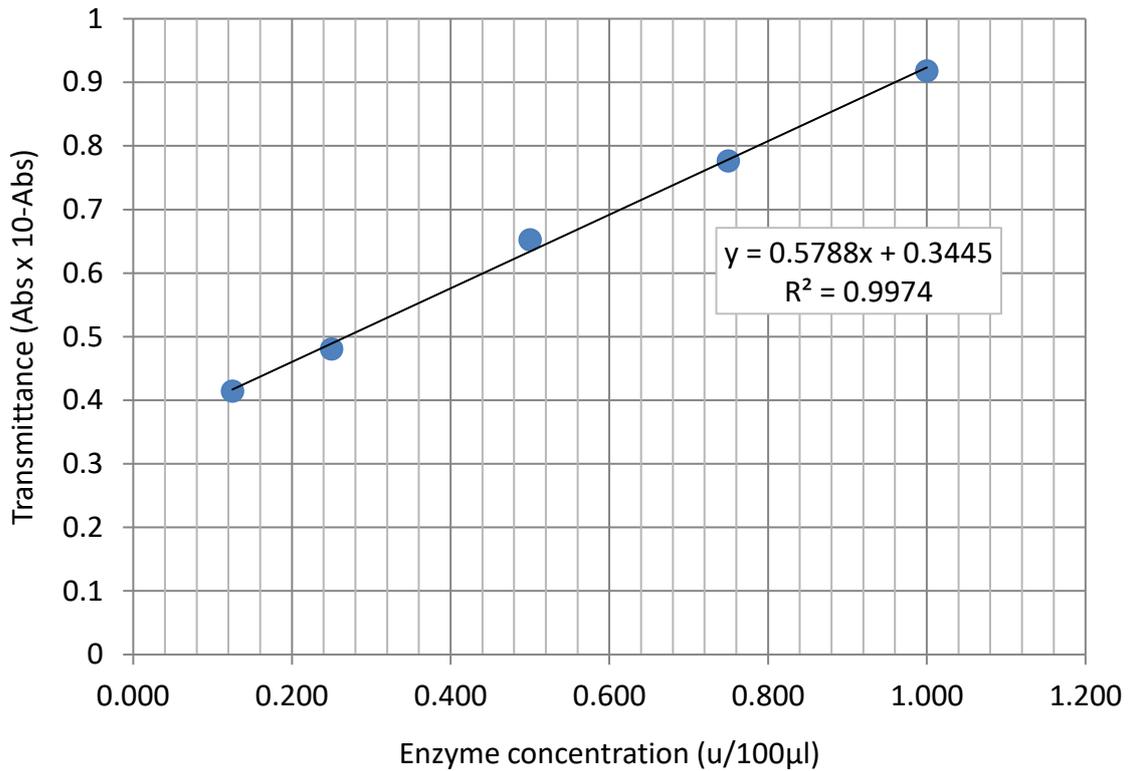


Figure 3.24. Enzyme-only concentration versus transmittance showing the linear region of the enzyme curve only. Mean values for transmittance were calculated from seven enzyme-only replicates to produce the above values.

Concentration (u/100ul)	Transmittance (10 ^{-Abs})	Standard error
2.000*	0.992*	0.002
1.500*	0.988*	0.003
1.000	0.918	0.014
0.750	0.777	0.022
0.500	0.652	0.011
0.250	0.481	0.008
0.125	0.414	0.009
0.000*	0.387*	0.010

Table 3.10. Values of transmittance as calculated across seven enzyme-only experiments conducted in parallel to sperm-based experiments. *Values excluded from construction of enzyme-only standard curve (Figure 3.24).

3.5.3. Results of sperm head hyaluronidase assay

Figure 3.25 shows enzyme concentration as calculated at each sperm concentration per animal; *Figure 3.26* shows these values as a mean across all concentrations. Standard error values of replicate measurements for each animal or across all individuals in populations are shown. Values of sperm concentration are as follows:

High = 1×10^6 cells/100 μ l

Medium = 5×10^5 cells/100 μ l

Low = 2.5×10^5 cells/100 μ l

Overall equivalent hyaluronidase enzyme concentration in YQ population sperm was determined to be 99.4% of that in WT. A two-tailed T-test was applied to values across all sperm concentrations to discern whether a significant difference in equivalent hyaluronidase enzyme concentration in sperm could be discerned between WT and YQ populations.

No significant differences were detected in the mean value of equivalent hyaluronidase enzyme concentration between WT and YQ populations. The data show that the null hypothesis should not be rejected. Both enzyme-only activity and equivalent sperm enzyme activity were tested for batch effects across the four data collection days and no batch effect was detected (data not shown).

Comparative enzyme concentration at high, medium and low sperm concentrations across WT and YQ individuals

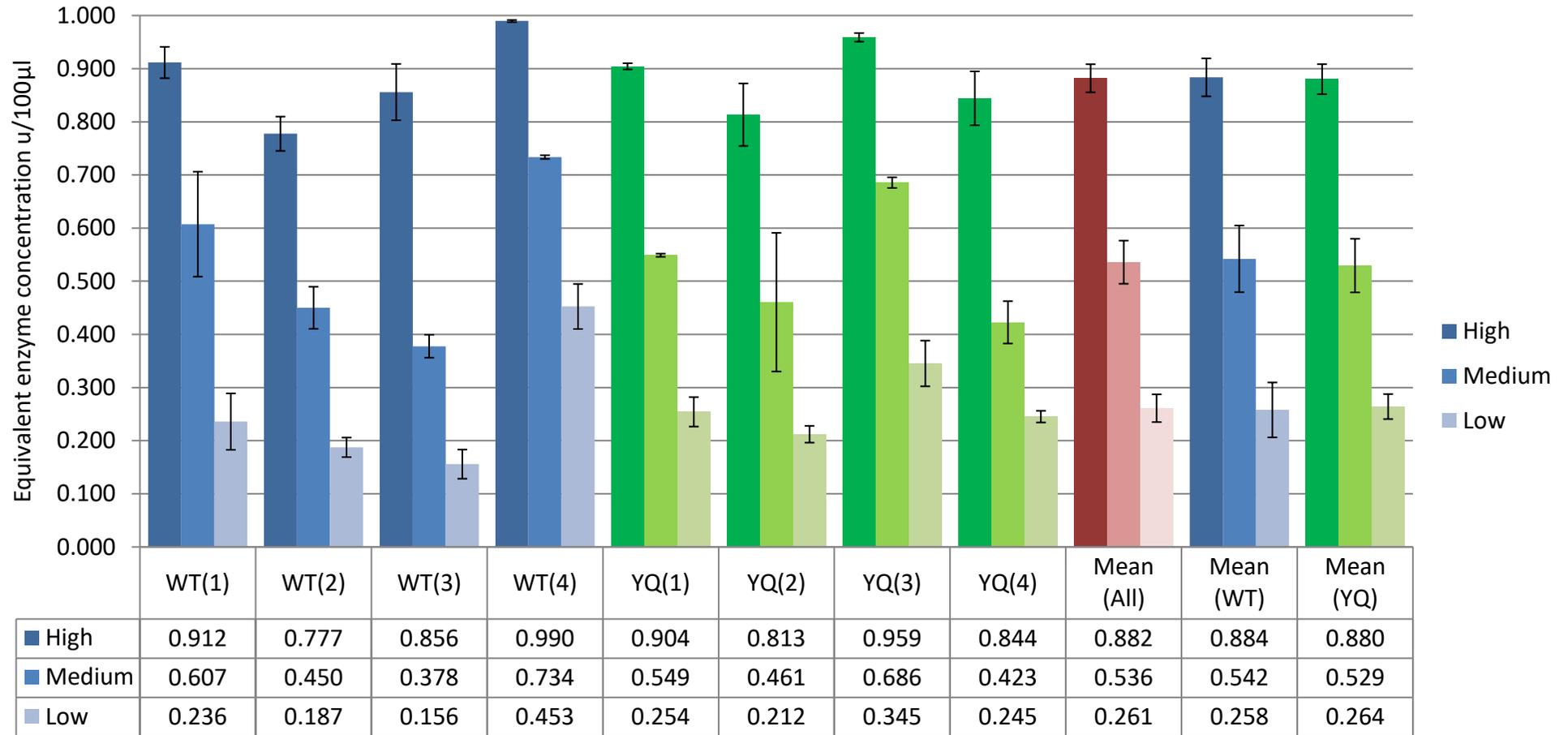


Figure 3.25. Equivalent hyaluronidase concentration of sperm samples from individual MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ) mice across all sperm concentrations (high, medium, low). The mean average value across all individuals and across the separate populations of WT and YQ are also shown. Standard error of replicate measurements for each animal/across all individuals in populations shown.

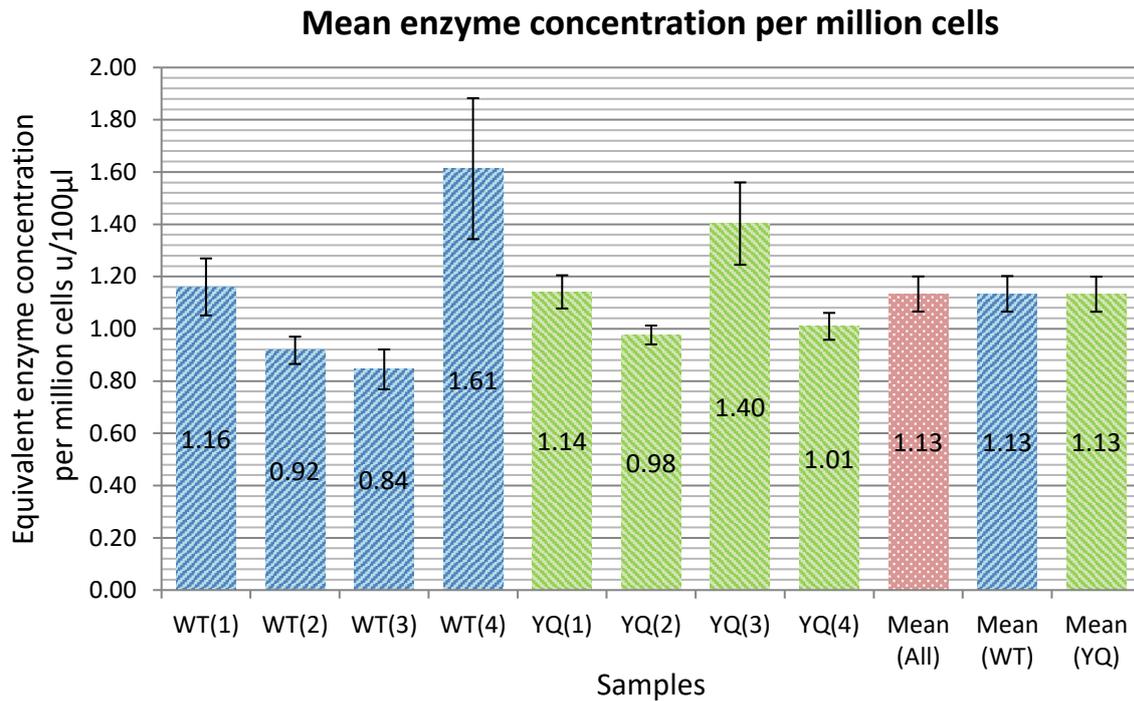


Figure 3.26. Equivalent hyaluronidase concentration of sperm samples from individual MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ) mice. Data is represented as the mean enzyme concentration per million cells taken across all sperm concentrations (high, medium, low) and displayed across all individuals. The mean average value across all individuals, and then across the separate populations of WT and YQ are also shown. Standard error of measurements across all concentrations is shown.

3.5.4. Discussion of findings of sperm head hyaluronidase assay

By assessing hyaluronidase activity in enzyme-only samples and constructing an enzyme activity curve to determine the equivalent enzyme concentration of sperm samples, it has been possible to compare enzyme activity in sperm from MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ). Observations and subsequent statistical testing have shown there to be no difference in the hyaluronidase concentration between sperm samples from the WT and YQ mouse models on this occasion.

Although high reproducibility between transmittance values in enzyme-only experiments across experimental days was achieved (standard deviation <0.058 at each enzyme concentration), and between replicates performed with the same mouse (standard error <0.06 for each mouse), the range of enzyme concentrations recorded between individuals within strains was large (up to 36% difference between lowest and highest concentrations from mice within the same strain). The difference in recorded enzyme activity between individuals is greater than the difference between the mean enzyme activity values of the two populations.

It may be the case that were Y-bearing sperm hyaluronidase activity reduced, the activity of X-bearing sperm may be relatively increased, potentially owing to the competitive uptake of hyaluronidase proteins or retention of transcripts (Griffiths et al. 2008). In this scenario, differences in hyaluronidase activity between X- and Y-bearing sperm would not be detectable using the above method, and sorting of mature sperm by sex chromosome content would be required to confirm these differences. Immunostaining for hyaluronidase proteins such as SPAM1 in sperm followed by sperm FISH both before and after exposure to oviductal fluid could provide detailed information on distribution amongst sperm populations.

3.6. Chapter Summary

In this chapter, I reported the findings of several distinct physiological investigations of the MF1XY^{RIII}qdel model with the aim of further understanding factors involved in reduced sperm function between X- and Y-bearing sperm within YQ sperm populations, and global differences between the YQ and WT strains. A sex ratio skew of 6% was noted in YQ offspring compared to WT (49.5% males in WT, 43.5% in YQ), and a reduced fecundity (as measured by mean litter size) from 12.4 to 11.1 offspring. Targeted investigations revealed a distinct ‘flattened acrosome’ morphology that was observed using both protein and DNA visualisation techniques. A reduced overall sperm size was noted in YQ sperm heads compared to WT, which to a lesser extent was also evident between YQ and WT subpopulations (X- and Y-bearing). Hyaluronidase assays did not reveal any differences in hyaluronidase activity between the two strains. Whilst hyaluronidase activity has been mentioned as a functional facilitator of transmission ratio distortion and sex ratio distortion in other mouse models (Zheng et al. 2001a), we have not found any evidence for this in MF1XY^{RIII}qdel.

Directing molecular investigations through physiological characterisation

The observation that morphological defects detected with protein stains (silver nitrate) were also evident following examination with DAPI, shows that morphological differences are not limited to the proteinaceous surrounds of the sperm head but extend to the nuclear material. The existence of distinctions in various aspects of shape and size between WTX and WTY which match in some aspects of character but not in scale, the distinctions seen between WT and YQ and between YQX and YQY might suggest that one or more sex-linked gene products involved in DNA and/or sperm head shaping and condensation showing restricted, but not a lack of sharing between developing spermatids could control aspects of sperm head nuclear shape in mice. These effects would then be amplified in the YQ where the expression of sex-chromatin is increased. This may be compensated for by sex-linked genes from the opposite chromosome (potentially genes involved in intragenomic conflict) that

under normal circumstances act to achieve balanced functional outcomes for the X- and Y-bearing populations. The major morphology defects appear present across the entire YQ sperm population, suggesting it may represent a competitive disadvantage when compared to normal morphology spermatozoa, but is not biased to one of either X- or Y-bearing sperm within the teratozoospermic population.

In humans, defects in sperm chromatin condensation and acrosomal hypoplasia (underdevelopment of the acrosome) are the two most common abnormalities in teratozoospermia (review by Chemes and Sedo 2012). A major advancement in fertility treatment has been the development of intracytoplasmic sperm injection (ICSI); the injection of sperm directly into the oocyte. This has meant that individuals with sperm motility defects have been able to seek effective treatment. A condition that is often not overcome by ICSI is teratozoospermia. Studies in humans have associated teratozoospermia with failure post-ICSI fertilisation, during embryo development (reviewed by Chemes and Sedo 2012). Given that our results (section 3.2.2) suggest a reduced success rate of natural matings in YQ males, and the results of Ward and Burgoyne (2006) show a reduced success of natural matings and IVF but not of ICSI, it may be that major defects of YQ Y-bearing sperm are motility or capacitation (as capacitation is required for hypermotility of sperm) related. Teratozoospermia is evident but does not appear to significantly affect the success of ICSI fertilisation.

The potential effects of parameters not measured

Two important sperm fitness measures were not made during the course of study; capacitation rate and motility. Capacitation rates and sperm motility could both have a role to play in the reduced competitiveness of YQY sperm. Indeed, some reduced capacitation has been noted in B10.BRYdel mice (Xian et al. 1992), though a sex-specific exploration has not been conducted.

Chapter 4

4. RNA Sequencing Study of Gene Expression in Spermatids

The data in this chapter was retrieved following the RNA sequencing of enriched pools of spermatids from selected developmental stages, and their subcellular fractions retrieved from MF1XY^{RIII} and MF1XY^{RIII}qdel mice. The MFXY^{RIII}qdel model displays a skewed offspring sex ratio towards female progeny of 6% (section 3.2.2). It has been shown that the offspring sex ratio skew does not come about as a result of differences in the number of X- and Y-bearing sperm, or by preferential destruction of male embryos, but as a result of functional differences in the X- and Y-bearing sperm pre-fertilisation. Genes identified as engaged in an intragenomic conflict behind the skew act specifically at the spermatid stage. Due to the syncytial nature of spermatid development, gene products responsible for the functional inequality must not be shared across connecting cytoplasmic bridges. The aim of this investigation was to conduct a broad genomic screen of differentially expressed genes in order to identify gene and gene product candidates for facilitating or causing sperm phenotypic and reproductive abnormalities in the MF1XY^{RIII}qdel mouse line. By enriching populations of early and late spermatids from testes extract collected from pooled populations of 4 x MF1XY^{RIII} and 4 x MF1XY^{RIII}qdel strains, it has been possible to examine RNA expression changes across stages of spermiogenesis in these two models. Subcellular fractionation has facilitated the identification of potentially non-shared candidates through examination of transcript distribution across cell regions and comparisons with known shared and non-shared gene products. From this data it was then possible to construct lists of candidate genes for further investigation. The RNA sequencing experiment detailed below has facilitated an investigative sweep of the genome to identify genes most evidently disrupted in the deletion model as compared to the wild type.

4.1. Introduction

The global up-regulation of sex chromatin in MSYqdel spermatids is known to result in a cascade of transcriptional differences between the wild-type and MSYqdel cells (Cocquet et al. 2012, Ellis et al, 2005), causing reproductive and sperm phenotypic abnormalities seen in MSYqdel males. It has been shown in previous work that sperm head morphological abnormalities occur in both a global (across both X- and Y-bearing sperm) and Y-bearing specific manner (section 3.4.5). The existence of a consistent sex ratio skew towards females (6% in investigations reported in this thesis) is facilitated by a difference in functional capacity that is not yet fully defined. In order to facilitate functional

inequivalence between sperm developing as a syncytium, one or more gene products responsible for their different functional outcomes must remain unshared between developing sister spermatids developing in syncytium. Given what is understood about shared and non-shared (cis-limited) transcripts, it is anticipated that other non-shared transcripts could be initially identified by two complementary means: separation of subcellular RNAs into free cytosolic RNA fractions (un-bound), and RNAs bound to sperm structures such as the cytoskeleton and nuclear membrane (bound); and comparison of transcript behaviour with that of known cis-limited genes. With results from previous physiological assays borne in mind, Short-listed candidates can be investigated for links to functional outcomes in the literature.

4.1.1. Investigating differential expression during spermiogenesis

Sperm are transcriptionally inactive (with a few exceptions) yet developing spermatids are highly active as they transition into mature functional gametes. During spermiogenesis, the transcriptional profile of spermatids and other testicular cells changes as cells progress through successive stages (Zimmermann et al. 2015). The heterogeneity of testis tissues presents a challenge for the assessment of molecular events during spermatogenesis. Testes are a complex tissue, containing a wide variety of cell types, not limited to post-meiotic cells. Using testicular tissue without further purification would not allow for sufficient definition for the examination of transcriptional difference between two mouse strains at predefined maturation stages. Therefore, enriched or purified cell populations of differing cell stages are advantageous for these lines of investigation. By enriching cell populations for spermatid maturational stages, we gain greater resolution of the transcriptional differences between WT and YQ spermatids, and the ability to attribute variation to different stages of spermiogenesis. There are a number of methods cited for cell type enrichment; most prominent amongst these are cell sorting by flow cytometry and centrifugal elutriation of heterogeneous cellular populations.

Flow cytometry has the advantage of being able to produce highly purified cellular populations (Basu et al. 2010, Morsy et al. 2005). However, it does lead to cells being exposed to increased temperatures as they travel through the flow cytometer and requires the use of intercalating dyes, predominantly Hoechst 33342, and detection by exposure to UV. It is also relatively costly to undertake and achieving consistency across sample sets without proper guidance is difficult. Flow cytometry cell sorting of testis cells is sometimes optimised in combination with other cell sorting techniques such as elutriation or centrifugation on a discontinuous percoll gradient (Lassalle et al. 1999).

Centrifugal elutriation has been used widely and successfully in the separation of testicular cells, including mouse testis cells (examples include: Chang et al. 2011, Aravin et al. 2006, Grabske et al.

1975), and in the enrichment of a wide variety of other items, from fluorescent polystyrene microbeads and mammalian blood cells (Morijiri et al. 2013) to saltwater phytoplankton species (Pomponi and Cucci 1989). The technique utilises differences in sedimentation velocity (a descriptor of size, density and shape) between desired populations for separation. The advantages of centrifugal elutriation are the large quantity of cells that can be sorted in a short time, a sustained temperature and the ability to process cells in physiological buffers, preventing any major physiological alterations. It is cost-effective, time-saving and, like flow-cytometry, allows for the preservation of RNAs for further analysis. It is also simple to replicate the exact processes and experimental parameters when processing cell populations from different treatment groups or genetic strains over a number of hours or days.

4.1.2. Investigation of transcript abundance in differing subcellular regions of spermatids in mice

Developing germ cells are defined as either X- or Y-bearing post-meiosis at the transition to haploid spermatids, when cells enter spermiogenesis. It is from this stage onwards that functional differences between X- and Y-bearing sperm may arise. Post-meiotic expression can lead to phenotypic differences in gametes, and result in preferential transmission of one set of alleles over another (See section 1.4 for examples of transmission ratio distortion). However, the identification of potential candidate genes is complicated by the nature of sperm development. In order to facilitate functional inequivalence between sperm developing in syncytia, candidate gene products must exhibit restricted or non-sharing across cytoplasmic bridges. *Sly* and *Slx* act during the spermatid stages of spermatogenesis (Cocquet et al. 2010), when partial release from PSCR facilitates the expression of sex-linked genes. SLX proteins that are normally out-competed by SLY proteins, block the recruitment of repressive epigenetic marks to the sex chromatin, meaning the sex chromosomes are relatively up-regulated in Yqdel mice compared to wild-type mice with the full complement of *Sly* and other Yq genes (Cocquet et al. 2012).

All haploid cells contain one of the equal copies of all autosomes and either one of the distinct X or Y chromosome, meaning that the dysregulation of sex chromatin could impose different effects on the normally functionally similar X- and Y-bearing sperm, if their gene products were not shared. Given the evolutionary implications of the intragenomic conflict between *Sly* and *Slx* (section 1.4.4), it is likely that non-shared transcripts facilitating the functional inequivalence are X-linked. Over-expression of sex chromatin facilitated by X-linked *Slx* favours transmission of the female sex chromosome, and we know that an array of important developmental and spermatid specific genes exist on the X chromosome. It could be the case however that a partner in the mechanism is X-linked

and has a downstream effect upon an autosomal gene that also encodes non-shared gene products by way of a distorter / effector complex similar to that responsible for T-complex transmission ratio distortion (see *Figure 4.1*).

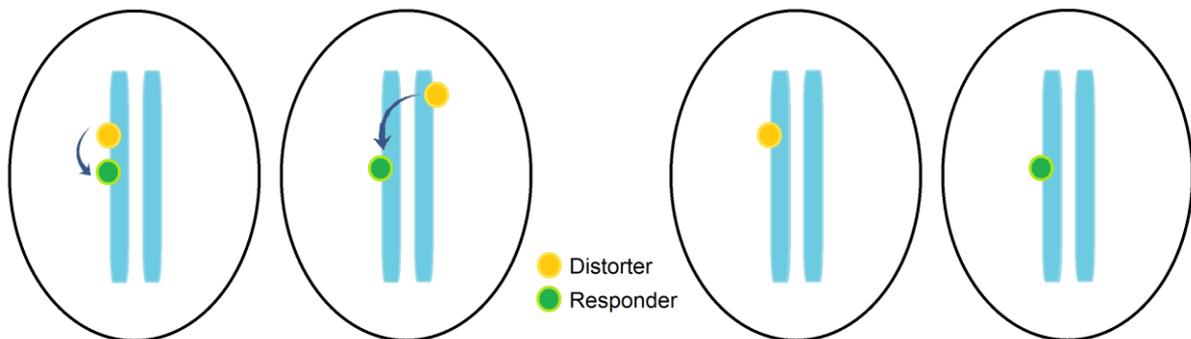


Figure 4.1. Demonstration of potential distorter / responder product interaction. Ellipses represent cells containing chromosomes. If an individual cell contains both the distorter and responder mechanism, an effect may result whether these are located on the same chromosome or not. If a cell does not contain either one of the distorter or responder within a complex, no effect can take place.

Gene products encoded by distorter genes must either be shared unevenly (or not shared) between sister spermatids or interact with other downstream factors that are themselves not shared (Ellis et al. 2011b). *Figure 4.2* demonstrates a number of possible scenarios describing what form a distorter / effector mechanism with at least one X-linked partner might take to result in functional inequality between X- and Y-bearing sperm.

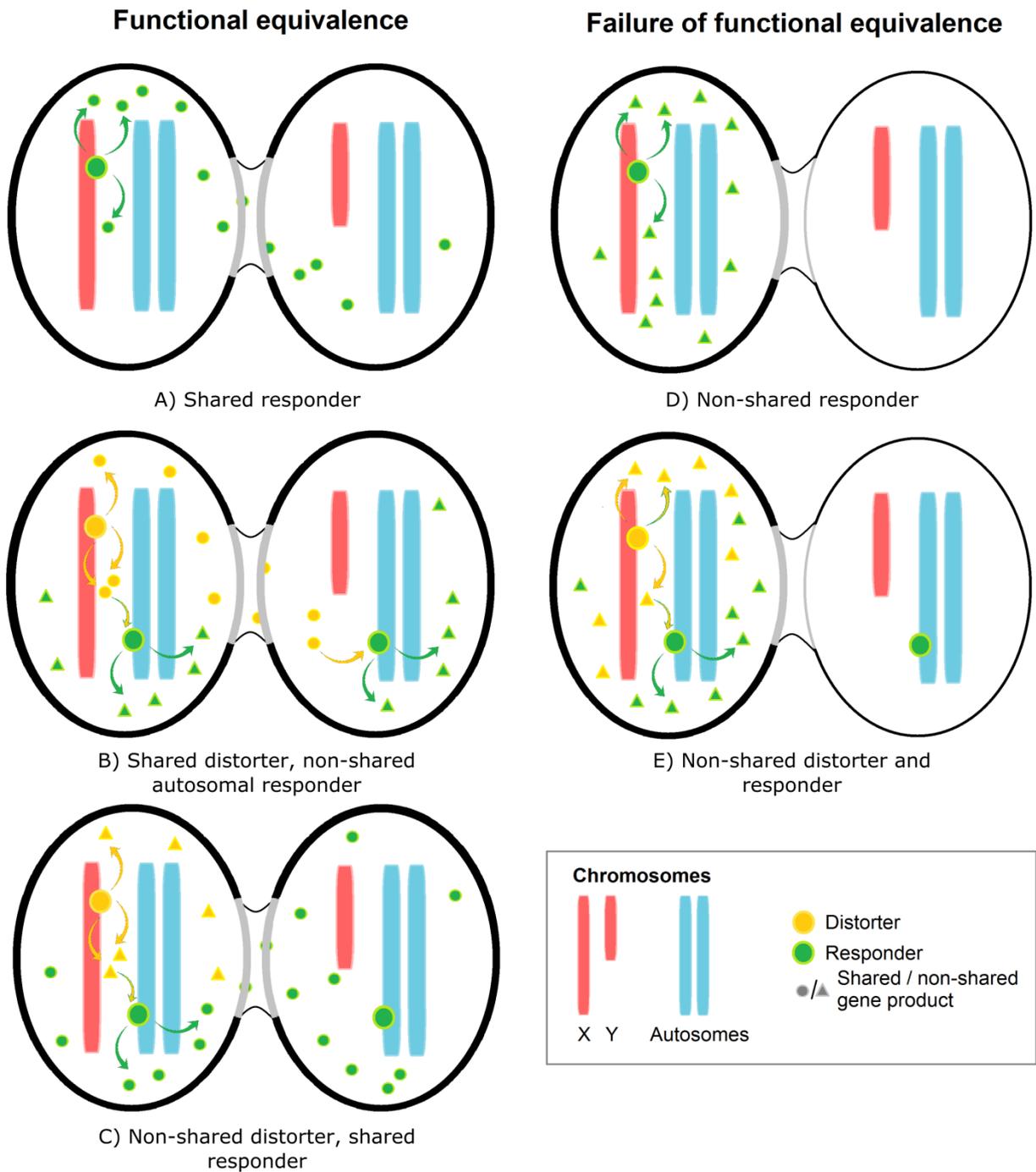


Figure 4.2. Demonstration of potential distorter / responder mechanisms implicated in offspring sex ratio distortion in MSYqdel mice. In this figure, ellipses represent developing sister spermatids connected by cytoplasmic bridges. Cells with functional equivalence are represented by the same thickness of cell wall.

4.1.3. Messenger RNA transcript localisation and sharing

The sharing of RNA transcripts has been linked to RNA distribution, and it is suggested that transcripts bound to cellular organelles, or tied to the actin cytoskeleton are less readily shared across cytoplasmic bridges than transcripts residing in the cytosol (review by Ventelä 2006, Morales et al. 2004, Ventelä et al. 2001). In addition, a number of genes exist for which sharing status of transcripts between developing spermatids is known (section 1.3.2). The distribution of these transcripts can be estimated through separate examination of the transcriptome of subcellular regions. The Hesketh fractionation series developed by JE Hesketh and colleagues and reported in Hesketh et al. 1991 allows for the separation of various cell types into the ‘bound’ and ‘unbound’ fractions (see table below). Hesketh et al. (1991) originally used the series to investigate cytoskeletal binding of polysomes in fibroblasts. Since then, several studies have utilised the fractionation series for the assessment of RNA transcript, polysome or protein distribution within a range cells (examples include: Searles et al. 2004, Vedeler and Hollås 2000). Fractionation of testicular RNAs by Hesketh lysis series was also used by Martin-DeLeon et al. (2005) to investigate the localisation of *Spam1* mRNAs in testis. It showed that the removal of free cytosolic RNAs could be achieved through fractionation to leave cytoskeletally- and organelle-bound mRNAs, and that *Spam1* appeared enriched in cytoskeletally-bound fractions.

Fraction name	Subcellular fraction recovered
Fraction 1 (F1)	Free cytosolic RNAs, soluble cytosolic RNPs, and cytosolic polysome-associated mRNAs
Fraction 2 (F2)	Cytoskeletally-associated polysome-bound mRNAs
Fraction 3 (F3)	Membrane-associated polysome-bound mRNAs and nuclear RNAs

Table 4.1. Approximation of subcellular fractions retrieved following application of the Hysketh lysis buffer series (Hesketh et al. 1991).

Messenger RNAs originate in the nucleus, where they are transcribed from DNA. These transcripts are then spliced to remove introns (with a single protein-coding gene often producing a variety of splice variants) and gain exon-junction protein complexes alongside other components which influence the translation and localisation of the now mature coding mRNA. Transcripts then exit the nucleus and form complexes with protein co-factors in the form of ribonucleoprotein (RNP) complexes (Parton et al. 2014). The localisation of mRNAs to specific subcellular regions provides a

mechanism for the regulation of gene expression via controlled release and translation of mRNAs. The localisation of mRNAs involved in the development of highly specialised cellular regions has been known for some time, for which numerous examples exist. For example, in *Drosophila*, the specific localisation of several mRNAs in the oocyte including *nanos*, *bicoid* and *oskar*, helps establish subsequent proper spatial organisation of the developing embryo (Martin and Ephrussi 2009, Johnstone and Lasko 2001). During brain development, neuronal axons are able to navigate more effectively to their synaptic partners through environmental cues facilitated by the local translation of specific mRNAs in axonal growth cones (Lin and Holt 2007). In *Drosophila* spermatids, the localised translation of aPKC at the apical end of elongating spermatids is essential for the establishment and maintenance of polarity of the cell, and subsequent asymmetric cellular development (Barr et al. 2016). Localisation is now understood to be an important factor in translational control, and in the production of asymmetric cells with highly specialised cellular features. Thorough reviews on the topic are provided by Parton et al. (2014) and Martin and Ephrussi (2009).

4.1.4. Application of RNA sequencing

RNA sequencing is a high-throughput sequencing technique that allows for the examination of multiple transcript types (see section 1.6). By examining transcriptional differences between the MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ) strains at different stages of spermiogenesis and across different subcellular fractions, the aim is to discover differences in overall expression and distribution of transcripts that may result in physiological differences between WT and YQ sperm, and to identify potentially cis-limited gene products as gene candidates for involvement in facilitating the known offspring sex ratio skew. Physiological changes may also result if transcripts are of similar abundance in the strains, but are altered in their subcellular location (compartmentalisation), potentially affecting translation. A matched comparison of subcellular fractions was conducted between strains. We now know that the distinct MF1XY^{RIII}qdel sperm head morphology is observed in both X- and Y-bearing sperm with an exacerbated phenotype in Y-bearing sperm (section 3.4). As a result, the implicated gene products (transcripts and proteins) for this phenotype are not expected to be entirely restricted to a subset of the sperm population. We therefore expect some transcriptional differences to be detected broadly between the two strains in genes not showing potential cis-limitation in one or more developmental stages.

RNA sequencing permits a flexible pipeline of analysis to be constructed, more appropriate for atypical datasets. EdgeR and associated packages provide suggested normalisation techniques most appropriate for single-replicate data. Combined with the cautious use of log fold change values, restricted to include only the most distinctly different and consistently differentially expressed genes

between the two strains, this allows for the construction of a candidate gene list for further investigation. As p-values are not produced for single replicate data, alternative methods of selection of gene candidates were conducted. Current knowledge of the effects of Yq deletions in mice, allows for anticipated changes to be measured alongside technical and biological controls. Control transcripts of known cell type origin were used to examine cellular contamination of enriched populations.

It is anticipated that an upregulation of X and Y chromatin in YQ spermatids (excepting the Yq genes contained within the deleted region), will be observable between direct WT and YQ comparisons in the data. The analysis was expected to yield a large number of differentially expressed candidates drawn from across the genome.

The resulting candidate gene lists contain both X-linked genes of interest and highly differentially expressed autosomal genes showing potential non-sharing behaviour expression patterns consistent with that of known non-shared genes.

4.1.5. Chapter Objectives

The aims of this chapter are as follows:

- O1. Identify sex-linked and autosomal candidate genes for potential involvement in the perceived offspring sex ratio skew in the MF1XY^{RIII}qdel strain.
 - To investigate hypothesised X-linked mediators of offspring sex ratio skew
 - To examine differentially expressed functional autosomal genes that may be involved in enacting functional asymmetry between X- and Y-bearing sperm in MF1XY^{RIII}qdel.
- O2. Identify candidate genes for potential involvement in previously identified broad morphological changes (Chapter 3) not limited to one of X- or Y-linked sperm.
- O3. Generate a list of genes suitable for further examination as candidates for producing non-shared (cis-limited) RNA transcripts testable by subsequent molecular analyses.

4.1.6. Notes for Chapter

Sample names

For the purposes of brevity and simplicity, RNA sample names have been abbreviated. The abbreviations are listed and explained below (*Table 4.2*) and an example given. The combined abbreviation describes the mouse strain from which the sample originated, the cellular population (following elutriation or alternative extraction), and the stage during the cellular fractionation procedure at which the sample was retrieved.

Mouse strain		Cell type		Cell fraction	
Full name	Abb.	Full name	Abb.	Full name	Abb.
MF1XY ^{RIII}	WT	Early spermatids	E	Whole cell	WC
MF1XY ^{RIII} qdel	YQ	Late spermatids	L	Fraction 1 (cytosolic)	F1
		Mature sperm	M	Fraction 2 (cytoskeletally-bound)	F2
				Fraction 3 (highly bound)	F3

Table 4.2. Explanation of sample name used in RNA sequencing chapter and abbreviations (Abb).

An example of a combined abbreviation: WTEF3 = a sample from MF1XY^{RIII} (wild-type), early spermatid stage, third cellular fraction.

4.2. Sample Preparation

RNA samples and Illumina sequencing libraries were prepared as described in this section. *Figure 4.3* provides a simple overview of sample preparation procedures.

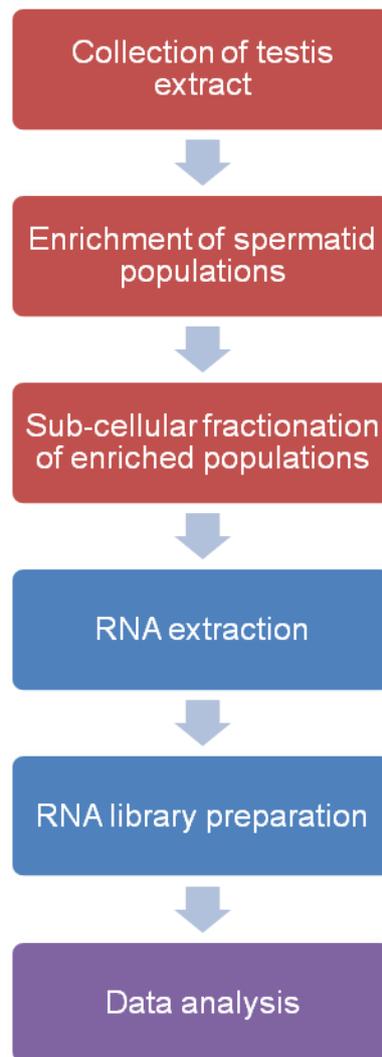


Figure 4.3. Overview of sample preparation procedures conducted for analysis of RNA sequencing data.

4.2.1. Cellular elutriation: examination of retrieved cell composition

Effective preparation of spermatid cells is important for the reduction of unwanted variation in later processes. Reducing contamination of desired populations by confounding cell types, encouraging consistent collection of cellular populations and subcellular fractions across mouse strains, and restricting RNA loss through best practise all contribute to successful examination of RNA composition across multiple conditions.

During sample preparation, testis extract from each of 5 x MF1XY^{R^{III}} (WT) and 5 x MF1XY^{R^{III}}qdel (YQ) mice was pooled by strain and processed by cellular elutriation to enrich for desired cell populations as described in section 2.4 and 4.1.1. Mature sperm were extracted by dissection (section

2.1.1). To determine the level of enrichment of cells achieved in each elutriation run, an amount of cells were collected prior to the subcellular fractionation stage for staining and examination using a light microscope and haemocytometer. Cells were first fixed in 4% PFA, stained with trypan blue and were counted and categorised according to cell morphology. Cell population composition was then estimated, and percentage enrichment of the desired cell type given. During the RNA sequencing sample preparation the following enrichments were observed across both strains:

Early spermatid: 80% round cells

Late spermatid: 85% elongating/condensing cells

Mature sperm: 100% mature sperm

Contamination in the early spermatid populations appeared primarily by later stage spermatids; there were some additional unidentified cell types. Contamination in late spermatid populations appeared composed of condensed (possibly mature) sperm and a minority of other testicular cell types. In subsequent experiments, where samples retrieved were used for microarray analysis, sample composition was examined using DAPI staining of cell populations and cell counting assisted by Cell Counter software (ImageJ) (section 2.4.4).

4.2.2. Further sample processing and subcellular fractionation

Following cellular elutriation, a proportion of cells from each population were placed immediately into QIAzol lysis reagent (Qiagen, UK) and homogenised. The remaining cells were taken for subcellular fractionation as described in section 2.4.2 and 4.1.2. This resulted in the retrieval of 18 RNA samples for sequencing; composed of the following total RNA samples from each set of mice (strain x 2):

EWC Early spermatid whole cell

EF1 Early spermatid fraction one

EF2 Early spermatid fraction two

EF3 Early spermatid fraction three

LWC Late spermatid whole cell

LF1 Late spermatid fraction one

LF2 Late spermatid fraction two

LF3 Late spermatid fraction three

MWC Mature sperm whole cell

4.2.3. RNA sample quality control

RNA sample quality and concentration was assessed prior to library preparation using Nanodrop 1000 (Thermo Scientific, US), Agilent 2100 Bioanalyzer (Agilent Technologies, UK), as recommended in the Qiagen miRNEasy protocol. These methods of sample evaluation rely on the presence of rRNAs, which are highly abundant cellular RNAs, for estimation of quality and concentration of other cellular RNAs. Due to the atypical RNA composition of some samples (for example, mature sperm contain very few rRNAs), some methods of assessing sample concentration are rendered less reliable (see section 2.5.2). Therefore the Qubit Fluorometric Quantitation system (Thermo Fisher Scientific, UK) was used alongside. All samples fell within the required range of minimum and maximum sample input for library preparation using the Illumina Total TruSeqLT kit.

4.2.4. RNA Sequencing

RNA library preparation and quality control was conducted using the Illumina Total TruSeqLT kit as described in sections 2.5 and 2.6.1. The opportunity arose to conduct a test run of total RNA libraries on the Illumina miSeq platform. This showed that two samples were overrepresented in the library pool and one under-represented. Library concentrations were adjusted prior to final sequencing. RNA sequencing was performed on the NextSeq 500 platform by Cambridge Genomics Services (Department of Pathology, University of Cambridge). Three runs of six RNA sample libraries were performed using single index, paired-end sequencing.

4.3. Data Processing and Quality Control

Data was prepared by Cambridge Genomics Services in consultation with the author. All outputs and plots prior to differential expression analysis were generated using a customised RNA sequencing pipeline, with alterations made appropriate for single replicate data, as described in later sections. All outputs were subsequently customised by the author in R (analysis scripts are available in the online repository⁶) or Microsoft Excel 2007 unless otherwise stated. *Figure 4.4* provides an overview of the data preparation and analysis pipeline post-sequencing.

⁶ Online repository: <https://www.github.com/EmmaEPJ/Repository>

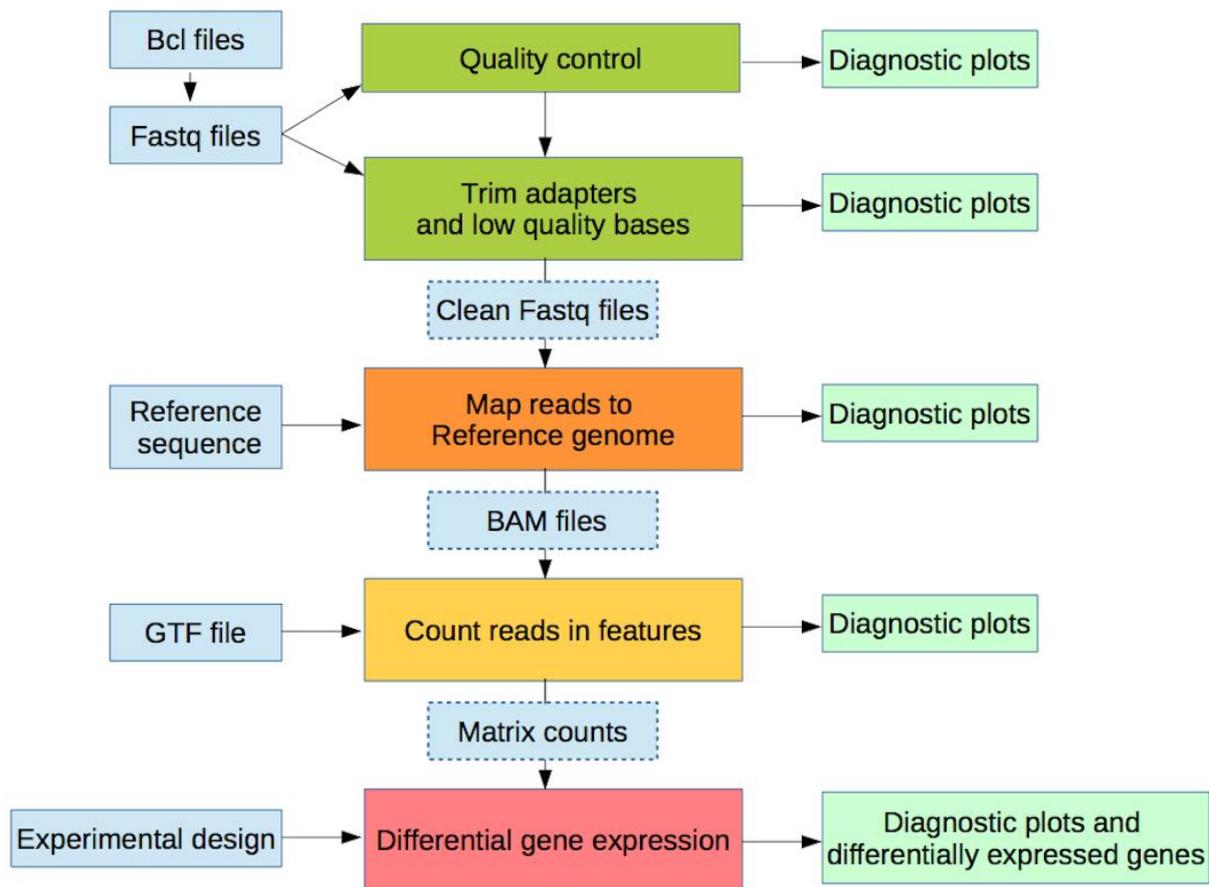


Figure 4.4. Overview of RNA data processing pipeline following RNA sequencing. Image taken from data analysis report produced by Cambridge Genomics Services (Department of Pathology, University of Cambridge).

4.3.1. Examination of technical and quality controls before trimming and mapping

The assessment of sequencing read quality and adjustment of subsequent data processing procedures accordingly is essential to retrieving reliable differential expression data. A number of quality assessment measures were used for the examination of all samples. *Figure 4.5* shows the number of reads retrieved from each sample as a percentage of reads across all samples. The larger the library (number of reads retrieved) the more diversity the data will likely contain, as low abundance reads are more likely to be detected. Samples were sequenced in sets of three (3 x 6) and were normalised for concentration before sequencing. Across all samples, reads retrieved varied from 3.2% to 12.8% of the entire read population. The most distinct matched samples (WTLF1 and YQLF1) show divergence resulting from library size prior to normalisation.

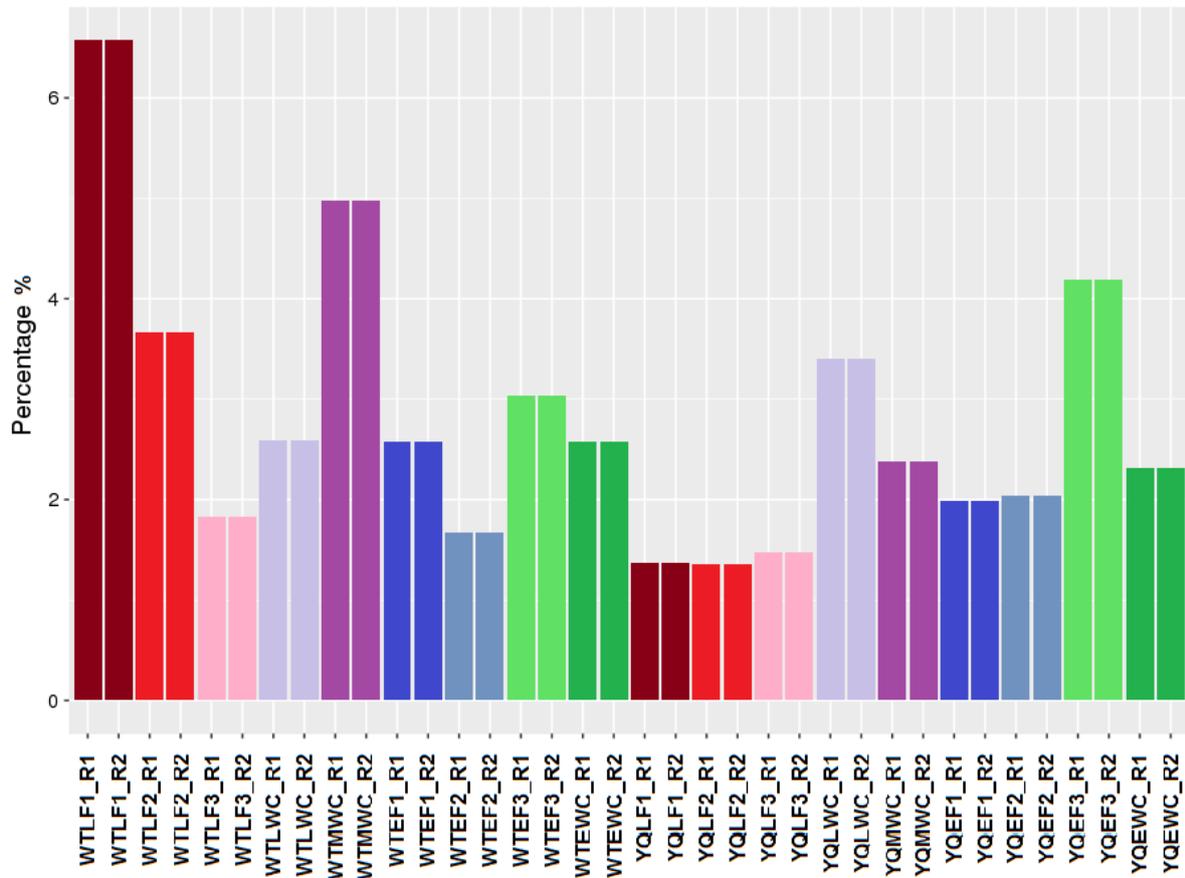


Figure 4.5. Reads retrieved from each sample as a percentage of the total number of reads retrieved across the whole dataset. Each sample is shown as two sets of reads as paired-end sequencing was performed. Colours indicate matched samples across the two mouse strains (WT and YQ).

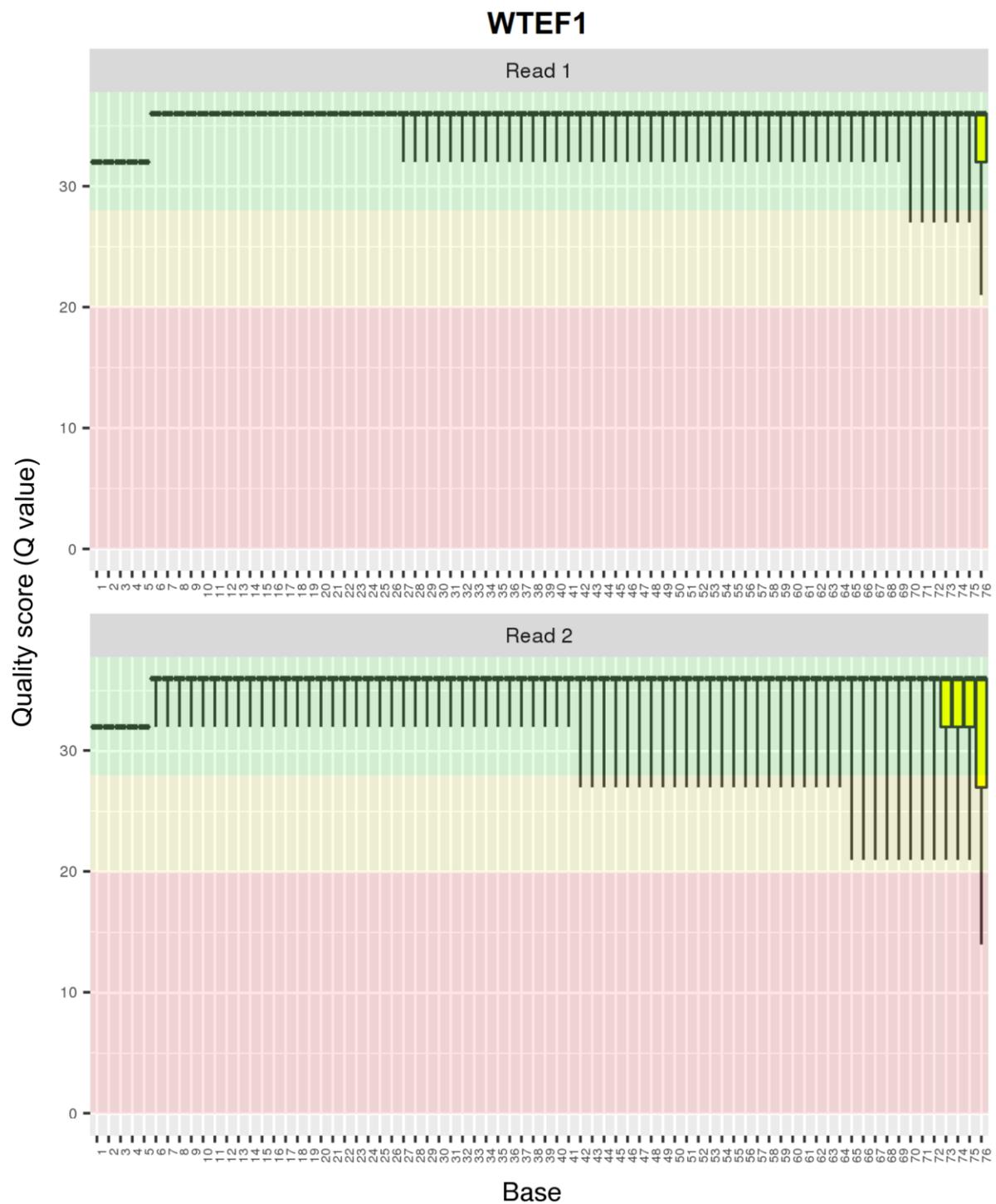


Figure 4.6. Boxplots of quality values distributed across each read of the sample examined. The y-axis denotes quality values representing expected error rate. This example is from RNA sample WTEF1. Images were adapted from those taken from RNA sequencing output as produced by Cambridge Genomics Services.

Figure 4.6 shows the per base quality score per sample. The example plot shown is from sample WTEF1. The y-axis displays a per-base quality score representing the number of errors as a ratio of bases read. A value of 30 represents an estimated error rate of 1 in 1000 bases; a lower score of 20 represents an estimated error rate of 1 in 100 bases. Quality normally drops towards the ends of the reads due to error accumulation along longer reads and the degradation of sequencing reagents due to heat exposure. One sample (WTLF3) showed reduced quality compared to other samples, which resulted in a greater proportion of trimmed reads in subsequent preparation steps. However, even in the lowest quality sample, <Q20 reads only represented a very small proportion of total reads. All other samples were of a high quality. Per sequence quality scores across all samples can be seen in *Figure 4.7*. Reads are expected to have a mean quality of 30 or greater, so a peak (y-axis) is expected to appear at or after a value of 30 (x-axis). All samples produced peak mean quality scores above 30. Read bases with values below Q20 were trimmed in subsequent stages.

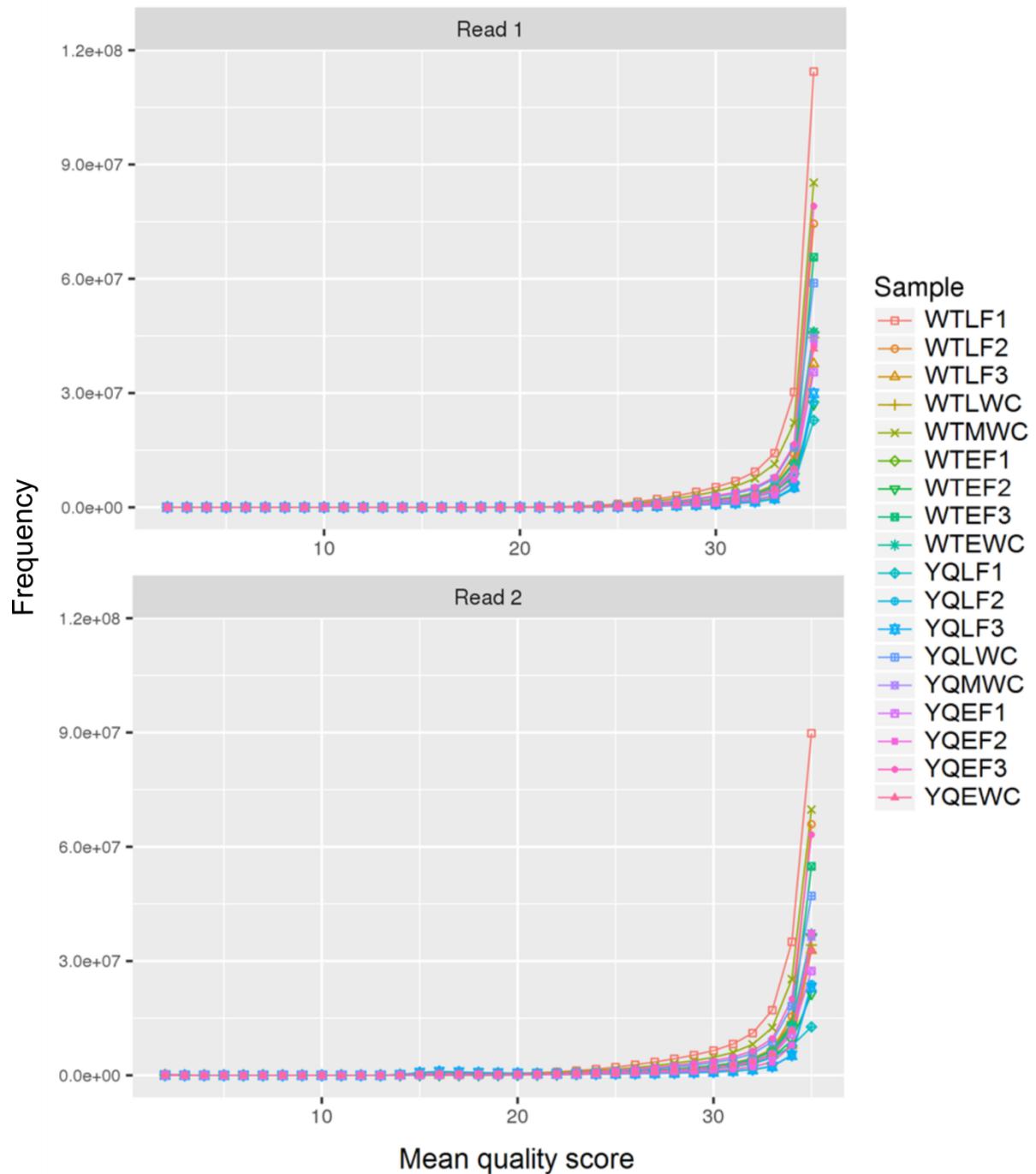


Figure 4.7. Per sequence mean quality scores. Images were adapted from those taken directly from RNA sequencing output as produced by Cambridge Genomics Services.

Other measures of sample read quality, such as sequence length distribution, GC content and sequence duplication levels were also examined. All samples satisfied quality control measures at this stage.

4.3.2. Trimming, mapping and quality control of reads

Trimming of reads

Reads were trimmed to remove low quality base calls and adapter sequences. Adapters were hybridised to each transcript in the sample to allow for collection and subsequent reading of sequences detected; however they do not contribute to information about the read and so need to be removed before further downstream analysis. Reads were trimmed using TrimGalore v.0.4.1 (Kreuger, Babraham Bioinformatics⁷). This package also removes low quality base calls from the 3' end of the reads and subsequently removes reads that are less than 20 bases long. The percentage of complete reads removed at this stage was less than 1.8% per sample. Only one sample, WTLF3, exceeded this (5.35% removed). This is likely due to reduced overall quality as observed in initial read quality control. This difference should not affect subsequent analysis, as low quality reads are trimmed from the data. Increased trimming of individual bases can also result from failure of full adapter trimming during .bcl to .fasta file conversion (*Figure 4.4*), which would mean that inappropriately retained adapter sequences would be trimmed at this later trimming stage instead.

Mapping of reads to the reference genome

Mapping of reads to a reference genome allows for identification of transcripts and the calculation of expression per gene; both essential for differential expression analysis. It also facilitates the annotation of genes with gene symbols, biotypes and other useful information for further analysis. Reads were mapped to the Ensembl *Mus musculus* GRCm38 fasta format file of the primary mouse genome assembly⁸ (release 84) using STAR v.2.5.2a (Dobin et al. 2013). Annotated transcripts were called from the equivalent Ensembl .gtf file. Mapping of reads can be used to assess read distribution across features. Picard Tools v.2.1.1 (Picard Tools) was used to determine the percentage of reads that mapped to different genomic regions per sample (*Figure 4.8*).

Filtering of rRNAs from the dataset

During physical RNA sample preparation, the Ribo-Zero rRNA Removal Kit Human/Mouse/Rat (Illumina, USA) was used to remove rRNAs from the samples. This works by exposing the RNA sample to antisense rRNA transcripts that are conjugated to magnetic beads, which can then be selectively removed from the sample. Ribosomal RNAs are by far the most abundant type of RNA present in cells but provide little useful information in expression analysis (unless they are the intended target). For this reason, if they are not removed, the useful information gained from

⁷ Trim Galore! - <https://github.com/FelixKreuger/TrimGalore>

⁸ Mouse assembly and gene annotation; Ensembl - http://www.ensembl.org/Mus_musculus/Info/Index

sequencing is severely limited due to rRNA abundance (>80% of reads) (Wilhelm and Landry 2009, Lodish et al. 2000). Removal allows for the investigator to focus more ably on the remaining diversity of the transcriptome. Although the use of rRNA removal kits is common practise in RNA sequencing sample preparation, they are not guaranteed to be 100% effective. Relatively high levels of rRNA (20-32% of all reads) were detected in five samples (*Figure 4.8*). This suggests that the rRNA removal step had been completed, but was not as successful as in other samples. These samples contained two between-strain matched sample sets (WTLF2 and YQLF2, WTLF3 and YQLF3) and one other sample (YQEF2). RNA extraction for these samples was performed simultaneously, so it is likely that human error or reagent failure is the cause of this batch effect. In order to retrieve a greater proportion of reads-of-interest from the samples, rRNA reads were filtered during data processing.

The rRNA filtering was performed on sample .bam files, before read counting. Under standard conditions, a .bed file is used to generate information on individual read location. In order to filter rRNA reads from the analysis, a separate .bed file was created that contained information pertaining to rRNAs only. This allowed for the identification of rRNA in the data and was used alongside the function 'intersect bed' (Intersect - bedtools) (Quinlan and Hall 2010) to reproduce the original data set without rRNAs. rRNA filtering was performed across the entire dataset. *Figure 4.8* shows the genomic context to per sample read distribution before and after rRNA filtering.

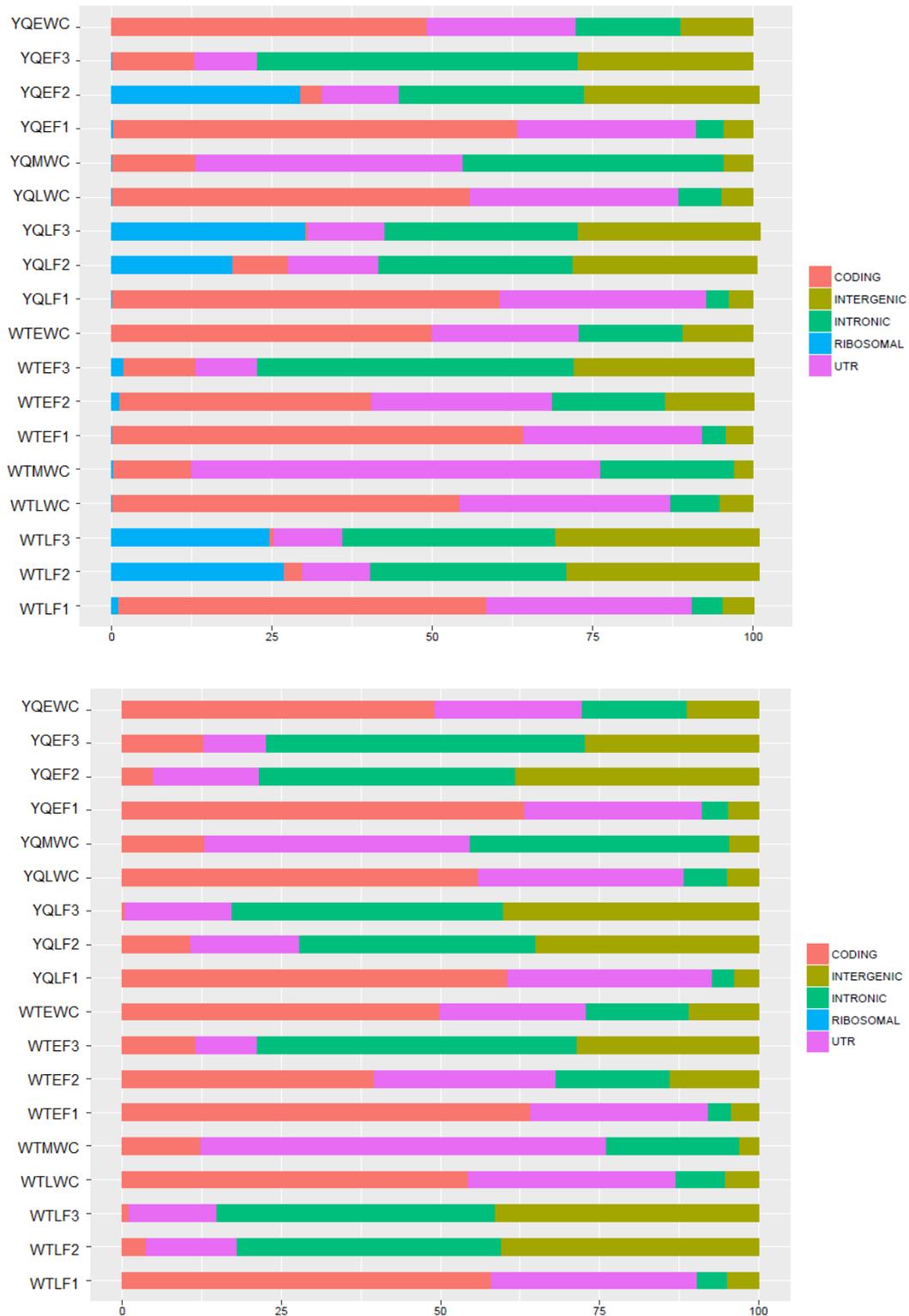


Figure 4.8. Per sample read distribution by read categories before (top) and after (bottom) rRNA filtering. Categories shown include: coding transcripts, non-coding, intronic, intergenic, ribosomal, and UTR. Images were adapted from those taken directly from RNA sequencing output as produced by Cambridge Genomics Services. Some sample distributions total more than 100%, due to the multiple allocation of a small minority of reads.

4.3.3. Single replicate data and generation of differential expression log fold change values

Normalisation by library size and differential expression analysis

Although examination of differential expression is ideally conducted with multiple biological replicates, there are a number of mechanisms for retrieving biological information from single replicate data. The examination of unique, rare, or fragile samples often requires that single replicate data is used and as a result techniques have been developed for appropriate analysis of such datasets. Several data analysis packages suggest that they are equipped to cope with single replicate data (Tarazona et al. 2011, Robinson et al. 2010, Anders and Huber 2010). RNA sequencing is still an expensive technology and where a great range of conditions need to be examined single replicate data can provide an enlightening base of information, provided it is handled correctly. The R package Empirical Analysis of Digital Gene Expression Data in R (EdgeR) is a widely used tool for the examination of differential expression in high-throughput sequencing datasets (Robinson et al. 2010). Although primarily intended for use with multiple biological replicates, it provides suggestions for the examination of single replicate data, including exploratory analysis and subsequent use of log fold change values.

Normalisation by library size and differential expression analyses were performed per comparison made e.g. WTEF1 vs YQEF1, to produce an adjusted 'log counts per million' (logCPM) value for each gene per sample within that particular comparison, from which a log fold change (logFC) value was then generated for easier examination of differential expression.

Library size = the number of counted reads per sample

Adjusted library size = library size (x) / 1,000,000

Counts per million = raw count for gene / adjusted library size (x)

At this stage, the counts per million (cpm) value represents a proportion of the adjusted library size. To calculate the log value of expression for each gene, a *scaled* prior count was applied, based on the library size.

Scaled prior count = (library size / mean library size across comparison samples) * prior count

The prior count value was used to proportionally shrink the resulting logCPM value for low count genes and was applied before the log transformation. Larger values for prior count shrink these lower count values more. High variability should not be ruled out where replicates are not available. A prior

count value of 2 (prior to scaling) was used in this analysis. The scaled prior count allows for differences in library size between the two samples to be taken into account. Finally, logFC was calculated by subtracting the logCPM value of sample A from the logCPM of sample B.

4.3.4. Overall trends in RNA sequencing data

Basic data visualisation is useful for understanding many aspects of RNA sequencing data. Visualisation can be used to easily reveal patterns, trends and potential problems in the data. Basic post-mapping data visualisation applied to the RNA sequencing dataset included principal components analysis, read biotype allocation and sample similarity visualisation by heatmap.

Principal components analysis (PCA)

Principal components analysis (PCA) is a statistical method used for exploring the sources of variation in a complex, multifaceted data set, by reducing the dimensionality of the data to more manageable components (compounded variables). This technique is especially useful in the examination of differential expression data. Given that both microarray and high-throughput sequencing experiment outputs have variables (gene expression levels) numbering well into the thousands and tens of thousands, the ability to reduce the dimensionality of the data and, as a result, the number of variables to be plotted, is extremely useful when broadly assessing differences between experimental samples and identifying sources of variance. Each principal component (PC1, PC2 etc.) can be described as being a ‘linear combination of the original variables’ (Raychaudhuri et al. 2000), accounting for as much variance as the original variables, whilst remaining uncorrelated to them. Sequential components account for decreasing proportions of overall original variation.

A convenient way to visualise principal components data is through the generation of PCA plots, where samples are placed at coordinates determined by two or more of the major principal components. The technique is useful for the visualisation of overall differences between multiple samples in a dataset.

Results of PCA shown in *Figure 4.9* reveal the greatest source of variance in the dataset to be cell population. Samples are broadly separated across the X-axis (PC1) according to progression through spermiogenesis. Whole cell (WC) samples can be seen to lie between fraction 1 (F1) and fraction 3 (F3) within each of the early spermatid and late spermatid cell populations. Mature sperm samples are distinctly different from both early and late spermatid cell population samples. Subcellular fractions represent the secondary source of variation, whilst originating mouse strain produces less variation.

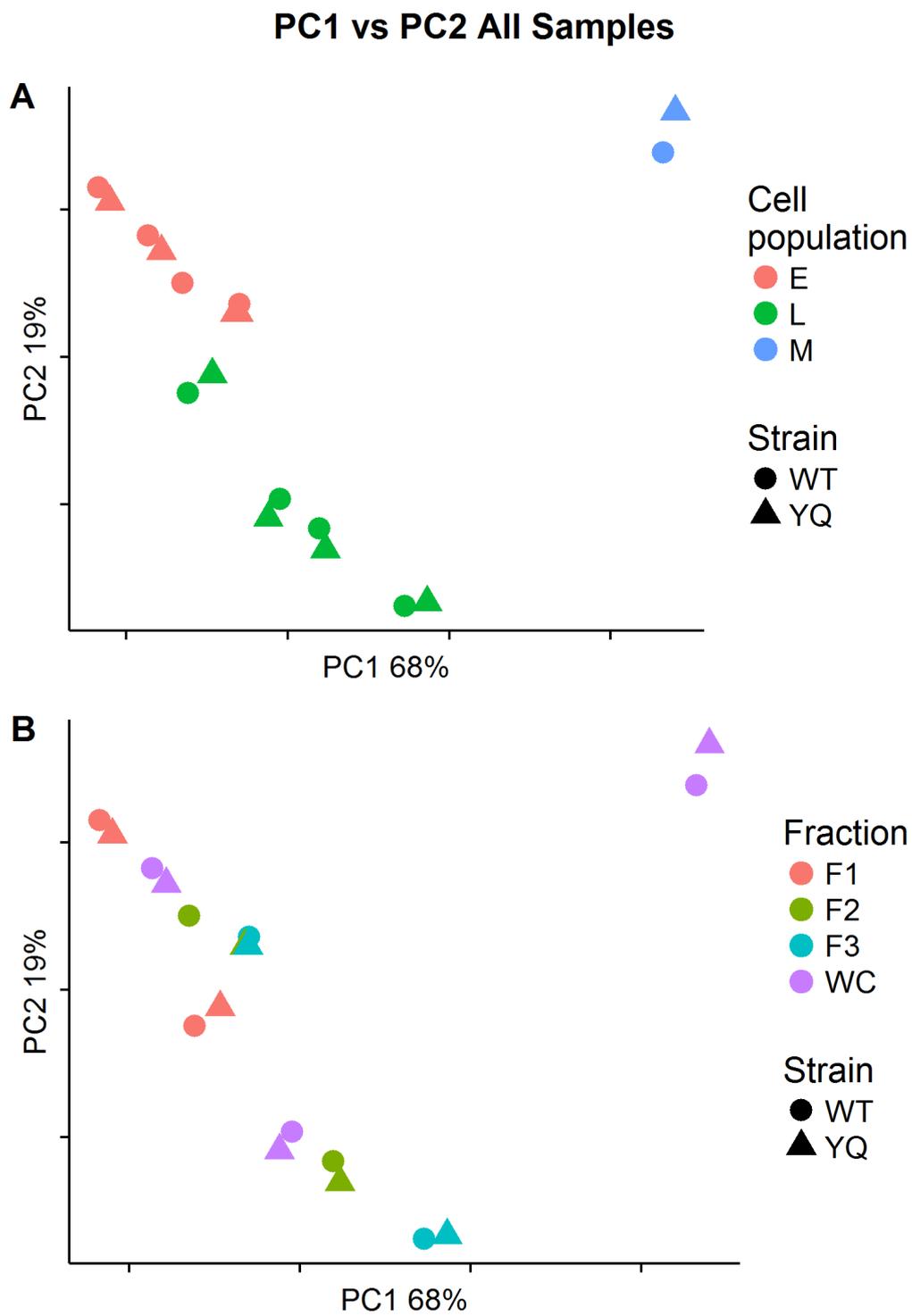


Figure 4.9. PCA plot of all RNA libraries post-data processing. This plot uses first and second principle components, which represent 87% of variance in total.

Sample YQEF2 shows some departure from all other early spermatid samples, towards the early spermatid fraction 3 and late spermatid populations (just discernible though partially hidden behind YQEF3 in the bottom plot of *Figure 4.9*). This may indicate sample contamination. Further examination of this sample reveals a transcript diversity profile highly similar to its wild-type counterpart, but a distinctly low ‘strandedness’ value compared to other samples from the same cell type. The strandedness value indicates the proportion of reads that originate from the expected strand of DNA for each detected transcript. This is made possible through the use of anchored oligo(dT) primers used in the production of the primary strands of cDNA (see Mills et al. 2013 for a review). In our data, this finding complies with other samples from the same RNA library preparation batch and results in an overall lower read count. Whilst this should not (and appears not to) affect overall RNA composition, it may affect the reliability of the comparisons with other samples with higher read counts. Normalisation by library size, as explained in section 4.3.3, should go some way in alleviating this. Further examination of the sample was conducted, and expression analysis was adapted to reflect these concerns. The cause of above symptoms has been identified as DNA contaminations and is discussed in the Supplementary Material.

Sample Similarity – Visualisation Using a Heatmap

Using Pearson correlation of log normalised counts to calculate sample similarity and perform subsequent hierarchical clustering, the relationship between each RNA library can be seen in the below heatmap (*Figure 4.10*). Samples cluster primarily by cell type, except in the case of late spermatid fraction three samples (WTLF3 and YQLF3); these cluster more closely to the subsequent mature sperm samples (WTMWC and YQMWC). Fraction three is expected to be primarily composed of nuclear-associated and membrane-bound RNAs. Late spermatid populations represent the final maturation stage before completion of spermiogenesis (as mature sperm). Mature sperm have shed their remaining cytoplasm and are composed primarily of nuclear material and specialised sperm-specific cell components. Given this, the observed similarity between mature sperm samples and late spermatid nuclear fractions is not unexpected. There is also a notable similarity between the first fractions across cell types. Sample YQEF2 (previously highlighted in PCA analysis) appears highly similar to the WT counterpart in this examination.

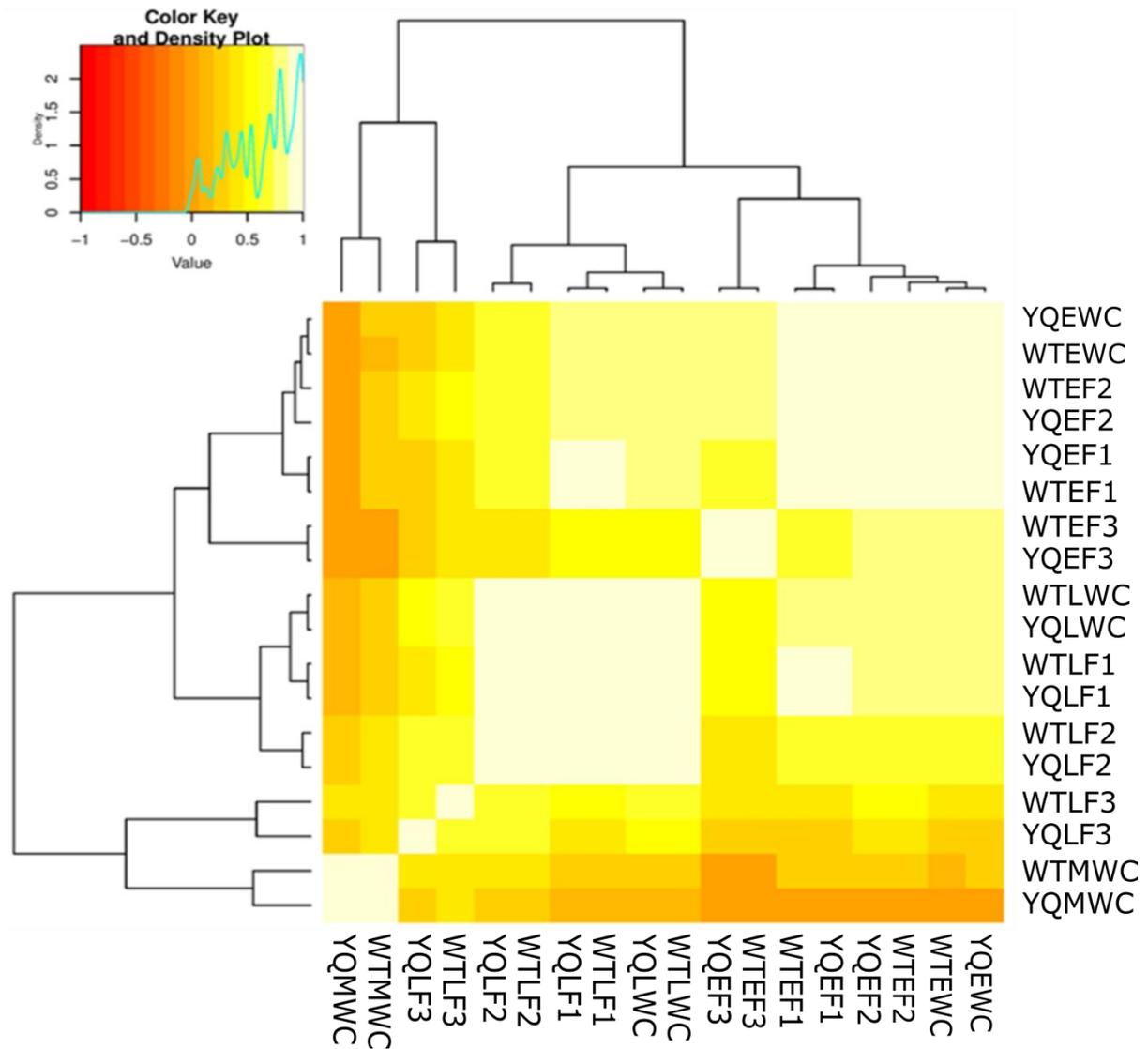


Figure 4.10. Heatmap comparing all RNA samples for similarity of RNA composition.

4.4. Results of Differential Expression Analysis of RNA Sequencing

Data

4.4.1. Increased expression of sex chromatin in MF1XY^{RIII}qdel

Figure 4.11 and Figure 4.12 show a broad upregulation of sex-linked genes (except those contained within the deleted region of the MSYq) observed in MF1XY^{RIII}qdel (YQ) spermatids when compared to counterpart samples from MF1XY^{RIII} (WT). This was noted across all subcellular fractions and cell types. The proportion of sex-linked genes showing increased expression reduced as spermiogenesis progressed. Autosomal genes remained broadly stable, with some exceptions.

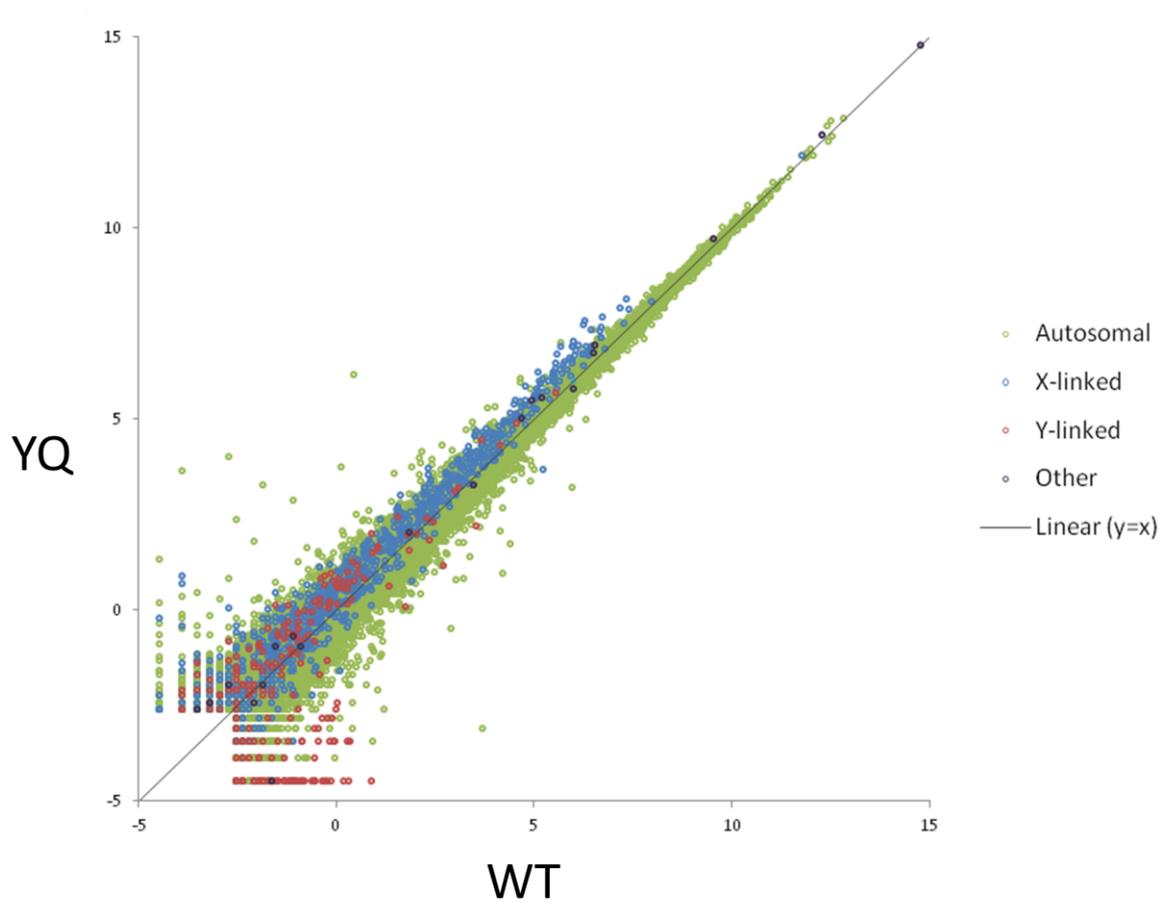


Figure 4.11. Comparison of overall expression of sex-linked and autosomal genes in early stage spermatid whole cell samples between WT and YQ models. Data points represent LogCPM values of individual genes in the comparison.

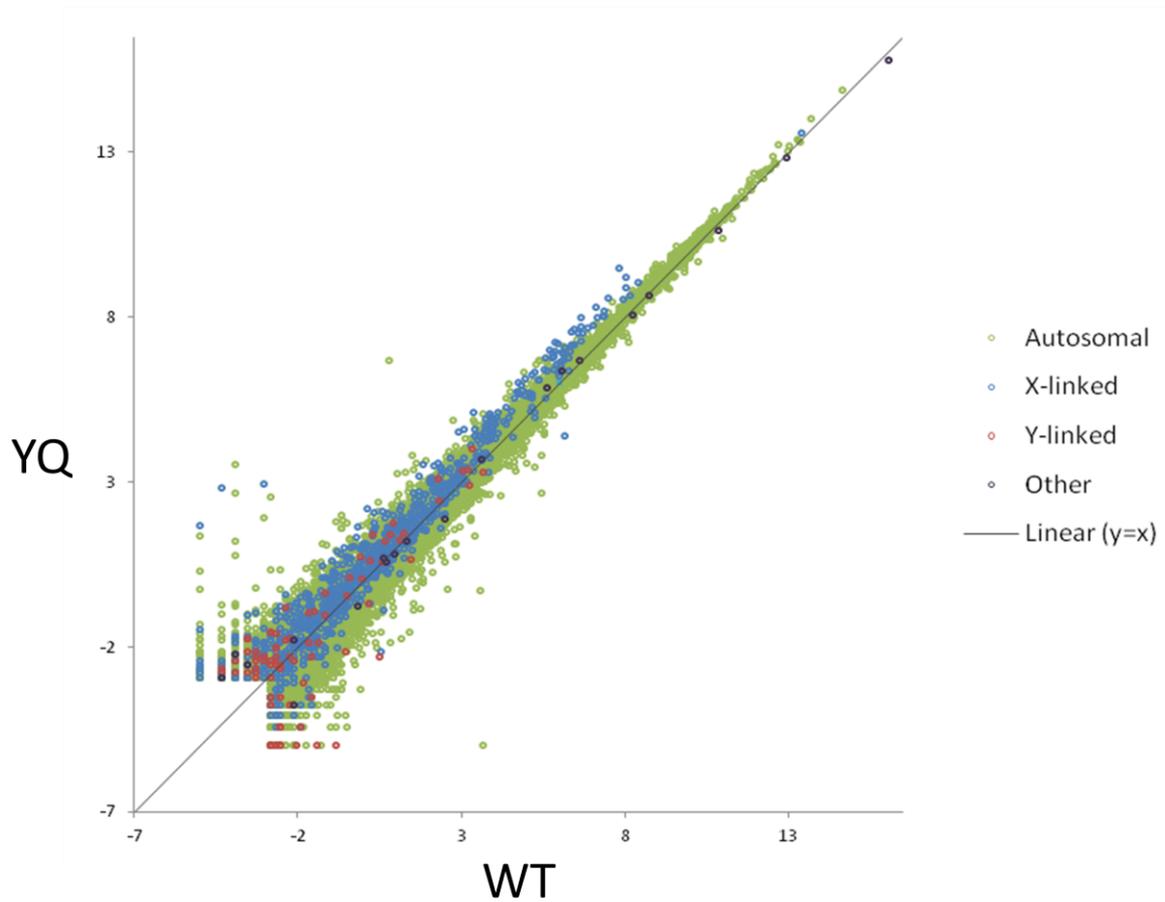


Figure 4.12. Comparison of overall expression of sex-linked and autosomal genes in late stage spermatids whole cell samples between WT and YQ models. Data points represent LogCPM values of individual genes in the comparison.

4.4.2. Expression patterns of control genes across the comparison data sets

Control transcripts of known cellular origin or transcript behaviour were used to establish certain characteristics of the experimental dataset and individual genes within it.

Transcripts from known cell type

Transcripts known to be specifically expressed by certain testicular cell types were used to estimate levels of contamination of enriched cell populations.

		Detection level (CPM values in whole cell sample comparisons)					
		Early		Late		Mature	
Gene symbol	Transcript origin	WT	YQ	WT	YQ	WT	YQ
<i>Rhox5</i>	Sertoli cells	0.23	0.53	2.60	3.37	ND	ND
<i>Sycp3</i>	Primary spermatocytes	41.64	46.76	10.66	6.88	1.06	0.56
<i>Acrv1</i>	Round spermatids	225.38	220.07	16.72	20.61	5.04	1.25
<i>Dbil5</i>	Elongating/condensing and round spermatids	1061.99	951.88	4883.45	4736.31	124.99	33.54
<i>Cyp11a1</i>	Leydig cells	2.36	5.35	21.62	34.12	0.46	0.49
<i>Hba-a1</i>	Erythrocytes	0.21	0.26	13.81	14.91	58.36	17.67
<i>Hbb-bt</i>	Erythrocytes	0.08	0.12	6.88	6.95	33.17	5.85
<i>Hbb-bs</i>	Erythrocytes	0.63	0.86	55.29	57.53	286.04	63.36
<i>Hbb-y</i>	Erythrocytes	ND	ND	0.05	0.15	ND	ND

Table 4.3. Table of CPM values of genes of known transcript origin (cell type) across all experimental cell types in whole cell sample comparisons. ND = not detected.

Direct comparisons between matched samples from WT and YQ were used for the retrieval of CPM values in Table 4.3. Table 4.4 contains CPM values compared between cell types within each genotype. This allowed for a comparison to be drawn with highly purified spermatid populations from Chang et al. 2011 for the assessment of cellular contamination of enriched populations. Comparative expression, in the form of expression value ratios, was used to compare between the findings of Chang et al. (2011) and the current dataset. Values cannot be directly compared as gene expression

was measured by different techniques and different data preparation methods were applied. However measurement within experiments was consistent, therefore ratio values were used to provide a proxy estimation of comparative enrichment.

Gene name	Transcript origin	Detection from Chang et al. 2011			Detection from experimental dataset	
		Early (Round)	Late (Elongating/condensing)	Ratio L:E	Ratio L:E in WT	Ratio L:E in YQ
Rhox5	Sertoli cells	0.630	0.630	1.000	0.089	0.156
Sycp3	Primary spermatocytes	0.540	0.150	3.600	3.906	6.793
Acrv1	Round spermatids	10.600	0.540	19.630	13.480	10.675
Dbil5	Elongating/condensing spermatids and round spermatids	1.720	4.580	0.376	0.217	0.201
Cyp11a1	Leydig cells	0.000	0.002	0.000	0.122	0.157

Table 4.4. Comparison of CPM values in late spermatid and early spermatid populations. The table includes ratios of CPM values across early and late spermatid stages in both WT and YQ samples, and a comparison with highly purified spermatid populations from Chang et al. 2011.

Contamination

Low level presence of *Sycp3*, a highly expressed gene specific to primary spermatocytes, was detected in early spermatid populations. Primary spermatocytes are of a more similar size to round spermatids than elongating/condensing spermatids. Given that samples were enriched via centrifugal elutriation, higher levels of *Sycp3* are therefore expected in early spermatid populations. Particularly low levels of *Rhox5* transcripts demonstrate negligible contamination by sertoli cells; these were absent in mature sperm populations. This also appears to be the case for leydig cells, with low level presence of *Cyp11a1* transcripts evident in late spermatid populations. Erythrocytes were, as expected, almost entirely absent from early spermatid populations, but show a notable presence in mature sperm populations. This may well be exaggerated due to the relatively low transcript abundance in these samples. Following normalisation as outlined above, counts were adjusted for library size (section 4.3.3) which makes data appropriate for examination between matched samples but not ideal for direct comparison with data not normalised concurrently. Transcript levels differ markedly between WT and YQ mature sperm populations. As this is evident across all transcripts within the mature cells, this is likely due to the dramatic increase in expression of a small set of transcripts in WT sperm vs YQ sperm. This is further discussed as a technical note in the Supplementary Material.

Enrichment

Comparisons between early and late stage spermatids in both WT and YQ (*Table 4.4*), show control transcripts expressed with some consistency across the two strains. Ratios of expression values are not dissimilar to those described in Chang et al. 2011. *Acrv1* enrichment was not as great in the experimental dataset as in the published counterpart, suggesting that more round cells may have been present in the late stage spermatid population within our dataset. *Sycp3* 'L:E ratio' was increased in YQ, however this was attributed to a reduced presence of *Sycp3* transcripts in late spermatid populations (*Table 4.3*). Early and late spermatid populations appeared broadly consistent across WT and YQ strains, with enriched populations displaying expression of expected transcripts.

Transcripts of known sharing/non-sharing behaviour

Heatmaps in *Figure 4.13* show the relative distribution (expression values) of control transcripts of known sharing behaviour. This is generated by comparing the logCPM value of a specified gene in one fraction (Sample A) with that from another fraction (Sample B) within the specified cell type, meaning logFC values of comparative expression for that sample combination are displayed. More information on sharing of transcripts across cytoplasmic bridges can be found in sections 1.3.3 and 4.1.2. The heatmaps in *Figure 4.13* show logFC values within wild-type early spermatid stages per gene as examples. *Table 4.5* describes how we might expect the genes to appear, given information from the literature and previous experimental work on their shared or non-shared behaviour.

Gene symbol	Behaviour	Likely enriched	Function	Reference
Akap4	Shared	F1	Fibrous sheath formation	Morales et al. 2002
Prm1	Shared	F1	Sperm protamine (histone replacement)	Caldwell and Handell 1991
Smok2b	Not-shared	F3, some F2	Sperm motility kinase	Veyron et al. 2009 (related) and previous lab work (see section 1.3.3)
Spam1	Likely not-shared	F2, some F3	Sperm head hyaluronidase	Martin-DeLeon et al. 2005

Table 4.5. Table of genes of known shared or non-shared behaviour, including expected transcript distributions and references to work where evidence of sharing status was first provided.

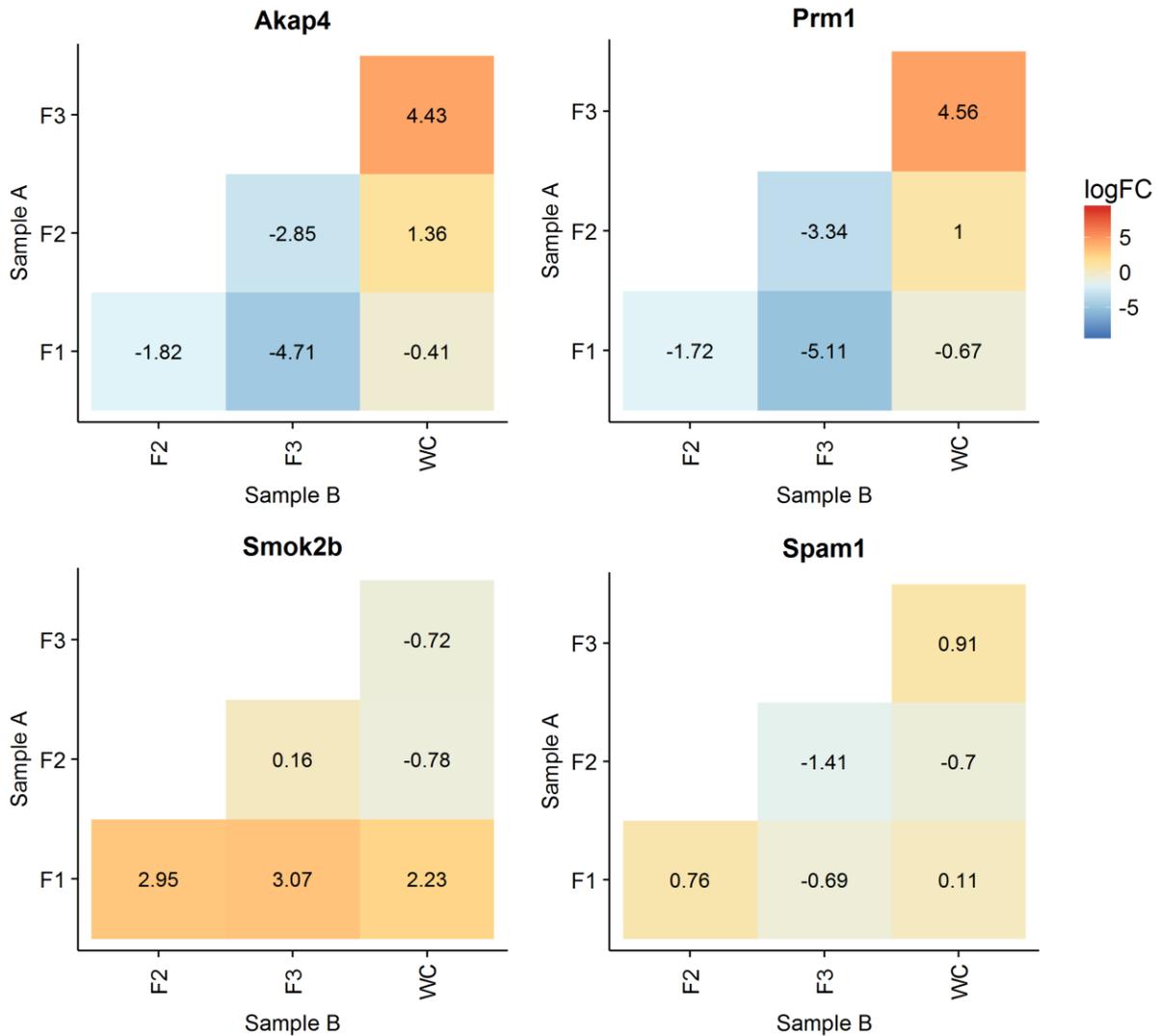


Figure 4.13. Heatmaps of log fold change (logFC) values. The samples compared originate from the wild-type early spermatid cell population only. Genes examined are of known sharing/non-sharing status.

In mice, *Akap4* and *Prm1* produce transcripts that are abundant and known to be shared across cytoplasmic bridges between spermatids (Nipper et al. 2005, Morales et al. 2002, Caldwell and Handell 1991). The transcripts of both genes appear highly enriched in the first (cytosolic) subcellular fraction (F1) and in whole cells when compared to F2 and F3 in the data, as expected.

Smok2b is an autosomal sperm motility kinase gene and has been shown, through RNA *in situ* experiments using heterozygous knockout mice (see section 1.3.3), to remain not-shared between developing spermatids. *Smok2b* shows enrichment in the second and third fractions of early spermatids. *Spam1* is an autosomal sperm head hyaluronidase gene and has been implicated in cases of sex ratio distortion and impaired sperm function in Robertsonian translocation mouse models

involving chromosome 6 (Deng et al. 1997). *Spam1* shows patterns of mild enrichment in the second (cytoskeletally-bound) fraction in the RNA sequencing data, but appears otherwise ambiguous. In the search for non-shared or reduced sharing (cis-limited) candidate genes, the expression patterns of *Smok2b* and *Spam1* (enrichment in F2 and F3) provide another basis for identifying these transcripts.

4.4.3. Investigation of differentially expressed genes: Identification of candidate genes in offspring sex ratio distortion in MF1XY^{RIII}qdel

The differences in sample diversity and library size between samples from different subcellular fractions means that comparing between sample types, e.g. EF1 vs EF3, has to be done with extra considerations. Low abundance reads will be over-represented in F3 compared to F1 as they will make up a larger proportion of detected reads due to a much reduced read diversity and library size in F3.

Selection of candidate genes for further investigation of cis-limitation: Identification of candidate genes in offspring sex ratio skew

Following the production of differential expression data, all analysis was performed by the author using R. All R scripts are available from the online repository⁹. As no p.value was available for the selection of candidate genes, log fold change values (logFC) were used to select genes of interest for further investigation. Genes were selected based on the magnitude of the observed logFC value to produce a list containing the most differentially expressed genes between WT and YQ. Genes were then subsequently filtered on enrichment in fractions 2 and 3 by comparing values in F1, F2 and F3 within each cell type. A comparison between mature sperm samples was performed separately.

The lack of p.values means that highly negative or positive logFC can be misconstrued as significant changes in gene expression when in fact they represent proportionally large changes in lowly expressed or sparsely detected genes. In order to overcome this, filtering based on cpm values was implemented. For each comparison, a first quartile, median, and third quartile cpm value was calculated across all genes, creating a 'cpm threshold' value. For the analysis discussed in this chapter, the third quartile was used. If a gene presented a cpm value of above the threshold value in either or both of the samples within the comparison, the gene would be retained in the dataset. This left only genes that displayed cpm values above cpm threshold values in at least one sample within the comparison. Although stringent filtering by expression values may mean that information on rare transcripts is lost, large changes in higher expressed genes are likely to have a more substantial impact

⁹ Online repository – <https://github.com/EmmaEPJ/Repository>

and result in subsequent phenotypic changes. For each comparison, genes with the twenty greatest (greatest value regardless of positive or negative direction) logFC values were short-listed for further analysis.

Current knowledge surrounding the non-sharing of gene products between sister spermatids is largely restricted to protein-coding genes; in addition, all control transcripts are protein-coding and form the experimental basis for the identification of new cis-limited transcripts. We also know that at least one functional protein-coding gene must be involved in the establishment of sperm phenotypic and reproductive abnormalities seen in YQ males. For these reasons, the major comparison dataset (comparing both spermatid populations and all fractions) was filtered to preserve only genes annotated as protein-coding to allow for additional focus on these genes. This was also used in functional protein association assessment. A separate but complementary comparison of whole cell samples was performed which included both coding and non-coding genes for which examination of potential cis-limitation was also performed (section 4.4.4). An overview of the process of candidate filtering can be seen in *Figure 4.14*.

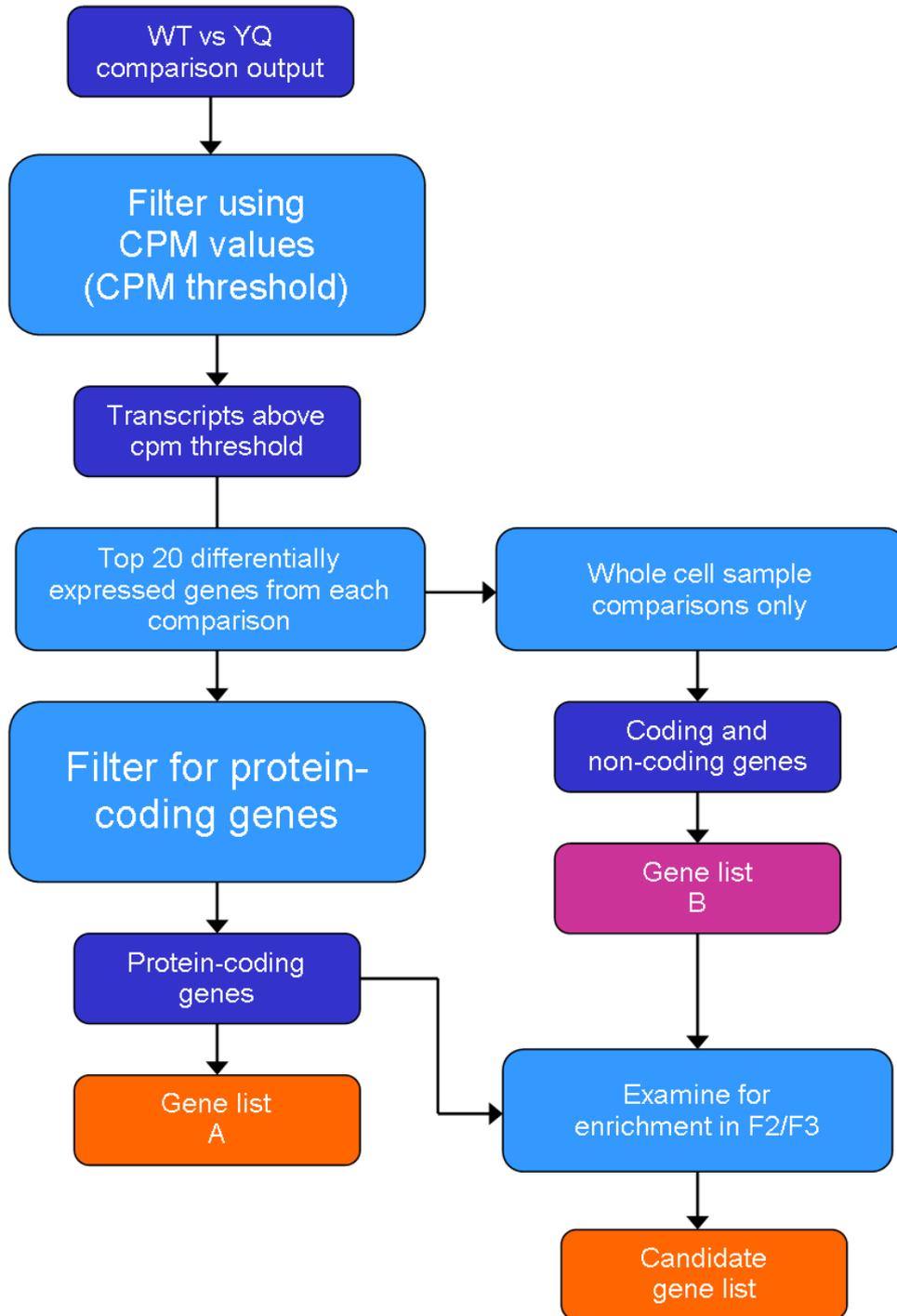


Figure 4.14. Overview of selection criteria for short-listing of genes in RNA sequencing data following differential expression analysis.

These methods resulted in a shortlist of 88 unique genes of interest, of which 63 were protein-coding); many of these genes were shortlisted across several comparisons. Heatmaps of logFC values were generated for simple visual examination of transcript distribution across all cell populations and subcellular compartments, an example is included in *Figure 4.15*. This allowed for the construction of a gene list of 30 genes for further examination of cis-limitation on the basis of enrichment in F2 and F3. Genes were then prioritised by the magnitude of logFC values with consideration of all comparisons, the frequency with which they had been shortlisted, and finally on association with spermiogenesis or implicated pathways for further investigation in the literature and inclusion in the final candidate gene list and discussion (section 4.5).

Table 4.6 shows Gene List A, generated as explained previously and in *Figure 4.14*. This lists genes shortlisted for further investigation and their chromosomal origin. A full table of shortlisted genes, their chromosomal origin, associated LogFC values and full gene name are included in the Supplementary Material. *Table 4.7* lists genes shortlisted from Gene List A that show transcript distribution patterns aligned with those of known non-shared transcripts (enrichment in F2 or F3 over F1 and a reduced presence in F1 compared to WC sample) for each cell type. All listed genes showed this pattern of transcript enrichment in at least one of the four assessed spermatid populations (WT early, WT late, YQ early, YQ late); this is indicated in the table per cell population by a red cell containing an 'X'. Given concerns regarding sample YQEF2, enrichment in F1 over F2 was not considered as an indicator of potential cis-limitation in the assessment of early stage spermatid populations. Gene names were retrieved from STRING v10.510 (Szklarczyk et al. 2015); where absent, names were obtained from MGI¹¹ and UniProt¹². Where MGI-obtained names were more informative than the original annotation, the original is provided in brackets within the 'Gene symbol' column.

¹⁰ Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) v10.5 - <https://string-db.org/>

¹¹ MGI - <http://www.informatics.jax.org/>

¹² Universal Protein Resource (UniProt) - <http://www.uniprot.org/>

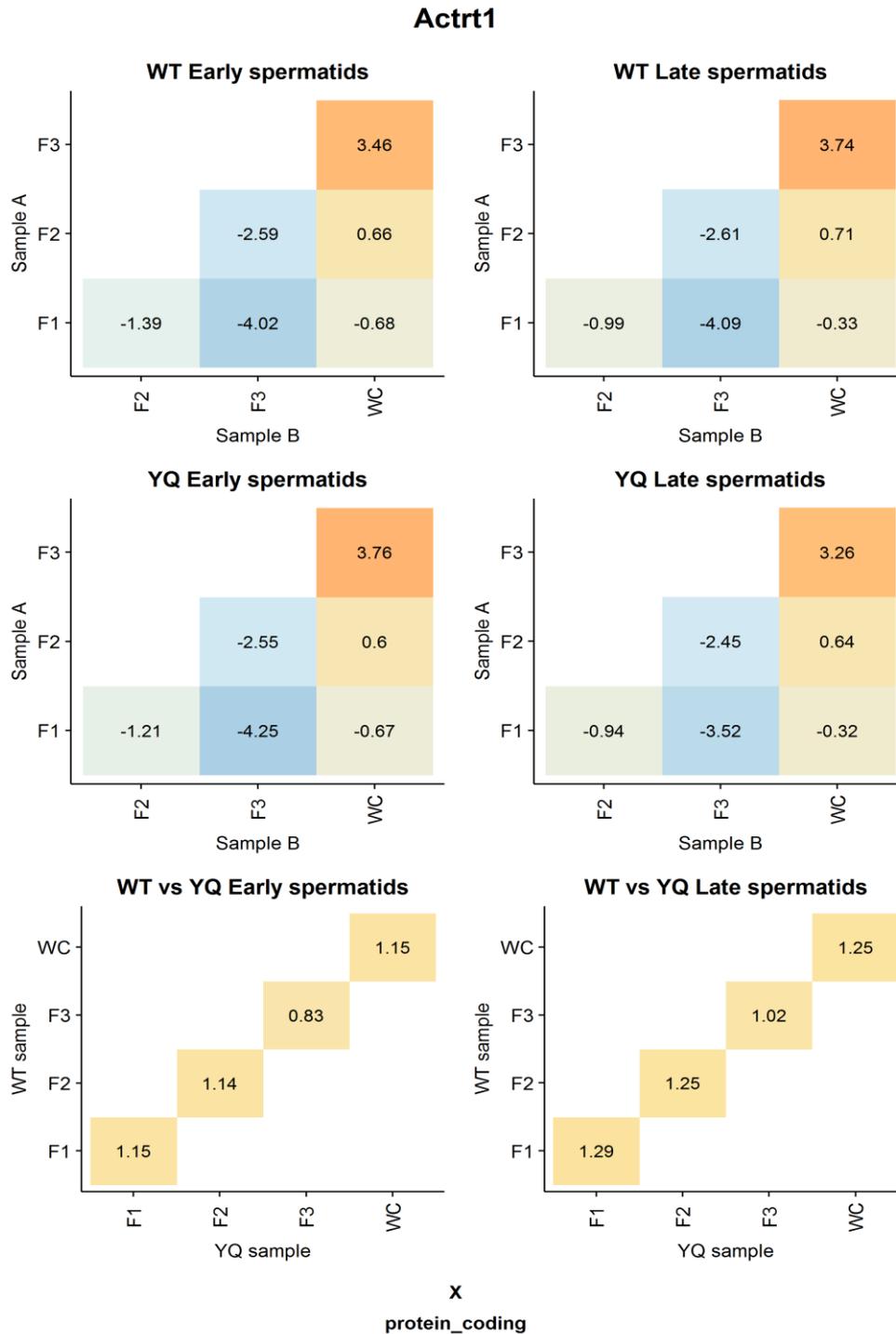


Figure 4.15. Heatmaps of log fold change (logFC) values across all spermatid sample comparisons (comparison of mature sperm samples not shown). Heatmaps allow for easy visual identification of relative transcript distribution on a gene-by-gene basis. LogFC values are values as calculated Sample A vs Sample B, or WT sample vs YQ sample (as in the bottom two tables). *Actrt1* shows transcript distribution typical of shared genes. This distribution pattern does not change across cell types or between strains; however, an overall upregulation of the gene can be seen when matched samples are compared between strains (bottom tables).

Gene list A: Protein-coding genes shortlisted on top 20 LogFC within sample comparisons					
Gene symbol	Chromosome	Gene symbol	Chromosome	Gene symbol	Chromosome
<i>2010109A12Rik</i>	5	<i>Fos</i>	12	<i>Myl12a</i>	17
<i>4930402K13Rik</i>	X	<i>Gm10260</i>	13	<i>Myl12b</i>	17
<i>4930451G09Rik</i>	16	<i>Gm10800</i>	2	<i>Neo1</i>	9
<i>Actrt1</i>	X	<i>Gm10801</i>	2	<i>Nlrc5</i>	8
<i>Antxr2</i>	5	<i>Gm11639</i>	11	<i>Nptx2</i>	5
<i>Axin2</i>	11	<i>H2a1k (Gm14501)</i>	X	<i>Pkd111</i>	11
<i>BC061237</i>	14	<i>Btbd35f27 (Gm3763)</i>	X	<i>Ppp2r5a</i>	1
<i>Cd55</i>	1	<i>Gm382</i>	X	<i>Rd3</i>	1
<i>Cers3</i>	7	<i>H2a1n (Gm5382)</i>	X	<i>Scfd2</i>	5
<i>Cldn34b4</i>	X	<i>Gm614</i>	X	<i>Shroom3</i>	5
<i>Cnga3</i>	1	<i>Gm904</i>	13	<i>Slc26a10</i>	10
<i>Cpne7</i>	8	<i>Gml2</i>	15	<i>Spata13</i>	14
<i>Cpxcr1</i>	X	<i>Gpkow</i>	X	<i>Tbl3</i>	17
<i>Cylc1</i>	X	<i>H2-D1</i>	17	<i>Tmc1</i>	19
<i>Dnah11</i>	12	<i>Hist1h2ba</i>	13	<i>Trappc2l</i>	8
<i>Dynlt1a</i>	17	<i>Hmcn1</i>	1	<i>Trio</i>	15
<i>Dynlt1f</i>	17	<i>Klk1b26</i>	7	<i>Tspan6</i>	X
<i>Ell</i>	8	<i>Lars2</i>	9	<i>Vars2</i>	17
<i>Fam122c</i>	X	<i>Lyz2</i>	10	<i>Vit</i>	17
<i>Fez1</i>	9	<i>Mdga2</i>	12	<i>Vsig1</i>	X
<i fn1<="" i=""></i>	1	<i>Meiob</i>	17	<i>Xirp1</i>	9

Table 4.6. List of protein-coding genes shortlisted by LogFC within sample comparisons. Genes showing differential expression in multiple sample comparisons are shown only once. Includes both sex-linked and autosomal genes.

Gene symbol	Chr	WTE	YQE	WTL	YQL	Gene symbol	Chr	WTE	YQE	WTL	YQL
<i>2010109A12Rik</i>	5		X			<i>Gpkow</i>	X	X	X	X	
<i>Antxr2</i>	5				X	<i>H2-D1</i>	17	X	X	X	X
<i>Axin2</i>	11	X	X	X	X	<i>Hmcn1</i>	1	X	X	X	X
<i>Cldn34b4</i>	X	X	X			<i>Lars2</i>	9	X	X	X	X
<i>Cpne7</i>	8	X	X	X	X	<i>Lyz2</i>	10	X	X		X
<i>Cylc1</i>	X	X		X		<i>Mir6236</i>	9	X	X	X	X
<i>Dnah11</i>	12	X	X			<i>Myl12a</i>	17				X
<i>Fn1</i>	1	X	X	X	X	<i>Myl12b</i>	17				X
<i>Fos</i>	12	X	X	X	X	<i>Neo1</i>	9				X
<i>Gm10800</i>	2	X	X	X	X	<i>Nlrc5</i>	8	X	X		X
<i>Gm10801</i>	2	X	X	X	X	<i>Pkd1l1</i>	11	X	X		
<i>Gm11639</i>	11	X	X			<i>Spata13</i>	14	X		X	X
<i>Gm382</i>	X			X		<i>Tmc1</i>	19	X	X		
<i>Gm904</i>	13	X	X	X	X	<i>Trio</i>	15	X	X	X	X
<i>Gml2</i>	15	X	X			<i>Vars2</i>	17			X	

Table 4.7. Genes shortlisted as potentially cis-limited based on criteria previously described. Red boxes containing an 'X' indicate cell populations in which these patterns of transcript distribution (enrichment in F2 or F3 over F1 and reduced presence of transcripts in F1 compared to WC) were evident.

4.4.4. Investigation of differential expression of protein-coding and non-coding genes across whole cell samples

All genes in early and late spermatid whole cell samples were examined for overall differential expression. Genes were first shortlisted based on cpm values across the comparison dataset. Genes showing expression equal to or greater than the third quartile in one or both samples, as calculated by the 'summary' function in R, were selected. The top differentially expressed genes were then shortlisted. This examination was restricted to whole cell spermatid samples in order to examine large scale dysregulation and, as fractionated samples implicitly represent part of the whole cell sample, to avoid repetition in detection. Both coding and non-coding genes were included in this examination and are listed below (Table 4.8 and Table 4.9). The genes were shortlisted using R; scripts can be found in the online repository. Full gene names were obtained from the Mouse Genome Informatics

(MGI) database¹³; where MGI-obtained names were more informative than the original annotation, the original is provided in brackets within the ‘Gene name’ column. Additional gene information was collected from UniprotKB/Swissprot¹⁴; this can be seen in the Supplementary Material.

¹³ MGI - <http://www.informatics.jax.org/>

¹⁴ UniProtKB/SwissProt - <http://www.uniprot.org/>

Top 20 differentially expressed genes in early stage spermatids					
	Gene symbol	Gene name	Gene biotype	Chr	LogFC
Differentially expressed in early stage spermatids	<i>Gm15772</i>	predicted gene 15772	processed pseudogene	5	5.72
	<i>Gm1821</i>	Predicted gene 1821	processed pseudogene	14	-2.74
	<i>BC061237</i>	cDNA sequence BC061237	protein coding	14	-2.66
	<i>Slc26a10</i>	Solute carrier family 26, member 10	protein coding	10	-1.53
	<i>H2al1n</i>	H2A histone family member L1N (Gm5382)	protein coding	X	-1.51
	<i>Gm5591</i>	Predicted gene 5591 (Gm6866)	transcribed unprocessed pseudogene	7	1.50
	<i>Hist1h2ba</i>	Histone cluster 1, H2ba	protein coding	13	1.47
	<i>Vit</i>	Vitrin	protein coding	17	1.45
	<i>Dynlt1a</i>	Dynein light chain Tctex-type 1A	protein coding	17	1.41
	<i>4930451G09Rik</i>	RIKEN cDNA 4930451G09 gene	protein coding	16	1.33
	<i>Trio</i>	Triple functional domain - PTPRF interacting	protein coding	15	-1.33
	<i>Shroom3</i>	Shroom family member 3	protein coding	5	-1.33
	<i>Klk1b26</i>	Kallikrein 1-related peptidase b26	protein coding	7	1.31
	<i>Gm4665</i>	predicted gene 4665	antisense	9	1.30
	<i>Vsig1</i>	V-set and immunoglobulin domain containing 1	protein coding	X	1.30
	<i>Dynlt1f</i>	Dynein light chain Tctex-type 1F;	protein coding	17	1.29
	<i>Tspan6</i>	Tetraspanin 6	protein coding	X	1.25
	<i>Gm15104</i>	predicted gene 15104	lincRNA	X	1.22
	<i>Trappc2l</i>	Trafficking protein particle complex 2-like	protein coding	8	-1.16
	<i>Actrt1</i>	Actin-related protein T1	protein coding	X	1.15

Table 4.8. Table containing the top twenty differentially expressed genes in YQ early stage spermatids when compared to WT.

Top 20 differentially expressed genes in late spermatids					
	Gene symbol	Gene name	Gene biotype	Chr	LogFC
Differentially expressed in late spermatids	<i>Rps13-ps1</i>	Ribosomal protein S13, pseudogene 1	processed pseudogene	8	-8.64
	<i>Gm10260</i>	Predicted gene 10260	protein coding	13	7.42
	<i>Gm15772</i>	Predicted gene 15772	processed pseudogene	5	5.92
	<i>2010109A12Rik</i>	RIKEN cDNA 2010109A12 gene	protein coding	5	-3.90
	<i>Gm1821</i>	Predicted gene 1821	processed pseudogene	14	-2.81
	<i>Gml2</i>	Glycosylphosphatidylinositol anchored molecule like 2	protein coding	15	2.29
	<i>Nptx2</i>	Neuronal pentraxin 2	protein coding	5	2.09
	<i>Gm15674</i>	Predicted gene 15674	antisense	1	-2.09
	<i>Cpne7</i>	Copine VII	protein coding	8	1.97
	<i>Mdga2</i>	MAM domain containing glycosylphosphatidylinositol anchor 2	protein coding	12	-1.80
	<i>H2al1n</i>	H2A histone family member L1N (Gm5382)	protein coding	X	-1.79
	<i>H2al1k</i>	H2A histone family member L1K (Gm14501)	protein coding	X	1.74
	<i>Cnga3</i>	Cyclic nucleotide gated channel alpha 3	protein coding	1	1.73
	<i>Btbd35f27</i>	BTB domain containing 35, family member 27 (Gm3763)	protein coding	X	1.70
	<i>Slc26a10</i>	Solute carrier family 26, member 10	protein coding	10	-1.68
	<i>Vsig1</i>	V-set and immunoglobulin domain containing 1	protein coding	X	1.63
	<i>Antxr2</i>	Anthrax toxin receptor 2	protein coding	5	-1.62
	<i>4930402K13Rik</i>	RIKEN cDNA 4930402K13 gene	protein coding	X	1.58
	<i>Cd55</i>	CD55 antigen	protein coding	1	1.52
	<i>Gm382</i>	Predicted gene 382	protein coding	X	1.51

Table 4.9. Table containing the top twenty differentially expressed genes in YQ late stage spermatids.

Mature sperm samples

Mature sperm are considered to be transcriptionally inactive, with the exception of a number of genes of mitochondrial origin (Grunewald et al. 2005, Premkumar and Bhargava 1972). During the later stages of spermiogenesis, the recruitment of protamines to the spermatid DNA leads to the ultra-tight packaging of chromatin with the sperm nucleus. Only a few nucleic acid sites within the sperm genome remain bound by nucleosomes, whilst the vast majority is tightly packaged up by protamines. Although transcriptionally silent, mature sperm do contain RNA, estimated between 10 and 100 fg RNA per cell (Krawetz 2005). This RNA is however thought to be highly diverse, with mRNAs, miRNAs, interference RNAs (iRNAs) and additional species found in mature sperm (reviewed by Hosken and Hodgson 2014, Peng et al. 2012); though little is known about the function of many of these RNAs. A recent boon in the study of small non-coding RNAs has shown that these are abundant in mature sperm. Small non-coding RNAs, including micro RNA and piwi RNA transcripts, in sperm have been implicated in disease and health outcomes (especially metabolic) for offspring; there are now a large number of studies published on this phenomenon in mice alone (Fullston et al. 2016, Grandjean et al. 2015, Rodgers et al. 2015, Sharma et al. 2015, Rodgers et al. 2013). However, the composition of genes in mature sperm is unlikely to relate directly to physiological differences noted in our model, due to the mature sperm being fully formed; it may however reveal differences in the composition of RNAs passed on to the oocyte and link to effects in developing embryos. Mature sperm samples were examined separately to spermatids. As the comparison between mature sperm is not the main focus of the investigation a list of the top 10, rather than 20, differentially expressed genes is listed in *Table 4.10*.

The top expressed genes in the mature cell comparison between strains were, in the vast majority, of mitochondrial origin. Two mitochondrially encoded rRNAs (*mt-Rnr2* and *mt-Rnr1*) that were not excluded during rRNA filtering occupied the top two places. Mitochondrially encoded protein coding genes *mt-Co1* and *mt-Nd1* were the top expressed protein-coding genes, alongside lincRNAs *Gm26917* and *Malat1*, and miRNA *Gm23935*. This fits well with the findings of other mouse sperm RNA work (Johnson et al. 2015, review by Jodar et al. 2013). An additional technical note on examination of mature sperm samples and the cautious interpretation of results is included in the Supplementary Material.

Top 10 differentially expressed genes in mature sperm					
	Gene symbol	Gene name	Gene biotype	Chr	LogFC
Differentially expressed genes in mature sperm	<i>Gm15772</i>	Predicted gene 15772	processed pseudogene	5	5.11
	<i>Snora78</i>	Small nucleolar RNA, H/ACA box 7	snoRNA	17	-4.59
	<i>Gm23442</i>	Predicted gene, 23442	snoRNA	17	-4.37
	<i>Tuba3b</i>	Tubulin, alpha 3B	protein coding	6	-4.24
	<i>Actl7b</i>	Actin-like 7b	protein coding	4	-3.91
	<i>Hils1</i>	Histone H1-like protein in spermatids 1	protein coding	11	-3.82
	<i>Mroh4</i>	Maestro heat-like repeat family member 4	protein coding	15	-3.82
	<i>Actl9</i>	Actin-like 9	protein coding	17	-3.78
	<i>Gm23935</i>	Predicted gene, 23935	miRNA	16	-3.78
	<i>Gm6483</i>	Predicted gene 6483	protein coding	8	3.76

Table 4.10. Table containing the top ten differentially expressed genes in YQ mature sperm.

A single additional gene was shortlisted as potentially cis-limited based on criteria previously described. *Gm6866* was enriched in F2 and F3 over F1 and showed reduced presence of transcripts in F1 compared to WC in round spermatids. This gene was reported as having a ‘biotype conflict’ between various databases; it was registered variously as a pseudogene and as protein-coding.

4.4.5. Functional Annotation Analysis of shortlisted protein-coding genes by STRING (Search Tool for the Retrieval of Interacting Genes/Proteins)

STRING¹⁵ v.10.5 utilises information from published experimental work, curated datasets, and text-mining to evaluate the searched terms for potential interactions (Szklarczyk et al. 2015).

Examination of differentially expressed protein-coding genes across all sample comparisons – gene list A

To provide insights into gene ontology and expose potential biochemical pathways for involvement in offspring sex ratio distortion, a functional protein association search was performed using the STRING platform (Szklarczyk et al. 2015). Sixty-two genes shortlisted for further examination as

¹⁵ Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) v.10.5 - <https://string-db.org/>

outlined in 4.4.3 were examined using this method. *Nlrc5* was excluded from STRING analysis as it, and its aliases, were not recognised by the STRING platform. *Figure 4.16.* and *Figure 4.17.* are graphical outputs of functional protein association networks identified through STRING.

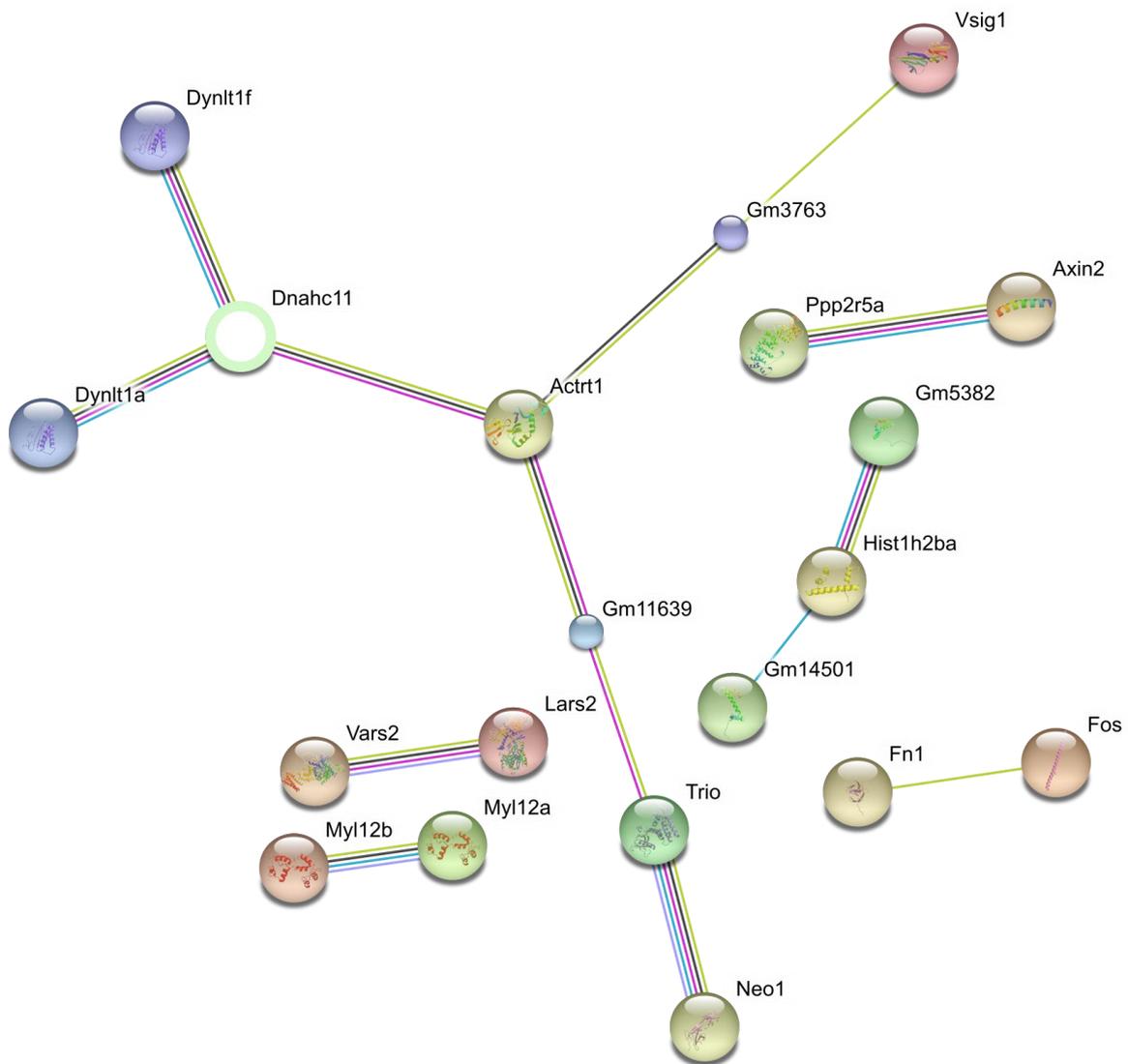


Figure 4.16. Graphical output of STRING protein functional association analysis at medium confidence. Unconnected nodes have been excluded from the image. Lines between nodes indicate the cooccurrence of connected nodes in curated datasets. Images within nodes indicate protein structure for proteins where structure is known or predicted.

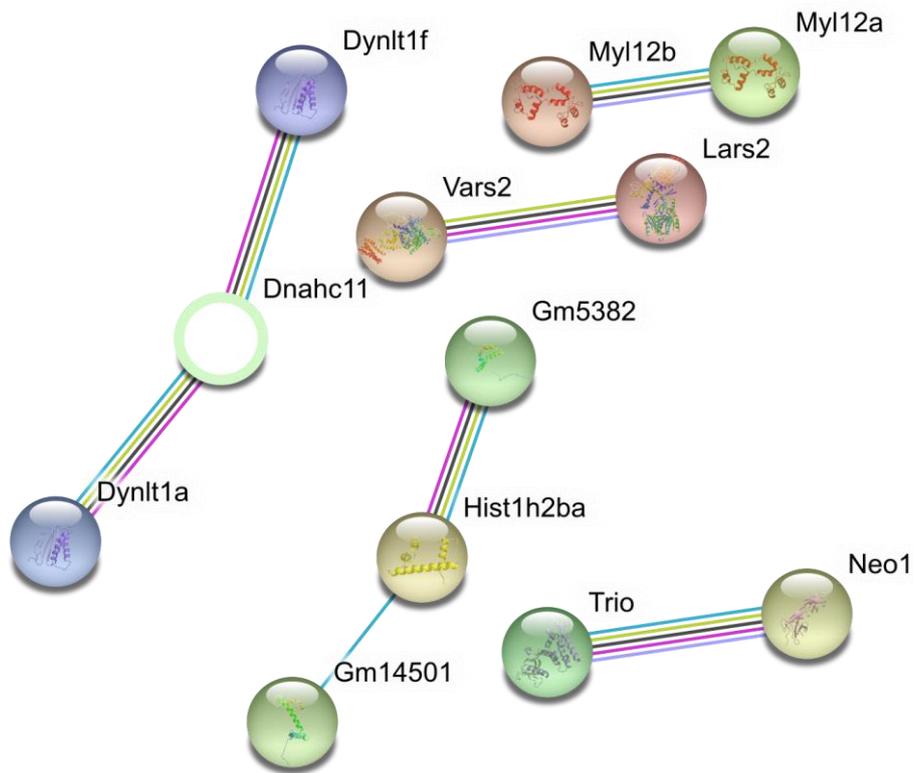


Figure 4.17. Remaining functional association networks at high confidence. Unconnected nodes have been excluded.

The functional association networks seen above (*Figure 4.16* and *Figure 4.17*), and therefore the associated gene list, showed no significant enrichment as determined by STRING analysis.

All genes included in the STRING analysis were inherently testis-expressed. This is likely to result in more connections than would otherwise be seen from a list of sixty-two randomly selected protein-coding genes from across the genome. Although no significant enrichment was found by this analysis, there are a number of connections which reveal interactions between individual genes. These are further discussed in section 4.5.

4.4.6. Examination of genes of interest in RNA sequencing dataset

Prior to commencement of experimental work, a number of genes of interest were identified from the literature (section 1.8). These include genes mentioned as potential instigators of transmission ratio distortion in other animal models, and genes implicated in similar physiological aberrations as noted in our model. A brief summary of their behaviour in this dataset is given below (*Table 4.11*). *Tcp11x2* shows strong up-regulation across YQ samples.

Gene symbol	Transcript distribution in RNA sequencing dataset
<i>Alkbh</i>	Eight members of the <i>Alkbh</i> family were recorded in the sequencing data. Only <i>Alkbh1</i> , the gene identified as being involved in a case of transmission ratio and associated sex ratio distortion (Nordstrand et al. 2010), showed enrichment in F3 and F2 in both early and late spermatids. <i>Alkbh5</i> showed similar (though weaker) enrichment in late spermatids. These genes did not show altered expression between WT and YQ spermatids across any comparison.
<i>Appbp2</i> , <i>Dkk11</i> , and <i>Kat5</i>	None of the genes showed enrichment in F2 or F3, and did not show altered expression between WT and YQ.
<i>Pgrmc1</i>	<i>Pgrmc1</i> showed strong enrichment in F3 and F2 in late spermatids, but did not show altered expression between strains.
<i>Speer</i>	Ten <i>Speer</i> genes were identified in the dataset. Only two of these did not show strong enrichment in F3 and F2 in early spermatids; <i>Speer2</i> and <i>Speer4cos</i> . All genes except <i>Speer2</i> showed some upregulation in at least one comparison (defined as LogFC of 0.5 or above).
<i>Tcp11</i>	<i>Tcp11</i> does not show either enrichment in F2 or F3, or altered expression between strains.
<i>Tcp11x2</i>	<i>Tcp11x2</i> was not shortlisted following differential expression analysis but does show upregulation across all comparisons. It does not show enrichment in either F2 or F3.

Table 4.11. Summary of transcript distribution of candidate genes identified from the literature.

4.5. Chapter Discussion

This chapter covers work performed and results retrieved from an investigative single replicate RNA sequencing transcriptome examination aimed at highlighting dysregulated genes in spermatids and mature sperm of the MSYq deletion model MF1XY^{RIII}del. The investigation sought to confirm proposed differences in transcript distribution between genes of known sharing or non-sharing status, and further indentify potential candidate genes for involvement in the previously described offspring sex ratio and sperm phenotypic abnormalities seen in the model. This was conducted through direct comparison of matched control strain MF1XY^{RIII} (WT) and deletion strain MF1XY^{RIII}qdel (YQ) spermatid RNA samples. Following short-listing of differentially expressed genes, comparisons of transcript abundance in subcellular fractions were made to assess transcript distribution within specific cell types and potential for non-sharing (cis-limitation) of RNA transcripts. This yielded a number of coding and non-coding genes for further investigation including several potentially cis-limited candidates.

Upregulation of sex chromatin across spermatid samples

Due to the known consequences of *Sly* deficiency and actions of *Sly* and *Slx* (see section 1.4.4 for more detail), an increase in the expression of sex chromatin is expected in MF1XY^{RIII}qdel (YQ) spermatids compared with paired samples from MF1XY^{RIII} (WT). This can be observed across all cell types and fractions. Additional upregulation of *Vsig1*, and mild up-regulation of *Speer* family members was also noted. This conforms to previous examinations of the spermatid transcriptome as a consequence of loss of *Sly* (Cocquet et al. 2012, Cocquet et al. 2009).

Examination of genes of interest identified in the literature and during physiological characterisation

Tcp11x2 and *Speer* family showed enrichment in YQ compared to WT. Enrichment of at least some members of the *Speer* cluster was expected and has been reported in previous work with mice with reduced *Sly* expression (Cocquet et al. 2012, Cocquet et al. 2009), however upregulation was not substantial enough for it to be shortlisted. *Tcp11x2* did show strong upregulation across all comparisons, suggesting potential importance in reported physiological differences. However it did not show patterns of enrichment in F2 or F3 or similar transcript abundance profiles to known non-shared genes, meaning it is not currently expected to be directly involved in causing functional inequalities in sperm leading to offspring sex ratio skewing. *Alkbh1*, known to be involved in transmission ratio distortion (TRD) and associated sex ratio skewing (Nordstrand et al. 2010) showed strong enrichment in F2 and F3, suggesting transcripts were not free in the cytosol (or bound to cytosolic polysomes). This observation fits well with their involvement in TRD.

Four X-linked genes show potential cis-limitation

Four of the genes shortlisted for potential cis-limitation were also X-linked. These were *Cldn34b4*, *Cylc1*, *Gm382*, and *Gpkow*. *Cldn34b4* is a claudin; claudins are major constituents of tight junctions and are especially important in the testis. Claudin 11 loss has been linked with infertility in humans and mice where deficiency has been shown to alter and destroy the sertoli cell epithelial phenotype (Mazaus-Guittot et al. 2010). *Cylc1* is a cyclin, a protein of the sperm head cytoskeleton. It is involved in formation of the cytoskeletal calyx (or post-acrosomal sheath), which is the basal region of the perinuclear theca, a condensed protein layer that surrounds the sperm nucleus except in the tail attachment region and makes up most of the cytosol remaining in the sperm head. This transcript appears expressed and upregulated in YQ in both spermatid cell populations. In addition, upregulation of CYLC1 and CYLC2 (along with a few other cell-cycle related genes) has been noted in men with treatable maturational arrest during sperm development (Shiraishi et al. 2017). *Gm382* is a predicted

gene with potential RNA-binding activity. A BLAST¹⁶ search reveals high sequence similarity (91% identity) to the X-linked RGD1563104 ('similar to Vigilin – high density lipoprotein-binding protein') in rat. *Gpkow* is the 'G patch domain and KOW motif' gene and is upregulated across all fractions in YQ early spermatids, especially fraction 3. As these genes are both X-linked and show enrichment in either F2 or F3 in one or more spermatid cell populations, they are now principal candidates for sperm inequality in sperm from the same ejaculate.

Wnt signalling pathway

The Wnt signalling pathway is a highly evolutionarily conserved pathway involving a number of different genes. Several components of the Wnt signalling pathway are expressed in the murine testes. Koch et al. (2015) showed that mature spermatozoa respond to Wnt signals released from the epididymis, and mice that were mutant for specific Wnt regulators were sterile due to immotile and malformed sperm. They also showed that Wnt signalling acts on mature sperm in part through inhibition of protein phosphatase to initiate sperm motility. The non canonical Wnt/calcium signalling pathway has been linked to cytoskeletal reorganisation and regulation of calcium inside the cell. Calcium signalling in sperm is essential for major sperm processes such as capacitation and hyperactivation, providing hypermotility and preparing the sperm to undergo the acrosome reaction (Breitbart 2002, review by Rahman et al. 2014). Previously, DKKL1 protein has been shown to interact with SLY1 in the testes and is part of the family of secreted antagonists of Wnt signal transduction (Reynard et al. 2009), although the outcome of this interaction is not fully understood. In our examination, a number of genes previously shown to play a role in Wnt-signalling were seen to be strongly upregulated in YQ spermatids. These included autosomal *Axin2* and *Fos*, which also showed potential for cis-limitation. Genes linked to protein-phosphorylation such as protein phosphatase 2 regulatory subunit *Ppp2r5a* also showed general low level down-regulation and altered transcript distribution (increased in 'bound' fraction 3) in YQ late spermatids. The connection between *Axin2* and *Ppp2r5a* was also shown following STRING analysis (section 4.4.5). In addition, several dysregulated genes show calcium binding or are involved in calcium mediated reactions. Due to roles in calcium availability for both mature sperm and maturing spermatids, roles in cytoskeletal reorganisation, and in epididymal sperm maturation, the Wnt signalling pathways present potential pathways of action for both X-/Y-bearing sperm specific differences and morphological abnormalities observed across the MF1XY^{RIII}qdel strain.

¹⁶ BLAST (Basic Local Alignment Search Tool) - <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Dysregulation of genes linked to chromatin condensation

Three shortlisted genes (*Hist1h2ba*, *Gm14501*, *Gm5382*) were shown to have high confidence connections following analysis by STRING. Predicted genes *Gm14501* and *Gm5382* were revealed, by subsequent annotation through MGI, to be *H2al1k* and *H2al1n* respectively. *Hist1h2ba* is a histone variant specifically required to direct the change of dissociating nucleosomes to protamine in male germ cells (Montellier et al. 2013). It forms part of a heterodimer in condensing spermatids that promotes the loading of transition proteins TNP1 and TNP2 prior to protamine recruitment. *H2al1k* and *H2al1n* are members of the H2A histone family. These are histone coding genes, proteins that form the nucleosome. Interestingly *H2al1k* and *H2al1n* genes are X-linked (*Hist1h2ba* is found on chromosome 13) and all three genes show strong dysregulation in YQ spermatids. Altered transcription of these genes is likely to result in altered chromatin packaging in maturing sperm cells. In addition all three genes were implicated in the *Hoxc4* pathway. A study examining aberrant methylation and DNA-packing in human sperm found that hypomethylation (a decrease in methylation of cytosine and adenosine residues) of *Hoxc4* was observed in sperm with poor motility (Pachecho et al. 2011). The implications of altered histone transcription are further investigated in Chapter 7.

Consistent up-regulation of predicted pseudogene *Gm15772* in YQ cells

Gm15772 appears strongly upregulated in both early and late spermatids in whole cell and F1 and F2 samples. *Gm15772* is a predicted pseudogene on chromosome 5 that sits within a CNV region (Henrichsen et al. 2009) and across a promoter. The promoter lies upstream of the gene on the forward strand. Using ORFfinder (Open Reading Frame finder), six potential open reading frames (ORFs) were identified. SmartBLAST revealed 100% amino acid identity of ORF2 to Human 60S ribosomal protein L26 from *RPL26*. Annotation of this gene in humans reveals that RPL26 forms part of the 60S subunit of the human ribosome, and the protein is located in the cytoplasm. It is involved in viral mRNA translation and rRNA processing in the nucleus and cytosol; leading to potential role for this gene in gene regulation. The strong dysregulation of this gene across the majority of samples suggests it may play a role in outcomes of MSYq deletion in this model. This gene will be examined in subsequent investigations.

Potential cis-limitation of non-coding RNAs

Although non-sharing of non-coding genes has not been shown in spermiogenesis, non-coding genes dysregulated in YQ were examined for transcript distribution in case further interest arises in subsequent investigations.

4.6. Chapter Summary

Through the enrichment and subcellular fractionation of spermatids, and subsequent RNA sequencing analysis outlined in this chapter, it has been possible to examine dysregulation of genes and relative transcript abundance within and between two spermatid developmental stages in the MSYq deletion model MF1XY^{RIII}qdel (YQ) and its congenic control MF1XY^{RIII} (WT). This investigation has resulted in the identification of candidate genes for involvement in both the observed offspring sex ratio skew (including four X-linked candidates) and sperm phenotypic abnormalities seen in the YQ model.

Amongst short-listed genes were a number of genes potentially linked to the morphological aberrations observed in MF1XY^{RIII}qdel sperm heads. Several of these genes exhibited potential cis-limitation, including the X-linked gene *Cylc1*. Others showed dysregulation across several matched samples without cis-limitation such as X-linked *Actrt1* and autosomal *Nptx2*; indicating potential roles in global sperm phenotypic abnormalities not linked to sex chromosome content. Both *Cylc1* and *Actrt1* show strong functional links to sperm head calyx formation in one or more mammalian species (Heid et al. 2002, Hess et al. 1993).

Genes showing dysregulation through RNA sequencing will be further examined in subsequent chapters.

Chapter 5

5. Microarray Study of Gene Expression in Spermatids

In this chapter, the results of a microarray study examining gene expression over three stages of sperm maturation (early spermatid, late spermatid, and mature sperm) and in subcellular fractions (broadly: cytosolic, cytoskeleton-bound, nuclear and organelle-bound) of spermatid populations are examined. Testis cell RNA samples were collected from MF1XY^{RIII} and MF1XY^{RIII}qdel mouse strains and examined between and within both strains. Each of the sets of testicular cell populations (three replicates per strain) was generated from separate pools of cells derived from four age-matched mice, between 89 and 93 days of age at the time of use (4 x MF1XY^{RIII} and 4 x MF1XY^{RIII}qdel per replicate set), generating 54 samples for analysis. Differential expression analysis across the comparison groups yielded several candidate genes for both observed physiological differences between strains and the recorded offspring sex ratio skew (see Chapter 3). Differential expression analysis was conducted across the full transcriptome. Whole cell and fractionated spermatid samples were examined for differences in overall expression and transcript localisation. Following differential expression analysis, an overall up-regulation of sex-linked genes (excepting genes contained within the MSYq deleted region) was noted in YQ samples. Shortlisted candidate genes were assessed for transcript non-sharing (cis-limitation) potential through examination of subcellular enrichment and comparison with control genes of known sharing/non-sharing behaviour.

5.1. Chapter Introduction

An extended introduction and theoretical background to the processes undertaken in this and other expression analysis chapters is given in Chapter 4 section 4.1. A brief chapter-specific introduction is included below.

5.1.1. Hypothesis and experimental design

Following on from the genome wide single-replicate transcriptome study reported in Chapter 4, in this chapter we aim to examine consistent, significant transcriptional differences across multiple replicate data. This study uses microarray expression analysis to examine RNA transcript abundance in developing spermatids and within subcellular regions of developing spermatids in the MF1XY^{RIII} and MF1XY^{RIII}qdel models. Microarray is a chip and probe based method of assessing RNA abundance in purified RNA samples and provides an alternative and distinct method to high-throughput RNA

sequencing (see section 1.6 for a comparison of the two techniques). In the previous chapter an examination of transcripts of known sharing status were used as an experimental basis for the identification of potential cis-limitation. In this chapter, this examination is repeated and divergence between transcript distributions in the controls confirmed. The identification of transcript distribution patterns allows for the identification of potentially cis-limited genes from this dataset. These results are compared to previous findings.

Good experimental design is an important aspect of gene expression studies. In a well designed experiment, specific changes are made to the input which can subsequently be reliably identified as being the source (or not) of any detectable changes in the results. Replicate data sets are required to determine consistency and significance. Potential undesired sources of variation were identified prior to sample collection, factors such as variation in reagent batches/production, sample preparation techniques, user to user differences, and were avoided as far as possible. Technical variance was accounted for by the sensible use of technical controls. Treatment and control probes are present on the same array chip so that technical noise can be more reliably separated from biological differences on an array to array basis.

In this microarray gene expression study, input variables are the different mouse strains, cell populations and subcellular fractions. The study utilises 3 replicate sets, originating from 4x different individuals per strain, equalling six full data sets, outlined in the flow chart below (*Figure 5.1*).

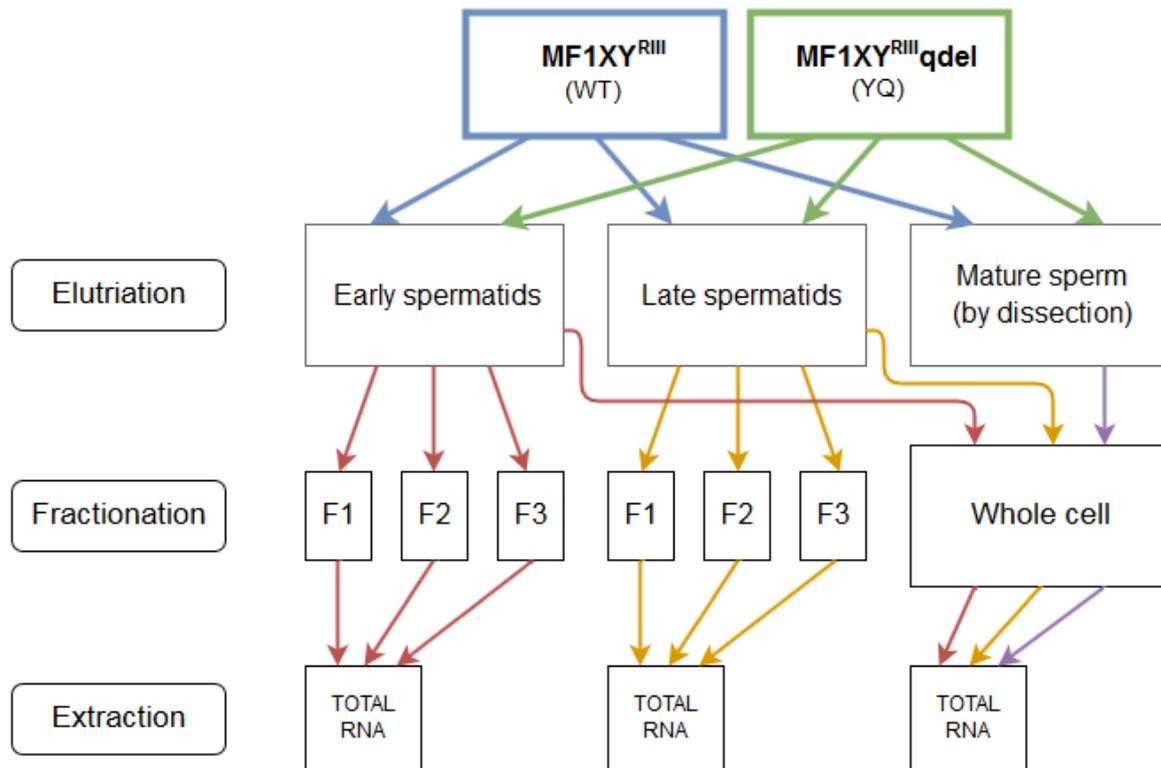


Figure 5.1. Overview of total RNA sample preparation for use in the gene expression study utilising microarray technology

5.1.2. Chapter objectives

The aims of this chapter are as follows:

- O1. To examine the effectiveness of cellular elutriation as a means of segregating and enriching cellular populations for the purpose of transcriptome analysis.
 - Report the level of cell type enrichment attained for each sample population.
- O2. To examine and understand the quality and reliability of the data retrieved from microarray experiments.
- O3. To identify genes that may be linked to physiological differences between MF1XY^{RIII} (WT) and MF1XY^{RIII}qdel (YQ) sperm (Chapter 3) by examining differential expression across all cellular stages.
 - Identify differentially expressed genes across all developmental stages in whole cell samples, between WT and YQ.

- Examine transcript distribution (relative abundance in subcellular fractions) within cell stage populations.
- O4. To identify genes that may be linked to the observed offspring sex ratio distortion.
- Examine differential expression between WT and YQ, to identify candidate genes.
 - Examine the distribution of transcripts across three subcellular fractions (F1, F2, F3) to determine which transcripts show similar patterns of expression to transcripts that are known to be cis-limited.
- O6. To compare and contrast candidates identified through the two major differential expression techniques used in this thesis; RNA sequencing (Chapter 4) and microarray (Chapter 5).
- O7. To consider findings in the context of the literature and the implications for future research.

5.1.3. Notes for Chapter

Sample names

Similarly to Chapter 4, sample names have been abbreviated in this chapter. The following names have been allocated to the combined elutriation and fractionation procedures that took place for each RNA sample collection. The names are used in this chapter when referring to the following procedures: animal collection and extraction; elutriation of whole testis cells; fractionation of elutriated cell collection; examination of cellular enrichment; array data processing and analysis. The abbreviations are listed and explained below (*Table 5.1.* and *Table 5.2*) and an example given. The combined abbreviation describes the sample replicate set, mouse strain from which the sample originated, the cellular population (following elutriation or alternative extraction), and the stage during the cellular fractionation procedure at which the sample was retrieved. In later sections, these abbreviations will also be used to describe results and conclusions drawn from the data collected.

Full Title	Abbreviation
Microarray sample collection, <i>unused</i>	RX
Microarray sample collection, replicate one	R1
Microarray sample collection, replicate two	R2
Microarray sample collection, replicate three	R3

Table 5.1. Abbreviations used to describe RNA sample replicate sets

Mouse strain		Cell type		Cell fraction	
Full	Abbreviated	Full	Abbreviated	Full	Abbreviated
MF1XY ^{RIII}	WT	Early spermatids	E	Whole cell	WC
MF1XY ^{RIII} qdel	YQ	Late spermatids	L	Fraction 1	F1
		Mature sperm	M	Fraction 2	F2
				Fraction 3	F3

Table 5.2. Abbreviations used to describe RNA samples

An example of a combined abbreviation: R2WTLF1 = a sample from replicate set two, MF1XY^{RIII} (wild-type) mouse strain, late spermatid stage, first cellular fraction.

Software

All data preparation, differential expression analysis and production of plots were conducted in R v.3.3.3 (R Core Team 2016) unless otherwise stated. R Scripts for data quality control and data normalisation were originally created by Cambridge Genomics Services (University of Cambridge). These scripts were adapted by the author and analyses repeated. These adapted scripts can be found at the online repository¹⁷. All scripts for the examination of differential expression subsequent to initial analysis and the production of candidate gene lists and associated visualisations were produced by the author for use explicitly with the described microarray RNA dataset. These scripts can also be found at the above stated location.

5.2. Sample Preparation

Figure 5.1 shows the sample preparation workflow for the microarray expression study. Testis cell extract from four individuals of each of MF1XY^{RIII} and MF1XY^{RIII}qdel mice were separately pooled to form WT (MF1XY^{RIII}) and YQ (MF1XY^{RIII}qdel) samples. Mature sperm samples were also taken. The pooled testis extract was then processed by centrifugal elutriation. After retaining a proportion of cells to form a whole cell (WC) sample, each of these populations was then separated into subcellular fractions through exposure to a lysis buffer series. This collection, from extraction to fractionation was repeated twice more using additional sets of animals to generate three replicate sets; fifty-four

¹⁷ Online repository – <http://www.github.com/EmmaEPJ/Repository>

samples in total. RNA was then purified from all fifty-four samples. Details of sample preparation are available in sections 2.4, 2.5 and 2.7.1.

All sample preparation was conducted by one individual (the author), except for cellular elutriation and fractionation where the same two people (Dr Benjamin Skinner and the author) prepared each sample set. Extraction and library preparation was conducted using the same techniques and equipment for each set, as outlined in this chapter.

5.2.1. Results of cellular population enrichment

To determine the level of enrichment of cells achieved in each elutriation run, an amount of cells were collected prior to the fractionation stage for staining and examination under the microscope.

Cells were fixed drop-wise in 4% PFA, and subsequently spun down onto slides. The slides were then incubated with a PNA-lectin stain and DAPI. A number of images were taken of each slide and images were collected using a fluorescence microscope (for full protocol, see sections 2.4.3 and 2.4.4). Examinations were conducted for the following stages: post-trypsin treatment and filtering of testis extract; post elutriation; post sperm extraction.

Following examination by fluorescence microscopy, it became evident that lectin staining had produced a high level of background and would not be useful in cell type identification. This channel was filtered out before images were analysed. In order to accurately group the cells using DAPI alone, images and descriptions from Nakata et al. (2015) and Dolnik et al. (2007) were used for visual reference. Dolnik et al. (2007) also refers to a written reference given in Oakberg (1956). This information was used to form the basis of the cell type attributions generated below (*Figure 5.2*).

Results of cell type enrichments from elutriation (as described in section 2.4.1) are given in *Table 5.3*. Example images of cell composition of each of the populations prior to and following elutriation are given in *Figure 5.3* below.

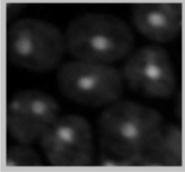
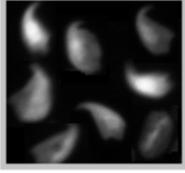
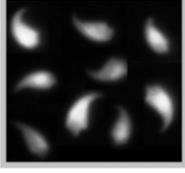
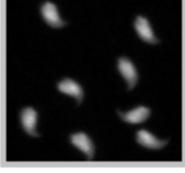
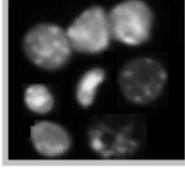
Attributed type	Description	Example
Type 1	Round spermatids	
Type 2	Elongating spermatids	
Type 3	Condensing spermatids	
Type 4	Mature spermatozoa	
Type 5	Other or unknown cell type	

Figure 5.2. Description of cell type attributions used for examination of cell composition. Example images are greyscale images of DAPI stained cells.

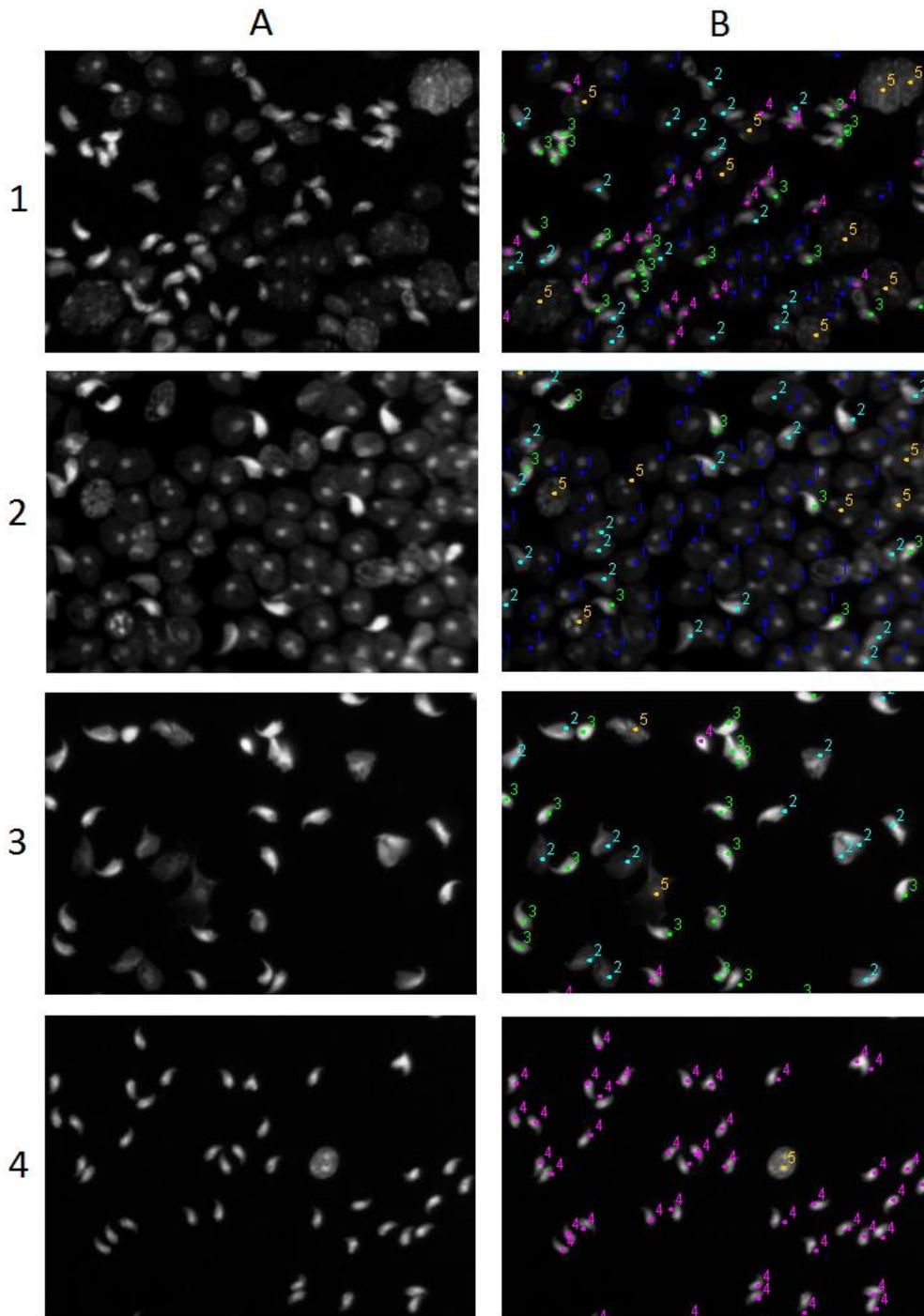


Figure 5.3. Examples of images used to determine the composition of pre- and post-elutriation cell populations. Column A contains selected regions of 20x images taken of DAPI stained cells prior to cell type allocation. Column B contains the same images following manual cell type allocation using ‘Cell Counter’¹⁸ in ImageJ. Cell type allocations are explained above (see *Figure 5.2*). Row 1) Direct testis extract (pre-elutriation); row 2) Early spermatid population (post-elutriation); row 3) Late spermatid population (post-elutriation); row 4) Mature sperm population following direct extraction from vas deferens.

¹⁸ Kurt De Vos, Cell Counter plugin for ImageJ (<https://imagej.nih.gov/ij/plugins/cell-counter.html>)

Following examination of the first microarray sample collection (Microarray sample collection run RX), it was found that poor enrichment had been achieved across all cell types, and that large cell masses, containing multiple cell types were visible. This lead us to believe that new trypsin (cell separation agent) was required before proceeding with further sample collection. Higher level cell type enrichment and the absence of large multi-cell clumps were noted following trypsin renewal. Therefore, the first sample set was discarded and subsequent sample collection sets were taken forward for microarray expression analysis. The following enrichment was attained for subsequent sample collections (*Table 5.3*), assessed as described above.

Percentage representation of cell type per population / %				
	Retrieved cell populations (pre- and post-elutriation)			
Attributed cell type	Extract	Early	Late	Mature (vas deferens)
Type 1	17.14 (1.51)	55.48 (2.03)	1.13 (0.32)	0 (0)
Type 2	19.88 (0.62)	23.08 (2.48)	35.34 (2.98)	0.03 (0.03)
Type 3	32.34 (1.37)	8.28 (1.29)	39.53 (2.77)	0.63 (0.19)
Type 4	14.16 (4.13)	2.44 (0.99)	14.76 (2.23)	93.43 (1.92)
Type 5	16.48 (3.37)	10.72 (1.97)	9.24 (0.83)	5.92 (1.74)

Table 5.3. Cell type attribution and percentage representation of cell type per population for microarray cell samples. Desired cell types are highlighted in blue and are dependent upon population. Secondary cell types, where there may be a small amount of cross-over, are highlighted in a pale pink. Standard error is written in brackets aside the percentage value; n=4 for early, late and mature populations, n=2 for testis extract.

The early spermatid population was enriched from 17.14% round cells to 55.48%. The late spermatid population was enriched from 52.22% elongating and condensing cells to 74.87%. Standard error did not exceed 2.98 for any enriched subtype. Mature sperm samples were extracted directly from vas deferens and were not enriched by elutriation. Mature sperm cells represented 93.43% of the cell population extracted from vas. Attribution of type 1 (round cells) was highly cautious due to the lack of lectin staining, as a result this enrichment is expected to be a minimal estimation of actual enrichment.

5.2.2. Subcellular fractionation of whole cell populations

Following cellular elutriation, a proportion of cells from each population were placed immediately into QIAzol Lysis Reagent (Qiagen, UK) and homogenised. The remaining cells were taken for subcellular fractionation as described in sections 2.4.2 and 4.1.2. Mature sperm were collected by dissection. This resulted in the retrieval of the following total RNA samples from each of the six sets of mice (Strain x 2, Replicate x 3):

EWC Early spermatid whole cell
EF1 Early spermatid fraction one
EF2 Early spermatid fraction two
EF3 Early spermatid fraction three
LWC Late spermatid whole cell
LF1 Late spermatid fraction one
LF2 Late spermatid fraction two
LF3 Late spermatid fraction three
MWC Mature sperm whole cell

This produced 18 RNA samples per replicate set; 54 samples in total.

5.2.3. RNA sample quality control

Low quality RNA can lead to missed or false expression calls, and result in misinterpretation of differential expression data downstream. This is particularly important in clinical diagnosis or clinical research settings, where RNA levels may be used as a diagnostic tool for potentially life-threatening conditions such as cancer (Vermuelen et al. 2011). In order to ensure good and consistent RNA quality, RNA sample quality control (QC) was performed on all samples prior to microarray analysis as described in section 2.5.2.

All samples returned an RNA integrity (RIN) value greater than 5.5, with only 10/54 samples with a RIN of below 7. Ribosomal RNA constitutes 80%+ of all cellular RNA (Wilhelm and Landry 2009, Lodish et al. 2000), and RIN is calculated using rRNA abundance as a base line and proxy for quality of other RNAs. These samples likely returned a lower RIN value as a result of the unusual rRNA levels expected (all ten samples were spermatid fraction 3 or mature sperm samples). All other measured parameters appeared normal. Quantitative PCR was used to determine RNA sample concentration prior to library preparation. This was performed by Cambridge Genomics Services (Department of Pathology, University of Cambridge).

Sample processing

All RNA samples successfully passed QC. Samples underwent library preparation and were run on the Affymetrix GeneChip™ Mouse Gene 2.1 ST Arrays, processed on the GeneTitan as described in section 2.7.1. RNA sample QC and library preparation was performed by Cambridge Genomic Services.

5.3. Normalisation and Data Treatment

In order for gene expression comparisons to be made between samples, the data must first be assessed for sample and hybridisation quality, normalised across samples and organised in such a way as to make the information accessible and informative to the user.

5.3.1. Data preparation

RNA samples were hybridised to the array and scanned as described in section 2.7.1. After scanning, the files generated by the scanner (CEL files) were loaded in R (R Core Team 2016) using the Oligo package from Bioconductor (Carvahlo et al. 2010). No background correction or normalisation was applied at this stage.

To assess the quality of the data, plots of control probes were generated along with MAplots, fluorescence intensity distribution plots, and other examinations of experimental data to allow visualisation of the sample hybridisation information. Results and explanation of technical controls can be seen below (section 5.3.2). Variation between biological replicates was investigated using clustering methods. The raw data were then processed using the Robust Multiarray Analysis (RMA) method (Irizarry et al. 2003). Once the data had been processed as above, comparisons were performed using the R Limma package and the results adjusted for multiple testing error using False Discovery Rate (FDR) (Benjamini and Hochberg 1995). The package version and R session information (R version 3.3.3) are provided in the reference section (R Core Team 2016). An overview of the array quality control workflow can be seen in *Figure 5.4*.

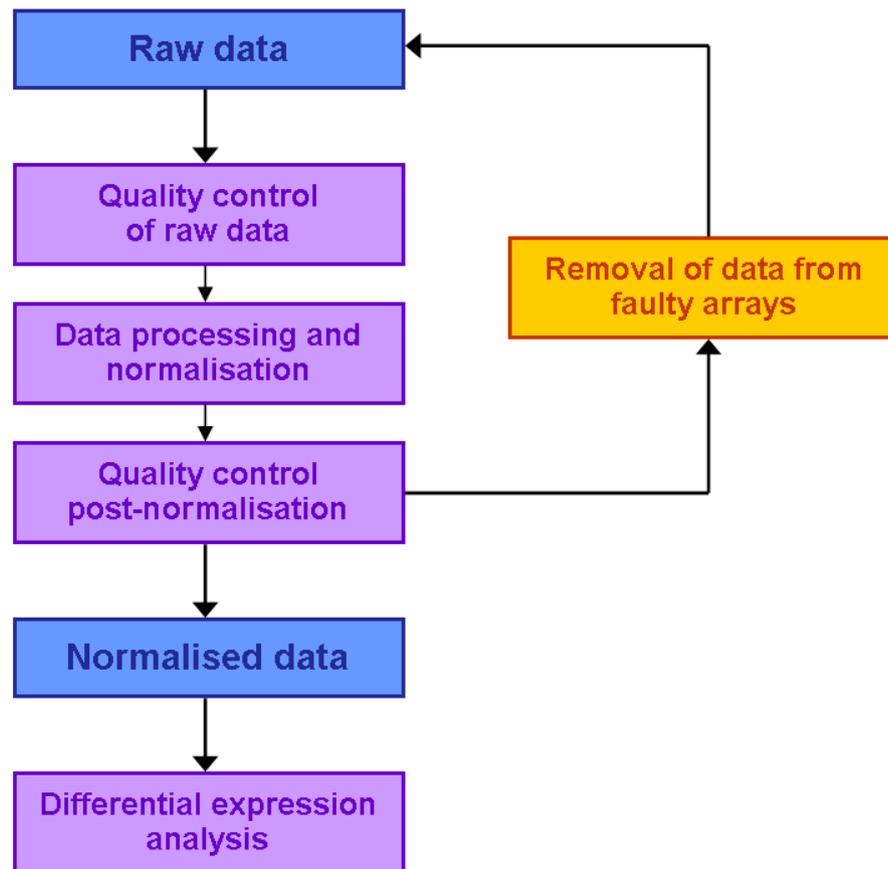


Figure 5.4. Array quality control workflow. Based on image from arrayanalysis.org

5.3.2. Examination of technical controls

A broad set of technical quality control assessments are performed to examine array to array differences in information capture and ensure that the data retrieved is of suitable quality for subsequent differential expression analysis. The Affymetrix arrays contain a number of specific integrated control probe sets for the examination of variation caused by technical noise and highlight potential problems encountered during array hybridisation. This allows for variation to be correctly attributed to either technical or other differences and provides insight for the appropriate application of normalisation methods. A technical QC report was produced by Cambridge Genomics Services; plots were automatically generated during array analysis. Array analysis and associated plots were subsequently edited and reproduced in R by the author.

None of the controls indicated hybridisation failure of any sample. All samples passed QC and were therefore deemed suitable for use in differential expression analysis.

5.3.3. Normalisation of microarray data: Raw data (pre-normalisation) and post-normalisation quality control

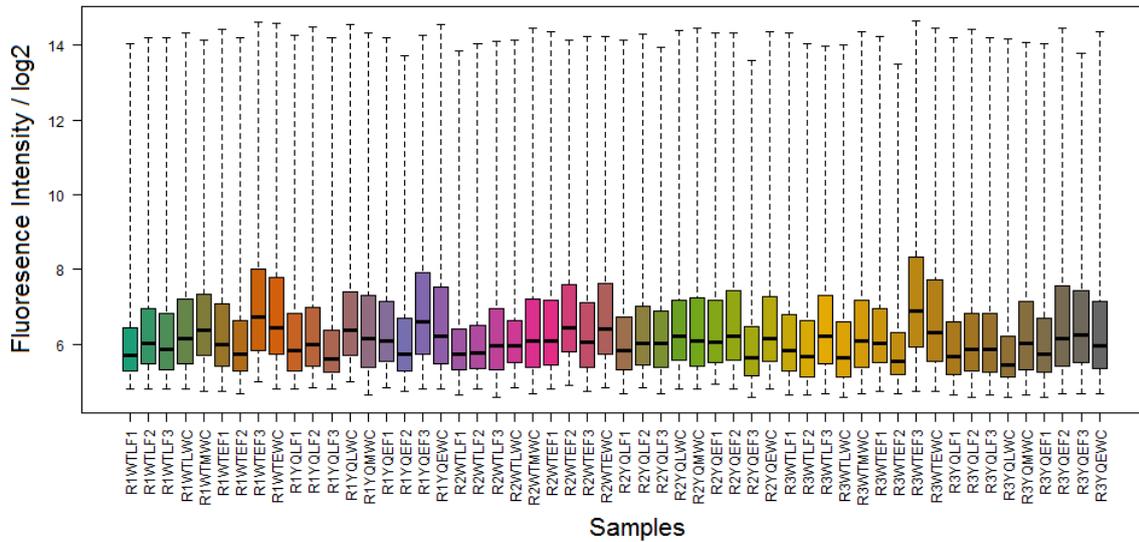
Normalisation is required for the elimination of variance caused by technical noise as quantified by technical QC steps, described above, and to allow the comparison of array data between samples by standardising the distribution of fluorescence intensities recorded across all samples. Variance (except for that caused by inappropriate sample preparation) can then be more reliably attributed to real biological variance between the samples. A pre- and post-normalisation QC are performed to assess the effects of normalisation on the experimental data. Some illustrative examples are given below.

The Robust Multiarray Analysis (RMA) method allows for the generation of a summary of intensities of all probes covering a gene, constructing a model containing a "probe" and a "chip" effect. This model uses all arrays available to give levels of normalisation appropriate across the whole experiment; including a greater number of arrays increases robustness when calculating normalisation to be applied. The probes are filtered by intensity level; failing to filter low expressing probes when performing the comparisons would have a detrimental effect on power as it increases the stringency of the False Discovery Rate (FDR) multiple testing correction. As a result of filtering, high levels of genes are recorded as non-expressing; non-normal distribution of probe intensities means that the median is the best estimator of the overall intensity of probes of interest.

Probe fluorescence intensity (intensity boxplots)

Boxplots of probe fluorescence intensity (*Figure 5.5.a* and *Figure 5.5.b*) are used for between-array comparison. Samples should display similar distribution intensities following normalisation.

Fluorescence Intensity Across All Samples - Raw Data (Pre-Normalisation)



Fluorescence Intensity Across All Samples - Post-Normalisation

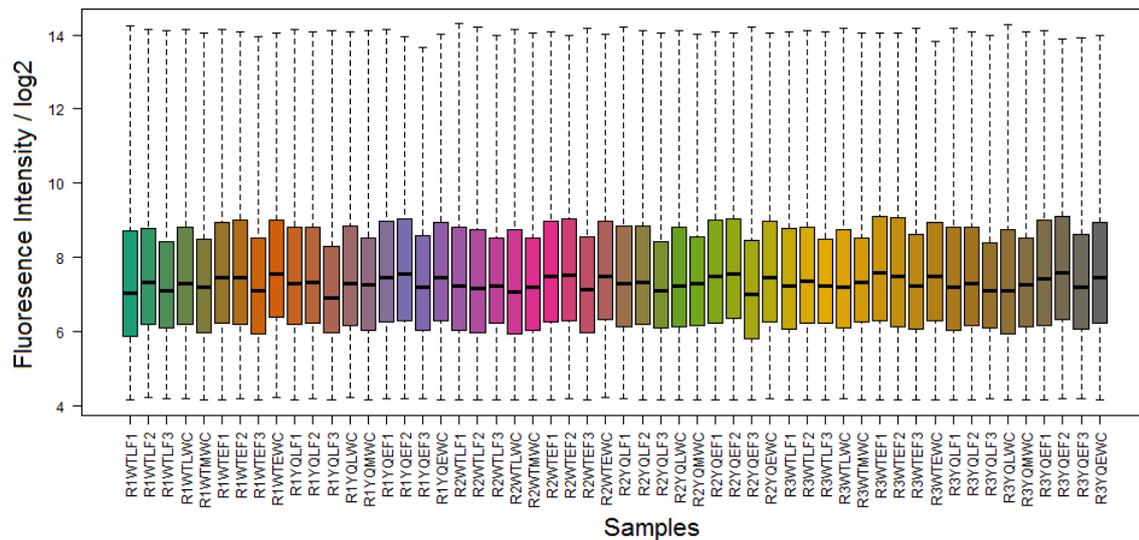


Figure 5.5. Boxplots of probe fluorescence intensity across all samples a) pre-normalisation (top) and b) post-normalisation (bottom). The top, middle, and bottom of the boxes represent the 75th percentile, median, and 25th percentile respectively. The intensity scale is log base 2. A larger version of these figures is included in Supplementary Material.

Other pre- and post-normalisation measures of intensity, density and per-sample group intensity distribution were conducted to compare samples in more detail. Here the raw data per sample and sample group was evaluated on MA and density plots, and re-plotted following normalisation. Replicates are expected to show limited variance post-normalisation and match tightly to the median line generated across the three replicates. All samples showed a median value of zero, and only one

sample, R3WTEF1, showed an inter-quartile range (IQR) variance value of above zero (0.024). The reason for this is unknown, but as the sample shows little variance from the median, and other measures showed quality to be high, there was no good or sufficient reason to reject the sample and it was included in downstream analysis. Data clustering was visualised using principal components analysis (PCA) plots (section 5.4.1) where samples can be seen to cluster primarily according to cell population. This indicates the variance between cell population types is greater than between mouse strains, which was expected. PCA provides a good opportunity to identify any sample batch effects; this is discussed in more detail in section 4.3.4. No samples were rejected as a result of failure to pass QC.

5.4. Overall Trends in Microarray Data

5.4.1. Principal Components Analysis (PCA) – clustering analysis

Principal components analysis (PCA) is a statistical method used for exploring the sources of variation in a complex, multifaceted dataset. A detailed explanation of PCA can be found in section 4.3.4.

Across all samples, the greatest source of variance is cell population. Clustering by fraction is seen within cell type populations, with fraction 3 (F3) appearing most distinct. Late stage spermatid F3 (nuclear/organelle bound fraction) samples appear clustered between mature sperm cells and late spermatid samples. Early stage spermatids, where the highest level of active transcription is expected, show the greatest variance between samples and most distinct profiles between fractions as a result. Early and late spermatid population whole cell samples lie within distributions of F1, F2 and F3 samples; as whole cells represent the combination of the three fractions per cell population, this is expected. Minimal clustering by replicate set is seen, and this is reduced further post-normalisation.

By separating samples into cell populations for additional PCA analysis, the greatest source of variance is excluded and greater resolution within populations can be observed. Distinct clustering by fractions is seen within the early spermatid population both pre- normalisation (*Figure 5.7*) and post-normalisation (*Figure 5.8*). Fraction-type clusters can also be seen in late spermatid groups post-normalisation. This suggests the fractionation process was consistent across replicates in both cell populations. No consistent clustering by replicate set is noted, suggesting the absence of sample preparation batch effects. This is later upheld when samples are examined alongside their different-strain counterparts (see Supplementary Material).

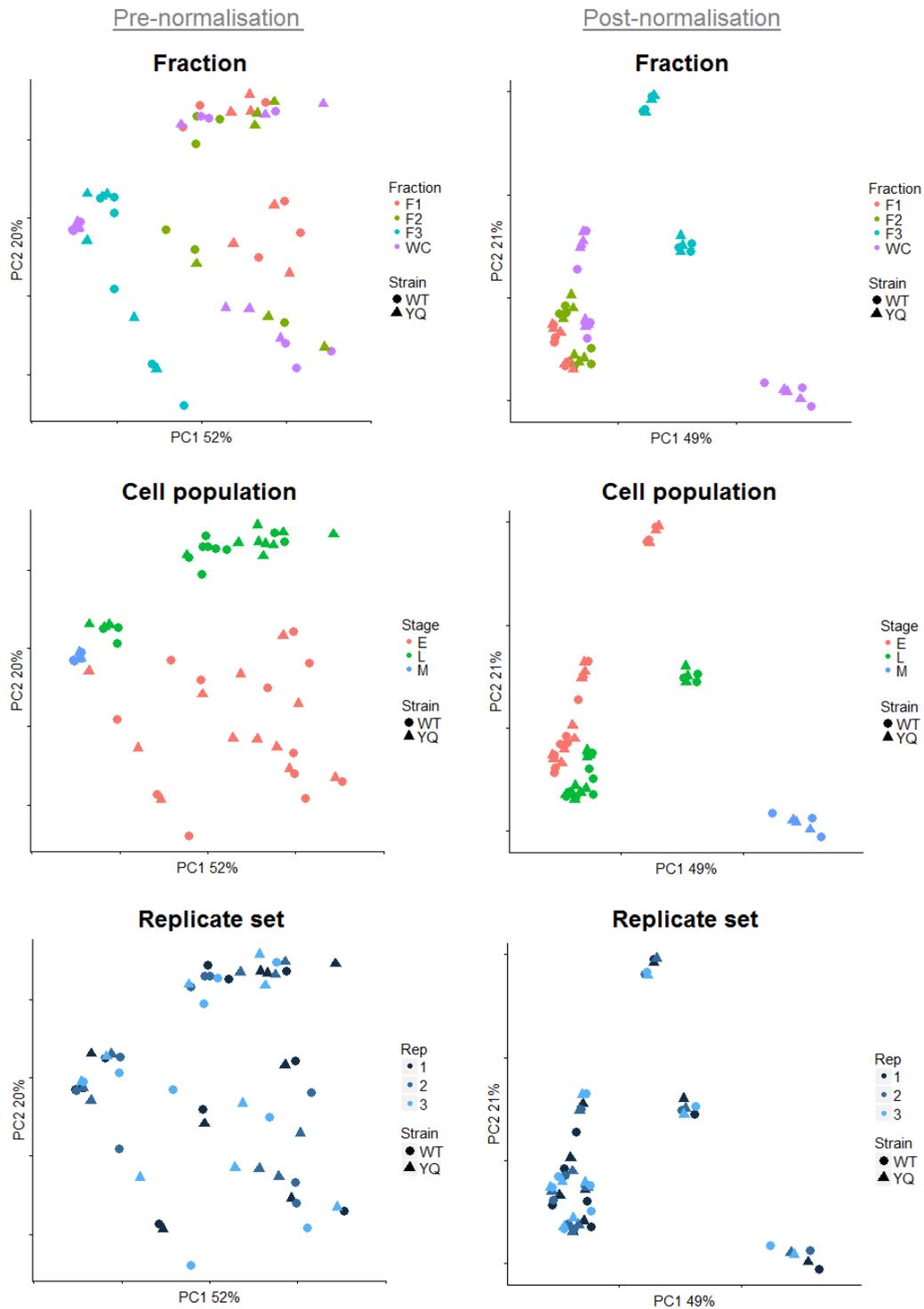


Figure 5.6. PCA plots containing all samples, pre-normalisation (left-hand column) and post-normalisation (right-hand column). Samples have been highlighted as follows; subcellular fraction (row 1), cell population (row 2), replicate set (row 3). Mouse strain from which samples originate is indicated by marker shape. Both pre- and post-normalisation shows samples clustering by cell type and subcellular fraction. These clusters are tighter following the removal of technical noise and normalisation of samples (post-normalisation).

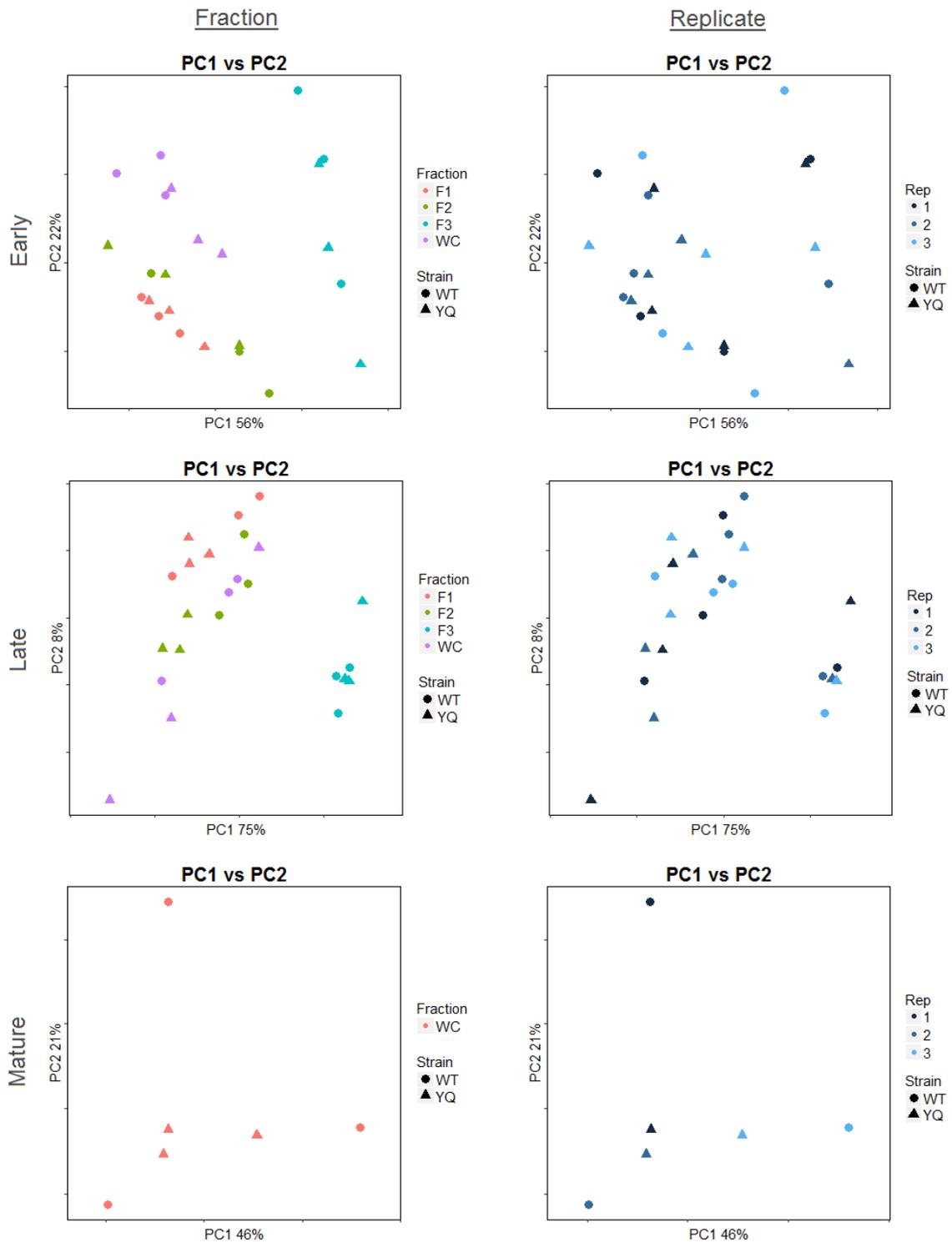


Figure 5.7. PCA plots of samples pre-normalisation. Samples have been segregated by cell population; early spermatid (row 1), late spermatid (row 2), mature sperm (row 3). Plots in the ‘fraction’ column highlight cell fractions (marker colour) within populations alongside the mouse strain from which they originate (marker shape). Plots in the ‘replicate’ column show the same comparisons, but with marker colour representing replicate set number. Spermatid samples already show strong separation by subcellular fraction within cell populations.

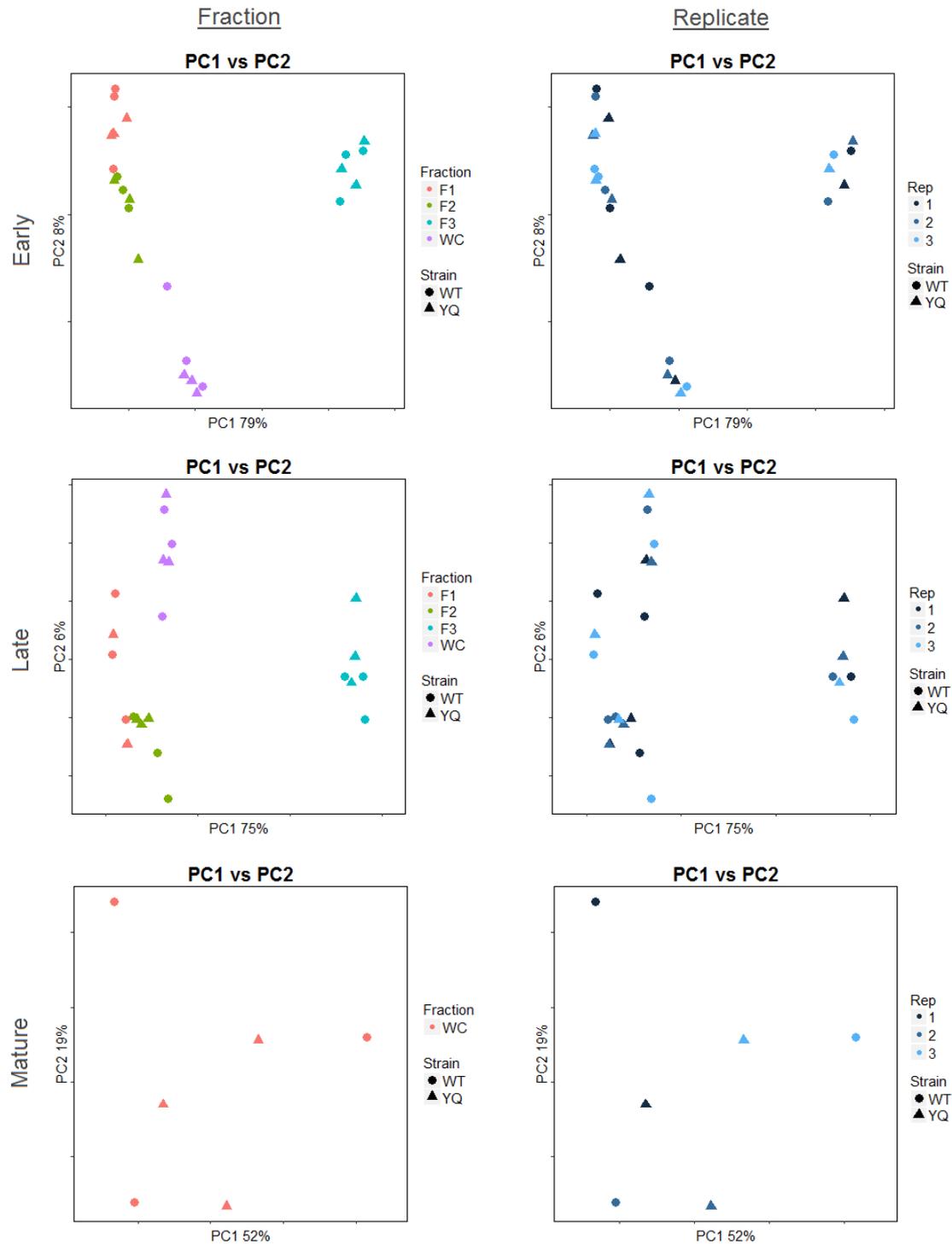


Figure 5.8. PCA plots of samples post-normalisation. Samples have been segregated by cell population; early spermatid (row 1), late spermatid (row 2), mature sperm (row 3). Plots in the ‘fraction’ column highlight cell fractions (marker colour) within populations alongside the mouse strain from which they originate (marker shape). Plots in the ‘replicate’ column show the same comparisons, but with marker colour representing replicate set number. A strong separation by subcellular fraction type can be seen within spermatid populations.

PCA reveals a separation of all samples by strain with primary and secondary principal components (accounting for greatest variance), except in mature sperm samples. Plots showing major variance between WT and YQ samples of the same type were also produced and can be found in the Supplementary Material. These plots showed that overall samples are clustered more tightly between strains than between replicates.

In the above plots, principal components 1 and 2 are shown only, as further components represented only a relatively small amount of variation (between 3% and 9%) and provided little further information when displayed graphically.

5.4.2. Upregulation of sex-linked genes

As anticipated, a global upregulation of X-linked genes and remaining Y-linked genes, and lower expression of reduced copy number Y-linked genes, is seen in the deletion model (YQ) (*Figure 5.9*). This up-regulation of remaining sex-linked genes is also seen across all fractions, but is not evident across the few genes that are detected in mature sperm samples, and is less distinct in F2 and F3 of late spermatid samples. While most autosomal genes appear largely unchanged, there are exceptions.

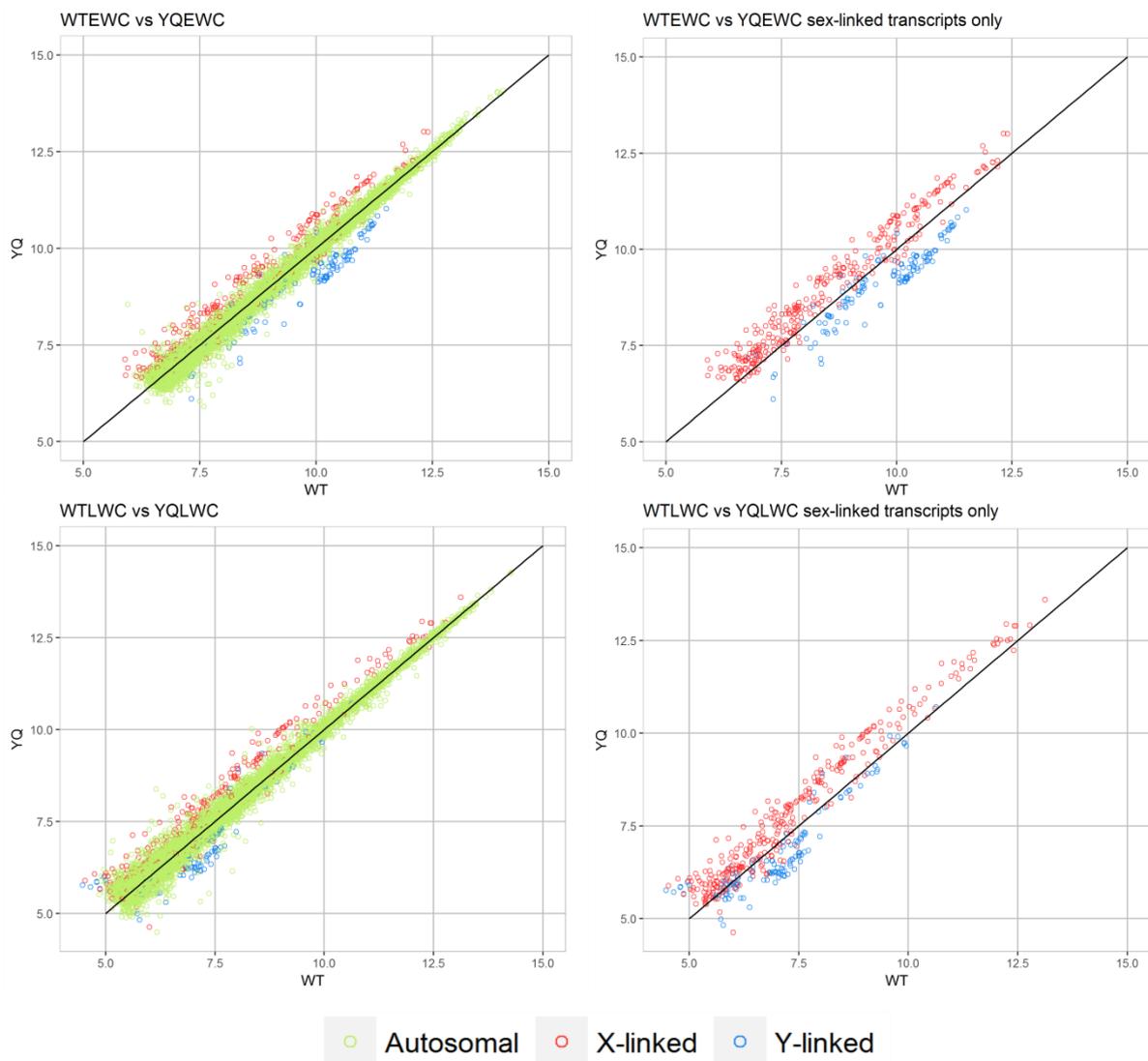


Figure 5.9. Scatterplots showing up-regulation of X-linked (red) genes and remaining Y-linked (blue) genes, and reduced expression of reduced copy number Y-linked (blue) genes in the deletion model (YQ) whole cell spermatid samples. Average array expression values for individual genes across all replicates are plotted.

5.5. Results of Differential Expression Analysis of MF1XY^{RIII} and MF1XY^{RIII}qdel Spermatid RNA Samples

Differential expression analysis was performed using the Limma package (Ritchie et al. 2015) in R. Matched comparisons were made between the samples e.g. (WTEWC vs YQEWC). This produced log fold change values and p-values for all detected genes within the comparison being made. FDR multiple testing correction was applied to calculate adjusted p-values. All replicates were included in the listed comparisons. Subsequent filtering by adjusted p-value of <0.05 to retain only genes with significant differential expression between the two sample sets yielded gene lists of varying length for each comparison (*Table 5.4* and *Table 5.5*). The arrays included all candidate genes of interest from the literature; these are discussed in section 5.5.5.

Differential expression across the entire genome

Differential expression was examined at high stringency. That is, a high minimum expression value was selected for genes to be carried forward. This expression threshold was set at 75%; the default threshold is 50%. The high number of hypothesis tests conducted in a microarray expression study warrants the use of multiple testing adjustments such as described in Benjamini and Hochberg (1995). However, in instances where the number of differentially expressed genes is expected to be limited to a relatively small proportion of examined genes, the application of multiple testing corrections can result in severely reduced power when testing for significance in differential expression.

The filtering of detected genes by fluorescence intensity at an increased stringency means that fewer hypotheses are tested (the number of genes examined is reduced) and as a result, significance testing power is increased (Bourgon et al. 2010, Hackstedt and Hess 2009). Hackstedt and Hess (2009) showed using Affy expression data that by filtering for variance the investigator is able to increase power while still controlling false discovery rate when the total number of differentially expressed genes was low.

In this dataset, it is expected that the transcriptome will remain largely unaltered between WT and YQ strains except for a number of sex-linked genes (dominated by multicopy X- and Y-linked transcripts) and few autosomal transcripts. The application of higher stringency filtering resulted in the reduction of the number of tests conducted from between 15851 and 18810 genes using standard filtering to between 8004 and 11097 genes per matched comparison. In direct strain to strain comparisons, this increased the number of significantly differentially expressed genes from 47 to 367 across all comparisons (301 of which were sex-linked). At standard stringency, differentially expressed genes

across matched comparisons consisted almost entirely of down-regulated Y-linked genes such as *Ssty2*, not unexpected because of the substantial Yq deletion. This additional stringency is not appropriate for the examination of between fraction comparisons where high numbers of genes are expected to be differentially expressed. In this instance (comparison SET 2), the default level of stringency was applied. Comparison set 2 was used exclusively for the examination of transcript distribution, and was not a primary source of candidate selection.

Table 5.4. COMPARISON SET 1 – differential expression between WT and YQ samples across all stages and fractions at high (75%) stringency

Sample A	Sample B	DE transcript clusters (at p<0.05)	No. of genes compared
WTEF1	YQEF1	1	9386
WTEF2	YQEF2	72	9341
WTEF3	YQEF3	37	9082
WTEWC	YQEWC	93	9294
WTLF1	YQLF1	0	9311
WTLF2	YQLF2	0	9349
WTLF3	YQLF3	22	9119
WTLWC	YQLWC	142	9319
WTMWC	YQMWC	0	8004

Table 5.5. COMPARISON SET 2 – differential expression between fractions and within strains at standard (50%) stringency

Sample A	Sample B	DE transcript clusters (at p<0.05)	No. of genes compared
WTEF1	WTEF2	279	18076
WTEF1	WTEF3	13238	18689
WTEF2	WTEF3	12913	18470
WTEF3	WTEWC	11844	18095
WTLF1	WTLF2	89	18150
WTLF1	WTLF3	11772	18302
WTLF2	WTLF3	11078	18261
WTLF3	WTLWC	11307	18130

YQEF1	YQEF2	260	18210
YQEF1	YQEF3	12535	18513
YQEF2	YQEF3	12661	18526
YQEF3	YQEW C	11361	17931
YQLF1	YQLF2	7	18303
YQLF1	YQLF3	11369	18424
YQLF2	YQLF3	11531	18435
YQLF3	YQLW C	10409	18169

Tables 5.4 and 5.5. Tables showing the number of genes classed as differentially expressed per comparison.

5.5.1. Examination of control transcripts

Transcripts of known behaviour

Transcript distributions of genes of known sharing status were examined using logFC values between fractions within each spermatid population. *Table 5.6* describes how we might expect the genes to appear, given information from the literature and previous experimental work. *Figure 5.10* shows the relative distribution (expression values) of these control transcripts, comparing one cellular fraction (sample A) to another fraction (sample B) within the same cell population using the log fold change (logFC) value of comparative expression.

Gene symbol	Behaviour	Likely enriched	Function	Reference
<i>Akap4</i>	Shared	F1	Fibrous sheath formation	Morales et al. 2002
<i>Pm1</i>	Shared	F1	Sperm protamine (histone replacement)	Caldwell and Handell 1991
<i>Smok2b</i>	Not-shared	F3, some F2	Sperm motility kinase	Previous lab work and Veyron et al. 2009 (see section 1.3.3)
<i>Spam1</i>	Not-shared or reduced sharing	F2, some F3	Sperm head hyaluronidase	Martin-DeLeon et al. 2005

Table 5.6. Table of genes of known shared or non-shared behaviour. The table includes expected distribution of transcripts across subcellular fractions and associated evidence for sharing status and expected distributions.

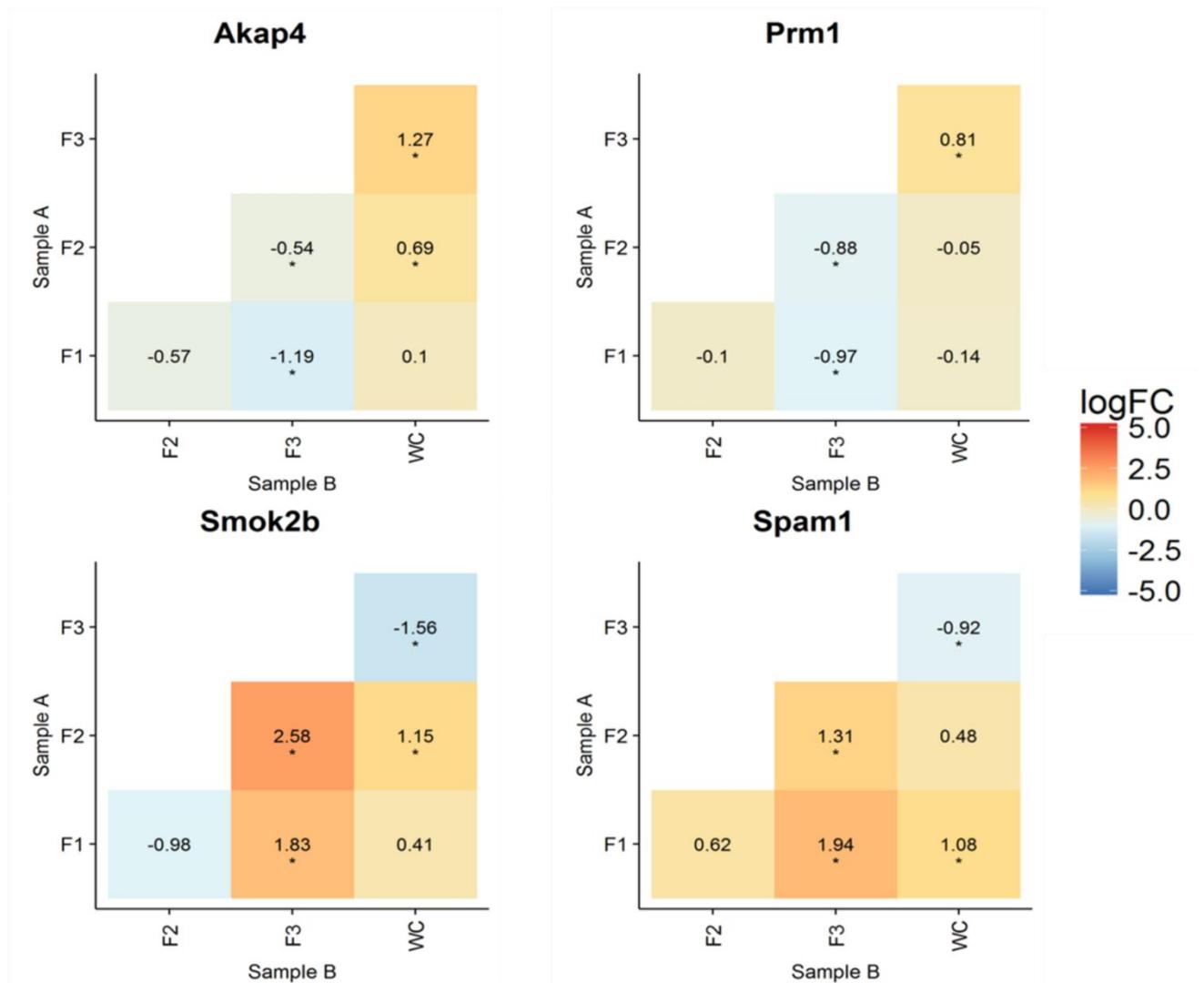


Figure 5.10. Heatmaps of log fold change (logFC) values of expression between fractionated samples in the same wild-type early spermatid cell populations. Each heatmap shows logFC values for an individual gene. * Significantly differentially expressed values at 50% stringency. Heatmaps are interpreted as follows; for instance in *Smok2b*, a logFC value of 2.58 is given in a comparison between F2 and F3. This shows a 2.58 logFC value increase in expression in F3 compared with F2 in this sample.

In mice, *Akap4* and *Prm1* produce transcripts that are abundant and known to be shared across cytoplasmic bridges between spermatids (section 1.3.3). The transcripts of both genes appear enriched in the first (unbound) subcellular fraction (F1) and in whole cells when compared to F2 and F3 (bound fractions) in the data, as expected.

Smok2b produces RNA transcripts that are known to remain not-shared between developing spermatids. *Spam1* is haploid expressed and Zheng et al. (2001a) provided support for the suggestion that the mRNA and protein remain compartmentalised within the cells and are not shared between

sister spermatids. Both of these genes show patterns of enrichment in the third fraction in our microarray data, which is especially strong in *Smok2b*. *Spam1* shows enrichment both in F3 and mild enrichment in F2. Comparisons F1 vs F3, F2 vs F3 and F3 vs WC showed significant and consistent differences between shared and non-shared control genes, suggesting enrichment in F3 as the major indicator of cis-limitation in this dataset.

5.5.2. Results of matched comparisons between strains – MF1XY^{RIII} (WT) vs MF1XY^{RIII}qdel (YQ)

The following tables list genes that were shown to be differentially expressed (p.value >0.05) in the comparisons stated. Genes are listed in order of logFC value. Where more informative gene symbol synonyms have been subsequently found, these are included in brackets.

Significant differential expression in early spermatids (p=0.05)

Table 5.7.1. WTEF1 vs YQEF1			
Transcript ID	Gene symbol	Chromosome	logFC
17226736	<i>Mapkapk2</i>	1	-1.84

Table 5.7.2. WTEF2 vs YQEF2							
Transcript ID	Gene symbol	Chr	logFC	Transcript ID	Gene symbol	Chr	logFC
17444498	<i>Nptx2</i>	5	1.65	17547394	<i>Gm20854</i>	Y	-1.19
17489923	<i>Gm21115</i>	7	1.37	17547464	<i>Gm20738</i>	Y	-1.19
17489943	<i>Gm21136</i>	7	1.35	17547385	<i>LOC101056308</i>	Y	-1.25
17489959	<i>Gm5591</i>	7	1.13	17547402	<i>Gm21799</i>	Y	-1.26
17489968	<i>Gm21129</i>	7	1.13	17547283	<i>Gm21180</i>	Y	-1.28
17286046	<i>4932702P03Rik</i>	13	1.12	17546457	<i>Gm20907</i>	Y	-1.3
17489962	<i>Gm1988</i>	7	1.09	17547231	<i>Gm20854</i>	Y	-1.34
17442173	<i>A930024E05Rik</i>	5	1.01	17546465	<i>Gm20879</i>	Y	-1.35
17338595	<i>4932415M13Rik</i>	17	0.81	17547360	<i>Gm20865</i>	Y	-1.47
17520116	<i>Gm10634</i>	9	0.78	17547277	<i>Gm20867</i>	Y	-1.52
17534730	<i>Fam122c</i>	X	0.71	17464312	<i>4930479D17Rik</i>	6	-1.84

17546633	<i>Gm20879</i>	Y	-1.05	A further 46 transcript clusters annotated as <i>Ssty2</i> showed logFC values ranging from -0.87 to -1.5.
17547380	<i>Gm20868</i>	Y	-1.13	
17546452	<i>Gm20815</i>	Y	-1.16	
17547261	<i>Gm21943</i>	Y	-1.19	

Table 5.7.3. WTEF3 vs YQEF3

Transcript ID	Gene symbol	Chr	logFC	Transcript ID	Gene symbol	Chr	logFC
17541140	<i>Btg1-ps2</i>	X	1.16	17547413	<i>Gm20842</i>	Y	-0.97
17542897	<i>Cldn34-ps</i>	X	1.02	17547402	<i>Gm20879</i>	Y	-1
17544385	<i>Tspan6</i>	X	0.94	17547291	<i>Gm20809</i>	Y	-1.09
17537527	<i>4921511C20Rik</i>	X	0.93	17547360	<i>Gm20867</i>	Y	-1.1
17543767	<i>4930519F16Rik</i>	X	0.92	17547172	NA	Y	-1.27
17534425	<i>Stag2</i>	X	0.87	17547238	NA	Y	-1.46
17533539	<i>Rp2</i>	X	0.86	17547454	NA	Y	-1.63
17543562	<i>Gm614</i>	X	0.83	17547182	NA	Y	-1.91
17533604	<i>Uba1</i>	X	0.77	17546570	NA	Y	-2.11
17230327	NA	1	-0.72	A further 18 transcript clusters annotated as <i>Ssty2</i> showed logFC values ranging from -0.87 to -1.13.			

Table 5.7.4. WTEWC vs YQEWC

Transcript ID	Gene symbol	Chr	logFC	Transcript ID	Gene symbol	Chr	logFC
17409804	<i>1700061117Rik</i>	3	1.57	17535175	<i>Ctag2</i>	X	0.64
17544273	<i>Tgif2lx1</i>	X	1.23	17548606	<i>Gm9271</i>	7	0.62
17533946	<i>Btbd35f5</i>	X	1.21	17286805	NA	13	-0.55
17540370	<i>Ppp1r2-ps9</i>	X	1.12	17395345	<i>Gm14419</i>	2	-0.59
17537347	<i>Tgif2lx1</i>	X	0.96	17446757	NA	5	-0.63
17533539	<i>Rp2</i>	X	0.89	17547203	<i>Gm20747</i>	Y	-0.7
17542897	<i>Cldn34-ps</i>	X	0.87	17546452	<i>Gm21180</i>	Y	-0.72

17534493	<i>Actrt1</i>	X	0.87	17547464	<i>Gm20738</i>	Y	-0.72
17548883	<i>4930453H23Rik</i>	X	0.85	17546465	<i>Gm21292</i>	Y	-0.74
17489940	<i>Gm21129</i>	7	0.84	17547191	<i>Gm20747</i>	Y	-0.74
17489959	<i>Gm29258</i>	7	0.83	17546471	NA	Y	-0.82
17489968	<i>Gm1988</i>	7	0.83	17546457	<i>Gm20877</i>	Y	-0.84
17536450	<i>Yipf6</i>	X	0.82	17547380	<i>Gm20868</i>	Y	-0.84
17538194	<i>Vsig1</i>	X	0.78	17547392	<i>Gm20738</i>	Y	-0.89
17541140	<i>Btg1-ps2</i>	X	0.78	17547256	<i>Gm20934</i>	Y	-0.92
17489923	<i>Gm4454</i>	7	0.76	17547428	<i>Gm20747</i>	Y	-0.94
17489943	<i>Gm21136</i>	7	0.75	17547283	<i>Gm20815</i>	Y	-0.95
17545514	<i>Gm15143</i>	X	0.71	17547124	NA	Y	-0.98
17542888	<i>Cldn34d</i>	X	0.7	17547360	<i>Gm20865</i>	Y	-0.99
17303125	NA	14	0.7	17547277	<i>Gm21180</i>	Y	-1.02
17545526	<i>Gm16445</i>	X	0.69	17547291	<i>Gm29024</i>	Y	-1.12
17538320	<i>Pak3</i>	X	0.68	17547454	NA	Y	-1.27
17540431	<i>Gm5123</i>	X	0.68	17464312	<i>4930479D17Rik</i>	6	-1.33
17416196	<i>C8b</i>	4	0.68	17226736	<i>Mapkapk2</i>	1	-1.58
17541041	<i>Gm9</i>	X	0.66	A further 44 transcript clusters annotated as <i>Ssty2</i> showed logFC values ranging from -0.74 to -1.20.			

Significant differential expression in late spermatids (p=0.05)

No significant differential expression detected in WTLF1 vs YQLF1 or WTLF2 vs YQLF2.

Table 5.7.5. WTLF3 vs YQLF3

Transcript ID	Gene symbol	Chr	logFC	Transcript ID	Gene symbol	Chr	logFC
17540370	<i>Ppp1r2-ps9</i>	X	1.69	17469607	<i>5031434C07Rik</i>	6	-0.87
17459095	NA	6	1.59	17410492	<i>Tet2</i>	3	-0.94
17532627	NA	M	1.36	17338231	NA	17	-0.99
17541140	<i>Btg1-ps2</i>	X	1.36	17547464	<i>Gm20898</i>	Y	-1

17533469	<i>Cypt8</i>	X	1.31	17547277	<i>Gm20809</i>	Y	-1.04
17537212	<i>Cytc1</i>	X	1.2	17547360	<i>Gm20823</i>	Y	-1.06
17541134	<i>Btg1-ps1</i>	X	1.19	17547283	<i>Gm20815</i>	Y	-1.07
17482924	NA	7	1.13	17546452	<i>Gm21180</i>	Y	-1.09
17335540	<i>Pim1</i>	17	1.1	17487211	<i>Fosb</i>	7	-1.16
17398719	NA	3	1.07	A further 3 transcript clusters annotated as <i>Ssty2</i> showed logFC values of -0.98.			

Table 5.7.6. WTLWC vs YQLWC

Transcript ID	Gene symbol	Chr	logFC	Transcript ID	Gene symbol	Chr	logFC
17444498	<i>Nptx2</i>	5	2.23	17325648	<i>Gm15802</i>	16	0.86
17532692	<i>LOC100861738</i>	Un*	1.54	17535175	<i>Ctag2</i>	X	0.84
17539750	<i>Btbd35f16</i>	X	1.51	17511018	NA	8	0.84
17339462	<i>4930471L23Rik</i>	17	1.47	17257819	<i>1700096J18Rik</i>	11	0.84
17545287	<i>Gm15104</i>	X	1.41	17533230	<i>H2al3</i>	X	0.84
17532714	<i>Btbd35f17</i>	X	1.39	17538320	<i>Pak3</i>	X	0.82
17533234	<i>Sytl5</i>	X	1.37	17545420	<i>Ribc1</i>	X	0.82
17537324	<i>Ube2dn11</i>	X	1.33	17245374	<i>Grip1os1</i>	10	0.81
17409804	<i>1700061I17Rik</i>	3	1.32	17246033	<i>Gm16230</i>	10	0.81
17534493	<i>Actrt1</i>	X	1.3	17352018	<i>4930465K10Rik</i>	18	0.81
17482924	NA	7	1.29	17285488	<i>Amph</i>	13	0.81
17237687	<i>4921513I03Rik</i>	10	1.29	17538461	<i>Ott</i>	X	0.8
17533946	<i>Btbd35f5</i>	X	1.26	17540203	<i>Rpgr</i>	X	0.78
17533195	<i>4930402K13Rik</i>	X	1.24	17538470	<i>Ott</i>	X	0.77
17541041	<i>Gm9</i>	X	1.14	17545785	<i>Cdkl5</i>	X	0.77
17544668	<i>Tcp11x2</i>	X	1.12	17546434	<i>Rbm31y</i>	Y	0.77
17258007	<i>1700092K14Rik</i>	11	1.12	17318083	<i>Ly6a</i>	15	0.77
17533469	<i>Cypt8</i>	X	1.11	17544102	<i>Brwd3</i>	X	0.76
17546355	<i>Rbmy</i>	Y	1.07	17534767	<i>Gm10488</i>	X	0.75

17539023	<i>Cypt3</i>	X	1.07	17536436	<i>Ar</i>	X	0.74
17546953	<i>Rbmy</i>	Y	1.07	17533858	<i>Gm14632</i>	X	0.73
17536450	<i>Yipf6</i>	X	1.06	17540846	<i>Gm10147</i>	X	0.73
17343823	<i>Gm20513</i>	17	1.05	17363572	<i>Tjp2</i>	19	0.72
17541190	<i>Cul4b</i>	X	1.05	17462705	<i>Foxj2</i>	6	0.71
17533937	<i>Btbd35f2</i>	X	1.05	17326944	<i>Gm34826</i>	16	0.7
17546333	<i>Rbmy</i>	Y	1.05	17533965	<i>Gm1993</i>	X	0.7
17542239	<i>Cetn2</i>	X	1.04	17540793	<i>Gm14525</i>	X	0.7
17416196	<i>C8b</i>	4	1.03	17367157	<i>Gm13266</i>	2	0.69
17546963	<i>Rbmy</i>	Y	1.02	17540766	<i>Gm10147</i>	X	0.68
17548883	<i>Gm20153</i>	X	1.02	17540838	<i>Gm4836</i>	X	0.68
17534486	<i>Tex13c1</i>	X	1.02	17538194	<i>Vsig1</i>	X	0.64
17540370	<i>Ppp1r2-ps9</i>	X	1.01	17412202	<i>Faxc</i>	4	0.64
17539548	<i>Piga</i>	X	1.01	17537163	<i>Fam46d</i>	X	0.63
17539754	<i>Btbd35f6</i>	X	1	17214025	NA (<i>Gm25360</i>)	1	0.63
17540885	<i>Btbd35f7</i>	X	1	17343675	<i>Gm19412</i>	17	0.6
17533539	<i>Rp2</i>	X	0.99	17511014	NA	8	0.6
17533806	<i>Gm4907</i>	X	0.99	17305315	<i>Gm8256</i>	14	0.6
17540902	<i>Btbd35f6</i>	X	0.98	17536997	<i>Zcchc13</i>	X	0.56
17532706	<i>Btbd35f2</i>	X	0.98	17547256	<i>Gm20806</i>	Y	-0.74
17532720	<i>Gm14367</i>	X	0.98	17390217	NA	2	-0.79
17539740	<i>Btbd35f4</i>	X	0.98	17547456	<i>Gm21672</i>	Y	-0.82
17540889	<i>Gm5925</i>	X	0.96	17546450	<i>Gm20809</i>	Y	-0.84
17539744	<i>Btbd35f21</i>	X	0.96	17547445	<i>Gm20747</i>	Y	-0.89
17532702	<i>Gm5925</i>	X	0.96	17464312	<i>4930479D17Rik</i>	6	-0.97
17536215	<i>LOC101055749</i>	X	0.95	17547392	<i>Gm20816</i>	Y	-0.99
17540011	<i>Wdr13</i>	X	0.94	17226736	<i>Mapkapk2</i>	1	-1.23
17548829	<i>Gm4916</i>	X	0.93	A further 16 transcript clusters annotated as <i>Ssty2</i> showed logFC values ranging from -0.81 to -1.07.			
17287249	NA	13	0.91	A further 20 transcript clusters annotated as <i>Slx</i>			

				showed logFC values ranging from 0.59 to 0.74.
17533939	<i>Btbd35f1</i>	X	0.88	A further 10 transcript clusters annotated as <i>Slx1</i> showed logFC values ranging from 0.69 to 0.79.
17546412	<i>Rbm31y</i>	Y	0.88	*Alternatively recorded as chromosome Y (MGI database)

Tables 5.7. Lists of genes determined as being differentially expressed at $p=0.05$ per comparison. Full tables with gene aliases and gene biotype information are available in Supplementary Material. No significant differential expression was detected in WTLF1 vs YQLF1, WTLF2 vs YQLF2, or WTMWC v YQMWC.

A number of the down-regulated poorly annotated Y-linked predicted (Gm****) and 'NA' transcript clusters are expected to originate from the massive repeat array on the Y long arm. Other known genes on the MSY are comparatively well annotated and are not expected to be down-regulated in this model.

5.5.3. STRING (search tool for the retrieval of interacting genes/proteins) analysis of differentially expressed protein-coding genes

STRING v.10.5¹⁹ (Szklarczyk 2015) requires a list of unique gene symbols or protein symbols as input for examination. A list of all differentially expressed genes across all between-strain comparisons was used; each gene symbol was included only once. Various synonyms exist for *Ssty2*, *Slx* and *Slx11* in the differentially expressed gene list. All identified synonyms, such as Gm20738 (*Ssty2*), Gm4836 (*Slx*), and Gm6660 (*Slx11*), of these genes were removed before String analysis was run; leaving a single representative of each gene. Some other predicted genes (Gm****) represented duplicates of gene symbols already represented in the list, and so were included under a single gene symbol. Only genes annotated as protein-coding were included in the analysis.

¹⁹ Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) v10.5 - <https://string-db.org/>

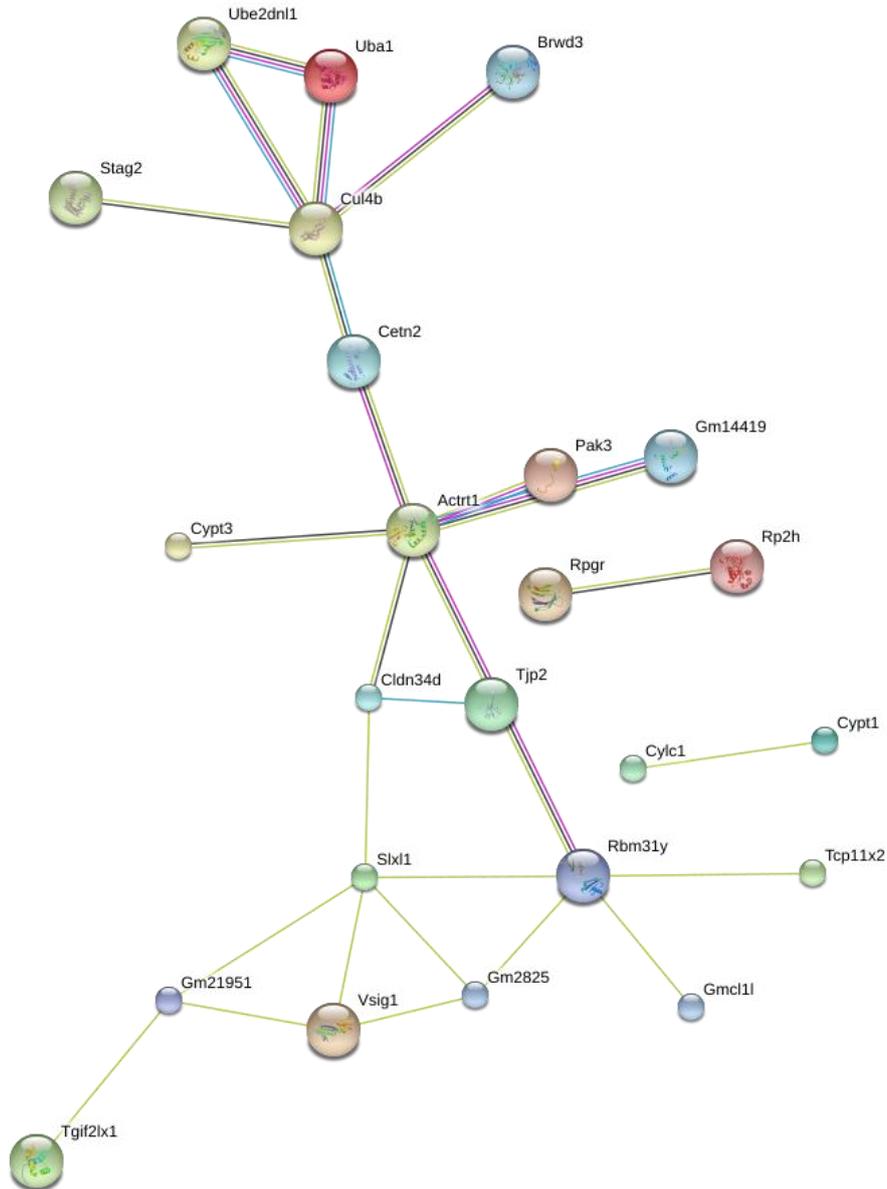


Figure 5.11. Graphical output of STRING protein functional association analysis at medium confidence. Unconnected nodes have been excluded from the image. Lines between nodes indicate the cooccurrence of connected nodes in curated datasets. Images within nodes indicate protein structure for proteins where structure is known or predicted.

STRING analysis of protein-coding genes revealed enrichment in BDB/POZ domain containing genes. Seven of the sixty-eight genes in the gene list were revealed to be BDB/POZ domain containing genes. These genes are discussed in section 5.6. Removal of BTB/POX domain containing genes did not reveal any further pathway enrichments. Gene names retrieved from STRING are included in *Table 5.8*.

Symbol	Gene name	Symbol	Gene name
<i>Actrt1</i>	Actin-related protein T1	<i>Ly6a</i>	Lymphocyte antigen 6 complex, locus A
<i>Amph</i>	Amphiphysin	<i>Mapkapk2</i>	MAP kinase-activated protein kinase 2
<i>Ar</i>	Androgen receptor	<i>4930402K13</i>	RIKEN cDNA 4930402K13 gene
<i>Brwd3</i>	Bromodomain and WD repeat domain containing 3	<i>Nptx2</i>	Neuronal pentraxin 2
<i>Btg1-ps1</i>	RIKEN cDNA 4930525M21 gene	<i>Ott</i>	Ovary testis transcribed
<i>Btg1-ps2</i>	RIKEN cDNA 4930430D24 gene	<i>Pak3</i>	P21 protein (Cdc42/Rac)-activated kinase 3
<i>C8b</i>	Complement component 8, beta polypeptide	<i>Piga</i>	Phosphatidylinositol glycan anchor biosynthesis, class A
<i>Cdkl5</i>	Cyclin-dependent kinase-like 5	<i>Pim1</i>	Proviral integration site 1
<i>Cetn2</i>	Centrin 2	<i>RBM</i>	RNA-binding motif protein, Y chromosome, family 1 member A1
<i>Cldn34d</i>	RIKEN cDNA 4930468A15 gene	<i>Rbm31y</i>	RNA binding motif 31, Y-linked
<i>Ctag2</i>	Cancer/testis antigen 2	<i>Ribc1</i>	RIB43A domain with coiled-coils 1
<i>Cul4b</i>	Cullin 4B	<i>Rp2h</i>	Retinitis pigmentosa 2 homolog (human)
<i>Cylc1</i>	Cylicin, basic protein of sperm head cytoskeleton 1	<i>Rpgr</i>	Retinitis pigmentosa GTPase regulator
<i>Cypt1</i>	Cysteine-rich perinuclear theca 1	<i>Slx</i>	Sycp3 like X-linked
<i>Cypt3</i>	Cysteine-rich perinuclear theca 3	<i>Slx1</i>	Slx-like 1
<i>4921511C20</i>	RIKEN cDNA 4921511C20 gene	<i>Stag2</i>	Stromal antigen 2
<i>Fam122c</i>	Family with sequence similarity 122, member C	<i>Syt15</i>	Synaptotagmin-like 5
<i>Fam46d</i>	Family with sequence similarity 46, member D	<i>Tcp11x2</i>	T-complex 11 X-linked protein 2
<i>Faxc</i>	RIKEN cDNA 6230409E13 gene	<i>Tet2</i>	Tet methylcytosine dioxygenase 2

<i>Fosb</i>	FBJ osteosarcoma oncogene B	<i>Tgif2lx1</i>	TGFB-induced factor homeobox 2-like, X-linked 1
<i>Foxj2</i>	Forkhead box J2	<i>Tjp2</i>	Tight junction protein 2
<i>Gm14419</i>	Predicted gene 14419	<i>Tspan6</i>	Tetraspanin 6
<i>Gm20852</i>	Ssty2 protein	<i>Uba1</i>	Ubiquitin-like modifier activating enzyme 1
<i>Gm362</i>	Predicted gene 362	<i>Ube2dn1</i>	Ubiquitin-conjugating enzyme E2D N-terminal like 1
<i>Gm4907</i>	Predicted gene 4907	<i>Vsig1</i>	V-set and immunoglobulin domain containing 1
<i>Gm5591</i>	Predicted gene 5591	<i>Wdr13</i>	WD repeat domain 13
<i>Gm614</i>	Predicted gene 614	<i>Yipf6</i>	Yip1 domain family, member 6
<i>Gm9</i>	MCG10360	<i>Zcchc13</i>	Zinc finger, CCHC domain containing 13
<i>Hypm</i>	RIKEN cDNA 1700054O13 gene		

Table 5.13. Minimal gene annotation of differentially expressed genes retrieved from STRING.

5.5.4. Predicting sharing behaviour of differentially expressed genes

To determine whether differentially expressed genes identified above (section 5.5.2) might be implicated in functional differences between X- and Y- bearing sperm, transcript distribution was evaluated using expression heatmaps to identify potential cis-limited transcripts. Expression patterns of genes of known sharing behaviour (section 5.5.1) were used as a reference. By comparing the transcript distributions, estimates of gene behaviour can be made; whether they might be more or less readily shared between developing spermatids.

It is expected that at least one X-linked gene is implicated in the sex ratio distortion due to the mechanism of action of genes involved in the intragenomic conflict. As haploid cells possess one copy of all autosomal genes and only one of either sets of X-linked or Y-linked genes, only sex-linked genes will be impacted by restricted sharing of genes. However, non-shared sex-linked genes may impact the expression of non-shared autosomal genes, which could result in a phenotypic effect through the actions of both the sex-linked and autosomal genes in question; and dysregulation of cis-limited autosomal genes could result in dosage differences between individual spermatids. Therefore,

both sex-linked and autosomal genes were assessed for subcellular fraction enrichment in this part of the investigation.

Examination of transcripts for potential cis-limitation or reduced sharing

The tables below were created for easy comparison of transcript distributions (by logFC values) between genes showing differential expression in microarray gene expression analysis. Three sample comparisons show the greatest significant differences in transcript abundance between genes of known sharing behaviour, as determined using genes listed in section 5.5.1:

- 1) F1 vs F3 2) F2 v F3 3) F3 v WC

Shared transcripts of *Akap4* and *Prm1* show significant negative logFC values in above comparisons 1 and 2, positive in 3 (increased presence in F1, F2 and WC compared to F3). Joint characteristics of non-shared transcripts *Smok2b* and *Spam1* are significant positive logFC values in 1 and 2, negative in 3 (Table 5.9).

	WT Early			YQ Early			WT Late			YQ Late		
Gene	1	2	3	1	2	3	1	2	3	1	2	3
<i>Akap4</i>	-1.2	-0.5	1.3	-1.1	-0.6	1.3	-0.6	-0.7	0.8	-0.6	-0.6	0.9
<i>Prm1</i>	-1.0	-0.9	0.8	-1.0	-0.9	0.8	0.0	0.0	0.0	0.0	0.0	0.0
<i>Smok2b</i>	1.8	2.6	-1.6	1.6	1.7	-1.4	0.0	1.1	0.0	0.6	0.6	0.0
<i>Spam1</i>	1.9	1.3	-0.9	2.1	1.7	-1.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 5.9. LogFC values from comparisons 1) F1 vs F3, 2) F2 vs F3, 3) F3 vs WC, of control genes within stated cell populations. Only logFC values that are significant at $p > 0.05$ at 50% stringency have been included. Other values are printed as zero.

Of the genes listed in section 5.5.2 as being differentially expressed between WT and YQ spermatids, genes listed in Table 5.10 showed patterns of transcript distribution most consistent with non-shared transcripts *Smok2b* and *Spam1*. Only logFC values that are significant at $p > 0.05$ at 50% stringency have been included. Other values are printed as zero. A number of X-linked genes appear in the selection. Of the genes listed, five genes were up-regulated in YQ late spermatid whole cell samples (*Ar*, *Brwd3*, *Cdkl5*, *LOC100861738*, *Lya6a*), one up-regulated in YQ early spermatid whole cell samples (*Cldn34-ps*), two up-regulated in YQ early spermatid fraction 3 (*Cldn34-ps* and *Stag2*), one up-regulated in YQ late spermatid fraction 3 (*Cylc1*) and two down-regulated in YQ late spermatid fraction 3 (*Fosb* and *Tet2*).

Gene symbol	Chr	WT Early			YQ Early			WT Late			YQ Late		
		1	2	3	1	2	3	1	2	3	1	2	3
<i>Ar</i>	X	1.6	1.8	-1.4	1.1	1.5	-1.2	3.6	2.7	-3.0	2.7	2.2	-1.9
<i>Brwd3</i>	X	2.4	2.1	-0.7	2.5	2.3	-0.9	1.5	1.0	0.0	1.4	1.1	0.0
<i>Cdkl5</i>	X	1.7	2.0	-0.5	1.9	1.8	-0.6	0.0	0.0	0.0	0.8	0.0	0.6
<i>Cldn34-ps</i>	X	1.2	0.9	-1.0	1.1	1.4	-1.2	0.0	0.0	0.0	0.0	0.0	0.0
<i>Cylc1</i>	X	3.2	3.1	-1.2	3.1	2.5	0.0	2.6	2.0	0.0	3.9	2.4	0.0
<i>Fosb</i>	7	1.7	1.4	-1.1	2.0	1.3	-1.2	2.2	1.7	-0.9	1.3	0.0	0.0
<i>LOC100861738*</i>	Y	2.1	1.8	-1.7	2.0	2.7	-1.5	2.5	2.7	-2.4	2.5	2.9	-1.9
<i>Ly6a</i>	15	1.7	1.6	-1.0	1.4	1.9	-1.1	4.3	3.1	-3.2	4.5	3.6	-3.2
<i>Stag2</i>	X	2.9	2.5	0.0	3.7	3.1	-0.5	3.0	2.8	-1.6	3.4	2.7	-1.2
<i>Tet2</i>	3	0.0	0.0	0.0	0.0	0.0	0.0	3.6	3.6	-3.2	2.6	3.1	-1.9

Table 5.10. LogFC values from comparisons 1) F1 vs F3, 2) F2 vs F3, 3) F3 vs WC, of DE genes within stated cell populations. All genes listed show similar patterns of transcript distribution to control non-shared genes (enrichment in F3). Only logFC values that are significant at $p > 0.05$ at 50% stringency have been included. Other values are printed as zero. *Gene also recorded as ‘eukaryotic translation elongation factor 2, pseudogene’

Several other genes, including lincRNAs, snRNAs, multiple poorly annotated pseudogenes and predicted genes also show this pattern of transcript distribution. These are listed in the Supplementary Material. All BTB domain containing genes (14 transcript clusters linked to 11 different gene IDs) showed this pattern across primarily early spermatid populations and, in some instances, late spermatid populations. Across all of these genes, this pattern was exacerbated in WT early spermatids compared to YQ and exacerbated in YQ late spermatids compared to WT, suggesting a change in transcript localisation between the two strains.

5.5.5. Genes of interest from the literature and previous work

Prior to commencement of experimental work, a number of genes of interest were identified from the literature (section 1.8). As in the previous chapter, the behaviour of these genes was examined in this dataset. Except for the genes listed (*Table 5.11*), all genes behaved as described in the previous dataset (*Table 4.11*).

Gene symbol	Transcript distribution in RNA sequencing dataset
<i>Speer</i>	Ten <i>Speer</i> genes were identified in the dataset. Only two of these did not show strong enrichment in F3 and F2 in early spermatids; <i>Speer2</i> and <i>Speer4cos</i> . Some <i>Speer</i> genes showed mild upregulation in one or more strain to strain comparisons; however this was not the case for all detected transcripts.
<i>Tcp11x2</i>	<i>Tcp11x2</i> shows enrichment in F3 in early stage spermatids (unlike in the previous), however it did not fully follow pattern of distribution as noted in control genes. As in the previous dataset, <i>Tcp11x2</i> shows upregulation across all matched sample comparisons, this time the gene was short-listed.

Table 5.11. Behaviour of genes of interest identified from the literature. Only genes that showed different behaviour to that recorded in Chapter 4 (Table 4.11) are included.

5.6. Chapter Discussion

This chapter covers work aimed at highlighting genes that are differentially expressed in the post-meiotic spermatids and mature sperm of MF1XY^{RIII}qdel mice compared to MF1XY^{RIII}. Through this investigation, the author sought to identify potential gene candidates for involvement in the observed offspring sex ratio distortion and sperm phenotypic abnormalities (Chapter 3). This investigation utilised three replicate sets of spermatid and sperm samples including fractionated samples of spermatid subcellular regions for the identification of potentially non-shared (cis-limited) genes by similarity of transcript distribution to known cis-limited genes. This also allowed for the investigation of differential transcript localisation by directly comparing expression of transcripts within three subcellular fractions between WT and YQ. Shortlisted genes were further examined in the literature and through the use of databases containing experimentally-backed and predicted protein interactions. This multi-replicate experiment follows on from work conducted in Chapter 4; comparisons were drawn between results generated from the two investigations. As with Chapter 4, results of physiological analyses conducted in Chapter 3 and information gained from investigation of the literature in Chapter 1 have been taken into account when considering which genes (and pathways) are suitable for further investigation.

Several genes exhibit differential expression and potential for cis-limitation

A number of differentially expressed genes showed patterns of transcript distribution aligned with those of non-shared (cis-limited) control transcripts. These genes are: *Ar*, *Brwd3*, *Cdkl5*, *Cldn34-ps*, *Cycl1*, *Fosb*, *LOC100861738*, *Ly6a*, *Stag2*, and *Tet2*. Of these genes, six were shown to be X-linked.

A gene annotation search was performed on all genes using STRING v.10.5²⁰ (Szklarczyk 2015), UniProt²¹ and NCBI²². A table containing extended annotations for these candidates is included in the Supplementary Material; where points of particular interest to the YQ model are highlighted. All of these X-linked genes could have consequences for sperm following dysregulation during spermiogenesis. Suggestions for further investigation of these genes, alongside potentially cis-limited X-linked candidates from other investigations, are discussed in section 7.1.

Enrichment of BDB/POZ domain containing genes in STRING functional protein association network analysis of differentially expressed genes

All differentially expressed BDB/POZ domain containing genes showed increased expression in YQ late whole cell spermatids compared to WT and some in early whole cell spermatids. In addition, all of these genes showed transcript distribution patterns highly similar to those of known non-shared transcripts and all are X-linked. All transcript clusters were members of the *Btbd35* family.

<u>Gene symbol</u>	<u>Aliases</u>
<i>Btbd35f1</i>	<i>Gclh</i> ; <i>Gmcl2</i> ; <i>Gmcl11</i> ; <i>1700017G21Rik</i>
<i>Btbd35f16</i>	<i>Gm3701</i>
<i>Btbd35f17</i>	<i>Gm10922</i>
<i>Btbd35f2</i>	<i>Gm2799</i>
<i>Btbd35f21</i>	<i>Mgclh</i> ; <i>Gm21951</i>
<i>Btbd35f4</i>	<i>Gm3750</i>
<i>Btbd35f5</i>	<i>Gclh</i> ; <i>Gm2825</i> ; <i>Gmcl11</i> ; <i>GMCL1P1</i>
<i>Btbd35f6</i>	<i>Gm2964</i>
<i>Btbd35f7</i>	<i>Gm14374</i>
<i>Btbd35f9</i>	<i>Gm21681</i>

BTB domains are highly prevalent protein-protein interaction motifs, occurring throughout eukaryotes. Family members listed above possess multiple aliases. Family members 1 and 5 are also known as *Gmcl11*. This gene is a known ampliconic gene on the X, suggesting that other members are additional copies of the same gene, which was shown to have undergone significant expansion in recent mouse evolution (Ellis et al. 2011). The functions and implications of this gene are currently unknown. However, common functions of BTB domain containing genes are in the recruitment of E3 ubiquitin ligase complex degradation targets and the regulation of transcription (Perez-Torrado et al. 2006), meaning that the protein could potentially be involved in the breakdown of protein structure during reconfiguration of the nucleus. These provide ideal candidates for further investigation.

²⁰ Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) v10.5 - <https://string-db.org/>

²¹ UniProt - <http://www.uniprot.org/>

²² NCBI - <https://www.ncbi.nlm.nih.gov/>

Genes showing down-regulation in matched sample comparisons

Genes down-regulated in YQ primarily consist of Y-linked predicted genes comprising part of the huge repeat array on the reduced MSYq. However, two autosomal genes, *Mapkapk2* and *4930479D17Rik* show strong down-regulation both in whole cell and one subcellular region of early stage spermatids. *Mapkapk2* is a mitogen-activated protein kinase-activated protein kinase (2) located on chromosome 1. It is regulated through direct phosphorylation by *Mapk14* (Mitogen activated protein kinase 14) (Kotlyarov et al. 1999). *Mapkapk2* was found to have known interactions with 23 *Mus musculus* genes (EBI – IntAct²³), however, none of these were X-linked or showed altered expression in YQ vs WT. *Spin*, a relative of the massively multicopy Y-linked gene family *Ssty* that showed extensively reduced expression in YQ, is known to be a substrate in the MAP kinase pathway in oocytes, where it is specifically expressed (Oh et al. 1998). Its phosphorylation is linked to the presence of MAP kinases. In mice lacking MAP kinase activity, only partial phosphorylation of SPIN occurs and association with the meiotic spindle is reduced, affecting its function. It is possible that *Ssty* family members may be phosphorylated by similar proteins. *4930479D17Rik* is an autosomal (Chr. 6) lincRNA that is located in close proximity to (not over-lapping) protein-coding gene *Med21*, a subunit of the ‘Mediator of RNA polymerase II transcription’ complex. Potential interactions of this lincRNA were investigated using the linc-mRNA interaction database lincRNAator²⁴ (Park et al. 2014), an extensive database of species-specific interactions and further annotations of lincRNAs. This confirmed high expression of this lincRNA in testis tissue and revealed one significant lincRNA-protein interaction with FUS/TLS, a RNA/DNA binding protein primarily associated with neurodegenerative diseases. The N-terminal of this protein contains a prion-like domain that is subsequently involved in transcriptional activation and the formation of RNA-containing granules (Kino et al. 2016). It is also suggested that FUS may play a role in genomic integrity (as determined by sequence similarity) (UniProt-KB search). *Fus* shows high levels of expression in mouse testis, though limited evidence is available for its specific expression in spermatids (Gene Expression Atlas – EBI). Reduced expression of this lincRNA in early stage spermatid nuclear/organelle bound fractions (EF3) could alter the subsequent activity of transcription-altering and RNA/DNA binding proteins with which it interacts.

Up-regulation of *Cldn34d* and a related pseudogene *Cldn34-ps*

Both *Cldn34d* and *Cldn34-ps* are up-regulated in early spermatids. The pseudogene is also up-regulated in EF3. They show distinctly different behaviours; *Cldn34d* shows shared behaviour,

²³ EBI – IntAct: <http://www.ebi.ac.uk/intact/>

²⁴ lincRNAator - <http://lincnator.ewha.ac.kr/index.htm>

Cldn34-ps shows cis-limited. A different claudin family member, *Cldn34b4*, was short-listed as a gene of interest following RNA sequencing (Chapter 4), being both X-linked and showing potential cis-limitation. Dysregulation of other claudins has been linked to altered sperm fitness outcomes (mentioned in section 4.5). A search of sequence similarity using NCBI BLAST²⁵ of the pseudogene mRNA with discontinuous Mega BLAST returns hits with >90% sequence identity to claudin variants in *Mus caroli*, *Mus pahari* and rat. The pseudogene is annotated as having no protein product, but the mRNA was shown to have a large intact open reading frame (ORF) so may well retain some functionality (examined by ORFfinder²⁶).

Neuronal pentraxin 2 (*Nptx2*) – Chr.5

Nptx2 expression was notably increased in YQ spermatids compared to WT. It was shown to be the top differentially expressed gene in comparisons between matched EF2 and LWC samples, respectively. It did not display patterns of transcript distribution similar to non-shared genes, but showed the greatest upregulation of any gene across all samples (logFC of 2.23 in LWC). This upregulation was also noted in late stage whole cell spermatids in RNA sequencing data (section 4.5) where *Nptx2* was also shortlisted as one of the top 10 positively differentially expressed genes in this cell type and a gene of interest. *Nptx2* is an autosomal gene and shows homology with pentraxins across several species, including human NPTX2, and *Am50* (an acrosomal matrix gene) from guinea pig (Buffone 2016). AM50 is a major structural component of the acrosomal matrix in guinea pig sperm. NPTX2, like other pentraxins contains a highly conserved calcium-binding domain and strongly binds calcium. Its further-studied homologue, AM50, is processed at sites undergoing acrosomal disassembly during the acrosome reaction following the acrosomal exocytosis and activation of acrosin which is implicated in its proteolytic disassembly (Kim et al. 2011). *Nptx2* (also referred to as *Np2*) knockout mice have been produced (Bjartmar et al. 2006) and were described as being fertile; however no mention of sperm parameters or sperm fitness was made beyond this. Given information from the literature, over-expression of *Nptx2* could lead to abnormal structural development of the acrosomal matrix or be involved in reduced efficiency of the acrosome reaction prior to fertilisation, depending upon resulting translation of additional transcripts. It may also, like other pentraxins, play an important role in the uptake of extracellular material (Hsu and Perin 1995). Given its calcium binding properties, this could be of particular importance in sperm functions including capacitation and the activation of hypermotility. Considering its consistent dysregulation within the same cell type across two gene expression platforms, a strong indication of functional relevance from the literature, and transcript distribution comparable to that of shared transcripts,

²⁵ NCBI Basic Local Alignment Search Tool (BLAST) - <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

²⁶ Open reading frame finder (ORFfinder) - <https://www.ncbi.nlm.nih.gov/orffinder/>

Nptx2 appears as a prime candidate for the structural sperm head deformities evident across YQ sperm and potentially involved in other unmeasured physiological parameters that might affect to some degree all sperm from YQ mice.

Morphological aberrations & global alterations

In addition to *Nptx2*, *Brwd3*, and *Cylc1*, a number of differentially expressed genes are strongly linked to cellular morphology, and often sperm head morphology. Given what we now know about the distribution and extent of morphological aberrations seen in YQ sperm, genes implicated in causing these abnormalities need not be cis-limited but could show some restricted sharing. *Actrt1* and *Cypt8* showed differential expression in whole cell samples and a strong functional link to developing cellular morphology in spermiogenesis. *Actrt1* was also short-listed during RNA sequencing analysis (Chapter 4). *Cylc1* has shown potential for cis-limitation in both RNA seq (Chapter 4) and microarray expression analyses, and strong upregulation in YQ in at least one sample comparison in both datasets. This gene was initially discussed in section 4.5, and is further discussed in Chapter 7.

Differential expression in early stage spermatids

Sly and *Slx* are known to exert their effects upon chromatin primarily in round spermatids. Round spermatids form the primary cell type of our early stage spermatid samples and are the most transcriptionally active post-meiotic phase. Several genes not previously mentioned were significantly upregulated in YQEWB samples compared to WT. These genes were further investigated using MGI, STRING, UNiProt and basic literature searches. Further annotation of these genes and suggested links to YQ sperm outcomes are listed in the Supplementary Material.

Upregulation of *Rbmy* in late stage spermatids

Rbmy is a Y-linked gene, the protein of which is a testis specific RNA-binding protein and an important factor in successful spermatogenesis. It was previously thought to be expressed only during early spermatogenesis and was not expected to be transcribed in spermatids. However, in mouse, *Rbmy* has been shown to be expressed both in spermatogonia and spermatids (Mahadevaiah et al. 1998). As a spermatid expressed gene on the Yp, we expect to see upregulation of this gene in YQ, owing to the interactions of *Sly*, *Slx* and global sex chromatin upregulation. Further examination of the literature showed that although *Rbmy* is now known to be transcribed in spermatids, it does not appear to be subsequently translated (Szot et al. 2003), leading to uncertainty about its purpose in post-meiotic cells. Removal or knockout of *Rbmy* is associated with meiotic arrest in the human germ line and sperm abnormalities in mice, however these abnormalities are thought to come about as a result of

failure of *Rbmy* during early spermatogenesis (Szot et al. 2003) prior to spermiogenesis and the actions of *Sly*. *Rbmy* copy number is not expected to be reduced in MF1XY^{RIII}qdel, and its notable up-regulation would suggest this is the case. Touré et al. (2004a) also showed that levels of RBMY protein in the related XY^{RIII}qdel model were normal, so it seems unlikely that *Rbmy* is a key source of abnormalities in the MF1XY^{RIII}qdel.

No transcript clusters were recorded as being significantly differentially expressed between the two mature sperm samples WTMWC and YQMWC. A note on expression in mature cells relating to this dataset is included in the Supplementary Material.

5.6.1. Comparison of results from RNA sequencing (Chapter 4) and microarray data (Chapter 5)

Although comparing similarly generated samples, the RNA sequencing (Chapter 4) and microarray data (Chapter 5) present distinct datasets due to the differences inherent in sequencing technologies, experimental design, desired objectives of each investigation and the adaptation of data processing techniques to the individual investigation. This is also evident in the differing patterns of expression of control genes of known sharing behaviour between RNA sequencing and microarray data. Although both sets (shared and non-shared) of transcripts show distinct distributions from one another as expected, the magnitude of the logFC values was reduced for non-shared transcripts and increased for shared transcripts in Chapter 4 compared to Chapter 5. The character of the distributions remained similar between datasets. This is likely caused by differences in data normalisation techniques and potentially due to DNA contamination discovered in five samples used in the RNA sequencing work. However, assessing these transcripts independently in both datasets allows for the adaptation of investigative techniques to reduce error and target candidate selection techniques for desired transcript traits.

Across RNA sequencing and microarray data covered in chapters 4 and 5 respectively; distinct sets of analysis were completed, these were modified and appropriate to the individual dataset. However, some comparisons may be drawn when considering matched cell sample differential expression analysis made between WT and YQ spermatids in both datasets. Genes detected as strongly and consistently differentially expressed across both datasets include 13 of the 88 unique genes short-listed from RNA sequencing data (*Table 5.12*). This includes *Nptx2*, a neuronal pentraxin strongly up-regulated in late spermatids, and *Vsig1*, V-set immunoglobulin containing 1 shown to be up-regulated in both early and late spermatids. A high level of similarity was not expected due to both the reasons mentioned above and the differing selection criteria used between the two techniques (top 20 LogFC

per comparison vs p-values across all samples). Although RNA sequencing arguably represents the more powerful and flexible technology, the microarray data presented in this chapter is composed of three replicates that showed high similarity and allowed for the assessment of significant differences in expression to be made.

Gene symbol	Chr	Gene biotype	Gene name
<i>4930402K13Rik</i>	X	Protein coding	RIKEN cDNA 4930402K13 gene
<i>Actrt1</i>	X	Protein coding	actin-related protein T1
<i>Cylc1</i>	X	Protein coding	cylicin, basic protein of sperm head cytoskeleton 1
<i>Fam122c</i>	X	Protein coding	family with sequence similarity 122, member C
<i>Gm14367</i>	X	Unprocessed pseudogene	predicted gene 14367
<i>Gm15104</i>	X	lincRNA	predicted gene 15104
<i>Gm25360</i>	1	snRNA	predicted gene, 25360
<i>Gm3763 (Gmcl11 multicopy)</i>	X	Protein coding	BTB domain containing 35, family member 27 (and other members)
<i>Gm614</i>	X	Protein coding	predicted gene 614
<i>Gm6866</i>	7	Transcribed unprocessed pseudogene	predicted gene 6866
<i>Nptx2</i>	5	Protein coding	neuronal pentraxin 2
<i>Tspan6</i>	X	Protein coding	tetraspanin 6
<i>Vsig1</i>	X	Protein coding	V-set and immunoglobulin domain containing 1

Table 5.12. Genes shortlisted in differential expression analysis from both major techniques; RNA sequencing (Chapter 4) and RNA microarray (Chapter 5).

5.7. Chapter Summary

Through the work covered in this chapter, we have identified a set of potentially cis-limited sex-linked genes which show dysregulation in the MF1XY^{RIII}qdel model for further investigation of their involvement in offspring sex ratio distortion. We have also identified a number of dysregulated candidates, which show enrichment in the cytosolic subcellular fractions, suggesting sharing across

cytoplasmic bridges. Some of these genes have strong functional links to abnormal sperm phenotypes noted across the entire MF1XY^{R^{III}}qdel sperm population. Of particular interest are *Nptx2* and *Actrt1*, which were also highlighted in RNA sequencing data. Prime X-linked candidates identified during this analysis, that is differentially expressed X-linked genes showing potential for cis-limitation, are *Gmcl1l* (and copies), *Ar*, *Brwd3*, *Cdkl5*, *Cldn34-ps*, *Cylc1* and *Stag2*. An examination of genes of interest from the literature highlighted *Tcp11x2* as being strongly up-regulated with enrichment in bound fractions. Additional potential implications of dysregulation of these genes is discussed in Chapter 7, along with further work that could be undertaken to advance understanding of their involvement in the phenotypes observed in MF1XY^{R^{III}}qdel mice.

Chapter 6

6. Micro RNA Expression Study of Spermatids and Mature Sperm

The data presented in this chapter were retrieved following RNA sequencing and micro RNA (miRNA) microarray examination of RNA samples from spermatids and mature sperm of MF1XY^{RIII} and MF1XY^{RIII}qdel mice. Spermatid populations were enriched for cell type by centrifugal elutriation and subcellular fractions were retrieved by exposure of cellular populations to RNA-preserving lysis buffers as previously described in Chapter 4 (section 4.1). This chapter examines the differential expression and distribution of micro RNAs in MF1XY^{RIII} and MF1XY^{RIII}qdel germ cells at different stages of maturation.

Previously described reproductive and sperm phenotypic abnormalities stem from a broad up-regulation of the sex chromatin during spermiogenesis, which has consequences for the competitiveness of Y-bearing sperm expected to be caused by altered expression of several protein-coding genes. As understanding of non-coding RNAs has expanded, micro RNAs (miRNAs) have been increasingly implicated in the altered expression of protein coding genes. The mechanisms resulting in functional inequivalence between X- and Y-bearing sperm are not yet fully understood and it is possible that miRNAs, which are relatively abundant in post-meiotic spermatids and mature sperm, may have a role in facilitating differential expression of proteins and subsequently physiological differences. Although an upregulation of sex-linked genes and few autosomal genes is expected as a consequence of *Sly* and *Ssty* loss in YQ spermatids, several genes (including X-linked genes) appear down-regulated in expression data (chapters 4 and 5), this could be facilitated by the increased expression of associated miRNAs. Trans-generational effects not linked to differences in chromatin content have been also noted in the offspring of MSYq mice, and might possibly be linked to differential loading of miRNAs in the mature sperm. The examination of miRNA content is performed in the light of previously defined physiological, offspring sex ratio and total RNA expression differences between the two strains; see chapters three, four and five for a detailed explanation of these findings.

The experiments and analysis detailed below have allowed for a broad examination of miRNA expression across both strains during several stages of spermiogenesis. This has led to the identification of a number of differentially expressed miRNAs tied to specific cellular maturational stages and subcellular regions. Some of these miRNAs have been identified as potential interacting

partners of previously identified differentially expressed protein-coding genes, providing a selection of biochemical pathways plausibly involved in both the offspring sex ratio distortion and broader physiological changes noted in MF1XY^{RIII}qdel sperm. Other differentially expressed miRNAs from the data have experimentally-validated interactions with spermatid-specific transcripts and enrichment of target mRNAs in pathways of interest. In addition, the findings of both sets of data discussed in this chapter (RNA sequencing and microarray) are compared and the two approaches evaluated.

6.1. Chapter Introduction

Control of gene expression by small noncoding RNA molecules was first observed in 1993, when a team of researchers discovered a small, double-stranded RNA (dsRNA) in nematode (*Caenorhabditis elegans*) larvae that complemented the sense strand of a larger mRNA and bound to its 3' untranslated region, thus inhibiting translation (Lee et al., 1993). Since then, a number of different mechanisms for translational control by small RNAs have been discovered.

Micro RNAs (miRNAs) are small (normally 18-22 nucleotides in length) nucleic acid sequences that are highly specific regulators of genes. Micro RNAs as a transcript biotype are highly conserved, having been found in many taxa, ranging from humans, worms and mice to plants and algae (Molnár et al. 2007, Reinhart et al. 2002, Lagos-Quintana et al. 2001, Pasquinelli et al. 2000), although with some major differences between those found in the plant and animal kingdoms (reviewed in Axtell et al. 2011). They have been shown to act primarily through the inhibition of mRNA translation into proteins via degradation, however a number of different mechanisms of action have now been revealed for metazoan mRNAs. Micro RNAs are now known to function through interactions with various genetic regions at both the DNA and RNA level and have a role in upregulation of genes as well as the hindrance of translation and selective degradation of mRNAs (Dharap et al. 2013, Place et al. 2008). Dysregulation of miRNAs has been implicated in a broad range of diseases, including cancers such as colorectal cancers and leukaemia (reviews by Liao et al. 2017, and Chi and Zhou 2016). They have also been implicated in essential cell functions such as cell death and differentiation (Xu et al. 2003, Brennecke et al. 2003, Lee et al. 1993). Micro RNAs exist in both pri-miRNA (immature), pre-miRNA (precursor) and miRNA (mature) forms and a number of maturation pathways exist for these transcripts (Cloonan 2015). *Figure 6.1* illustrates the major pathway for the maturation of miRNAs. A large proportion of recognized miRNAs are known to reside within introns of other protein-coding and non-coding genes (review by Gao et al. 2012, Rodriguez et al. 2004); and miRNAs are generally divided into 'intronic' and 'intergenic' depending upon their genomic location.

The existence of many miRNA-target interactions have been computationally predicted, and there are several miRNA knowledge databases that incorporate target-prediction programs (Wong and Wang 2015, Cho et al. 2013, Hsu et al. 2008). However, until recently, few of these had been validated by experimental work. MiRTarBase is a database that provides access to information on experimentally-validated miRNA-target interactions (EVMTIs); between the 2014 and 2016 updates to the database, the number of EVMTIs increased nearly 7-fold from just over 50,000 to 348,007 (Chou et al. 2016).

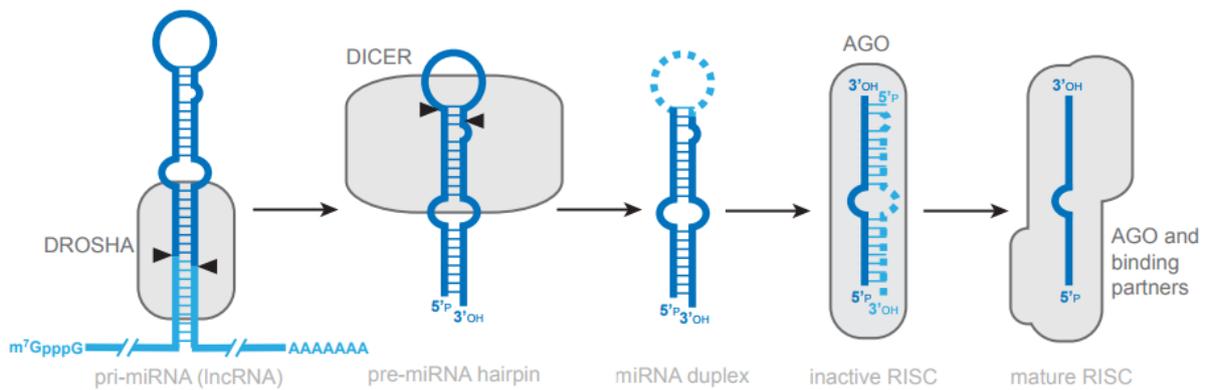


Figure 6.1. Major enzymatic steps for the maturation of primary transcripts into mature miRNAs. Only the major canonical pathway is shown. Image adapted from Cloonan (2015).

Since 1993, when control of gene expression by a small non-coding RNA was first discovered, a number of different mechanisms for the disruption of gene expression by these small nucleic acids have been observed. For example, many small interfering RNAs (siRNAs) form a siRNA-protein complex (the RNA-induced silencing complex, or RISC complex) with an argonaute protein which specifically binds to and results in the cleavage of complementary messenger RNA. Micro RNAs interact with target mRNAs and genes through specific base-pairing by a major domain of the miRNA called the ‘seed’ region. (Yao et al. 2015). MicroRNA-induced silencing complexes (miRISCs) more commonly act by binding to the 3’ region of a mature mRNA, preventing translation into a functioning protein (Ling et al. 2013). These processes are referred to as RNA interference (RNAi). However there are now known to be a variety of mechanisms of action by which miRNAs act on DNA and RNA. *Figure 6.2* shows mechanisms by which miRNAs may act on mRNA.

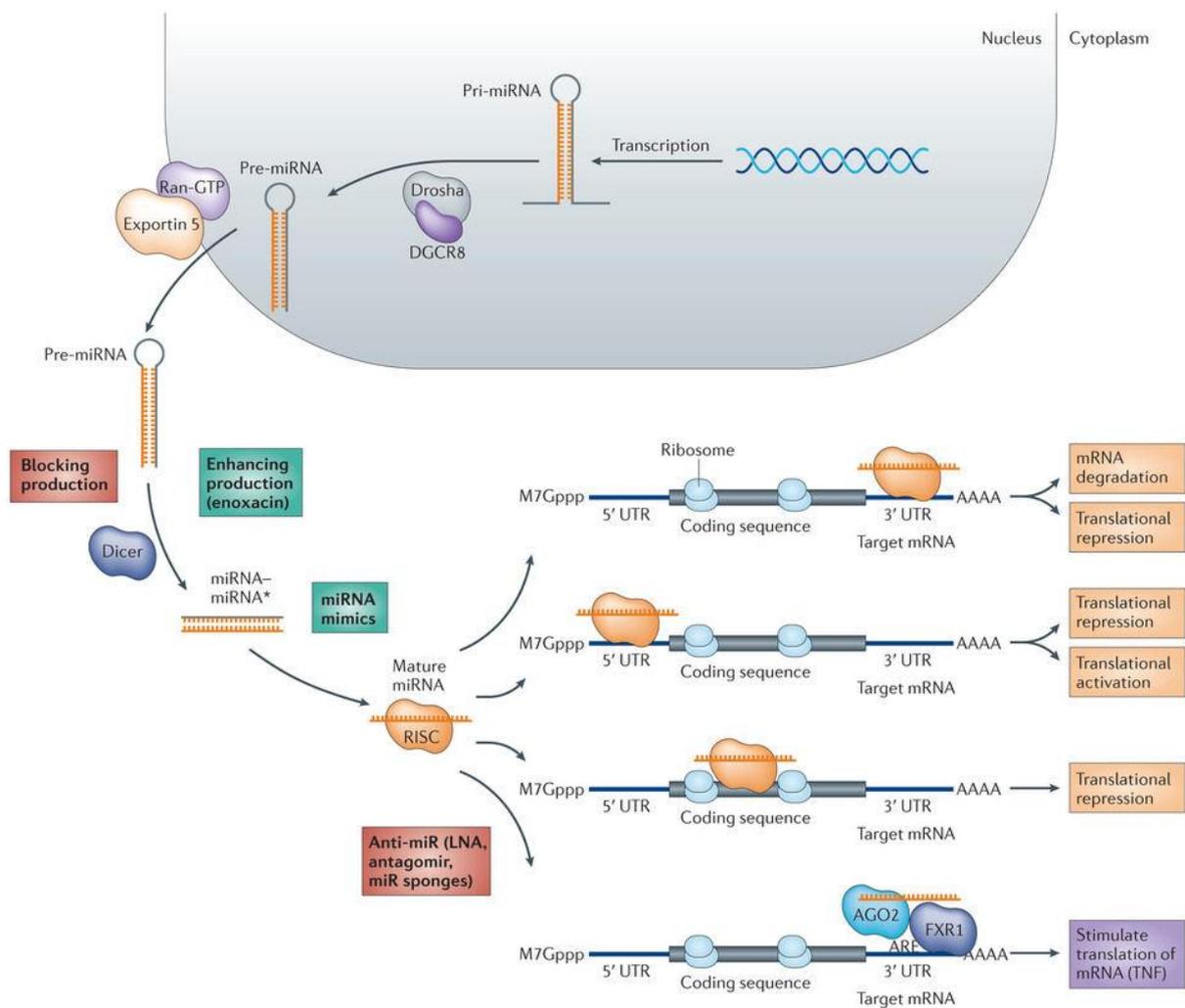


Figure 6.2. The conserved maturation pathway of miRNAs within a cell and the subsequent diverse mechanisms of miRNA action upon mRNAs. Primary miRNAs (pri-miRNAs) are transcribed in the nucleus and processed into precursor miRNAs (pre-miRNAs) by a Drosha and DGCR8 complex. Following pre-miRNA export from the nucleus by an exportin complex, an enzyme of the RNase III family known as Dicer converts immature pre-miRNAs into mature functional miRNAs. Dicer is also key in the processing of long double stranded DNA (dsDNA) molecules into small interfering RNAs (siRNAs). In the figure, green and red boxes present strategies that increase miRNA activity and strategies that reduce it, respectively. Micro RNAs are then inducted into the RISC (RNA induced silencing complex) and affect mRNA translation through a variety of mechanisms as shown. Figure taken from Ling et al. (2013).

Evidence is accumulating for the cell stage specific expression of miRNAs in mammalian spermatogenesis (Tan et al. 2014, Khazaie and Nsar Eshfahani 2014, Huszar and Payne 2013). The knockdown or knockout of *Dicer*, which is essential for the successful processing and maturation of miRNAs and other sncRNAs, results in a variety of reproductive abnormalities in spermatogenesis (Korhonen et al. 2011, Romero et al. 2011, Hayashi et al. 2008). Following the deletion of *Dicer* in prospermatogonia, examination of the testes revealed delayed progression to and increased apoptosis at the spermatocyte stage, reduction in the number of round spermatids produced, and morphological and motility abnormalities in the few mature spermatozoa; all resulting in infertility (Romero et al. 2011). In haploid spermatids roles for miRNAs in the regulation of chromatin condensation have been revealed. *MiR-122a* has been shown to reduce the expression of *Tnp2* (transition protein 2) in late stage spermatids (Yu et al. 2005); transition proteins are essential for the successful transition of histone bound chromatin to the highly-condensed protamine bound form found in mature sperm. MiRNAs have also been implicated in the reduced translation of TNP2 and PRM2 (protamine 2); down-regulation of *miR-469* was linked to temporally-aberrant translation of Tnp2 and Prm2 (protamine 2) leading to failure to produce mature sperm in DDX25 knockout mice (Dai et al. 2011). As mentioned in Chapter 4, mature sperm are considered to be transcriptionally inactive, with the exception of a number of genes of mitochondrial origin (Grunewald et al. 2005, Premkumar and Bhargava 1972). The chromatin of sperm must undergo highly regulated tight packing for maturation to be completed successfully. This packaging is essential for fully functioning sperm, allowing the DNA to fit into the streamlined and dorso-ventrally flattened sperm head which permits effective motility. Incorrect DNA packaging leads to problems with genome integrity, embryonic development, and can result in infertility (Liu et al. 2012, Vavouri and Lehner 2011, Cho et al. 2003, Cho et al. 2001, de Yebra et al. 1993). However, mature sperm have been shown to contain RNA (Krawetz 2005) and this RNA is highly diverse in nature (reviewed by Hosken and Hodgen 2014, Peng et al. 2012). As previously described (Chapter 4 section 4.4.4), micro RNAs in mature sperm have been implicated in several sets of disease and health outcomes for offspring, and there are a large number of studies on this phenomenon in mice alone (Fullston et al. 2016, Grandjean et al. 2015, Rodgers et al. 2015, Sharma et al. 2015, Rodgers et al. 2013). This transgenerational inheritance even extends to increased sensitivity to certain stimuli resulting from traumatic experience of the parent (Dias and Ressler 2014, Gapp et al. 2014). Indeed, the RNA load of sperm has been shown to affect offspring phenotypes (Chen et al. 2016a, review by Chen et al. 2016b), with Gapp et al. (2014) providing the first evidence of sperm RNAs being responsible for the transmittance of an acquired phenotype to offspring.

Sperm-borne RNA transcripts can transmit experience-dependent information from parent to offspring via sperm. As previously discussed (section 1.3), the transcriptional profile of spermatids changes as they progress through spermiogenesis. For transcriptionally inactive mature sperm, it is RNAs retained from late spermatid stages that form the majority of their RNA payload. The RNA profile of sperm can however, be slightly misleading. Small non-coding RNAs tend to be retained in the cytoplasm (as this is where snRNAs act on mRNA molecules following interaction with nuclear export protein exportin), yet when the mature sperm fuses with an oocyte, the minimal cytoplasm that remains is shed. Nuclear RNAs are more likely to impact embryo development and the health of future generations as these are retained and transferred to the oocyte (Johnson et al. 2015, Johnson et al. 2011). *Figure 6.3* provides an overview of snRNA composition in mouse germ cells during spermiogenesis.

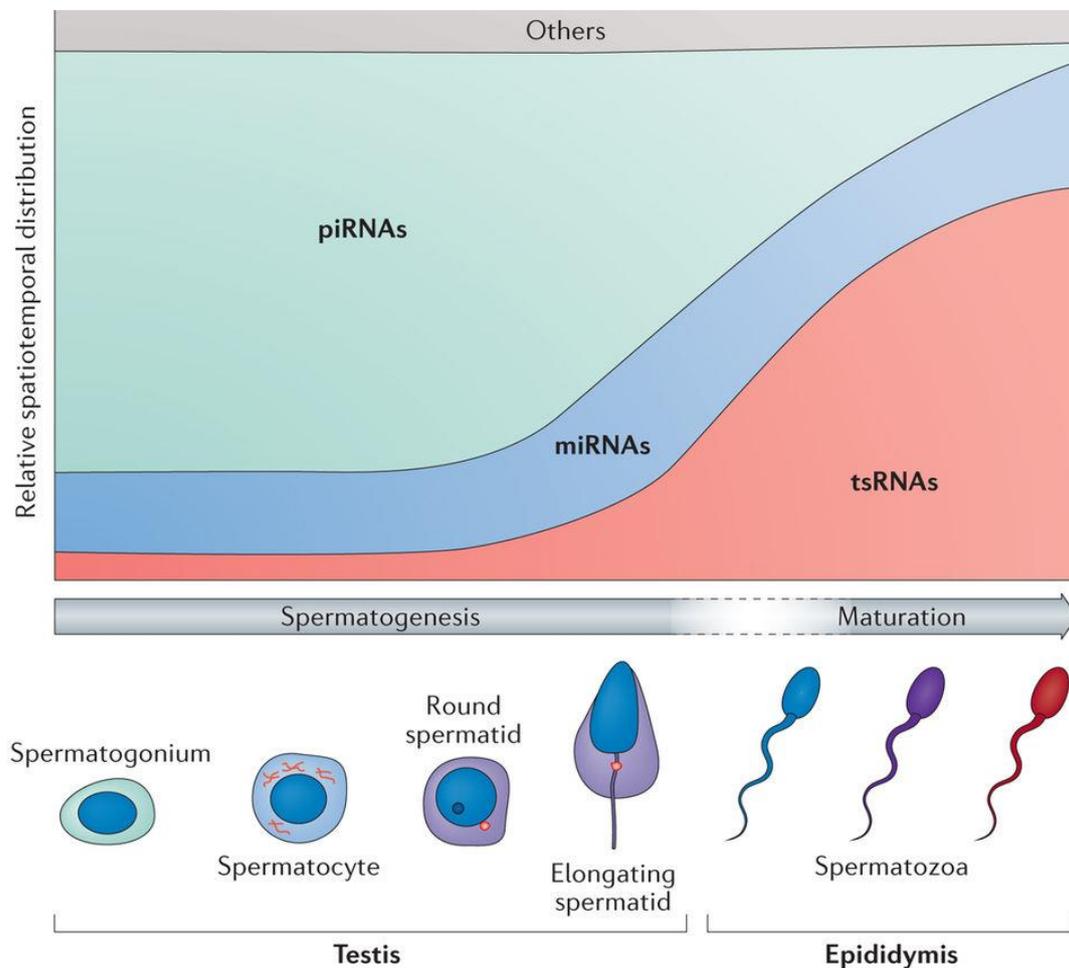


Figure 6.3. The changing spatiotemporal distribution of different types of small non-coding RNAs as cells progress through spermatogenesis. Piwi RNAs (piRNAs) are highly abundant during early spermatogenesis and in early haploid spermatids. As spermiogenesis progresses, tRNA-derived small RNAs (tsRNAs) become proportionally more abundant. Micro RNAs (miRNAs) exist as a relatively stable share of snRNAs throughout spermatogenesis. Image taken from Chen et al. (2016b)

6.1.1. Hypothesis and experimental Design

An introduction to cellular enrichment of spermatids and the process of fractionation is given in detail in Chapter 4 section 4.1. This section (section 6.1.1) deals with the miRNA investigation specific hypotheses and objectives.

Several small non-coding RNAs have been shown to affect transcription of protein-coding genes. We know that the interaction of *Sly* and *Slx* results in altered methylation of sex chromatin. If regions of the genome containing miRNAs are affected, this could result in altered expression of protein-coding genes. As miRNAs do not always directly alter transcription, but instead can affect translation of already transcribed mRNA, a mechanism facilitating MF1XY^{RIII}qdel-specific sperm and sex ratio phenotypes might not be immediately apparent when studying mRNAs alone. Substantial changes in the miRNA profile of developing spermatids between MF1XY^{RIII} and MF1XY^{RIII}qdel strains could reveal pathways enabling dysregulation of functionally important transcripts or proteins that might not otherwise appear differentially expressed in a mRNA-only examination of the transcriptome. An increasing number of studies link aberrant miRNA expression profiles to disease. Treatments for serious conditions based on miRNA therapeutics are now being trialled (though with mixed results) or are already in production (Beg et al. 2017, Christopher et al. 2016).

Micro RNAs have been shown to facilitate trans-generational effects through fertilisation without alteration of the chromatin content of the cells (Gapp et al. 2014, Rodgers et al. 2015). Trans-generational effects not linked to DNA content have been noted in other models of MSYq deletions, in which female offspring of MSYq deletion males exhibit increased progesterone production in cumulus-oocyte complexes and increased hardening of the layers surrounding the oocyte (Kotarska et al. 2014, Kotarska et al. 2013, Kotarska and Styrna 2011). Whilst this may be linked to altered DNA methylation by other mechanisms in this instance, there is a growing database of evidence for the effects of miRNA content upon subsequent generations. We have already shown differential expression of a small number of miRNAs in previous total RNA sequencing and microarray studies of expression in the MF1XY^{RIII}qdel model (Chapters 4 and 5).

Analysis of small RNAs was restricted to miRNAs; miRNAs become increasingly important as cells progress through spermiogenesis, taking over from piRNAs (see *Figure 6.3*). Unlike both piRNAs and tsRNAs, their presence is stable and prominent throughout all stages of interest within spermiogenesis (*Figure 6.3*). Research on the importance and implications of differential small RNA abundance is rapidly expanding, however miRNAs remain the best studied and characterised, with

several databases, such as miRTarBase²⁷ and miRWalk2.0²⁸, dedicated to recording and uncovering their interactions and providing a large number of experimentally validated targets (Chou et al. 2016, Dweep and Gretz 2015). Examination of miRNAs provides a focused dataset for which the best quality functional annotation can be retrieved. Restriction of the sequencing dataset to miRNAs also provides a basis for comparison to miRNA-only microarray data.

6.1.2. Chapter Objectives

The aims of this chapter are as follows:

- O1. To identify differentially expressed miRNAs across all developmental stages using between-strain (MF1XY^{RIII} and MF1XY^{RIII}qdel matched samples).
- O2. To identify differentially expressed miRNAs that may be implicated in the differential expression of shortlisted mRNAs (or their associated proteins) from total RNA sequencing (Chapter 4) and microarray (Chapter 5) experiments using databases of published miRNA interactions and annotations.
- O3. To produce a list of candidate genes that may be implicated in morphological difference and/or sex ratio distortion according to the above criteria.
- O4. To consider findings in the context of the literature and the implications for future research.

6.1.3. Notes for chapter

As was the case for Chapter 4 and Chapter 5, sample names have been abbreviated. The abbreviations are listed and explained below (*Table 6.1*). All samples in this chapter refer to either Small RNA libraries as prepared for RNA sequencing or miRNA libraries as prepared for Toray miRNA microarray expression analysis.

Full Title	Abbreviation
RNA sequencing sample collection, single replicate	R0
Microarray sample collection, replicate one	R1
Microarray sample collection, replicate two	R2
Microarray sample collection, replicate three	R3

²⁷ miRTarBase - <http://mirtarbase.mbc.nctu.edu.tw/>

²⁸ miRWalk2.0 - <http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/> and <http://129.206.7.150/>

Mouse strain		Cell type		Cell fraction	
Full name	Abbreviated	Full name	Abbreviated	Full name	Abbreviated
MF1XY ^{RIII}	WT	Early spermatids	E	Whole cell	WC
MF1XY ^{RIII} qdel	YQ	Late spermatids	L	Fraction 1	F1
		Mature sperm	M	Fraction 2	F2
				Fraction 3	F3

Table 6.1. Table of sample name abbreviations to be used throughout the chapter. These names describe sample collection number, mouse train, cell population and cellular/subcellular fraction.

6.2. Sample Preparation: Library Preparation, RNA Sample and RNA Library QC, Sequencing and Array Hybridisation of Samples

Primary sample preparation was performed as outlined for total RNA samples in previous chapters. Details of sample preparation can be found in the Methods and Materials chapter (Chapter 2): sections 2.1.2, 2.4, 2.5, 2.6.2 and are further outlined in the introductory section of Chapter 4. RNA extraction was performed from spermatid whole cell populations and subcellular fractions using the miRNeasy mini kit, for the purpose of retaining small RNAs.

Small RNA for RNA sequencing analysis

Following the extraction of total RNA (including small RNAs), library preparation specific to small RNAs was performed using the Illumina TruSeq Small RNA Library Preparation kit (Illumina, UK), as per manufacturer's instructions. RNAs were gel-selected for size, sample by sample, as recommended. This yielded 18 samples for small RNA sequencing.

All Small RNA sequencing libraries were assessed for concentration and quality using the Agilent Bioanalyzer 2100 and a High Sensitivity DNA Assay chip as recommended in the Illumina Small RNA library Preparation protocol. All samples appeared of an acceptable quality and abundant enough in RNAs to continue with processing. Single-end RNA sequencing of Illumina small RNA libraries was conducted using a NextSeq 500 (Illumina, Chesterford, UK). Two sequencing runs of nine samples (totalling eighteen samples) were conducted by Cambridge Genomic Services.

RNA for miRNA microarray expression analysis

Labelling of RNA samples for miRNA analysis was performed using the 3D-Gene miRNA labelling kit (Toray, Tokyo, Japan) as per manufacturer's instructions. Toray miRNA labelling preparation was performed by Cambridge Genomics Services (Department of Pathology, University of Cambridge). An additional two RNA samples from sample set R0 were also processed to produce miRNA microarray samples. This resulted in 56 samples for miRNA microarray analysis.

Quality control of miRNA microarray libraries was performed by Cambridge Genomics Services (Department of Pathology, University of Cambridge). All samples appeared of an acceptable quality and abundant enough in RNAs to continue with processing. The Toray 3D-gene system utilised for the work in this thesis offers multi- and single species chips for miRNA analysis; the mouse specific chip (Toray version 21 miRNA microarray chip) contains 1900 unique miRNA mounted probes.

6.3.Preparation of Small RNA Sequencing Data

Quality control of samples and differential gene expression was performed using a customised pipeline developed by Cambridge Genomics Services (Department of Pathology, University of Cambridge) in consultation with the author. Analysis scripts were produced by Maria Gomez (Cambridge Genomic Services). An overview of data preparation and the analysis pipeline can be seen in *Figure 6.4*. The number of reads generated per sample and subsequently entered for analysis was first measured across the dataset. *Figure 6.5* shows the percentage of reads originating from each sample as a proportion of total yield across the entire dataset. This provides indications of how subsequent normalisation techniques might affect individual samples.

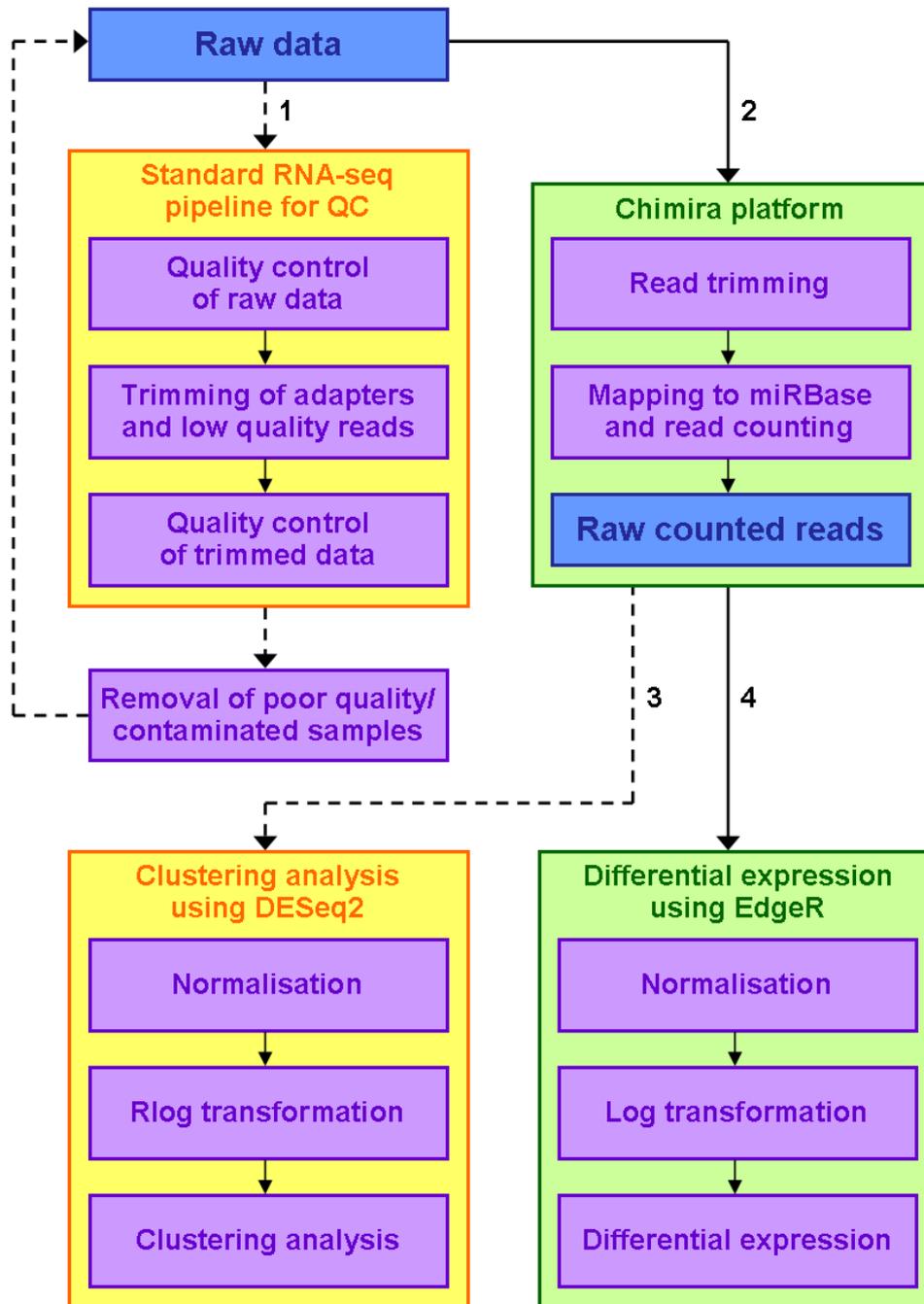


Figure 6.4. Overview of Small RNA sequencing data preparation and analysis pipeline. Numbers represent two distinct phases of analysis. Dashed line indicates movement through QC and data exploration steps (phase 1 and 3) and removal of the equivalent poor quality or contaminated RNA samples (in original raw data) before processing of the data (in phase 2 and 4).

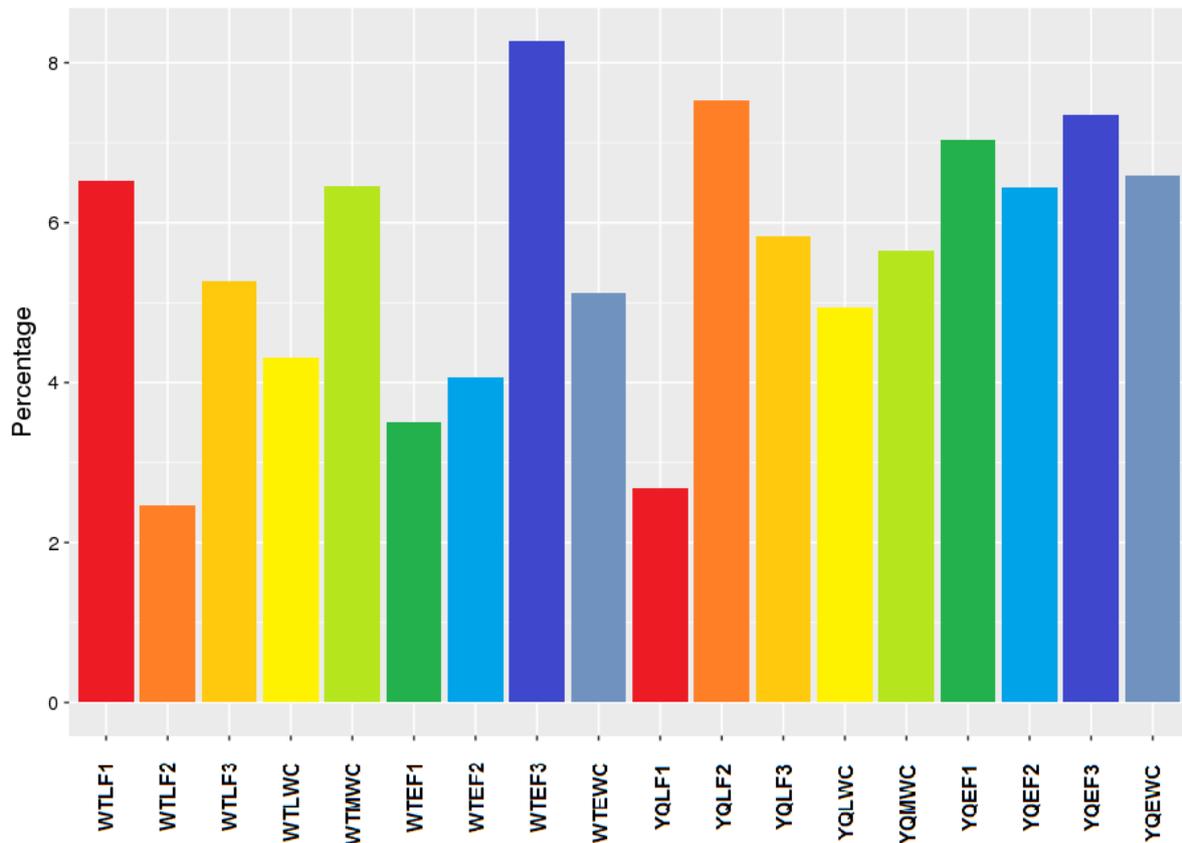


Figure 6.5. Reads retrieved from each sample as a percentage of the total number of reads retrieved across the whole dataset. Colours indicate matched samples across the two mouse strains (WT and YQ). Samples were moderately well balanced, with the greatest difference between matched samples seen in WT- and YQLF2.

6.3.1. Raw data quality control

The quality control of the reads was performed using FastQC v.0.11.4. (Babraham Bioinformatics 2011). Per base quality scores were generated to assess read quality of individual sample reads. Examination of plotted scores revealed reduced quality of a single sample, WTEWC (*Figure 6.6*). *Figure 6.7* shows means quality scores across all samples and demonstrates that this sample produced a lower mean quality than other samples. This may be due to sample degradation or faults during sequencing. Read bases with a quality score of less than 20 are trimmed at the subsequent read trimming step (adapters and low quality trimming) prior to analysis. All samples were re-assessed following read trimming.

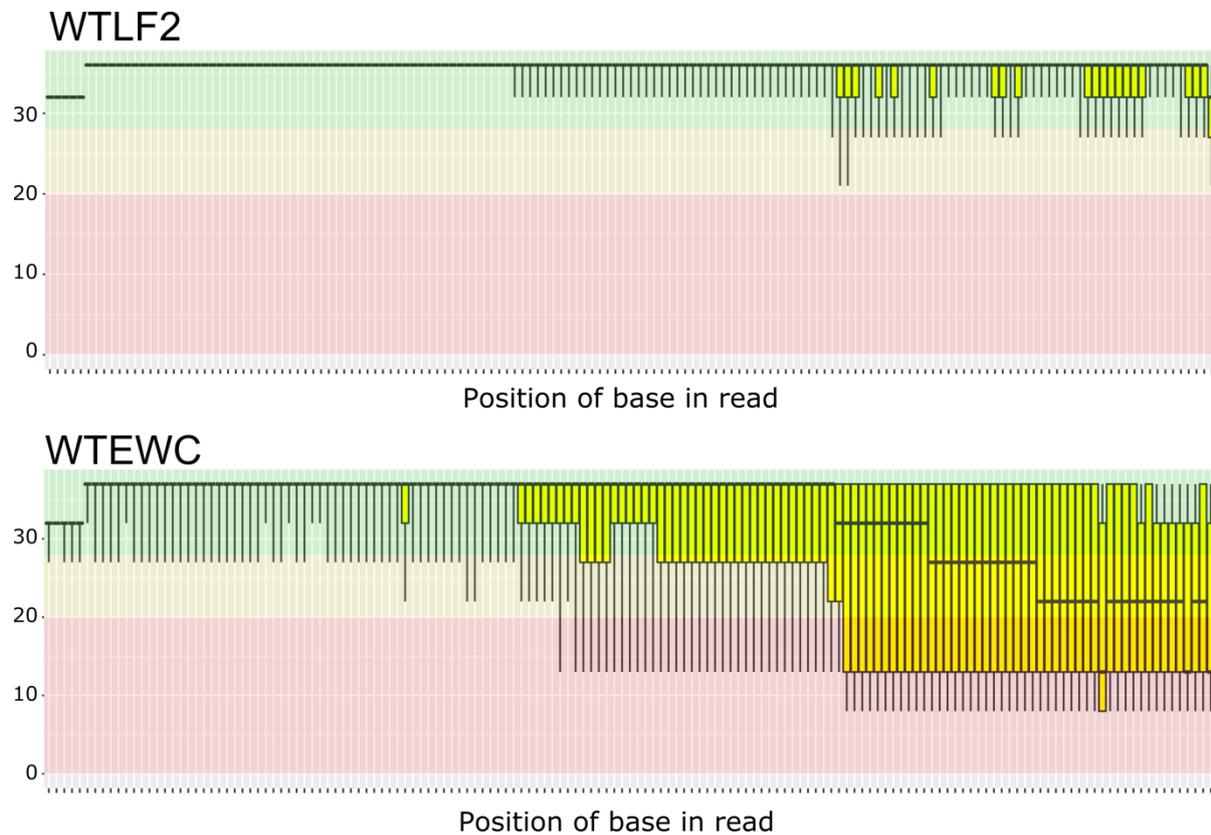


Figure 6.6. Boxplots of quality (Q) values distributed across each read of the sample examined. The y-axis denotes quality values representing expected error rate. Base position numbers are not given; they extend from position '1' on the extreme left of the x-axis to '151' on the extreme right of the same axis. Sample WTLF2 (top) provides an example of overall high read quality, with a low error rate. Sample WTEWC (bottom) was identified as possessing overall lower quality reads following examination of per base quality scores. Images were adapted from those taken directly from small RNA sequencing output as produced by Cambridge Genomics Services.

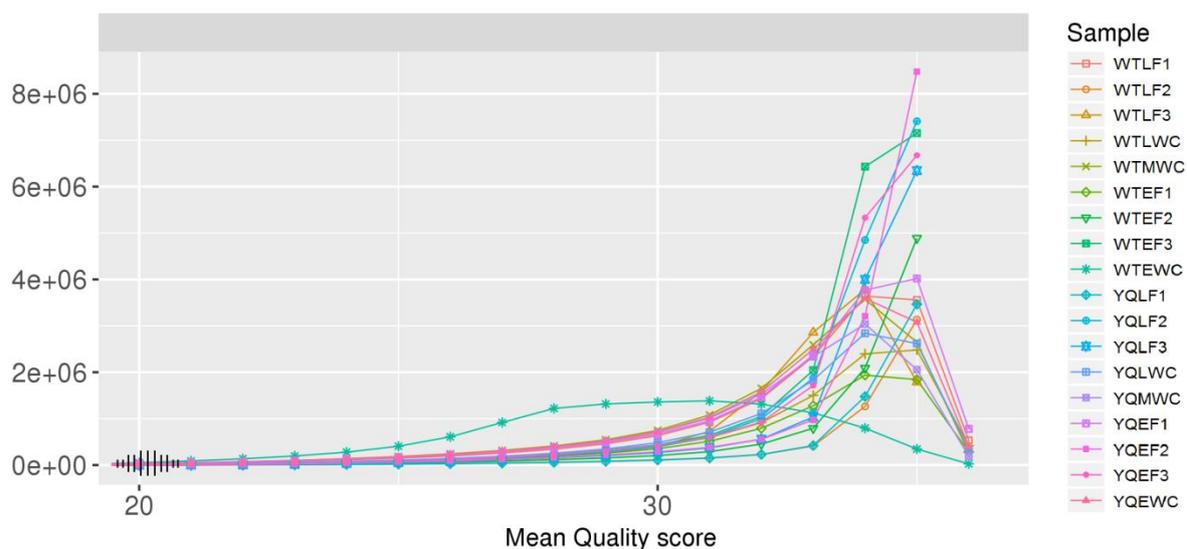


Figure 6.7. Mean quality score across all small RNA samples. Sample WTEWC has a greater proportion of reads with a quality score below 30. Sample reads are expected to show a mean quality score distribution that peaks at or after Q30 to obtain acceptable read reliability. The x-axis has been shortened in this image. The image was adapted from that taken directly from small RNA sequencing output as produced by Cambridge Genomics Services.

6.3.2. Trimming of low quality bases and adapters

Reads were trimmed using TrimGalore v. 0.4.1²⁹. TrimGalore removes low quality base calls from the 3' end of the read towards the 5' end prior to adapter removal. TrimGalore will continue removing bases until the quality base call exceeds Q20. The percentage of complete reads removed was less than five percent per sample. WTEWC did not show a greater proportion of reads removed than other samples, suggesting that base trimming was achieved without significant loss of reads. In order to check subsequent read quality, per sequence quality scores were examined following trimming. *Figure 6.8* shows that read trimming has removed bases showing poor quality scores across all samples without significant read loss. Reads from sample WTEWC are now seen to be of acceptable quality to sensibly progress to read counting and data normalisation.

²⁹ Trim Galore! By Babraham Bioinformatics - http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

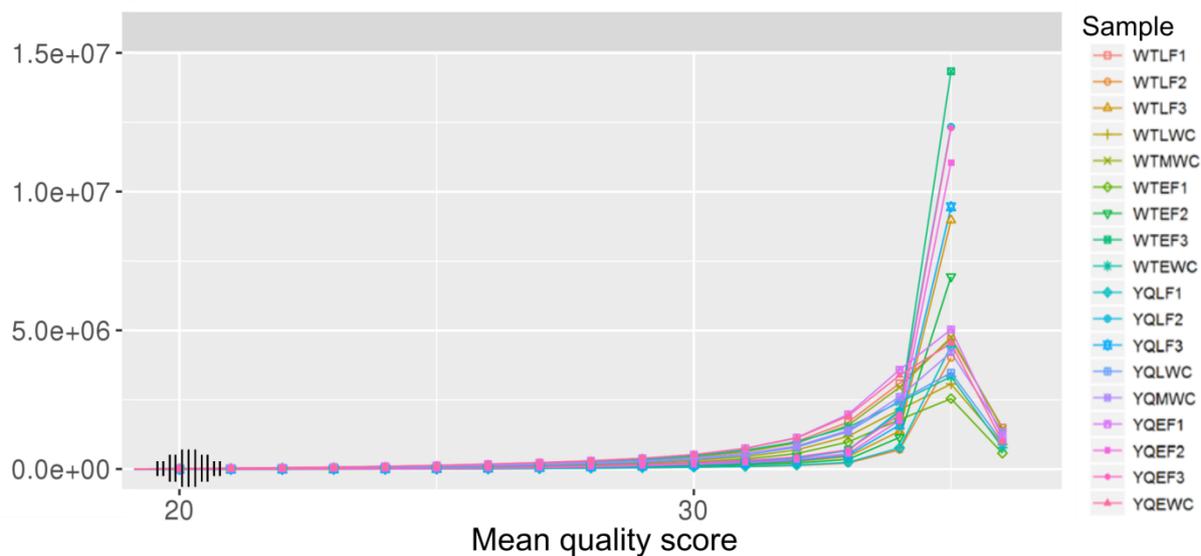


Figure 6.8. Mean quality score across all small RNA samples post-trimming. The x-axis has been shortened in this image.

6.3.3. Mapping, counting, normalisation and differential expression analysis

Mapping and counting of reads using the Chimira platform

Chimira v.1.0³⁰ is a web-based tool for the analysis of small RNA sequencing data (Vitsios and Enright 2015). It facilitates the mapping of miRNA reads to miRNA database miRBase³¹ using either trimmed or untrimmed reads to determine miRNA content of samples. Chimira selects only for miRNAs. Mapping and counting of reads was performed in Chimira.

Normalisation and post-normalisation quality control in DESeq2

Normalisation and subsequent quality control was performed in DESeq2 release 3.5 (Love et al. 2014). Applying the rlog transformation to the data in DESeq2 shrinks the variability of the lowest expressed miRNAs to produce principal components analysis that is more representative of overall miRNA content rather than the variation between individual genes which may or may not make up a substantial proportion of total miRNA composition.

Principal components analysis of data pre- and post-normalisation

Principal components analysis (PCA) provides a convenient method for visualising differences between multi-dimensional data sets (a detailed explanation of PCA is available in section 4.3.4).

³⁰ Chimira - <http://wwwdev.ebi.ac.uk/enright-dev/chimira/index.php>

³¹ Micro RNA database (miRBase) - <http://www.mirbase.org/>

Examination of first and second principal components revealed cell population and subcellular fraction as the primary source of variation in the samples (*Figure 6.9*). Whole cell samples lie within distributions of subcellular fractions as anticipated. This was also seen following examination of the sample similarity heatmap (*Figure 6.10*), where all samples clustered primarily by cell type, and secondarily by subcellular fraction except in the case of spermatid cell fraction 3 samples, which lie close to each other and mature sperm samples.

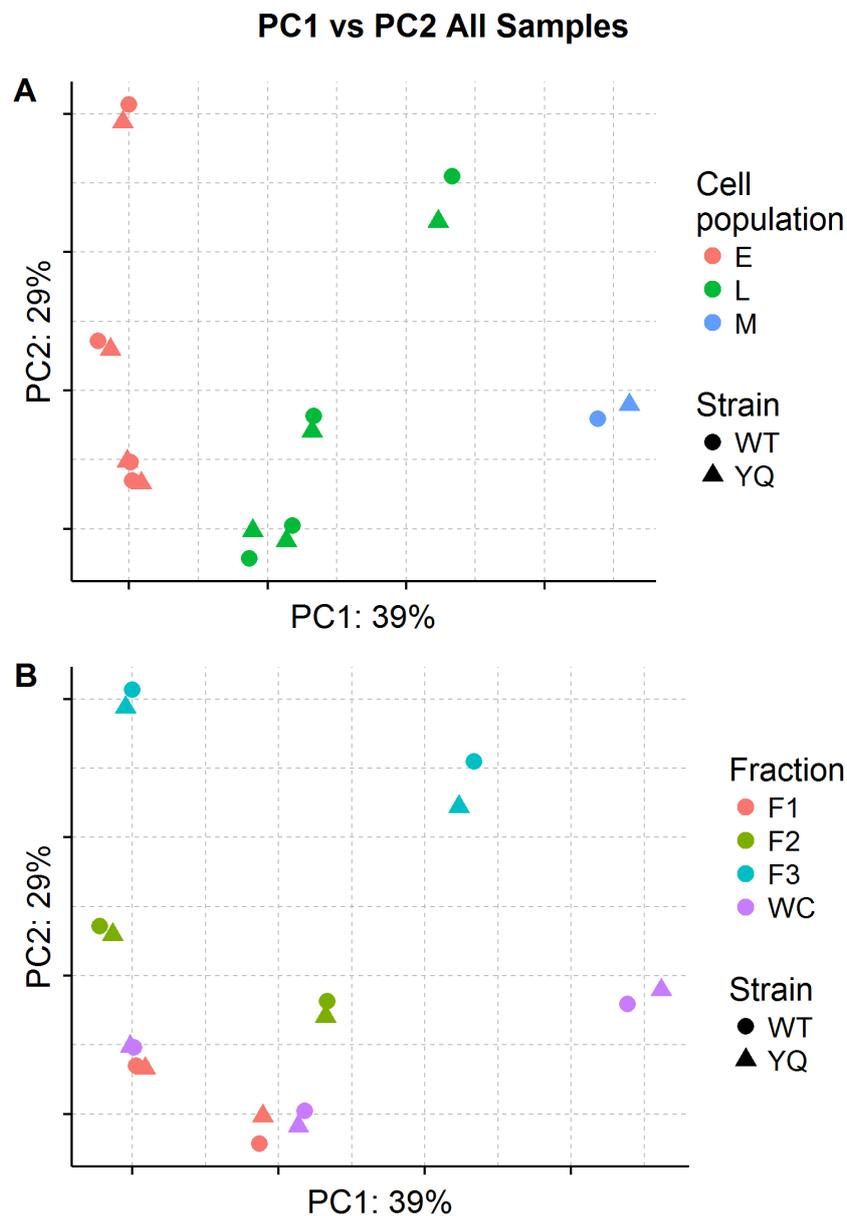


Figure 6.9. PCA plot of all samples post-normalisation and transformation in DESeq2. This plot uses first and second principle components, which represent 68% of variance in total.

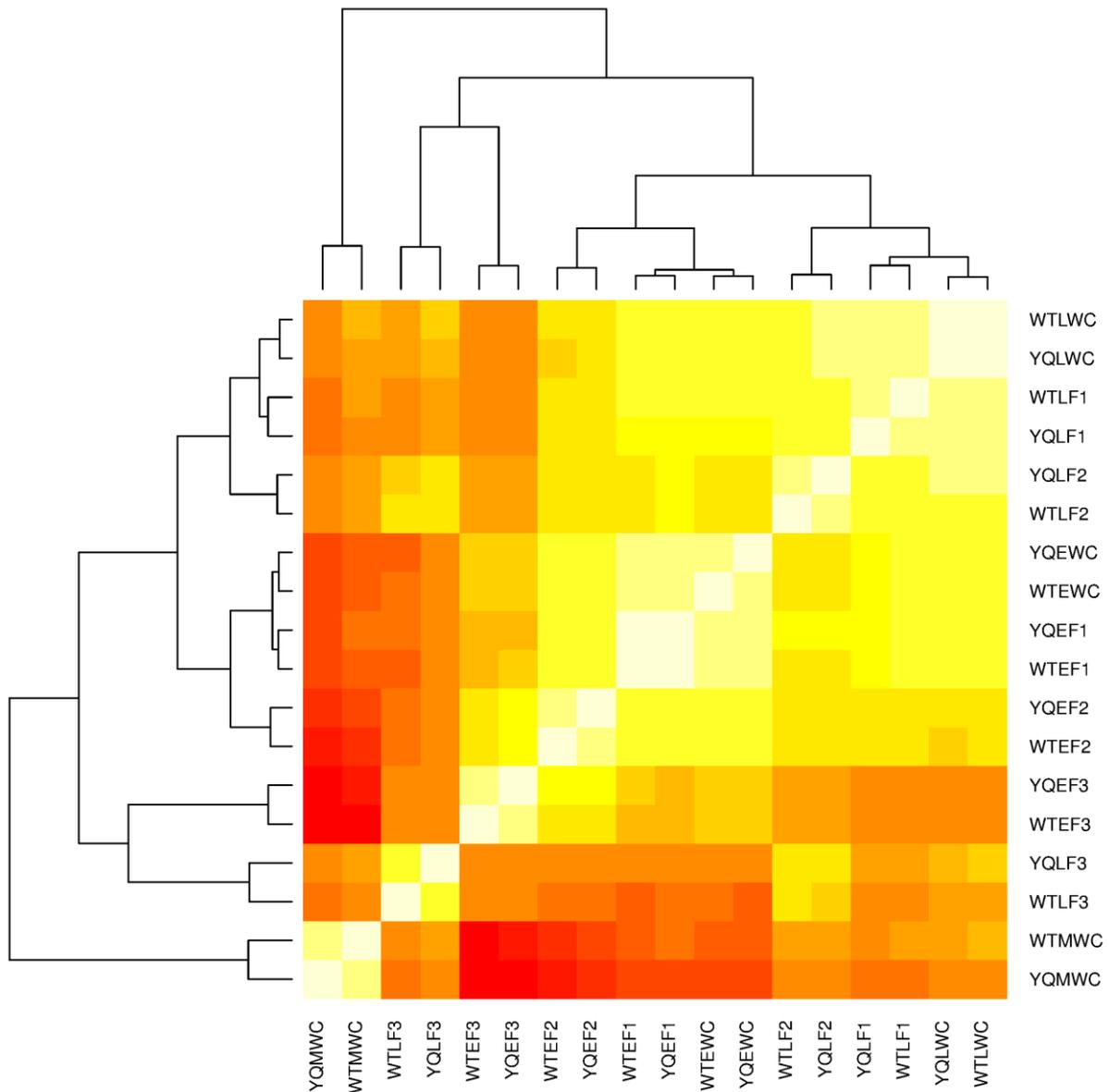


Figure 6.10. Heatmap comparing read composition for similarity, as determined by Pearson correlation, across WT and YQ samples. Samples cluster primarily by cell type, and secondarily by subcellular fraction except in the case of early and late spermatid fraction 3 samples, which lie close to each other and mature sperm samples.

Differential expression analysis using EdgeR

Counted reads were again loaded into R and subsequently normalised and log transformed using edgeR (Robinson et al. 2010). Counts are converted to counts per million (cpm) values as described in section 4.3.3. The log fold change value (logFC) is calculated per comparison using the normalised-log cpm values.

6.4. Results of Small RNA Sequencing Analysis

Differential expression analysis produced logFC values for each detected gene between matched samples (e.g. WTEF1 and YQEF1). The dataset used in small RNA sequencing is single replicate, meaning that significance values were not available to distinguish between differentially expressed and variable miRNAs. It is not possible to assess all miRNAs detected, however in these instances a logFC value can be used to substitute as a cut off for inclusion in further analysis. A logFC value of magnitude 3 was chosen as this produced a list of genes of a manageable size for detailed subsequent analysis. The number of miRNAs shortlisted with logFC values ≥ 3 or ≤ -3 is shown below in *Table 6.2*. Further information on these miRNAs is given in *Table 6.3*.

	Number of miRNAs showing altered expression
Sample group	logFC ≥ 3
LFC1	3
LFC2	1
LFC3	1
LWC	0
MWC	1
EF1	0
EF2	0
EF3	0
EWC	1

Table 6.2. The number of genes showing logFC values of a magnitude equal to or greater than given value. Negative values have been converted to positive values for inclusion in this table.

	LogFC >= 3 or <=-3	
Comparison	miRNA Symbol	LogFC value
WTLF1 vs YQLF1	mmu-mir-1961	4.37
	mmu-mir-145b	4.17
	mmu-mir-465d-5p	-4.07
WTLF2 vs YQLF2	mmu-mir-5108	3.06
WTLF3 vs YQLF3	mmu-mir-145b	3.08
WTMWC vs YQMWC	mmu-mir-5118	-3.34
WTEWC vs YQEWC	mmu-mir-145b	-5.11

Table 6.3. Details of miRNAs shortlisted for subsequent analysis across all sample comparisons from small RNA sequencing data.

6.5.Preparation of miRNA Microarray Data

Quality control of samples and differential gene expression was performed using a customised R pipeline developed by Cambridge Genomics Services in consultation with the author. Analysis scripts were produced by Maria Gomez (Cambridge Genomic Services). These scripts were passed to the author for subsequent customisation. Data preparation and differential expression analysis followed a similar overall workflow as total RNA microarray expression analysis (Chapter 5). The workflow diagram has been reproduced below (*Figure 6.11*). Different normalisation and annotation methods were used for miRNA analysis; these are explained in subsequent sections.

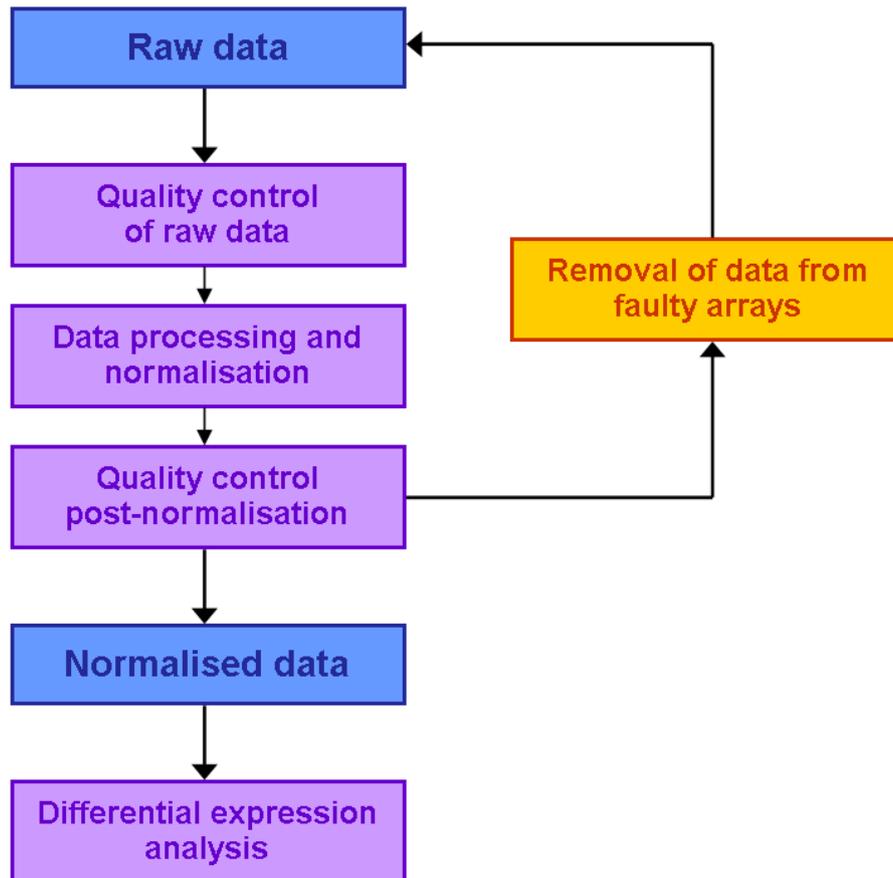


Figure 6.11. Simplified overview of miRNA data processing pipeline following collection of raw data.

6.5.1. Micro RNA microarray raw data quality control

Background (technical) noise is calculated using empty probe spaces. This background value is calculated per sample (per array), and is subtracted from all probe fluorescence intensity values before data normalisation is performed. Raw data quality control revealed poor adherence of several samples to replicate expression distribution and sample diversity. Scatterplots of raw data are shown below (*Figure 6.12*). Data normalisation was performed before removal of samples to ensure samples are not discarded unnecessarily; data from all arrays was included in initial normalisation and subsequent quality control.

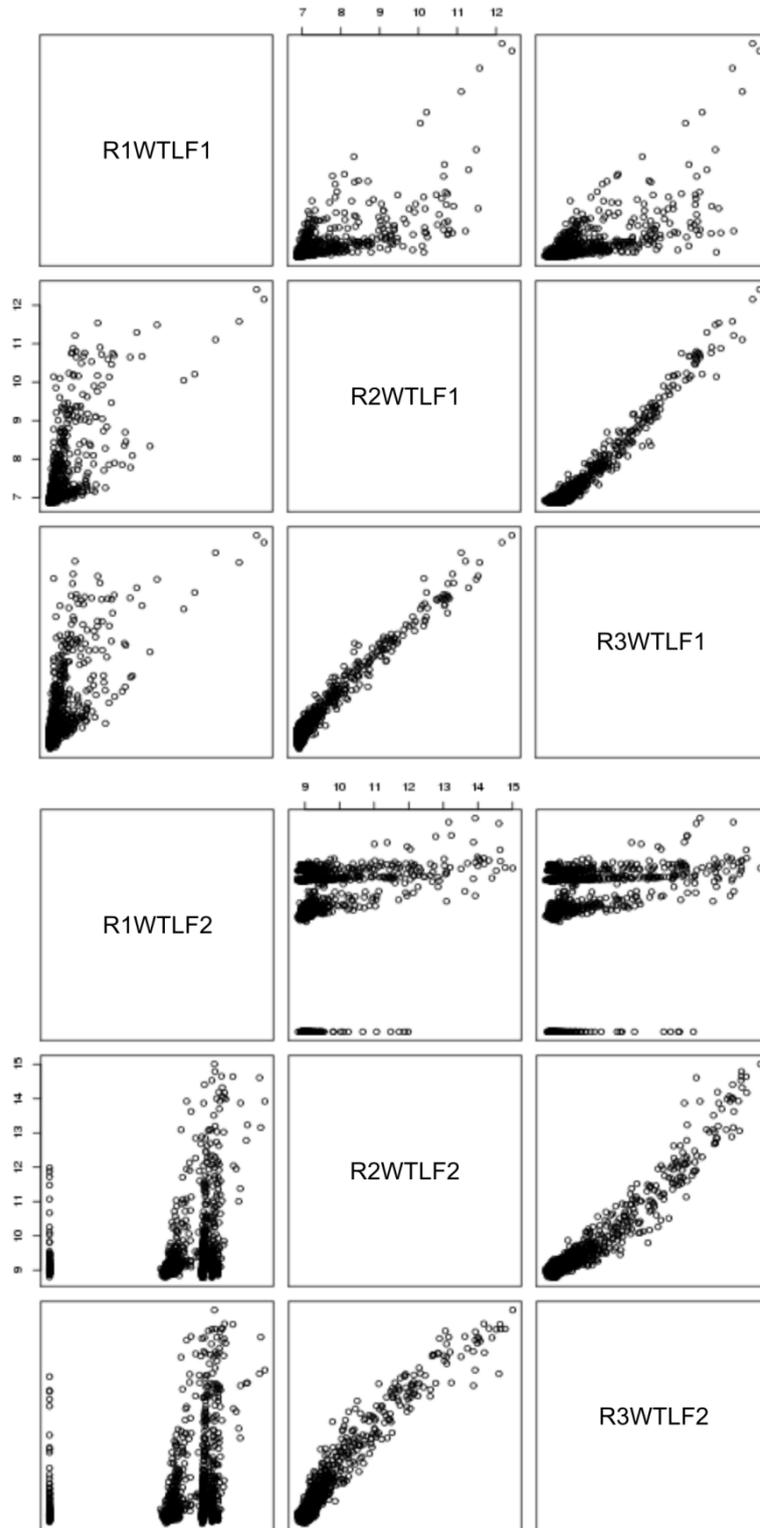


Figure 6.12. Scatterplots of raw detected miRNA content prior to background subtraction from three separate arrays of the same sample type. Two samples are shown; WTLF1 (top) and WTLF2 (bottom). Two aligned replicates can be seen per sample group and one distinctly different sample.

6.5.2. Normalisation and post-normalisation quality control

Following background correction, quantile normalisation was used across all comparisons using the ‘normaliseBetweenArrays’ function from the Limma package (Ritchie et al. 2015) in R. Quantile normalisation is a standard normalisation technique used in microarray data analysis. The aim of quantile normalisation is to standardise the distribution of probe intensities for each array in a set of arrays to allow for more reliable and representative comparison between samples. *Figure 6.13* gives a simple graphical representation of the outcome of quantile normalisation. A clear and detailed explanation of quantile normalisation (and other common array normalisation methods) can be found in Bolstad et al. (2003).

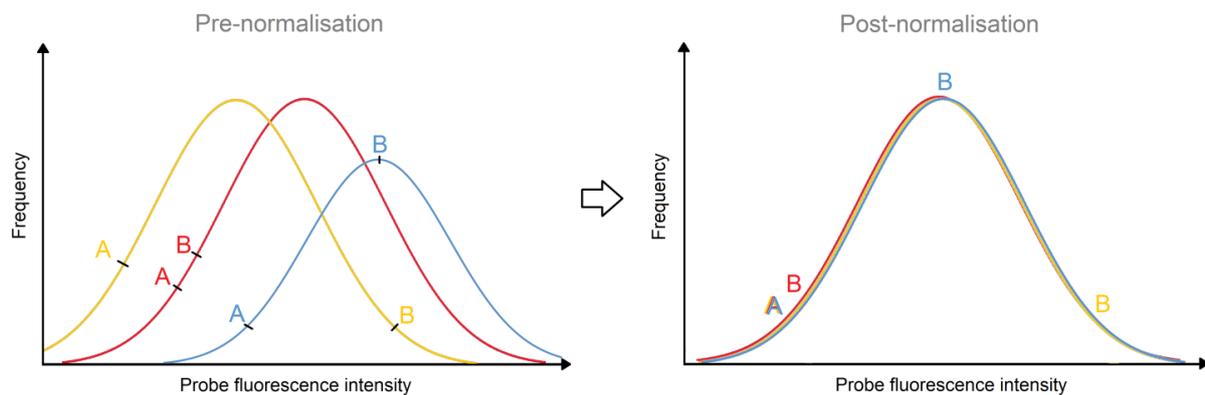


Figure 6.13. Simple representation of the effect of quantile normalisation across arrays. Curves of different colours represent probe fluorescence intensity distributions across three arrays (three separate samples). Examples of genes showing similar (A) and dissimilar (B) resulting levels of expression are given as examples.

Following normalisation, a simple visual comparison of sample composition was performed using scatterplots of detected miRNA content (similar to those in *Figure 6.12*). Following identification of 16 apparently aberrant samples, it was discovered that 12 of the samples represented all occupants of three separate microarray chips (four on each chip). This strongly suggested that these three chips were either defective or problems were encountered during array processing that cannot be corrected at this stage. The four remaining samples which were identified as differing considerably from other replicates were not all processed on the same chip, implying that problems relating to the array itself are not the cause. The additional R0 (additional replicate) mature sperm samples are RNA samples originally prepared for the pilot RNA sequencing experiment (Chapter 4) and therefore were older and had been thawed and refrozen more times than samples derived specifically for the microarray. This may have been a cause of some discrepancy. No obvious reason can be determined for the nature of the remaining two samples R3YQEF1 and R3YQEF2. It is possible that the two were mislabelled at some point during processing, though they bear little resemblance to the other’s replicates either.

Nonetheless, it is evident that they do not reasonably concur with the remaining two replicates of each sample, which can be seen to be highly similar to one another. For this reason the samples listed in *Table 6.4* were excluded from subsequent analyses. A minimum of two replicates remained for each sample group examined. Four of eighteen samples (WTEF3, YQEF3, WTLF3 and YQLF3) retained all three original replicates. This exclusion results in a reduction of statistical power, but is greatly preferable to including potentially aberrant samples. Normalisation was repeated following the removal of these samples, as reported in section 6.5.2.

Sample name	Chip no.	Sample name	Chip no.
R1WTLF1	09	R1YQLF1	09
R1WTLF2	09	R1YQLF2	09
R3YQEW C	17	R3WTEWC	17
R3WTLWC	17	R3YQLWC	17
R3WTEF1	18	R3WTMWC	18
R3WTEF2	18	R3YQMWC	18
R3YQEF1	19	R0YQMWC	22
R3YQEF2	19	R0WTMWC	22

Table 6.4. A list of aberrant samples and the chips that they were hybridised to. These samples were excluded from further analyses.

Following removal of aberrant samples listed above (*Table 6.4*), all array data processing was repeated. Principal components analysis (PCA) was performed pre- and post-normalisation to determine the major sources of variance between samples. *Figure 6.14* shows PCA plots of samples post-normalisation. Plots of principal components 2 and 3 were also included as these represented a higher proportion of overall variance than in previous investigations. Samples showed separation by both cell population type and subcellular fraction. For a more detailed explanation of PCA, see section 4.3.4.

PCA of data post-normalisation

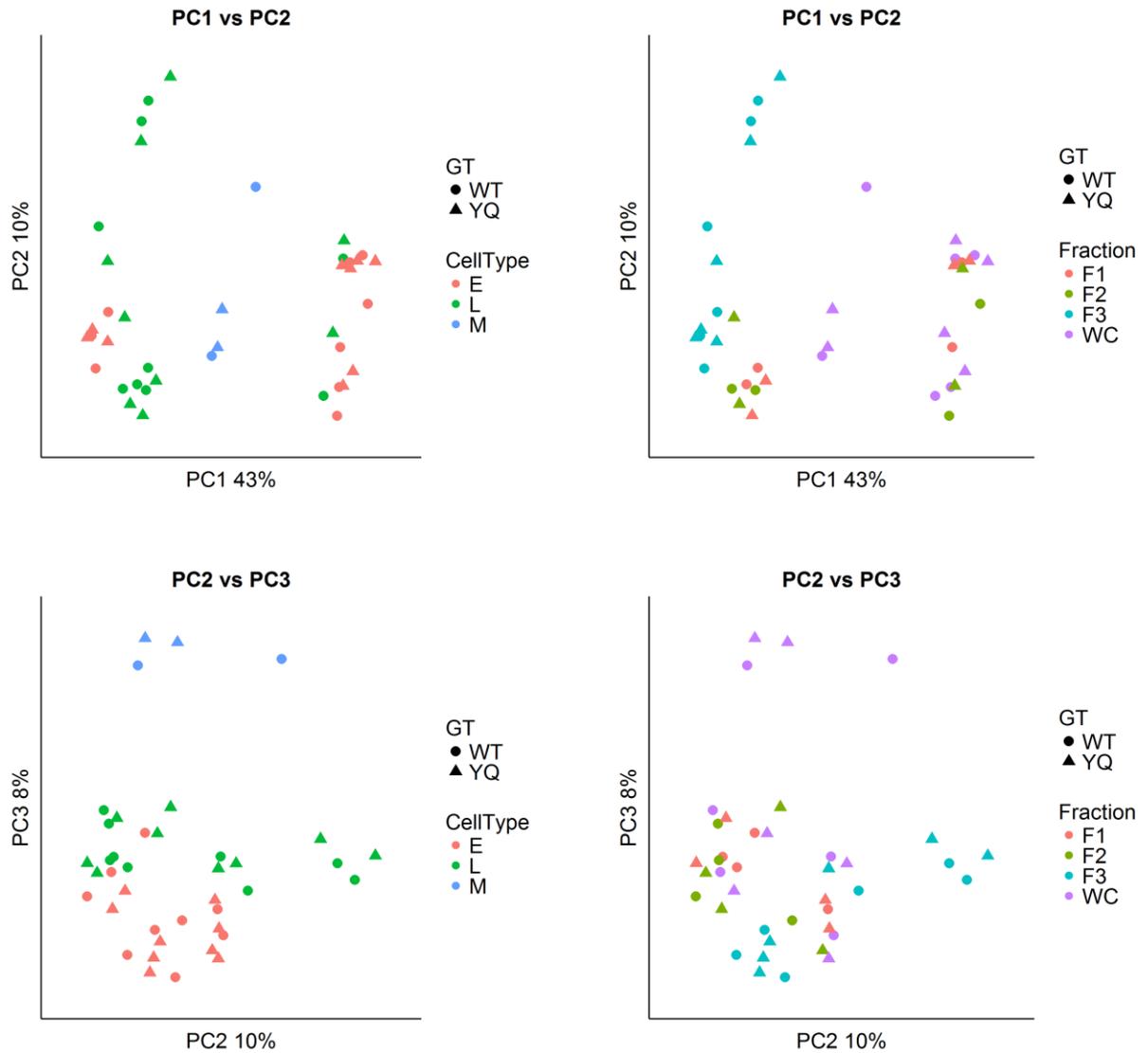


Figure 6.14. PCA plots of all miRNA microarray samples post-normalisation. Following normalisation, samples show separation by both cell type and subcellular fraction. Fraction 3 showed the greatest distinction from other samples; mature sperm were also separated from other cell populations by the primary principal component (PC1). Cell types are more easily distinguished along PC3, suggesting a combination of cell type and subcellular fraction make up major sources of variation in the miRNA dataset. Fraction 3 of early spermatids shows strong similarity to late stage spermatid samples.

6.6. Results of miRNA Microarray Expression Analysis

Following differential expression analysis as described, only two sample comparisons of nine conducted produced miRNAs showing significant differential expression. All significantly differentially expressed miRNAs are listed in *Table 6.5*. As in section 6.4, an examination of differentially expressed miRNAs and their involvement in known biological pathways was conducted and is described in more detail in section 6.7.

miRNAs showing significant differential expression			
Comparison	miRNA symbol	logFC	P.value
WTEF1 vs YQEF1	mmu-mir-466h-3p	6.04	0.012
	mmu-mir-3067-5p	5.78	0.012
	mmu-mir-7028-5p	5.82	0.012
	mmu-mir-6907-5p	-3.64	0.018
	mmu-mir-7020-5p	6.11	0.020
WTEWC vs YQEWC	mmu-mir-7216-5p	-4.55	0.027
	mmu-mir-7011-5p	-4.41	0.027
	mmu-mir-504-3p	4.26	0.027
	mmu-mir-467g	-4.50	0.027
	mmu-mir-5131	4.33	0.027
	mmu-mir-6916-5p	-4.62	0.027
	mmu-mir-148a-3p	4.22	0.027
	mmu-mir-26b-5p	4.65	0.027
	mmu-mir-470-5p	4.26	0.029
	mmu-mir-1198-5p	4.36	0.031
	mmu-mir-7020-5p	4.50	0.031
	mmu-mir-7082-5p	-4.65	0.031
	mmu-mir-466m-3p	4.06	0.031
	mmu-mir-669a-5p, mmu-mir-669p-5p	-4.67	0.050

Table 6.5. A list of miRNAs showing significant differential expression across all sample comparisons. Only comparison between EF1 samples and EWC samples yielded any significantly differentially expressed genes.

6.7. Further Examination of Differentially Expressed miRNAs

Small RNA sequencing and miRNA microarray technologies were used to produce separate assessments of differences in miRNA abundance between WT and YQ spermatids and mature sperm in three whole cell samples and six subcellular fraction samples. Both technologies retrieved low numbers of differentially expressed genes. A significance value of $p < 0.05$ was used to determine differential expression of miRNAs in microarray samples, whereas a logFC change value of ≥ 3 or ≤ -3 was used to select miRNAs across small RNA sequencing samples. The results of each investigation are further investigated in this section.

MiRTarBase v.6.0³² is a searchable database containing experimentally validated curated miRNA-target interactions (MTIs) for a number of species (Chou et al. 2016). It holds information on 370,211 individual MTIs; 41,588 of which are from mouse alone. The most recent release (v.6.0) saw the number of validated MTIs rise 7-fold from just over 50,000. The database provides the option to download a file containing all curated mRNA-miRNA interactions in a species specific manner alongside associated PubMed references of published validation experiments. The mouse-specific file (release 6.1) was downloaded from miRTarBase³³. A script produced in R (available in the online repository³⁴) was used to manipulate the dataframe to recover information associated with differentially expressed miRNAs from the above described experimental work (sections 6.4 and 6.6). A large number of MTIs were retrieved for the differentially expressed miRNAs. These were subsequently matched with mRNAs shown to be differentially expressed in previous RNA sequencing and microarray experiments in chapters 4 and 5; this produced two interacting mRNAs (Table 6.6).

miRNA	Target gene	Validation	miRNA source	mRNA source
mmu-miR-1198-5p	<i>Trio</i>	HITS-CLIP	miRNA array	RNA seq
mmu-miR-466m-3p	<i>Piga</i>	HITS-CLIP	miRNA array	microarray

Table 6.6. Differentially expressed miRNAs returning experimentally-validated interactions with shortlisted differentially expressed mRNAs from RNA sequencing and microarray chapters. Both miRNAs listed were shown to be differentially expressed in miRNA microarray data.

³² Micro RNA Target Interactions Database (MiRTarBase) v.6.0 - <http://mirtarbase.mbc.nctu.edu.tw/index.php>

³³ Mouse-specific miRNA-target interactions download, available at <http://mirtarbase.mbc.nctu.edu.tw/php/download.php>

³⁴ Thesis online repository – <http://www.github.com/EmmaEPJ/Repository>

Trio is an autosomal gene located on chromosome 15. It was substantially down-regulated in YQ in both early spermatid whole cell and fraction 3 samples (-1.33 and -1.82 LogFC respectively) in RNA sequencing data. *Trio* also showed potential cis-limitation in both RNA and microarray data. *Piga* is an X-linked phosphatidylinositol glycan anchor biosynthesis gene. It showed no evidence of cis-limitation, but like many other X-linked genes was up-regulated in the YQ late spermatid population. These interactions are further discussed in section 6.8.

All listed messenger RNA targets of differentially expressed miRNAs, not restricted to mRNAs shortlisted in previous chapters, were examined for spermatid expression using MGI Gene Expression Database (GEX)³⁵. Although far from comprehensive, this database provided confirmation of the spermatid-specific expression of several associated mRNAs (Table 6.7).

Gene symbol	Gene name	Interacting miRNA
<i>Tdrd1</i>	Tudor domain containing 1	mmu-miR-5108

Gene symbol	Gene name	Interacting miRNA
<i>Kpna2</i>	Karyopherin (importin) alpha 2	mmu-miR-26b-5p
<i>Pou5f1</i>	POU domain, class 5, transcription factor 1	mmu-miR-470-5p
<i>Piwi2</i>	Piwi-like homolog 2 (Drosophila)	mmu-miR-467g, mmu-miR-466m-3p
<i>Acvr2a</i>	Activin receptor IIA	mmu-miR-466m-3p
<i>Tdrd1</i>	Tudor domain containing 1	mmu-miR-467g, mmu-miR-466m-3p

Table 6.7. Spermatid-expressed mRNAs showing interaction with differentially expressed miRNAs retrieved from small RNA sequencing data (top) and miRNA microrarray data (bottom). All listed interactions were retrieved from miRTarBase.

A second analysis was performed using mirPath v.3.0³⁶ from DIANA tools (Vlachos et al. 2015) to examine known miRNA-pathway interactions. DIANA-mirPath v.3.0 is a miRNA pathway web-server that utilises predicted and experimentally validated miRNA targets (validated targets are collated from miRTarBase) alongside KEGG pathways database³⁷ (Kanehisa et al. 2017, Kanehisa et al. 2016, Kanehisa et al. 2000) to allow for the discovery of enriched pathways based on a list of

³⁵ MGI Gene Expression Database (GEX) - <http://www.informatics.jax.org/expression.shtml>

³⁶ DIANA-mirPath v.3.0 - <http://snf-515788.vm.okeanos.grnet.gr/>

³⁷ Kyoto Encyclopedia of Gene and Genomics (KEGG) pathways database - <http://www.genome.jp/kegg/>

miRNAs as an input. This allowed for the identification of biological pathways for which one or more of the differentially expressed miRNAs is involved or predicted to be involved. Pathways of particular interest from microarray detected miRNA-interactions included ‘ECM-receptor interaction’ (ECM = extra cellular matrix) and ‘Wnt signalling pathway’. These were the top two enriched pathways as determined by a combination of predicted and experimentally-validated MITs. Examination of RNA sequencing shortlisted miRNAs also returned ‘ECM-receptor interaction’ in the top two enriched pathways by the same method.

6.8. Chapter Discussion

This chapter covers work aimed at highlighting miRNAs that are differentially expressed in the post-meiotic spermatids and mature sperm of MF1XY^{RIII}qdel mice. The investigation sought to determine whether miRNA-mediated regulation of genes and associated mRNAs might be involved in facilitating various phenotypes observed in mice with substantial MSYq deletions. By using two distinct methods of differential expression analysis, small RNA sequencing and miRNA microarray, differentially expressed miRNAs were identified for further examination of interactions with protein-coding genes. Eighteen individual miRNAs were shown to be differentially expressed across all samples in miRNA microarray data, and five miRNAs were selected for further investigation based on LogFC values between WT and YQ from RNA seq data. Experimentally-backed database miRTarBase v.6.0³⁸ (Chou et al. 2016), alongside miRNA investigation web-server DIANA-mirPath v.3.0 were used to examine miRNA involvement in potentially implicated pathways and interaction with candidate genes previously identified in chapters 4 and 5, in both sex ratio distortion and other observed reproductive abnormalities. A few of the short-listed miRNAs revealed potentially interesting interactions in a testis-specific context.

Interaction of upregulated *mmu-miR-1198-5p* and down-regulated mRNA *Trio*

Through examination of miRNA-target interactions in miRTarBase in combination with customised cross-examination with dysregulated mRNAs retrieved from previous chapters, *Trio* was shown to be an interacting partner of up-regulated *mmu-miR-1198-5p*. Additional annotation using miRBase (Kozomara and Griffiths-Jones 2014, Griffith-Jones 2004) revealed *mmu-miR-1198* (stem loop) to be X-linked. *Trio* is an autosomal gene that shows substantial down-regulation in YQ early spermatids in RNA seq data compared with WT; its interacting partner *miR-1198-5p* shows substantial upregulation (LogFC +4.36) in the same sample type. This suggests *miR-1198-5p* may be inhibiting transcription

³⁸ MicroRNA-target interactions database (miRTarBase) - <http://mirtarbase.mbc.nctu.edu.tw/>

of its partner *Trio*. Interestingly, the X-linked *miR-1198-5p* has been shown to belong to an miRNA cluster that has undergone recent expansion in the mouse genome (Zheng et al. 2011), which might reveal a selective influence much like that seen for sex-linked *Sly* and *Slx* as part of the intragenomic conflict (Cocquet et al. 2012). *MiR-1198-5p* is known to be testis expressed and in a study by Nixon et al. (2015) showed reducing expression as sperm progressed through epididymal maturation. *Trio* is Triple function domain protein (PTPRF interacting), and, amongst other functions, coordinates actin remodelling and is a guanine nucleotide exchange factor for RHOA and RAC1 GTPases (UniProt, by similarity with rat and human homologues), as discovered during an examination of TRIO loss of function and associated intellectual disability in rats (Ba et al. 2016). As a stimulator of Rho GTPase activity, the down-regulation of Trio could have a detrimental effect on Rho-mediated actin assembly, which is required for formation of the actin cytoskeleton and tethering of membrane bound vesicles (May-Simera and Kelley 2012, Sit and Manser 2011); Rho-GTPases also regulate formation of the microtubule cytoskeleton (Hall and Lalli 2010, Schmidt and Hall 2002, Wittmann and Waterman-Storer 2001). Proper formation of the actin and microtubule cytoskeleton is essential for a number of sperm functions, including motility.

Interaction of upregulated *mmu-miR-466m-3p* and upregulated *Piga*

Piga is a N-acetylglucosaminyl-phosphatidylinositol biosynthetic protein, these are early pre-cursors in GPI anchor formation. Whilst X-linked *Piga* appears up-regulated in our data, it may be that the interacting miRNA is having a stabilising effect on the mRNA, and subsequently inhibiting translation. GPI-anchor proteins are essential for full sperm function, permitting the localisation of proteins to the sperm surface during epididymal maturation for the purpose of cell-to-cell communication and implementation of the acrosome reaction (Fujihara et al. 2014, Watanabe and Kondoh 2011).

Micro RNA interaction with tudor-domain containing mRNAs and spermatid transcription ‘master gene’ homologue *Piwil2*

Examination of spermatid-specific and experimentally-validated interactions of miRNAs revealed several targets of interest, including importin, activin, and three miRNAs (miR-5108, miR-467g, and miR-466m-3p) that interact with autosomal *Tdrd1*, a tudor-domain containing gene that does not show cis-limitation in our data but does have important roles in spermatogenesis. *Tdrd6*^{-/-} knockout mice are infertile, with spermatogenesis blocked at the round spermatid stage. *Tdrd6* protein was found to localise to the chromatoid body and form a complex with *Tdrd1* and *Tdrd7*, although localisation of these proteins seems to be controlled independently (Vasileva et al. 2009). *Tdrd1* itself has been shown to be essential in mouse spermatogenesis, with a mixed phenotype seen in knock out

mice. Spermatogenesis was found to be highly disorganised with some tubules containing cells that did not reach beyond the pachytene-diplotene stage. Others produced round spermatids, but cellular elongation and nuclear condensation did not occur (Chuma et al. 2006). TDRD1 protein was also found to localise to the chromatoid body, suggesting the protein is also shared across cytoplasmic bridges. Tudor domain genes were not amongst short-listed mRNAs in our data, but given that miRNAs exhibit multiple mechanisms of action (*Figure 6.2*), primarily resulting in inhibition of translation, evidence of altered Tudor domain containing genes may be found following examination of the proteome. The up-regulated *mmu-miR-466m-3p* and down-regulated *mmu-miR-467g* were found to interact with spermatid-expressed *Piwil2*. *Piwil2* knockout in mice has been shown to almost completely ablate piwi RNA expression in female germ cells (Kabayama et al. 2017), and male mice lacking *Piwil2* do not undergo full spermatogenesis, arresting at the round spermatid stage (Deng and Lin 2002, review by Bak et al. 2011). Alteration of this pathway could have substantial consequences for the control of transcriptional regulation in mouse spermatids. These relationships are further investigated in Chapter 7.

6.9. Chapter Summary

The investigation of altered miRNA expression in MF1XY^{RIII}qdel and MF1XY^{RIII} spermatids and sperm has led to the identification of two mRNA-miRNA interacting pairs, both elements of which appear dysregulated in the MF1XY^{RIII}qdel model. Of particular interest is the interaction between upregulated *mmu-miR-1198-5p* and down-regulated mRNA *Trio*. In addition, pathway analysis revealed a potential role for the dysregulated miRNAs in the Wnt signalling and ECM interactions. The two methods provided distinct results, which may result from the differing selection techniques but may also reflect some variation in miRNA expression between replicates. The findings of this chapter are further discussed in Chapter 7, and considered alongside findings of other chapters.

Chapter 7

7. Discussion and Further Work

The work reported in this thesis has provided a further understanding of the consequences of MSYq deletions in mice. The MF1XY^{RIII}qdel model which possesses a Y-chromosome with 66% MSYq loss was used in comparative physiological and gene expression studies alongside its congenic full Y chromosome counterpart MF1XY^{RIII}. The major outcomes of work reported in results chapters 3, 4, 5 and 6 are discussed in this chapter, with consideration of how all findings might contribute to MSYq deletion consequences.

Potentially cis-limited X-linked genes provide candidates for causing functional differences between X- and Y-bearing sperm in MF1XY^{RIII}qdel

One of the primary aims of this investigation was to discover dysregulated and potentially non-shared X-linked genes that could facilitate inequality of sperm function between X- and Y-bearing sperm in YQ mice. Using methods previously outlined, the following genes have been shown to be suitable candidates: *Gmcl1l* (and variants), *Cylc1*, *Cldn34b4*, *Gm382*, *Gpkow*, *Ar*, *Brwd3*, *Cdkl5*, and *Cldn34-ps*. All genes show potentially important functions in spermatid development. Of particular interest are the multicopy gene *Gmcl1l*, sperm head calyx protein cylicin (*Cylc1*), and bromo domain and wd-domain containing protein (*Brwd3*). *Gmcl1l* and *Cylc1* have been previously described (section 4.5 and 5.6), and *Brwd3* family members are known to affect chromatin condensation and have been linked to the aberrant expression of histone family genes during spermatogenesis (Shang et al. 2007). Further annotation can be found in the Supplementary Material. Investigation of these genes should now progress to confirming cis-limitation, and investigating the distribution of proteins by methods outlined in the further work section (section 7.1).

Both shared and cis-limited gene products may be responsible for the morphological differences seen in sperm heads

The consistency in the character of sperm head phenotypes observed between studies of MSYq mice (Fischer et al. 2017, Ward and Burgoyne 2006, Styrna et al. 2003, Styrna et al. 2002, sections 3.3.2 and 3.4.4) suggests that the dysregulation imposed by reduction of *Sly* is specific and that the targets of both *Sly* and *Slx* are somewhat consistent. Several genes linked to the developing morphology and chromatin condensation of sperm heads were shown to be dysregulated in our data. Distinctive

amongst these were a collection of histone variants and interacting partners (*Hist1h2ba*, *H2al1k*, *H2al1n*, *H2al3*) which were highlighted during STRING analysis, and *H2al3* added following analysis of microarray data. *H2al1k*, *H2al2n* and *H2al3* are all X-linked, and whilst both *H2al1k* and *H2al3* showed increased expression in YQ spermatids, *H2al2n* was down-regulated across many sample comparisons in RNA sequencing data, unusual for an X-linked gene in this model. The H2A family are spermatid-specific histone variants, and it is likely that their altered expression is associated with the abnormal chromatin condensation expected following our sperm head morphology analysis (section 3.4). *Hist1h2ba* is known to be intimately involved and essential in histone to protamine transition in sperm (Montellier et al. 2013). During late spermiogenesis, histones are sequentially replaced by protamines which result in extreme condensation of the nuclear chromatin (section 1.3). The removal and replacement of histones is an active process, facilitated in part by the ubiquitin proteasome (Chemes and Sedo 2012). Recently, Moretti et al. (2017) have shown that for at least one H2A variant (*H2afb3*), an increase in H2A.3B incorporation into chromatin resulted following increased expression of the gene (Moretti et al. 2017). During quantitative morphological assessments, it was noted that YQ sperm are not only a distinct shape, but also far smaller than WT sperm (-10.13% by area in square microns). Size differences were also evident in both WT and YQ populations between X- and Y-bearing sperm, and that these differences were not accounted for by differences in DNA content (section 3.4.4). In their recent substantial paper focusing on the interactive mechanisms of SLY protein, Moretti et al. (2017) also noted the upregulation of several H2A variants (*H2afb3*, *H2al1*, and *H1fnt*) and showed higher H2A.B3 incorporation into chromatin in their *Sly*-KD model, strongly suggesting a link to the observed physiological changes in our data. As proposed, it may be that restricted (but not absent) sharing of one or more of the X-encoded H2A variants leads to differential chromatin condensation between sperm of differing sex chromosome content in both WT and YQ, and that these differences are merely exacerbated in YQ mice due to the upregulation of H2A genes. The sharing potential of up-regulated H2A variants was assessed, but none of the detected H2A variants appeared to show restricted sharing. Given the population-wide nature of the morphological sperm head defects seen in YQ, it is unlikely that these changes are the sole direct source of functional inequality between the X- and Y-bearing sperm in this model. However, these defects may have an effect on the overall fecundity of MSYqdel mice, as is suggested from our examination of litter sizes (section 3.2.2), which revealed a significantly lower average litter size for MSYqdel male.

Other strongly and consistently dysregulated genes including *Nptx2*, *Actr1l*, *Cylc1* also showed strong functional links to sperm morphological outcomes. These genes variously show patterns of transcript distribution similar to shared (*Nptx2*, *Actr1l*) and non-shared (*Cylc1*) transcripts. Further explanation

of their function is given in previous chapters (sections 4.5 and 5.6). The combination of shared and non-shared behaviours in combination with altered histone expression may contribute both to the global and sex-content specific morphological differences seen in sperm populations. Additional genes of interest that may not be related to morphology outcomes are *Tcp11x2* and *Mapkapk2*. Neither gene showed indications of cis-limitation, but both have functional relevance. *Tcp11x2* is an X-linked relative of *Tcp*, an allele known to be involved in the t-complex; and it shows increased expression in YQ compared to WT. MAP kinases and MAPK-activated protein kinases, such as *Mapkapk2*, are implicated in a number of functional processes relevant to sperm (reviews by Almog and Naor 2010, Li et al. 2009). Members of the MAP kinase family have also been implicated in Wnt signalling, providing a range of potential outcomes for down-regulation of this autosomal gene.

Enrichment of both functional and non-coding genes for interactions with Wnt signalling pathways

Wnt signalling was shown as an enriched interaction pathway for miRNAs short-listed by miRNA microarray expression analysis. Previous gene expression work (chapters 4 and 5) have also shown potential links to the Wnt/Ca⁺ signalling pathways based on differentially expressed protein coding genes. The Wnt signalling pathways are highly important pathways that exist in a variety of forms, contain a large number of interacting partners and having functions in a large variety of cell types. The non-canonical Wnt pathway helps regulate the production of the cytoskeleton and thus the shape of a cell. The Wnt/Ca⁺ pathway regulates the entry and exit of calcium into the cell; a function that is hugely important for many aspects of sperm function, including capacitation and hypermotility (review by Publicover et al. 2007). Wnt signalling is primarily known as a regulator of transcription; it has also shown to have post-transcriptional effects, especially in the stabilisation of proteins (Taelman et al. 2010). In 2015, Koch et al. demonstrated that post-transcriptional Wnt signalling was essential for the proper coordination of epididymal maturation of spermatozoa. Post-transcriptional Wnt signalling impacts sperm through a number of mechanisms mediated by GSK3 including, as outlined by Koch et al. (2015): the reduction of global protein poly-ubiquitination; the inhibition of *Septin4* phosphorylation (to establish a membrane diffusion barrier in the sperm tail); and the inhibition of protein phosphatase 1 via protein phosphatase 1 regulatory inhibitor subunit 2 (*Ppp1r2*) to initiate sperm motility. Two Protein phosphatases were found to be strongly upregulated in spermatids in YQ mice, autosomal *Ppp2r5a* was short-listed from RNA sequencing data and *Ppp1r2-ps9*, an X-linked *Ppp1r2* pseudogene was shown to be significantly up-regulated across a number of comparisons in microarray data. Although a pseudogene, an examination of *Ppp1r2-ps9* by ORFFinder reveals a number of intact open reading frames; including the largest with >90% identity

to ORFs in PPP1R2 in human and mouse, suggesting a possible functional role for this protein. If this up-regulated gene were to produce a functional protein, it could impact on the inhibition/activation of sperm motility through the pathways outlined above as the mouse homologue has been shown to have an inhibitory effect on protein phosphatase 1 (Koch et al. 2015, Brüchert et al. 2008). In addition, several other short-listed genes show links to Wnt signalling, including autosomal *Axin2*, which showed strong (+3.26 and +4.8 LogFc in early and late spermatids respectively) upregulation and F3 enrichment in RNA seq data and plays an important role in the stability of β -catenin. It is also an inhibitor of the canonical Wnt pathway. Its transcription and translation has been shown to be induced by activation of Wnt signalling (Jho et al. 2002). Wnt-specific examinations in post-meiotic spermatids have shown that loss of (normally pervasive) β -catenin led to subsequent acrosomal defects, release of immature spermatozoa, disruption of junctions connecting elongating spermatids and sertoli cells, compromised sperm motility and abnormal chromatin compaction (Chang et al. 2011). Potential down-regulation of Rho-mediated actin assembly by implicated *Trio* (section 6.7), also has interacting partners. Wnt signalling implicated genes shortlisted in our data show variously shared and cis-limited potential, meaning that effects could be both specific and global to X- and Y-bearing sperm. MiRPath analysis of differentially expressed miRNAs showed 13 shortlisted miRNAs to have a number of predicted interacting partners in the Wnt pathways; these included miRNAs both up- and down-regulated in YQ. The interactions of these miRNAs could result in disruption to primary Wnt signalling proteins such as WNT1 by inhibition of translation. The various incarnations of the Wnt signalling pathway do employ a large number of genes; however, the recurrent enrichment of coding and non-coding genes linked to Wnt signalling in sperm and spermatogenesis across a variety of platforms suggests that some perturbation of this pathway may occur in the YQ model. The most likely outcomes of abnormal Wnt signalling for sperm are failure of individual sperm to receive calcium signals and undergo capacitation (leading to delayed acrosomal reaction and reduced motility) and sub-motility due to aberrant phosphorylation of sperm tail/cytoskeletal proteins and altered stability of β -catenin.

Several significant consistencies were seen between our data and that reported in similar experiments (upregulation of *Vsig1* and H2A variants for example) but some variation was seen within our own data and that reported by others. *Speer* genes, stated as up-regulated in Cocquet et al. 2012, were shown not to be upregulated in mice with *Sly* loss following RT-qPCR investigations by Moretti et al. (2017). Our results reflected this, with very mild upregulation noted in some samples in RNA sequencing and little or no upregulation noted in microarray data. None of the dysregulation of *Speer* encountered in our dataset was significant. Some of this can be accounted for by differences in

cellular preparation and techniques utilised in gene expression studies. Additional ways to investigate this are given in the following section.

7.1. Further Work

The investigations outlined in this thesis have resulted in the identification of several X-linked candidate genes with potential roles in causing functional inequivalence between sperm populations from the same ejaculate and the resulting sex ratio skew. To confirm or refute the suspected cis-limitation of these genes, RNA *in situ* of the transcripts in testis could be conducted to determine whether X-linked transcripts appear present in all spermatids or simply in X-bearing spermatid populations. This should be tested in both WT and YQ testis, as altered cis-limitation appears evident between the strains for some genes. RNA fluorescence *in situ* hybridisation (RNA-FISH) could be performed to provide a more quantitative view of transcripts where sharing may be reduced but not entirely ablated (Zimmerman et al. 2013).

Vsig1, *H2al1* and *Gmcl11* have previously been investigated as offspring sex specific antigens (OSSAs) (Ellis et al. 2005). The results of an RNA *in situ* experiment showed these transcripts were present across all round spermatids at appropriate tubule stages, potentially ruling them out as mediators of the sex ratio distortion. However, these investigations do not rule out the possibility of more subtle differences in transcript abundance between neighbouring spermatids. In the work disclosed in this thesis, *Vsig1* and *H2al1* did not show transcript distribution pattern akin to that of non-shared gene products, however the multi-copy gene *Gmcl11* (including all detected copies) did show this pattern. This makes *Gmcl11*, a multi-copy X-linked gene with a currently unknown function, an ideal candidate for experimental work such as single cell sequencing or RNA FISH (as described above), where individual cells can be examined for transcript abundance.

Although separation of mature sperm cells by sex chromosome content using flow cytometry has proven successful in large agricultural animals such as cattle and pigs, the characteristic hook-shape of mouse sperm complicates this process. This capability would be highly beneficial in the examination of variation in protein (perhaps caused by altered mRNA or miRNA expression seen in our results) and RNA content between X- and Y- bearing sperm within and between various strains, and further investigation of miRNA content in sperm. Depending on the effects to processed sperm, it might also be possible to conduct sex chromosome content specific assessments of sperm motility.

From the examination of sperm head morphological differences between X- and Y-bearing sperm populations in WT and YQ sperm, and examination of differentially expressed genes, it is expected

that aberrant chromatin condensation has a significant role to play in the distinct morphology seen in YQ sperm. This could be further investigated through the use of chromatin condensation test dyes such as aniline blue, that provide evidence of histone retention in the mature sperm nucleus (Sati and Huszar 2013). Examination of the Z-axis of these separate populations by imaging fine layer z-axis slices through sperm prior to sex-chromosome FISH would provide information about potential consequences of aberrant chromatin condensation; and its potential impacts on streamlining of the sperm head for improved motility. Given the pathways and genes implicated by both broad and specific transcriptional changes in the YQ reported in this thesis, it may be that both capacitation (related to Ca^{2+} signalling) and motility (related to microtubule/cytoskeletal formation and Ca^{2+} signalling) are sources of functional differences between and within the YQ population. These can be investigated in a number of ways, including CASA (computer-assisted sperm analysis) for motility (review by Amann and Waberski 2014) and chlortetracycline assay (Fraser and Herod 1990) for capacitation.

An investigation of the outcomes of apparent dysregulation of potentially cis-limited mRNA *Trio* by up-regulated and X-linked *miR-1198-5p* could reveal another potential mechanism of functional inequivalence between X- and Y-bearing sperm in YQ. An investigation of TRIO protein abundance, either by western blot of whole sperm lysates, or through immunohistochemistry of mouse testis would be (theoretically) simple to perform. Confirmation of the miRNA-mRNA interaction could be performed through the use of luciferase reporter constructs or artificial over-expression of the miRNA in a spermatid cell line (Kuhn et al. 2008). Potential outcomes relate to Wnt/ Ca^{2+} signalling or cytoskeletal/microtubule formation. These could have deleterious outcomes for sperm motility and/or the calcium dependent capacitation reaction as outlined above. These analyses could also be applied to other MTIs revealed in investigations outlined in Chapter 6.

Gm15772 was shown to be strongly upregulated in YQ in RNA sequencing, and was seen to be highly expressed and upregulated in microarray data. This gene was not short-listed in microarray expression analysis because although upregulated, the change was not significant, as it showed some variability between replicates. The lack of replicates in the RNA seq data means that significance values cannot be assigned. RT-qPCR of this and other genes that were short-listed in one dataset but not in another would provide further clarity and confidence in gene expression results.

Chapter 8

8. Conclusion

Through the work outlined in this thesis we have gained further insights into the consequences of MSYq deletion in mice. The intragenomic conflict between X- and Y-linked *Sycp3*-like genes results in the dysregulation of hundreds of genes following depletion or absence of *Sly* transcripts. Amongst these consequences is an offspring sex ratio skew toward females, facilitated by unequal functional capacity of X- and Y-bearing sperm. The aims of this investigation were to further understand the physiological consequences of *Sly* depletion and identify candidate genes for further investigation based on their potential involvement in sex ratio distortion and the perceived physiological differences between wild-type and MSYq deletion mice. Examination of physiological characteristics in MF1XY^{RIII}qdel (YQ) mice provides a highly detailed examination of abnormal sperm nuclear morphology likely linked to altered chromatin packaging. This revealed differences of size and region-specific shape variation in the YQ model. Novel investigations revealed nuclear shape differences linked to sex chromosome content for the first time. A reduced litter size and offspring sex ratio skew of 6% towards females was also revealed for YQ sires. Subsequent molecular investigations of the transcriptome by RNA sequencing and microarray facilitated the investigation of differential expression and transcript distribution within enriched spermatid populations and subcellular fractions to retrieve potentially shared and non-shared candidate genes for involvement in reproductive and sperm phenotypic consequences in YQ. Suggestions for further investigation of these candidates and physiological outcomes of MSYq deletion are also given and further explored. The understanding of transcript sharing between spermatids, effects of Y-deletions in mammals and contributions to the search for sex-specific markers in sperm are greatly beneficial to both fertility and agricultural research.

Appendix A

Publications and Presentations

Publications

Skinner, B. M., & Johnson, E. E. P. (2017). Nuclear morphologies: their diversity and functional relevance. *Chromosoma*, 126(2), 195–212. <https://doi.org/10.1007/s00412-016-0614-5>

Skinner BM, Rathje CC, Bacon J, Johnson EEP, Yousafazi G, Mills BP, Affara NA, Ellis PJI. Methods for assessing sperm nuclear morphology in mice. In preparation.

Rathje CC, Silvestri G, Johnson EEP, Bacon J, Ellis PJI, Skinner BM. Deletion of ampliconic mouse Y genes affects sperm morphology and function. In preparation.

Randle SJ, Skinner BM, Majumdar A, Nelson DE, Baker J, Vlazaki M, Bacon J, Johnson E, Rathje C, Ellis P, Affara N, Laman H. Loss of *Fbxo7* expression causes male sterility associated with defective spermatocyte remodelling. In preparation.

Presentations

Johnson EEP, Skinner BM, Ellis PJI, Affara NA (2016); Oral and poster presentations by Johnson EEP; Sex Ratio Distortion in a Y-Chromosome Deletion Model; 21st International Chromosome Conference, Foz do Iguaçu, Brazil

Johnson EEP (2016) All sperm are equal but some are more equal than others: understanding sex ratio skews; poster presentation; SET for Britain, House of Commons, UK

Johnson EEP, Skinner BM, Ellis PJI, Affara NA (2015); Poster presentation by Johnson EEP; Investigating physiological and molecular mechanisms of sex ratio distortion; 20th International Chromosome Conference, Canterbury, UK.

Johnson EEP, Skinner BM, Ellis PJI, Affara NA (2015); Poster presentation by Johnson EEP; Investigating physiological and molecular mechanisms of sex ratio distortion; Annual Pathology Symposium, Cambridge, UK

Appendix B

Supplementary Material

R scripts used for data analysis as cited throughout this thesis are included at the following address:
www.github.com/EmmaEPJ/Repository

Supplementary Material for Chapter 3

Technical note on probe efficiency from section 3.4.5.

There is a question of preferential detection or greater efficiency of probes in either X or Y which might result in differential sampling. Observations from preparation of this data suggest that where probe detection failed, it failed across the entire image because of obstruction or slide damage. Probes were also tested on metaphase spleen cells and showed no difference in hybridisation efficiency (*Figure 1.15*). A small amount of X probe was noted to bind to a minute autosomal region; this was also observed in Y-bearing sperm. The autosomal fluorescence was highly distinct from X chromosome fluorescence in these instances.

Supplementary Material for Chapter 4

Technical note on DNA contamination in sample YQEF2

Sample YQEF2 was noted as diverging slightly from the main early spermatid sample cluster in PCA analysis and as displaying (along with four other samples) some ribosomal RNA contamination (*Figure 4.8*). Further investigation revealed a distinctly low strandedness value for this and the four other samples which were processed in a single batch. Measures of strandedness in RNA sequencing allow the user to determine how many of the detected reads map to the expected DNA strand, meaning it is easier to discern between genes of similar composition and allocate reads to the correct genomic origin. The additional four samples were paired samples WTLF2, YQLF2, WTLF3, and YQLF3. Mapping of all reads from these samples and YQEF2 to the mouse genome (conducted by Julien Bauer, Cambridge Genomic Services) revealed some good alignment including in intronic regions, strongly suggesting DNA contamination. Communications with Illumina confirmed that DNA contamination may in turn have been a cause of higher rRNA presence, as ribosomal depletion kits work less well in the presence of DNA. In order to overcome some of the problems associated with DNA contamination, reads were filtered by strandedness, meaning that only reads originating

from the anticipated strand were used in the analysis. In affected samples, this leads to a reduced library size, which is somewhat alleviated by the specific normalisation technique applied to this dataset. However, this reduced library size will lead to an under-sampling of data from that sample. Four of the five samples are matched and were compared directly with their counterpart from the other mouse strain; this resulted in libraries of similar sizes within the pairs. Though not ideal, comparisons between these pairs will be more robust than comparisons made between YQEF2 and its counterpart which was not affected by DNA contamination. Two of the samples (WTLF3 and YQLF3) were expected to contain primarily nuclear/organelle bound transcripts, being derived from fraction 3 (see section 4.1.2), which might have gone some way to explaining an increased presence of DNA compared to other samples. However, other primarily nuclear samples did not show this contamination and sample YQEF2, which is comparable to WTEF2, also showed contamination whilst its counterpart did not. This suggests, as all affected samples were processed in the same batch, that a failure of DNase treatment, perhaps previously incorrectly stored or damaged mechanically, was at least partly responsible for the reported contamination. Suitable measures were taken to use the data to its greatest benefit given these factors, however further gene expression work is required to resolve transcriptional changes across the examined populations.

Note on examination of mature sperm samples in RNA sequencing data: Differentially expressed genes in mature sperm and uneven distribution of contaminants in mature cell populations

The existence of two extremely highly expressed mitochondrial rRNAs (mt-Rnr2 and mt-Rnr1) accounted for the majority of expression in the mature sperm. Their comparison between strains recorded a substantially lower count value in YQ (though not recorded as a top 10 differentially expressed gene, as proportionally other genes showed greater levels of differential expression), and may have resulted in a skewing of the data recorded in mature samples as these transcripts accounted for around half of all counts recorded across either sample. The results from differential expression analysis of these two samples show a skew towards higher magnitude negative logFC values (reflected in the examination of mean and median values), this may be a reflection of true biological differences simply amplified by the skew. DE genes from this comparison should be treated cautiously but positive logFC values in particular are expected to relate to real differences. In future examinations a removal of the two highly expressed transcripts (preferably by biological depletion rather than computational) may provide more insight from these samples, however these samples would primarily benefit from an increase in replicate number.

Differentially expressed genes: Gene list A: Protein-coding genes

Protein-coding genes from top 20 differentially expressed; WTEF1 vs YQEF1			
Gene symbol	Chr	Gene / protein name	LogFC
<i>4930451G09Rik</i>	16	RIKEN cDNA 4930451G09 gene	1.88
<i>Hist1h2ba</i>	13	Histone cluster 1, H2ba	1.53
<i>Vit</i>	17	Vitrin	1.51
<i>Dynlt1a</i>	17	Dynein light chain Tctex-type 1A	1.39
<i>Vsig1</i>	X	V-set and immunoglobulin domain containing 1	1.39
<i>H2al1k</i>	X	H2A histone family member L1K (Gm14501)	1.30
<i>Klk1b26</i>	7	Kallikrein 1-related peptidase b26	1.24
<i>Dynlt1f</i>	17	Dynein light chain Tctex-type 1F	1.19
<i>Ell</i>	8	Elongation factor RNA polymerase II	1.17
<i>Actrt1</i>	X	Actin-related protein T1	1.15
<i>Cpxcr1</i>	X	CPX chromosome region, candidate 1	1.10
<i>Gm11639</i>	11	Predicted gene 11639	1.08
<i>Trappc2l</i>	8	Trafficking protein particle complex 2-like	-1.08
<i>Neo1</i>	9	Neogenin	-1.15
<i>H2al1n</i>	X	H2A histone family member L1N (Gm5382)	-1.33
<i>Shroom3</i>	5	Shroom family member 3	-1.34

Protein-coding genes from top 20 differentially expressed; WTEF2 vs YQEF2			
Gene symbol	Chr	Gene / protein name	LogFC
<i>Lars2</i>	9	leucyl-tRNA synthetase, mitochondrial	6.68
<i fn1<="" i=""></i>	1	Fibronectin 1	2.10
<i>Hist1h2ba</i>	13	Histone cluster 1, H2ba	1.80
<i>Lyz2</i>	10	Lysozyme 2	1.61
<i>Cldn34b4</i>	X	Claudin / RIKEN cDNA 4930428D18 gene	1.37
<i>Fos</i>	12	FBJ osteosarcoma oncogene	1.32
<i>Vsig1</i>	X	V-set and immunoglobulin domain containing 1	1.25
<i>4930451G09Rik</i>	16	RIKEN cDNA 4930451G09 gene	1.23
<i>Cers3</i>	7	LAG1 homolog, ceramide synthase 3	1.14

<i>Actrt1</i>	X	Actin-related protein T1	1.14
<i>Shroom3</i>	5	Shroom family member 3	-1.47

Protein-coding genes from top 20 differentially expressed; **WTEF3 vs YQEF3**

Gene symbol	Chr	Gene / protein name	LogFC
<i>Vars2</i>	17	valyl-tRNA synthetase 2, mitochondrial (putative)	3.56
<i>Fn1</i>	1	Fibronectin 1	2.45
<i>Meiob</i>	17	Meiosis specific with OB domains	2.37
<i>Cpne7</i>	8	Copine VII	1.68
<i>Gpkow</i>	X	G patch domain and KOW motifs	1.40
<i>H2-D1</i>	17	Histocompatibility 2, D region locus 1	1.40
<i>Scfd2</i>	5	Sec1 family domain containing 2	-1.40
<i>Hmcn1</i>	1	Hemicentin 1	-1.40
<i>Slc26a10</i>	10	Solute carrier family 26, member 10	-1.43
<i>Pkd1l1</i>	11	Polycystic kidney disease 1 like 1	-1.62
<i>Tmc1</i>	19	Transmembrane channel-like gene family 1	-1.69
<i>Trio</i>	15	Triple functional domain (PTPRF interacting)	-1.82
<i>Xirp1</i>	9	Xin actin-binding repeat containing 1	-2.08
<i>Lars2</i>	9	leucyl-tRNA synthetase, mitochondrial	-4.62

Protein-coding genes from top 20 differentially expressed; **WTEWC vs YQEWC**

Gene symbol	Chr	Gene / protein name	LogFC
<i>Hist1h2ba</i>	13	Histone cluster 1, H2ba	1.47
<i>Vit</i>	17	Vitrin	1.45
<i>Dynlt1a</i>	17	Dynein light chain Tctex-type 1A	1.41
<i>4930451G09Rik</i>	16	RIKEN cDNA 4930451G09 gene	1.33
<i>Klk1b26</i>	7	Kallikrein 1-related peptidase b26	1.31
<i>Vsig1</i>	X	V-set and immunoglobulin domain containing 1	1.30
<i>Dynlt1f</i>	17	Dynein light chain Tctex-type 1F	1.29
<i>Tspan6</i>	X	Tetraspanin 6	1.25
<i>Actrt1</i>	X	Actin-related protein T1	1.15
<i>Trappc2l</i>	8	Trafficking protein particle complex 2-like	-1.16

<i>Shroom3</i>	5	Shroom family member 3	-1.33
<i>Trio</i>	15	Triple functional domain (PTPRF interacting)	-1.33
<i>H2al1n</i>	X	H2A histone family member L1N (Gm5382)	-1.51
<i>Slc26a10</i>	10	Solute carrier family 26, member 10	-1.53
<i>BC061237</i>	14	cDNA sequence BC061237	-2.66

Protein-coding genes from top 20 differentially expressed; **WTLF1 vs YQLF1**

Gene symbol	Chr	Gene / protein name	LogFC
<i>Gml2</i>	15	Hematopoietic cell transcript 1	3.10
<i>Cd55</i>	1	CD55 antigen	1.88
<i>Nptx2</i>	5	Neuronal pentraxin 2	1.88
<i>H2al1k</i>	X	H2A histone family member L1K (Gm14501)	1.69
<i>4930402K13Rik</i>	X	RIKEN cDNA 4930402K13 gene	1.65
<i>Vsig1</i>	X	V-set and immunoglobulin domain containing 1	1.61
<i>Cylc1</i>	X	Cylicin, basic protein of sperm head cytoskeleton 1	1.50
<i>Trappc2l</i>	8	Trafficking protein particle complex 2-like	-1.65
<i>H2al1n</i>	X	H2A histone family member L1N (Gm5382)	-1.70
<i>Slc26a10</i>	10	Solute carrier family 26, member 10	-1.82
<i>Mdga2</i>	12	MAM domain containing glycosylphosphatidylinositol anchor 2	-1.92
<i>Antxr2</i>	5	Anthrax toxin receptor 2	-2.82
<i>2010109A12Rik</i>	5	RIKEN cDNA 2010109A12 gene	-4.41
<i>Gm10800</i>	2	Predicted gene 10800	-4.74

Protein-coding genes from top 20 differentially expressed; **WTLF2 vs YQLF2**

Gene symbol	Chr	Gene / protein name	LogFC
<i>Lyz2</i>	10	Lysozyme 2	2.91
<i>Gm904</i>	13	Predicted gene 904	1.87
<i>Gm614</i>	X	Predicted gene 614	1.78
<i>Cd55</i>	1	CD55 antigen	1.60
<i>Fam122c</i>	X	Family with sequence similarity 122, member C	1.55
<i>Vsig1</i>	X	V-set and immunoglobulin domain containing 1	1.54
<i>Fez1</i>	9	Fasciculation and elongation protein zeta 1 (zygin I)	-1.46

<i>Mdga2</i>	12	MAM domain containing glycosylphosphatidylinositol anchor 2	-1.54
<i>Trappc2l</i>	8	Trafficking protein particle complex 2-like	-1.63
<i>Slc26a10</i>	10	Solute carrier family 26, member 10; Chloride/bicarbonate exchanger	-1.67
<i>H2al1n</i>	X	H2A histone family member L1N (Gm5382)	-1.92
<i>Rd3</i>	1	Retinal degeneration 3	-1.97
<i>Shroom3</i>	5	Shroom family member 3	-2.51

Protein-coding genes from top 20 differentially expressed; **WTLF3 vs YQLF3**

Gene symbol	Chr	Gene / protein name	LogFC
<i>Axin2</i>	11	Axin2	4.80
<i>Lars2</i>	9	leucyl-tRNA synthetase, mitochondrial	4.66
<i>Gm10800</i>	2	Predicted gene 10800	3.76
<i>Gm10801</i>	2	Predicted gene 10801	3.33
<i>Myl12a</i>	17	Myosin, light chain 12A, regulatory, non-sarcomeric	2.48
<i>Ppp2r5a</i>	1	Protein phosphatase 2, regulatory subunit B (B56), alpha isoform	2.26
<i>Gm904</i>	13	Predicted gene 904	2.10
<i>Myl12b</i>	17	Myosin, light chain 12B, regulatory	2.05
<i>Cpne7</i>	8	Copine VII	1.82
<i>Nlrc5</i>	8	NLR family, CARD domain containing 5	1.78
<i>Vsig1</i>	X	V-set and immunoglobulin domain containing 1	1.75
<i>Tbl3</i>	17	Transducin (beta)-like 3	1.70
<i>Spata13</i>	14	Spermatogenesis associated 13	1.61
<i>Dnah11</i>	12	Dynein, axonemal, heavy chain 11	-1.88

Protein-coding genes from top 20 differentially expressed; **WTLWC vs YQLWC**

Gene symbol	Chr	Gene / protein name	LogFC
<i>Gm10260</i>	13	Predicted gene 10260	7.42
<i>Gml2</i>	15	Hematopoietic cell transcript 1	2.29
<i>Nptx2</i>	5	Neuronal pentraxin 2	2.09
<i>Cpne7</i>	8	Copine VII	1.97
<i>H2al1k</i>	X	H2A histone family member L1K (Gm14501)	1.74

<i>Cnga3</i>	1	Cyclic nucleotide gated channel alpha 3	1.73
<i>Gm3763</i>	X	Predicted gene 3763	1.70
<i>Vsig1</i>	X	V-set and immunoglobulin domain containing 1	1.63
<i>4930402K13Rik</i>	X	RIKEN cDNA 4930402K13 gene	1.58
<i>Cd55</i>	1	CD55 antigen	1.52
<i>Gm382</i>	X	Predicted gene 382	1.51
<i>Antxr2</i>	5	Anthrax toxin receptor 2	-1.62
<i>Slc26a10</i>	10	Solute carrier family 26, member 10	-1.68
<i>H2a1n</i>	X	H2A histone family member L1N (Gm5382)	-1.79
<i>Mdga2</i>	12	MAM domain containing glycosylphosphatidylinositol anchor 2	-1.80
<i>2010109A12Rik</i>	5	RIKEN cDNA 2010109A12 gene	-3.90

Gene list A: Pseudogenes and Non-coding RNAs

Non-coding genes and pseudogenes from top 20 differentially expressed; **WTEF1 VS YQEF1**

Gene symbol	Gene biotype	Chr	Gene name	LogFC
<i>Gm15772</i>	Processed pseudogene	5	Predicted gene 15772	5.67
<i>Gm1821</i>	Processed pseudogene	14	Predicted gene 1821	-2.59
<i>Gm4665</i>	Antisense	9	Predicted gene 4665	1.51
<i>Gm15104</i>	LincRNA	X	Predicted gene 15104	1.29

Non-coding genes and pseudogenes from top 20 differentially expressed; **WTEF2 VS YQEF2**

Gene symbol	Gene biotype	Chr	Gene name	LogFC
<i>Gm25360</i>	SnRNA	1	Predicted gene 25360	6.73
<i>Mir6236</i>	MiRNA	9	microRNA 6236	5.97
<i>Gm15564</i>	Antisense	16	Predicted gene 15564	5.53
<i>Gm15772</i>	Processed pseudogene	5	Predicted gene 15772	5.34
<i>Gm23935</i>	MiRNA	16	Predicted gene 23935	3.62
<i>Gm6866</i>	Transcribed unprocessed pseudogene	7	Predicted gene 6866	1.57
<i>Gm43263</i>	lincRNA	8	Predicted gene 43263	1.32
<i>Gm24407</i>	SnRNA	5	Predicted gene 24407	-1.19
<i>Gm24265</i>	SnRNA	5	Predicted gene 24265	-1.18

Non-coding genes and pseudogenes from top 20 differentially expressed; **WTEF3 VS YQEF3**

Gene symbol	Gene biotype	Chr	Gene name	LogF C
<i>Gm25360</i>	SnRNA	1	Predicted gene 25360	5.32
<i>Gm25813</i>	SnRNA	8	Predicted gene 25813	4.91
<i>RP23-268E17.1</i>	lincRNA	7	RIKEN cDNA 4930533N22 gene	-3.66
<i>Gm6866</i>	Transcribed unprocessed pseudogene	7	Predicted gene 6866	1.93
<i>Gm23935</i>	MiRNA	16	Predicted gene 23935	-1.63
<i>Rnu12</i>	SnRNA	15	RNA U12, small nuclear	1.37

Non-coding genes and pseudogenes from top 20 differentially expressed; **WTEWC VS YQEWC**

Gene symbol	Gene biotype	Chr	Gene name	LogF C
<i>Gm15772</i>	Processed pseudogene	5	Predicted gene 15772	5.72
<i>Gm1821</i>	Processed pseudogene	14	Predicted gene 1821	-2.74
<i>Gm6866</i>	Transcribed unprocessed pseudogene	7	Predicted gene 6866	1.50
<i>Gm4665</i>	antisense	9	Predicted gene 4665	1.30
<i>Gm15104</i>	lincRNA	X	Predicted gene 15104	1.22

Non-coding genes and pseudogenes from top 20 differentially expressed; **WTLF1 VS YQLF1**

Gene symbol	Gene biotype	Chr	Gene name	LogFC
<i>Gm15772</i>	Processed pseudogene	5	Predicted gene 15772	6.05
<i>Gm1821</i>	Processed pseudogene	14	Predicted gene 1821	-3.02
<i>1700063H06Rik</i>	LincRNA	9	RIKEN cDNA 1700063H06 gene	1.54
<i>Gm14367</i>	Unprocessed pseudogene	X	Predicted gene 14367	1.50
<i>Gm11753</i>	Sense intronic	11	Predicted gene 11753	-1.49
<i>Gm15104</i>	lincRNA	X	Predicted gene 15104	1.48

Non-coding genes and pseudogenes from top 20 differentially expressed; **WTLF2 VS YQLF2**

Gene symbol	Gene biotype	Chr	Gene name	LogFC
<i>Gm15772</i>	Processed pseudogene	5	Predicted gene 15772	4.85
<i>Gm1821</i>	Processed pseudogene	14	Predicted gene 1821	-2.49

<i>Gm24270</i>	MiRNA	9	Predicted gene 24270	-1.57
<i>C2cd6b</i>	Unprocessed pseudogene	1	C2 calcium dependent domain containing 6B (Als2cr11b). Biotype conflict - variously annotated as lincRNA, unprocessed pseudogene and protein coding	-1.55
<i>4930570D08Rik</i>	Processed transcript	X	RIKEN cDNA 4930570D08 gene	1.50
<i>Gm24265</i>	SnRNA	5	Predicted gene 24265	1.45
<i>Gm4665</i>	Antisense	9	Predicted gene 4665	1.45

Non-coding genes and pseudogenes from top 20 differentially expressed; **WTLF3 VS YQLF3**

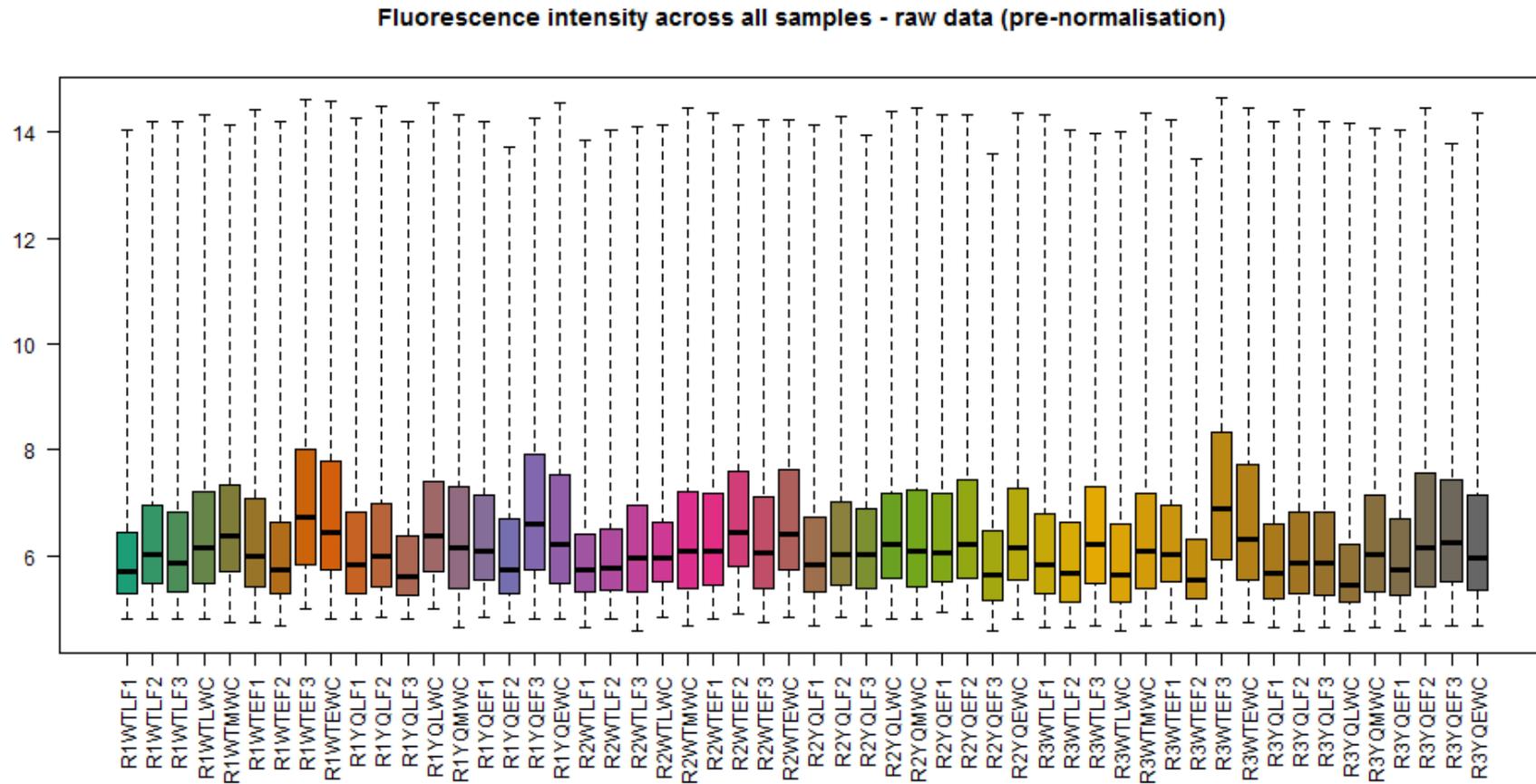
Gene symbol	Gene biotype	Chr	Gene name	LogFC
<i>9330121J05Rik</i>	Antisense	3	RIKEN cDNA 9330121J05 gene	6.70
<i>Gm15564</i>	Antisense	16	Predicted gene 15564	5.29
<i>Gm25732</i>	MiRNA	4	Predicted gene 25732	3.67
<i>Mir6236</i>	MiRNA	9	MicroRNA 6236	2.85
<i>Gm24270</i>	MiRNA	9	Predicted gene 24270	1.69
<i>Gm24265</i>	SnRNA	5	Predicted gene 24265	1.63

Non-coding genes and pseudogenes from top 20 differentially expressed; **WTLWC VS YQLWC**

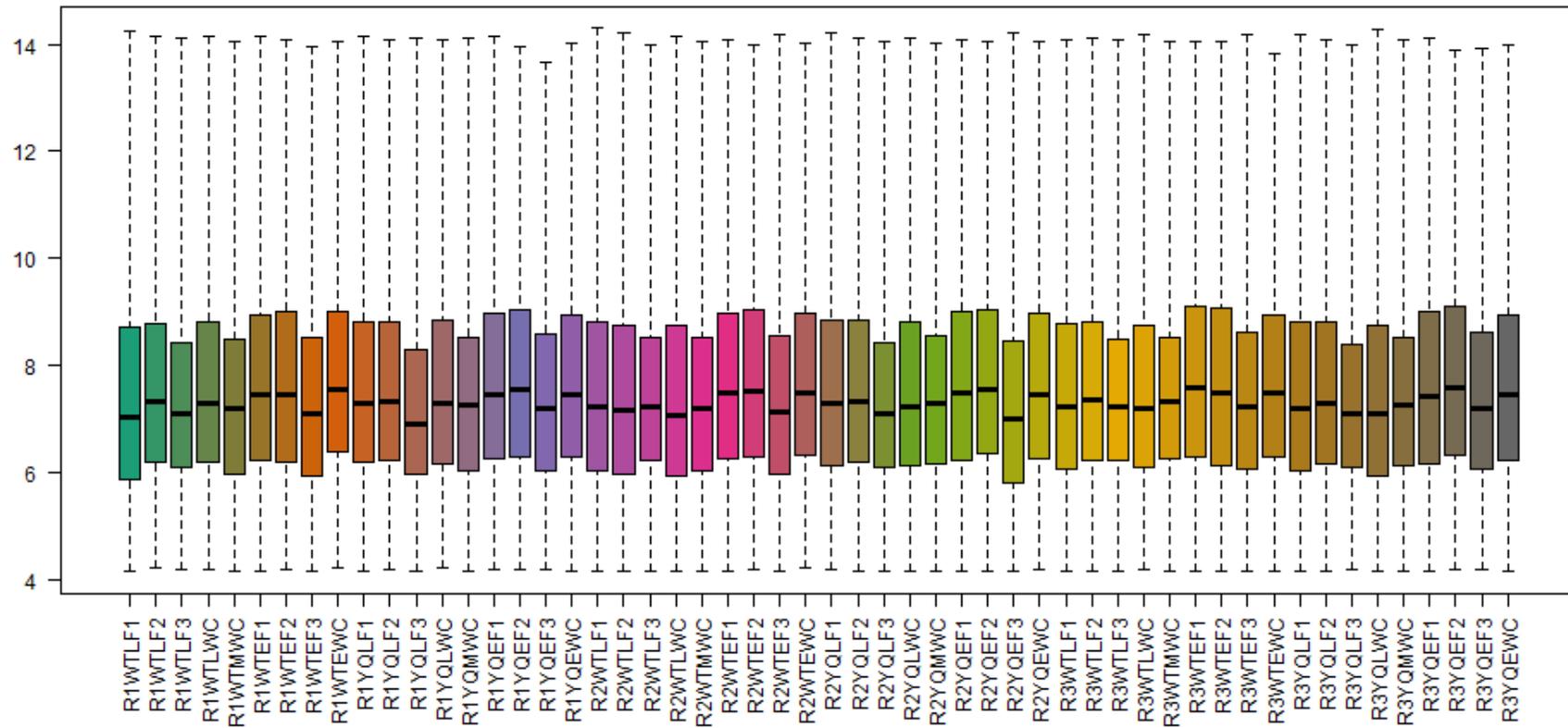
Gene symbol	Gene biotype	Chr	Gene name	LogFC
<i>Rps13-ps1</i>	Processed pseudogene	8	Ribosomal protein S13, pseudogene 1	-8.64
<i>Gm15772</i>	Processed pseudogene	5	Predicted gene 15772	5.92
<i>Gm1821</i>	Processed pseudogene	14	Predicted gene 1821	-2.81
<i>Gm15674</i>	Antisense	1	Predicted gene 15674	-2.09

Supplementary Material for Chapter 5

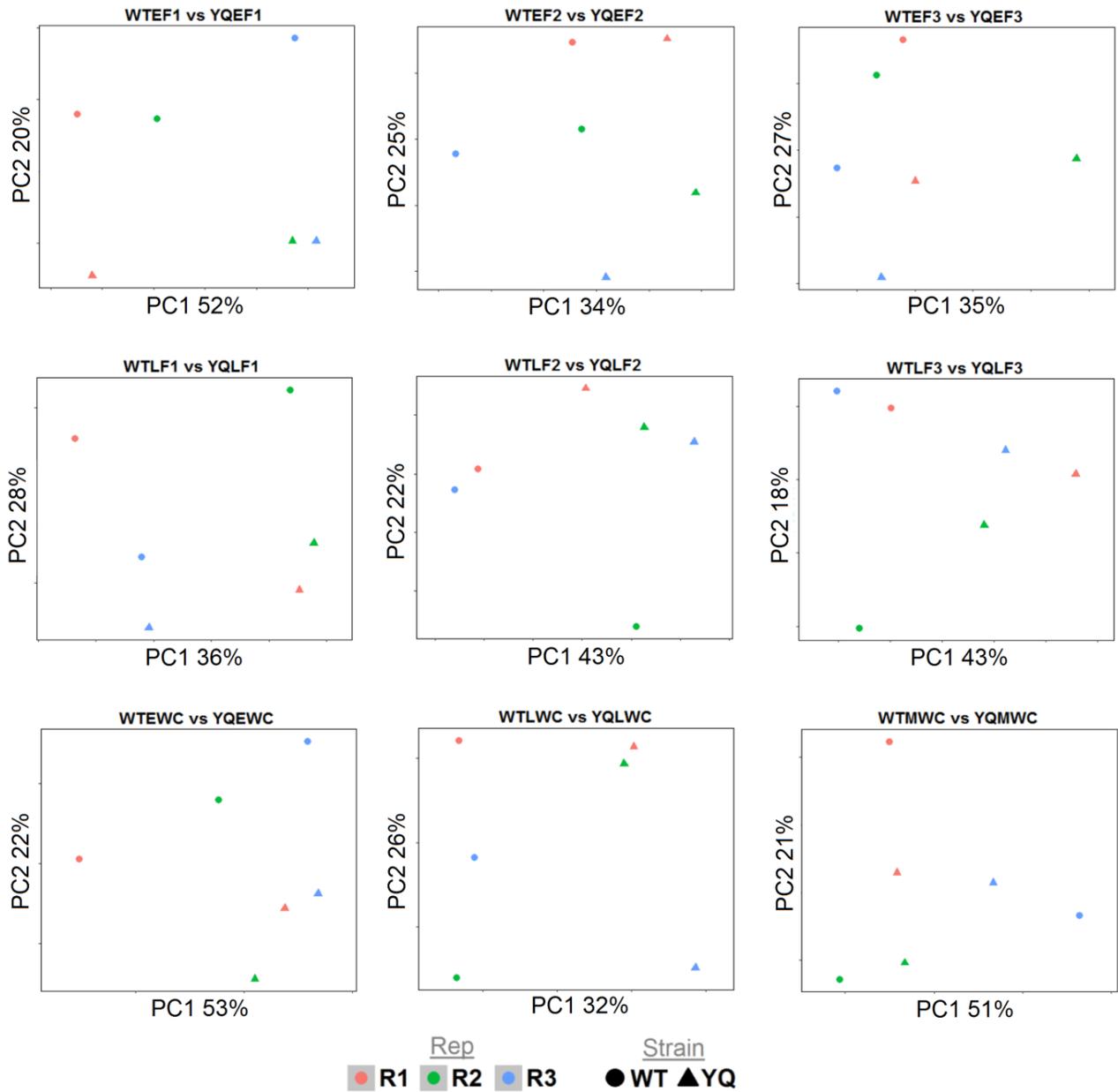
Probe fluorescence intensity (intensity boxplots)



Fluorescence intensity across all samples - post-normalisation



PCA plots showing major variance in direct comparisons of WT and YQ samples: Each sample is paired with its counterpart from the other mouse strain. All replicates are shown. Samples are arranged by cell type and fraction; top row = early spermatid samples, middle row = late spermatid samples, bottom row = whole cell samples.



All differentially expressed genes from across all sample comparisons. These tables include Transcript IDs from the Affymetrix array, gene symbols, aliases (retrieved from MGI), chromosomal origin, gene biotype and logFC value of that transcript cluster in YQ compared to WT.

WTEF1 vs YQEF1					
Transcript ID	Gene symbol	Aliases	Chr	Gene biotype	logFC
17226736	<i>Mapkapk2</i>		1	protein coding	-1.84

WTEF2 vs YQEF2					
Transcript ID	Gene symbol	Aliases	Chr	Gene biotype	logFC
17444498	<i>Nptx2</i>		5	protein_coding	1.65
17489923	<i>Gm21115</i>	<i>Gm4454</i>	7	unprocessed_pseudogene	1.37
17489943	<i>Gm21136</i>		7	unprocessed_pseudogene	1.35
17489959	<i>Gm5591</i>	<i>Gm6818, Gm29258, Gm21129 ...</i>	7	unprocessed_pseudogene, protein_coding, processed_pseudogene	1.13
17489968	<i>Gm21129</i>	<i>Gm5591, Gm6818, Gm6833 ...</i>	7	unprocessed_pseudogene, protein_coding, processed_pseudogene	1.13
17286046	<i>4932702P03Rik</i>		13	antisense	1.12
17489962	<i>Gm1988</i>		7	unprocessed_pseudogene	1.09
17442173	<i>A930024E05Rik</i>		5	antisense	1.01
17338595	<i>4932415M13Rik</i>		17	lincRNA	0.81
17520116	<i>Gm10634</i>		9	lincRNA, antisense, processed_transcript	0.78
17534730	<i>Fam122c</i>		X	protein_coding	0.71
17547235	<i>Ssty2</i>	<i>Gm21672, Gm20809, Gm20868 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.87
17547355	<i>Ssty2</i>	<i>Gm21672, Gm20816, Gm20898 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.87
17547389	<i>Ssty2</i>	<i>Gm21180, Gm20815, Gm21394 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.87
17546509	<i>Ssty2</i>	<i>Gm20823, Gm21394, Gm20877 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.96
17547223	<i>Ssty2</i>	<i>Gm20809, Gm21943, Gm20877 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.97

17547341	<i>Ssty2</i>	<i>Gm21672, Gm20823, Gm21180 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.97
17547375	<i>Ssty2</i>	<i>Gm20809, Gm21242, Gm21292 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.97
17547407	<i>Ssty2</i>	<i>Gm21394, Gm20809, Gm20867 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.97
17547410	<i>Ssty2</i>	<i>Gm20816, Gm20815, Gm21394 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.97
17546668	<i>Ssty2</i>	<i>Gm20867, Gm20823, Gm20868 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.97
17547226	<i>Ssty2</i>	<i>Gm20823, Gm20816, Gm21242 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.98
17547346	<i>Ssty2</i>	<i>Gm21394, Gm21242, Gm20868 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.98
17546633	<i>Gm20879</i>	<i>Gm20747</i>	Y	unprocessed_pseudogene, protein_coding	-1.05
17547380	<i>Gm20868</i>		Y	NA	-1.13
17547281	<i>Ssty2</i>	<i>Gm20917, Gm21180, Gm21672 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding ...	-1.13
17547251	<i>Ssty2</i>	<i>Gm20934, Gm20865, Gm20816 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-1.13
17546452	<i>Gm20815</i>	<i>Gm20879, Gm20816, Gm20867 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-1.16
17547368	<i>Ssty2</i>	<i>Gm20816, Gm20823, Gm20854 ...</i>	Y	protein_coding	-1.17
17547420	<i>Ssty2</i>	<i>Gm20738, Gm20868, Gm20806 ...</i>	Y	protein_coding	-1.17
17546450	<i>Ssty2</i>	<i>Gm21799, Gm21180, Gm21394 ...</i>	Y	protein_coding	-1.19
17547261	<i>Gm21943</i>	<i>Gm20806, Gm20842, Gm20815 ...</i>	Y	protein_coding	-1.19
17547394	<i>Gm20854</i>	<i>Gm20842, Gm20907, Gm21292 ...</i>	Y	protein_coding	-1.19
17547464	<i>Gm20738</i>	<i>Gm20898, Gm20865, Gm20809 ...</i>	Y	NA	-1.19
17547246	<i>Ssty2</i>	<i>Gm20815, Gm21180,</i>	Y	transcribed_unprocessed_p	-1.20

		<i>Gm20806</i> ...		seudogene, protein_coding	
17547229	<i>Ssty2</i>	<i>Gm21394, Gm21292, Gm21242</i> ...	Y	transcribed_unprocessed_p seudogene, protein_coding ...	-1.21
17547443	<i>Ssty2</i>	<i>Gm21180, Gm21943, Gm20867</i> ...	Y	transcribed_unprocessed_p seudogene, protein_coding ...	-1.21
17547450	<i>Ssty2</i>	<i>Gm21242, Gm20867, Gm20898</i> ...	Y	transcribed_unprocessed_p seudogene, protein_coding ...	-1.21
17546467	<i>Ssty2</i>	<i>Gm20868, Gm20852, Gm20907</i> ...	Y	protein_coding, unprocessed_pseudogene	-1.21
17546469	<i>Ssty2</i>	<i>Gm21658, Gm20868, Gm20842</i> ...	Y	NA	-1.21
17547256	<i>Ssty2</i>	<i>Gm20867, Gm20934, Gm21292</i> ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-1.23
17547285	<i>Ssty2</i>	<i>Gm21943, Gm21394, Gm20823</i> ...	Y	protein_coding	-1.23
17547295	<i>Ssty2</i>	<i>Gm21242, Gm21394, Gm20738</i> ...	Y	protein_coding	-1.23
17546673	<i>Ssty2</i>	<i>Gm21292, Gm20806, Gm20747</i> ...	Y	transcribed_unprocessed_p seudogene, protein_coding	-1.23
17547423	<i>Ssty2</i>	<i>Gm20898, Gm20865, Gm20806</i> ...	Y	transcribed_unprocessed_p seudogene, protein_coding	-1.24
17547385	<i>LOC101056308</i>	<i>Gm20854, Gm20921, Gm21394</i> ...	Y	NA	-1.25
17547351	<i>Ssty2</i>	<i>LOC101056308, Gm20907, Gm29024</i> ...	Y	NA	-1.25
17547378	<i>Ssty2</i>	<i>Gm20907, Gm20806, Gm20879</i> ...	Y	NA	-1.25
17547402	<i>Gm21799</i>	<i>Gm21394, Gm21242, Gm20826</i> ...	Y	transcribed_unprocessed_p seudogene, protein_coding	-1.26
17547417	<i>Ssty2</i>	<i>Gm20865, Gm20823, Gm20854</i> ...	Y	protein_coding	-1.28
17547365	<i>Ssty2</i>	<i>Gm21180, Gm20815, Gm20868</i> ...	Y	protein_coding	-1.28
17547283	<i>Gm21180</i>	<i>Gm20809, Gm20738, Gm20854</i> ...	Y	transcribed_unprocessed_p seudogene, protein_coding	-1.28
17547445	<i>Ssty2</i>	<i>Gm21118, Gm20806, Gm20868</i> ...	Y	protein_coding	-1.28

17546454	<i>Ssty2</i>	<i>Gm21242, Gm20854, Gm21180 ...</i>	Y	protein_coding	-1.28
17547271	<i>Ssty2</i>	<i>Gm21799, Gm20934, Gm20809 ...</i>	Y	protein_coding	-1.29
17546457	<i>Gm20907</i>	<i>Gm21835, Gm21242, Gm20809 ...</i>	Y	protein_coding	-1.30
17547415	<i>Ssty2</i>	<i>Gm20854, Gm20865, Gm20898 ...</i>	Y	protein_coding	-1.30
17547428	<i>Ssty2</i>	<i>Gm20867, Gm20823, Gm21394 ...</i>	Y	protein_coding	-1.31
17547413	<i>Ssty2</i>	<i>Gm20747, Gm20917, Gm21672 ...</i>	Y	protein_coding	-1.32
17547288	<i>Ssty2</i>	<i>Gm20852, Gm21394, Gm21943 ...</i>	Y	protein_coding	-1.33
17547362	<i>Ssty2</i>	<i>Gm21394, Gm20917, Gm21242 ...</i>	Y	protein_coding	-1.33
17547231	<i>Gm20854</i>	<i>Gm29024, Gm21245, Gm20865 ...</i>	Y	protein_coding	-1.34
17546498	<i>Ssty2</i>	<i>Gm20867, Gm20917, Gm21394 ...</i>	Y	protein_coding	-1.34
17546461	<i>Ssty2</i>	<i>Gm20806, Gm21292, Gm20823 ...</i>	Y	protein_coding, transcribed_unprocessed_p seudogene	-1.35
17546463	<i>Ssty2</i>	<i>Gm20898, Gm20865, Gm20852 ...</i>	Y	protein_coding, transcribed_unprocessed_p seudogene	-1.35
17546465	<i>Gm20879</i>	<i>Gm20868, Gm20806, Gm21292 ...</i>	Y	protein_coding, transcribed_unprocessed_p seudogene	-1.35
17546502	<i>Ssty2</i>	<i>Gm20921, Gm20868, LOC101056308 ...</i>	Y	unprocessed_pseudogene	-1.37
17547291	<i>Ssty2</i>	<i>Gm20816, Gm20868, Gm20842 ...</i>	Y	protein_coding	-1.41
17547360	<i>Gm20865</i>	<i>Gm20854, Gm21180, Gm20809 ...</i>	Y	NA	-1.47
17547456	<i>Ssty2</i>	<i>Gm20815, Gm21242, Gm21180 ...</i>	Y	protein_coding	-1.50
17547277	<i>Gm20867</i>	<i>Gm20815, Gm20738, Gm20809 ...</i>	Y	unprocessed_pseudogene, transcribed_unprocessed_p seudogene ...	-1.52
17464312	<i>4930479D17Rik</i>		6	lincRNA	-1.84

WTEF3 vs YQEF3					
Transcript ID	Gene symbol	Aliases	Chr	Gene biotype	logFC
17541140	<i>Btg1-ps2</i>		X	protein_coding	1.16
17542897	<i>Cldn34-ps</i>		X	processed_pseudogene	1.02
17544385	<i>Tspan6</i>		X	protein_coding	0.94
17537527	<i>4921511C20Rik</i>		X	protein_coding	0.93
17543767	<i>4930519F16Rik</i>		X	unitary_pseudogene	0.92
17534425	<i>Stag2</i>		X	protein_coding	0.87
17533539	<i>Rp2</i>		X	protein_coding	0.86
17543562	<i>Gm614</i>		X	protein_coding	0.83
17533604	<i>Uba1</i>		X	protein_coding	0.77
17230327	NA		1	processed_pseudogene	-0.72
17546450	<i>Ssty2</i>	<i>Gm21658, Gm20852, Gm20867 ...</i>	Y	protein_coding	-0.87
17547261	<i>Ssty2</i>	<i>Gm21672, Gm20854, Gm21835 ...</i>	Y	protein_coding	-0.87
17547394	<i>Ssty2</i>	<i>Gm21180, Gm20823, Gm20738 ...</i>	Y	protein_coding	-0.87
17546498	<i>Ssty2</i>	<i>Gm20917, Gm21943, Gm21292 ...</i>	Y	protein_coding	-0.88
17546469	<i>Ssty2</i>	<i>Gm21242, Gm20842, Gm21799 ...</i>	Y	NA	-0.92
17547351	<i>Ssty2</i>	<i>Gm20854, LOC101056308, Gm20917 ...</i>	Y	NA	-0.94
17547378	<i>Ssty2</i>	<i>Gm20806, Gm20921, Gm21242 ...</i>	Y	NA	-0.94
17547385	<i>Ssty2</i>	<i>Gm20921, Gm20917, Gm20806 ...</i>	Y	NA	-0.96
17547231	<i>Ssty2</i>	<i>Gm21658, Gm21394, Gm21242 ...</i>	Y	protein_coding	-0.96
17547413	<i>Gm20842</i>	<i>Gm20806, Gm20809, Gm20907 ...</i>	Y	protein_coding	-0.97
17547428	<i>Ssty2</i>	<i>Gm20823, Gm21658, Gm20852 ...</i>	Y	protein_coding	-0.98
17547445	<i>Ssty2</i>	<i>Gm21394, LOC101056308,</i>	Y	protein_coding	-0.99

		<i>Gm21943 ...</i>			
17547456	<i>Ssty2</i>	<i>Gm21292, Gm20815, Gm29024 ...</i>	Y	protein_coding	-0.99
17547402	<i>Gm20879</i>	<i>Gm20826, Gm20898, Gm20868 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-1.00
17547288	<i>Ssty2</i>	<i>Gm20868, Gm20747, Gm20898 ...</i>	Y	protein_coding	-1.03
17547362	<i>Ssty2</i>	<i>Gm20815, Gm20854, Gm20852 ...</i>	Y	protein_coding	-1.03
17547271	<i>Ssty2</i>	<i>Gm21242, Gm21180, Gm20868 ...</i>	Y	protein_coding	-1.03
17547285	<i>Ssty2</i>	<i>Gm21943, Gm21242, Gm20867 ...</i>	Y	protein_coding	-1.03
17547295	<i>Ssty2</i>	<i>Gm20816, Gm20823, Gm21943 ...</i>	Y	protein_coding	-1.03
17547291	<i>Gm20809</i>	<i>Gm20898, Gm20815, Gm20823 ...</i>	Y	protein_coding	-1.09
17547360	<i>Gm20867</i>	<i>Gm20823, Gm20809, Gm21180 ...</i>	Y	NA	-1.10
17546502	<i>Ssty2</i>	<i>LOC101056308, Gm29024, Gm21118 ...</i>	Y	unprocessed_pseudogene	-1.13
17547172	NA		Y	protein_coding	-1.27
17547238	NA		Y	unprocessed_pseudogene	-1.46
17547454	NA		Y	unprocessed_pseudogene	-1.63
17547182	NA		Y	NA	-1.91
17546570	NA		Y	protein_coding	-2.11

WTEWC vs YQEWC	Transcript ID	Gene symbol	Aliases	Chr	Gene biotype	logFC
	17409804	<i>1700061117Rik</i>		3	lincRNA	1.57
	17544273	<i>Tgif2lx1</i>		X	protein_coding	1.23
	17533946	<i>Btbd35f5</i>		X	protein_coding	1.21
	17540370	<i>Ppp1r2-ps9</i>		X	transcribed_processed_pseudogene	1.12
	17537347	<i>Tgif2lx1</i>	<i>Tgif2lx2</i>	X	protein_coding	0.96
	17533539	<i>Rp2</i>		X	protein_coding	0.89

17542897	<i>Cldn34-ps</i>		X	processed_pseudogene	0.87
17534493	<i>Actrt1</i>		X	protein_coding	0.87
17548883	<i>4930453H23Rik</i>	<i>Gm15104, Gm20153</i>	X	NA	0.85
17489940	<i>Gm21129</i>		7	unprocessed_pseudogene	0.84
17489959	<i>Gm29258</i>	<i>Gm6818, Gm5591, Gm21129 ...</i>	7	unprocessed_pseudogene, protein_coding ...	0.83
17489968	<i>Gm1988</i>	<i>Gm6818, Gm5591, Gm6833 ...</i>	7	unprocessed_pseudogene, protein_coding ...	0.83
17536450	<i>Yipf6</i>		X	protein_coding	0.82
17538194	<i>Vsig1</i>		X	protein_coding	0.78
17541140	<i>Btg1-ps2</i>		X	protein_coding	0.78
17489923	<i>Gm4454</i>	<i>Gm21115</i>	7	unprocessed_pseudogene	0.76
17489943	<i>Gm21136</i>		7	unprocessed_pseudogene	0.75
17545514	<i>Gm15143</i>		X	unprocessed_pseudogene	0.71
17542888	<i>Cldn34d</i>		X	protein_coding	0.70
17303125	NA		14	protein_coding	0.70
17545526	<i>Gm16445</i>		X	unprocessed_pseudogene	0.69
17538320	<i>Pak3</i>		X	protein_coding	0.68
17540431	<i>Gm5123</i>	<i>Gm5755</i>	X	NA	0.68
17416196	<i>C8b</i>		4	protein_coding	0.68
17541041	<i>Gm9</i>		X	protein_coding	0.66
17535175	<i>Ctag2</i>		X	protein_coding	0.64
17548606	<i>Gm9271</i>	<i>Gm4884, Gm6124, Gm21136 ...</i>	7	unprocessed_pseudogene	0.62
17286805	NA		13	antisense	-0.55
17395345	<i>Gm14419</i>	<i>Gm14391, Gm14305, Gm14430 ...</i>	2	protein_coding	-0.59
17446757	NA		5	antisense	-0.63
17547203	<i>Gm20747</i>	<i>Gm20879</i>	Y	unprocessed_pseudogene	-0.70
17546452	<i>Gm21180</i>	<i>Gm20867, Gm20747, Gm20854 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.72
17547464	<i>Gm20738</i>	<i>Gm21180, Gm20867, Gm21672 ...</i>	Y	NA	-0.72
17546461	<i>Ssty2</i>	<i>Gm20815, Gm20867, Gm20934 ...</i>	Y	protein_coding, transcribed_unprocessed_p	-0.74

				seudogene	
17546463	Ssty2	Gm21394, Gm20852, Gm21672 ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-0.74
17546465	Gm21292	Gm29024, Gm20852, Gm21943 ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-0.74
17547191	Gm20747	Gm20879	Y	unprocessed_pseudogene	-0.74
17547389	Ssty2	Gm20868, Gm20867, Gm20816 ...	Y	transcribed_unprocessed_p seudogene, protein_coding	-0.75
17547235	Ssty2	Gm21394, Gm20823, Gm20865 ...	Y	transcribed_unprocessed_p seudogene, protein_coding	-0.75
17547355	Ssty2	Gm20809, Gm21943, Gm21672 ...	Y	transcribed_unprocessed_p seudogene, protein_coding	-0.75
17547226	Ssty2	Gm21180, Gm21292, Gm21672 ...	Y	transcribed_unprocessed_p seudogene, protein_coding	-0.78
17547346	Ssty2	Gm20738, Gm20823, Gm21394 ...	Y	transcribed_unprocessed_p seudogene, protein_coding	-0.78
17547223	Ssty2	Gm21394, Gm21672, Gm20815 ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-0.78
17547341	Ssty2	Gm20815, Gm20877, Gm21943 ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-0.78
17547375	Ssty2	Gm21394, Gm20816, Gm21242 ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-0.78
17547407	Ssty2	Gm20898, Gm21394, Gm21672 ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-0.78
17547410	Ssty2	Gm20809, Gm20815, Gm20865 ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-0.78
17546668	Ssty2	Gm20816, Gm21394, Gm20877 ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-0.78
17546471	NA		Y	unprocessed_pseudogene	-0.82
17546469	Ssty2	Gm20907, Gm20842, Gm20815 ...	Y	NA	-0.83
17546509	Ssty2	Gm20867, Gm21292, Gm20868 ...	Y	protein_coding, transcribed_unprocessed_p seudogene	-0.83

17546457	<i>Gm20877</i>	<i>Gm21835, Gm20816, Gm21180 ...</i>	Y	protein_coding	-0.84
17547415	<i>Ssty2</i>	<i>Gm20747, Gm20806, Gm20852 ...</i>	Y	protein_coding	-0.84
17547380	<i>Gm20868</i>		Y	NA	-0.84
17546498	<i>Ssty2</i>	<i>Gm20868, Gm20898, Gm20823 ...</i>	Y	protein_coding	-0.84
17547351	<i>Ssty2</i>	<i>Gm20868, Gm20747, Gm21943 ...</i>	Y	NA	-0.85
17547378	<i>Ssty2</i>	<i>Gm21943, Gm20806, Gm21242 ...</i>	Y	NA	-0.85
17547251	<i>Ssty2</i>	<i>Gm20815, Gm20852, Gm20917 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.86
17547281	<i>Ssty2</i>	<i>Gm20917, Gm20823, Gm20816 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding ...	-0.87
17547229	<i>Ssty2</i>	<i>Gm21672, Gm21943, Gm20854 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding ...	-0.88
17547443	<i>Ssty2</i>	<i>Gm21180, Gm21943, Gm20815 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding ...	-0.88
17547450	<i>Ssty2</i>	<i>Gm20867, Gm21292, Gm20917 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding ...	-0.88
17547392	<i>Gm20738</i>	<i>Gm20816, Gm20809, Gm20854 ...</i>	Y	unprocessed_pseudogene, protein_coding	-0.89
17547385	<i>Ssty2</i>	<i>Gm21799, Gm20826, Gm20934 ...</i>	Y	NA	-0.92
17547256	<i>Gm20934</i>	<i>Gm21835, Gm21180, Gm20815 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.92
17546502	<i>Ssty2</i>	<i>Gm20868, Gm21658, Gm20879 ...</i>	Y	unprocessed_pseudogene	-0.93
17547246	<i>Ssty2</i>	<i>Gm20809, Gm20806, Gm21943 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.93
17547428	<i>Gm20747</i>	<i>Gm20806, Gm21835, Gm20907 ...</i>	Y	protein_coding	-0.94
17547445	<i>Ssty2</i>	<i>Gm20865, Gm20815, Gm20823 ...</i>	Y	protein_coding	-0.94
17547423	<i>Ssty2</i>	<i>Gm21242, Gm21292, Gm20907 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.94

17547368	<i>Ssty2</i>	<i>Gm20823, Gm21242, Gm20868 ...</i>	Y	protein_coding	-0.94
17547420	<i>Ssty2</i>	<i>Gm21672, Gm21180, Gm20917 ...</i>	Y	protein_coding	-0.94
17547402	<i>Ssty2</i>	<i>Gm20879, Gm20917, Gm20747 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.94
17547283	<i>Gm20815</i>	<i>Gm20738, Gm20823, Gm20865 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.95
17546673	<i>Ssty2</i>	<i>Gm29024, Gm20879, Gm21180 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.96
17547231	<i>Ssty2</i>	<i>Gm21245, Gm21658, Gm20865 ...</i>	Y	protein_coding	-0.97
17547124	NA		Y	lincRNA	-0.98
17547456	<i>Ssty2</i>	<i>Gm21943, Gm20865, Gm29024 ...</i>	Y	protein_coding	-0.99
17547360	<i>Gm20865</i>	<i>Gm20854, Gm21180, Gm20815 ...</i>	Y	NA	-0.99
17547417	<i>Ssty2</i>	<i>Gm20747, Gm20738, Gm21242 ...</i>	Y	protein_coding	-1.00
17547285	<i>Ssty2</i>	<i>Gm20865, Gm21943, Gm20867 ...</i>	Y	protein_coding	-1.00
17547295	<i>Ssty2</i>	<i>Gm20816, Gm21180, Gm20738 ...</i>	Y	protein_coding	-1.00
17547365	<i>Ssty2</i>	<i>Gm20868, Gm21394, Gm21242 ...</i>	Y	protein_coding	-1.01
17547277	<i>Gm21180</i>	<i>Gm20809, Gm20816, Gm20865 ...</i>	Y	unprocessed_pseudogene, transcribed_unprocessed_pseudogene ...	-1.02
17546454	<i>Ssty2</i>	<i>Gm21672, Gm20738, Gm20806 ...</i>	Y	protein_coding	-1.02
17547271	<i>Ssty2</i>	<i>Gm20898, Gm20738, Gm20907 ...</i>	Y	protein_coding	-1.09
17547413	<i>Ssty2</i>	<i>Gm21242, Gm20879, Gm20747 ...</i>	Y	protein_coding	-1.09
17547291	<i>Gm29024,</i>	<i>Gm20868, Gm21943, Gm20816 ...</i>	Y	protein_coding	-1.12
17547288	<i>Ssty2</i>	<i>Gm20815, Gm21242, Gm20854 ...</i>	Y	protein_coding	-1.20
17547362	<i>Ssty2</i>	<i>Gm20852, Gm21943, Gm20917 ...</i>	Y	protein_coding	-1.20

17547454	NA		Y	unprocessed_pseudogene	-1.27
17464312	4930479D17Rik		6	lincRNA	-1.33
17226736	Mapkapk2		1	protein_coding	-1.58

WTLF3 vs YQLF3					
Transcript ID	Gene symbol	Aliases	Chr	Gene biotype	logFC
17540370	<i>Ppp1r2-ps9</i>		X	transcribed_processed_pseudogene	1.69
17459095	NA		6	rRNA	1.59
17532627	NA		M	Mt_tRNA, protein_coding	1.36
17541140	<i>Btg1-ps2</i>		X	protein_coding	1.36
17533469	<i>Cypt8</i>	<i>Cypt1, Cypt7</i>	X	protein_coding	1.31
17537212	<i>Cylc1</i>		X	protein_coding	1.20
17541134	<i>Btg1-ps1</i>		X	protein_coding	1.19
17482924	NA		7	TEC	1.13
17335540	<i>Pim1</i>		17	protein_coding	1.10
17398719	NA		3	snRNA	1.07
17469607	5031434C07Rik		6	antisense	-0.87
17410492	<i>Tet2</i>		3	protein_coding	-0.94
17547229	<i>Ssty2</i>	<i>Gm21943, Gm20747, Gm20816 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding ...	-0.98
17547443	<i>Ssty2</i>	<i>Gm20815, Gm20738, Gm21943 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding ...	-0.98
17547450	<i>Ssty2</i>	<i>Gm20738, Gm21292, Gm21672 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding ...	-0.98
17338231	NA		17	rRNA	-0.99
17547464	<i>Gm20898</i>	<i>Gm21292, Gm20823, Gm21672 ...</i>	Y	NA	-1.00
17547277	<i>Gm20809</i>	<i>Gm21672, Gm20738, Gm21180 ...</i>	Y	unprocessed_pseudogene, transcribed_unprocessed_pseudogene ...	-1.04
17547360	<i>Gm20823</i>	<i>Gm20738, Gm20815, Gm21672 ...</i>	Y	NA	-1.06

17547283	<i>Gm20815</i>	<i>Gm20854, Gm20816, Gm21292 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-1.07
17546452	<i>Gm21180</i>	<i>Gm20738, Gm21672, Gm21292 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-1.09
17487211	<i>Fosb</i>		7	protein_coding	-1.16

WTLWC vs YQLWC					
Transcript ID	Gene symbol	Aliases	Chr	Gene biotype	logFC
17444498	<i>Nptx2</i>		5	protein_coding	2.23
17532692	<i>LOC100861738</i>		Un*	NA	1.54
17539750	<i>Btbd35f16</i>		X	protein_coding	1.51
17339462	<i>4930471L23Rik</i>		17	NA	1.47
17545287	<i>Gm15104</i>		X	lincRNA	1.41
17532714	<i>Btbd35f17</i>		X	protein_coding	1.39
17533234	<i>Sytl5</i>		X	protein_coding	1.37
17537324	<i>Ube2dnl1</i>		X	protein_coding	1.33
17409804	<i>1700061117Rik</i>		3	lincRNA	1.32
17534493	<i>Actrt1</i>		X	protein_coding	1.30
17482924	NA		7	TEC	1.29
17237687	<i>4921513I03Rik</i>		10	lincRNA	1.29
17533946	<i>Btbd35f5</i>		X	protein_coding	1.26
17533195	<i>4930402K13Rik</i>		X	protein_coding	1.24
17541041	<i>Gm9</i>		X	protein_coding	1.14
17544668	<i>Tcp11x2</i>		X	protein_coding	1.12
17258007	<i>1700092K14Rik</i>		11	processed_transcript	1.12
17533469	<i>Cypt8</i>	<i>Cypt1, Cypt7</i>	X	protein_coding	1.11
17546355	<i>Rbmy</i>	<i>Gm10352, Gm4064, Gm21693 ...</i>	Y	protein_coding	1.07
17539023	<i>Cypt3</i>		X	protein_coding	1.07
17546953	<i>Rbmy</i>	<i>Gm21677, Gm4064, Gm10256 ...</i>	Y	protein_coding	1.07
17536450	<i>Yipf6</i>		X	protein_coding	1.06
17343823	<i>Gm20513</i>		17	lincRNA	1.05

17541190	<i>Cul4b</i>		X	protein_coding	1.05
17533937	<i>Btbd35f2</i>		X	protein_coding	1.05
17546333	<i>Rbmy</i>	<i>Gm4064, Gm10352, Gm21677 ...</i>	Y	protein_coding	1.05
17542239	<i>Cetn2</i>		X	protein_coding	1.04
17416196	<i>C8b</i>		4	protein_coding	1.03
17546963	<i>Rbmy</i>	<i>Gm3376, Gm21693, Gm10352 ...</i>	Y	protein_coding	1.02
17548883	<i>Gm20153</i>	<i>Gm15104, 4930453H23Rik</i>	X	NA	1.02
17534486	<i>Tex13c1</i>		X	protein_coding	1.02
17540370	<i>Ppp1r2-ps9</i>		X	transcribed_processed_pseudogene	1.01
17539548	<i>Piga</i>		X	protein_coding	1.01
17539754	<i>Btbd35f6</i>	<i>Btbd35f4, Btbd35f15, Btbd35f28 ...</i>	X	protein_coding, processed_pseudogene ...	1.00
17540885	<i>Btbd35f7</i>	<i>Btbd35f9, Btbd35f10, Gm5925 ...</i>	X	protein_coding, processed_pseudogene ...	1.00
17533539	<i>Rp2</i>		X	protein_coding	0.99
17533806	<i>Gm4907</i>		X	protein_coding	0.99
17540902	<i>Btbd35f6</i>		X	protein_coding	0.98
17532706	<i>Btbd35f2</i>	<i>Btbd35f7, Btbd35f20, Btbd35f22 ...</i>	X	protein_coding, processed_pseudogene ...	0.98
17532720	<i>Gm14367</i>	<i>Btbd35f26, Btbd35f22, Btbd35f17 ...</i>	X	protein_coding, processed_pseudogene ...	0.98
17539740	<i>Btbd35f4</i>	<i>Btbd35f12, Btbd35f22, Btbd35f23 ...</i>	X	protein_coding, processed_pseudogene ...	0.98
17540889	<i>Gm5925</i>	<i>Btbd35f9, Btbd35f22, Btbd35f23 ...</i>	X	protein_coding, processed_pseudogene ...	0.96
17539744	<i>Btbd35f21</i>	<i>Btbd35f3, LOC100048813, Btbd35f16 ...</i>	X	protein_coding, processed_pseudogene ...	0.96
17532702	<i>Gm5925</i>	<i>Btbd35f21, Btbd35f3, Btbd35f1 ...</i>	X	protein_coding, processed_pseudogene ...	0.96
17536215	<i>LOC101055749</i>		X	lincRNA	0.95
17540011	<i>Wdr13</i>		X	protein_coding	0.94
17548829	<i>Gm4916</i>		X	transcribed_processed_pseudogene, TEC	0.93

17287249	NA		13	NA	0.91
17533939	<i>Btbd35f1</i>		X	protein_coding	0.88
17546412	<i>Rbm31y</i>	<i>Gm21379</i>	Y	protein_coding, TEC	0.88
17325648	<i>Gm15802</i>		16	antisense	0.86
17535175	<i>Ctag2</i>		X	protein_coding	0.84
17511018	NA		8	antisense	0.84
17257819	<i>1700096J18Rik</i>		11	lincRNA	0.84
17533230	<i>H2al3</i>		X	protein_coding	0.84
17538320	<i>Pak3</i>		X	protein_coding	0.82
17545420	<i>Ribc1</i>		X	protein_coding	0.82
17245374	<i>Grip1os1</i>		10	antisense	0.81
17246033	<i>Gm16230</i>		10	antisense	0.81
17352018	<i>4930465K10Rik</i>		18	pseudogene	0.81
17285488	<i>Amph</i>		13	protein_coding	0.81
17538461	<i>Ott</i>	<i>Luzp4, Gm15107, Gm15114 ...</i>	X	protein_coding, unprocessed_pseudogene	0.80
17534843	<i>Slx1</i>	<i>Gm14594, Gm14595, Gm16430 ...</i>	X	protein_coding, unprocessed_pseudogene	0.79
17540203	<i>Rpgr</i>		X	protein_coding	0.78
17538470	<i>Ott</i>	<i>Luzp4, Gm15128, Gm15080 ...</i>	X	protein_coding, unprocessed_pseudogene	0.77
17545785	<i>Cdkl5</i>		X	protein_coding	0.77
17534781	<i>Slx1</i>	<i>Gm16404, Gm16405, Gm16430 ...</i>	X	protein_coding, unprocessed_pseudogene	0.77
17546434	<i>Rbm31y</i>	<i>Gm21379</i>	Y	protein_coding, TEC	0.77
17318083	<i>Ly6a</i>		15	protein_coding	0.77
17534849	<i>Slx1</i>	<i>Gm14590, Gm14594, Gm14596 ...</i>	X	protein_coding, unprocessed_pseudogene	0.77
17544102	<i>Brwd3</i>		X	protein_coding	0.76
17534820	<i>Slx1</i>	<i>Gm6660, Gm14590, Gm14596 ...</i>	X	protein_coding, unprocessed_pseudogene	0.76
17534767	<i>Gm10488</i>	<i>Gm6121, Gm4836, Gm6660 ...</i>	X	protein_coding, unprocessed_pseudogene	0.75
17533896	<i>Slx</i>	<i>Gm5169, Gm4836, Gm7391 ...</i>	X	unprocessed_pseudogene, protein_coding	0.74

17536436	<i>Ar</i>		X	protein_coding	0.74
17533843	<i>Slx</i>	<i>Gm14819, Gm5168, Gm5169 ...</i>	X	unprocessed_pseudogene, protein_coding	0.73
17533858	<i>Gm14632</i>	<i>Gm14819, Gm7391, Gm2012 ...</i>	X	unprocessed_pseudogene, protein_coding	0.73
17533882	<i>Slx</i>	<i>Gm10096, Gm10487, Gm1993 ...</i>	X	unprocessed_pseudogene, protein_coding	0.73
17533926	<i>Slx</i>	<i>Gm10147, Gm5169, Gm7391 ...</i>	X	unprocessed_pseudogene, protein_coding	0.73
17540846	<i>Gm10147</i>	<i>Gm5168, Gm10058, Gm4836 ...</i>	X	unprocessed_pseudogene, protein_coding	0.73
17540906	<i>Slx</i>	<i>Gm2003, Gm5934, Gm2030 ...</i>	X	unprocessed_pseudogene, protein_coding	0.73
17540634	<i>Slx</i>	<i>Gm10488, Gm1993, Gm14632 ...</i>	X	protein_coding, unprocessed_pseudogene	0.72
17363572	<i>Tjp2</i>		19	protein_coding	0.72
17534774	<i>Slx1</i>	<i>Gm16430, Gm6664, Gm16405 ...</i>	X	protein_coding, unprocessed_pseudogene	0.72
17540814	<i>Slx</i>	<i>Gm14632, Gm10147, Gm7391 ...</i>	X	unprocessed_pseudogene, protein_coding	0.72
17462705	<i>Foxj2</i>		6	protein_coding	0.71
17534827	<i>Slx1</i>	<i>Gm6660, Gm14594, Gm14625 ...</i>	X	protein_coding, unprocessed_pseudogene	0.71
17540804	<i>Slx</i>	<i>Gm10147, Gm5934, Gm2003 ...</i>	X	unprocessed_pseudogene, protein_coding	0.71
17540875	<i>Slx</i>	<i>Gm14819, Gm14525, Gm14632 ...</i>	X	unprocessed_pseudogene, protein_coding	0.71
17540774	<i>Slx</i>	<i>Gm2012, Gm1993, Gm6121 ...</i>	X	unprocessed_pseudogene, protein_coding	0.71
17326944	<i>Gm34826</i>	<i>Gm7735, Gm9789, Gm35004 ...</i>	16	protein_coding	0.70
17533911	<i>Slx</i>	<i>Gm10486, Gm2012, Gm6121 ...</i>	X	unprocessed_pseudogene, protein_coding	0.70
17533965	<i>Gm1993</i>	<i>Gm4836, Gm10486, Gm10487 ...</i>	X	unprocessed_pseudogene, protein_coding	0.70
17533981	<i>Slx</i>	<i>Gm5168, Gm10096, Gm14525 ...</i>	X	unprocessed_pseudogene, protein_coding	0.70
17540793	<i>Gm14525</i>	<i>Gm10486, Gm2030, Gm4836 ...</i>	X	unprocessed_pseudogene, protein_coding	0.70

17540824	<i>Slx</i>	<i>Gm5934, Gm1993, Gm10096 ...</i>	X	unprocessed_pseudogene, protein_coding	0.70
17534797	<i>Slx1</i>	<i>Gm14596, Gm14590, Gm14594 ...</i>	X	protein_coding, unprocessed_pseudogene	0.69
17367157	<i>Gm13266</i>		2	processed_transcript	0.69
17534791	<i>Slx1</i>	<i>Gm14596, Gm6660, Gm16430 ...</i>	X	protein_coding, unprocessed_pseudogene	0.69
17534806	<i>Slx1</i>	<i>Gm16430, Gm14590, Gm14596 ...</i>	X	protein_coding, unprocessed_pseudogene	0.69
17534813	<i>Slx1</i>	<i>Gm16405, Gm16404, Gm14625 ...</i>	X	protein_coding, unprocessed_pseudogene	0.69
17540766	<i>Gm10147</i>	<i>Gm10058, LOC102634388, Gm4297 ...</i>	X	unprocessed_pseudogene, protein_coding	0.68
17540838	<i>Gm4836</i>	<i>Gm14525, Gm2003, Gm4297 ...</i>	X	unprocessed_pseudogene, protein_coding	0.68
17538194	<i>Vsig1</i>		X	protein_coding	0.64
17412202	<i>Faxc</i>		4	protein_coding	0.64
17540713	<i>Slx</i>	<i>Gm6121, Gm2030, Gm5935 ...</i>	X	unprocessed_pseudogene, protein_coding	0.64
17533827	<i>Slx</i>	<i>Gm10230, Gm10096, Gm5934 ...</i>	X	unprocessed_pseudogene, protein_coding	0.63
17533866	<i>Slx</i>	<i>Gm5935, Gm10096, Gm10058 ...</i>	X	unprocessed_pseudogene, protein_coding	0.63
17540862	<i>Slx</i>	<i>Gm1993, Gm14525, Gm10096 ...</i>	X	unprocessed_pseudogene, protein_coding	0.63
17540921	<i>Slx</i>	<i>Gm5935, Gm10230, Gm14819 ...</i>	X	unprocessed_pseudogene, protein_coding	0.63
17537163	<i>Fam46d</i>		X	protein_coding	0.63
17214025	NA		1	snRNA	0.63
17540753	<i>Slx</i>	<i>Gm10487, Gm10096, Gm10488 ...</i>	X	protein_coding, unprocessed_pseudogene	0.61
17343675	<i>Gm19412</i>		17	antisense	0.60
17511014	NA		8	antisense, protein_coding	0.60
17305315	<i>Gm8256</i>	<i>1700091H14Rik, Gm7970, Gm8220 ...</i>	14	protein_coding	0.60
17540736	<i>Slx</i>	<i>Gm2030, Gm5935, Gm10486 ...</i>	X	unprocessed_pseudogene, protein_coding	0.59
17536997	<i>Zcchc13</i>		X	protein_coding	0.56

17547251	<i>Ssty2</i>	<i>Gm20806, Gm29024, Gm20879 ...</i>	Y	transcribed_unprocessed_pseudogene, protein_coding	-0.67
17547256	<i>Gm20806</i>	<i>Gm20816, Gm21118, Gm21180 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.74
17547385	<i>Ssty2</i>	<i>Gm20917, Gm20868, Gm21245 ...</i>	Y	NA	-0.75
17390217	NA		2	NA	-0.79
17546509	<i>Ssty2</i>	<i>Gm21180, Gm20815, Gm21292 ...</i>	Y	protein_coding, transcribed_unprocessed_pseudogene	-0.81
17547456	<i>Gm21672</i>	<i>Gm20815, Gm20852, Gm20747 ...</i>	Y	protein_coding	-0.82
17546450	<i>Gm20809</i>	<i>Gm21394, Gm20867, Gm20816 ...</i>	Y	protein_coding	-0.84
17547261	<i>Ssty2</i>	<i>Gm21672, Gm20865, Gm20809 ...</i>	Y	protein_coding	-0.84
17547394	<i>Ssty2</i>	<i>Gm21658, Gm21245, Gm20809 ...</i>	Y	protein_coding	-0.84
17547351	<i>Ssty2</i>	<i>Gm21943, Gm21658, Gm21835 ...</i>	Y	NA	-0.85
17547378	<i>Ssty2</i>	<i>Gm20921, Gm21245, Gm20854 ...</i>	Y	NA	-0.85
17547368	<i>Ssty2</i>	<i>Gm20815, Gm20747, Gm21180 ...</i>	Y	protein_coding	-0.89
17547420	<i>Ssty2</i>	<i>Gm20852, Gm21180, Gm21242 ...</i>	Y	protein_coding	-0.89
17546467	<i>Ssty2</i>	<i>Gm21394, Gm21799, Gm21943 ...</i>	Y	protein_coding, unprocessed_pseudogene	-0.89
17547445	<i>Gm20747</i>	<i>Gm20907, Gm20865, Gm20826 ...</i>	Y	protein_coding	-0.89
17547291	<i>Ssty2</i>	<i>Gm21245, Gm20868, Gm20879 ...</i>	Y	protein_coding	-0.91
17464312	<i>4930479D17Rik</i>		6	lincRNA	-0.97
17547285	<i>Ssty2</i>	<i>Gm20868, Gm20816, Gm20865 ...</i>	Y	protein_coding	-0.99
17547295	<i>Ssty2</i>	<i>Gm20868, Gm20867, Gm20815 ...</i>	Y	protein_coding	-0.99
17547392	<i>Gm20816</i>	<i>Gm21672, Gm20867, Gm21180 ...</i>	Y	unprocessed_pseudogene, protein_coding	-0.99

17546502	<i>Ssty2</i>	<i>Gm21799, Gm20934, Gm21943 ...</i>	Y	unprocessed_pseudogene	-1.03
17547231	<i>Ssty2</i>	<i>Gm21835, Gm21292, Gm20934 ...</i>	Y	protein_coding	-1.06
17547428	<i>Ssty2</i>	<i>Gm20738, Gm20898, Gm20852 ...</i>	Y	protein_coding	-1.07
17226736	<i>Mapkapk2</i>		1	protein_coding	-1.23

Differentially expressed non-coding genes and pseudogenes showing potential cis-limitation in microarray data

Gene symbol / Transcript cluster ID	Biotype	Chr	WT Early			YQ Early			WT Late			YQ Late		
			1	2	3	1	2	3	1	2	3	1	2	3
<i>1700061117Rik</i>	lincRNA	3	1.3	1.7	0.0	2.2	2.8	-1.1	-1.4	0.0	1.6	0.0	0.0	1.0
<i>4932415M13Rik</i>	lincRNA	17	4.2	3.6	-2.1	4.1	3.6	-2.6	2.4	2.7	-1.9	1.7	1.6	-0.9
<i>5031434C07Rik</i>	antisense	6	0.0	0.0	0.0	0.0	0.0	0.0	2.4	2.9	-2.6	1.0	1.4	-1.1
<i>A930024E05Rik</i>	antisense	5	3.3	3.6	-1.9	3.1	2.7	-1.8	1.9	2.2	-1.2	0.8	1.1	-0.6
<i>Gm10147</i>	unprocessed pseudogene	X	2.5	1.9	-1.0	2.1	1.7	-0.7	1.2	1.1	-0.7	1.6	1.2	-0.8
<i>Gm10147</i>	unprocessed pseudogene	X	2.3	1.8	-0.9	2.0	1.7	-0.7	1.0	0.7	-0.6	1.2	1.1	-0.7
<i>Gm10634</i>	lincRNA, antisense	9	4.8	4.9	-2.3	4.8	4.2	-2.2	4.3	4.2	-2.8	4.2	4.2	-2.8
<i>Gm13266</i>	processed transcript	2	0.5	1.0	-0.9	0.0	1.3	-1.2	-1.0	0.0	0.7	-1.0	-1.2	0.6
<i>Gm14525</i>	unprocessed pseudogene	X	2.4	1.7	-0.9	2.0	1.6	-0.6	1.0	0.7	-0.5	1.4	1.1	-0.7
<i>Gm14632</i>	unprocessed pseudogene	X	2.3	1.8	-0.9	2.0	1.7	-0.7	1.0	0.7	-0.6	1.2	1.1	-0.7
<i>Gm1988</i>	unprocessed pseudogene	7	3.6	2.3	-2.1	3.2	1.9	-1.7	0.0	0.0	0.0	0.0	0.0	0.0
<i>Gm1988</i>	unprocessed pseudogene	7	3.0	2.4	-1.9	3.2	1.9	-1.7	2.2	1.6	-1.7	2.0	1.7	-1.5
<i>Gm1993</i>	unprocessed pseudogene	X	2.4	1.7	-0.9	2.0	1.6	-0.6	1.0	0.7	-0.5	1.4	1.1	-0.7
<i>Gm20747</i>	unprocessed pseudogene	Y	1.0	0.7	-0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Gm20747</i>	unprocessed pseudogene	Y	1.1	0.8	-0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Gm20809</i>	unprocessed pseudogene	Y	1.3	0.7	-0.9	0.0	1.1	-0.8	0.0	0.0	0.0	0.0	0.0	0.0
<i>Gm20815</i>	transcribed unprocessed pseudogene	Y	1.3	0.8	-0.8	0.0	1.2	-0.8	0.0	0.0	0.0	0.0	0.0	0.0

<i>Gm20816</i>	unprocessed pseudogene	Y	1.1	0.8	-0.8	0.0	1.0	-0.7	0.0	0.0	0.0	0.0	0.0	0.0
<i>Gm20823</i>	NA	Y	1.2	0.8	-0.8	0.0	1.1	-0.7	0.0	0.0	0.0	0.0	0.0	0.0
<i>Gm20868</i>	NA	Y	1.3	0.9	-0.9	0.0	0.8	0.0	0.0	0.0	0.8	0.0	-0.8	0.0
<i>Gm20879</i>	unprocessed pseudogene	Y	0.9	0.6	-0.5	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Gm20898</i>	NA	Y	1.3	0.8	-0.8	0.0	1.0	-0.8	0.0	0.0	0.0	0.0	0.0	0.0
<i>Gm21129</i>	unprocessed pseudogene	7	3.1	2.5	-1.9	2.9	1.8	-1.5	1.6	1.0	-1.1	1.4	1.0	-0.7
<i>Gm21136</i>	unprocessed pseudogene	7	3.3	2.7	-2.2	3.4	2.0	-1.9	0.0	0.0	-1.0	1.4	0.9	-0.8
<i>Gm21180</i>	transcribed unprocessed pseudogene	Y	1.3	0.9	-0.9	0.8	1.2	-0.9	0.0	0.0	0.0	0.0	0.0	0.0
<i>Gm29258</i>	unprocessed pseudogene	7	3.0	2.4	-1.9	3.2	1.9	-1.7	2.2	1.6	-1.7	2.0	1.7	-1.5
<i>Gm4454</i>	unprocessed pseudogene	7	3.4	2.7	-2.1	3.3	2.1	-1.9	1.1	0.7	0.0	1.6	1.1	-1.4
<i>Gm4836</i>	unprocessed pseudogene	X	2.5	1.9	-1.0	2.1	1.7	-0.7	1.2	1.1	-0.7	1.6	1.2	-0.8
<i>Gm9271</i>	unprocessed pseudogene	7	3.0	2.5	-1.8	3.0	2.0	-1.4	0.0	0.0	0.0	0.0	1.0	-0.5
<i>LOC100861738</i>	NA	Y	2.1	1.8	-1.7	2.0	2.7	-1.5	2.5	2.7	-2.4	2.5	2.9	-1.9
17547238	unprocessed pseudogene	Y	3.2	2.6	-1.8	1.5	2.6	-1.7	0.0	1.1	0.0	0.0	0.0	0.0
17547182	NA	Y	4.3	4.0	-2.3	2.9	2.9	-1.3	1.8	2.1	-1.7	1.5	0.0	0.0
17547124	lincRNA	Y	5.5	3.7	-2.2	4.7	3.5	-1.8	4.3	3.8	-2.6	2.5	2.8	-2.2
17511018	antisense	8	1.2	1.3	-1.2	0.7	0.9	-1.0	0.0	0.0	0.0	-1.0	-1.2	0.9
17511014	antisense	8	2.9	3.0	-1.4	2.4	3.1	-1.5	1.0	0.9	-0.6	0.0	0.0	0.0
17398719	snRNA	3	0.9	1.2	-1.8	1.0	1.6	-1.8	1.2	1.5	-1.0	3.2	3.4	-2.9
17338231	rRNA	17	0.0	0.0	0.0	0.0	0.0	0.0	4.1	2.6	-2.4	2.9	1.8	-1.6

17287249	NA	13	3.7	3.5	-2.5	3.1	2.9	-2.4	4.1	3.4	-2.9	3.6	2.9	-2.3
17286805	antisense	13	1.1	1.7	-1.0	1.0	1.6	-1.4	0.0	0.0	0.0	-0.8	0.0	0.0
17214025	snRNA	1	2.3	1.3	-2.3	1.9	1.5	-2.7	3.1	2.2	-3.1	3.3	2.6	-2.9

Candidate genes short-listed for potential involvement in sex ratio distortion as identified through gene expression analysis by microarray. Candidates listed below show significant differential expression in at least one cell type when compared between WT and YQ samples, and transcript distributions that indicate potential cis-limitation. Gene annotations from STRING and UniProt are included. Some annotations have been trimmed for brevity. Points of particular interest to the investigation of sex ratio distortion in the YQ model are highlighted in **bold**.

Gene Symbol	Chr	Annotation
<i>Ar</i>	X	Androgen receptor; Steroid hormone receptors are ligand-activated transcription factors that regulate eukaryotic gene expression and affect cellular proliferation and differentiation in target tissues. Transcription factor activity is modulated by bound coactivator and corepressor proteins. Transcription activation is down-regulated by NR0B2. Activated, but not phosphorylated, by HIPK3 and ZIPK/DAPK3
<i>Brwd3</i>	X	Bromodomain and WD repeat domain containing 3; Plays a role in the regulation of cell morphology and cytoskeletal organization. Required in the control of cell shape (By similarity)
<i>Cdkl5</i>	X	Cyclin-dependent kinase-like 5 ; Mediates phosphorylation of MECP2
<i>Cldn34-ps</i>	X	Claudin 34 (pseudogene); RIKEN cDNA 4930468A15 gene
<i>Cylc1</i>	X	Cylicin, basic protein of sperm head cytoskeleton 1
<i>Fosb</i>	7	FBJ osteosarcoma oncogene B; FosB interacts with Jun proteins enhancing their DNA binding activity
<i>LOC100861738</i>	Y	eukaryotic translation elongation factor 2 (pseudogene)
<i>Ly6a</i>	15	Lymphocyte antigen 6 complex, locus A; T-cell activation
<i>Stag2</i>	X	Stromal antigen 2; Component of cohesin complex , a complex required for the cohesion of sister chromatids after DNA replication. The cohesin complex apparently forms a large proteinaceous ring within which sister chromatids can be trapped. At anaphase, the complex is cleaved and dissociates from chromatin, allowing sister chromatids to segregate. The cohesin complex may also play a role in spindle pole assembly during mitosis (By similarity)
<i>Tet2</i>	3	Tet methylcytosine dioxygenase 2; Dioxygenase that catalyzes the conversion of the modified genomic base 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC) and plays a key role in active DNA demethylation . Has a preference for 5-hydroxymethylcytosine in CpG motifs. In addition to its role in DNA demethylation, also involved in the recruitment of the O-GlcNAc transferase OGT to CpG-rich transcription start sites of active genes, thereby promoting histone H2B GlcNAcylation by OGT.

Genes showing strong up-regulation in YQ early stage spermatids. Included below are gene annotations retrieved from STRING, UNiProt and basic literature searches; alongside chromosomal origin and potential links to physiological consequences relevant to the YQ model. In addition to genes discussed below were several further genes for which no further functional information seemed available. These comprise *Rik and predicted (Gm*) genes, including several predicted genes located on chromosome 7 within a CNV region which were also up-regulated in YQEF2 and YQEWC.

Gene symbol	Chr	Name
<i>1700061117Rik</i>	3	RIKEN cDNA 1700061117 gene
Little information available; primarily testis expressed and predicted to interact with a large number of miRNAs.		
<i>Tgif2lx1</i>	X	TGFB-induced factor homeobox 2-like, X-linked 1
Known spermatid expressed gene which exists in two copies on the X-chromosome (both are seen to be upregulated in our data) and posses a Y-linked homologue. The TGIF family encodes transcription factors likely important during spermatogenesis (Stouffs et al. 2009).		
<i>Ppp1r2-ps9</i>	X	protein phosphatase 1, regulatory (inhibitor) subunit 2, pseudogene 9
Ppp1r2-p9 possesses homologues by the same name in several murid species, but limited sequence similarity to other protein phosphatase genes. Little information is available about the function of this gene.		
<i>Rp2</i>	X	retinitis pigmentosa 2 homolog
A GTP-ase-activating protein (GAP) involved in trafficking between the golgi and the ciliary membrane. Involved in the localisation of proteins. Also acts as a GAP for tubulin together with tubulin-specific chaperone C. Tubulin is a key component of both the sperm manchette and axoneme.		
<i>Yipf6</i>	X	Yip1 domain family, member 6
Yipf6 is a member of a gene family that, in yeast, is known to regulate vesicular transport. Its function in mammals is unknown. However, Brandl et al. (2012) showed a role tissue-specific role for Yipf6 in the maintenance of intestinal homeostasis in mice.		
<i>Vsig1</i>	X	V-set and immunoglobulin domain containing 1
Vsig1 has previously been recognised as be strongly up-regulated in the round spermatids of MSYq males, and has been investigated as an offspring sperm specific antigen (Ellis et al. 2011, Ellis et al. 2005). Work by Kim et al. (2010) showed that VSIG1 is localised on the germ cell surface and interacts with Sertoli cells by heterophilic adhesion. It also specifically binds with ZO-1, the central structural protein of the tight junction, via its cytoplasmic domain.		
<i>Btg1-ps2</i>	X	B cell translocation gene 1, anti-proliferative, pseudogene 2
Ensembl39, the genomics reference database, lists Btg1-ps2 as having a protein-coding transcript variant, which is not distinguishable from the detected 'pseudogene' here listed. It is implicated in the negative regulation of cell proliferation.		

³⁹ Ensembl - <http://www.ensembl.org/index.html>

No detected up-regulation of sex-linked genes in mature sperm samples

As transcription of protein-coding genes in mature sperm is known to be minimal, an upregulation of sex-linked genes is not expected. Many transcripts remaining are likely to be those retained from previous developmental stages, although the composition of mRNAs within sperm has been linked to differential sperm fitness parameters (Lambard et al. 2004, Lambard et al. 2003). No transcript clusters were recorded as being significantly differentially expressed between the two samples. The remaining protein-coding genes detected may be influenced by the transcriptional activity of the few identified contaminating cells, such as leydig cells, alongside retention of mRNAs from previous maturation stages.

Bibliography

- Abbott, J. K., Nordén, A. K., & Hansson, B. (2017). Sex chromosome evolution: historical insights and future perspectives. *Proceedings of the Royal Society B: Biological Sciences*, 284(1854). <https://doi.org/10.1098/rspb.2016.2806>
- Abou-Haila, A., & Tulsiani, D. R. (2000). Mammalian sperm acrosome: formation, contents, and function. *Archives of Biochemistry and Biophysics*, 379(2), 173–182. <https://doi.org/10.1006/abbi.2000.1880>
- Affymetrix (retrieved 2017). Mouse Gene Array Datasheet - GeneChip® Mouse Gene 2.0 ST Array. Retrieved from: http://tools.thermofisher.com/content/sfs/brochures/mouse_gene_array_datasheet.pdf on 6th July 2017.
- Alm, K., Peltoniemi, O., Koskinen, E., & Andersson, M. (2006). Porcine Field Fertility with Two Different Insemination Doses and the Effect of Sperm Morphology. *Reproduction in Domestic Animals*, 41(3), 210–213. <https://doi.org/10.1111/j.1439-0531.2005.00670.x>
- Almog, T., & Naor, Z. (2010). The role of Mitogen activated protein kinase (MAPK) in sperm functions. *Molecular and Cellular Endocrinology*, 314(2), 239–243. <https://doi.org/10.1016/j.mce.2009.05.009>
- Amann, R. P., & Waberski, D. (2014). Computer-assisted sperm analysis (CASA): Capabilities and potential developments. *Theriogenology*, 81(1), 5–17.e3. <https://doi.org/10.1016/j.theriogenology.2013.09.004>
- Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11(10), R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
- Andraszek, K., & Smalec, E. (2011). The use of silver nitrate for the identification of spermatozoon structure in selected mammals. *Canadian Journal of Animal Science*, 91(2), 239–246. <https://doi.org/10.4141/CJAS10052>
- Andrews et al (2010). Fastqc: A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Aranha, I. P., & Martin-DeLeon, P. A. (1991). The murine Rb(6.16) translocation: evidence for sperm selection and a modulating effect of aging. *Human Genetics*, 87(3), 278–284. <https://doi.org/10.1007/BF00200904>
- Aranha, I. P., & Martin-DeLeon, P. A. (1995). Mouse chromosome 6 in Rb translocations: consequences in singly and doubly heterozygous males. *Cytogenetic and Genome Research*, 69(3–4), 253–259. <https://doi.org/10.1159/000133975>
- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., ... Sheridan, R. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*; London, 442(7099), 203–7. <https://doi.org/http://dx.doi.org/10.1038/nature04916>
- Atlan, A., Capillon, C., Derome, N., Couvet, D., & Montchamp-Moreau, C. (2003). The evolution of autosomal suppressors of sex-ratio drive in *Drosophila simulans*. *Genetica*, 117(1), 47–58.

- Axtell, M. J., Westholm, J. O., & Lai, E. C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biology*, 12, 221. <https://doi.org/10.1186/gb-2011-12-4-221>
- Ba, W., Yan, Y., Reijnders, M. R. F., Schuurs-Hoeijmakers, J. H. M., Feenstra, I., Bongers, E. M. H. F., ... De Vries, B. B. A. (2016). TRIO loss of function is associated with mild intellectual disability and affects dendritic branching and synapse function. *Human Molecular Genetics*, 25(5), 892–902. <https://doi.org/10.1093/hmg/ddv618>
- Baba, D., Kashiwabara, S., Honda, A., Yamagata, K., Wu, Q., Ikawa, M., ... Baba, T. (2002). Mouse Sperm Lacking Cell Surface Hyaluronidase PH-20 Can Pass through the Layer of Cumulus Cells and Fertilize the Egg. *Journal of Biological Chemistry*, 277(33), 30310–30314. <https://doi.org/10.1074/jbc.M204596200>
- Babraham Bioinformatics (2011). FastQC: a quality control tool for high throughput sequence data. Babraham Institute, Cambridge, United Kingdom: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Bak, C. W., Yoon, T.-K., & Choi, Y. (2011). Functions of PIWI proteins in spermatogenesis. *Clinical and Experimental Reproductive Medicine*, 38(2), 61–67. <https://doi.org/10.5653/cerm.2011.38.2.61>
- Baker, R. H., Narechania, A., DeSalle, R., Johns, P. M., Reinhardt, J. A., & Wilkinson, G. S. (2016). Spermatogenesis Drives Rapid Gene Creation and Masculinization of the X Chromosome in Stalk-Eyed Flies (Diptera). *Genome Biology and Evolution*, 8(3), 896–914. <https://doi.org/10.1093/gbe/evw043>
- Banaszewska, D., Andraszek, K., Czubaszek, M., & Biesiada-Drzazga, B. (2015). The effect of selected staining techniques on bull sperm morphometry. *Animal Reproduction Science*, 159, 17–24. <https://doi.org/10.1016/j.anireprosci.2015.06.019>
- Barr, J., Yakovlev, K. V., Shidlovskii, Y., & Schedl, P. (2016). Establishing and maintaining cell polarity with mRNA localization in *Drosophila*. *BioEssays*, 38(3), 244–253. <https://doi.org/10.1002/bies.201500088>
- Bartoov, B., Berkovitz, A., Eltes, F., Kogosowski, A., Menezes, Y., & Barak, Y. (2002). Real-Time Fine Morphology of Motile Human Sperm Cells is Associated With IVF-ICSI Outcome. *Journal of Andrology*, 23(1), 1–8. <https://doi.org/10.1002/j.1939-4640.2002.tb02595.x>
- Basu, S., Campbell, H. M., Dittel, B. N., & Ray, A. (2010). Purification of Specific Cell Population by Fluorescence Activated Cell Sorting (FACS). *Journal of Visualized Experiments: JoVE*, (41). <https://doi.org/10.3791/1546>
- Beg, M. S., Brenner, A. J., Sachdev, J., Borad, M., Kang, Y.-K., Stoudemire, J., ... Hong, D. S. (2017). Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors. *Investigational New Drugs*, 35(2), 180–188. <https://doi.org/10.1007/s10637-016-0407-y>
- Beletti, M., da Fontoura Costa, L., & Mendes Guardieiro, M. (2005). Morphometric features and chromatin condensation abnormalities evaluated by Toluidine Blue staining in bull spermatozoa. *Braz J Morphol Sci*, 22.

- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289-300, 1995.
- Beukeboom, L. W., & Perrin, N. (2014). *The Evolution of Sex Determination*. Oxford University Press.
- Birkhead, T. R., Hosken, D. J., & Pitnick, S. S. (2008). *Sperm Biology: An Evolutionary Perspective*. Academic Press.
- Bjartmar, L., Huberman, A. D., Ullian, E. M., Rentería, R. C., Liu, X., Xu, W., ... Perin, M. S. (2006). Neuronal Pentraxins Mediate Synaptic Refinement in the Developing Visual System. *Journal of Neuroscience*, 26(23), 6269–6281. <https://doi.org/10.1523/JNEUROSCI.4212-05.2006>
- Bohrer, B. M. (2017). Review: Nutrient density and nutritional value of meat products and non-meat foods high in protein. *Trends in Food Science & Technology*, 65, 103–112. <https://doi.org/10.1016/j.tifs.2017.04.016>
- Bolstad, B. M., Irizarry, R. A., Åstrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2), 185–193. <https://doi.org/10.1093/bioinformatics/19.2.185>
- Bongso, T. A. (1983). Comparative Silver Staining Patterns of Water Buffalo, Goat, and Pig Spermatozoa. *Archives of Andrology*, 11(1), 13–17. <https://doi.org/10.3109/01485018308987453>
- Bourgon, R., Gentleman, R., & Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences*, 107(21), 9546–9551. <https://doi.org/10.1073/pnas.0914005107>
- Branco, A. T., Tao, Y., Hartl, D. L., & Lemos, B. (2013). Natural variation of the Y chromosome suppresses sex ratio distortion and modulates testis-specific gene expression in *Drosophila simulans*. *Heredity*, 111(1), 8–15. <https://doi.org/10.1038/hdy.2013.5>
- Brandl, K., Tomisato, W., Li, X., Nepl, C., Pirie, E., Falk, W., ... Beutler, B. (2012). Yip1 domain family, member 6 (Yipf6) mutation induces spontaneous intestinal inflammation in mice. *Proceedings of the National Academy of Sciences*, 109(31), 12650–12655. <https://doi.org/10.1073/pnas.1210366109>
- Braun, R. E., Behringer, R. R., Peschon, J. J., Brinster, R. L., & Palmiter, R. D. (1989). Genetically haploid spermatids are phenotypically diploid. *Nature*, 337(6205), 373–376. <https://doi.org/10.1038/337373a0>
- Breed, W., M Leigh, C., Aplin, K., Shahin, A., & Avenant, N. (2014). Morphological Diversity and Evolution of the Spermatozoon in the Mouse-Related Clade of Rodents. *Journal of Morphology*, 275, 540–547. <https://doi.org/10.1002/jmor.20236>
- Breitbart, H. (2002). Intracellular calcium regulation in sperm capacitation and acrosomal reaction. *Molecular and Cellular Endocrinology*, 187(1), 139–144. [https://doi.org/10.1016/S0303-7207\(01\)00704-3](https://doi.org/10.1016/S0303-7207(01)00704-3)

- Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B., & Cohen, S. M. (2003). bantam Encodes a Developmentally Regulated microRNA that Controls Cell Proliferation and Regulates the Proapoptotic Gene hid in *Drosophila*. *Cell*, 113(1), 25–36. [https://doi.org/10.1016/S0092-8674\(03\)00231-9](https://doi.org/10.1016/S0092-8674(03)00231-9)
- Brewer, L. R., Corzett, M., & Balhorn, R. (1999). Protamine-Induced Condensation and Decondensation of the Same DNA Molecule. *Science*, 286(5437), 120–123. <https://doi.org/10.1126/science.286.5437.120>
- Brewer, L., Corzett, M., & Balhorn, R. (2002). Condensation of DNA by Spermatid Basic Nuclear Proteins. *Journal of Biological Chemistry*, 277(41), 38895–38900. <https://doi.org/10.1074/jbc.M204755200>
- Brüchert, N., Mavila, N., Boknik, P., Baba, H. A., Fabritz, L., Gergs, U., ... Neumann, J. (2008). Inhibitor-2 prevents protein phosphatase 1-induced cardiac hypertrophy and mortality. *American Journal of Physiology - Heart and Circulatory Physiology*, 295(4), H1539–H1546. <https://doi.org/10.1152/ajpheart.00515.2008>
- Buffone, M. G. (2016). *Sperm Acrosome Biogenesis and Function During Fertilization*. Springer.
- Burgoyne, P. S., Mahadevaiah, S. K., Sutcliffe, M. J., & Palmer, S. J. (1992). Fertility in mice requires X-Y pairing and a Y-chromosomal ‘Spermiogenesis’ gene mapping to the long arm. *Cell*, 71(3), 391–398. [https://doi.org/10.1016/0092-8674\(92\)90509-B](https://doi.org/10.1016/0092-8674(92)90509-B)
- Burgoyne, P. S., Mahadevaiah, S. K., & Turner, J. M. A. (2009). The consequences of asynapsis for mammalian meiosis. *Nature Reviews Genetics*, 10(3), 207–216. <https://doi.org/10.1038/nrg2505>
- Caldwell, K. A., & Handel, M. A. (1991). Protamine transcript sharing among postmeiotic spermatids. *Proceedings of the National Academy of Sciences of the United States of America*, 88(6), 2407–2411.
- Caneparo L, Pantazis P, Dempsey W, Fraser SE (2011) Intercellular Bridges in Vertebrate Gastrulation. *PLOS ONE* 6(5): e20230. doi: 10.1371/journal.pone.0020230
- Carvalho, A. B., Vaz, S. C., & Klaczko, L. B. (1997). Polymorphism for Y-linked suppressors of sex-ratio in two natural populations of *Drosophila mediopunctata*. *Genetics*, 146(3), 891–902.
- Carvalho B. S., and Irizarry, R. A. 2010. A Framework for Oligonucleotide Microarray Preprocessing. *Bioinformatics*. V.26(19):2363-7, doi: 10.1093/bioinformatics/btq431, Package version: 1.38.0
- Case, L. K., Wall, E. H., Osmanski, E. E., Dragon, J. A., Saligrama, N., Zachary, J. F., ... Teuscher, C. (2015). Copy number variation in Y chromosome multicopy genes is linked to a paternal parent-of-origin effect on CNS autoimmune disease in female offspring. *Genome Biology*, 16, 28. <https://doi.org/10.1186/s13059-015-0591-7>
- Chang, Y.-F., Lee-Chang, J. S., Panneerdoss, S., MacLean, J. A., & Rao, M. K. (2011). Isolation of Sertoli, Leydig, and spermatogenic cells from the mouse testis. *BioTechniques*, 51(5), 341–342, 344. <https://doi.org/10.2144/000113764>
- Charlat, S., Hurst, G. D. D., & Merçot, H. (2003). Evolutionary consequences of Wolbachia infections. *Trends in Genetics: TIG*, 19(4), 217–223. [https://doi.org/10.1016/S0168-9525\(03\)00024-6](https://doi.org/10.1016/S0168-9525(03)00024-6)

- Chemes, H. E., & Rawe, V. Y. (2003). Sperm pathology: a step beyond descriptive morphology. Origin, characterization and fertility potential of abnormal sperm phenotypes in infertile men. *Human Reproduction Update*, 9(5), 405–428. <https://doi.org/10.1093/humupd/dmg034>
- Chemes, H. E., & Alvarez Sedo, C. (2012). Tales of the Tail and Sperm Head Aches Changing concepts on the prognostic significance of sperm pathologies affecting the head, neck and tail. *Asian Journal of Andrology*, 14(1), 14–23. <https://doi.org/10.1038/aja.2011.168>
- Chen, H., Griffiths, G., Galileo, D. S., & Martin-DeLeon, P. A. (2006). Epididymal SPAM1 Is a Marker for Sperm Maturation in the Mouse. *Biology of Reproduction*, 74(5), 923–930. <https://doi.org/10.1095/biolreprod.105.048587>
- Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., ... Zhou, Q. (2016a). Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science (New York, N.Y.)*, 351(6271), 397–400. <https://doi.org/10.1126/science.aad7977>
- Chen, Q., Yan, W., & Duan, E. (2016b). Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nature Reviews Genetics*, 17(12), 733–743. <https://doi.org/10.1038/nrg.2016.106>
- Cherr, G. N., Yudin, A. I., & Overstreet, J. W. (2001). The dual functions of GPI-anchored PH-20: hyaluronidase and intracellular signaling. *Matrix Biology*, 20(8), 515–525. [https://doi.org/10.1016/S0945-053X\(01\)00171-8](https://doi.org/10.1016/S0945-053X(01)00171-8)
- Chi, Y., & Zhou, D. (2016). MicroRNAs in colorectal carcinoma - from pathogenesis to therapy. *Journal of Experimental & Clinical Cancer Research*, 35, 43. <https://doi.org/10.1186/s13046-016-0320-4>
- Cho, C., Willis, W. D., Goulding, E. H., Jung-Ha, H., Choi, Y. C., Hecht, N. B., & Eddy, E. M. (2001). Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nature Genetics*, 28(1), 82–86. <https://doi.org/10.1038/88313>
- Cho, C., Jung-Ha, H., Willis, W. D., Goulding, E. H., Stein, P., Xu, Z., ... Eddy, E. M. (2003). Protamine 2 deficiency leads to sperm DNA damage and embryo death in mice. *Biology of Reproduction*, 69(1), 211–217. <https://doi.org/10.1095/biolreprod.102.015115>
- Cho, S., Jang, I., Jun, Y., Yoon, S., Ko, M., Kwon, Y., ... Lee, S. (2013). MiRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Research*, 41(Database issue), D252–257. <https://doi.org/10.1093/nar/gks1168>
- Chou, C.-H., Chang, N.-W., Shrestha, S., Hsu, S.-D., Lin, Y.-L., Lee, W.-H., ... Huang, H.-D. (2016). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Research*, 44(D1), D239–D247. <https://doi.org/10.1093/nar/gkv1258>
- Christopher, A. F., Kaur, R. P., Kaur, G., Kaur, A., Gupta, V., & Bansal, P. (2016). MicroRNA therapeutics: Discovering novel targets and developing specific therapy. *Perspectives in Clinical Research*, 7(2), 68–74. <https://doi.org/10.4103/2229-3485.179431>

- Chuma, S., Hosokawa, M., Kitamura, K., Kasai, S., Fujioka, M., Hiyoshi, M., ... Nakatsuji, N. (2006). Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. *Proceedings of the National Academy of Sciences*, 103(43), 15894–15899. <https://doi.org/10.1073/pnas.0601878103>
- Cloonan, N. (2015). Re-thinking miRNA-mRNA interactions: Intertwining issues confound target discovery. *BioEssays*, 37(4), 379–388. <https://doi.org/10.1002/bies.201400191>
- Cocquet, J., Ellis, P. J. I., Yamauchi, Y., Mahadevaiah, S. K., Affara, N. A., Ward, M. A., & Burgoyne, P. S. (2009). The Multicopy Gene Sly Represses the Sex Chromosomes in the Male Mouse Germline after Meiosis. *PLoS Biol*, 7(11), e1000244. <https://doi.org/10.1371/journal.pbio.1000244>
- Cocquet, J., Ellis, P. J. I., Yamauchi, Y., Riel, J. M., Karacs, T. P. S., Rattigan, A., ... Burgoyne, P. S. (2010). Deficiency in the multicopy Sycp3-like X-linked genes Slx and Slx1l causes major defects in spermatid differentiation. *Molecular Biology of the Cell*, 21(20), 3497–3505. <https://doi.org/10.1091/mbc.E10-07-0601>
- Cocquet, J., Ellis, P. J. I., Mahadevaiah, S. K., Affara, N. A., Vaiman, D., & Burgoyne, P. S. (2012). A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. *PLoS Genet*, 8(9), e1002900. <https://doi.org/10.1371/journal.pgen.1002900>
- Comptour, A., Moretti, C., Serrentino, M.-E., Auer, J., Ialy-Radio, C., Ward, M. A., ... Cocquet, J. (2014). SSTY proteins co-localize with the post-meiotic sex chromatin and interact with regulators of its expression. *The FEBS Journal*, 281(6), 1571–1584. <https://doi.org/10.1111/febs.12724>
- Conway, S. J., Mahadevaiah, S. K., Darling, S. M., Capel, B., Rattigan, A. M., & Burgoyne, P. S. (1994). Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome. *Mammalian Genome*, 5(4), 203–210. <https://doi.org/10.1007/BF00360546>
- Cooper, T. G. (2011). The epididymis, cytoplasmic droplets and male fertility. *Asian Journal of Andrology*, 13(1), 130–138. <https://doi.org/10.1038/aja.2010.97>
- Cornwall, G. A., & Horsten, H. H. von. (2007). Sperm Maturation in the Epididymis. In D. T. Carrell (Ed.), *The Genetics of Male Infertility* (pp. 211–231). Humana Press. Retrieved from http://link.springer.com/chapter/10.1007/978-1-59745-176-5_13
- Dacheux, J.-L., & Dacheux, F. (2014). New insights into epididymal function in relation to sperm maturation. *Reproduction*, 147(2), R27–R42. <https://doi.org/10.1530/REP-13-0420>
- Dai, L., Tsai-Morris, C.-H., Sato, H., Villar, J., Kang, J.-H., Zhang, J., & Dufau, M. L. (2011). Testis-specific miRNA-469 Up-regulated in Gonadotropin-regulated Testicular RNA Helicase (GRTH/DDX25)-null Mice Silences Transition Protein 2 and Protamine 2 Messages at Sites within Coding Region: Implications of its role in germ cell development. *Journal of Biological Chemistry*, 286(52), 44306–44318. <https://doi.org/10.1074/jbc.M111.282756>

- Darszon, A., Nishigaki, T., Beltran, C., & Treviño, C. L. (2011). Calcium Channels in the Development, Maturation, and Function of Spermatozoa. *Physiological Reviews*, 91(4), 1305–1355. <https://doi.org/10.1152/physrev.00028.2010>
- de Boer, P., de Vries, M., & Ramos, L. (2015). A mutation study of sperm head shape and motility in the mouse: lessons for the clinic. *Andrology*, 3(2), 174–202. <https://doi.org/10.1111/andr.300>
- de Yebra, L., Balleascà, J. L., Vanrell, J. A., Bassas, L., & Oliva, R. (1993). Complete selective absence of protamine P2 in humans. *The Journal of Biological Chemistry*, 268(14), 10553–10557.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.
- Deng, X., Moran, J., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Primakoff, P., & Martin-DeLeon, P. A. (1997). The Mouse Spam1 maps to proximal Chromosome 6 and is a candidate for the sperm dysfunction in Rb(6.16)24Lub and Rb(6.15)1Ald heterozygotes. *Mammalian Genome*, 8(2), 94–97. <https://doi.org/10.1007/s003359900365>
- Deng, X., Czymmek, K., & Martin-DeLeon, P. A. (1999). Biochemical maturation of Spam1 (PH-20) during epididymal transit of mouse sperm involves modifications of N-linked oligosaccharides. *Molecular Reproduction and Development*, 52(2), 196–206. [https://doi.org/10.1002/\(SICI\)1098-2795\(199902\)52:2<196::AID-MRD11>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1098-2795(199902)52:2<196::AID-MRD11>3.0.CO;2-A)
- Deng, W., & Lin, H. (2002). miwi, a Murine Homolog of piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis. *Developmental Cell*, 2(6), 819–830. [https://doi.org/10.1016/S1534-5807\(02\)00165-X](https://doi.org/10.1016/S1534-5807(02)00165-X)
- Dharap, A., Pokrzywa, C., Murali, S., Pandi, G., & Vemuganti, R. (2013). MicroRNA miR-324-3p Induces Promoter-Mediated Expression of RelA Gene. *PLOS ONE*, 8(11), e79467. <https://doi.org/10.1371/journal.pone.0079467>
- Dias, B. G., & Ressler, K. J. (2014). Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nature Neuroscience*, 17(1), 89–96. <https://doi.org/10.1038/nn.3594>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.
- Dolnik, A. v., Pochukalina, G. n., Parfenov, V. n., Karpushev, A. v., Podgornaya, O. i., & Voronin, A. p. (2007). Dynamics of satellite binding protein CENP-B and telomere binding protein TRF2/MTBP in the nuclei of mouse spermatogenic line. *Cell Biology International*, 31(4), 316–329. <https://doi.org/10.1016/j.cellbi.2007.01.017>
- Dunn, C. A., & Mager, D. L. (2005). Transcription of the human and rodent SPAM1 / PH-20 genes initiates within an ancient endogenous retrovirus. *BMC Genomics*, 6, 47. <https://doi.org/10.1186/1471-2164-6-47>

- Dweep, H., & Gretz, N. (2015). miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nature Methods*, 12(8), 697–697. <https://doi.org/10.1038/nmeth.3485>
- Eddy E. M (2002) Male germ cell gene expression. *Recent Progress in Hormone Research*, Vol. 2002;57:103–128.
- Edwards, A. W. (2000). Carl Düsing (1884) on the regulation of the sex-ratio. *Theoretical Population Biology*, 58(3), 255–257. <https://doi.org/10.1006/tpbi.2000.1482>
- Egel, R., & Penny, D. (2007). On the Origin of Meiosis in Eukaryotic Evolution: Coevolution of Meiosis and Mitosis from Feeble Beginnings. In *Recombination and Meiosis* (pp. 249–288). Springer, Berlin, Heidelberg. https://doi.org/10.1007/7050_2007_036
- Elliott, D. J., Ma, K., Kerr, S. M., Thakrar, R., Speed, R., Chandley, A. C., & Cooke, H. (1996). An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells. *Human Molecular Genetics*, 5(7), 869–874.
- Ellis, P. J. I., Clemente, E. J., Ball, P., Touré, A., Ferguson, L., Turner, J. M. A., ... Burgoyne, P. S. (2005). Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. *Human Molecular Genetics*, 14(18), 2705–2715. <https://doi.org/10.1093/hmg/ddi304>
- Ellis, P. J., Ferguson, L., Clemente, E. J., & Affara, N. A. (2007). Bidirectional transcription of a novel chimeric gene mapping to mouse chromosome Yq. *BMC Evolutionary Biology*, 7(1), 1–19. <https://doi.org/10.1186/1471-2148-7-171>
- Ellis, P. J. I., Bacon, J., & Affara, N. A. (2011a). Association of *Sly* with sex-linked gene amplification during mouse evolution: a side effect of genomic conflict in spermatids? *Human Molecular Genetics*, 20(15), 3010–3021. <https://doi.org/10.1093/hmg/ddr204>
- Ellis, P. J. I., Yu, Y., & Zhang, S. (2011b). Transcriptional dynamics of the sex chromosomes and the search for offspring sex-specific antigens in sperm. *Reproduction*, 142(5), 609–619. <https://doi.org/10.1530/REP-11-0228>
- Fabricant, J. D., & Parkening, T. A. (1982). Sperm morphology and cytogenetic studies in ageing C57BL/6 mice. *Journal of Reproduction and Fertility*, 66(2), 485–489.
- Fahrenkrug, S. C., Blake, A., Carlson, D. F., Doran, T., Eenennaam, A. V., Faber, D., ... Glenn, B. P. (2010). Precision genetics for complex objectives in animal agriculture. *Journal of Animal Science*, 88(7), 2530–2539. <https://doi.org/10.2527/jas.2010-2847>
- Finch, K. A., Fonseka, K. G. L., Abogrein, A., Ioannou, D., Handyside, A. H., Thornhill, A. R., ... Griffin, D. K. (2008). Nuclear organization in human sperm: preliminary evidence for altered sex chromosome centromere position in infertile males. *Human Reproduction* (Oxford, England), 23(6), 1263–1270. <https://doi.org/10.1093/humrep/den112>

- Firman, R. C., & Simmons, L. W. (2009). Sperm competition and the evolution of the sperm hook in house mice. *Journal of Evolutionary Biology*, 22(12), 2505–2511. <https://doi.org/10.1111/j.1420-9101.2009.01867.x>
- Firman, R. C., Bentley, B., Bowman, F., Marchant, F. G.-S., Parthenay, J., Sawyer, J., ... O'Shea, J. E. (2013). No evidence of sperm conjugate formation in an Australian mouse bearing sperm with three hooks. *Ecology and Evolution*, 3(7), 1856–1863. <https://doi.org/10.1002/ece3.577>
- Fischer, M., Kosyakova, N., Liehr, T., & Dobrowolski, P. (2017). Large deletion on the Y-chromosome long arm (Yq) of C57BL/6JBomTac inbred mice. *Mammalian Genome*, 28(1–2), 31–37. <https://doi.org/10.1007/s00335-016-9669-0>
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*
- Fisher, R. A. (1931). The Evolution of Dominance. *Biological Reviews*, 6(4), 345–368. <https://doi.org/10.1111/j.1469-185X.1931.tb01030.x>
- Fisher, H. S., & Hoekstra, H. E. (2010). Competition drives cooperation among closely related sperm of deer mice. *Nature*, 463(7282), 801–803. <https://doi.org/10.1038/nature08736>
- Fisher, H. S., Giomi, L., Hoekstra, H. E., & Mahadevan, L. (2014). The dynamics of sperm cooperation in a competitive environment. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1790), 20140296. <https://doi.org/10.1098/rspb.2014.0296>
- Foster, J. W., & Graves, J. A. (1994). An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. *Proceedings of the National Academy of Sciences of the United States of America*, 91(5), 1927–1931.
- Fraser, L. R., & Herod, J. E. (1990). Expression of capacitation-dependent changes in chlortetracycline fluorescence patterns in mouse spermatozoa requires a suitable glycolysable substrate. *Journal of Reproduction and Fertility*, 88(2), 611–621.
- Fraser, L. R., Hosseini, R., Hanyalogou, A., Talmor, A., & Dudley, R. K. (1997). TCP-11, the product of a mouse t-complex gene, plays a role in stimulation of capacitation and inhibition of the spontaneous acrosome reaction. *Molecular Reproduction and Development*, 48(3), 375–382. [https://doi.org/10.1002/\(SICI\)1098-2795\(199711\)48:3<375::AID-MRD11>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1098-2795(199711)48:3<375::AID-MRD11>3.0.CO;2-V)
- Fraser, L. R., & Dudley, K. (1999). New insights into the t-complex and control of sperm function. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 21(4), 304–312. [https://doi.org/10.1002/\(SICI\)1521-1878\(199904\)21:4<304::AID-BIES6>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1521-1878(199904)21:4<304::AID-BIES6>3.0.CO;2-H)
- Fujihara, Y., Satouh, Y., Inoue, N., Isotani, A., Ikawa, M., & Okabe, M. (2012). SPACA1-deficient male mice are infertile with abnormally shaped sperm heads reminiscent of globozoospermia. *Development (Cambridge, England)*, 139(19), 3583–3589. <https://doi.org/10.1242/dev.081778>

- Fujihara, Y., Okabe, M., & Ikawa, M. (2014). GPI-anchored protein complex, LY6K/TEX101, is required for sperm migration into the oviduct and male fertility in mice. *Biology of Reproduction*, 90(3), 60. <https://doi.org/10.1095/biolreprod.113.112888>
- Fullston, T., Ohlsson-Teague, E. M. C., Print, C. G., Sandeman, L. Y., & Lane, M. (2016). Sperm microRNA Content Is Altered in a Mouse Model of Male Obesity, but the Same Suite of microRNAs Are Not Altered in Offspring's Sperm. *PLOS ONE*, 11(11), e0166076. <https://doi.org/10.1371/journal.pone.0166076>
- Gao, X., Qiao, Y., Han, D., Zhang, Y., & Ma, N. (2012). Enemy or partner: Relationship between intronic micrnas and their host genes. *IUBMB Life*, 64(10), 835–840. <https://doi.org/10.1002/iub.1079>
- Gapp, K., Jawaid, A., Sarkies, P., Bohacek, J., Pelczar, P., Prados, J., ... Mansuy, I. M. (2014). Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nature Neuroscience*, 17(5), 667–669. <https://doi.org/10.1038/nn.3695>
- Garner, D. L. (2009). Hoechst 33342: The dye that enabled differentiation of living X-and Y-chromosome bearing mammalian sperm. *Theriogenology*, 71(1), 11–21. <https://doi.org/10.1016/j.theriogenology.2008.09.023>
- Gibson, A. K., Delph, L. F., & Lively, C. M. (2017). The two-fold cost of sex: Experimental evidence from a natural system. *Evolution Letters*, 1(1), 6–15. <https://doi.org/10.1002/evl3.1>
- Goodenough, U., & Heitman, J. (2014). Origins of Eukaryotic Sexual Reproduction. *Cold Spring Harbor Perspectives in Biology*, 6(3), a016154. <https://doi.org/10.1101/cshperspect.a016154>
- Grabske, R. J., Lake, S., Gledhill, B. L., & Meistrich, M. L. (1975). Centrifugal elutriation: Separation of spermatogenic cells on the basis of sedimentation velocity. *Journal of Cellular Physiology*, 86(1), 177–189. <https://doi.org/10.1002/jcp.1040860119>
- Grandjean, V., Fourré, S., De Abreu, D. A. F., Derieppe, M.-A., Remy, J.-J., & Rassoulzadegan, M. (2015). RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Scientific Reports*, 5, 18193. <https://doi.org/10.1038/srep18193>
- Graves, J. A. M. (2013). How to evolve new vertebrate sex determining genes. *Developmental Dynamics*, 242(4), 354–359. <https://doi.org/10.1002/dvdy.23887>
- Graves, J. A. M. (2016). Did sex chromosome turnover promote divergence of the major mammal groups? *Bioessays*, 38(8), 734–743. <https://doi.org/10.1002/bies.201600019>
- Greenbaum, M. P., Iwamori, T., Buchold, G. M., & Matzuk, M. M. (2011). Germ cell intercellular bridges. *Cold Spring Harbor Perspectives in Biology*, 3(8), a005850. <https://doi.org/10.1101/cshperspect.a005850>
- Griffiths, G. S., Miller, K. A., Galileo, D. S., & Martin-DeLeon, P. A. (2008). Murine SPAM1 is secreted by the estrous uterus and oviduct in a form that can bind to sperm during capacitation: acquisition enhances hyaluronic acid-binding ability and cumulus dispersal efficiency. *Reproduction*, 135(3), 293–301. <https://doi.org/10.1530/REP-07-0340>

- Griffiths-Jones, S. (2004). The microRNA Registry. *Nucleic Acids Research*, 32(Database issue), D109-111. <https://doi.org/10.1093/nar/gkh023>
- Grossfeld, R., Klinc, P., Sieg, B., & Rath, D. (2005). Production of piglets with sexed semen employing a non-surgical insemination technique. *Theriogenology*, 63(8), 2269–2277. <https://doi.org/10.1016/j.theriogenology.2004.10.004>
- Grunewald, S., Paasch, U., Glander, H.-J., & Anderegg, U. (2005). Mature human spermatozoa do not transcribe novel RNA. *Andrologia*, 37(2–3), 69–71. <https://doi.org/10.1111/j.1439-0272.2005.00656.x>
- Grzmił, P., Gołas, A., Müller, C., & Styrna, J. (2007). The influence of the deletion on the long arm of the Y chromosome on sperm motility in mice. *Theriogenology*, 67(4), 760–766. <https://doi.org/10.1016/j.theriogenology.2006.10.007>
- Gunnarsson, B., Goodacre, S. L., & Hewitt, G. M. (2009). Sex ratio, mating behaviour and Wolbachia infections in a sheetweb spider. *Biological Journal of the Linnean Society*, 98(1), 181–186. <https://doi.org/10.1111/j.1095-8312.2009.01247.x>
- Guzick, D. S., Overstreet, J. W., Factor-Litvak, P., Brazil, C. K., Nakajima, S. T., Coutifaris, C., ... Vogel, D. L. (2001). Sperm Morphology, Motility, and Concentration in Fertile and Infertile Men. *New England Journal of Medicine*, 345(19), 1388–1393. <https://doi.org/10.1056/NEJMoa003005>
- Hackstadt, A. J., & Hess, A. M. (2009). Filtering for increased power for microarray data analysis. *BMC Bioinformatics*, 10, 11. <https://doi.org/10.1186/1471-2105-10-11>
- Halary, S., Malik, S.-B., Lildhar, L., Slamovits, C. H., Hijri, M., & Corradi, N. (2011). Conserved meiotic machinery in *Glomus* spp., a putatively ancient asexual fungal lineage. *Genome Biology and Evolution*, 3, 950–958. <https://doi.org/10.1093/gbe/evr089>
- Hall, A., & Lalli, G. (2010). Rho and Ras GTPases in Axon Growth, Guidance, and Branching. *Cold Spring Harbor Perspectives in Biology*, 2(2), a001818. <https://doi.org/10.1101/cshperspect.a001818>
- Hamilton, W. (1967). Extraordinary Sex Ratios. *Science*, 156(3774), 477-488. Retrieved from <http://www.jstor.org/stable/1721222>
- Hartsoeker N (1694). *Essay de Dioptrique. :: Color and Optics*. Retrieved 3 July 2017, from The Linda Hall Library Digital Collections: <http://lhldigital.lindahall.org/cdm/ref/collection/color/id/32952>
- Hayashi, K., Lopes, S. M. C. de S., Kaneda, M., Tang, F., Hajkova, P., Lao, K., ... Surani, M. A. (2008). MicroRNA Biogenesis Is Required for Mouse Primordial Germ Cell Development and Spermatogenesis. *PLOS ONE*, 3(3), e1738. <https://doi.org/10.1371/journal.pone.0001738>
- Heid, H. W., Figge, U., Winter, S., Kuhn, C., Zimbelmann, R., & Franke, W. W. (2002). Novel Actin-Related Proteins Arp-T1 and Arp-T2 as Components of the Cytoskeletal Calyx of the Mammalian Sperm Head. *Experimental Cell Research*, 279(2), 177–187. <https://doi.org/10.1006/excr.2002.5603>

- Heller, M. J. (2002). DNA Microarray Technology: Devices, Systems, and Applications. *Annual Review of Biomedical Engineering*, 4(1), 129–153. <https://doi.org/10.1146/annurev.bioeng.4.020702.153438>
- Henrichsen, C. N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., ... Reymond, A. (2009). Segmental copy number variation shapes tissue transcriptomes. *Nature Genetics*, 41(4), 424–429. <https://doi.org/10.1038/ng.345>
- Henry, L., Schwander, T., & Crespi, B. J. (2012). Deleterious Mutation Accumulation in Asexual Timema Stick Insects. *Molecular Biology and Evolution*, 29(1), 401–408. <https://doi.org/10.1093/molbev/msr237>
- Hesketh, J. E., Campbell, G. P., & Whitelaw, P. F. (1991). c-myc mRNA in cytoskeletal-bound polysomes in fibroblasts. *Biochemical Journal*, 274(Pt 2), 607–609.
- Hess, H., Heid, H., & Franke, W. W. (1993). Molecular characterization of mammalian cyclin, a basic protein of the sperm head cytoskeleton. *The Journal of Cell Biology*, 122(5), 1043–1052.
- Hoheisel, J. D. (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nature Reviews Genetics*, 7(3), 200–210. <https://doi.org/10.1038/nrg1809>
- Hollister, J. D., Greiner, S., Wang, W., Wang, J., Zhang, Y., Wong, G. K.-S., ... Johnson, M. T. J. (2015). Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. *Molecular Biology and Evolution*, 32(4), 896–905. <https://doi.org/10.1093/molbev/msu345>
- Hosken, D. J., & Hodgson, D. J. (2014). Why do sperm carry RNA? Relatedness, conflict, and control. *Trends in Ecology & Evolution*, 29(8), 451–455. <https://doi.org/10.1016/j.tree.2014.05.006>
- Hsu, Y.-C., & Perin, M. S. (1995). Human Neuronal Pentraxin II (NPTX2): Conservation, Genomic Structure, and Chromosomal Localization. *Genomics*, 28(2), 220–227. <https://doi.org/10.1006/geno.1995.1134>
- Hsu, C.-W., Juan, H.-F., & Huang, H.-C. (2008). Characterization of microRNA-regulated protein-protein interaction network. *PROTEOMICS*, 8(10), 1975–1979. <https://doi.org/10.1002/pmic.200701004>
- Huang, L. O., Labbe, A., & Infante-Rivard, C. (2013). Transmission ratio distortion: review of concept and implications for genetic association studies. *Human Genetics*, 132(3), 245–263. <https://doi.org/10.1007/s00439-012-1257-0>
- Huszar, G., & Vigue, L. (1993). Incomplete development of human spermatozoa is associated with increased creatine phosphokinase concentration and abnormal head morphology. *Molecular Reproduction and Development*, 34(3), 292–298. <https://doi.org/10.1002/mrd.1080340309>
- Huszar, G., Patrizio, P., Vigue, L., Willets, M., Wdjker, C., Adhoot, D., & Johnson, L. (1998). Cytoplasmic Extrusion and the Switch From Creatine Kinase B to M Isoform are Completed by the Commencement of Epididymal Transport in Human and Stallion Spermatozoa. *Journal of Andrology*, 19(1), 11–20. <https://doi.org/10.1002/j.1939-4640.1998.tb02465.x>

Huszar, J. M., & Payne, C. J. (2013). MicroRNA 146 (Mir146) Modulates Spermatogonial Differentiation by Retinoic Acid in Mice. *Biology of Reproduction*, 88(1). <https://doi.org/10.1095/biolreprod.112.103747>

Immler, S., Moore, H. D. M., Breed, W. G., & Birkhead, T. R. (2007). By Hook or by Crook? Morphometry, Competition and Cooperation in Rodent Sperm. *PLoS ONE*, 2(1), e170. <https://doi.org/10.1371/journal.pone.0000170>

intersect — bedtools 2.26.0 documentation. Retrieved 13 July 2017, from <http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html>

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., & Speed, T. P. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research*, 31(4), e15.

Jackson Laboratory. (1966). *Biology of the laboratory mouse*. (E. L. Green, Ed.). Blakiston Division, McGraw-Hill

Jaquiéry, J., Rispe, C., Roze, D., Legeai, F., Trionnaire, G. L., Stoeckel, S., ... Simon, J.-C. (2013). Masculinization of the X Chromosome in the Pea Aphid. *PLOS Genetics*, 9(8), e1003690. <https://doi.org/10.1371/journal.pgen.1003690>

Jho, E., Zhang, T., Domon, C., Joo, C.-K., Freund, J.-N., & Costantini, F. (2002). Wnt/ β -Catenin/Tcf Signaling Induces the Transcription of Axin2, a Negative Regulator of the Signaling Pathway. *Molecular and Cellular Biology*, 22(4), 1172–1183. <https://doi.org/10.1128/MCB.22.4.1172-1183.2002>

Jin, M., Fujiwara, E., Kakiuchi, Y., Okabe, M., Satouh, Y., Baba, S. A., ... Hirohashi, N. (2011). Most fertilizing mouse spermatozoa begin their acrosome reaction before contact with the zona pellucida during *in vitro* fertilization. *Proceedings of the National Academy of Sciences*, 108(12), 4892–4896. <https://doi.org/10.1073/pnas.1018202108>

Jodar, M., Selvaraju, S., Sendler, E., Diamond, M. P., Krawetz, S. A., & Network, for the R. M. (2013). The presence, role and clinical use of spermatozoal RNAs. *Human Reproduction Update*, 19(6), 604–624. <https://doi.org/10.1093/humupd/dmt031>

Johnson, G. D., Lalancette, C., Linnemann, A. K., Leduc, F., Boissonneault, G., & Krawetz, S. A. (2011). The sperm nucleus: chromatin, RNA, and the nuclear matrix. *Reproduction*, 141(1), 21–36. <https://doi.org/10.1530/REP-10-0322>

Johnson, G. D., Mackie, P., Jodar, M., Moskovtsev, S., & Krawetz, S. A. (2015). Chromatin and extracellular vesicle associated sperm RNAs. *Nucleic Acids Research*, 43(14), 6847–6859. <https://doi.org/10.1093/nar/gkv591>

Johnstone, O., & Lasko, P. (2001). Translational Regulation and RNA Localization in *Drosophila* Oocytes and Embryos. *Annual Review of Genetics*, 35(1), 365–406. <https://doi.org/10.1146/annurev.genet.35.102401.090756>

Jonge, C. J. D., & Barratt, C. (2006). *The Sperm Cell: Production, Maturation, Fertilization, Regeneration*. Cambridge University Press.

Kabayama, Y., Toh, H., Katanaya, A., Sakurai, T., Chuma, S., Kuramochi-Miyagawa, S., ... Sasaki, H. (2017). Roles of MIWI, MILI and PLD6 in small RNA regulation in mouse growing oocytes. *Nucleic Acids Research*, 45(9), 5387–5398. <https://doi.org/10.1093/nar/gkx027>

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27–30.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457–D462. <https://doi.org/10.1093/nar/gkv1070>

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(Database issue), D353–D361. <https://doi.org/10.1093/nar/gkw1092>

Kang, W., Zhou, C., Koga, Y., & Baba, T. (2010). Hyaluronan-Degrading Activity of Mouse Sperm Hyaluronidase Is Not Required for Fertilization? *Journal of Reproduction and Development*, 56(1), 140–144.

Khazaie, Y., & Nasr Esfahani, M. H. (2014). MicroRNA and Male Infertility: A Potential for Diagnosis. *International Journal of Fertility & Sterility*, 8(2), 113–118.

Khil, P. P., & Camerini-Otero, R. D. (2005). Molecular features and functional constraints in the evolution of the mammalian X chromosome. *Critical Reviews in Biochemistry and Molecular Biology*, 40(6), 313–330. <https://doi.org/10.1080/10409230500356703>

Kierszenbaum, A. L. (2002). Sperm axoneme: a tale of tubulin posttranslation diversity. *Molecular Reproduction and Development*, 62(1), 1–3. <https://doi.org/10.1002/mrd.10139>

Kim, E., Baba, D., Kimura, M., Yamashita, M., Kashiwabara, S., & Baba, T. (2005). Identification of a hyaluronidase, Hyal5, involved in penetration of mouse sperm through cumulus mass. *Proceedings of the National Academy of Sciences of the United States of America*, 102(50), 18028–18033. <https://doi.org/10.1073/pnas.0506825102>

Kim, E., Lee, Y., Kim, J.-S., Song, B.-S., Kim, S.-U., Huh, J.-W., ... Chang, K.-T. (2010). Extracellular domain of V-set and immunoglobulin domain containing 1 (VSIG1) interacts with sertoli cell membrane protein, while its PDZ-binding motif forms a complex with ZO-1. *Molecules and Cells*, 30(5), 443–448. <https://doi.org/10.1007/s10059-010-0138-4>

Kim, K.-S., Foster, J. A., Kvasnicka, K. W., & Gerton, G. L. (2011). Transitional States of Acrosomal Exocytosis and Proteolytic Processing of the Acrosomal Matrix in Guinea Pig Sperm. *Molecular Reproduction and Development*, 78(12), 930–941. <https://doi.org/10.1002/mrd.21387>

- Kino, Y., Washizu, C., Kurosawa, M., Yamada, M., Doi, H., Takumi, T., ... Nukina, N. (2016). FUS/TLS acts as an aggregation-dependent modifier of polyglutamine disease model mice. *Scientific Reports*, 6, srep35236. <https://doi.org/10.1038/srep35236>
- Koch, S., Acebron, S. P., Herbst, J., Hatiboglu, G., & Niehrs, C. (2015). Post-transcriptional Wnt Signaling Governs Epididymal Sperm Maturation. *Cell*, 163(5), 1225–1236. <https://doi.org/10.1016/j.cell.2015.10.029>
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., & Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry. *Nature*, 351(6322), 117–121. <https://doi.org/10.1038/351117a0>
- Korhonen, H. M., Meikar, O., Yadav, R. P., Papaioannou, M. D., Romero, Y., Da Ros, M., ... Kotaja, N. (2011). Dicer is required for haploid male germ cell differentiation in mice. *PloS One*, 6(9), e24821. <https://doi.org/10.1371/journal.pone.0024821>
- Kotarska, K., & Lenartowicz, M. (2011). Sperm migration and selection in the reproductive tract of female mice is mostly affected by male genotype. *Folia Biologica*, 59(1–2), 71–75.
- Kotarska, K., Lenartowicz, M., Przybyło, M., Gołas, A., & Styrna, J. (2013). Increased prostaglandin E2–EP2 signalling in cumulus cells of female mice sired by males with the Y-chromosome long-arm deletion. *Reproduction, Fertility and Development*. Retrieved from <http://agris.fao.org/agris-search/search.do?recordID=US201400150303>
- Kotarska, K., Galas, J., Przybyło, M., Bilińska, B., & Styrna, J. (2014). Increased Progesterone Production in Cumulus-Oocyte Complexes of Female Mice Sired by Males With the Y-Chromosome Long Arm Deletion and its Potential Influence on Fertilization Efficiency. *Reproductive Sciences*. <https://doi.org/10.1177/1933719114537717>
- Kotlyarov, A., Neining, A., Schubert, C., Eckert, R., Birchmeier, C., Volk, H. D., & Gaestel, M. (1999). MAPKAP kinase 2 is essential for LPS-induced TNF-alpha biosynthesis. *Nature Cell Biology*, 1(2), 94–97. <https://doi.org/10.1038/10061>
- Kozomara, A., & Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 42(D1), D68–D73. <https://doi.org/10.1093/nar/gkt1181>
- Krawetz, S. A. (2005). Paternal contribution: new insights and future challenges. *Nature Reviews. Genetics*, 6(8), 633–642. <https://doi.org/10.1038/nrg1654>
- Krueger, F. (2013). Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisulfite-Seq) libraries. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- Krzanowska, H. (1969). Factor responsible for spermatozoan abnormality located on the Y chromosome in mice. *Genetical Research*, 13(1), 17–24.

- Kuhn, D. E., Martin, M. M., Feldman, D. S., Terry, A. V., Nuovo, G. J., & Elton, T. S. (2008). Experimental Validation of miRNA Targets. *Methods* (San Diego, Calif.), 44(1), 47–54. <https://doi.org/10.1016/j.ymeth.2007.09.005>
- Łabaj, P. P., & Kreil, D. P. (2016). Sensitivity, specificity, and reproducibility of RNA-Seq differential expression calls. *Biology Direct*, 11, 66. <https://doi.org/10.1186/s13062-016-0169-7>
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., & Tuschl, T. (2001). Identification of Novel Genes Coding for Small Expressed RNAs. *Science*, 294(5543), 853–858. <https://doi.org/10.1126/science.1064921>
- Lai, S.-L., Chien, A. J., & Moon, R. T. (2009). Wnt/Fz signaling and the cytoskeleton: potential roles in tumorigenesis. *Cell Research*, 19(5), 532–545. <https://doi.org/10.1038/cr.2009.41>
- Lambard, S., Galeraud-Denis, I., Bouraïma, H., Bourguiba, S., Chocat, A., & Carreau, S. (2003). Expression of aromatase in human ejaculated spermatozoa: a putative marker of motility. *MHR: Basic Science of Reproductive Medicine*, 9(3), 117–124. <https://doi.org/10.1093/molehr/gag020>
- Lambard, S., Galeraud-Denis, I., Martin, G., Levy, R., Chocat, A., & Carreau, S. (2004). Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: relationship to sperm motility and capacitation. *MHR: Basic Science of Reproductive Medicine*, 10(7), 535–541. <https://doi.org/10.1093/molehr/gah064>
- Larson, E. L., Vanderpool, D., Keeble, S., Zhou, M., Sarver, B. A. J., Smith, A. D., ... Good, J. M. (2016). Contrasting Levels of Molecular Evolution on the Mouse X Chromosome. *Genetics*, 203(4), 1841–1857. <https://doi.org/10.1534/genetics.116.186825>
- Lassalle, B., Ziyat, A., Testart, J., Finaz, C., & Lefèvre, A. (1999). Flow cytometric method to isolate round spermatids from mouse testis. *Human Reproduction*, 14(2), 388–394. <https://doi.org/10.1093/humrep/14.2.388>
- Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75(5), 843–854.
- Lee, S. C., Ni, M., Li, W., Shertz, C., & Heitman, J. (2010). The Evolution of Sex: a Perspective from the Fungal Kingdom. *Microbiology and Molecular Biology Reviews*, 74(2), 298–340. <https://doi.org/10.1128/MMBR.00005-10>
- Lehtonen, J., Jennions, M. D., & Kokko, H. (2012). The many costs of sex. *Trends in Ecology & Evolution*, 27(3), 172–178. <https://doi.org/10.1016/j.tree.2011.09.016>
- Lehtonen, J., & Parker, G. A. (2014). Gamete competition, gamete limitation, and the evolution of the two sexes. *Molecular Human Reproduction*, 20(12), 1161–1168. <https://doi.org/10.1093/molehr/gau068>
- Lercher, M. J., Urrutia, A. O., & Hurst, L. D. (2003). Evidence that the human X chromosome is enriched for male-specific but not female-specific genes. *Molecular Biology and Evolution*, 20(7), 1113–1116. <https://doi.org/10.1093/molbev/msg131>

- Li, M. W. M., Mruk, D. D., & Cheng, C. Y. (2009). Mitogen-activated protein kinases in male reproductive function. *Trends in Molecular Medicine*, 15(4), 159–168. <https://doi.org/10.1016/j.molmed.2009.02.002>
- Lin, A. C., & Holt, C. E. (2007). Local translation and directional steering in axons. *The EMBO Journal*, 26(16), 3729–3736. <https://doi.org/10.1038/sj.emboj.7601808>
- Lin, Y., Mahan, K., Lathrop, W. F., Myles, D. G., & Primakoff, P. (1994). A hyaluronidase activity of the sperm plasma membrane protein PH-20 enables sperm to penetrate the cumulus cell layer surrounding the egg. *The Journal of Cell Biology*, 125(5), 1157–1163. <https://doi.org/10.1083/jcb.125.5.1157>
- Lin, Y.-W., Hsu, T.-H., & Yen, P. H. (2013). Mouse sperm acquire a new structure on the apical hook during epididymal maturation. *Asian Journal of Andrology*, 15(4), 523–528. <https://doi.org/10.1038/aja.2013.46>
- Ling, H., Fabbri, M., & Calin, G. A. (2013). MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nature Reviews Drug Discovery*, 12(11), 847–865. <https://doi.org/10.1038/nrd4140>
- Liu, W.-M., Pang, R. T. K., Chiu, P. C. N., Wong, B. P. C., Lao, K., Lee, K.-F., & Yeung, W. S. B. (2012). Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proceedings of the National Academy of Sciences of the United States of America*, 109(2), 490–494. <https://doi.org/10.1073/pnas.1110368109>
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). Processing of rRNA and tRNA. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK21729/>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLOS Computational Biology*, 13(5), e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>
- Lyon, M. F. (2003). Transmission Ratio Distortion in Mice. *Annual Review of Genetics*, 37(1), 393–408. <https://doi.org/10.1146/annurev.genet.37.110801.143030>
- Lyon, M. F. (1986). Male sterility of the mouse t-complex is due to homozygosity of the distorter genes. *Cell*, 44(2), 357–363. [https://doi.org/10.1016/0092-8674\(86\)90770-1](https://doi.org/10.1016/0092-8674(86)90770-1)
- Lyon, M. F. (1984). Transmission ratio distortion in mouse t-haplotypes is due to multiple distorter genes acting on a responder locus. *Cell*, 37(2), 621–628.
- MacBride, M. M., Navis, A., Dasari, A., & Perez, A. V. (2017). Mild reproductive impact of a Y chromosome deletion on a C57BL/6J substrain. *Mammalian Genome*, 28(5–6), 155–165. <https://doi.org/10.1007/s00335-017-9680-0>
- Mahadevaiah, S. K., Odorisio, T., Elliott, D. J., Rattigan, Á., Szot, M., Laval, S. H., ... Burgoyne, P. S. (1998). Mouse Homologues of the Human AZF Candidate Gene RBM Are Expressed in Spermatogonia and

Spermatids, and Map to a Y Chromosome Deletion Interval Associated with a High Incidence of Sperm Abnormalities. *Human Molecular Genetics*, 7(4), 715–727. <https://doi.org/10.1093/hmg/7.4.715>

Malik, S.-B., Pightling, A. W., Stefaniak, L. M., Schurko, A. M., & Logsdon, J. M. (2008). An Expanded Inventory of Conserved Meiotic Genes Provides Evidence for Sex in *Trichomonas vaginalis*. *PLoS ONE*, 3(8). <https://doi.org/10.1371/journal.pone.0002879>

Manser, A., Lindholm, A. K., König, B., & Bagheri, H. C. (2011). Polyandry and the Decrease of a Selfish Genetic Element in a Wild House Mouse Population. *Evolution*, 65(9), 2435–2447. <https://doi.org/10.1111/j.1558-5646.2011.01336.x>

Mantione, K. J., Kream, R. M., Kuzelova, H., Ptacek, R., Raboch, J., Samuel, J. M., & Stefano, G. B. (2014). Comparing Bioinformatic Gene Expression Profiling Methods: Microarray and RNA-Seq. *Medical Science Monitor Basic Research*, 20, 138–141. <https://doi.org/10.12659/MSMBR.892101>

Martin, K. C., & Ephrussi, A. (2009). mRNA Localization: Gene Expression in the Spatial Dimension. *Cell*, 136(4), 719–730. <https://doi.org/10.1016/j.cell.2009.01.044>

Martin-DeLeon, P. A., Zhang, H., Morales, C. R., Zhao, Y., Rulon, M., Barnoski, B. L., ... Galileo, D. S. (2005). Spam1-associated transmission ratio distortion in mice: Elucidating the mechanism. *Reproductive Biology and Endocrinology*, 3(1), 32. <https://doi.org/10.1186/1477-7827-3-32>

Martin-DeLeon, P. A. (2006). Epididymal SPAM1 and its impact on sperm function. *Molecular and Cellular Endocrinology*, 250(1–2), 114–121. <https://doi.org/10.1016/j.mce.2005.12.033>

May-Simera, H. L., & Kelley, M. W. (2012). Cilia, Wnt signaling, and the cytoskeleton. *Cilia*, 1, 7. <https://doi.org/10.1186/2046-2530-1-7>

Maynard Smith, J., 1971 The origin and maintenance of sex, pp. 163–175 in *Group Selection*, edited by G. C. Williams. Aldine-Atherton, Chicago

Maynard-Smith, J. (1978). *The Evolution of Sex*. CUP Archive

Mazaud-Guittot, S., Meugnier, E., Pesenti, S., Wu, X., Vidal, H., Gow, A., & Le Magueresse-Battistoni, B. (2010). Claudin 11 deficiency in mice results in loss of the Sertoli cell epithelial phenotype in the testis. *Biology of Reproduction*, 82(1), 202–213. <https://doi.org/10.1095/biolreprod.109.078907>

Mazeyrat, S., Saut, N., Grigoriev, V., Mahadevaiah, S. K., Ojarikre, O. A., Rattigan A, null, ... Burgoyne, P. S. (2001). A Y-encoded subunit of the translation initiation factor Eif2 is essential for mouse spermatogenesis. *Nature Genetics*, 29(1), 49–53. <https://doi.org/10.1038/ng717>

McCull, K., Clarke, B., & Doran, T. (2013). Role of genetically engineered animals in future food production. *Australian Veterinary Journal*, 91(3), 113–117. <https://doi.org/10.1111/avj.12024>

- Meistrich, M.L. (1989). Histone and Basic Nuclear Protein Transitions in Mammalian Spermatogenesis. In "Histones and Other Basic Nuclear Proteins" (L.S.Hnicilia, G.S.Stein, and J.L.Stein, eds), pp. 165-182. CRC Pres, Orlando, Florida
- Menkveld, R. (2010). Clinical significance of the low normal sperm morphology value as proposed in the fifth edition of the WHO Laboratory Manual for the Examination and Processing of Human Semen. *Asian Journal of Andrology*, 12(1), 47–58. <https://doi.org/10.1038/aja.2009.14>
- Menkveld, R., Holleboom, C. A., & Rhemrev, J. P. (2011). Measurement and significance of sperm morphology. *Asian Journal of Andrology*, 13(1), 59–68. <https://doi.org/10.1038/aja.2010.67>
- Merçot, H., Llorente, B., Jacques, M., Atlan, A., & Montchamp-Moreau, C. (1995). Variability within the Seychelles cytoplasmic incompatibility system in *Drosophila simulans*. *Genetics*, 141(3), 1015–1023.
- Merton, H. (1940). XIV.—Studies on Reproduction in the Albino Mouse. II. Contributions on the Maturation of the Sperm Cells. *Proceedings of the Royal Society of Edinburgh*, 59, 145–152. <https://doi.org/10.1017/S0370164600012232>
- Mills, J. D., Kawahara, Y., & Janitz, M. (2013). Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling. *Current Genomics*, 14(3), 173–181. <https://doi.org/10.2174/1389202911314030003>
- Mochida, K., Tres, L. L., & Kierszenbaum, A. L. (1999). Structural and biochemical features of fractionated spermatid manchettes and sperm axonemes of the *Azh/Azh* mutant mouse. *Molecular Reproduction and Development*, 52(4), 434–444. [https://doi.org/10.1002/\(SICI\)1098-2795\(199904\)52:4<434::AID-MRD13>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1098-2795(199904)52:4<434::AID-MRD13>3.0.CO;2-D)
- Molnár, A., Schwach, F., Studholme, D. J., Thuenemann, E. C., & Baulcombe, D. C. (2007). miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature*, 447(7148), 1126–1129. <https://doi.org/10.1038/nature05903>
- Montchamp-Moreau, C., Ginhoux, V., & Atlan, A. (2001). The y chromosomes of *Drosophila simulans* are highly polymorphic for their ability to suppress sex-ratio drive. *Evolution*, 55(4), 728–737. [https://doi.org/10.1554/0014-3820\(2001\)055\[0728:TYCODS\]2.0.CO;2](https://doi.org/10.1554/0014-3820(2001)055[0728:TYCODS]2.0.CO;2)
- Montellier, E., Boussouar, F., Rousseaux, S., Zhang, K., Buchou, T., Fenaille, F., ... Khochbin, S. (2013). Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes & Development*, 27(15), 1680–1692. <https://doi.org/10.1101/gad.220095.113>
- Moore, H., Dvoráková, K., Jenkins, N., & Breed, W. (2002). Exceptional sperm cooperation in the wood mouse. *Nature*, 418(6894), 174–177. <https://doi.org/10.1038/nature00832>
- Morales, C. R., Lefrancois, S., Chennathukuzhi, V., El-Alfy, M., Wu, X., Yang, J., ... Hecht, N. B. (2002). A TB-RBP and Ter ATPase Complex Accompanies Specific mRNAs from Nuclei through the Nuclear Pores and into Intercellular Bridges in Mouse Male Germ Cells. *Developmental Biology*, 246(2), 480–494. <https://doi.org/10.1006/dbio.2002.0679>

- Morales, C. R., Badran, H., El-Alfy, M., Men, H., Zhang, H., & Martin-Deleon, P. A. (2004). Cytoplasmic localization during testicular biogenesis of the murine mRNA for Spam1 (PH-20), a protein involved in acrosomal exocytosis. *Molecular Reproduction and Development*, 69(4), 475–482. <https://doi.org/10.1002/mrd.20177>
- Moreau, J., & Rigaud, T. (2000). Operational sex ratio in terrestrial isopods: interaction between potential rate of reproduction and Wolbachia-induced sex ratio distortion. *Oikos*, 91(3), 477–484. <https://doi.org/10.1034/j.1600-0706.2000.910308.x>
- Moretti, C., Vaiman, D., Tores, F., & Cocquet, J. (2016). Expression and epigenomic landscape of the sex chromosomes in mouse post-meiotic male germ cells. *Epigenetics & Chromatin*, 9, 47. <https://doi.org/10.1186/s13072-016-0099-8>
- Moretti, C., Serrentino, M.-E., Ialy-Radio, C., Delessard, M., Soboleva, T. A., Tores, F., ... Cocquet, J. (2017). SLY regulates genes involved in chromatin remodeling and interacts with TBL1XR1 during sperm differentiation. *Cell Death & Differentiation*, 24(6), 1029–1044. <https://doi.org/10.1038/cdd.2017.32>
- Morgan, A. P., & Villena, F. P. M. de. (2017). Sequence and structural diversity of mouse Y chromosomes. *bioRxiv*, 096297. <https://doi.org/10.1101/096297>
- Morijiri, T., Yamada, M., Hikida, T., & Seki, M. (2013). Microfluidic counterflow centrifugal elutriation system for sedimentation-based cell separation. *Microfluidics and Nanofluidics*, 14(6), 1049–1057. <https://doi.org/10.1007/s10404-012-1113-5>
- Moriwaki K, Suh D, Styrna J. Genetic factors affecting sperm morphology in the mouse. *Mouse Newsl.* 1988;82:138
- Morsy, M. A., Norman, P. J., Mitry, R., Rela, M., Heaton, N. D., & Vaughan, R. W. (2005). Isolation, purification and flow cytometric analysis of human intrahepatic lymphocytes using an improved technique. *Laboratory Investigation*, 85(2), 285–296. <https://doi.org/10.1038/labinvest.3700219>
- Mueller, J. L., Mahadevaiah, S. K., Park, P. J., Warburton, P. E., Page, D. C., & Turner, J. M. A. (2008). The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature Genetics*, 40(6), 794–799. <https://doi.org/10.1038/ng.126>
- Muller, H. J. (1932). Some Genetic Aspects of Sex. *The American Naturalist*, 66(703), 118–138. <https://doi.org/10.1086/280418>
- Myles, D. G., & Primakoff, P. (1997). Why did the sperm cross the cumulus? To get to the oocyte. Functions of the sperm surface proteins PH-20 and fertilin in arriving at, and fusing with, the egg. *Biology of Reproduction*, 56(2), 320–327. <https://doi.org/10.1095/biolreprod56.2.320>
- Nakata, H., Wakayama, T., Takai, Y., & Iseki, S. (2015). Quantitative Analysis of the Cellular Composition in Seminiferous Tubules in Normal and Genetically Modified Infertile Mice. *Journal of Histochemistry and Cytochemistry*, 63(2), 99–113. <https://doi.org/10.1369/0022155414562045>

- Nelson, N. J. (2001). Microarrays Have Arrived: Gene Expression Tool Matures. *JNCI: Journal of the National Cancer Institute*, 93(7), 492–494. <https://doi.org/10.1093/jnci/93.7.492>
- Nipper, R. W., Chennothukuzhi, V., Tutuncu, L., Williams, C. J., Gerton, G. L. and Moss, S. B. (2005), Differential RNA expression and polyribosome loading of alternative transcripts of the Akap4 gene in murine spermatids. *Mol. Reprod. Dev.*, 70: 397–405. doi:10.1002/mrd.20224
- Nixon, B., Stanger, S. J., Mihalas, B. P., Reilly, J. N., Anderson, A. L., Tyagi, S., ... McLaughlin, E. A. (2015). The MicroRNA Signature of Mouse Spermatozoa Is Substantially Modified During Epididymal Maturation. *Biology of Reproduction*, 93(4). <https://doi.org/10.1095/biolreprod.115.132209>
- Nordstrand, L. M., Svärd, J., Larsen, E., Nilsen, A., Ougland, R., Furu, K., ... Klungland, A. (2010). Mice Lacking Alkbh1 Display Sex-Ratio Distortion and Unilateral Eye Defects. *PLoS ONE*, 5(11), e13827. <https://doi.org/10.1371/journal.pone.0013827>
- Oakberg, E. F. (1956). Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. *American Journal of Anatomy*, 99(3), 507–516. <https://doi.org/10.1002/aja.1000990307>
- Oh, B., Hampl, A., Eppig, J. J., Solter, D., & Knowles, B. B. (1998). SPIN, a substrate in the MAP kinase pathway in mouse oocytes. *Molecular Reproduction and Development*, 50(2), 240–249. [https://doi.org/10.1002/\(SICI\)1098-2795\(199806\)50:2<240::AID-MRD15>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1098-2795(199806)50:2<240::AID-MRD15>3.0.CO;2-A)
- Organisation for Economic Co-operation and Development (2016). *OECD-FAO Agricultural Outlook* (No. 1563–447). Paris. Retrieved from <http://www.oecd-ilibrary.org/content/serial/19991142>
- Pacheco, S. E., Houseman, E. A., Christensen, B. C., Marsit, C. J., Kelsey, K. T., Sigman, M., & Boekelheide, K. (2011). Integrative DNA Methylation and Gene Expression Analyses Identify DNA Packaging and Epigenetic Regulatory Genes Associated with Low Motility Sperm. *PLOS ONE*, 6(6), e20280. <https://doi.org/10.1371/journal.pone.0020280>
- Pal, A., & Vicoso, B. (2015). The X Chromosome of Hemipteran Insects: Conservation, Dosage Compensation and Sex-Biased Expression. *Genome Biology and Evolution*, 7(12), 3259–3268. <https://doi.org/10.1093/gbe/evv215>
- Palmer, N. O., Bakos, H. W., Owens, J. A., Setchell, B. P., & Lane, M. (2012). Diet and exercise in an obese mouse fed a high-fat diet improve metabolic health and reverse perturbed sperm function. *American Journal of Physiology - Endocrinology and Metabolism*, 302(7), E768–E780. <https://doi.org/10.1152/ajpendo.00401.2011>
- Parisi, M., Nuttall, R., Naiman, D., Bouffard, G., Malley, J., Andrews, J., ... Oliver, B. (2003). Paucity of Genes on the *Drosophila* X Chromosome Showing Male-Biased Expression. *Science*, 299(5607), 697–700. <https://doi.org/10.1126/science.1079190>

- Park, C., Yu, N., Choi, I., Kim, W., & Lee, S. (2014). lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics*, 30(17), 2480–2485. <https://doi.org/10.1093/bioinformatics/btu325>
- Parker, G. A. (2014). The sexual cascade and the rise of pre-ejaculatory (Darwinian) sexual selection, sex roles, and sexual conflict. *Cold Spring Harbor Perspectives in Biology*, 6(10), a017509. <https://doi.org/10.1101/cshperspect.a017509>
- Parker, G. A. (1970). Sperm Competition and Its Evolutionary Consequences in the Insects. *Biological Reviews*, 45(4), 525–567. <https://doi.org/10.1111/j.1469-185X.1970.tb01176.x>
- Parton, R. M., Davidson, A., Davis, I., & Weil, T. T. (2014). Subcellular mRNA localisation at a glance. *J Cell Sci*, 127(10), 2127–2133. <https://doi.org/10.1242/jcs.114272>
- Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., ... Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808), 86–89. <https://doi.org/10.1038/35040556>
- Peng, H., Shi, J., Zhang, Y., Zhang, H., Liao, S., Li, W., ... Duan, E. (2012). A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Research*, 22(11), 1609–1612. <https://doi.org/10.1038/cr.2012.141>
- Pereira, P. M. de C. C., & Vicente, A. F. dos R. B. (2013). Meat nutritional composition and nutritive role in the human diet. *Meat Science*, 93(3), 586–592. <https://doi.org/10.1016/j.meatsci.2012.09.018>
- Perez-Torrado, R., Yamada, D., & Defossez, P.-A. (2006). Born to bind: the BTB protein–protein interaction domain. *BioEssays*, 28(12), 1194–1202. <https://doi.org/10.1002/bies.20500>
- Perry, J., Palmer, S., Gabriel, A., & Ashworth, A. (2001). A Short Pseudoautosomal Region in Laboratory Mice. *Genome Research*, 11(11), 1826–1832. <https://doi.org/10.1101/gr.203001>
- Pertile, M. D., Graham, A. N., Choo, K. H. A., & Kalitsis, P. (2009). Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability. *Genome Research*, 19(12), 2202–2213. <http://doi.org/10.1101/gr.092080.109>
- Picard Tools V.2.1.1 (accessed 2015). <http://broadinstitute.github.io/picard>.
- Pietraszek, M., & Styrna, J. (1995). Analysis of Hya locus in B10.BR males with partial deletion of the Y chromosome. *Folia Biologica*, 43(3–4), 89–91.
- Pinto-Correia, C. (2015). Strange Tales of Small Men: Homunculi in Reproduction. *Perspectives in Biology and Medicine*, 42(2), 225–244. <https://doi.org/10.1353/pbm.1999.0046>
- Place, R. F., Li, L.-C., Pookot, D., Noonan, E. J., & Dahiya, R. (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences*, 105(5), 1608–1613. <https://doi.org/10.1073/pnas.0707594105>

- Pomponi, S. A., & Cucci, T. L. (1989). Separation and concentration of phytoplankton populations using centrifugal elutriation. *Cytometry*, 10(5), 580–586. <https://doi.org/10.1002/cyto.990100513>
- Premkumar, E., & Bhargava, P. M. (1972). Transcription and Translation in Bovine Spermatozoa. *Nature New Biology*, 240(100), 139–143. <https://doi.org/10.1038/newbio240139a0>
- Price, T. A. R., Bretman, A. J., Avent, T. D., Snook, R. R., Hurst, G. D. D., & Wedell, N. (2008a). Sex ratio distorter reduces sperm competitive ability in an insect. *Evolution; International Journal of Organic Evolution*, 62(7), 1644–1652. <https://doi.org/10.1111/j.1558-5646.2008.00386.x>
- Price, T. a. R., Hodgson, D. J., Lewis, Z., Hurst, G. D. D., & Wedell, N. (2008b). Selfish Genetic Elements Promote Polyandry in a Fly. *Science*, 322(5905), 1241–1243. <https://doi.org/10.1126/science.1163766>
- Publicover, S., Harper, C. V., & Barratt, C. (2007). [Ca²⁺]_i signalling in sperm — making the most of what you've got. *Nature Cell Biology*, 9(3), 235–242. <https://doi.org/10.1038/ncb0307-235>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Rahman, M. S., Kwon, W.-S., & Pang, M.-G. (2014). Calcium Influx and Male Fertility in the Context of the Sperm Proteome: An Update [Research article]. <https://doi.org/10.1155/2014/841615>
- Rath, D., Barcikowski, S., Graaf, S. de, Garrels, W., Grossfeld, R., Klein, S., ... Washausen, S. (2013). Sex selection of sperm in farm animals: status report and developmental prospects. *Reproduction*, 145(1), R15–R30. <https://doi.org/10.1530/REP-12-0151>
- Raychaudhuri, S., Stuart, J. M., & Altman, R. B. (2000). Principal components analysis to summarize microarray experiments: application to sporulation time series. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 455–466.
- Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., & Bartel, D. P. (2002). MicroRNAs in plants. *Genes & Development*, 16(13), 1616–1626. <https://doi.org/10.1101/gad.1004402>
- Reynard, L. N., Cocquet, J., & Burgoyne, P. S. (2009). The Multi-Copy Mouse Gene Sycp3-Like Y-Linked (*Sly*) Encodes an Abundant Spermatid Protein That Interacts with a Histone Acetyltransferase and an Acrosomal Protein. *Biology of Reproduction*, 81(2), 250–257. <https://doi.org/10.1095/biolreprod.108.075382>
- Reynard, L. N., & Turner, J. M. A. (2009). Increased sex chromosome expression and epigenetic abnormalities in spermatids from male mice with Y chromosome deletions. *Journal of Cell Science*, 122(22), 4239–4248. <https://doi.org/10.1242/jcs.049916>
- Rice, W. R. (1984). Sex Chromosomes and the Evolution of Sexual Dimorphism. *Evolution*, 38(4), 735–742. <https://doi.org/10.2307/2408385>

- Rigaud, T., Moreau, J., & Juchault, P. (1999). Wolbachia infection in the terrestrial isopod *Oniscus asellus*: sex ratio distortion and effect on fecundity. *Heredity*, 83(4), 469–475. <https://doi.org/10.1038/sj.hdy.6885990>
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43(7), e47. Package version: 3.30.13
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rodgers, A. B., Morgan, C. P., Bronson, S. L., Revello, S., & Bale, T. L. (2013). Paternal Stress Exposure Alters Sperm MicroRNA Content and Reprograms Offspring HPA Stress Axis Regulation. *Journal of Neuroscience*, 33(21), 9003–9012. <https://doi.org/10.1523/JNEUROSCI.0914-13.2013>
- Rodgers, A. B., Morgan, C. P., Leu, N. A., & Bale, T. L. (2015). Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proceedings of the National Academy of Sciences*, 112(44), 13699–13704. <https://doi.org/10.1073/pnas.1508347112>
- Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L., & Bradley, A. (2004). Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Research*, 14(10a), 1902–1910. <https://doi.org/10.1101/gr.2722704>
- Romero, Y., Meikar, O., Papaioannou, M. D., Conne, B., Grey, C., Weier, M., ... Nef, S. (2011). Dicer1 Depletion in Male Germ Cells Leads to Infertility Due to Cumulative Meiotic and Spermiogenic Defects. *PLOS ONE*, 6(10), e25241. <https://doi.org/10.1371/journal.pone.0025241>
- Romrell, L. J., Bellvé, A. R., & Fawcett, D. W. (1976). Separation of mouse spermatogenic cells by sedimentation velocity. *Developmental Biology*, 49(1), 119–131. [https://doi.org/10.1016/0012-1606\(76\)90262-1](https://doi.org/10.1016/0012-1606(76)90262-1)
- Rosegrant, M.W. et al. (2009). Looking into the future for agriculture and AKST (Agricultural Knowledge, Science and Technology). *Agriculture at a Crossroads*, 307–376.
- Rosenfeld, C. S., Grimm, K. M., Livingston, K. A., Brokman, A. M., Lamberson, W. E., & Roberts, R. M. (2003). Striking variation in the sex ratio of pups born to mice according to whether maternal diet is high in fat or carbohydrate. *Proceedings of the National Academy of Sciences*, 100(8), 4628–4632. <https://doi.org/10.1073/pnas.0330808100>
- Royo, H., Seitz, H., Ellnati, E., Peters, A. H. F. M., Stadler, M. B., & Turner, J. M. A. (2015). Silencing of X-Linked MicroRNAs by Meiotic Sex Chromosome Inactivation. *PLoS Genetics*, 11(10), e1005461. <https://doi.org/10.1371/journal.pgen.1005461>
- Russell, L. D. (1990). *Histological and histopathological evaluation of the testis*. Cache River Press.

- Safronova, L. D., Kudryavtsev, I. V., & Kudryavtsev, P. I. (2002). Sterility of Males Determined by Functional Features of the Mouse Spermatozoa Bearing t-Complex. *Russian Journal of Developmental Biology*, 33(3), 131–135. <https://doi.org/10.1023/A:1015674422803>
- Sati, L., & Huszar, G. (2013). Methodology of aniline blue staining of chromatin and the assessment of the associated nuclear and cytoplasmic attributes in human sperm. *Methods in Molecular Biology* (Clifton, N.J.), 927, 425–436. https://doi.org/10.1007/978-1-62703-038-0_36
- Schmidt, A., & Hall, A. (2002). Guanine nucleotide exchange factors for Rho GTPases: turning on the switch. *Genes & Development*, 16(13), 1587–1609. <https://doi.org/10.1101/gad.1003302>
- Searles, C. D., Ide, L., Davis, M. E., Cai, H., & Weber, M. (2004). Actin Cytoskeleton Organization and Posttranscriptional Regulation of Endothelial Nitric Oxide Synthase During Cell Growth. *Circulation Research*, 95(5), 488–495. <https://doi.org/10.1161/01.RES.0000138953.21377.80>
- Setti, A. S., Paes de Almeida Ferreira Braga, D., Iaconelli, A., Aoki, T., & Borges, E. (2013). Twelve years of MSOME and IMSI: a review. *Reproductive Biomedicine Online*, 27(4), 338–352. <https://doi.org/10.1016/j.rbmo.2013.06.011>
- Shang, E., Nickerson, H. D., Wen, D., Wang, X., & Wolgemuth, D. J. (2007). The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation. *Development* (Cambridge, England), 134(19), 3507–3515. <https://doi.org/10.1242/dev.004481>
- Sharma, U., Conine, C. C., Shea, J. M., Boskovic, A., Derr, A. G., Bing, X. Y., ... Rando, O. J. (2015). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science*, aad6780. <https://doi.org/10.1126/science.aad6780>
- Sharpe, J. C., & Evans, K. M. (2009). Advances in flow cytometry for sperm sexing. *Theriogenology*, 71(1), 4–10. <https://doi.org/10.1016/j.theriogenology.2008.09.021>
- Shiraishi, K., Oka, S., & Matsuyama, H. (2017). Predictive Factors for Sperm Recovery after Varicocele in Men with Nonobstructive Azoospermia. *The Journal of Urology*, 197(2), 485–490. <https://doi.org/10.1016/j.juro.2016.08.085>
- Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., ... Goodfellow, P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, 346(6281), 240–244. <https://doi.org/10.1038/346240a0>
- Sit, S.-T., & Manser, E. (2011). Rho GTPases and their role in organizing the actin cytoskeleton. *J Cell Sci*, 124(5), 679–683. <https://doi.org/10.1242/jcs.064964>
- Skinner, B. M., & Johnson, E. E. P. (2017). Nuclear morphologies: their diversity and functional relevance. *Chromosoma*, 126(2), 195–212. <https://doi.org/10.1007/s00412-016-0614-5>

- Soh, Y. Q. S., Alföldi, J., Pyntikova, T., Brown, L. G., Graves, T., Minx, P. J... Page, D. C. (2014). Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. *Cell*, 159(4), 800–813. <https://doi.org/10.1016/j.cell.2014.09.052>
- Solari, A. J. (1969). The evolution of the ultrastructure of the sex chromosomes (sex vesicle) during meiotic prophase in mouse spermatocytes. *Journal of Ultrastructure Research*, 27(3), 289–305. [https://doi.org/10.1016/S0022-5320\(69\)80018-3](https://doi.org/10.1016/S0022-5320(69)80018-3)
- Sperry, A. O. (2012). The dynamic cytoskeleton of the developing male germ cell. *Biology of the Cell / under the Auspices of the European Cell Biology Organization*, 104(5). <https://doi.org/10.1111/boc.201100102>
- Styrna, J., Imai, H. T., & Moriwaki, K. (1991a). An increased level of sperm abnormalities in mice with a partial deletion of the Y chromosome. *Genetics Research*, 57(2), 195–199. <https://doi.org/10.1017/S0016672300029268>
- Styrna, J., Klag, J., & Moriwaki, K. (1991b). Influence of partial deletion of the Y chromosome on mouse sperm phenotype. *Journal of Reproduction and Fertility*, 92(1), 187–195.
- Styrna, J., & Krzanowska, H. (1995). Sperm Select Penetration Test Reveals Differences in Sperm Quality in Strains with Different Y Chromosome Genotype in Mice. *Systems Biology in Reproductive Medicine*, 35(2), 111–118. <https://doi.org/10.3109/01485019508987861>
- Styrna, J., Bilińska, B., & Krzanowska, H. (2002). The effect of a partial Y chromosome deletion in B10.BR-Ydel mice on testis morphology, sperm quality and efficiency of fertilization. *Reproduction, Fertility and Development*, 14(2), 101–108.
- Styrna, J., Kilarski, W., & Krzanowska, H. (2003). Influence of the CBA genetic background on sperm morphology and fertilization efficiency in mice with a partial Y chromosome deletion. *Reproduction*, 126(5), 579–588. <https://doi.org/10.1530/rep.0.1260579>
- Sutter, A., & Lindholm, A. K. (2015). Detrimental effects of an autosomal selfish genetic element on sperm competitiveness in house mice. *Proc. R. Soc. B*, 282(1811), 20150974. <https://doi.org/10.1098/rspb.2015.0974>
- Sutton, E., Hughes, J., White, S., Sekido, R., Tan, J., Arboleda, V., ... Thomas, P. (2011). Identification of SOX3 as an XX male sex reversal gene in mice and humans. *The Journal of Clinical Investigation*, 121(1), 328–341. <https://doi.org/10.1172/JCI42580>
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., ... von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, 43(Database issue), D447–452. <https://doi.org/10.1093/nar/gku1003>
- Szot, M., Grigoriev, V., Mahadevaiah, S. K., Ojarikre, O. A., Touré, A., von Glasenapp, E., ... Burgoyne, P. S. (2003). Does Rbmy have a role in sperm development in mice? *Cytogenetic and Genome Research*, 103(3–4), 330–336. <https://doi.org/76821>

- 't Hoen, P. A. C., Friedländer, M. R., Almlöf, J., Sammeth, M., Pulyakhina, I., Anvar, S. Y., ... Lappalainen, T. (2013). Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nature Biotechnology*, 31(11), 1015–1022. <https://doi.org/10.1038/nbt.2702>
- Taelman, V. F., Dobrowolski, R., Plouhinec, J.-L., Fuentealba, L. C., Vorwald, P. P., Gumper, I., ... Robertis, E. M. D. (2010). Wnt Signaling Requires Sequestration of Glycogen Synthase Kinase 3 inside Multivesicular Endosomes. *Cell*, 143(7), 1136–1148. <https://doi.org/10.1016/j.cell.2010.11.034>
- Tan, T., Zhang, Y., Ji, W., & Zheng, P. (2014). miRNA signature in mouse spermatogonial stem cells revealed by high-throughput sequencing. *BioMed Research International*, 2014, 154251. <https://doi.org/10.1155/2014/154251>
- Tao, Y., Masly, J. P., Araripe, L., Ke, Y., & Hartl, D. L. (2007a). A sex-ratio Meiotic Drive System in *Drosophila simulans*. I: An Autosomal Suppressor. *PLOS Biology*, 5(11), e292. <https://doi.org/10.1371/journal.pbio.0050292>
- Tao, Y., Araripe, L., Kingan, S. B., Ke, Y., Xiao, H., & Hartl, D. L. (2007b). A sex-ratio Meiotic Drive System in *Drosophila simulans*. II: An X-linked Distorter. *PLOS Biology*, 5(11), e293. <https://doi.org/10.1371/journal.pbio.0050293>
- Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., & Conesa, A. (2011). Differential expression in RNA-seq: A matter of depth. *Genome Research*, 21(12), 2213–2223. <https://doi.org/10.1101/gr.124321.111>
- Thaler, C. D., & Cardullo, R. A. (1995). Biochemical Characterization of a Glycosylphosphatidylinositol-Linked Hyaluronidase on Mouse Sperm. *Biochemistry*, 34(24), 7788–7795. <https://doi.org/10.1021/bi00024a002>
- The World Population Prospects: 2015 Revision*. (2015). (Report). Retrieved from world-population-prospects-2015-revision.html
- Toray (retrieved 2017), miRNA Oligo Chip webpage – 3D-Gene®. Retrieved from: <http://www.3d-gene.com/en/products/dna/index.html> on 6th July 2017.
- Touré, A., Szot, M., Mahadevaiah, S. K., Rattigan, Á., Ojarikre, O. A., & Burgoyne, P. S. (2004a). A New Deletion of the Mouse Y Chromosome Long Arm Associated With the Loss of Ssty Expression, Abnormal Sperm Development and Sterility. *Genetics*, 166(2), 901–912. <https://doi.org/10.1534/genetics.166.2.901>
- Touré, A., Grigoriev, V., Mahadevaiah, S. K., Rattigan, Á., Ojarikre, O. A., & Burgoyne, P. S. (2004b). A protein encoded by a member of the multicopy Ssty gene family located on the long arm of the mouse Y chromosome is expressed during sperm development. *Genomics*, 83(1), 140–147. [https://doi.org/10.1016/S0888-7543\(03\)00216-7](https://doi.org/10.1016/S0888-7543(03)00216-7)
- Touré, A., Clemente, E. J., Ellis, P., Mahadevaiah, S. K., Ojarikre, O. A., Ball, P. A. F., ... Affara, N. A. (2005). Identification of novel Y chromosome encoded transcripts by testis transcriptome analysis of mice with

deletions of the Y chromosome long arm. *Genome Biology*, 6(12), R102. <https://doi.org/10.1186/gb-2005-6-12-r102>

Tourmente, M., Zarka-Trigo, D., & Roldan, E. R. S. (2016). Is the hook of muroid rodent's sperm related to sperm train formation? *Journal of Evolutionary Biology*, 29(6), 1168–1177. <https://doi.org/10.1111/jeb.12857>

Trivers, R. L., & Willard, D. E. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science (New York, N.Y.)*, 179(4068), 90–92.

Turner, J. M., Mahadevaiah, S. K., Benavente, R., Offenberger, H. H., Heyting, C., & Burgoyne, P. S. (2000). Analysis of male meiotic 'sex body' proteins during XY female meiosis provides new insights into their functions. *Chromosoma*, 109(6), 426–432.

Turner, J. M. A. (2007). Meiotic sex chromosome inactivation. *Development*, 134(10), 1823–1831. <https://doi.org/10.1242/dev.000018>

van der Horst, G., Maree, L., Kotzé, S. H., & O'Riain, M. J. (2011). Sperm structure and motility in the eusocial naked mole-rat, *Heterocephalus glaber*: a case of degenerative orthogenesis in the absence of sperm competition? *BMC Evolutionary Biology*, 11, 351. <https://doi.org/10.1186/1471-2148-11-351>

Varea-Sánchez, M., Bastir, M., & Roldan, E. R. S. (2013). Geometric Morphometrics of Rodent Sperm Head Shape. *PLOS ONE*, 8(11), e80607. <https://doi.org/10.1371/journal.pone.0080607>

Vasileva, A., Tiedau, D., Firooznia, A., Müller-Reichert, T., & Jessberger, R. (2009). Tdrd6 Is Required for Spermiogenesis, Chromatoid Body Architecture, and Regulation of miRNA Expression. *Current Biology*, 19(8), 630–639. <https://doi.org/10.1016/j.cub.2009.02.047>

Vavouri, T., & Lehner, B. (2011). Chromatin Organization in Sperm May Be the Major Functional Consequence of Base Composition Variation in the Human Genome. *PLOS Genetics*, 7(4), e1002036. <https://doi.org/10.1371/journal.pgen.1002036>

Vedeler, A., & Hollås, H. (2000). Annexin II is associated with mRNAs which may constitute a distinct subpopulation. *Biochemical Journal*, 348(3), 565–572. <https://doi.org/10.1042/bj3480565>

Ventelä, S., Toppari, J., & Parvinen, M. (2003). Intercellular Organelle Traffic through Cytoplasmic Bridges in Early Spermatids of the Rat: Mechanisms of Haploid Gene Product Sharing. *Molecular Biology of the Cell*, 14(7), 2768–2780. <http://doi.org/10.1091/mbc.E02-10-0647>

Ventelä, S. (2006). Cytoplasmic Bridges as Cell-Cell Channels of Germ Cells. In *Cell-Cell Channels* (pp. 208–216). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-0-387-46957-7_15

Véron, N., Bauer, H., Weiße, A. Y., Lüder, G., Werber, M., & Herrmann, B. G. (2009). Retention of gene products in syncytial spermatids promotes non-Mendelian inheritance as revealed by the t complex responder. *Genes & Development*, 23(23), 2705–2710. <https://doi.org/10.1101/gad.553009>

- Vermeulen, J., De Preter, K., Lefever, S., Nuytens, J., De Vloed, F., Derveaux, S., ... Vandesompele, J. (2011). Measurable impact of RNA quality on gene expression results from quantitative PCR. *Nucleic Acids Research*, 39(9), e63. <https://doi.org/10.1093/nar/gkr065>
- Vitsios, D. M., & Enright, A. J. (2015). Chimira: analysis of small RNA sequencing data and microRNA modifications. *Bioinformatics* (Oxford, England), 31(20), 3365–3367. <https://doi.org/10.1093/bioinformatics/btv380>
- Wang, P. J., McCarrey, J. R., Yang, F., & Page, D. C. (2001). An abundance of X-linked genes expressed in spermatogonia. *Nature Genetics*, 27(4), 422–426. <https://doi.org/10.1038/86927>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, 10(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Ward, M. A., & Burgoyne, P. S. (2006). The effects of deletions of the mouse Y chromosome long arm on sperm function—intracytoplasmic sperm injection (ICSI)-based analysis. *Biology of Reproduction*, 74(4), 652–658. <https://doi.org/10.1095/biolreprod.105.048090>
- Watanabe, H., & Kondoh, G. (2011). Mouse sperm undergo GPI-anchored protein release associated with lipid raft reorganization and acrosome reaction to acquire fertility. *Journal of Cell Science*, 124(Pt 15), 2573–2581. <https://doi.org/10.1242/jcs.086967>
- Wei, K. H.-C., & Barbash, D. A. (2015). Never Settling Down: Frequent Changes in Sex Chromosomes. *PLoS Biology*, 13(4). <https://doi.org/10.1371/journal.pbio.1002077>
- Wilhelm, B. T., & Landry, J.-R. (2009). RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, 48(3), 249–257. <https://doi.org/10.1016/j.ymeth.2009.03.016>
- Wilkins, A. S., & Holliday, R. (2009). The Evolution of Meiosis From Mitosis. *Genetics*, 181(1), 3–12. <https://doi.org/10.1534/genetics.108.099762>
- Wittmann, T., & Waterman-Storer, C. M. (2001). Cell motility: can Rho GTPases and microtubules point the way? *Journal of Cell Science*, 114(21), 3795–3803.
- Wong, N., & Wang, X. (2015). miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Research*, 43(Database issue), D146–152. <https://doi.org/10.1093/nar/gku1104>
- World Health Organisation (WHO) (2010). WHO laboratory manual for the examination and processing of human semen.. Retrieved 3 June 2017, from <http://www.who.int/reproductivehealth/publications/infertility/9789241547789/en/>
- Xian, M., Azuma, S., Naito, K., Kunieda, T., Moriwaki, K., & Toyoda, Y. (1992). Effect of a partial deletion of Y chromosome on *in vitro* fertilizing ability of mouse spermatozoa. *Biology of Reproduction*, 47(4), 549–553. <https://doi.org/10.1095/biolreprod47.4.549>

- Xu, P., Vernooy, S. Y., Guo, M., & Hay, B. A. (2003). The Drosophila MicroRNA Mir-14 Suppresses Cell Death and Is Required for Normal Fat Metabolism. *Current Biology*, 13(9), 790–795. [https://doi.org/10.1016/S0960-9822\(03\)00250-1](https://doi.org/10.1016/S0960-9822(03)00250-1)
- Yamauchi, Y., Riel, J. M., Stoytcheva, Z., Burgoyne, P. S., & Ward, M. A. (2010). Deficiency in mouse Y chromosome long arm gene complement is associated with sperm DNA damage. *Genome Biology*, 11, R66. <https://doi.org/10.1186/gb-2010-11-6-r66>
- Yamauchi, Y., Riel, J. M., Ruthig, V. A., Ortega, E. A., Mitchell, M. J., & Ward, M. A. (2016). Two genes substitute for the mouse Y chromosome for spermatogenesis and reproduction. *Science*, 351(6272), 514–516. <https://doi.org/10.1126/science.aad1795>
- Yan, W., & McCarrey, J. R. (2009). Sex Chromosome Inactivation in the Male. *Epigenetics : Official Journal of the DNA Methylation Society*, 4(7), 452–456.
- Yanagimachi, R. (2011). Mammalian Sperm Acrosome Reaction: Where Does It Begin Before Fertilization? *Biology of Reproduction*, 85(1), 4–5. <https://doi.org/10.1095/biolreprod.111.092601>
- Yao, R., Ito, C., Natsume, Y., Sugitani, Y., Yamanaka, H., Kuretake, S., ... Noda, T. (2002). Lack of acrosome formation in mice lacking a Golgi protein, GOPC. *Proceedings of the National Academy of Sciences*, 99(17), 11211–11216. <https://doi.org/10.1073/pnas.162027899>
- Yao, C., Liu, Y., Sun, M., Niu, M., Yuan, Q., Hai, Y., ... He, Z. (2015). MicroRNAs and DNA methylation as epigenetic regulators of mitosis, meiosis and spermiogenesis. *Reproduction*, 150(1), R25–R34. <https://doi.org/10.1530/REP-14-0643>
- Yoshida, S., Sukeno, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T., & Nabeshima, Y. (2006). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. *Development*, 133(8), 1495–1505. <https://doi.org/10.1242/dev.02316>
- Yu, Z., Raabe, T., & Hecht, N. B. (2005). MicroRNA Mirn122a Reduces Expression of the Posttranscriptionally Regulated Germ Cell Transition Protein 2 (Tnp2) Messenger RNA (mRNA) by mRNA Cleavage. *Biology of Reproduction*, 73(3), 427–433. <https://doi.org/10.1095/biolreprod.105.040998>
- Zackowski, J. L., & Martin-DeLeon, P. A. (1989). Segregation products of male mice doubly heterozygous for the RB(6.16) and RB(16.17) translocations: Influence of sperm karyotype on fertilizing competence under varying mating frequencies. *Gamete Research*, 22(1), 93–107. <https://doi.org/10.1002/mrd.1120220110>
- Zahn, C. T., & Roskies, R. Z. (1972). Fourier Descriptors for Plane Closed Curves. *IEEE Transactions on Computers*, C-21(3), 269–281. <https://doi.org/10.1109/TC.1972.5008949>
- Zhang, H., & Martin-DeLeon, P. A. (2003). Mouse Spam1 (PH-20) Is a Multifunctional Protein: Evidence for Its Expression in the Female Reproductive Tract. *Biology of Reproduction*, 69(2), 446–454. <https://doi.org/10.1095/biolreprod.102.013854>

- Zheng, Y., & Martin-Deleon, P. A. (1999). Characterization of the genomic structure of the murine Spam1 gene and its promoter: evidence for transcriptional regulation by a cAMP-responsive element. *Molecular Reproduction and Development*, 54(1), 8–16. [https://doi.org/10.1002/\(SICI\)1098-2795\(199909\)54:1<8::AID-MRD2>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1098-2795(199909)54:1<8::AID-MRD2>3.0.CO;2-D)
- Zheng, Y., Deng, X., & Martin-DeLeon, P. A. (2001a). Lack of Sharing of Spam1 (Ph-20) among Mouse Spermatids and Transmission Ratio Distortion. *Biology of Reproduction*, 64(6), 1730–1738. <https://doi.org/10.1095/biolreprod64.6.1730>
- Zheng, Y., Deng, X., Zhao, Y., Zhang, H., & Martin-DeLeon, P. A. (2001b). Spam1 (PH-20) mutations and sperm dysfunction in mice with the Rb(6.16) or Rb(6.15) translocation. *Mammalian Genome*, 12(11), 822–829. <https://doi.org/10.1007/s00335-001-1008-3>
- Zheng, G. X. Y., Ravi, A., Gould, G. M., Burge, C. B., & Sharp, P. A. (2011). Genome-wide impact of a recently expanded microRNA cluster in mouse. *Proceedings of the National Academy of Sciences*, 108(38), 15804–15809. <https://doi.org/10.1073/pnas.1112772108>
- Zhou, C., Kang, W., & Baba, T. (2012). Functional Characterization of Double-knockout Mouse Sperm Lacking SPAM1 and ACR or SPAM1 and PRSS21 in Fertilization. *Journal of Reproduction and Development*, 58(3), 330–337.
- Zimmer, C. (2009). On the Origin of Sexual Reproduction. *Science*, 324(5932), 1254–1256. https://doi.org/10.1126/science.324_1254
- Zimmerman, S. G., Peters, N. C., Altaras, A. E., & Berg, C. A. (2013). Optimized RNA ISH, RNA FISH and protein-RNA double labeling (IF/FISH) in *Drosophila* ovaries. *Nature Protocols*, 8(11), 2158–2179. <https://doi.org/10.1038/nprot.2013.136>
- Zimmermann, C., Stévant, I., Borel, C., Conne, B., Pitetti, J.-L., Calvel, P., ... Nef, S. (2015). Research resource: the dynamic transcriptional profile of sertoli cells during the progression of spermatogenesis. *Molecular Endocrinology (Baltimore, Md.)*, 29(4), 627–642. <https://doi.org/10.1210/me.2014-1356>