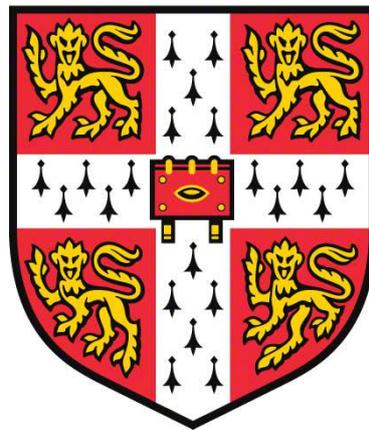


Modelling neurodegenerative diseases in human iPSC-derived neurons

Ryan Prestil



**UNIVERSITY OF
CAMBRIDGE**

Darwin College

28th of July, 2021

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Degree Committee for the Faculties of Clinical and Veterinary Medicine of 60,000 words.

Ryan Prestil

28th of July, 2021

Abstract

Neurodegeneration is a pathology shared by a varied class of diseases, and many of the mutations that are known to cause such diseases have been linked to protein aggregation and autophagy dysfunction. Improvements to gene editing and neuronal differentiation strategies have enabled the derivation of *in vitro* disease models using human iPSC-derived neurons to provide a more accurate understanding of how disease mutations affect neuronal health.

I first sought to model the polyglutamine disease spinal and bulbar muscular atrophy (SBMA), detailed in Chapter 4. Using iPSCs derived from a healthy donor and an SBMA patient, the CAG repeat of the endogenous androgen receptor (*AR*) gene was CRISPR-edited to encode a series of lengths or an early stop codon. However, *AR* expression was silenced upon transcription factor-mediated differentiation to a lower motor neuron-like state, and chemical differentiation prevented ligand-induced AR nuclear translocation. Deriving the cell lines in this work highlighted that purification of transgenic cells is a key bottleneck to gene editing. I therefore adapted a synthetic marker gene that presents a streptavidin binding peptide (SBP) tag on the extracellular membrane, detailed in Chapter 3. Expression of this tag in iPSCs enabled transient fluorophore staining and effective sorting of mixed populations with magnetic streptavidin beads. Finally, Chapter 5 establishes *L1CAM* as a novel autophagy modulator; iPSC-derived neurons showed that reduction of the *L1CAM* transcript with shRNAs, but not genetic knockout of the L1CAM protein, is sufficient to reduce transcription of the *ATG8* gene family, which are core components of macroautophagy.

This work exemplified both the strengths and weaknesses of iPSC-derived neurons; namely, they are tractable and able to recapitulate neural phenotypes, but deriving new model lines requires a high initial investment, so adequate proof-of-concept is crucial.

Acknowledgements

Thank you to everyone who has been kind to me, who has helped me, and who has made unforgettable memories with me. Thank you in particular:

To my supervisors, Kenneth Fischbeck & David Rubinsztein. You are both inspirations, and you have each made me a better scientist and a better person.

To the NIH Oxford-Cambridge Scholars Program and the Cambridge Trust. Thank you for supporting me and funding me throughout this process.

To those who inspired me, taught me, and believed in me when I had no idea what I was doing, especially Travis Cordie, Krishanu Saha, Marvin Wickens, Hernando Martínez Vergara, and Detlev Arendt. Without your influence, I would not have even begun a PhD. Thank you for your patience, for imparting your knowledge and passion, and for being excellent role models.

To everyone that has supported me and my work, especially: at the NIH—George Harmison, Christopher Grunseich, Michael Ward, Kory Johnson, Martha Kirby & Stacie Anderson (NHGRI flow cytometry core), and Steven Coon & Tianwei Li (NICHD molecular genomics core), and at Cambridge—Victoria Barratt; Farah Siddiqi; Lidia Wrobel; Sandra Malmgren Hill; Sungmin Son; Reiner Schulte, Chiara Cossetti, and Gabriela Grondys-Kotarba (CIMR flow cytometry core); and Matthew Gratian & Mark Bowen (CIMR microscopy core). You have all been instrumental in helping me and teaching me. Thank you for answering my many questions.

To Rachel Movsas, for trusting me to teach you and guide you as a summer student. I hope the experience was as rewarding for you as it was for me being your mentor.

To all those who have been sounding boards, confidants, and friends, especially: Sarah Carlo, Madeline Epping, Michael Fernandopulle, Stewart Humble, Rajan Patel, Ling Hao, Mathias Bellaiche, Fátima Ruiz, Cameron Gardner, Alvin Djajadikerta, Laura

Ryan, Swati Keshri, Ruby Zhu, Cansu Karabiyik, Jun Rui He, Matea Rob, Jonathon Nixon-Abell, Oliver Fleck, Stephanie Metzger, and Arthur Davis. Each of you has influenced who I am, and each of you has inspired who I want to become. I could not have made it through the hard times without your support, and I can think of no better companions with whom to celebrate the good times.

To all of my fellow Darwinians—I have never felt so at home.

And to my parents and family, for your unwavering support, belief, and encouragement to follow my dreams, even when they have taken me half a world away.

Table of Contents

Abstract	i
Acknowledgements	ii
Table of Contents.....	iv
Index of Figures.....	xi
Index of Tables.....	xiii
Preface.....	xv
Chapter 1: Introduction.....	1
1.1 Neurodegenerative diseases.....	1
1.1.1 Pathology and genetics	1
1.1.2 Protein aggregation.....	2
1.1.3 Polyglutamine diseases.....	3
1.2 Autophagy.....	4
1.2.1 Degradation pathways	4
1.2.2 Autophagosome synthesis in humans	6
1.2.3 The <i>ATG8</i> gene family	7
1.2.4 Transcriptional control of macroautophagy genes.....	11
1.3 Development of <i>in vitro</i> human neuronal models	12
1.3.1 Advantages of a human model.....	12
1.3.2 Advantages of a neuronal model	13
1.3.3 Human stem cells – past and future.....	15
1.4 Neuronal differentiation strategies.....	17
1.4.1 The epigenetic landscape of cell state conversions	17
1.4.2 Chemical differentiation.....	18
1.4.3 Transcription factor-mediated differentiation	19
1.5 Gene editing strategies	22
1.6 Modelling disease with iPSCs	24
1.7 Summary & Aims.....	25

Chapter 2: Materials & Methods	27
2.1 Reagents.....	27
2.2 Plasmids	27
2.2.1 Gifted & purchased plasmids	27
2.2.2 Lab-constructed plasmids	28
2.2.3 RNA expression plasmids.....	31
2.2.4 Transformation	36
2.3 iPSC culture.....	36
2.3.1 iPSC line derivation.....	37
2.3.2 Maintenance culture	37
2.3.3 Plasmid transfection	38
2.3.4 Initial selection & enrichment.....	39
2.3.5 Clonal isolation	39
2.3.6 Genotyping	40
2.3.7 Karyotyping.....	42
2.4 Neural differentiation of iPSCs.....	43
2.4.1 Transcription factor-mediated differentiation	43
2.4.2 Chemical differentiation.....	45
2.5 HeLa & HEK 293T culture	48
2.5.1 Maintenance culture	48
2.5.2 Plasmid transfection	49
2.5.3 siRNA transfection.....	49
2.5.4 Lentiviral preparation	50
2.5.5 Transduction of iPSCs	50
2.6 Drug treatments	51
2.7 Antibodies	52
2.8 Western blotting.....	54
2.8.1 Sample preparation.....	54
2.8.2 Polyacrylamide gel electrophoresis.....	54

2.8.3	Transferring, marking, & imaging.....	55
2.8.4	Quantification & analysis.....	56
2.9	Microscopy	57
2.9.1	Immunocytochemistry.....	57
2.9.2	Imaging.....	58
2.9.3	Cell state determination.....	59
2.10	Flow cytometry	60
2.11	Optimization of SBP- Δ LNGFR	61
2.11.1	Definitions.....	61
2.11.2	Experimental design.....	62
2.11.3	Primary measures	65
2.11.4	Number of washes.....	68
2.11.5	Bead volume	69
2.11.6	Incubation time.....	69
2.11.7	Cell density.....	69
2.12	Aggregating protein clearance assays	70
2.13	Proteasome activity reporter assay.....	70
2.14	Transcriptional analyses.....	70
2.14.1	RNA isolation	70
2.14.2	RT-qPCR.....	71
2.14.3	RNAseq.....	72
2.15	Statistical analyses	74
Chapter 3: Sorting transgenic iPSCs by streptavidin affinity		75
3.1	Introduction	75
3.1.1	Motivation	75
3.1.2	The selection gene toolbox.....	76
3.1.3	Magnetic bead cell sorting	78
3.1.4	Aims & Hypotheses.....	79
3.2	SBP- Δ LNGFR validation	80

3.2.1	Stable cell line derivation.....	80
3.2.2	SBP- Δ LNGFR co-expresses with mApple and binds to streptavidin	81
3.2.3	Proof-of-concept of enrichment	85
3.3	Protocol optimization	88
3.3.1	Aims	88
3.3.2	Number of washes.....	88
3.3.3	Bead volume	92
3.3.4	Incubation time.....	93
3.3.5	Cell density.....	93
3.3.6	Summary of optimization.....	97
3.4	Assessing efficacy.....	100
3.4.1	Experimental design.....	100
3.4.2	Enrichment efficiency.....	100
3.4.3	Depletion efficiency.....	103
3.5	Bayesian analysis of measured & derived values	104
3.5.1	Percentages as probabilities	104
3.5.2	Confidence of flow cytometry data	104
3.5.3	Significance of selection.....	106
3.5.4	Derived test characteristics	107
3.5.5	Characteristics of enrichment.....	109
3.5.6	Characteristics of depletion.....	112
3.5.7	Confidence of sensitivity and specificity.....	113
3.6	Modelling enrichment and depletion.....	116
3.6.1	Receiver operating characteristic (ROC)	116
3.6.2	Area under the ROC curve	118
3.6.3	Modelling selection as a function of %Ip.....	119
3.6.4	Area under the beta model curve.....	121
3.7	Discussion.....	123
3.7.1	Summary of findings	123

3.7.2	Enrichment and depletion in practice	123
3.7.3	Comparison of analysis approaches.....	125
3.7.4	Future directions.....	127
Chapter 4: Derivation of AR polyQ allelic series to model SBMA		129
4.1	Introduction	129
4.1.1	Spinal and bulbar muscular atrophy.....	129
4.1.2	Limitations of existing models	130
4.1.3	Aims & hypotheses	133
4.2	Editing the endogenous <i>AR</i> locus.....	134
4.2.1	Gene editing strategy & plasmid construction	134
4.2.2	Clonal isolation & validation	136
4.3	Mag-hNIL does not induce i ³ LMN differentiation	144
4.3.1	Insertion & genotyping of Mag-hNIL.....	144
4.3.2	A simple assay to monitor rtTA3G activity.....	147
4.3.3	Mag-hNIL cells are insensitive to dox treatment	149
4.3.4	Cre excision of selection genes does not restore differentiability	151
4.4	Mag2-hNIL induces differentiation but silences <i>AR</i>	154
4.4.1	Insertion & validation of Mag2-hNIL differentiation.....	154
4.4.2	Differentiated Mag2-hNIL i ³ LMNs lack AR expression.....	156
4.4.3	Transcriptional analysis of SBMA model lines as iPSCs & i ³ LMNs.....	158
4.5	Chemical differentiation produces DHT-insensitive LMN-like cells	163
4.6	Discussion.....	169
4.6.1	Summary of findings	169
4.6.2	Retrospective on AR silencing in i ³ LMNs	171
4.6.3	Retrospective on DHT insensitivity in chemical differentiation	175
4.6.4	Lessons learned	178
Chapter 5: <i>L1CAM</i> is a novel regulator of autophagy.....		180
5.1	Introduction	180
5.1.1	L1CAM in health & disease.....	180

5.1.2	Biology of L1CAM.....	181
5.1.3	Identification of <i>L1CAM</i> as a gene of interest	185
5.1.4	Aims & Hypotheses.....	191
5.2	LC3B is reduced by individual siRNAs but not pooled siRNAs	191
5.2.1	Pooled siRNAs reduce L1CAM but not LC3B-II in HeLa cells.....	192
5.2.2	Individual siRNAs reduce L1CAM and LC3B in HeLa cells	194
5.3	Individual siRNAs impair autophagic clearance	197
5.3.1	Mutant protein degradation is reduced in HeLa cells.....	197
5.3.2	The proteasome is not impaired by L1CAM reduction	202
5.4	Modelling loss of L1CAM in iPSC-derived neurons.....	205
5.4.1	Cell line derivation.....	205
5.4.2	CRISPRi-mediated transcriptional silencing	208
5.4.3	shRNA-mediated degradation of <i>L1CAM</i> mRNA	211
5.4.4	CRISPR-mediated knockout	215
5.5	Characterizing i ³ Neuron model lines.....	223
5.5.1	LC3B is unaffected in iPSCs.....	223
5.5.2	Pluripotency erasure is not affected by reduction of L1CAM	225
5.5.3	Neuronal acquisition is not affected by reduction of L1CAM	227
5.6	L1CAM shRNAs reduce transcription of the <i>ATG8s</i>	234
5.6.1	L1CAM shRNAs reduce levels of GABARAP propeptides	234
5.6.2	Transcriptional analysis of <i>L1CAM</i> and <i>ATG8</i> genes.....	239
5.6.3	shRNA knockdown reduces mRNA of <i>L1CAM</i> and <i>ATG8s</i>	242
5.6.4	L1CAM KO does not induce nonsense-mediated decay	242
5.7	<i>L1CAM</i> transcript, not protein, regulates <i>LC3B</i>	243
5.7.1	shRNA expression in KO i ³ Neurons reduces LC3B protein	243
5.7.2	L1CAM overexpression does not rescue LC3B.....	247
5.7.3	L1CAM overexpression does not significantly affect LC3B.....	249
5.8	Discussion.....	252
5.8.1	Summary of findings	252

5.8.2	Key outstanding questions	252
5.8.3	Implications of <i>ATG8</i> reduction in disease	258
Chapter 6: Discussion		260
6.1	Overall summary & future directions	260
6.2	Limitations of iPSC-derived neuron models	262
6.2.1	Model line derivation requires high up-front investment.....	262
6.2.2	Unexpected gene expression patterns	263
6.2.3	Difficulties manipulating and maintaining differentiated neurons.....	264
6.3	Advantages of iPSC-derived neuron models	265
6.3.1	Genetic stability produces better controls.....	265
6.3.2	Enhanced reproducibility	266
6.3.3	Recapitulation of neuron-specific phenotypes.....	267
6.4	Tools of the trade.....	268
References.....		269
Appendix		A-1
A.1: Index of Abbreviations		A-1
A.2: Reagents		A-9
A.3: Media Formulations.....		A-13

Index of Figures

Figure 1.1:	Initial steps of canonical autophagosome synthesis.	7
Figure 2.1:	Maps of Mag-hNIL and Mag2-hNIL plasmids.	29
Figure 2.2:	Maps of gRNA & shRNA-expressing plasmids.	34
Figure 2.3:	Timeline of transcription factor-mediated differentiation.	44
Figure 2.4:	Timeline of chemical differentiation.....	47
Figure 2.5:	Parameters for flow cytometry analysis.....	63
Figure 3.1:	Strategy of selection via SBP- Δ LNGFR.....	82
Figure 3.2:	Characterization of SBP- Δ LNGFR <i>in vitro</i>	84
Figure 3.3:	Establishing the enrichment protocol.	87
Figure 3.4:	Optimization of the number of washes.	89
Figure 3.5:	Optimization of bead volume and incubation time.....	94
Figure 3.6:	Optimization of cell density.....	96
Figure 3.7:	Protocol recommendations.....	99
Figure 3.8:	Quantifying enrichment and depletion.....	101
Figure 3.9:	Derived test characteristics by %Ip.	111
Figure 3.10:	Adjusted 95% confidence intervals by %Ip.	115
Figure 3.11:	ROC of enrichment & depletion.	117
Figure 3.12:	Modeling enrichment and depletion.....	122
Figure 4.1:	AR editing strategy.	135
Figure 4.2:	Genotyping strategy.	138
Figure 4.3:	Validation of AR protein expression and genomic integrity.	142
Figure 4.4:	Genotyping strategy of Mag-hNIL insertion.	146
Figure 4.5:	Mag-hNIL cells lack dox sensitivity.....	147
Figure 4.6:	Mag-hNIL cells lack functional rtTA3G.	149
Figure 4.7:	Cre excision of does not restore differentiability.	153
Figure 4.8:	Characterization of Mag2-hNIL differentiation.....	155
Figure 4.9:	Differentiation timecourse of AR expression.....	157

Figure 4.10:	Expression of lineage marker genes in parental iPSCs and i ³ LMNs..	161
Figure 4.11:	Characterization of chemically differentiated cells.....	164
Figure 4.12:	Chemical differentiation disrupts AR nuclear translocation.	164
Figure 4.13:	Lack of nuclear AR staining was not due to technical issues.	168
Figure 4.14:	Protein half-lives in i ³ Neurons.....	174
Figure 5.1:	Map of <i>L1CAM</i> features and RNA targets.....	183
Figure 5.2:	Diagram of known L1CAM cleavage products.....	184
Figure 5.3:	Identification of L1CAM as a potential autophagy modulator.	187
Figure 5.4:	Pooled siRNAs reduce L1CAM but not LC3B-II.....	193
Figure 5.5:	Deconvolution of L1CAM siRNAs.	195
Figure 5.6:	Individual siRNAs impair eGFP-A53T clearance.	198
Figure 5.7:	Individual siRNAs impair eGFP-HTT74 clearance.....	201
Figure 5.8:	Individual siRNAs do not impair proteasomal clearance.	202
Figure 5.9:	Stable CRISPRi and shRNA line derivation.....	206
Figure 5.10:	CRISPRi fails to silence <i>L1CAM</i> in i ³ Neurons.	210
Figure 5.11:	shRNAs reduce L1CAM in i ³ Neurons.....	211
Figure 5.12:	shRNA knockdown reduces LC3B-I & LC3B-II.	214
Figure 5.13:	Sequence traces of L1CAM KO clones.	216
Figure 5.14:	CRISPR-mediated KO of L1CAM.	217
Figure 5.15:	L1CAM KO reduces all cleavage products.	220
Figure 5.16:	L1CAM KO & DMSO treatment does not affect LC3B.....	221
Figure 5.17:	L1CAM KO & baf treatment does not reduce LC3B.....	222
Figure 5.18:	iPSCs lack L1CAM & exhibit no LC3B effects.	223
Figure 5.19:	L1CAM reduction does not prevent pluripotency erasure.	225
Figure 5.20:	L1CAM & GFAP immunofluorescence in iPSCs.....	228
Figure 5.21:	L1CAM & GFAP immunofluorescence in i ³ Neurons.	228
Figure 5.22:	MBP & β 3-Tubulin immunofluorescence in i ³ Neurons.	232
Figure 5.23:	NeuN & MAP2 immunofluorescence in i ³ Neurons.....	233
Figure 5.24:	pro-GABARAPs are reduced in shRNA i ³ Neurons.	235

Figure 5.25:	pro-GABARAPs appear reduced in siRNA HeLas.	237
Figure 5.26:	pro-GABARAPs appear similar in KO and WT i ³ Neurons.....	238
Figure 5.27:	RT-qPCR analysis of i ³ Neurons.	241
Figure 5.28:	LC3B is reduced in KO clone 5+shRNAs.	245
Figure 5.29:	LC3B is not rescued by L1CAM reconstitution.....	248
Figure 5.30:	LC3B is not affected by L1CAM overexpression.	250
Figure 5.31:	Hypothesized mechanism of <i>ATG8</i> transcriptional regulation.....	258

Index of Tables

Table 2.1:	gRNA target sequences.....	32
Table 2.2:	siRNA & shRNA target sequences.....	33
Table 2.3:	Plasmid construction primers.	35
Table 2.4:	Genotyping primers	41
Table 2.5:	Treatment conditions used in cell culture.....	51
Table 2.6:	Primary antibodies.	52
Table 2.7:	Secondary antibodies.	53
Table 2.8:	Primary measures & principal calculations.....	66
Table 2.9:	qPCR primers	71
Table 3.1:	Common selection genes and associated strengths & weaknesses.	77
Table 4.1:	Summary of isogenic series derivation.	140
Table 4.2:	Top exonic potential off-target loci & PCR primers.....	143
Table A.2:	Reagents.....	A-9
Table A.3.1:	HeLa & HEK 293T maintenance culture medium.....	A-13
Table A.3.2:	iPSC maintenance culture medium	A-13
Table A.3.3:	Transcription factor-mediated differentiation media	A-14
Table A.3.4:	Chemical differentiation media.....	A-15

Preface

At the time of this writing, portions of Chapters 2, 3, and 4 have been published in modified form, including shared data and adapted figures (Fernandopulle, Prestil et al., 2018).

In Chapter 3, plasmid construction was performed in collaboration with Christopher Grunseich and Michael Ward. Fig. 3.1a is used from Matheson et al. (2014), and Fig. 3.8b is adapted from the same source.

In Chapter 4, the strategy of editing *AR*, including design of gRNAs and inserts, was developed and carried out in collaboration with Xia Feng and Jill Hakim. Assessment of potential off-target mutagenesis (Table 4.2) was performed by Ronald Wang and Xia Feng. Fig. 4.14 is from unpublished work that I performed in collaboration with Ling Hao and Michael Ward.

In Chapter 5, Fig. 5.2 is adapted from Konar et al. (2018). Data identifying L1CAM as a gene of interest was obtained by Eleanna Stamatakou (Fig. 5.3), and preliminary experiments were performed in collaboration with So Yeong Cheon (Figs. 5.4 & 5.5).

Chapter 1: Introduction

1.1 Neurodegenerative diseases

1.1.1 Pathology and genetics

Cognitive deterioration and motor dysfunction have long been recognized as a common feature of diseases characterized by progressive loss of neurons, leading to their classification as neurodegenerative diseases. Among the most common and well-studied of these are Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), and Huntington's disease (HD). Each of these diseases affect different subsets of neurons, leading to their distinct clinical manifestations (Dugger & Dickson, 2017). In addition, many neurodegenerative diseases have been linked to dominant causative mutations. Since onset of neurodegeneration in these diseases does not typically occur until middle to late adulthood, the identification of such genetic causes has led to predictive genotyping and the future possibility of preventative therapies.

However, most ALS cases have no known genetic cause, and a range of genetic factors have been identified in the familial subset of cases (Tsai & Manley, 2021). Similarly mixed sporadic risk factors and familial genetic causes are also known for AD and PD (Forman et al., 2004), so these diseases can be best understood as subclasses of similar pathologies under the larger neurodegenerative disease umbrella. Several ALS disease genes have also been found to overlap with those identified in frontotemporal dementia (FTD; Liscic et al., 2020). This suggests that shared molecular mechanisms underpin different neurodegenerative diseases; defining these commonalities may lead to generalized treatment strategies.

1.1.2 Protein aggregation

It has been over a century since Alois Alzheimer first reported the discovery of dense protein deposits in the cerebral tissue of dementia patients, now known as amyloid plaques and neurofibrillary tangles, leading to the recognition of his eponymous disease (Cipriani et al., 2011). It was subsequently found that protein aggregation is a shared trait of nearly all late-onset neurodegenerative diseases (Taylor et al., 2002).

While the specific mechanisms by which aggregation-prone proteins cause disease remains an open question, enhancing the degradation of disease-associated mutant proteins by induction of autophagy has been shown to reduce neurotoxicity in animal and cell models (Menzies et al., 2015; Ravikumar et al., 2004), and many genes associated with neurodegenerative disease have been linked to autophagy regulation, as explored below. Neurons also feature complex morphology and have limited regenerative capacity, so these cells are particularly sensitive to disruptions to proteostasis. This led to the idea that disruption of protein quality control is a key aspect of neurodegeneration; aggregates may form because degradative mechanisms are impaired. Improving mutant protein clearance via autophagy induction may thus be protective against neurotoxicity and thereby delay neurodegenerative disease onset (Djajadikerta et al., 2020).

Nevertheless, these diseases have high prevalence and poor prognosis, and the only established treatments are palliative. It is perhaps this very intransigence of neurodegenerative diseases to traditional pharmacological intervention that makes these diseases good candidates for the development of new therapeutic modalities. Recently, aducanumab was approved by the FDA as the first monoclonal antibody therapy in AD, although clinical trial results did not conclusively show efficacy (Mullard, 2021). In addition, gene therapy has been explored, for example with viral vectors to modify or replace disease genes in familial ALS, HD, and spinal muscular atrophy (SMA; Sun & Roy, 2021).

1.1.3 Polyglutamine diseases

Nine different neurodegenerative diseases have each been linked to an expansion of a CAG trinucleotide repeat in a disease-associated gene, which results in an extended polyglutamine (polyQ) tract in the protein product (Orr & Zoghbi, 2007). These diseases (abbreviation; *disease-associated gene*) are: HD (*HTT*), spinal and bulbar muscular atrophy (SBMA; *AR*), dentatorubral-pallidoluyian atrophy (DRPLA; *ATN1*), and six different spinocerebellar ataxias (SCA types 1, 2, 3, 6, 7, and 17; *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, and *TBP*, respectively). SBMA is explored in further detail in Chapter 4.

Interestingly, normal functions of these genes vary, and other mutations to these genes, including knockouts, are not associated with neurodegeneration, suggesting that the CAG repeat expansion and resulting increased polyQ tract length confer a toxic gain-of-function that is necessary for neurotoxicity (Sambataro & Pennuto, 2017). Furthermore, it has been shown that long polyQ repeats predispose these proteins to aggregation, polyQ diseases often present with both cytoplasmic and nuclear inclusions, and mutant proteins in nuclei can disrupt gene regulation while escaping degradation (Li et al., 2016; Perez et al., 1998). Aggregation may be a protective mechanism to reduce the available surface area, and thus binding potential, of these mutant proteins, but the formation of such aggregates in nuclei may prevent their export and clearance.

In addition, multiple polyQ-expanded proteins have been found to directly bind to and destabilize beclin 1 via the polyQ tract (Ashkenazi et al., 2017). This study revealed that the ataxin 3 polyQ repeat functions normally to enable beclin 1 binding and that polyQ-expanded proteins may directly impair autophagy and thereby inhibit their own degradation.

1.2 Autophagy

1.2.1 Degradation pathways

Autophagy is a critical and evolutionarily conserved system of intracellular protein degradation. In general, autophagy refers to any process that acts to import substrates from the cytoplasm into an organelle containing degradative enzymes, such as protein and lipid hydrolases (Huang & Klionsky, 2002; Tanida, 2011). Unlike the ubiquitin-proteasome system (UPS), the other major degradative pathway which degrades individual peptides, autophagy is capable of degrading whole organelles and protein complexes, including oligomers of mutant proteins thought to be precursors for insoluble aggregates, contributing to its recognition as relevant to neurodegenerative diseases (Scotter et al., 2014).

The most well-understood process of autophagy is macroautophagy, which involves the collection of a portion of cytoplasm, including proteins, vesicles, and organelles (e.g., mitochondria), into a double-membrane structure known as an autophagosome. Once the autophagosome is filled and closed, it is then trafficked and fused with lysosomes to degrade the inner membrane along with the cargo inside (Dikic & Elazar, 2018). Autophagy in general is often conflated with macroautophagy in particular, but this practice diminishes the important contributions of alternative autophagy pathways, such as microautophagy, which involves direct engulfment of cargo by the lysosome, and chaperone-mediated autophagy, in which cargo is translocated into the lysosome via specific receptors on the lysosomal membrane (Cuervo & Wong, 2014; Oku & Sakai, 2018). This work therefore seeks to maintain specificity by using autophagy only in reference to the general process and using macroautophagy when this particular pathway is implicated.

Foundational work in yeast first enumerated the genes necessary for macroautophagy, later termed the autophagy-related (*ATG*) genes (Klionsky et al., 2003; Tsukada &

Ohsumi, 1993). It was subsequently found that orthologs of these genes are ubiquitous throughout the eukaryotes, and the core process of macroautophagy has remained remarkably consistent across vast evolutionary time. Perhaps the most notable difference is the locus of degradation—while plants, fungi (including yeast), and many other eukaryotes deliver autophagosomes to a large vacuole for digestion, animal species have exchanged vacuoles in favor of lysosomes. These vesicles are relatively small, but they are highly acidified by transmembrane vacuolar ATPases, which in turn activates internal acid hydrolases that digest lysosomal contents. This key difference affects the location and machinery involved in autophagosome synthesis, trafficking, and fusion; for instance, early synthesis occurs on the vacuole membrane and returns to the vacuole for degradation in yeast (Mizushima 2007), while animal cells can traffic the lysosomes themselves to distal cellular processes for degradation *in situ*, which is essential to the maintenance of neuronal axons and dendrites (Liao et al., 2019).

Interactions between the ATG proteins were largely discovered and have been characterized in the most detail in yeast, forming a complex model of the molecular mechanisms responsible for canonically regulating and enacting macroautophagy (Levine & Kroemer, 2019). In many species however, the *ATG* genes have undergone multiple duplications, leading to their independent specification of function and regulation. Differences in the number of autophagy-related genes have been noted in plants, fungi, and animals; for instance, the one yeast gene *ATG8* has two orthologs in *Drosophila*, seven in humans, and twelve in *Brassica rapa*. Thus, while the overall structure of macroautophagy appears consistent across eukaryotes, the specific machinery and regulatory mechanisms involved varies greatly.

To add yet another layer of complexity, the rate of protein degradation varies between cell types within an organism, with cells specializing in phagocytosis and rapidly proliferative cells turning over much of their proteome in hours while cells with extensive structures and long lives can take days to degrade the same proteins. The

following subsections specifically focus on the formation of the nascent autophagosome in human cells due to its relevance to Chapter 5, and a slower rate of flux in neurons compared with iPSCs and HeLa cells is also noted in this chapter.

1.2.2 Autophagosome synthesis in humans

Macroautophagy is initiated by the nucleation of the phagophore, which then elongates and is closed to form the autophagosome (Mizushima, 2007). The initial stages of phagophore formation are dependent on formation and activation of the ULK complex, which is regulated by the target of rapamycin complex 1 (TORC1) and consists of ATG13, FIP200, ATG101, and ULK1 or ULK2 (Nakatogawa, 2020). When active, the ULK complex phosphorylates numerous targets and forms a scaffold to recruit other ATG proteins (Fig. 1.1a). One such target is the PI3K complex I, which consists of VPS34, VPS15, ATG14L, NRBF2, and beclin 1 (an Atg6 homolog); when activated, this complex binds to intracellular membranes and phosphorylates the lipid phosphatidylinositol (PI) to produce PI-3-phosphate (PI3P; Funderburk et al., 2010; Itakura et al., 2008). PI3P enables recognition and binding by the four WIPI proteins, and WIPI2 specifically recruits ATG16L1 to the phagophore membrane (Fig. 1.1b; Wilson et al., 2014).

Meanwhile, ATG7 functions as an E1 ubiquitin-activating enzyme to prime the conjugation of ATG12 and the ATG8 protein family (Fig. 1.1c; Mizushima et al., 1998). ATG12 is passed from ATG7 to ATG10 and is finally conjugated to ATG5 (Tanida et al., 2001). ATG5-12 then complexes with ATG16L1 at the phagophore as a dimer (Dooley et al., 2014). With this step, all of the components are in place to discuss the ATG8s.

1.2.3 The *ATG8* gene family

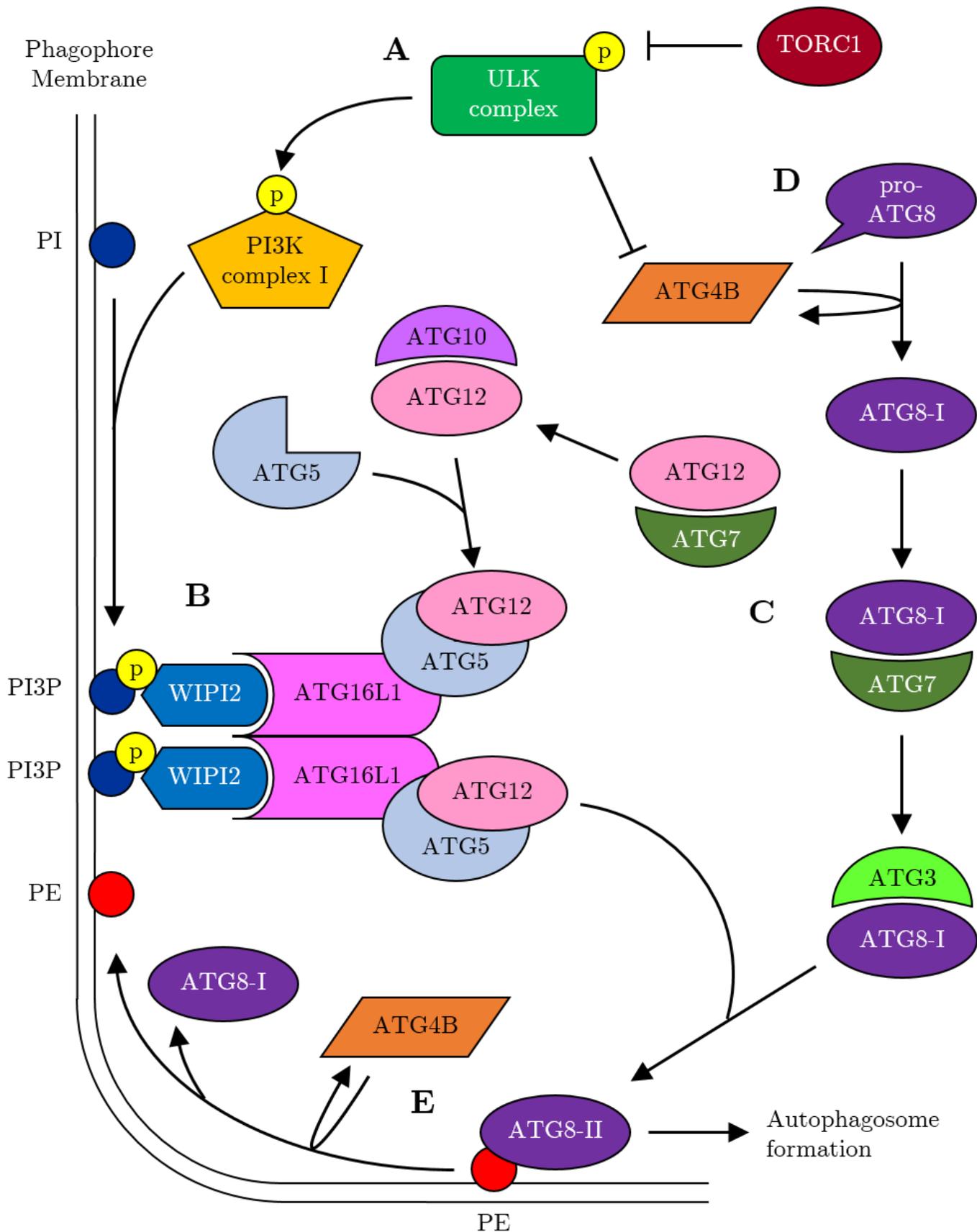
The *ATG8* family comprises seven known genes in humans, all evolutionarily related to the *Atg8* gene identified in yeast and conserved throughout the eukaryotes (Shpilka et al., 2011). These genes fall into two major subfamilies: the microtubule-associated protein 1 light chain 3s (*MAP1LC3s*, or *LC3s*), comprised of *LC3A*, *LC3B*, *LC3B2* (nearly identical to *LC3B*), and *LC3C*, and the γ -aminobutyric acid receptor-associated proteins (*GABARAPs*), comprised of *GABARAP*, *GABARAPL1*, and *GABARAPL2* (also known as *GATE-16*). Of these, the most studied protein is LC3B, which is commonly simply referred to as LC3; for the sake of clarity, this work will specifically use the name LC3B, and the name ATG8 is used to refer to the family in general.

Figure 1.1: Initial steps of canonical autophagosome synthesis.

- (a) When active, the ULK complex initiates phagophore formation by phosphorylating the PI3K complex I, which in turn phosphorylates PI to PI3P.
- (b) PI3P enables WIPI2 binding, which recruits ATG16L1 to the membrane.
- (c) ATG7 activates ATG12, which is conjugated by ATG10 to ATG5, and this complexes with ATG16L1.
- (d) pro-ATG8s are cleaved by ATG4B to become ATG8-I, which is activated by ATG7 and conjugated by ATG3 and the ATG5-12-16L1 complex to PE to form ATG8-II. This form promotes autophagosome maturation and binds a variety of adaptor proteins.
- (e) ATG4B can delipidate ATG8-II; active ULK complex inhibits ATG4B in the vicinity of the phagophore.

(continued)

Figure 1.1 (*continued*): Initial steps of canonical autophagosome synthesis.



The ATG8s are highly similar, small ubiquitin-like proteins that are associated with intracellular membranes, most notably—but not only—autophagosomes (Nieto-Torres et al., 2021a). After translation, a full-length ATG8 propeptide (known as pro-ATG8; e.g., pro-LC3B) is cleaved at the C terminus by one of the four mammalian ATG4 homologs to form ATG8-I (e.g., LC3B-I or GABARAP-I); ATG4B has been shown to be most active in general (Fig. 1.1d; Li et al., 2011). Conjugation is then activated by ATG7, which passes the ATG8-I to ATG3, which acts alongside the ATG5-12-16L1 complex to ligate the ATG8-I to the lipid phosphatidylethanolamine (PE) in the phagophore membrane as ATG8-II (e.g., LC3B-II or GABARAP-II) (Tanida et al., 2004). The ATG4s are also targets of the ULK complex, and phosphorylation by ULK1 inhibits their protease activity which likely functions to protect the ATG8-IIs from subsequent deconjugation from the membrane (Fig. 1.1e; Pengo et al., 2017).

Once conjugated, the ATG8s play an important role in phagophore elongation (LC3s) and closure (GABARAPs; Weidberg et al., 2010). Many proteins have also been found to contain LC3-interacting region (LIR) motifs, and these interactions are a key component of cargo selection, autophagosome trafficking, and lysosomal fusion (Johansen & Lamark, 2020). Mutations in several of these LIR-containing adapter proteins have been directly linked to neurodegenerative diseases, including HTT, SQSTM1, OPTN, and UBQLN2 (Deng et al., 2017; Rui et al., 2015). This underscores the importance of macroautophagy in the maintenance of neuronal health and how its dysfunction leads to disease.

While most research on the ATG8s focuses on LC3B specifically, there are notable differences in the roles of the individual proteins, their functions are not redundant, and overexpression of one ATG8 may or may not rescue the effects of a reduction of a different family member (Weidberg et al., 2010). For instance, an important role of the ATG8s is to mediate autophagosome fusion with lysosomes, which is largely mediated by the GABARAP subfamily (Nguyen et al., 2016). Specialization has also been shown for cargo selectivity, as LC3C is crucial for autophagic degradation of

invading *Salmonella* bacteria, a process termed xenophagy (von Muhlinen et al., 2012). While PE conjugation occurs on the C-terminus, the N-terminus contains two α -helices that can bind membranes, and the LIR-docking site is a part of the ubiquitin-like core (Weidberg et al., 2011). Thus, differences between the ATG8s in these domains, along with posttranslational modifications, can impart different affinities (Huang et al., 2015; Nieto-Torres et al., 2021b; Rogov et al., 2014; Wirth et al., 2019). Recent studies have suggested other modes of interaction that can further separate the binding partners of the ATG8s (Wesch et al., 2020)

In addition to double-membrane autophagosomes, ATG8s can be conjugated to single-membrane vesicles that are associated with degradation, such as phagosomes and endosomes, in much the same manner (Florey et al., 2011). Such vesicles in microglia were shown to degrade amyloid β and reduce neurotoxicity in a mouse model of AD (Heckmann et al., 2019), suggesting that impairment of the ATG8s may predispose neurodegeneration in a manner beyond impairment of macroautophagy. Furthermore, mammalian GABARAP and GABARAPL1 as well as yeast Atg8 are involved in trafficking proteins and vesicles between the endoplasmic reticulum, Golgi apparatus, and extracellular membrane (Leil, 2004; Mansuy et al., 2004). GABARAPL2 is involved in membrane fusion in the Golgi as well (Sagiv et al., 2000).

Also of note, alternative mechanisms of macroautophagy initiation have been reported which do not require ATG5, ATG7, or conjugation of ATG8s, although the ULK complex and PI3K complex I are required (Arakawa et al., 2017; Nishida et al., 2009). These mechanisms are not as well-characterized as the canonical pathways described above and are induced by DNA damage but not conventional stimuli such as TORC1 inhibition. It is therefore unlikely that these mechanisms can fully make up for a loss of canonical macroautophagy.

1.2.4 Transcriptional control of macroautophagy genes

Research on macroautophagy often focuses on protein-protein interactions that influence the rate of autophagosome formation or lysosomal degradation, but for these interactions to take place, the necessary components must be available. Understanding transcriptional regulation is therefore vital, and many pathways can influence the transcription of autophagy genes in mammalian cells (Füllgrabe et al., 2016). Most relevant for this thesis, members of the transcription factor E (TFE) family TFEB, TFE3, and TFEC, and the forkhead box class O (FOXO) family FOXO1, FOXO3, FOXO4, and FOXO6 can drive transcription of multiple core macroautophagy genes, including the *ATG8s* (Lapierre et al., 2015; Shpilka et al., 2011).

Within each family, target genes are often shared, and similar stimuli, such as starvation, can activate both transcription factor families (Di Malta et al., 2019). Orthologs of both families also cooperate in *C. elegans*, suggesting that co-activation of both pathways may be important for macroautophagy maintenance or induction more broadly (Lin et al., 2018). Cells overexpressing TFEB or constitutively active forms of FOXO1 or FOXO3 also have increased levels of multiple *ATG8s* (Mammucari et al., 2007; Sengupta et al., 2009; Settembre et al., 2011; Zhao et al., 2007).

Often pictured as opposing the TFE family, ZKSCAN3 binds to and represses transcription of several core macroautophagy genes, including the *ATG8s*, in basal conditions, but it is exported to the cytoplasm upon nutrient deprivation (Chauhan et al., 2013). Similarly, FXR (aka NR1H4) represses the transcription of many core macroautophagy genes, including *TFEB* and *ATG8s*, by preventing CREB and PPAR α from binding to these genes (Lee et al., 2014; Seok et al., 2014). Again, starvation deactivates FXR, thereby allowing CREB and PPAR α to drive transcription. While not a comprehensive survey of all known transcriptional regulators, these pathways have the widest known influence on macroautophagy genes.

1.3 Development of *in vitro* human neuronal models

1.3.1 Advantages of a human model

As mentioned above, our understanding of the molecular machinery driving macroautophagy in mammalian cells is based on research first performed in yeast. Macroautophagy is a process that is ubiquitous throughout eukaryotes, so findings in different species have largely been able to generalize through genetic orthologs. Biological research has historically used numerous model organisms, and *in vitro* cell culture has existed for over a century, but tractable human cell and tissue models are recent and ongoing developments.

Previously, a draw to animal models was the greater flexibility of experimentation and simplicity of the system. Every neuron and synapse in *C. elegans* is known, the *Drosophila* genome was essentially mapped long before high-throughput sequencing was invented, and mouse breeding programs have developed widely available knockout lines. The simple availability of antibodies and annotated genome sequences is something taken for granted in established model systems but requires massive up-front investment. The gap in experimental possibilities between human and animal models, once quite wide, has narrowed alongside the development of tools for human cell models. Such tools are often a combination of new technologies and an extension of the tools made possible by seminal research using animal models.

Performing experiments in human cells directly, rather than in an animal model, is advantageous for understanding mechanisms of human disease and how stimuli affect human biology generally. In many cases, the genes that underlie neurodegenerative diseases are not present or have different functions in common model organisms; for instance, there is no ortholog to *HTT* in *Drosophila*, and mice lack an extended polyglutamine expansion in their androgen receptor gene (both of which are reviewed in greater detail in Chapter 4). In such cases, hallmarks of disease can often be

recapitulated by transgenic overexpression of human disease proteins, but the environment that these transgenes are in is fundamentally nonhuman.

There is simply no guarantee that the molecular pathologies in these models, such as changes to gene expression and proteostasis, is an accurate representation of what occurs in a human cell, so findings in animal models must be replicated in a human model before medical applications can be pursued. The current state is that animal models are the ideal for understanding physiological complexity in a manner inaccessible for cells in a dish and with experimental tractability that human subjects do not allow; however, if the goal is to understand human biology at a cellular or molecular scale, human cells are preferable.

1.3.2 Advantages of a neuronal model

Initially, the only human-derived cells that could be cultured long-term were tumor cells, such as HeLa cells, and cells from human embryos, such as HEK 293 cells. Later developments enabled the generation of immortalized patient-specific cell lines by treating explants with radiation, carcinogens, or oncogenic viruses. However, each of these methods results in cells with abnormal and unstable karyotypes as well as irregular gene expression patterns (Adey et al., 2013; Lacoste et al., 2010; Lin et al., 2014). While such models have proven indispensable, it is more accurate to consider such cell lines as model organisms in their own right rather than truly representative of human biology.

Even with a healthy human genome, however, neurons feature transcriptomic and proteomic profiles that are highly different from other cell types, so understanding the biology of neuron-specific functions is best done in neurons directly. For instance, neurons grow complex projections and synaptic connections with other neurons for receiving and transmitting signals, the formation and maintenance of which require

extensive vesicular trafficking and metabolic regulation at great distances from the cell body and nucleus. Neurons are also postmitotic and feature extremely long protein half-lives, leading to a more stable cell state but also sensitizing neurons to dysfunctions in proteostasis and damage from reactive oxygen species relative to proliferative cells or tissues with a regenerative niche of stem cells. In addition, neurons express a variety of unique ion transport channels, neurotransmitter production and receptor machinery, and transcription factors that enable neuronal function and determine neuronal identity.

Neurodegenerative diseases also often affect different neuronal subtypes in diverse ways. For instance, mutations in *PRKN* cause PD, which principally affects the dopaminergic neurons in the substantia nigra, and several ALS-linked mutations result in motor neuron death while leaving the rest of the cortex unscathed. In both cases, the same DNA is present in every cell in the body, but the unique epigenetics of each cell type determines both its function and its susceptibility to disease. A fly is the ideal model to study fly genetics but not a disease unique to humans, and likewise a cholinergic motor neuron may be ideal to study ALS but not PD.

Fundamentally, it is possible to gain insight using an unideal or less physiologically relevant model system, and it can even be advantageous if the system chosen is more tractable or the process studied is known to be generalizable. However, the results from such systems will still need to be replicated in a more relevant model, much like how the *Atg* genes in yeast had to be mapped to human orthologs and their functions reestablished in a process greatly informed and facilitated by the knowledge gained in yeast. In the end, it is the responsibility of the researcher to determine the scope of their study and to identify the appropriate model system, and it is my opinion that a more immediately relevant system for studying a human disease is preferable and worth an investment of time and resources.

1.3.3 Human stem cells – past and future

Postmortem tissue samples have helped to understand the end state of neurodegenerative pathologies, but preventing disease onset is a much more attractive therapeutic strategy than neuronal regeneration and requires understanding the early stages of pathogenesis. Probable disease mechanisms have been identified in animal models and non-neuronal cells; however, it has been a longstanding challenge to reproduce these findings directly in live human neurons. This is in part due to the inaccessibility of primary cells, but even if human neuronal explants were widely available, the postmitotic and fragile nature of these cells greatly hinders experimentation. Starting with a cell type that can be genetically manipulated, expanded, and later converted to neuron-like cells is thus an attractive approach.

Human embryonic stem cells (hESCs) were first derived by isolating and culturing the inner cell mass of *in vitro* fertilized human blastocysts (Thomson et al., 1998). These cells were found to proliferate and maintain a stable genome in culture indefinitely as well as retain pluripotency (the ability to differentiate into the three somatic germ layers), enabling the study of human biology and early development directly. By applying knowledge from evolutionary and developmental biology, processes to direct the differentiation of cultured hESCs into somatic cell types arising from each germ layer were developed using recombinant proteins and small molecules to mimic developmental signals, and these have been gradually refined into a variety of tractable protocols (Zakrzewski et al., 2019).

Another major development occurred when it was found that the overexpression of just four transcription factors, *Oct4*, *Sox2*, *Klf4*, and *c-Myc*, was sufficient to reprogram adult mouse fibroblasts into a stem cell-like state, termed induced pluripotent stem cells (iPSCs; Takahashi & Yamanaka, 2006). This process was soon thereafter extended to human cells using either the human orthologs of the same factors (Park et al., 2008b; Takahashi et al., 2007) or the cadre of *OCT4*, *SOX2*, *NANOG*, and *LIN28* (Yu et al.,

2007). Each of these foundational studies extensively characterized the resulting human iPSCs, showing that they share key characteristics with hESCs (i.e., proliferation, genetic stability, and pluripotency).

Subsequent advances have yielded fully defined culture conditions and improved methods that remain widely used and are necessary for the production of clinical-grade cell therapies (Chen et al., 2011; Ludwig & Thomson, 2007). In addition, the development of human iPSCs enabled the derivation of human disease models via reprogramming and differentiating patient cells, revealing disease features at the cellular and molecular level (Rowe & Daley, 2019). The scalability and differentiability of iPSCs also led to their widespread use as platforms for drug discovery and genome-wide screens (Chia et al., 2010).

An apparent limitation of current iPSC-derived cell models is their lack of tissue-level and organismal biology as compared to animal models. However, this shortcoming is being addressed; for example, protocols have been developed for differentiating three-dimensional replicas of tissues *in vitro*, known as organoids (Rossi et al., 2018). These often rely on cell masses in suspension or on synthetic biomaterials to act as a scaffold for cell attachment, followed by cell self-patterning in a manner reminiscent of development (Cordie et al., 2014; Daly et al., 2021; Salick et al., 2021). In a simpler case, specific cell types can be differentiated and co-cultured, a strategy that has produced functional models of the blood-brain barrier and neuromuscular junctions, among others (Delsing et al., 2020; Lin et al., 2019).

1.4 Neuronal differentiation strategies

1.4.1 The epigenetic landscape of cell state conversions

In the case of both differentiation of stem cells to somatic cell types and reprogramming of somatic cells to an iPSC state, cells must first deactivate the epigenetic programs that perpetuate an existing cell state, a process known as erasure (Cantone & Fisher, 2013). This may be seen as an epigenetic inertia or activation energy that must be overcome in order for cells to lose an existing identity and enter a transitory intermediate state (Gulati et al., 2020; Mikkelsen et al., 2008; Wang et al., 2011). Only after this occurs can cells acquire the gene expression profiles of subsequent cell states.

Reprogramming from somatic cells to iPSCs requires overcoming considerable inertia, resulting in slow progress, low efficiency, and cell type variability (Hanna et al., 2009). Once reprogrammed, iPSCs require specific culture conditions to maintain their cell state, including a combination of growth factors in the media, proximity to other iPSCs, and sufficient space on a specialized protein substrate onto which to grow. Lacking these, iPSCs are known to readily undergo apoptosis or to lose pluripotency through spontaneous differentiation. Furthermore, extraneous signals can induce iPSCs to differentiate even in culture conditions that otherwise support pluripotency, suggesting that erasure of pluripotency and subsequent differentiation requires overcoming considerably less inertia than reprogramming (Thoma et al., 2012). To use Waddington's metaphor of an epigenetic landscape, reprogramming is pushing a ball up a hill, but once at the top, a slight nudge can send it rolling back down (Waddington, 1966).

1.4.2 Chemical differentiation

Cell types from all three germ layers have been differentiated from iPSCs, but the most relevant to neurodegenerative diseases are neurons. As mentioned, differentiation protocols initially relied on a cocktail of small molecules and recombinant developmental morphogens to drive iPSCs out of pluripotency, through a neural progenitor state, and finally into neuronal commitment (Wichterle et al., 2002). Specific chemical combinations vary, but the earliest studies found that simply withdrawing FGF- β from the culture media was sufficient to induce hESC lines to differentiate to a neuron-like state (Zhang et al., 2001). However, other lineages arose, highlighting the need to direct differentiation specifically towards a desired fate.

It was later recognized that a neuroepithelial progenitor-like cell (NPC) state could be induced via specific inhibition of activin/TGF- β and BMP4 signaling (Chambers et al., 2009; Chambers et al., 2012). Depending on the desired final neuron type, WNT/ β -catenin may also be inhibited, as in the forebrain, or activated, as in the hindbrain and spinal cord (Kirkeby & Parmar, 2012). The NPC state can be maintained and expanded by culturing with FGF- β , and interestingly, differentiation to a cortical neuron lineage can then be achieved simply by withdrawing FGF- β , suggesting that neurogenesis is the default process for NPCs (Muñoz-Sanjuán & Brivanlou, 2002).

Specific neuronal cell types can also be patterned at this stage along the rostrocaudal (head-to-tail) and dorsoventral (back-to-front) axes. Retinoic acid promotes both terminal differentiation and caudalization of developing neurons, so it is commonly used to stimulate spinal or peripheral neuron identities (Maden, 2002). Meanwhile, dorsal patterning depends on activation of WNT and BMP, while ventral identities (e.g., lower motor neurons) are induced by sonic hedgehog signaling (Petros et al., 2011). Neurotrophic factors, such as brain-derived neurotrophic factor (BDNF) and neurotrophin-3 (NT-3), may also be added to media to generally promote differentiation and maturation.

However, chemical differentiation also results in heterogeneous cell populations for the very reason that the pathways that are induced are designed to mirror normal development, and the NPC state is shared between neurons and glia *in vivo*. Subsequent induction of the neuronal cell state can be incomplete, and some cells may commit to a glial state instead. Additional refinements have helped inhibit glial outgrowth and promote neuronal maturation (e.g., with mitotic inhibitors; Patani, 2016), but these protocols often extend over several weeks and require daily attention, limiting throughput and increasing the probability of culture contamination or cell detachment.

1.4.3 Transcription factor-mediated differentiation

The discovery of iPSC reprogramming was in part inspired by the previous discovery that cell state conversions could be induced by the targeted expression of a small number of key transcription factors, which can activate epigenetic networks that greatly alter cell phenotype, termed transdifferentiation. For example, mouse fibroblasts were transdifferentiated to myoblasts by the expression of a single complementary DNA (cDNA), *MyoD* (Davis et al., 1987). In the same way, expression of key developmental genes can drive iPSC differentiation or enact direct transdifferentiation without a pluripotent intermediate (Prasad et al., 2016). While such approaches may not fully reflect development *in vivo* (e.g., by skipping over progenitor states) they are usually faster and more pure than chemical methods, have elucidated the genetic logic behind cell fate determination, and provide potential for cell therapy applications (Vierbuchen et al., 2010). Recent computational approaches have expanded the scope of transcription factor-mediated cell state conversions to include potentially hundreds of cell types (Rackham et al., 2016).

For neuronal differentiation, brief overexpression of key human neurodevelopmental genes including *NGN2* (aka *NEUROG2*) was found sufficient to trigger a cascade of additional neuronal transcription factors that cause sustained commitment to a neuronal cell fate without transitioning through an intermediate progenitor state (Pang et al., 2011; Thoma et al., 2012). Likewise, it was found that overexpression of human *NGN2*, *ISL1*, and *LHX3* (hNIL) was sufficient to drive differentiation specifically to a lower motor neuron (LMN)-like state (Hester et al., 2011; Mazzoni et al., 2013). The choice of these genes was not random; most differentiation pathways *in vivo* are highly evolutionarily conserved. For instance, it has been proposed that homologs of *ISL1* and *LHX3* were responsible for motor neuron patterning in early ancestors of Bilateria nearly 500 million years ago, even before nervous system centralization (Thor & Thomas, 2002; Vergara et al., 2017).

In each case, cells were shown to recapitulate gene expression highly similar to neurons *in vivo*, specifically cortical excitatory neurons for *NGN2* alone and cholinergic LMNs for hNIL. Cells were also shown to have electrophysiological activity and form functional synapses within two weeks of differentiation, indicating neuronal specification and maturity in a short timeframe relative to previous techniques. Moreover, both the neuron yield (i.e., the percent of cells seeded that were recoverable as neuron-like cells) and neuron subtype purity were shown to be nearly 100% with *NGN2* induction (Zhang et al., 2013b).

Subsequent developments have led to even more tractable systems, termed isogenic, integrated and inducible (i³)Neurons and i³LMNs (Wang et al., 2017). i³Neurons have a defined transgene integration at the commonly-used adeno-associated virus integration site 1 (*AAVS1*) locus; *NGN2* cDNA is placed under the transcriptional control of a third-generation tetracycline response element (TRE3G) promoter, and the third-generation reverse tetracycline transactivator (rtTA3G) is expressed constitutively by the commonly-used chimeric promoter composed of the cytomegalovirus early enhancer, chicken β -actin promoter, and rabbit β -globin splice

acceptor (CAG promoter). When a tetracycline such as doxycycline (dox) is present, rtTA3G is activated and induces expression of the TRE3G promoter (Das et al., 2016). Thus, *NGN2*-driven differentiation can be induced by treatment with dox, and control of expression level and duration is possible by adjusting dox treatment conditions.

i³LMNs have a similarly integrated, dox-inducible construct that drives the polycistronic expression of the hNIL factors separated by 2A cleavage peptides. Originally discovered in viruses, these linkers induce ribosomal skipping, which enables stoichiometric expression of multiple proteins from the same transcript (Liu et al., 2017; Ryan et al., 1991). While i³LMNs have been described with integration at the same *AAVS1* locus as above (Song et al., 2019), the i³LMNs used in this thesis feature integration in the second intron of the *CLYBL* gene, a genetic locus known to permit transgene insertion with minimal effects on cell biology and low rates of epigenetic silencing (Cerbini et al., 2015).

Although high cell type purity and electrophysiological activity have been reported for i³Neurons and i³LMNs, the protocols lack additional signals after dox induction to direct maturation other than the recommended (optional) addition of BDNF and NT-3. These generally promote maturation and synapse formation, but they do not specify neuronal subtype development (Binder & Scharfman, 2004). While most reports claim that differentiated i³Neurons resemble glutamatergic forebrain neurons with high purity (Zhang et al., 2013b), more detailed observation has shown that heterogeneity in neuronal subtypes may arise based on *NGN2* expression level and duration of dox treatment as well as subsequent culture (Lin et al., 2020). Many different types of neurons are known to exist with a shared evolutionary and developmental origin (Arendt et al., 2019); therefore, it is critical for reproducibility that transcription factor-mediated approaches be standardized (e.g., Fernandopulle, Prestil et al., 2018).

1.5 Gene editing strategies

Traditionally, incorporation of transgenic DNA into mammalian cell and animal models relied on integration of recombinant plasmids into the genome, mediated by oocyte microinjection, transfection, or viral transduction (Gordon et al., 1980; Milone & O’Doherty, 2018). None of these strategies target integration to a specific genetic locus, leading to heterogeneity in the location and number of insertions, and non-viral techniques have poor efficiency (Brinster et al., 1985). Viral integration of transgenes, particularly with lentiviruses, is a commonly-used strategy since it yields high rates of integration and expression; however, proper controls must be included (Elegheert et al., 2018).

Homologous recombination is a type of homology-directed repair (HDR), which is a set of innate mechanisms for repairing DNA breaks with high fidelity using matching DNA sequences as a template (Pardo et al., 2009). Providing such a template on the insert plasmid with homology arms can target integration to a precise genomic location (Smithies et al., 1985). However, this strategy is less efficient (often <0.1% integration), and off-target integration can occur at higher rates than on-target integration (Vasquez et al., 2001; Würtele et al., 2003). This may be due to the prevalence of non-homologous end joining (NHEJ), an innate repair mechanism for DNA double-stranded breaks that does not use a template, and is thus highly error-prone (Davis & Chen, 2013). NHEJ machinery recognizes ends of DNA strands and may incorporate nearby exogenous DNA fragments, including linearized plasmids (Smith, 2001). In hESCs, the efficiency of homologous recombination-based integration was found to be between 1×10^{-7} and 5×10^{-5} , too low for precise gene editing to be tractable (Zwaka & Thomson, 2009).

The discovery of site-directed endonucleases was recognized as a strategy both to increase efficiency of HDR integration and to induce mutations via NHEJ, but such enzymes were rare and not specific for single genomic locations (Rouet et al., 1994). It

was later discovered that proteins with custom DNA-binding domains could be engineered to recognize specific nucleotide sequences, leading to the development of zinc-finger nucleases (ZFNs; Urnov et al., 2010) and transcription factor-like effector nucleases (TALENs; (Boch et al., 2009; Hockemeyer et al., 2011). These proteins enabled the recognition of specific loci, and additional domains could be attached to cut DNA, promote transgene insertion, or influence epigenetic modifications to the DNA and histones (Gaj et al., 2013). For example, a double-stranded break could be induced in the coding sequence of a gene, leading to NHEJ-mediated frameshift mutagenesis and thereby genetic knockout (Bibikova et al., 2002). Alternatively, scarless integration of a plasmid with homology arms flanking the cut site could be achieved by HDR with higher efficiency than previous methods (Song et al., 2014).

However, these endonucleases required redesign and reconstruction for each genomic target. This limitation was not shared by systems using clustered regularly interspaced short palindromic repeats (CRISPR), originally discovered in bacteria as an antiviral system that consists of a single endonuclease, CRISPR-associated (Cas) 9, that is guided by two short RNAs to recognize and cut homologous DNA sequences (Sorek et al., 2008). CRISPR was adapted as a programmable system for use in a variety of organisms, including human cells, by fusing the two RNAs to form a single guide RNA (gRNA) and co-expressing with Cas9 (Jinek et al., 2012; Ran et al., 2013).

The use of CRISPR has been rapidly adopted throughout biology for targeted mutagenesis and transgene insertion, and mutation of the Cas9 nuclease domain produced single-strand nickase and nuclease-dead varieties enabling transcriptional activation (CRISPRa) or repression/interference (CRISPRi), imaging, and other applications (Barrangou & Doudna, 2016; Doudna & Charpentier, 2014). Genome-wide screens can now be performed using either CRISPR cutting or transcriptional modulation (Horlbeck et al., 2016; Shalem et al., 2014; Tian et al., 2019). A variety of potential medical applications have also been identified, and as of this writing the first clinical trials are underway (Ledford, 2020). In addition, the first germline gene-edited

humans, with CRISPR knockout of *CCR5*, were reported in 2018 (Cyranoski & Ledford, 2018). The ethics of germline gene editing in humans remains highly controversial (Jasanoff et al., 2019; Prestil, 2020).

1.6 Modelling disease with iPSCs

Soon after the development of iPSC technology, human iPSCs were derived from patients with a variety of genetic diseases, including ALS and HD (Dimos et al., 2008; Park et al., 2008a), but the use of these to study disease mechanisms was initially limited by several technical factors (Saha & Jaenisch, 2009). Improved differentiation techniques and culture platforms, along with better characterization of the cell types they produce, have enhanced the physiological relevance of iPSC-derived disease models (Rowe & Daley, 2019). More recently, scalable platforms for iPSC-derived neurons like i³Neurons have identified new genetic risk factors and therapeutic targets in familial and sporadic neurodegenerative diseases (Kondo et al., 2017; Tian et al., 2019; Wang et al., 2017).

All experiments require controls, but the many genetic variations between individuals confounds the direct comparison of cells with different genetic backgrounds (Torrance et al., 2001). With the development of precise gene editing, disease mutations can be introduced into healthy control iPSC lines or corrected in patient-derived iPSC lines in order to provide isogenic pairs (McTague et al., 2021). With their shared genetic background, an observed phenotype can be assessed in isogenic controls to show that a specific mutation is necessary (if correcting the mutation corrects the phenotype) and sufficient (if inducing the mutation reproduces the phenotype). Isogenic controls are therefore preferable for screening, -omics studies, and similar discovery-based assays that are critical for identifying phenotypes and formulating hypotheses, while

genetic diversity (i.e., assessing phenotypes in multiple cell lines with different genetic backgrounds) is best applied in testing these hypotheses for generalizability.

Of course, the generation of such isogenic controls for disease models requires the identification of specific disease-causing mutations. There has thus been a proliferation of isogenic models of familial monogenic neurodegenerative diseases, while polygenic and sporadic diseases rely on healthy donor iPSCs as controls (Valadez-Barba et al., 2020; Xie & Zhang, 2015). The ongoing Answer ALS project has generated and evaluated hundreds of new iPSC lines from healthy donors and ALS patients with the goal of synthesizing clinical data with multi-omics analysis, and this approach is ideally suited to address sporadic ALS and ALS/FTD (Rothstein et al., 2020). Meanwhile, the iPSC Neurodegenerative Disease Initiative ambitiously plans to model 134 different mutations linked to AD and other dementias in isogenic iPSCs, amounting to 682 planned lines, followed by similar multi-omics analyses (Ramos et al., 2021). The scale of both projects is a testament to the prospective utility of iPSC-based disease modelling.

1.7 Summary & Aims

Despite the insight that neurodegenerative diseases typically involve protein aggregation, they have proven difficult to study and unyielding to treatment. Animal and cell disease models have revealed a shared link to proteostasis, specifically autophagy, as a mechanism that ameliorates disease when enhanced and exacerbates disease when inhibited. A deeper understanding of autophagy has shown that many known neurodegenerative disease-associated genes are directly involved in regulating and enacting autophagy. Mutations in these genes may thus impair autophagy, which could reduce degradation of these mutant proteins and thereby lead to a cytotoxic feedback loop.

Patient-specific and gene-edited iPSCs that are differentiated into cell types of interest are among the most tractable and physiologically relevant disease models to date, and *in vitro* neuronal differentiation has enabled experiments directly in cells resembling human neurons for the first time. This work therefore sought to combine these capabilities in order to investigate the molecular mechanisms underpinning neurodegenerative diseases and, in the process, to evaluate and improve the methods that enable this approach.

In the course of this work, I became aware of the bottleneck of selecting successfully-edited transgenic cells from a mixed population. I therefore aimed to expand the tools available for selection of transgenic iPSCs. I also collaborated with Xia Feng and Jill Hakim to derive novel isogenic iPSC models of SBMA, and I aimed to differentiate these cells to an LMN-like state in order to compare gene expression in these lines and thereby gain new insight into disease mechanisms. Finally, I aimed to validate and characterize the action of a candidate macroautophagy-regulating gene, *L1CAM*.

Chapter 2: **Materials & Methods**

2.1 **Reagents**

Source information for all reagents, including suppliers and product numbers, is available in the Appendix (A.2). Media formulations are summarized in the Appendix (A.3).

2.2 **Plasmids**

Plasmid maps and primer sequences were designed and maintained with Benchling. Figures of constructs were prepared with SnapGene.

2.2.1 **Gifted & purchased plasmids**

Plasmids encoding TALENs that target the CLYBL locus (pZT-C13-L1 & pZT-C13-R1) were gifts from Jizhong Zou (Addgene 62196 & 62197 respectively; Cerbini et al., 2015).

pCAG-eCas9-EGFP-U6-gRNA was a gift from Jizhong Zou (Addgene 79145).

pGL3-U6-sgRNA-BFP was a gift from Xingxu Huang (Addgene 107722; Liu et al., 2018).

pBI-MCS-EGFP was a gift from Bert Vogelstein (Addgene 16542; Yu et al., 1999).

pCAG-Cre was a gift from Connie Cepko (Addgene 13775; Matsuda & Cepko, 2007).

Lentiviral packaging vectors psPAX2 and pMD2.G were gifts from Didier Trono (Addgene 12260 & 12259). pAdVantage (Promega) was also used to enhance transgene expression.

pEGFP-C1 is from Clontech (discontinued).

pHL1A-pcDNA3 (referred to in this work as L1-OE) was a gift from Vance Lemmon (Addgene 12307; Hlavin & Lemmon, 1991).

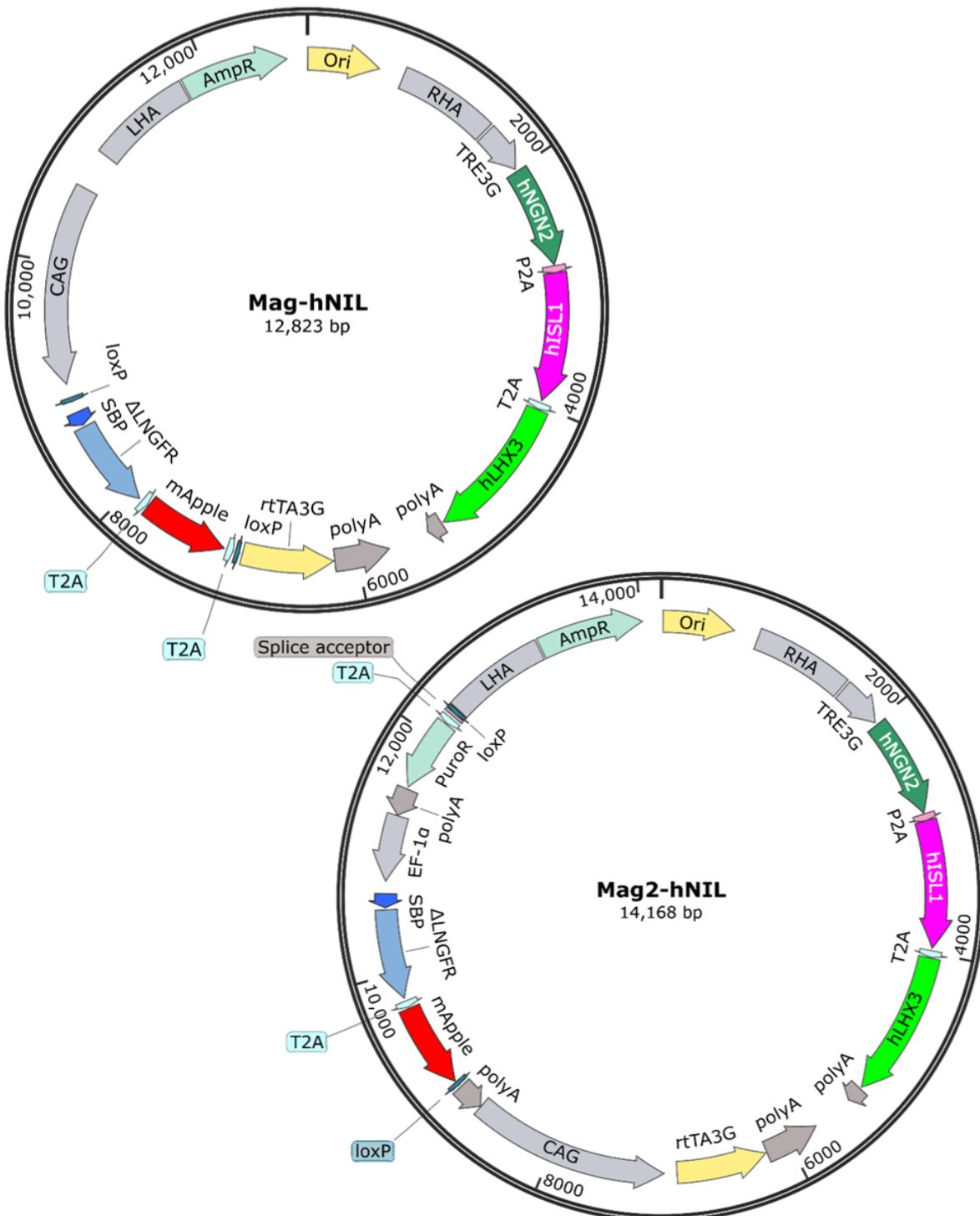
2.2.2 Lab-constructed plasmids

The Mag-hNIL and Mag2-hNIL plasmids were assembled in collaboration with Christopher Grunseich and Michael Ward. In particular, Michael Ward synthesized SBP- Δ LNGFR as a gene block (IDT DNA), and Christopher Grunseich used Gibson Assembly to insert SBP- Δ LNGFR and a PCR-cloned T2A-mApple construct into an hNIL backbone plasmid between the CAG promoter and rtTA3G, constituting Mag-hNIL (Fig. 2.1). This backbone was similar to pUCM-CLYBL-hNIL (Addgene 105841), but lacked NeoR and the EF-1 α -driven mCherry cassettes.

The plasmid pUCM-AAVS1-TO-hNGN2 (Addgene 105840) was also modified in this way to replace the EF-1 α -driven selection cassette from mCherry to SBP- Δ LNGFR-T2A-mApple, forming AAVS1-EF1-Mag-hNGN2. I digested this plasmid and pUCM-CLYBL-hNIL (Addgene 105841) with AvrII and MluI-HF in CutSmart buffer, and the resulting fragments were separated by agarose gel electrophoresis, cut at the correct band size, and purified with the QIAquick gel extraction kit. I then ligated these fragments using the DNA Ligation Kit Mighty Mix to constitute Mag2-hNIL (Fig. 2.1; Addgene 105842; Fernandopulle, Prestil et al., 2018).

Figure 2.1: Maps of Mag-hNIL and Mag2-hNIL plasmids.

Abbreviations: Origin of replication (ori); right/left homology arms (R/LHA); third-generation tetracycline response element (TRE3G); chimeric promoter composed of cytomegalovirus early enhancer, chicken β -actin promoter, and rabbit β -globin splice acceptor (CAG promoter); streptavidin binding peptide (SBP); truncated low-affinity nerve growth factor receptor, codon-optimized for amino acids 28-274 (Δ LNGFR); third-generation reverse tetracycline transactivator (rtTA3G).



The following AR-targeting gRNA plasmids and SBMA insert plasmids were designed by Xia Feng, and the plasmids were constructed by Xia Feng and Jill Hakim. Simplified maps of these plasmids are presented in Fig. 4.1b. The pCAG-eCas9-EGFP-U6-gRNA plasmid was employed to transiently express the eSpCas9(1.1) nuclease as well as an eGFP tag as an expression reporter on the same cassette, separated from the nuclease by a T2A cleavage peptide. The eSpCas9(1.1) nuclease was used because it showed greater specificity than other versions of Cas9 available at the time, reducing the likelihood of off-target effects (Slaymaker et al., 2016). On the same plasmid, expression of a CRISPR gRNA is driven by a human U6 promoter; target sequences for AR-g1 and AR-g3 were cloned into the gRNA scaffold. In the same manner, AR-g2 was ligated into pGL3-U6-sgRNA-BFP, which expresses BFP under a human PGK promoter. Target sequences for gRNAs are summarized in Table 2.1.

SBMA insert plasmids were constructed using the pCAGGS-mCherry plasmid backbone, which expresses mCherry under a CAG promoter (a gift from Phil Sharp; Addgene 41583; Gurtan et al., 2012). The genomic region between the targets of AR-g1 and AR-g2 was codon-optimized, and a mixture of mainly CAA and 5-6 CAG codons was inserted at each desired polyQ repeat length (23, 40, 54, or 68 codons). Each of these sequences was synthesized by GeneArt (Thermo), ligated with left and right homology arms of 768 bp and 802 bp upstream and downstream of the gRNA cut sites, respectively, and inserted into the plasmid.

Plasmids containing constitutive expression cassettes for eGFP-HTT exon 1-Q74 (known as eGFP-HTT74; Addgene 40262; Narain et al., 1999) and eGFP-A53T- α -synuclein (known as eGFP-A53T; Addgene 40823; Furlong et al., 2000) were previously constructed in the Rubinsztein lab.

2.2.3 RNA expression plasmids

I used the following process to introduce *L1CAM*-targeting gRNAs and shRNAs into pMK1334, a second-generation lentiviral insert that was modified to express PuroR-T2A-2×NLS-BFP from an EF-1 α promoter and a short RNA from a U6 promoter (Fig. 2.2). This backbone was a gift from Martin Kampmann (Addgene 127965; Tian et al., 2019). The use of this backbone facilitated viral preparation and subsequent genomic integration for constitutive CRISPRi and shRNA expression, and the L1-KO gRNA plasmid was transiently transfected alongside Cas9 for effective CRISPR-mediated mutagenesis. Target sequences for gRNAs were determined for CRISPRi using a published database of *in silico*-designed targets (Horlbeck et al., 2016) and for CRISPR cutting using Benchling (Table 2.1). Targets for shRNAs were identified using the Invivogen online design tool; shRNA sequences were chosen that target all known transcript variants and were screened by BLAST to ensure that no off-target loci share more than 15bp (Table 2.2).

Target sequences were inserted into a sequence that preserved restriction digest sites and 20 bp overhangs on both ends to facilitate insertion into the backbone, and both the forward and reverse complement sequences were ordered as custom oligonucleotides (Thermo; Table 2.3). Oligos for gRNAs also preserved expression of the gRNA constant region to enable gRNA binding to Cas9, while shRNA oligos were longer due to including two palindromic target sequences connected by a linker sequence, followed by seven T nucleotides as a transcription stop sequence prior to its 3' overhang (Fig. 2.2). In this way, the U6 promoter expressed the shRNAs as a hairpin without extraneous nucleotides (Gao et al., 2018b).

Table 2.1: gRNA target sequences

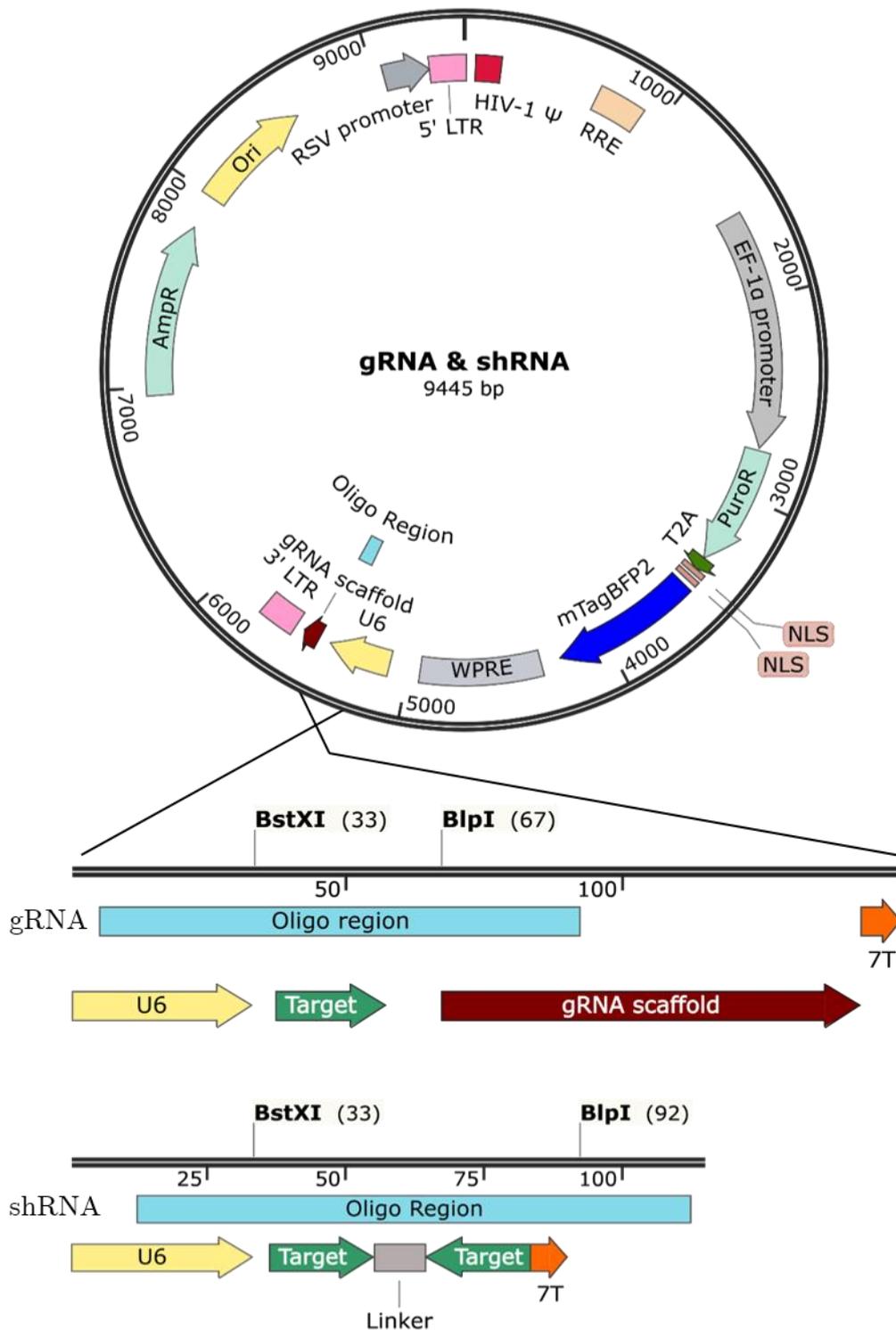
Application	Name	Target Sequence (5'-3')	Top 3 off-target genes with PAM (bp mismatches)
CRISPR cutting gRNAs	AR-g1	GGATCACTTCGCGCACGCTC	JAK3 (4), CD82 (4), KRT73 (4)
	AR-g2	GCCTGTGGGGCCTCTACGAT	KCNIP3 (4), MCF2L (4), DNAJA3 (4)
	AR-g3 (KO)	GCGCGAAGTGATCCAGAACC	DEDD (3), VPS16 (4), BRSK1 (4)
	NTg1	GGAGGCTAGGACGCAATGCA	INTS6 (4), FKBP15 (4), CHRM1 (4)
	NTg2	GTCCAGCTTATGATTGGCGC	DHX58 (4), PTPRE (4), ARL16 (4)
	L1-Screen 1	AAGCCCCAGAGCCATCTATA	HBS1L (4), ANKRD34C (4), SLC22A2 (4)
	L1-Screen 2	CGCCTGGACTGCCAAGTCCA	AGPAT2 (4), ZBTB7B (4), DPF1 (4)
	L1-KO	GCCTGCTTATCCAGATCCCCG	IFI27 (4), PLA2G4B (4), SLC5A9 (4)
CRISPRi Nontargeting gRNA	NTig	GACTCACGTAGCAGTGGAAA	GCN1L1 (3), SGCZ (4), EPN2 (4)
CRISPRi gRNAs	L1-ig1	GCATCGCAGACGCGCTCGGG	SPINT1 (4), PATZ1(4), WNT2 (4)
	L1-ig2	GGTCTGCGATGCCGATGCTG	ADAP1 (3), DDX56 (4), TMPO (4)
	L1-ig3	GGATCTGGATAAGCAGGCAG	RAPGEF1 (2), PFKM (3), DDHD1 (4)

Table 2.2: siRNA & shRNA target sequences

Application	Name	Target Sequence (5'-3')	Top 3 off-target genes (bp mismatches)
ON-TARGETplus Control Pool Non-Targeting siRNAs	NTsi	UGGUUUACAUGUCGACUAA	DEFB4A (5), RPS6KA2 (5), PHF14 (6)
		UGGUUUACAUGUUGUGUGA	PPM1B (3), AKAP6 (4), ZNF91 (5)
		UGGUUUACAUGUUUUCUGA	N4BP2L2 (3), SGCD (3), CLIP1 (4)
		UGGUUUACAUGUUUCCUA	TSPAN2 (2), BTAF1 (3), ASAH2 (4)
L1CAM siRNAs	L1-si1	CACUACACCUUUAGGGUUA	TRIM2 (5), TMEM159 (6), RECQL5 (6)
	L1-si2	GCAAGAGACAUAUCCACAA	LITAF (5), ASPM (5), OR4K2 (5)
	L1-si3	GAUACAAUGUGACGUACUG	TMEM209 (3), PRR16 (5), PCDHGB3 (6)
	L1-si4	ACACAAUGGUGACCCAAUG	PPP6R2 (3), ATRIP (5), TREX1 (5)
Nontargeting shRNA	NTsh	GACTCACGTAGCAGTGGAA	LAMA5 (6), CMPK1 (6), EPHX4 (6)
L1CAM shRNAs	L1-sh1	GGACGAACGCTTCTTCCCC	ACRBP (5), TKT (6), PHKA1 (6)
	L1-sh2	GGGTTACTGCCATAAACAA	RWDD2B (4), C8orf34 (4), ARHGEF7 (5)
	L1-sh3	GCGGATACAATGTGACGTA	PRR16 (5), MIR3165 (5), GPR139 (6)

Figure 2.2: Maps of gRNA & shRNA-expressing plasmids.

Abbreviations: Rous sarcoma virus (RSV), long terminal repeat (LTR), lentiviral packaging element (Ψ), Rev response element (RRE), Nuclear localization sequence (NLS), Woodchuck hepatitis virus posttranscriptional regulatory element (WPRE), seven T nucleotides as a U6 termination sequence (7T).



Of the backbone plasmid, 10 μg was digested using 0.5 μL BspI and 1 μL BstXI in 50 μL 1 \times NEBuffer 2.1 for 1.5 h at 37°C. Twice as much BstXI was used to account for its reduced efficiency in NEBuffer 2.1; this is the best buffer condition for this dual digest. Linearized plasmid product was then purified with Qiaquick PCR purification kit and eluted in water. Complete digest was confirmed by agarose gel electrophoresis, which showed a single band that ran slower than the undigested plasmid.

Next, each pair of oligos was annealed by combining at 1 μM in water, incubating at 95°C for 5 min, and cooling slowly to room temperature (RT). This product was then diluted to 100 μL in water, and 1.25 μL was combined with 150 ng of digested backbone in a 20 μL NEBuilder reaction. This mix was incubated at 50°C for 15 min, chilled on ice, and transformed as described below. Sanger sequencing was performed by Genewiz using the SV40 polyA Reverse Universal primer to ensure the plasmid integrated the target oligos as desired (Table 2.4).

Table 2.3: Plasmid construction primers.

Target sequences are as noted in Tables 2.1 & 2.2; RC=Reverse complement.

Application	Forward oligo (5'-3')	Reverse oligo (5'-3')
L1CAM gRNA construction	TATAAGTATCCCTTGGAGAACCACC TTGTTG(Target)GTTTAAGAGCTAAG CTGGAAACAGCATAGCAAGTT	RC of Forward
L1CAM shRNA construction	ATCCCTTGGAGAACCACCTTGTTG (Target)TTCAAGAGA(RC-Target)TTTT TTTGCTAAGCTGGAAACAGCATAGC	RC of Forward

2.2.4 Transformation

Plasmids were transformed into XL-10 Gold Ultracompetent *E. coli* cells by mixing 1 μ L plasmid with 50 μ L cells on ice for 30 min, heat shocking at 42°C for 45 sec, resting on ice for 2 min, propagating in SOC by shaking at 37°C for 1 h, and streaking on LB agar plates containing either 100 μ g/mL Ampicillin or 30 μ g/mL Kanamycin depending on the plasmid backbone. Plates were grown overnight at 37°C, and single bacterial colonies were picked and grown by shaking at 37°C overnight in 5 mL liquid LB with the same concentration of either Ampicillin or Kanamycin.

Glycerol stocks were made by mixing 25% glycerol, 25% ultrapure water, and 50% bacterial culture in a 1.5 mL tube and storing at -80°C. Bacterial cultures were then centrifuged, lysed, and purified using the Qiaprep Spin Miniprep kit, and new constructs were Sanger sequenced (Genewiz) to ensure correct assembly before further expansion. Plasmids were propagated for use in cell culture by growing overnight in 250 mL liquid LB with either 100 μ g/mL Ampicillin or 30 μ g/mL Kanamycin, centrifuged, and purified with the PureLink HiPure plasmid filter maxiprep kit using TE buffer for elution. Concentration was assessed with a Nanodrop UV spectrophotometer.

2.3 iPSC culture

Detailed protocols for many of the techniques related to iPSCs and differentiated neurons that are used in this work were published previously (Fernandopulle, Prestil et al., 2018). Methodology and reagents used are provided in brief below, and additional methods and any alterations to the protocols as published are expanded further.

2.3.1 iPSC line derivation

Fibroblasts from a healthy donor (H23) and an SBMA patient (S68) were previously acquired and reprogrammed to iPSCs via lentiviral expression of *OCT4*, *SOX2*, *KLF4*, and *MYC*. These lines were previously used in studies that performed chemical differentiation to LMN-like cells and phenotypic validation (Grunseich et al., 2014a; Grunseich et al., 2014b). The WTC11 line was previously acquired from a healthy male donor and reprogrammed to iPSCs via episomal expression of *OCT4*, *SOX2*, *KLF4*, *LIN28*, *MYCL*, and an shRNA against *p53* (Hayashi et al., 2016; Kreitzer et al., 2013; Okita et al., 2011).

The G3 line was previously derived from WTC11 by CRISPR-mediated insertion of pUCM-AAVS1-TO-hNGN2 (Addgene 105840), containing a doxycycline-inducible NGN2 cassette driven by a TRE3G promoter and a constitutively expressed rtTA3G driven by a CAG promoter, at the *AAVS1* locus. After clonal selection and validation, transient expression of Cre was used to excise selection genes, followed by subcloning. The G3-dCas9 line was previously derived by further adding a CAG promoter-driven dCas9-2A-NLS-BFP cassette to the *CLYBL* locus. Its selection genes were also excised with Cre followed by subcloning. The G3 and G3-dCas9 iPSC lines were gifts from Michael Ward.

2.3.2 Maintenance culture

Human iPSCs were maintained on matrigel-coated polystyrene multiwell plates in Essential 8 (E8) medium with complete media changes daily or every two days at low confluency (Table A.3.2). Cells were passaged at approximately 80% confluency by washing with Ca²⁺/Mg²⁺-free PBS (Thermo) and incubating in 0.5 mM EDTA in PBS at RT for 7-10 minutes, until colonies were opaque to the naked eye. Cells remained

attached during EDTA aspiration but detached in clumps by rinsing with E8. Cells were typically split at a 1:12 ratio onto fresh matrigel-coated plates. Media was supplemented with 2.5 μM Y-27632 Rho-associated protein kinase (ROCK) inhibitor (RI) following maintenance passages to promote cell survival and reduce selection pressures which can lead to genetic drift and the promotion of aberrant genotypes.

When singularization was required, cells were washed with $\text{Ca}^{2+}/\text{Mg}^{2+}$ -free PBS and incubated in accutase for 5 minutes at 37°C. Accutase and dissociated cells were then collected into a 15 mL conical tube, additional PBS was used to wash the plate and added to the tube, and the mixture was triturated to ensure full dissociation. The tube was centrifuged at 300 \times g for 5 minutes, and the supernatant was aspirated. The cell pellet was resuspended in fresh E8 supplemented with 10 μM RI (essential for the survival of singularized iPSCs). Cells were counted by taking a 10 μL sample, mixing with 10 μL Trypan blue in a microcentrifuge tube, and counting with a Countess II automated cell counter. The desired number of cells were then plated.

2.3.3 Plasmid transfection

Cells were grown on 6-well plates until at least 50% confluent. For each well, 2.5 μg total plasmid DNA was added to 250 μL Opti-MEM and incubated for 5 min at RT. 10 μL Lipofectamine Stem was added and mixed thoroughly, then incubated for 15 min at RT. The Opti-MEM transfection solution was then added dropwise to the culture medium.

2.3.4 Initial selection & enrichment

Media was replaced with fresh E8 the day after transfection. If a puromycin resistance gene was expressed on the transfected plasmid, this media may be supplemented with 1 µg/mL puromycin for 1-2 days. Alternatively, if a fluorescent protein was expressed, cells may be split with accutase, filtered, and purified by FACS. See Chapter 3 for details regarding magnetic streptavidin bead affinity-based cell sorting.

2.3.5 Clonal isolation

Following transfection and enrichment, single-cell clones were isolated by accutase singularization followed by serial dilution at low density (1,000-5,000 cells per well of a 6-well plate). These cells were plated and grown in E8 supplemented with 10 µM RI until distinct colonies were evident. After outgrowth, colonies were manually picked with a 1000 µL pipette using a Lumascope picking scope (EtaLuma) to individual wells of a 24-well plate.

For the SBMA isogenic lines (Chapter 4), since no selection marker was present, large, round, and isolated colonies were picked in order to prioritize pure clonal populations, and when fluorophores were present, pure positive colonies were marked on a fluorescent microscope and selectively picked. After growth, clones were split with EDTA; half of the cells were transferred to a 12-well plate for further growth, and half were transferred to a microcentrifuge tube for genotyping.

2.3.6 Genotyping

The cells set aside above were centrifuged, the supernatant was removed, and the pellet was resuspended in 30 μL QuickExtract (Epicentre Bio). In order to isolate PCR-ready genomic DNA (gDNA), samples were incubated at 65°C for 6 minutes followed by 98°C for 2 minutes and then stored at -20°C until use. The process of validating clones with AR polyQ edits is described in detail in Chapter 4, and the processes for validating other edits are described below. Clones that failed any of the following steps were discarded, and potential positives were tested further while live cell cultures were expanded and frozen in E8 with 10% DMSO.

PCR was performed using 1 μL of gDNA solution added to a strip tube containing 5 μL of 2 \times PCR master mix (PfuUltra for the *AR* edits, Platinum SuperFi for hNIL insertions, or Q5 for the L1CAM KO and lentiviral insertions), 0.5 μL of each of the forward and reverse primers (from a 10 μM stock; Table 2.4), and 3 μL of nuclease-free water. Reactions were run following manufacturer's instructions, with annealing temperatures dependent on primer design. PCR products were tested for amplification by adding 2 μL of 6 \times purple loading dye and performing electrophoresis through a 1% agarose gel. PCR tests were as follows:

For HDR-mediated transgene insertions (i.e., for hNIL insertions):

1. A pair of primers overlapping the insertion site and within the homology arms. A small band indicates an unedited allele, but the plasmid insert is too long to amplify effectively without lengthening the PCR extension step.
2. A primer upstream of the homology arm paired with a primer unique to the insert. Amplification indicates insertion at the desired locus.

Table 2.4: Genotyping primers

Application	Target	Forward/Reverse primers (5'-3')
Plasmid sequencing (Genewiz)	SV40pA-R	GAAATTTGTGATGCTATTGC
	CMV-Forward	CGCAAATGGGCGGTAGGCGTG
SBMA model lines	<i>AR</i> CAG repeat	TAGGGCTGGGAAGGGTCTAC
		CAGCTGAGTCATCCTCGTCC
	<i>AR</i> insert locus	GCCCTTTCCTCTTCGGTGAAGT
		CTCTACGATGTGCCTGAGGGCTG
hNIL insertion to CLYBL locus	Wildtype allele	TGACTAAACACTGTGCCCA
		AGGCAGGATGAATTGGTGGA
	Upstream & Mag- hNIL Insert	CAGACAAGTCAGTAGGGCCA
		TGCCAAGTGGGCAGTTTAC
	Upstream & Mag2-hNIL Insert	CAGACAAGTCAGTAGGGCCA
		AGGCCTTCCATCTGTTGCT
Inserts & downstream	AGTGTTGTGGAATTGCTCCAG	
	GCAAAAGGACTACCTGGATGAC	
Cre excision	Mag-hNIL selection markers	TTTGTCCCAAATCTGTGCGG
		GATGCTCAAGGGGCTTCATG
	Mag2-hNIL selection markers	TGCCAAGTGGGCAGTTTAC
		TGACTAAACACTGTGCCCA
L1CAM KO	L1CAM cut site	TAGTCACTAACGTCCTTCCG
		ATGGGGACAAGACTTGAACA
gRNA & shRNA viral insertion	Genomic integration	TATCGTTTCAGACCCACCTC
		TCTAACCAGAGAGACCCAGT

For Cre excision:

1. A pair of primers overlapping the sequence to be excised. A small band indicates excision, a large band indicates no excision, and dual bands indicate a mixed population.

For NHEJ-mediated gene knockouts:

1. A pair of primers upstream and downstream of the target site. Sanger sequencing was performed to assess mutagenesis.

For CRISPRi gRNA & shRNA lentiviral insertions:

1. A pair of primers within the insert, flanking the expressed RNA. Sanger sequencing was performed to ensure integration of the desired sequence.

Sanger sequencing was done by purifying 20 μ L PCR product with the QIAquick PCR Purification Kit and submitting a premixed tube with the same forward or reverse primer to Genewiz.

2.3.7 Karyotyping

Live cell cultures in sealed 75 cm² flasks were shipped overnight to WiCell (Madison, WI) for karyotyping. All SBMA model lines used in subsequent experiments were shown to harbor no signs of chromosomal abnormalities.

2.4 Neural differentiation of iPSCs

2.4.1 Transcription factor-mediated differentiation

Protocols used in this work are highly similar to those described in Fernandopulle, Prestil et al., 2018; however, I have reconciled the i³Neuron and i³LMN protocols as the single protocol presented here. A few steps have also been further optimized, such as replating on day 3 directly into the laminin-coating medium, but the key factors of the duration and concentration of dox treatment and the overall duration of the protocol were unchanged.

Chapter 4 followed Fernandopulle, Prestil et al., 2018 (Basic Protocols 7-8) without modification; following represents the process used in Chapter 5, and any deviations from the previously published protocol are specified (Fig. 2.1). Media formulations are as in A.3.3; Induction Medium (IM) without dox or RI was prepared in 500mL batches, and dox and RI was added fresh to aliquots as needed. Cortical neuron medium (CM) and motor neuron medium (MM) were prepared fresh in 50mL batches as needed. No frozen partially-differentiated stocks were used in this work.

iPSCs with an integrated dox-inducible hNGN2 or hNIL cassette were singularized with accutase, and 1×10^6 cells were plated to each well of a matrigel-coated 6-well plate in IM supplemented with 2 $\mu\text{g}/\text{mL}$ dox and 10 μM RI. Each day for the next two days, media was aspirated, cells were rinsed with PBS, and media was replaced with fresh IM (+dox, without RI).

One day after plating, new plates were coated overnight at 37°C with sterile-filtered 0.1 mg/mL poly-L-ornithine (PLO) dissolved in borate buffer. The following day, PLO-coated plates were washed three times with sterile water and allowed to dry completely in a culture hood. Plates were then coated overnight at 37°C with one-half culture

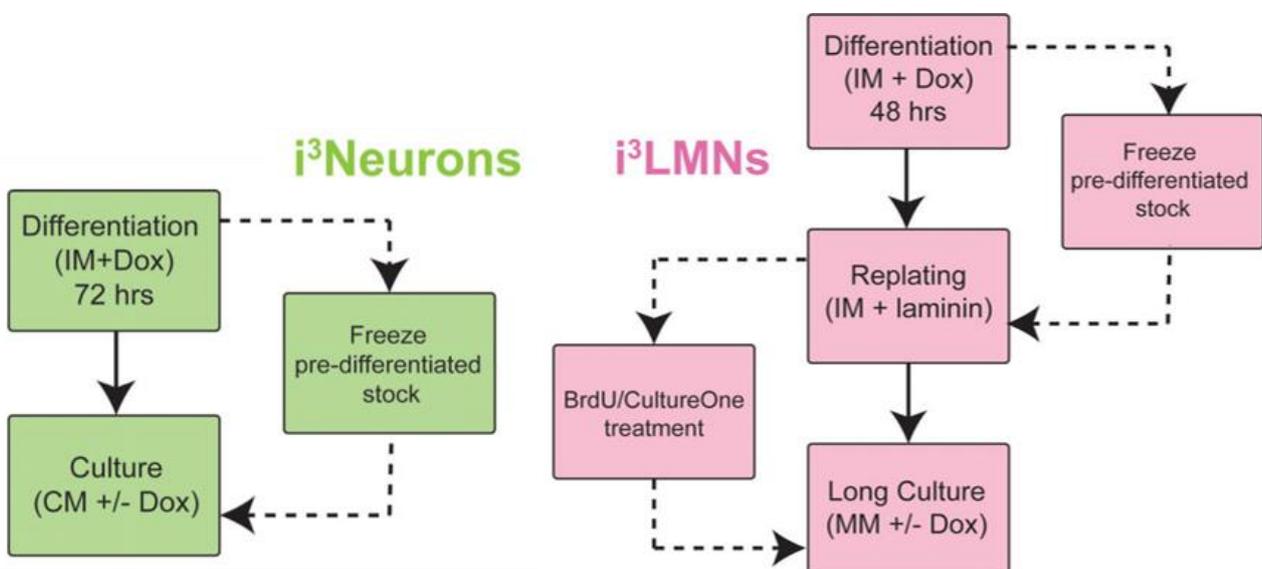
volume of CM that was prepared with double the concentration of each supplement and 5 $\mu\text{g}/\text{mL}$ laminin.

On the third day, cells were dissociated by accutase, centrifuged, and resuspended in Neurobasal medium without any supplements. Cells were counted, and additional medium was added to dilute to 2×10^6 cells/mL. One-half culture volume of cells was then seeded directly to the PLO+laminin coated plates without aspirating the $2 \times$ CM laminin-coating media. Thus, the CM was diluted to $1 \times$ concentration and the number of cells plated was 2×10^6 per well of 6-well plates, 1×10^6 per well of 12-well plates, and 2.5×10^5 per well of 8-well chamber slides.

The following day, an additional one-half culture volume of CM was added to all multiwell plates. Cells were then allowed to mature for ten days (totaling fourteen days after starting dox induction). Every three days during this period (d7 and d10), half of the CM was removed from each well by carefully pipetting from the middle of the wells in order to most effectively clear debris. Fresh CM was then added to the wells.

Figure 2.3: Timeline of transcription factor-mediated differentiation.

From Fernandopulle, Prestil et al., 2018.



On d13, another half medium change was performed on untreated cells (such as samples for RNA collection or IF), and a full medium change was performed on treated cells in order to standardize the volume (and thus drug concentration) in each well. A normal culture volume was used for treatment (as opposed to the 150% volume used throughout neuron maturation). First, spent medium was carefully aspirated, and medium with the desired treatment was slowly added dropwise to the center of each well while the plate was tilted. This helped to minimize shear forces on the edges of the well and thus prevented cells from detaching from the substrate as a sheet. Another advantage of this technique was that adding medium dropwise to the center helped to displace cell debris from the network of neurites. Once about half of the total volume was added, the rest was slowly pipetted down the side of the well.

The following day, medium was aspirated, and two PBS washes were performed in the same manner as the full medium change described above. Cells were then lysed or fixed depending on application.

2.4.2 Chemical differentiation

iPSCs were differentiated to motor neuron-like cells using a four-stage approach based on the protocol used in Hall et al., 2017. This protocol was selected because it reports >85% of cells expressing the lower motor neuron markers neurofilament and ChAT, low total expression of iPSC, progenitor, and glial markers, and a high percentage of cells demonstrating electrophysiological activity. Media formulations are as in A.3.4; basal medium (BM) was prepared in 500mL batches, and the four final media formulations were prepared fresh as needed by adding the respective small molecules. A summary of the protocol is provided in Fig. 2.2.

First, iPSCs were split with EDTA to matrigel-coated 6-well plates and grown to near-100% confluency in E8 medium. Cells were then washed with PBS, and medium was

replaced with twice the normal culture volume (4 mL/well) of neural induction medium (NIM) to mark d0 of the protocol. The NIM stage directs differentiation to an NPC state by simultaneous inhibition of activin/TGF- β with SB431542, inhibition of BMP4 with LDN 193189, and inhibition of GSK-3 β resulting in activation of WNT with CHIR99021 (Chambers et al., 2009; Chambers et al., 2012). Medium was changed daily for three days.

On d4, cells were washed with PBS and incubated with 1 mL dispase solution (1 mg/mL in DMEM/F12, sterile filtered) at 37°C for 7 minutes. Dispase was carefully aspirated so as to not detach cells, and wells were washed twice with 3 mL PBS+0.5 mM EDTA to release colonies. Cells in PBS were centrifuged at 200 \times g for 5 min, and the PBS was carefully aspirated. For each well split, two new wells were seeded of fresh matrigel-coated 6-well plates using 2 mL/well NIM supplemented with 10 μ M RI. After cells attached, an additional 2 mL/well NIM was added, and medium was changed daily for the following three days.

On d8, the medium was aspirated, cells were washed with PBS, and 4 mL/well Patterning Medium (PM) was added. The PM stage produces motor neuron progenitors by caudalizing cells with retinoic acid and ventralizing with purmorphamine, an agonist of sonic hedgehog (Li et al., 2008). Medium was changed daily for the following three days; neural rosettes were typically evident by this time. On d12, cells were split with dispase in the same manner as above and replated 1:2 on fresh matrigel-coated 6-well plates in PM. On the following three days, half-medium changes were performed daily.

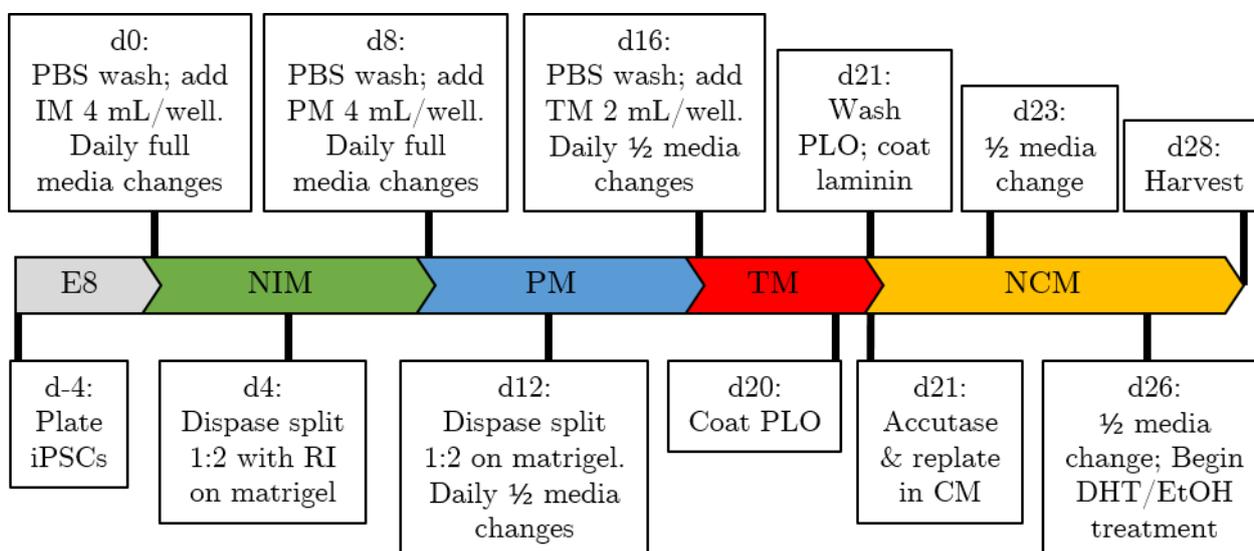
On d16, the medium was aspirated, cells were washed with PBS, and 2 mL/well Terminal Medium (TM) was added, which featured a reduced purmorphamine concentration that mimics a tapering off of sonic hedgehog signaling in order to drive cells towards the final motor neuron state; complete restriction can instead promote

interneuron differentiation (Ericson et al., 1996; Roelink et al., 1995). The following four days, half-medium changes were performed daily.

On d20, plates and slides were coated with PLO as previously described, and on d21 the plates were washed, dried, and coated with 5 $\mu\text{g}/\text{mL}$ laminin in $2\times$ NCM at 37°C for at least 1 h. NCM contained both the γ -secretase inhibitor Compound E and CultureOne supplement to promote cell cycle arrest. Compound E specifically promotes neuronal maturation by inhibiting Notch signaling (Lasky & Wu, 2005; Pierfelice et al., 2008). Cells were then dissociated with accutase, counted, and diluted in BM to 2×10^6 cells/mL. BM with cells was added directly to the $2\times$ NCM to dilute medium and seed 2×10^6 cells/well of 6-well plates, 1×10^6 cells/well of 12-well plates, and 2.5×10^5 cells/well of 8-well chamber slides.

Two days after plating, on d23, a half medium change was performed, and on d26 another half-medium change was done for untreated cells. For treated cells, a full medium change was done on d26 with NCM supplemented with the treatment condition (i.e., DHT or EtOH), as described for i^3 Neurons above. After the duration of treatment, cells were washed once with PBS and lysed or fixed depending on application.

Figure 2.4: Timeline of chemical differentiation.



2.5 HeLa & HEK 293T culture

2.5.1 Maintenance culture

HeLa and HEK 293T cells were cultured on polystyrene 10 cm dishes or multiwell plates in DMEM supplemented with 10% FBS and 1% L-glutamine (Table A.3.1). Optionally, 1% penicillin-streptomycin was also added for one week following FACS, clonal isolation, or when contamination was reported by colleagues in other cultures, but this was removed prior to experimentation. Cells were incubated at 37°C with 5% CO₂. Mycoplasma testing was performed at least once per month by PCR, and no cells tested positive.

HeLa cells were also used that had been previously undergone CRISPR-mediated knockout of ATG7 or ATG16L1 (Bento et al., 2016). HeLa cells that had been subject to the same clonal selection process were used as the controls for these experiments. Additionally, wildtype and ATG7 KO HeLa cells were used that had previously undergone stable integration of eGFP-A53T, and HeLa cells were used that had previously undergone stable integration of GFP fused to the CL1 degenon (constituting the line GFP-dg HeLa; Bence et al., 2001; Dantuma et al., 2000; Greussing et al., 2012).

Cells were passaged at approximately 90% confluency by washing with Ca²⁺/Mg²⁺-free PBS (Sigma) and incubating in Trypsin-EDTA at 37°C for 5 minutes. Culture medium was then added to deactivate the trypsin, and cells were serially diluted to approximately 1:100 for seeding.

2.5.2 Plasmid transfection

Cells were seeded on 6-well plates and allowed to attach for at least one day prior to transfection. For each well, 2.5 μg plasmid DNA was added to 250 μL Opti-MEM and incubated for 5 min at RT. To this, 7.5 μL TransIT-2020 was added and mixed thoroughly, then incubated for 15 min at RT. Culture medium was changed, and transfection solution was added dropwise directly to wells. Medium was changed again the following day.

2.5.3 siRNA transfection

Target sequences for siRNAs are summarized in Table 2.2. Lyophilized siRNAs were reconstituted at 20 μM in 1 \times siRNA Buffer (diluted from 5 \times with RNase-free water), aliquoted, and stored at -80°C until use. Cells were grown to 20-30% confluency on 6-well plates prior to transfection. Aliquots of siRNAs were thawed on ice, and for each well, two tubes were prepared with 250 μL Opti-MEM each. To one tube, 2.5 μL siRNA was added, and 4 μL Lipofectamine 2000 was added to the other. Tubes were incubated separately for 5 min at RT and then combined and mixed thoroughly. This solution was then incubated for 20 min at RT.

Cells were then washed with PBS, 500 μL Opti-MEM was added to each well, and the transfection solution was finally added dropwise. Media was mixed by gently swirling and rocking plates, and cells were incubated at 37°C for 2 h. At this time, an additional 1 mL culture medium (without Pen/Strep) was added, and cells were incubated overnight at 37°C . Media was changed the following day, and cells were harvested 72 h after transfection.

All siRNA experiments included a non-targeting siRNA pool as negative control. This included equal concentrations of four siRNAs with no on-target homology in the human

genome; however, these are not a proper scramble control since they contain different numbers of each nucleotide compared with the *L1CAM*-targeting siRNAs. In Chapter 5, it was also found that siRNA knockdown of *L1CAM* severely reduced cell attachment and growth following trypsin dissociation and impaired subsequent plasmid transfection efficiency, as compared to non-targeting siRNAs. Therefore, technical replicates in separate wells were transfected with the same volume from the same preparation of transfection solution rather than transfecting a single well followed by a split, and plasmids were transfected prior to or alongside siRNAs.

2.5.4 Lentiviral preparation

HEK 293T cells were cultured on polystyrene 6-well plates until at least 90% confluent. Each well was then transfected with 1.6 μg psPAX2, 0.6 μg pMD2.G, 0.2 μg pAdvantage, and 2.4 μg of the desired lentiviral insert plasmid prepared in Opti-MEM using 14.4 μL TransIT-2020 transfection reagent. Media was changed the following day, and expression of fluorescent proteins from the viral insert vector was observed. Two days later, media was collected into sterile 1.5 mL tubes and centrifuged at top speed for 10 min at 4°C to pellet cell debris. The supernatant was then transferred to a new tube and stored at -80°C until use.

2.5.5 Transduction of iPSCs

Viral supernatant prepared as above was thawed on ice and mixed 1:1 with E8. This mixture was added to iPSCs at approximately 50% confluency for 24 h, followed by washing cells with PBS and replacing with fresh E8. Transduced cells were then cultured for at least four days to ensure that no viral particles or transiently expressed proteins remained.

2.6 Drug treatments

All drug treatments were prepared in fresh medium and conducted alongside controls that were treated with an equal volume of solvent only (Table 2.5). Treatment with DHT was performed for only the final 48 hours of culture so as to avoid potential extraneous effects of androgen signaling during differentiation that could influence cell state and maturation. A concentration of 10 nM DHT has been shown to be effective at inducing AR activation and resultant gene expression within 24 h in multiple cell types (Cai et al., 2011).

Treatment with Bafilomycin A1 (baf) at 400nM for four hours was sufficient to induce a large increase in LC3B-II in HeLa cells and iPSCs, indicative of a blockage of lysosomal degradation. In contrast, differentiated i³Neurons were found to require at least 24 h treatment to induce a noticeable effect, although LC3B-I remained the predominant form (see Chapter 5). MG-132 treatment was performed at 10 μ M for 6 h in GFP-dg HeLa cells.

Table 2.5: Treatment conditions used in cell culture

Drug	Purpose	Solvent	Concentration	Duration
5 α -dihydro-testosterone (DHT)	A derivative of testosterone and potent agonist of AR	EtOH	10 nM	48 h
Bafilomycin A1 (Baf)	Inhibitor of lysosomal acidification	DMSO	400 nM	4 h (HeLa/iPSC); 24 h (i ³ Neuron)
MG-132	Proteasome inhibitor	DMSO	10 μ M	6 h

2.7 Antibodies

The following antibodies were used in the indicated application (Tables 2.6 & 2.7)

Table 2.6: Primary antibodies.

Poly=polyclonal, CST=Cell Signaling Technology, SCBT=Santa Cruz Biotechnology, WB=western blot, IF=immunofluorescence.

Antigen	Clone	Host Species	Source	Product Number	Use	Dilution
AR	D6F11	Rabbit	CST	5153	WB	1:1000
					IF	1:250
AR	Poly	Rabbit	SCBT	sc-13062	WB	1:2000
ATG16L1	D6D5	Rabbit	CST	8089	WB	1:1000
α -Tubulin	Poly	Rabbit	CST	2144	WB	1:2000
α -Tubulin	DM1A	Mouse	Sigma	T9026	WB	1:10000
β -actin	Poly	Rabbit	Sigma	A2066	WB	1:10000
β 3-Tubulin	D71G9	Rabbit	CST	5568	WB	1:10000
					IF	1:500
ChAT	CL3173	Mouse	Atlas	AMAb91130	WB	1:1000
					IF	1:200
GABARAP	Poly	Rabbit	Abgent	AP1821a	WB	1:1000
GABARAPL1	Poly	Rabbit	Protein Tech	11010-1-AP	WB	1:1000
GABARAPL2	Poly	Rabbit	MBL	PM038	WB	1:1000
GAPDH	6C5	Mouse	Abcam	ab8245	WB	1:10000
GFAP	D1F4Q	Rabbit	CST	12389	WB	1:1000
					IF	1:200
GFP	Poly	Rabbit	abcam	ab6556	WB	1:1000
Hb9 (MNX1)	F-5	Mouse	SCBT	sc-575769	WB	1:500
					IF	1:250
L1CAM	5G3	Mouse	SCBT	sc-33686	WB	1:1000
					IF	1:250
L1CAM	D-5	Mouse	SCBT	sc-374046	WB	1:1000
L1CAM	74-5H7	Mouse	SCBT	sc-59868	WB	1:1000
Lamin B	Poly	Goat	SCBT	sc-6216	WB	1:1000
					IF	1:250

Table 2.6 (*continued*): Primary antibodies.

Antigen	Clone	Host Species	Source	Product Number	Use	Dilution
LC3B	EPR18709	Rabbit	abcam	ab192890	WB	1:3000
					IF	1:400
MAP2	D5G1	Rabbit	CST	8707	WB	1:1000
					IF	1:250
MBP	E9P7U	Mouse	CST	83683	WB	1:1000
Nanog	D73G4	Rabbit	CST	4903	WB	1:1000
					IF	1:200
NeuN	1B7	Mouse	Novus	NBP192693	IF	1:1000
NGN2	D2R3D	Rabbit	CST	13144	WB	1:1000
					IF	1:250

Table 2.7: Secondary antibodies.

AF=Alexa Fluor, HRP=Horseradish peroxidase.

Host Species	Target Species	Conjugated	Source	Product Number	Use	Dilution
Goat	Rabbit	AF488	Thermo	A11008	IF	1:500-1000
		AF555	Thermo	A21245	IF	1:500-1000
		DyLight 680	Thermo	355568	WB	1:5000-20000
		DyLight 800	Thermo	SA535571	WB	1:5000-20000
	Mouse	AF488	Thermo	A11001	IF	1:500-1000
		AF555	Thermo	A21422	IF	1:500-1000
		AF594	Thermo	A11005	IF	1:500-1000
		DyLight 680	Thermo	35518	WB	1:5000-20000
		DyLight 800	Thermo	SA535521	WB	1:5000-20000
	Donkey	Goat	AF488	Thermo	A32814	IF
Rabbit		AF555	Thermo	A31527	IF	1:500-1000
		HRP	GE	NA934V	WB	1:5000
Sheep	Mouse	HRP	GE	NA931V	WB	1:5000
Rabbit	Goat	HRP	Thermo	611620	WB	1:5000

2.8 Western blotting

2.8.1 Sample preparation

Lysis buffers were prepared using the following recipes:

RIPA: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.5 mM EGTA, 1% v/v Triton X-100, 0.1% w/v sodium deoxycholate, 0.1% w/v SDS, 140 mM NaCl, and 1× cOmplete protease inhibitor cocktail in UltraPure water.

4× Laemmli: 30% v/v Glycerol, 4.6% w/v SDS, 250 mM Tris (pH 6.8), 10% v/v β-mercaptoethanol, and 0.02% w/v bromophenol blue in UltraPure water.

RIPA was stored at 4°C and 4× Laemmli was stored at -20°C. Buffers were mixed 3:1 immediately prior to use.

Cells were washed twice with PBS and lysis buffer was added at 250 μL/well for 12-well plates and 400 μL/well for 6-well plates. Plates were then scraped and pipetted to 1.5 mL tubes to collect total protein. Samples were heated at 95°C for 5 minutes and stored at -80°C until use. Samples were thawed at RT and mixed well prior to PAGE.

2.8.2 Polyacrylamide gel electrophoresis

Blots in Chapter 4 were performed using pre-cast NuPAGE Bis-Tris gels and transferred to an iBlot2 nitrocellulose membrane. Blots in Chapter 5 used fresh aqueous polyacrylamide gels at various percentages that were cast using the Bio-Rad Mini-PROTEAN Tetra Cell Casting Module in 1.5 mm, 15-well format (plates: 1653312 & 1653308, combs: 1653366, casting stand: 1658051). Stacking gels were prepared with 5% w/v acrylamide, 125 mM Tris (pH 6.8), 0.1% w/v SDS, 0.1% w/v APS, and 0.01%

v/v TEMED. Running gels were prepared with 8-15% w/v acrylamide, 375 mM Tris (pH 8.8), 0.1% w/v SDS, 0.1% w/v APS, and 0.005% v/v TEMED.

After casting, gels were either used immediately or wrapped in wet paper towels and saran wrap and stored overnight at 4°C. Gels were loaded into running modules (Bio-Rad, buffer tank: 1658040, running modules: 1658037 & 1658038) and immersed in running buffer consisting of 25 mM Tris, 192 mM glycine, and 0.1% w/v SDS in RO water. Of the sample, 10 µL was added to each lane, and 3 µL of PageRuler Plus pre-stained ladder was added to the edge wells. Electrophoresis proceeded at 80 V for 20 min while samples were in the stacking gel, followed by 120 V until the dye front approached the bottom of the gel (approximately 80 min).

2.8.3 Transferring, marking, & imaging

Gels were transferred using Immobilon-FL PVDF membranes that were cut to size, immersed in MeOH, adhered to gels, and any air bubbles removed. Gels and membranes were sandwiched in filter paper and sponges, pressed into a loading cassette, and immersed in a tank filled with transfer buffer consisting of 25 mM Tris, 192 mM glycine, and 20% v/v MeOH in RO water. Electrophoresis proceeded at 80 V for 90 min.

After transfer, membranes were dried on filter paper, cut into strips, and blocked in 3% w/v BSA in PBS for 1 h at RT on a shaker. Membranes were then incubated with primary antibodies diluted in 3% w/v BSA and 0.1% v/v Tween-20 in PBS (PBS-T) overnight at 4°C on a shaker. The following day, primary antibody solutions were removed and blots were washed with PBS-T at least three times for at least 5 min each. Secondary antibodies conjugated to horseradish peroxidase (HRP), DyLight 680, or DyLight 800 were then diluted in 3% w/v BSA in PBS-T, blots were incubated in this solution for 1 h at RT on a shaker, and an additional three washes were performed.

Finally, blots were washed one additional time with PBS. For HRP secondaries, enhanced chemiluminescence (ECL) substrate was added and blots were imaged with a ChemiDoc (BioRad). For fluorophore secondaries, blots were imaged by a LI-COR Odyssey CLx scanner.

2.8.4 Quantification & analysis

For both chemiluminescent and fluorescent signals, band intensity was quantified via LI-COR Image Studio Lite v5.2. The same size rectangle was used to measure a band for all samples on a membrane, and the background was subtracted by calculating the median surrounding each shape. Quantitative data were exported to Microsoft Excel for analysis. First, each signal was divided by the associated loading control (i.e., GAPDH, α -Tubulin, or β -actin), and then data were normalized by dividing each value by the mean of the associated control condition. The mean values of all experimental replicates were combined, and significance was assessed using a one-sample, paired two-tailed Student's t-test with a threshold of significance set as $p < 0.05$. Unless otherwise specified, all quantitative data represents at least three independent experiments, and qualitative blot figures are from one representative experiment.

2.9 Microscopy

2.9.1 Immunocytochemistry

Because differentiated neurons are easily detached from the culture substrate, the protocol for immunofluorescence differs slightly between these neurons and the more strongly adherent iPSCs and HeLas. In general, neuron staining is done more slowly and carefully, such as by using a micropipette rather than a vacuum aspirator and by performing fewer washes with longer incubation times. Because of these differences, protocols for each are separated below. Unless specified, all replicates of each staining condition, including DAPI, that are presented in this work were stained, imaged, and presented with identical settings between all cell lines and cell states.

2.9.1.1 Differentiated neurons

Cells were grown on μ -slide 8-well chamber slides and fixed by adding 100 μ L of 8% w/v PFA in PBS directly into the culture medium for 15 min at RT. Cells were then washed once with PBS for 15 min and permeabilized with 0.2% v/v Triton X-100 in PBS for 15 min at RT. Cells were again washed once with PBS for 15 min and then incubated in blocking buffer (1% w/v BSA in PBS) for 1 h. Meanwhile, primary antibodies were diluted in blocking buffer, and cells were incubated with these primaries overnight at 4°C. The following day, cells were washed with PBS once and washed again with blocking buffer for 15 min each. Secondary antibodies were diluted in blocking buffer and added to slides for 1 h at RT. Slides were kept in the dark and/or wrapped in foil from this point forward. Cells were then washed once with PBS, incubated with 1 μ g/mL DAPI in PBS for 15 min at RT to stain DNA, and washed once more with PBS. Cells were stored in PBS at 4°C until imaging.

2.9.1.2 iPSCs & HeLa cells

Cells were grown on μ -slide 8 well chamber slides and fixed with 4% w/v PFA in PBS for 15 min at RT. Cells were then washed three times with PBS and permeabilized with 0.2% v/v Triton X-100 in PBS for 10 min at RT. Cells were washed twice with PBS and then incubated in blocking buffer (1% w/v BSA in PBS) for 1 h. Meanwhile, primary antibodies were diluted in blocking buffer, and cells were incubated with these primaries overnight at 4°C. The following day, cells were washed with PBS twice and washed again with blocking buffer for 5 min each. Secondary antibodies were diluted in blocking buffer and added to slides for 1 h at RT. Slides were kept in the dark and/or wrapped in foil from this point forward. Cells were then washed once with PBS, incubated with 1 μ g/mL DAPI in PBS for 15 min at RT to stain DNA, and washed once more with PBS. Cells were stored in PBS at 4°C until imaging.

2.9.2 Imaging

Live imaging of bright-field (BF) and fluorescence was performed with a Lumascope picking scope (EtaLuma) or an Evos XL Core (Thermo). Slides were imaged using a Zeiss LSM710 or LSM880 confocal microscope using exposure settings standardized across all samples in each staining condition. Unless otherwise indicated, all images were collected with a Plan-Apochromat 63 \times /1.4 oil immersion objective (Zeiss). Negative controls were also imaged that were incubated with secondary antibodies but lacked primary antibody treatment, and any experimental condition that did not feature cellular staining more intense than this threshold were considered undetectable. Images were processed with Zen (Zeiss) and ImageJ.

2.9.3 Cell state determination

To assess the differentiation state of cells in Chapters 4 and 5, a number of marker genes and cellular features were identified, and immunofluorescence images were obtained. By imaging both iPSCs and i³LMNs/i³Neurons, changes to gene expression and cell morphology can be more easily assessed, and the different cell states serve variably as controls for each other.

Pluripotency was primarily assessed with NANOG, a transcription factor necessary for pluripotency induction and maintenance (Heurtier et al., 2019). All iPSCs were thus expected to feature high levels of nuclear expression of NANOG, while differentiation protocols feature the removal of growth factors that are necessary for pluripotency maintenance, so it is expected that differentiated cells should lose NANOG expression.

Glial fibrillary acidic protein (GFAP) was used as a marker of neural progenitors and of astrocytes to assess incomplete and incorrect differentiation (Ahmed et al., 2012). Myelin basic protein (MBP) was used as a marker of oligodendrocytes (Barbarese et al., 1988). Differentiated cells that properly acquire a neuronal cell state should demonstrate very little expression of either.

Positive neuronal markers include β 3-Tubulin (also known as Tuj1) which, while not strictly neuron-specific, is primarily expressed in neurons and is used widely as a neuronal marker (Lee et al., 1990). The neuronal nuclear marker NeuN (aka RBFOX3) was also used, as was MAP2, which is principally expressed in neurons and specifically localizes to dendrites (Gusel'nikova & Korzhevskiy, 2015; Soltani et al., 2005).

Beyond a general neuronal identity, LMNs can be specifically identified by expression of Hb9 (aka MNX1), a transcription factor important for LMN development (Arber et al., 1999), and choline acetyltransferase (ChAT), which is indicative of acetylcholine production for cholinergic neurotransmission (Strauss et al., 1991).

It is also worth noting that changes to the characteristics of cell nuclei are indicative of a loss of pluripotency, such as a reduced size, brighter DAPI signal suggestive of tighter packing of DNA, a less circular morphology, and a lack of cells undergoing mitosis (Gurdon & Graham, 1967; Molugu et al., 2020). Neurons also customarily feature small cell bodies and long, thin projections that form a dense network when cultured. While not used as a positive determining factor on its own, morphological features of cells and nuclei worked alongside staining for marker genes to increase confidence in cell state determination, and a lack neuronal morphology was used as exclusionary criteria to determine that cells are not neuron-like.

Thus, cells that lacked staining for iPSC and glial markers (NANOG, GFAP, MBP), positively stained for neuronal markers (β 3-Tub, NeuN, and/or MAP2), and featured neuronal morphology were considered to be neuron-like, and those that additionally stained for Hb9 and/or ChAT were considered LMN-like.

2.10 Flow cytometry

In Chapters 3-4, I performed endpoint flow cytometry experiments at the National Human Genome Research Institute (NHGRI) flow cytometry core using a CytoFLEX S (Beckman Coulter). For FACS, I prepared cells, and Stacie Anderson and Martha Kirby sorted cells using a FACSAria (BD Biosciences) at the same facility.

In Chapter 5, I performed endpoint flow cytometry experiments at the Cambridge Institute for Medical Research (CIMR) flow cytometry core using an Attune NxT (Thermo). For FACS, I prepared cells, and Reiner Schulte, Chiara Cossetti, and Gabriela Grondys-Kotarba sorted cells using an Influx (BD Biosciences) at the same facility.

Data were analyzed and images were prepared with FlowJo and FCS Express.

2.11 Optimization of SBP- Δ LNGFR

The following establishes abbreviations used throughout this section and Chapter 3 and describes the experimental and analytical procedures used for the establishment, optimization, and assessment of the enrichment and depletion protocols. Detailed methods of the optimization experiments are also provided below, the results of which are presented in Chapter 3.3.

2.11.1 Definitions

For simplicity, the following abbreviations are used throughout 2.11 and in Chapter 3; respective symbols are in *italics*:

or *%*: the number of cells or the percent of cells, respectively, of a given type.

Initial, *Negative*, or *Positive*: the population or fraction of cells measured. *Initial* is an aliquot of the population set aside after bead incubation but prior to selection, *Negative* is the supernatant fraction, and *Positive* is the resuspended final magnetized pellet fraction.

positive, *negative*, or *total*: post-selection assessment by flow cytometry of mApple+, mApple-, or all cells in a given fraction, respectively.

For example, *#Ip* are the number of mApple+ cells in the initial population, and the *%Nn* is the percent of negative cells in the negative fraction (equal to $\#Nn/\#Nt$).

2.11.2 Experimental design

Pure clones of Mag-hNIL cells with a wildtype genetic background were isolated and validated as described in Chapter 3.2.1, and populations of these cells and the parental line without Mag-hNIL were mixed at various ratios in order to isolate and optimize individual variables in the enrichment protocol as well as to develop a protocol for depletion of positive cells from the negative fraction. An aliquot of the initial population was set aside in a round-bottom 96-well plate after bead incubation but before selection, and selected positive and negative fractions were deposited into individual wells of the same plate following selection. Flow cytometry was then performed on each well, with equal sample volumes and flow rates in order to account for differences in cell density and enable direct comparisons between negative and positive fractions.

Flow cytometry data was analyzed by first removing suspected debris by gating for cells with moderate to high forward and side scatter. Interestingly, after the addition of beads, it was noted that a distinct population of small objects with high side scatter formed in samples, and the side scatter of a subset of cells increased dramatically (Fig. 2.5a). The small population was present as the only population in a sample consisting of only beads in rinsing buffer and was not found in samples of cells or rinsing buffer without beads added, so this small population was attributed to unbound beads. The proportion of side scatter-shifted cells also increased substantially in the positive fraction and decreased in the negative fraction, and these shifted cells were preferentially mApple+ (Fig. 2.5b). This shift was thus attributed to cells that were bound to beads.

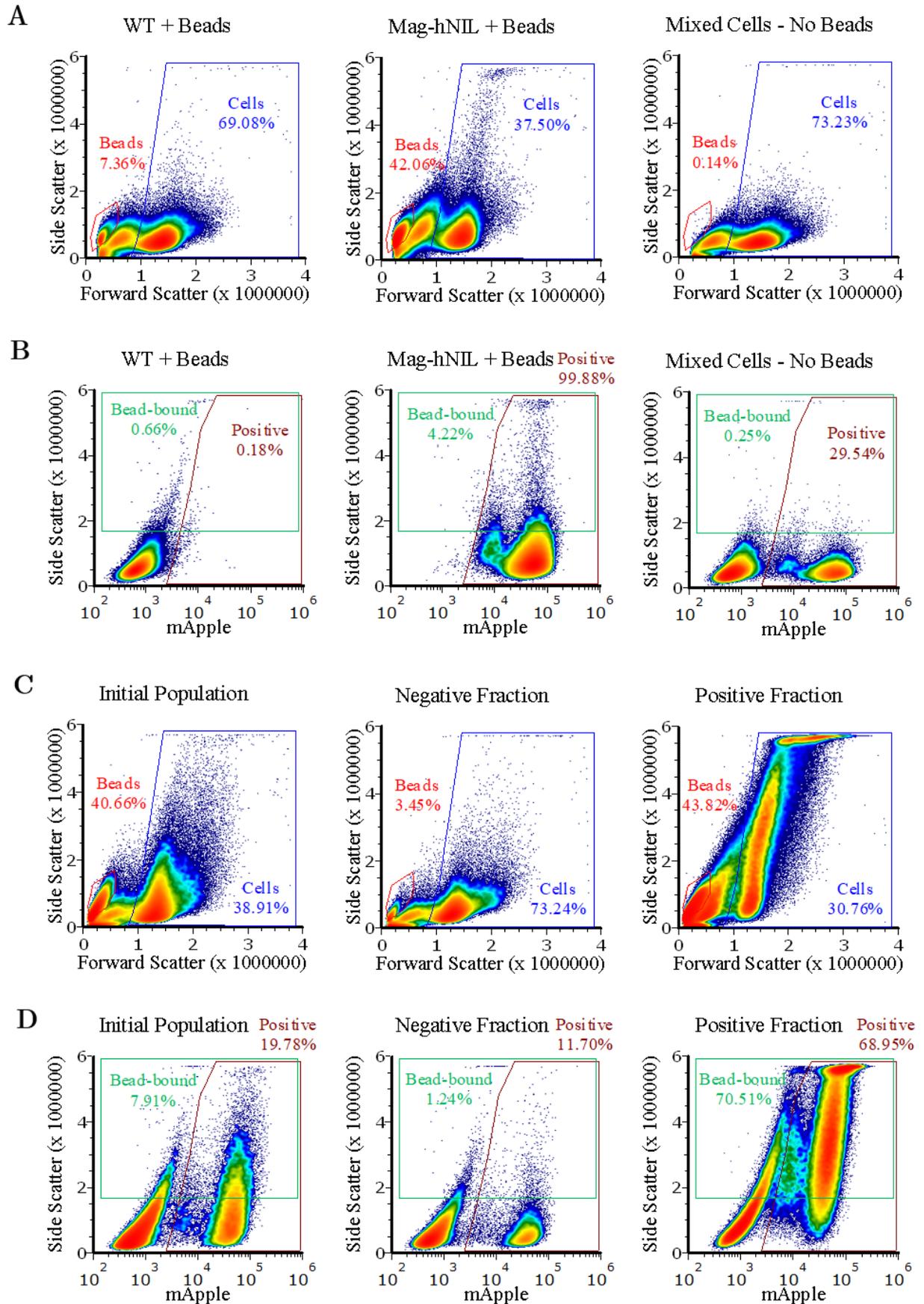
While beads interacting with negative cells increased fluorescence in the same channel as mApple, this was accounted for by gating by both mApple fluorescence and side scatter to prevent spectral overlap, and the gating shown and used for subsequent experiments yielded >99% accuracy against pure positive and negative populations (Fig. 2.5b-d). Because these analyses intended to assess the practical efficiency of enrichment and depletion, data are shown for all cells as identified by size in flow cytometry without regard to potential multiplicity.

Figure 2.5: Parameters for flow cytometry analysis.

- (a) WT cells or Mag-hNIL cells were incubated as pure populations with beads or mixed together without beads. Visualizing forward scatter and side scatter revealed a small population that was only present in samples with beads; this was also present for beads alone and was thus attributed to unbound beads. A subset of cells with beads also showed a side-scatter shift that was not present in any cells without beads, and more of the positive cells exhibited this shift. This shift was thus attributed to bead-bound cells.
- (b) The same samples as in (a) visualized with mApple fluorescence on the horizontal axis were gated to distinguish mApple+ and bead-bound cells based on mApple fluorescence and side scatter. A population with intermediate mApple intensity was noted in the Mag-hNIL clone; these were treated as positive cells.
- (c) An example of data from an enrichment trial with an aliquot of the initial population and equal volumes of the negative and positive fraction. The distribution of events in total flow cytometry data by forward and side scatter showed that both unbound beads and bead-bound cells were preferentially selected to the positive fraction.
- (d) Cells that were mApple+ or bound to beads were both preferentially sorted into the positive fraction.

(continued)

Figure 2.5 (*continued*): Parameters for flow cytometry analysis.



2.11.3 Primary measures

With these parameters, two key values were measured from each fraction: the total cell count $\#t$ and the positive cell count $\#p$. For certain analyses described below, the number of bead-bound cells was also measured using gates for high side-scatter regardless of whether the cells expressed mApple. It was then possible to identify several key characteristics of these populations and to relate them to established methods for assessing the quality of selection processes and diagnostic tests (Fawcett, 2006; Tharwat, 2021; Whitman et al., 2020). Table 2.8 outlines the following calculations, and alongside each is given a frequentist interpretation for this context. A Bayesian interpretation is explored later.

The prevalence may be measured directly as the percent of positive cells in the initial population ($\%Ip$) or calculated after selection by combining the counts of the selected fractions. For depletion, this is straightforwardly: $(\#Pp + \#Np) / (\#Pt + \#Nt)$ since both fractions are approximately 1 mL in volume, and equal volumes were sampled by flow cytometry. For enrichment however, the combined negative fraction amounts to 1 mL/wash (3 mL total), while the positive fraction is 1 mL, so the observed cell counts are artificially skewed. A corrective volume factor $v = 3$ was applied by multiplying the cell counts in the negative fraction by the number of washes, amounting to: $(\#Pp + 3 \times \#Np) / (\#Pt + 3 \times \#Nt)$. Post-selection prevalence was highly similar and correlated with $\%Ip$ (enrichment: mean squared error = 0.26%, maximum squared error = 0.86%, Pearson's R = 0.994, n = 21; depletion: mean squared error = 0.47%, maximum squared error = 1.4%, Pearson's R = 0.989, n = 17).

Furthermore, the positive predictive value (or precision) was just the percent of true positive cells in the positive fraction ($\%Pp$), and the false omission rate was the percent of positive cells in the negative fraction ($\%Np$). The false discovery rate was therefore $(1 - \%Pp) = \%Pn$, or the percent of cells in the positive fraction that did not express mApple, and the negative predictive value was similarly $\%Nn = (1 - \%Np)$.

Table 2.8: Primary measures & principal calculations.

The number of positive cells and total number of cells counted was applied to a confusion matrix to calculate measures of test efficacy. Prevalence, sensitivity, specificity, and accuracy require a correction for fraction volume in enrichment by multiplying observed cell counts by a volume factor (v) equal to the number of washes mixed as the negative fraction. Colors are used to group related measures.

		Bead Selection			Derived Characteristics
		Initial Population $\#It$	Positive Fraction $\#Pt$	Negative Fraction $\#Nt$	
Flow Cytometry	mApple Positive	$\#Ip$	$\#Pp$	$\#Np$	Sensitivity: $se = \frac{\#Pp}{\#Pp + v\#Np}$
	mApple Negative	$\#In$ $= \#It - \#Ip$	$\#Pn$ $= \#Pt - \#Pp$	$\#Nn$ $= \#Nt - \#Np$	Specificity: $sp = \frac{v\#Nn}{\#Pn + v\#Nn}$
Fraction Characteristics		Prevalence: $\%Ip$ $\approx \frac{\#Pp + v\#Np}{\#Pt + v\#Nt}$	Positive predictive value (precision): $\%Pp = \frac{\#Pp}{\#Pt}$	False Omission Rate: $\%Np = \frac{\#Np}{\#Nt}$	Accuracy: $acc = \frac{\#Pp + v\#Nn}{\#Pt + v\#Nt}$
		Recovery: $\frac{\#Pt}{\#Pt + v\#Nt}$	False Discovery Rate: $\%Pn = \frac{\#Pn}{\#Pt}$	Negative predictive value: $\%Nn = \frac{\#Nn}{\#Nt}$	Positive Likelihood Ratio: $\frac{se}{(1 - sp)}$
		Relative Enrichment: $\frac{\%Pp - \%Ip}{\%Ip}$	Total Enrichment: $\%Pp - \%Ip$	Total Depletion: $\%Ip - \%Np$	Relative Depletion: $1 - \frac{\%Ip - \%Np}{\%Ip}$

From these, the total enrichment could be derived, defined as the difference between $\%Pp$ and $\%Ip$, followed by the relative enrichment, defined as the total enrichment divided by the $\%Ip$. Likewise, total depletion was the reduction of $\%Np$ compared with $\%Ip$, and this is transformed into relative depletion by dividing by $\%Ip$ and subtracting from 1 in order to give a base value of 100%.

It was further possible to derive test characteristics that provide valuable insight into how the protocols function. As mentioned, sensitivity (se) is the percent of mApple+ cells that are correctly sorted into the positive fraction, and specificity (sp) is the percent of mApple- cells that are correctly sorted into the negative fraction. Working alongside these measures is the false negative rate ($FNR = 1-se$), or the percent of mApple+ cells that drop out to the negative fraction, and the false positive rate ($FPR = 1-sp$), or the percent of mApple- cells that are retained in the positive fraction. Accuracy (acc) can be conceived as a weighted average of sensitivity and specificity, analogous to the percent of all cells that are correctly sorted into their respective fractions.

While depletion explicitly aimed to purify the negative fraction by increasing the $\%Nn$, it was decided to keep the definitions of all measures consistent between enrichment and depletion for simplicity. Thus, a false positive was still a negative cell in the positive fraction, and a false negative was still a positive cell in the negative fraction for the sake of calculating test characteristics. Optimization processes specifically seek to adjust these characteristics; for example, specificity is more important than sensitivity for low prevalence, since false positives can easily outnumber true positives, while the opposite is true for high prevalence.

2.11.4 Number of washes

Populations of approximately 1%*Ip* and 5%*Ip* (equating to a 1:99 and 1:19 ratio of positive to negative cells, respectively) were chosen as snapshots of relatively rare populations analogous to the expected range of HDR-mediated insertion. Approximately 5×10^6 cells were prepared per sample and incubated for 30 min with 10 μ L Dynabeads. An aliquot of the initial population was set aside, the enrichment protocol was performed with seven washes, and each supernatant fraction was separated into individual wells of a round-bottom 96-well plate. The final positive fraction was also resuspended in rinsing buffer and added to the plate, and all fractions were analyzed by flow cytometry.

Because equal volumes were sampled, each fraction can be combined with all subsequent fractions by simply adding the cell counts to simulate the effect of ending the protocol at that wash step, forming a synthetic positive fraction. Likewise, adding the cell counts of each fraction with all previous fractions simulates ending the protocol at the following step, forming a synthetic negative fraction. Measured cell density was not affected because the sampled volume increased proportionally with the number of samples combined, so no corrective volume factor (v) needed to be applied. Thus, prevalence was just the %*p* from synthetic positive fraction 0 as it combined all future fractions and was thus analogous to the total initial population (i.e., if no selection or separation was performed). Synthetic positive fraction 7 was identical to the final positive fraction.

Because sparse samples make subsequent culture more difficult, a measure of the rate at which the total cell number declined was desired. Recovery was defined as the total number of cells in each synthetic positive fraction divided by the total cells in all fractions (Table 2.8).

2.11.5 **Bead volume**

A 10%*Ip* mixture of cells for enrichment and an 80%*Ip* mixture of cells for depletion (chosen as models of moderately-rare populations for each) was prepared, separated into several different tubes with 5×10^6 cells each and incubated for 30 min at RT with the indicated volume of beads. As before, an aliquot of the initial population was set aside, and, for the 10% mixture, enrichment was performed with three washes. For the 80% mixture, the first negative fraction was collected, and the remaining positive fraction was resuspended in rinsing buffer. Flow cytometry analysis was performed on each fraction for each sample, and test characteristics were calculated as above.

2.11.6 **Incubation time**

Populations of 5×10^6 cells at approximately 10%*Ip* were prepared, and enrichment was performed as above with 10, 20, or 30 minute incubations with beads spinning in suspension at RT.

2.11.7 **Cell density**

Initial populations were mixed at 1% and 10%*Ip* to test enrichment and at 90%*Ip* to test depletion. Total cell density was adjusted to 1×10^7 cells/mL, and samples were serially diluted to 5×10^6 , 1×10^6 , 5×10^5 , and 1×10^5 cells/mL in rinsing buffer. Since 1 mL rinsing buffer was used for each sample, cell density and cell number were interchangeable. Beads were added and washes were performed as per the previous recommendations, and initial, positive, and negative fractions were collected and quantified by flow cytometry. The change to the percent of mApple+ cells was calculated along with other test characteristics as above.

2.12 Aggregating protein clearance assays

Wildtype A53T-eGFP HeLa cells and ATG7 KO A53T-eGFP HeLa cells were transfected with siRNAs and cultured for 72 h before lysis. Cells were then dissociated by trypsin for flow cytometry.

Wildtype HeLa cells and ATG16L1 KO HeLa cells were co-transfected with empty pEGFP-C1, which acts as a transfection loading control, along with eGFP-74Q or eGFP-A53T. The following day, cells were transfected with siRNAs, and after 72 h in culture, cells were lysed for western blotting.

2.13 Proteasome activity reporter assay

GFP-dg HeLa cells were transfected with siRNAs and cultured for 72 h. For the final 6 h, cells were treated with 10 μ M MG-132 or DMSO. Cells were then either dissociated by trypsin for flow cytometry or fixed with PFA for imaging.

2.14 Transcriptional analyses

2.14.1 RNA isolation

Cells on a 6-well plate were washed once with PBS and lysed with 400 μ L/well TRIzol. RNA was then purified using the Direct-zol RNA miniprep kit, which includes treatment with DNase I. Aliquots of each sample were assessed for concentration and quality using a Nanodrop UV spectrophotometer and stored at -80°C until use.

2.14.2 RT-qPCR

Reverse transcription was performed with the SuperScript III First Strand Synthesis System using oligo(dT)₂₀ primers per the manufacturer's protocol. An equal amount of RNA was added to the RT reaction for every sample in each experimental replicate. cDNA was diluted to 3 ng/ μ L in DNase-free water, and 1 μ L of this template was loaded into 384-well PCR plates (Bio-Rad) along with a mix of 5 μ L PowerUp SYBR Green Master Mix, 0.4 μ L forward and reverse primers for each reaction (Table 2.9), and 3.2 μ L DNase-free water for 10 μ L total reaction volume. Each reaction for each sample was performed in technical triplicate on the same plate, and three experimental replicates of each cell line were included. Amplification was performed using a CFX384 Touch Real-Time PCR Detection System (Bio-Rad), and C_T values were obtained using CFX Maestro software (Bio-Rad).

Table 2.9: qPCR primers

Target	Forward primer (5'-3')	Reverse primer (5'-3')
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
ACTB	CACCATTGGCAATGAGCGGTTC	AGGTCTTTGCGGATGTCCACGT
L1CAM N-terminal	TGCTTATCCAGATCCCCGAG	TCTGTGGGGAAGACAACCAG
L1CAM Middle	TCGCCCTATGTCCACTACACCT	ATCCACAGGGTTCTTCTCTGGG
L1CAM ICD	GGCCCGACCGATGAAAG	TTGATGTCCCCGTTGAGC
LC3A	GCGAGTTGGTCAAGATCATC	TTCTCCTGCTCGTAGATGTC
LC3B	TCCTTGTACCTGACCATGTC	TCTGAGATTGGTGTGGAGAC
GABARAP	GAAGCGAATTCATCTCCGAG	TTCTTCATGGTGTTCCTGGT
GABARAPL1	ACCATCCCTTTGAGTATCGG	TTCTCTACAATCACGGGGAC
GABARAPL2	CTGAAAAGGCGATCTTCCTG	TTCTCGTAAAGCTGTCCCAT

Quantification of relative mRNA expression was performed by averaging the technical triplicate reactions and normalizing each sample by subtracting the mean C_T value of *GAPDH* and *ACTB*, both of which were included as reference genes on each plate, giving the ΔC_T . Next, each sample was normalized to the relevant controls by subtracting the mean ΔC_T value of the control cell line (i.e., NT for shRNA cells or WT for KO cells), giving the $\Delta\Delta C_T$. Finally, this value was log-transformed by calculating $2^{-\Delta\Delta C_T}$ to yield the amount of mRNA relative to the control line. The mean and SEM of the three biological replicates was calculated and a two-tailed paired T-test was performed to calculate p -values using the ΔC_T values.

2.14.3 **RNAseq**

Total purified RNA was submitted to the NICHD Genomics core for processing and sequencing. Tianwei Li and Steven Coon performed the following steps: RNA integrity number (RIN) was assessed using a BioAnalyzer (Agilent), samples with acceptable RIN values were processed by depleting rRNA with the Ribo-Zero Gold Kit (Illumina), followed by fragmentation and whole transcriptome library preparation with random primers using the TruSeq Stranded Total RNA Library Prep kit (Illumina), and samples were pooled and RNA sequencing was performed with an Illumina HiSeq using the paired-end Rapid Cluster Kit v2 (Illumina). This yielded total transcriptome data not limited to polyadenylated mRNAs (e.g., including lncRNAs). At least 60×10^6 read-pairs were generated per sample.

Kory Johnson processed raw data in the following manner: reads from each run were separated by sample barcode and combined across all runs, and quality control was performed with the CLC Genomics Workbench (Qiagen). Reads were trimmed and filtered, then mapped against the hg38 human reference genome followed by transcript expression analysis calculated as transcripts per million (TPM).

I then performed the following analysis, with assistance and guidance from Kory Johnson: transcripts not detected in any samples were removed, and TPM values were scaled by $\log_2(\text{TPM}+2)$, which provided a floor value of 1. Values were then quantile normalized across all samples to produce an even distribution. Next, noise modelling was performed by calculating the mean expression value for each transcript across the three replicates and plotting this against the respective coefficient of variation (CV). This permitted selection of a noise filter that removed highly variable and low-expressed transcripts. In this case, any transcripts for which no samples expressed $\log_2(\text{TPM}+2) > 3$ were removed. Any remaining transcripts with $\log_2(\text{TPM}+2) < 3$ were redefined as equal to 3 for those samples in order to remove statistical ambiguity below the noise threshold. The resulting mean of each condition was then calculated, and the difference between each mean gave the $\log_2(\text{fold change})$, or $\log_2\text{FC}$. Significance of this difference was determined by a two-tailed unpaired t-test followed by false discovery rate (FDR) correction using the Benjamini-Hochberg procedure.

I also analyzed expression of individual genes in the following manner: post-quantile normalization but pre-noise thresholding $\log_2(\text{TPM}+2)$ values were expanded to TPM and added for all annotated transcripts for each gene of interest. Since only a small number of mostly highly-expressed genes were analyzed by this method, total transcripts were rescaled by $\log_2(\sum(\text{TPM})+1)$ to yield a new floor of 0 rather than instituting a noise threshold. Averages and SEM were calculated using these values, and significance was determined by a two-tailed unpaired t-test followed by false discovery rate (FDR) correction using the Benjamini-Hochberg procedure.

2.15 Statistical analyses

The following are the default characteristics of the data throughout this work and can be assumed true unless otherwise noted. Calculations are rounded to three significant figures, values shown are the means of at least three experimental replicates, and error bars represent plus or minus the standard error of the mean (\pm SEM). Significance was assessed with two-tailed Student's *t*-tests thresholded as $p < 0.05$. Paired tests were used when comparing means of experimental replicates, while unpaired tests were used when biological replicates were run together (i.e., transcriptional analyses). Conditions that are significantly changed from the associated control condition are indicated in figures with a single asterisk regardless of the calculated *p*-value. Correlation assessments use the Pearson correlation coefficient (*R*). Additional tests and differing significance thresholds are explained in more detail when applicable. Data from representative experiments with fewer than three replicates are shown as the mean of the technical replicates; error bars represent plus or minus the standard deviation (\pm SD), and no significance tests were performed for these experiments.

In experiments with multiple comparisons, especially RNAseq and qPCR, statistical significance was assessed by applying a false discovery rate (FDR) correction using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). This is performed by ranking *p*-values from smallest to largest and calculating the adjusted critical *p*-value by multiplying 0.05 by the rank divided by the total number of tests. Then, the largest rank for which the measured *p*-value is less than the adjusted critical value is deemed significant, along with all ranks below this.

Chapter 3: **Sorting transgenic iPSCs by streptavidin affinity**

3.1 Introduction

3.1.1 Motivation

In the course of the work described in Chapter 4, I performed numerous gene edits on iPSCs, including stable integration of transgenes. Because mosaicism could mask chromatin changes resulting in silencing of transgenes, off-target mutagenesis, and genome instability leading to abnormal karyotypes, manual single-cell cloning was deemed necessary prior to genotyping and functional validation.

It rapidly became apparent that the cloning, genotyping, and validation process is the longest and most laborious stage of generating stable transgenic cell lines. While transient transfection of human iPSCs with plasmids can achieve expression in >50% of cells, targeted insertions via HDR usually yield stable, scarless genomic integration in <1% of cells (Miyaoaka et al., 2016). Alterations to the nuclease and delivery systems can improve the efficiency somewhat, but it rarely rises above low single-digit percentages (Carlson-Stevermer et al., 2016).

Selection genes are often included in transgenic expression cassettes to enable enrichment of a heterogeneous population, increasing the proportion of cells containing a desired edit and thereby widening this bottleneck. A variety of selection genes have long been used, but each has its own strengths and weaknesses (Debnath et al., 2010). I therefore sought to contribute to the development of new tools in this area, which would be both immediately applicable to my own efforts to derive the model lines in Chapter 4 as well as a valuable tool for any transgene insertion.

3.1.2 The selection gene toolbox

Selection genes may be used for positive selection, in which cells expressing the gene are desired, or for negative selection, in which cells lacking the gene are desired (Table 3.1). For positive selection, selection genes are often used that code for proteins that confer antibiotic resistance (enabling survival in the presence of, e.g., puromycin, blasticidin, or neomycin) or fluorescence (e.g., GFP and its relatives). These genes are then ligated into insert plasmids and transcribed under constitutive promoters such that they are highly expressed in cells that have integration of the plasmid into the genome.

However, there are disadvantages to these markers that leave room for different approaches. For instance, antibiotic resistance genes require optimization of antibiotic concentration depending on cell line and expression levels to ensure that all unedited cells die. This is often accomplished by using a non-transfected negative control alongside the transfected population to ensure complete cell death in the negative control. Despite this, negative cells in the center of colonies can be protected by nearby positive cells, limiting how well antibiotic selection can purify populations. Without a way to visualize the gene product directly, it is impossible to measure the purity of populations at a glance.

Antibiotics are also often used over a long term in culture media to select against silencing or excision, particularly in cells with unstable genomes. However, such long-term antibiotic use can inhibit differentiation potential in iPSCs and lead to unforeseen epigenetic changes (Farzaneh, 2020). In addition, cells that receive several copies of the resistance gene or express the gene to a greater extent may thereby be better protected against the antibiotic, so this selection strategy may inadvertently select for cells with off-target insertions. While this is a potential problem shared by all selection markers, without a way to easily compare the levels of resistance protein present in individual cells, antibiotic selection is at a particular disadvantage, whereas fluorescent

protein expression can be quantified during FACS and populations with different levels of expression can be separated.

Fluorescent proteins also have disadvantages. Constitutive expression of fluorescent proteins can interfere with immunocytochemistry, contributing to higher background and potentially removing the associated fluorescence channel from use, although this may be overcome by using different fluorescence channels that do not overlap with the excitation/emission spectrum of the expressed fluorescent protein. Moreover, fluorescent proteins can be prone to aggregation, which can reduce cell viability and produce extraneous variables in disease models (Wiedenmann et al., 2009).

Table 3.1: Common selection genes and associated strengths & weaknesses.

Positive Selection	Strengths	Weaknesses
Antibiotic resistance	Fast; inexpensive; sterile	Requires tuning; not visible; can select for incorrect insertion loci
Fluorescent protein	Easy to visualize and quantify; Multiple channels available	Requires FACS machine and/or picking scope; interferes with IF staining; sterility is difficult
Negative Selection	Strengths	Weaknesses
HSV thymidine kinase (TK)	One of only a few negative selectors; mirrors antibiotics	Requires tuning; TK can diffuse & sensitize non-expressing cells

Many of the problems with both antibiotic resistance and fluorescent proteins can be ameliorated by adding loxP sites on either side of the selection cassette and transiently transfecting with Cre to remove them after stable and validated clones have been derived. While this is recommended for the creation of important base models, as was the case for the WTC11 and G3 lines used in Chapter 5, this is a similarly laborious task since it requires restarting the selection process.

There are currently limited options for negative selection. If using only antibiotic selection, clones will need to be isolated blindly, split into two wells, and tested for antibiotic susceptibility in one well while keeping the other for expansion. This is somewhat easier with fluorescent proteins, as cells that lack fluorescence can be selected by FACS, and clones can be picked by marking purely dark colonies. The herpesvirus thymidine kinase (TK) has been used as a “suicide gene” since cells that express TK are sensitized to ganciclovir treatment (Wang et al., 2004). However, this requires an additional gene to be added to the selection cassette, may induce cell death in non-expressing cells, and carries with it all of the same issues as antibiotic resistance.

3.1.3 Magnetic bead cell sorting

In the field of immunology, highly heterogeneous cell populations in suspension are routinely separated into their constitutive cell types for study, and surface proteins are typically used as labels to identify these subpopulations. Antibodies with functional conjugates such as fluorophores or biotin enable identification and purification through FACS or by incubating cells with streptavidin-coated magnetic beads, and magnetic beads directly coated in primary antibodies can also be used for this purpose (Miltenyi & Schmitz, 2000; Weil et al., 2017). However, neither of these techniques is useful for separating transgenic cells out of an otherwise isogenic population, since the profile of

surface markers should not necessarily differ. In addition, the use of primary antibodies and antibody-conjugated beads for sorting is limited by high cost.

Recently, a study demonstrated a novel transgene tag that enabled enrichment of gene-edited lymphocytes in suspension using magnetic streptavidin-coated beads to select cells (Matheson et al., 2014). The authors employed a streptavidin binding peptide (SBP) fused to the N-terminus of the low-affinity nerve growth factor receptor (LNGFR) that was codon-optimized for amino acids 28-274, truncated just after the transmembrane domain which removed the intracellular domain to minimally impact cell biology, termed Δ LNGFR (Fig. 3.1a). The SBP tag has been used extensively in biochemistry as a fusion protein to mediate co-immunoprecipitation with streptavidin-coated magnetic beads (Pezzi et al., 2018; Yang et al., 2015). LNGFR is a cell-surface receptor highly expressed in neurons; its inclusion effectively localizes the tag to the extracellular membrane (Yamamoto et al., 1993). Together, cells that express the SBP- Δ LNGFR tag were found to bind magnetic streptavidin-coated beads on the cell surface and thereby enable selection (Matheson et al., 2014).

3.1.4 Aims & Hypotheses

In this project, I first sought to extend the use of the SBP- Δ LNGFR construct as a marker gene to iPSCs. Since their survival requires adherence to a specialized matrix and proximity to a colony, it was not clear whether iPSCs would tolerate extended isolation in suspension, binding to beads, and magnetic selection. Further, I hypothesized that the SBP- Δ LNGFR tag could be used for negative selection, enabling purification of a fraction depleted of cells expressing the tag; however, this functionality had not been demonstrated previously.

After a proof-of-concept of SBP- Δ LNGFR expression enabling magnetic bead enrichment was achieved (Chapter 3.2), I then aimed to optimize selection parameters.

An ideal enrichment process would capture all the positive cells while not capturing any negative cells. For instance, in rare populations, negative cells heavily outnumber positive cells, so allowing even a small percentage of negative cells into the positive fraction can drastically reduce enrichment efficiency. I hypothesized that varying the number of wash steps in the protocol, cell density, bead volume, and incubation time would affect the efficiency of the selection process, so I set to characterize the effects of these variables in test populations.

Once optimized protocols were established for both enrichment and depletion (Chapter 3.3), I further aimed to better understand the dynamics of these processes. To this end, I hypothesized that methods for the assessment of diagnostic tests and probabilistic modelling could be adapted to quantify selection efficacy over a range of input conditions and to predict the results of future trials (Chapter 3.4 – 3.6).

3.2 SBP- Δ LNGFR validation

3.2.1 Stable cell line derivation

The constructs Mag-hNIL and Mag2-hNIL were constructed as described in Chapter 2. The Mag selection cassette featured the SBP- Δ LNGFR construct and the gene for mApple, a red fluorescent protein, separated by 2A cleavage peptides and flanked by loxP sites to enable excision by transient Cre expression, upstream of the rtTA3G under the same CAG promoter (Fig. 3.1b). The Mag2 selection cassette featured the SBP- Δ LNGFR-T2A-mApple cassette under its own EF-1 α promoter (Fig. 3.1c). These plasmids are discussed in more detail in Chapter 4 regarding differentiation capacity and functional utility; the scope of this chapter was to assess the viability and efficacy of magnetic streptavidin bead-based cell sorting using the SBP- Δ LNGFR tag in iPSCs.

A wildtype iPSC line (H23) was transfected with 1.25 μg Mag-hNIL and 0.75 μg of each of the two *CLYBL*-targeting TALEN plasmids pZT-C13-L1 & pZT-C13-R1. After culture for seven days to allow transiently-expressed proteins to degrade, only cells with genomic insertion were expected to express SBP- Δ LNGFR, mApple, and rtTA3G from the polycistronic cassette.

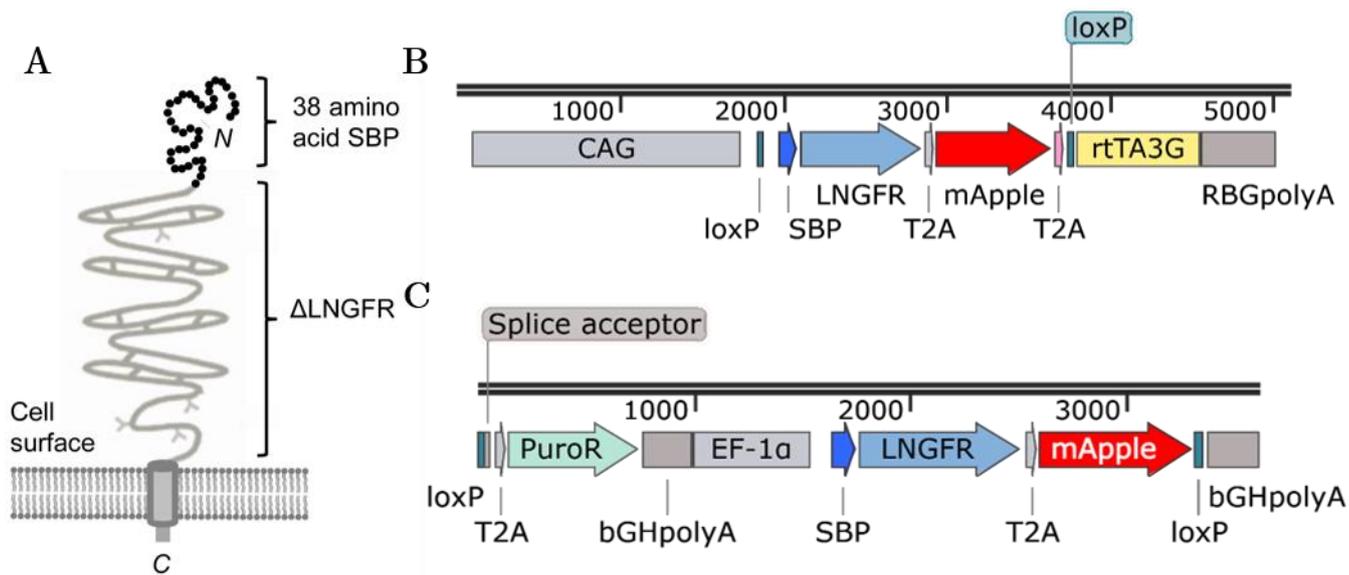
Afterwards, clones were isolated and validated by PCR and Sanger sequencing of genomic DNA at the insertion site (See Chapter 2 & 4). Clones were selected that had heterozygous insertion by PCR of the wildtype allele to ensure that expression originated from one copy of the plasmid. Only clones that sequenced with an unedited second allele were used further to facilitate future transgene insertion if desired.

3.2.2 SBP- Δ LNGFR co-expresses with mApple and binds to streptavidin

First, the level of co-expression of extracellular-facing SBP- Δ LNGFR and cytoplasmic mApple was assessed by dissociating cells with EDTA and incubating populations of unedited or Mag-hNIL clonal cells with streptavidin (SA)-conjugated Alexa Fluor (AF) 488 (SA488) for 15 minutes, followed by two PBS washes with centrifugation for 5 min at 200 \times g. In the Mag-hNIL clone, a clear ring of SA488 was evident on the surface of over 99% of mApple-positive (mApple+) cells (601/606 cells were mApple+; 595/601 mApple+ were also AF488+), while no SA488 was observed on any mApple-negative (mApple-) cells (0/5 mApple- cells in the Mag-hNIL clone were SA488+; 0/630 wildtype cells were mApple+ or SA488+) by fluorescent microscopy (Fig. 3.2a).

Figure 3.1: Strategy of selection via SBP- Δ LNGFR.

- (a) Diagram of the SBP- Δ LNGFR construct. SBP is presented on the exterior of the cell membrane, while the C terminus is truncated just after the transmembrane domain. From Matheson et al. (2014).
- (b) Diagram of the Mag selection cassette present on the Mag-hNIL plasmid (see Chapter 2 for a full plasmid map). The same CAG promoter drives expression of the SBP- Δ LNGFR, mApple, and rtTA3G genes separated by 2A cleavage peptides.
- (c) Diagram of the Mag2 selection cassette present on the Mag2-hNIL plasmid (see Chapter 2 for a full plasmid map). The puromycin resistance gene (PuroR) is driven by the endogenous CLYBL promoter, while the EF-1 α promoter drives SBP- Δ LNGFR and mApple.

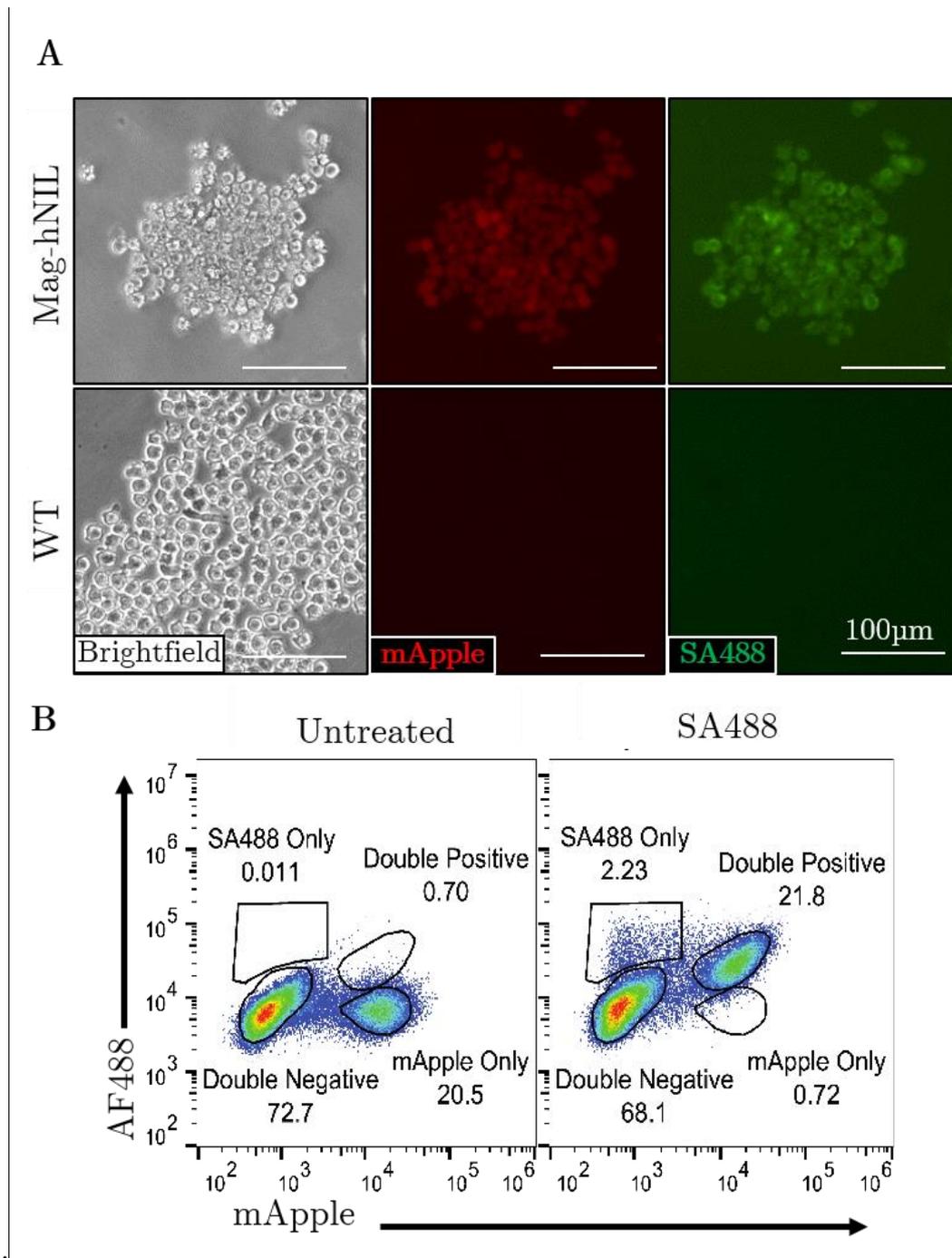


Flow cytometry also showed high correspondence of SA488 labeling and mApple expression (Fig. 3.2b). Sensitivity (*se*), or the percent of mApple+ cells that were SA488+, was 90.7%, while the specificity (*sp*), or the percent of mApple- cells that were SA488-, was 98.9%. Together, this experiment demonstrated a 96.8% accuracy (*acc*) of SA488 to label (and only label) mApple+ cells. Sensitivity, specificity, and accuracy are measures commonly used to assess the efficacy of diagnostic tests (Lalkhen & McCluskey, 2008; Parikh et al., 2008; Whitman et al., 2020); these measures are discussed in more detail below. However, it is important to note that determining them in this manner assumes that flow cytometry is a gold standard (i.e., a perfect test with *se*, *sp*, & *acc* all equal to 1). More complex methods exist that do not make this assumption (W. O. Johnson et al., 2001; Joseph et al., 1995); however, previous studies have demonstrated near-perfect sensitivity and specificity of distinguishing mApple against a negative background by flow cytometry (Kleeman et al., 2018), so this assumption was deemed valid.

These data provided compelling proof-of-concept that cells positive for SBP- Δ LNGFR can be identified using mApple expression as a proxy, as expected due to their expression under the same promoter. The high accuracy of SA488 labeling also suggested that this marker could be used for an SA affinity-dependent process, namely population enrichment via SA-coated magnetic beads. Finally, these data also suggested that transient labeling by fluorophore-conjugated SA could replace the utility of constitutive fluorescent protein expression in future constructs.

Figure 3.2: Characterization of SBP- Δ LNGFR *in vitro*.

- (a) Visualization of clonal Mag-hNIL cells and the unedited parental cell line (WT). Cells were EDTA split, treated for 15 minutes with streptavidin conjugated to Alexa Fluor 488 (SA488), washed twice with PBS, and visualized in suspension.
- (b) Accuracy of streptavidin binding was assessed by mixing Mag-hNIL and WT cells, separating the same population into two aliquots, and labeling one with SA488 as above. Flow cytometry was performed; SA488 demonstrated high sensitivity (90.7%) and specificity (98.9%) for labeling only mApple+ cells.



3.2.3 Proof-of-concept of enrichment

To establish the enrichment protocol, parameters were initially chosen similar to those used by Matheson et al. (2014). However, this protocol was developed for primary human CD4⁺ T cells, so further experiments were planned to optimize this protocol for iPSCs. It was also desired to establish if the negative fraction (i.e., the supernatant removed from the selected positive fraction) could be depleted of positive cells so as to extend the use of SBP- Δ LNGFR to depletion of the negative fraction.

Accutase dissociation resulted in a loss of SA488 binding, suggesting that enzymatic dissociation digested SBP- Δ LNGFR and is therefore incompatible with this method (Fig. 3.3a). Because of this, pure populations of cells were dissociated by washing with PBS followed by incubation with 0.5 mM EDTA in PBS for 15 minutes at RT, resulting in a mostly-singularized population and effectively removing biotin that could otherwise compete for SA binding, which is present in E8 medium and matrigel. These cells were collected in PBS, mixed, counted, and 5×10^6 cells were aliquoted into 15 mL conical tubes and centrifuged for 5 min at $200 \times g$. After removing the supernatant, pellets were resuspended with 1 mL autoMACS rinsing buffer (Miltenyi), consisting of a sterile-filtered solution of PBS and 0.5 mM EDTA, supplemented with 0.5% w/v BSA (Miltenyi).

Meanwhile, a 50 μ L aliquot of Dynabeads MyOne Streptavidin C1 (Thermo) beads was prepared for each sample in a 1.5 mL microcentrifuge tube. Beads tended to settle to the bottom of the bottle, so the beads were resuspended by vigorous inversion immediately before use. Additionally, beads were stored in a buffer containing Tween-20 and sodium azide as preservatives, and it is possible that unconjugated streptavidin could slough off beads and compete for SBP binding. To remove these, beads were washed by adding 1 mL rinsing buffer. These tubes were then magnetized on a DynaMag 2 magnetic tube rack until the supernatant was clear, and the supernatant

was removed by pipetting. This wash was repeated two additional times before adding the rinsing buffer containing cells.

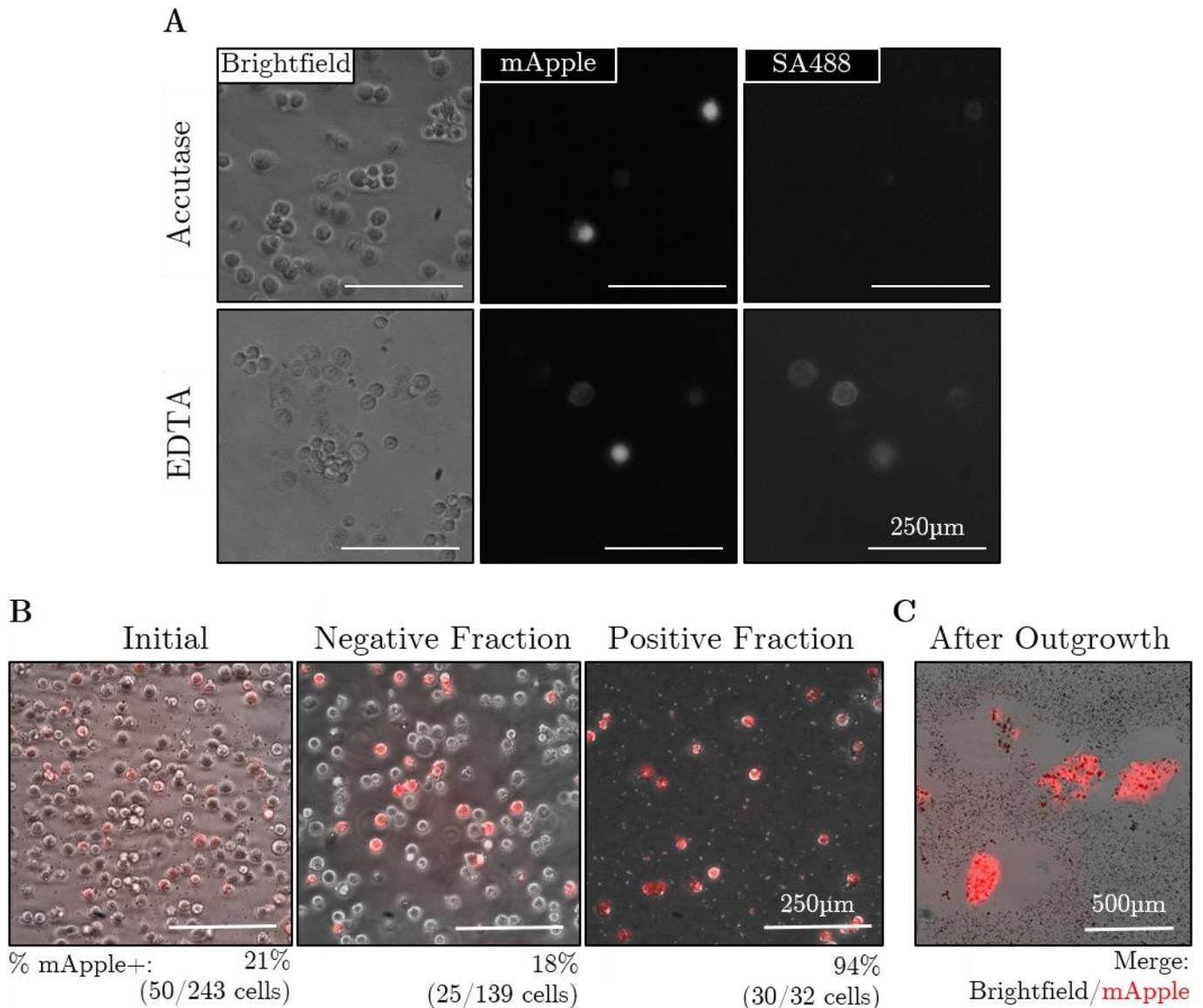
Cells and beads were spun together on a bench-top spinner at approximately 10 rpm at RT for 30 min. Tubes were then magnetized, and the rack was inverted several times to mix buffer from the inside of the tube lid. High cell density made the samples opaque, so magnetization continued for at least 2 minutes, until the sample was mostly white, with the brown beads against the side of the tube. The supernatant (negative fraction) was removed by pipetting down the outside wall of the tube, being careful not to disturb or remove beads, particularly at the bottom. The tube was then removed from the magnet, and the positive fraction was resuspended in 1 mL fresh rinsing buffer. This process was repeated two additional times, and following the third wash the final positive fraction was resuspended in fresh E8 supplemented with 10 μ M ROCK inhibitor (RI) and directly plated to a matrigel-coated 6-well plate.

Preliminarily, efficacy was assessed by imaging cells in suspension directly following selection. High numbers of mApple+ cells remained in the negative fraction, suggesting that depletion would require optimization of the protocol. While the density of cells was decreased in the positive fraction, its purity was substantially improved over the initial population (Fig. 3.3b).

It was possible to release beads from cells by incubating the final positive fraction with E8 supplemented with an excess of biotin in suspension, since biotin outcompeted the SBP for bead binding. Cells could then be isolated by re-magnetization and plating the supernatant. However, it was found that plating cells and beads together did not affect survival or outgrowth (Fig. 3.3c). Instead, beads tended to embed in the matrigel and were removed at the next passage. After plating, media was changed daily, and 10 μ g/mL RI was added until colonies were established, typically 2-3 days after plating.

Figure 3.3: Establishing the enrichment protocol.

- (a) Dissociating non-clonal Mag-hNIL cells with accutase for 5 min resulted in diminished binding to SA488 as compared to EDTA dissociation for 15 min.
- (b) Pure positive and negative populations were mixed, and the sample enrichment protocol was performed. Aliquots of each fraction were plated and imaged, and the percent of mApple+ cells was estimated by counting. An initial fraction with 21% mApple+ cells resulted in an 18% mApple+ negative fraction (each flow-through combined) and a 94% mApple+ final positive fraction.
- (c) After directly plating the positive fraction in E8+RI, several pure mApple+ colonies were evident. Background specks were beads; they did not inhibit cell growth and were removed from culture after one passage.



3.3 Protocol optimization

3.3.1 Aims

While the proof-of-concept enrichment protocol was found to be effective at a relatively high initial percent of positive cells ($%Ip$), it was less effective with rarer populations which are common with stable genomic integrations. In addition, it was noted that the negative population was not effectively depleted using this protocol. It was thus necessary to establish a depletion protocol and to maximize efficiency in both cases to investigate the effect of varying several key parameters of the selection protocol to characterize the effect of these variables.

In the following subsections, the key results are presented, while more detailed descriptions of the methods are provided in Chapter 2.11 for concision. Chapter 2.11 also establishes definitions for the abbreviations, calculations, and output measures used throughout this chapter. The experiments below were performed in the order presented; results from each experiment informed the conditions of the subsequent experiments. The optimal conditions for enrichment and depletion were carried forward for each protocol independently, resulting in the two distinct protocols presented below (Chapter 3.3.6).

3.3.2 Number of washes

Because the wash steps were the most labor-intensive step of the protocol, it was first desired to determine the effect of adjusting the number of washes and ascertain the ideal number. Recovery was found to decline rapidly for the first few washes and then level off in a manner suggestive of exponential decay (Fig. 3.4a).

By combining the positive cell counts in the same manner, the percent of positive cells in each synthetic fraction, and thus relative enrichment, could be calculated (Fig. 3.4b). This provided a measure of the efficacy of each wash step relative to the initial conditions. In particular, while additional washes increased the relative enrichment, this proceeded in a linear manner, so marginal returns decreased and variability increased with more washes.

The proportion of cells in each individual fraction that were bead-bound was also determined to assess the leakiness of magnetic selection itself (Fig. 3.4c). The percent of bead-bound cells increased as wash steps proceeded, but the bulk of bead-bound cells remained in the final positive fraction. This provided strong evidence that bead-bound cells were preferentially retained in the positive pellet as desired.

Figure 3.4: Optimization of the number of washes.

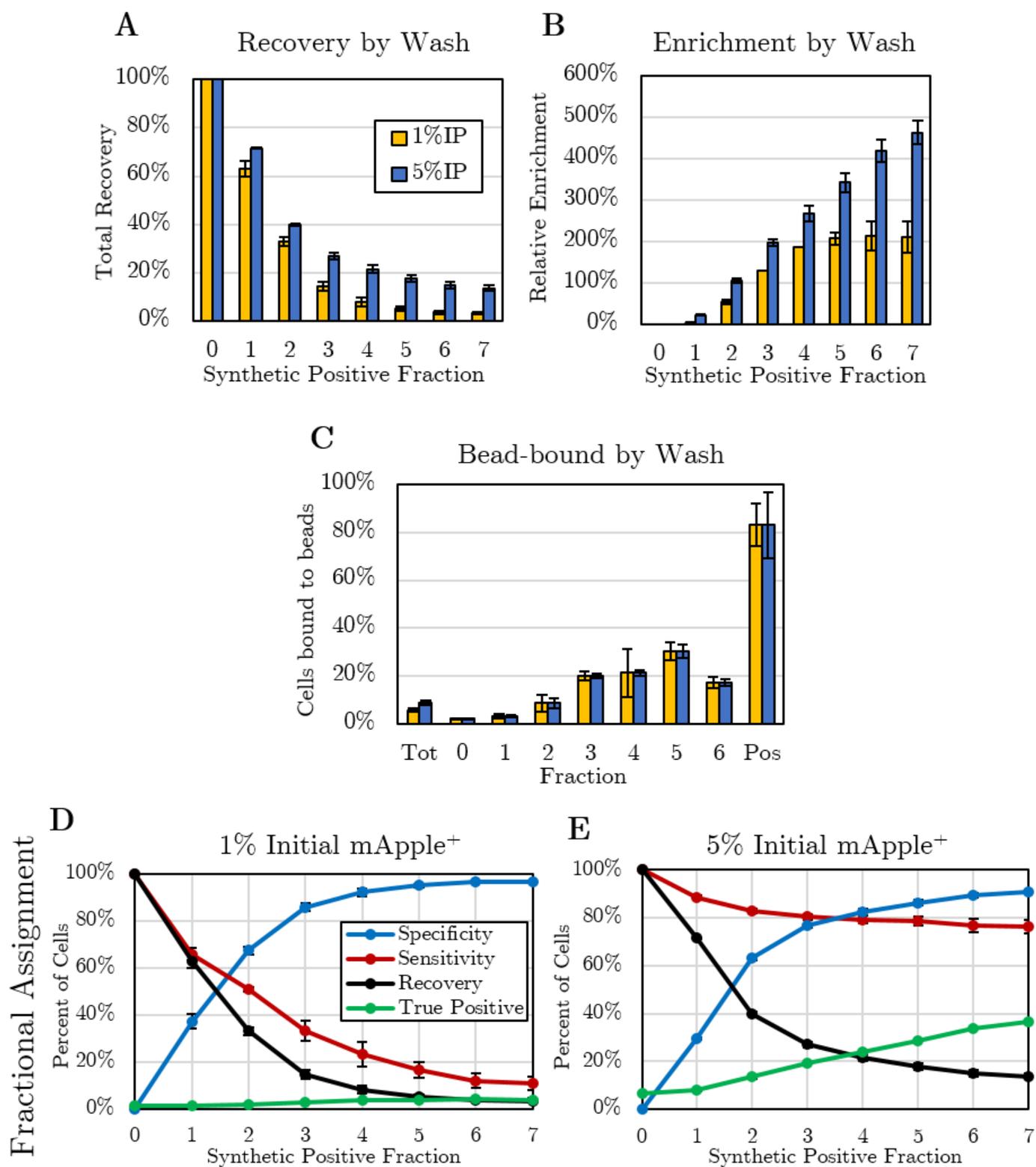
Pure clones with or without Mag-hNIL were mixed at the indicated ratios, the enrichment protocol was performed with six washes, and each wash was collected and measured by flow cytometry separately. Cell counts were combined for each fraction with all future fractions to form a synthetic positive fraction, simulating the result of ending the experiment at that step. Error bars are \pm SD of 2 technical replicates for each %Ip

- (a) Recovery (percent of all cells remaining) decreased rapidly as washes increased.
- (b) Relative enrichment (change to % mApple+ normalized by %Ip) increased gradually as washes increased.
- (c) The percent of bead-bound cells in the total initial population (Tot) and in individual (not combined) fractions. While more bead-bound cells leaked into later washes, a majority were retained in the final positive fraction (Pos).

(continued)

Figure 3.4 (*continued*): Optimization of the number of washes.

(d-e) Rates of fractional assignment for 1% (d) and 5% (e) initial mApple+ cells by synthetic positive fraction. Additional washes increased specificity (percent of all negative cells removed) and percent True Positive ($\%Pp$) but decreased sensitivity (percent of all positive cells remaining) and recovery.



In this context, for a given fraction x , sensitivity was analogous to the percent of all positive cells recovered in synthetic positive fraction x , and specificity was analogous to the percent of negative cells that dropped out to the respective synthetic negative fraction $x-1$. Thus, sensitivity began at 1 because all positive cells were present in the initial fraction, and specificity began at 0 because none of the negative cells had been removed. As washes proceeded, sensitivity declined as more of the positive cells were lost to the negative fraction, but specificity climbed as fewer false positives remained (Fig. 3.4d-e). Because bead-bound cells were preferentially retained, and beads preferentially bound to mApple+ cells, sensitivity was predicted to decline more slowly than overall recovery and thus promote enrichment. In addition, the rate of decline of sensitivity was noted to be slower in the 5%*Ip* samples than for 1%*Ip*.

As a crude measure, the point at which sensitivity and specificity crossed can be considered the point of diminishing returns, and little is to be gained from continuing past where they start to level off. The exponential reduction to recovery paired with only modest increases in enrichment resulted in a general recommendation that three washes (equivalent to the third synthetic positive fraction in Fig. 3.4) was ideal; however, it may be advantageous to perform more washes when samples have higher %*Ip* and higher initial cell densities. In this way, better enrichment may be attained, albeit at the expense of reduced recovery and a higher rate of positive cells leaking into the negative fraction.

Additionally, this experiment provided a proof-of-concept for the depletion protocol, suggesting that a strategy of isolating the first supernatant fraction as the sole negative fraction (equivalent to choosing the first synthetic positive fraction as the final positive fraction) could be used. Further washes would be unnecessary, as each subsequent wash would provide a reduced marginal benefit of fewer negative cells while reducing purity by permitting more positive cells into the negative fraction.

3.3.3 Bead volume

Next, the effects of varying bead volume were tested. Accuracy of selection decreased as bead volume increased in 10%*Ip* enrichment, but it increased over the same range for 80%*Ip* depletion (Fig. 3.5a). At the same time, enrichment was only efficient at low bead volumes, with high bead volumes faring little better than no beads. Conversely, depletion demonstrated greater efficiency at higher bead volumes, equating to smaller %*Np* values as bead volume increased (Fig. 3.5b).

There may be more interaction of beads with negative cells at higher bead densities; looking at only those cells that were bound to beads in the initial population, the percent of mApple+ cells peaks at low bead volume and decreases rapidly for low %*Ip*, potentially indicating that saturation of positive cells enabled off-target bead binding (Fig. 3.5c). Likewise, high %*Ip* showed highest bead selectivity for mApple+ cells at low bead volume, but it decreased slower as bead volume increased. The 8-fold higher ratio of positive cells provided more binding sites to which beads could attach, thereby preventing saturation.

Evidently for enrichment, less was more, while depletion favored higher volumes. Intuitively, fewer beads meant that cells must compete for limited SA-binding sites, which preferentially bind to positive cells. However, at high bead volumes, nonspecific interactions can compromise the selectivity of enrichment, preferentially trafficking all cells to the positive fraction regardless of whether they are truly positive. Likewise, while high bead volumes benefit depletion by saturating the positive cells, too many beads lead to diminished recovery and inflated reagent cost. Thus, the volume of beads recommended was 10 μ L for enrichment and 100 μ L for depletion.

3.3.4 Incubation time

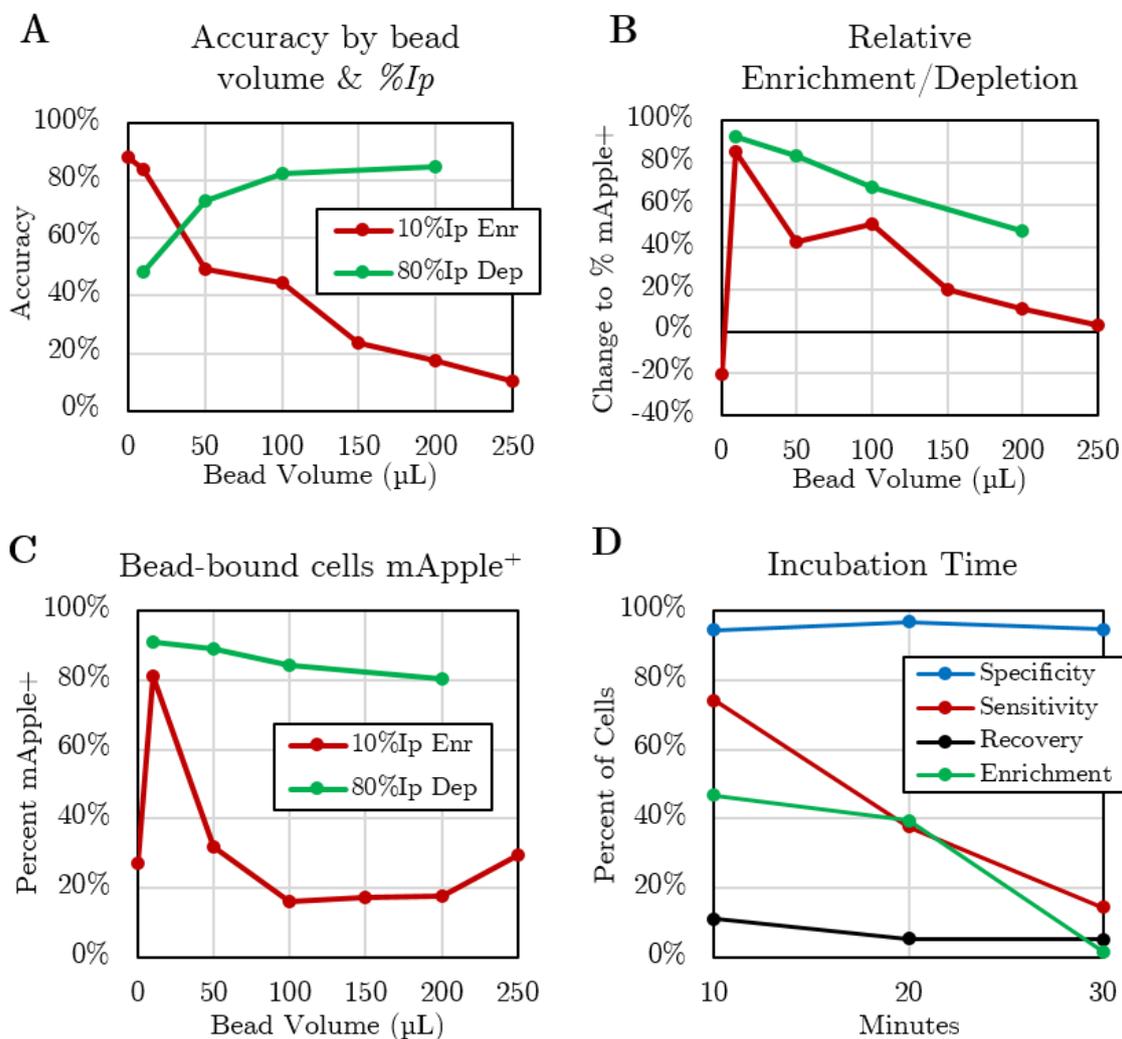
It was then sought to ascertain whether the recommended incubation time of 30 minutes could be decreased to increase protocol throughput. It was hypothesized that some time was necessary for proper attachment, but it may be possible to reach a steady state more quickly. Surprisingly, specificity and recovery remained consistent, while sensitivity and effective enrichment actually decreased as time went on (Fig. 3.5d). As sensitivity decreased, the false negative rate increased proportionally, so depletion also suffered from increased incubation time. The recommended incubation time for both protocols was thus reduced to 10 minutes.

3.3.5 Cell density

The ability to sort more cells would increase throughput and may overcome challenges of low recovery or sensitivity; thus, the effects of varying cell density were assessed. It was originally hypothesized that high cell density would reduce singularization, leading to reduced sorting accuracy. Contrariwise, enrichment efficiency improved at high cell density, with no upper bound established, while depletion improved at low density (Fig. 3.6a). Further, the percent of all cells that were single cells (as assessed by forward scatter height vs. area) remained high across all densities in the initial population (Fig. 3.6b), with similar results for each fraction after selection. The rate of cell recovery also remained surprisingly stable and was mostly dependent on the $%Ip$. However, recovery in the depleted negative fraction was very low, resulting in too few selected cells for this protocol to be reliable when fewer than 5×10^5 cells were loaded (Fig. 3.6c).

Figure 3.5: Optimization of bead volume and incubation time.

- (a) Mixed populations of 5×10^6 cells at 10%*Ip* or 80%*Ip* were subjected to enrichment or depletion, respectively, using the indicated volume of beads. Data shown indicate the accuracy, equivalent to the percent of all cells correctly assigned to their respective fraction.
- (b) The increase (from 0%) of mApple+ cells in the enriched positive fraction or the decrease (from 100%) of mApple+ cells in the depleted negative fraction, relative to %*Ip*.
- (c) The percent of bead-bound cells that are also mApple+ was maximized at low bead volumes for both low and high %*Ip*, but it dropped rapidly for low %*Ip*.
- (d) Samples with 5×10^6 cells at 10%*Ip* were prepared and incubated with beads for the indicated duration, enrichment was performed, and test characteristics were determined. Surprisingly, sensitivity and enrichment decreased after increased duration, while specificity and recovery remained consistent.



To gain insight into how the depletion protocol functions, key characteristics were plotted (Fig. 3.6d). These measures were calculated in the same manner as for enrichment (i.e., sensitivity was the percent of mApple+ cells that were sorted into the positive fraction, while sensitivity was the percent of mApple- cells that were sorted into the negative fraction). Interestingly, specificity increased alongside cell density, while sensitivity decreased. Since the initial population used here was approximately 90% positive, sensitivity had a proportionally stronger effect on the purity of the negative fraction, thus accounting for decreased efficiency at higher cell densities despite increased specificity (and accuracy).

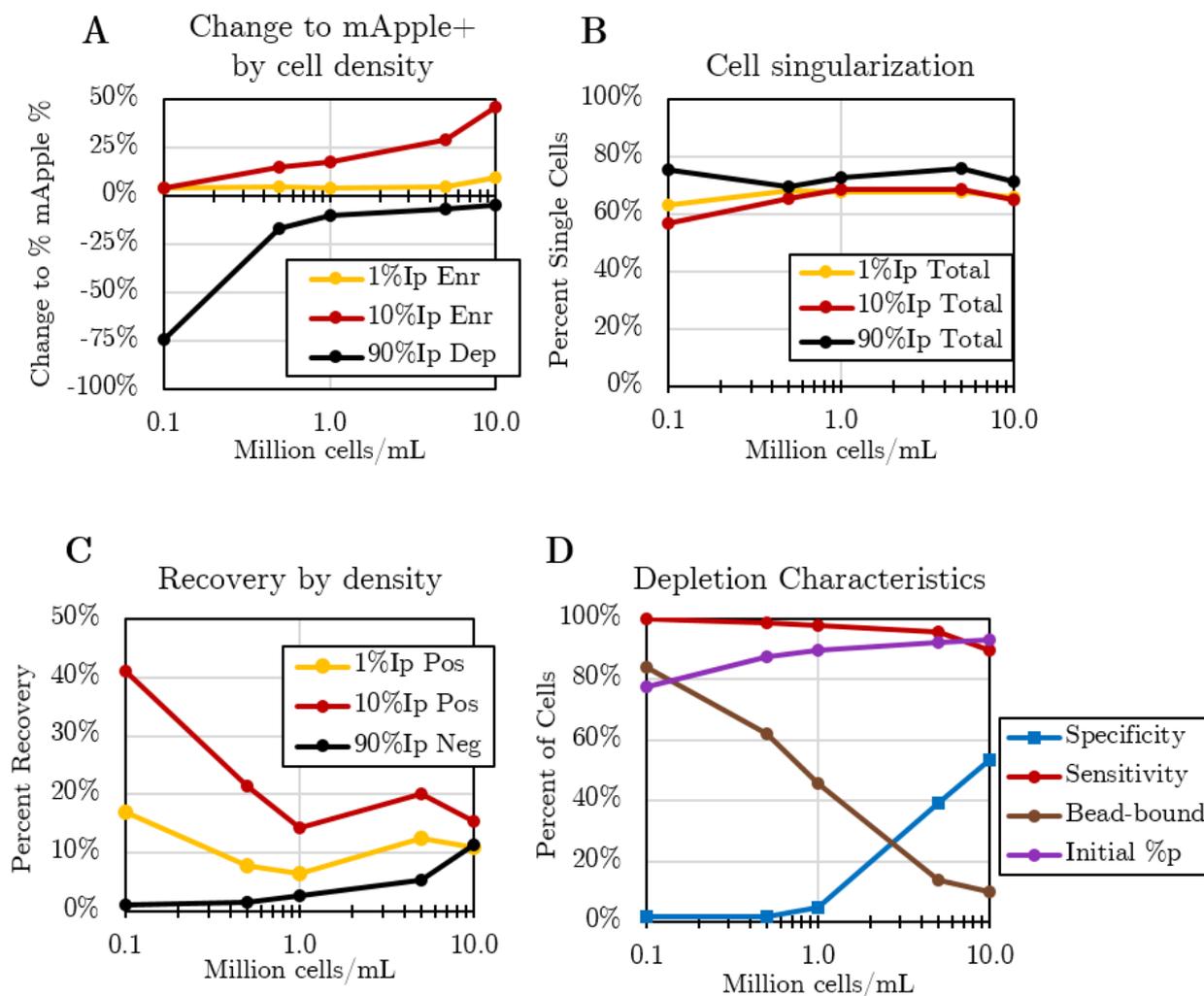
In addition, the proportion of all cells that were bound to beads decreased with high density despite an approximately equal initial prevalence of positive cells. This indicated that higher cell densities allowed more mApple+ cells to escape bead binding, potentially due to beads being saturated by potential SBP binding sites, explaining the loss of sensitivity. This would also explain the increased specificity, as fewer free beads were available to nonspecifically interact with negative cells.

Taken as a whole, increased cell density was beneficial for enrichment; efficacy was enhanced at no appreciable loss to singularization or selectivity. Recovery remained stable, so more cells could be selected at no additional cost. While no upper limit was established here, the enrichment protocol recommends 1×10^7 cells/mL since it approximately corresponds to a nearly-confluent 10 cm dish, and higher densities may be unwieldy. Meanwhile, depletion efficacy benefitted from lower cell density, but at the high cost of decreased recovery making the resulting negative fraction proportionally smaller. As a result, the depletion protocol recommends 1×10^6 cells/mL, or about one well of a 6-well plate, since this provides an adequate reduction in positive cells as well as enough cells overall to reliably centrifuge and culture.

Figure 3.6: Optimization of cell density.

Mixed populations at the indicated density and %*Ip* were subjected to enrichment using 10 μ L beads or depletion using 100 μ L beads. Density is presented on a log₁₀ scale as millions of cells per 1 mL rinsing buffer.

- Total enrichment of the positive fraction (Enr) or depletion of the negative fraction (Dep).
- Rates of cell singularization (number of single cells divided by total cells counted) in the total initial population. Note that 1 and 10 %*Ip* conditions are overlapping.
- Percent of all cells recovered in the enriched positive fraction (Pos) or depleted negative fraction (Neg).
- Summary of depletion test characteristics for the 90 %*Ip* depleted negative fraction. Sensitivity and specificity are calculated as for enrichment. Bead-bound refers to the percent of all cells in the initial population bound to beads.



3.3.6 Summary of optimization

The experiments discussed above explored a wide range of variables and revealed much about the dynamics of bead binding and magnetic selection. It was also demonstrated that SBP- Δ LNGFR can be used for negative selection (depletion) with only slight modifications to the overall protocol. A visual diagram of these protocols is provided in Fig. 3.7a, and an example experimental workflow for a hypothetical genetic insertion followed by Cre excision of the selection markers incorporating both enrichment and depletion is provided in Fig. 3.7b. Essentially, this is an extension to the utility of SBP- Δ LNGFR as a selection gene, and as such this is designed as an extension to the workflow proposed by Matheson et al. (2014).

Enrichment efficiency is increased by using more cells with fewer beads for a shorter time and then performing more washes. Conceptually, this strategy increases competition for a limited number of bead binding sites and prioritizes those cells that bind quickly and strongly. Many mApple⁺ cells may end up in the negative fraction, but few mApple⁻ cells remain in the positive fraction; in other words, this strategy prioritizes specificity over sensitivity, as false negatives are preferable to false positives. As a result, the final recommendations for the enrichment protocol are: 1×10^7 cells and 10 μ L Dynabeads spun for 10 min, followed by 3 washes (Fig. 3.7c).

It was also found that the opposite end of the effective range for most variables increased efficiency for depletion. Using fewer cells and more beads permitted an excess of beads that bound mApple⁺ cells more completely, and using only the first supernatant as the negative fraction made selection a one-step process with no washes. As with enrichment, many mApple⁻ cells may be caught in the bead pellet, but few mApple⁺ cells remain in the supernatant. Depletion is thus recommended to use 1×10^6 cells and 100 μ L Dynabeads spun for 10 min, and only the first fractionated supernatant is taken without any washes (Fig. 3.7c). However, unlike the enrichment

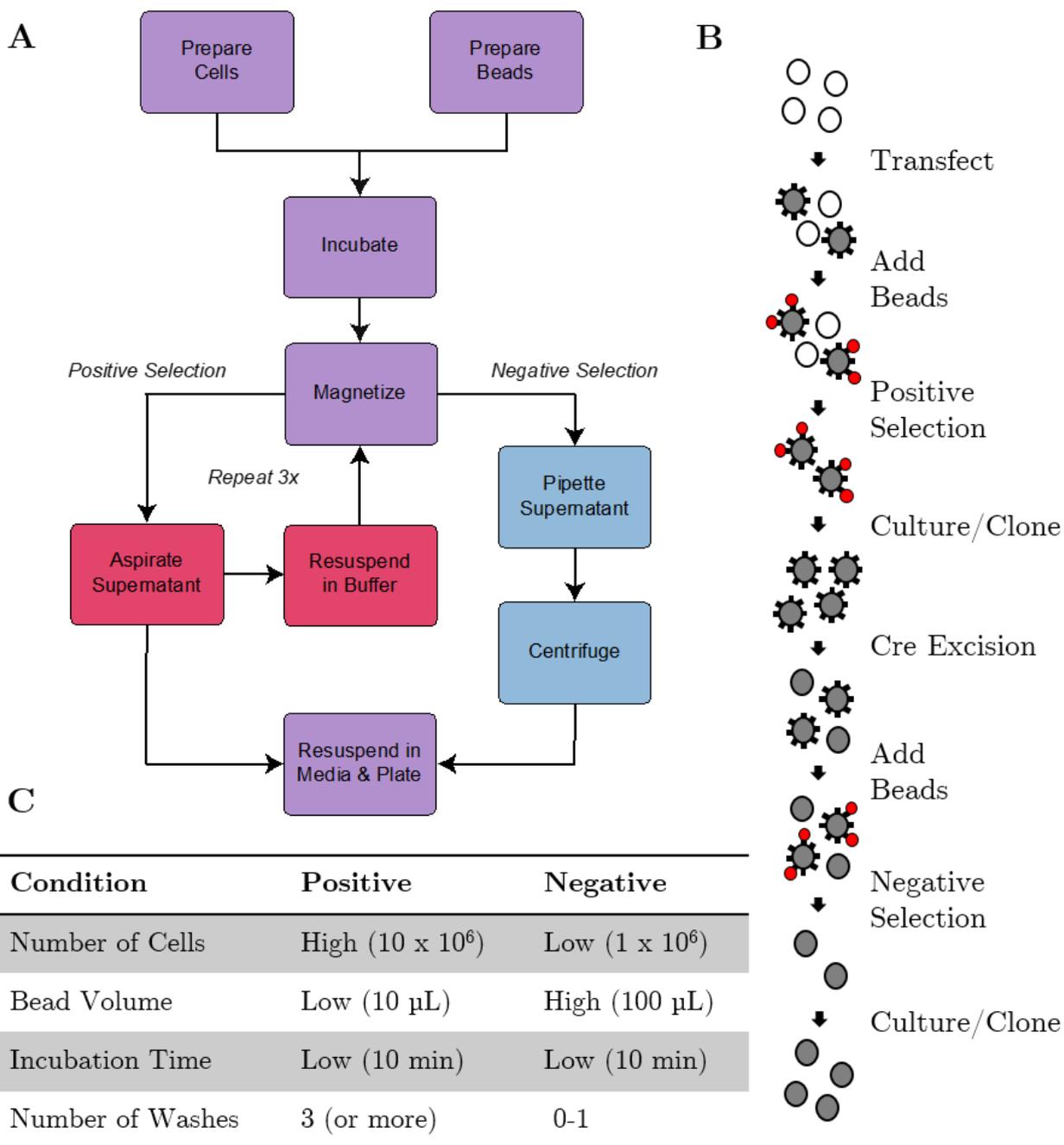
protocol, depletion results in cells suspended in the washing buffer, so they must be centrifuged before replating in fresh media.

While these recommendations work well in general, results may be improved further by adjusting for individual contexts. In particular, selection from an initial population with a low $\%Ip$ is affected more strongly by specificity, as negative cells outnumber positives, while sensitivity dominates when prevalence is high. For instance, increased specificity in enrichment may be achieved by performing more washes; the low resulting recovery may be abrogated by loading a higher initial cell density (which was itself shown to further increase specificity).

Bead volume could be adjusted gradually to avoid oversaturation in either direction, or it may be adjusted intentionally to prioritize either sensitivity (more beads) or specificity (fewer beads). Meanwhile, the strategies that increase sensitivity also tend to increase recovery to the positive fraction, which translates to low cell density in the depleted negative fraction. While depletion can be highly effective, it must be acknowledged that subsequent culture is more challenging with low density, so it may be desirable to sacrifice some efficacy (e.g., by performing one wash or increasing cell density) in order to increase cell density in the depleted negative fraction.

Figure 3.7: Protocol recommendations.

- (a) Summary schematic of streptavidin-coated magnetic bead selection protocols.
- (b) Example experimental workflow. Adapted from Matheson et al. (2014).
- (c) Efficiency of positive selection (enrichment) is maximized by using more cells, fewer beads, and more washes, while the opposite is optimal for negative selection (depletion). Both appear to benefit from decreased incubation time.



3.4 Assessing efficacy

3.4.1 Experimental design

It was noted in the experiments above that selection efficiency, in the absence of other variables, appeared to depend heavily on the %*Ip*. It was thus desired to test the efficacy of enrichment and depletion across a broader range of initial conditions in order to investigate this relationship. As before, populations of wildtype and Mag-hNIL positive cells were mixed at various ratios and either enrichment or depletion was performed using the recommended protocols. An aliquot of the initial population was again set aside after bead incubation but before selection. For enrichment, the three supernatant fractions were mixed to constitute the negative fraction, and the final positive fraction was resuspended in rinsing buffer to constitute the enriched positive fraction. For depletion, the first supernatant was separated as the depleted negative fraction, and all remaining cells and beads were resuspended in rinsing buffer to constitute the positive fraction. An equal volume of each fraction was then sampled by flow cytometry with the same flow rate, and the same gating as the optimization experiments was applied to determine characteristics as in Table 2.8.

3.4.2 Enrichment efficiency

Plotting data as a function of %*Ip*, it is evident that total enrichment is low for low %*Ip* since positive cells are simply outnumbered. However, total enrichment rapidly increased such that 12%-50%*Ip* yielded the best increases of approximately 40%, while higher %*Ip* approached 100% (Fig. 3.8a). Meanwhile, relative enrichment revealed exponentially higher efficiency for very rare populations, where a %*Ip* of <1% regularly yielded approximately 10-fold increases (e.g., from 0.5%*Ip* to 5%*Pp*), while the

marginal gains attained by enriching more equally-mixed populations (above 40%*Ip*) crashed against the upper bound of near-purity (Fig. 3.8b).

As a result, relative enrichment for a 30%*Ip* population was found to be just 1.5-fold; however, this resulted in >90%*Pp* which cannot reliably be increased further. These findings suggest that bead selection may be performed multiple times, and that repeated rounds of enrichment will converge toward 100%*Pp*. For example, a rare population of 0.5%*Ip* can be reliably enriched to 5%*Pp* in one round, a total enrichment of 4.5% but a relative enrichment of 9-fold. This can be plated and expanded, and a second round can bring the new 5%*Ip* to 30%*Pp*, a total enrichment of 25% and relative enrichment of 5-fold. One more round increases 30%*Ip* to 85%*Pp*, a total enrichment of 55% but relative enrichment of just 1.83-fold. Further rounds have little marginal utility, as 100% (true purity) is a limit that can only be approached with this method.

Figure 3.8: Quantifying enrichment and depletion.

Pure wildtype and Mag-hNIL cells were mixed at various ratios, and the indicated protocol was performed as recommended. %*p* is shown for the indicated fraction, and prevalence was calculated using cell counts as in Table 2.8.

- (a) Enrichment demonstrated a rapid increase in the positive fraction for low %*Ip* and then levelled off as the positive fraction neared purity. The negative fraction demonstrated a slight decrease for moderate %*Ip* but was not substantially depleted.
- (b) Comparing relative enrichment to %*Ip* highlights increased relative efficiency for populations in which mApple+ cells are rare.

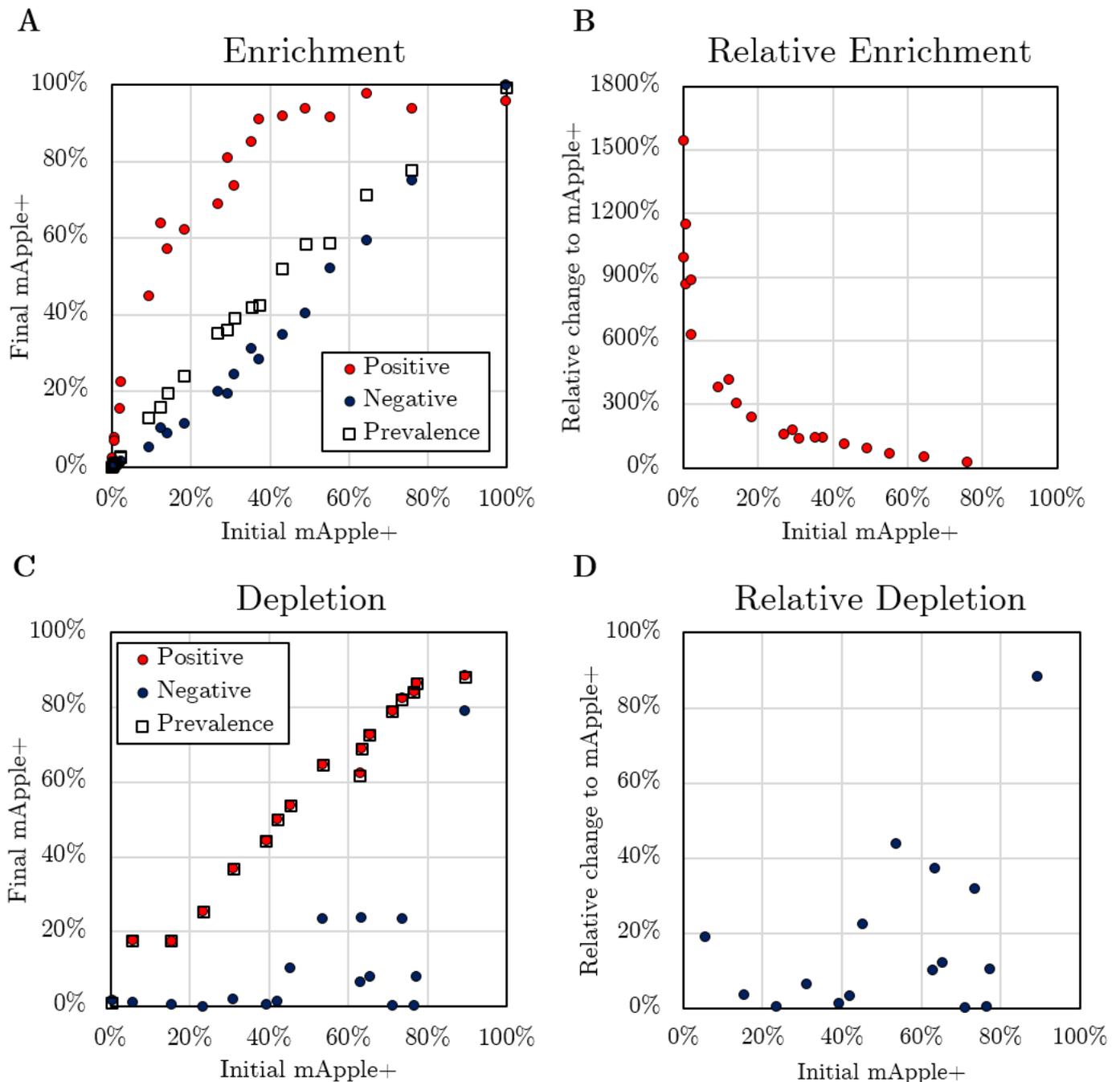
(continued)

Figure 3.8 (*continued*): Quantifying enrichment and depletion.

(c) Depletion demonstrated a rapid decrease in the negative fraction for high $\%Ip$.

The positive fraction recovered a majority of cells and showed little change from the initial population; note calculated prevalence and $\%Pp$ overlap.

(d) Comparing relative depletion to $\%Ip$ highlights high variability due to low cell density, but most samples were highly depleted of mApple+ cells.



3.4.3 Depletion efficiency

Similar observations were made with depletion, with the caveat that observed %*Ip* tended to be lower than intended when preparing populations, providing less data on high %*Ip*. In addition, low density in the negative fraction led to increased variability, which was reflected by the calculated prevalence skewing towards the positive fraction (Fig. 3.8c). However, the depleted fraction demonstrated a consistently low percent of mApple+ cells, even at 80%*Ip*. To use the previous workflow as an example, this would correspond to a 20% rate of Cre excision, which is easily achievable with transient Cre plasmid expression (Araki et al., 1997). Overall, peak relative depletion was reached in this 70-80%*Ip* range (Fig. 3.8d).

Decreased total depletion at higher %*Ip* might be expected due to beads being saturated by positive cells, leaving more unbound positive cells despite the increased bead volume called for by the optimized depletion protocol. However, increases to bead volume would become uneconomical and were shown to only marginally improve depletion efficiency while further reducing negative fraction cell density. Of more concern, negative fractions regularly accounted for less than 1% of the total cell population (i.e., recovery to the positive fraction >99%), reiterating that it may be practically advantageous to sacrifice depletion efficiency in exchange for more consistent results (e.g., by increasing initial cell density and performing one wash step).

3.5 Bayesian analysis of measured & derived values

3.5.1 Percentages as probabilities

The previous analyses intentionally simplified calculations by assuming that the measured % p for each fraction was a fixed value, known as a frequentist perspective. However, a more nuanced approach is to treat % p for each fraction as a random variable measuring probabilities, known as a Bayesian perspective (Turner et al., 2020). For example, the % Ip was explained above as the percent of positive cells in the initial fraction. Its analogous interpretation is that the measured % Ip is an estimate of the probability of picking a positive cell out of the total initial fraction at random; in other words, $p(mApple+)$. Following this line, the % Pp is the probability of picking a positive cell at random given that it was sorted to the positive fraction, or $p(mApple+ | P)$. Meanwhile, sensitivity can be described as the probability that any given positive cell is sorted into the positive fraction $p(P | mApple+)$, and specificity as the probability that any given negative cell is sorted into the negative fraction $p(N | mApple-)$.

As before, these definitions were kept consistent for both enrichment and depletion. While the change in perspective from percentages to probabilities does not affect the values or calculations directly, it does hold important implications for subsequent analyses and provides a deeper understanding of the relationships between experimental measures that will be explored in this section.

3.5.2 Confidence of flow cytometry data

By using flow cytometry on each fraction after selection, large numbers of cells can be sampled quickly, granting high confidence that the measured number of positive and negative cells reflect the true proportions in the full population. In particular, flow

cytometry sampling in this binary manner (positive or negative) is a Bernoulli process that should follow the hypergeometric distribution because it samples without replacement; however, proper modelling would require knowing the total number of positive and negative cells in each population *a priori*. Instead, sample and population sizes are large enough that a conservative estimate of a discrete confidence interval of each tested population can be reached using the binomial distribution.

Defining “success” as picking a positive cell, the measured $\%p$ of each fraction can be used as the probability of success in order to determine the maximum and minimum expected successes over a given number of trials, $\#t$. Using the cumulative binomial distribution, it is thus possible to calculate the maximum (or minimum) $\#p$ that bound the cumulative probability of a chosen critical value. For instance, 99.99% confidence intervals were determined using critical values of .99995 and 0.00005. In other words, there is a 0.01% probability that the observed $\#p$ would fall outside of this range. Simply dividing these maximum and minimum $\#p$ by $\#t$ converts this to a range of $\%p$ that the “true” $\%p$ is expected to fall within. As an aside, this is equivalent to a cumulative beta distribution with $\alpha = \#p+1$ and $\beta = \#n+1$. The beta is discussed in more detail below.

The binomial is also far more stringent than approximating confidence intervals using a normal distribution, and it has the added advantage of not assuming the distribution to be symmetrical (so it is valid even at values of $\%p$ near 0% and 100%). This stringency can be seen most clearly with small sample sizes; in particular, the fractions imaged in Fig. 3.3b were counted and the approximate $\%p$ was found. The binomial test shows a 95% confidence interval of 15.6-25.9%*I* p , 84.4-100%*P* p , and 11.5-24.5%*N* p . Meanwhile, a normal approximation (using the binomial standard deviation of $(\#t \times \%p \times \%n)^{1/2}$ for each fraction) yields the same 95% confidence interval of $20.6 \pm 0.4\%I$ p , $93.8 \pm 2.0\%P$ p and $18.0 \pm 0.7\%N$ p . Both show a $p < 0.05$ significant increase in the positive fraction, but the normal approximation also shows a $p < 0.05$ significant decrease in the negative fraction (both compared with the initial fraction). Both

confidence intervals shrink as sample sizes increase, but the binomial is always wider than the normal approximation; thus, to preserve stringency, confidence intervals and significance tests below were calculated with the binomial distribution.

3.5.3 Significance of selection

If enrichment or depletion is selecting cells at random, the probability of success after selection ($\%Pp$ or $\%Np$) would be the same as in the initial population ($\%Ip$). Additionally, if either protocol is worse than chance, the probability of success would decrease, which would run counter to the point of the protocol. Thus, the null hypotheses were that $\%Pp \leq \%Ip$ and $\%Ip \leq \%Np$ for each trial of enrichment and depletion, respectively.

For enrichment, the probability of the null hypothesis was determined by calculating the binomial probability that a population of $\%Ip$ would observe at least the number of successes in the number of trials performed as was actually observed in the positive fraction. This was equivalent to 1 minus the probability of the inverse case; namely, of observing any number up to $\#Pp-1$ positive cells in $\#Pt$ trials. Thus the p -value is 1 minus the cumulative binomial distribution for $\#Pp-1$ successes in $\#Pt$ trials. For all samples except for the control case of $100\%Ip$, this test returned $p < 0.0001$.

For depletion, a similar test can be applied by calculating the cumulative binomial distribution up to the observed $\#Np$, given $\#Nt$ trials and a probability of $\%Ip$. As in enrichment, all samples demonstrated $p < 0.0001$ except for the control case of 0% initial positive cells and one condition with the lowest measured $\%Ip$ of 5.6% . The measured $\%Np$ for this sample was 1.1% , and the binomial p -value is just below 0.0014 . Despite there being little room for improvement, depletion was still successful by standard definitions of significance and improved relative purity by over 80% .

Together, these data allow the null hypotheses to be rejected with a high degree of confidence, even at very low or very high $\%Ip$. Furthermore, the failure of completely pure control populations to pass this significance test underscores its utility and provides increased confidence in the implicit assumptions that the control populations are actually pure and that flow cytometry is a gold standard.

Finally, while there are many reasons why the assumptions underlying the T-test make it inappropriate, including that the data is not expected to follow a normal distribution as discussed above, it is possible to normalize the $\%p$ for each fraction by simply dividing by $\%Ip$, and to perform a one-sample, paired two-tailed T-test as a basic assessment of whether $\%p$ is changing. Using the samples shown in Fig. 3.8, both enrichment and depletion demonstrate $p < 0.001$.

3.5.4 Derived test characteristics

Sensitivity and specificity can be derived in the frequentist perspective by directly combining cell counts as in Table 2.8; however, this method relies on the assumption that sample density was normalized by sampling equal volumes and by properly correcting for fraction volume; this worked well only when recovery was high because stochastic differences in the sampling rates of sparse samples can be ignored. The Bayesian perspective provides an additional method to validate these calculations that does not require direct comparisons of different fractions but instead only compares proportions within each fraction and the initial population (Branscum et al., 2005; Drewe et al., 2010; Maceneaney & Malone, 2000; Vilar et al., 2015). This removes the need for a sample density assumption and for a volume correction factor. Using both methods, a higher degree of confidence can be gained if they correspond, while issues with the protocol can be identified if they vary; for instance, in near-pure populations or when low recovery led to skewed sample sizes.

To derive the Bayesian method, it is helpful to start with the traditional form of Bayes' theorem, where A and B are some events:

$$(1) \quad p(A | B) = p(B | A) \times p(A) / p(B)$$

This states that the probability that A happens given B is equal to the probability that B happens given A multiplied by the independent probability of A and divided by the independent probability of B .

Next, define A as picking a positive cell at random, and B as a cell being assigned to the positive fraction. Then, the probability that a cell is positive given that it was assigned to the positive fraction ($\%Pp$) is equal to the probability of a cell being positive ($\%Ip$) multiplied by the probability that a cell is assigned to the positive fraction given that it is positive (se) divided by the probability of any given cell being selected to the positive fraction, or the percent of cells recovered. However, recovery also relies on comparing cell counts rather than only ratios within individual fractions, so this denominator can be further broken down into the cases of selection; that is, the $\%Ip$ multiplied by the probability of selection given a cell is positive (se), added to the $\%In$ multiplied by the probability of selection given the cell is negative (the FPR , or $1-sp$). In summary:

$$(2) \quad \%Pp = (\%Ip \times se) / (\% \text{ recovery})$$

$$(3) \quad \%Pp = (\%Ip \times se) / (\%Ip \times se + \%In \times (1-sp))$$

While intuitive, this construction does not allow for separation of se and sp for analysis. Instead, dividing both sides of the formula by the equivalent formula for the $\%Pn$ (with numerator $\%In \times (1-sp)$ and the same denominator) yields the odds form of Bayes' theorem, in which the odds of picking a positive cell are substituted for probability. In this way, denominator of the formula ($\% \text{ recovery}$) and of each probability ($\#t$) is canceled out, and what remains is the number of positive cells divided by the number of negative cells, precisely the odds. These are referred to by fraction as $O(\text{Initial},$

Positive, or Negative). Furthermore, the same formulae can be derived for the negative fraction, so Bayes' theorem simplifies to:

$$(4) \quad O(P) = O(I) \times se / (1-sp)$$

$$(5) \quad O(N) = O(I) \times (1-se) / sp$$

Thus, sp and se can be isolated (as positive and negative likelihood ratios, respectively), dependent wholly on one condition compared to the initial. Finally, isolating for se and sp directly:

$$(6) \quad se = (O(P) \times (O(I) - O(N)) / (O(I) \times (O(P) - O(N)))$$

$$(7) \quad sp = ((1-se) \times O(I)) / O(N)$$

Substituting the odds of each fraction with the definition $\%p / (1-\%p)$ and simplifying further yields:

$$(8) \quad se = (\%Pp \times (\%Ip - \%Np)) / (\%Ip \times (\%Pp - \%Np))$$

From this, sensitivity can be calculated directly, and specificity can then be calculated using formula 7.

3.5.5 Characteristics of enrichment

It was previously noted for enrichment that the negative fraction contained a large number of positive cells, suggesting that it had a high false negative rate, because there was a high probability that a positive cell chosen at random was assigned to the negative fraction. It was thus expected that enrichment had a low sensitivity, given that $FNR = 1-se$. Meanwhile, strong selectivity was observed in the positive fraction, suggesting that there was a low false positive rate. Likewise, if there were few false positives, the rate of negative cells being sorted into the negative fraction was high, so it was expected that enrichment had a high specificity ($FPR = 1-sp$).

Indeed, sensitivity was relatively low and specificity was high across a range of $\%Ip$ as calculated both through cell counts and the Bayesian method described above (Fig. 3.9a). Interestingly, sensitivity tended to decrease as $\%Ip$ increased, while specificity remained high throughout. This may account for the high relative enrichment in low $\%Ip$, and decreasing sensitivity may be in part due to increasing competition for bead binding sites among mApple+ cells. Luckily, specificity was maintained at high $\%Ip$, indicating that negative cells had very little bead-binding capacity. When positive cells were sparse, there were more beads to go around, so competition decreased.

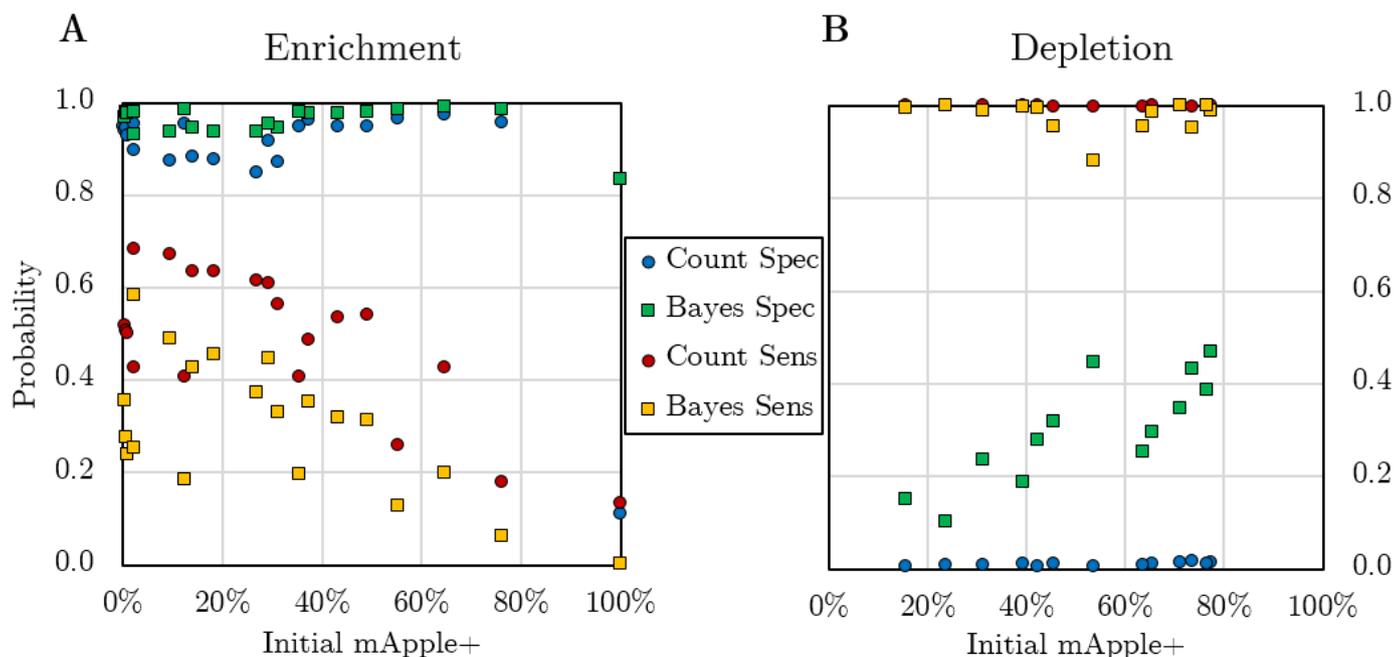
Sensitivity showed loose negative linear correlation to $\%Ip$ for both calculation methods (Pearson's $R = -0.71$ for cell counts and -0.66 for Bayesian) with approximately equal slope (-0.39 and -0.35 , respectively). Where they differed, however, was that the cell count sensitivity was always higher than the Bayesian, indicating a positive skew to cell counts even with the volume correction. Reapplying a volume correction of 7 brought the two sets into near-perfect alignment, dropping the mean square difference to just 0.18% rather than 3.73% as seen for a volume factor of 3. Simultaneously, increasing the volume factor raises the measured specificity into better alignment with the Bayesian specificity.

There are several possible reasons why the two measures differed; for instance, some volume of buffer was often carried over between wash steps to prevent removing beads, relative volumes used could vary by pipetting error or evaporation, or cells that were bound to beads could passively separate in solution from their less dense unbound counterparts and thus affect sampling rates during flow cytometry. This is not to say that using a volume factor of 3 here was incorrect, nor should the volume factor be set afterwards based on the fit of the data. Indeed, the Bayesian sensitivity may be systematically underestimating its value for different reasons. However, the insulation of the Bayesian calculations from differences in selected fraction density and sample size lends increased weight to these values.

As discussed in the protocol optimization experiments above, specificity is the most important factor for improving purity in the positive fraction. While the loss of positive cells through low sensitivity can reduce overall recovery, this result was still preferable to the retention of negative cells through low specificity, especially at low %Ip when negative cells already outnumbered the positives. Particularly if clonal isolation was desired, technically only a few dozen positive cells needed to be recovered out of potentially millions of total cells in the initial population in order to support subsequent genotyping and experimentation. It was thus reassuring to know that specificity remained high across all tested samples and showed little dependence on %Ip.

Figure 3.9: Derived test characteristics by %Ip.

Sensitivity (Sens) and specificity (Spec) as calculated using a frequentist approach using cell counts (Count) or through a Bayesian approach (Bayes) for enrichment (a) and depletion (b). Note overlapping data points for depletion sensitivities.



3.5.6 Characteristics of depletion

It was previously noted that, following depletion, the positive fraction contained many negative cells, suggestive of a high false positive rate and thus low specificity, and that the negative fraction contained few positives, equivalent to high sensitivity. It was thus expected that the opposite of enrichment was true for depletion; namely, that sensitivity was high and specificity was low, and the data supported these predictions (Fig. 3.9b).

In the data collected here, high variability between Bayesian and cell count values was evident due to the much smaller relative samplings for the negative fraction. As mentioned, the negative fraction often made up less than 1% of observed cells, resulting in high variability in the Bayesian calculations and pushing the cell count values towards $se = 1$ and $sp = 0$. The Bayesian calculations, being insensitive to cell density, were more balanced and suggested an upward trend in specificity as $\%Ip$ increases (Pearson's $R = 0.82$, slope = 0.45) while sensitivity remained high throughout. More samples would be needed at high and low $\%Ip$ for this relationship to be robust, though.

Logically, selection only exhibits pressure towards the positive fraction, actively retaining bead-bound cells via magnetization, while the negative fraction is filled passively with those cells that are unaffected by the magnet. This suggests that the primary action of depletion is not to actively select for negative cells, but rather to exclude positive cells, and the high observed sensitivity supports this. Further, low specificity tracks with the observation that selected fractions are often sparse, because in this case all cells are preferentially sorted into the positive fraction.

As with enrichment above, the purity of the selected negative fraction was greatest when the opposite type of cell was blocked from it, rather than by attracting more of the desired cells at the expense of general permissiveness. The ideal scenario for depletion was that an excess of bead binding sites existed, saturating the positive cells, and the protocol was intentionally designed to promote this. Regardless of how many

mApple⁻ cells were pulled into the magnetic pellet through nonspecific streptavidin binding or by simply being surrounded by nearby mApple⁺ cells, the first supernatant was largely made up of mApple⁻ cells that remained unbound to beads.

While enrichment benefited from greater cell number to counteract the resulting loss of recovery, depletion did not. Consistently low cell density in the negative fraction reduced the utility of the protocol and confounded analysis, so it would be useful for future experiments to intentionally sacrifice some sensitivity in exchange for greater specificity; as recommended above, density may be improved by loading more cells or by combining the supernatants of one or two wash steps in order to dislodge additional negative cells.

3.5.7 Confidence of sensitivity and specificity

Another advantage of the Bayesian perspective is that it allows modelling of sensitivity and specificity as random variables themselves, and thus confidence in their values can be determined (Mossman & Berger, 2001; Srinivasan et al., 2012). The beta distribution was mentioned above as an alternative to the binomial test that directly computes the confidence interval of sampling in terms of percentages rather than cell counts because the beta directly uses probability as its input parameter rather than the number of total trials ($#t$) with a given probability of success.

Instead, the beta incorporates test data via its two shape parameters, α and β . For binomial confidence intervals, these are equivalent to the number of successes (positive cells) and failures (negative cells), respectively, but to apply this method to the calculated specificity and sensitivity, these shape parameters can be calculated with the observed mean (μ) and variance (σ^2) of the data via the following relationships (Ferrari & Cribari-Neto, 2004):

$$(9) \quad v = \alpha + \beta = (\mu \times (1-\mu) / \sigma^2) - 1$$

$$(10) \quad \alpha = \mu \times v$$

$$(11) \quad \beta = (1-\mu) \times v$$

This yielded the respective values for α and β , from which the 2.5th and 97.5th percentiles were calculated to give 95% confidence intervals (Table 3.2). Parameters calculated in this way can be thought of as a synthetic number of successes and failures; in $\alpha+\beta$ cells of a given type (positive for sensitivity, negative for specificity), the expected number of cells correctly sorted is α , and the expected number incorrectly sorted is β . The scaling of α and β additionally reflected the variability observed; as $\alpha+\beta$ increased, confidence intervals narrowed. This analysis enabled better comparisons between the sensitivity and specificity calculated in either manner, and demonstrates that enrichment specificity was high in and depletion sensitivity was high for all %Ip.

Table 3.2: Summary of key derived characteristics.

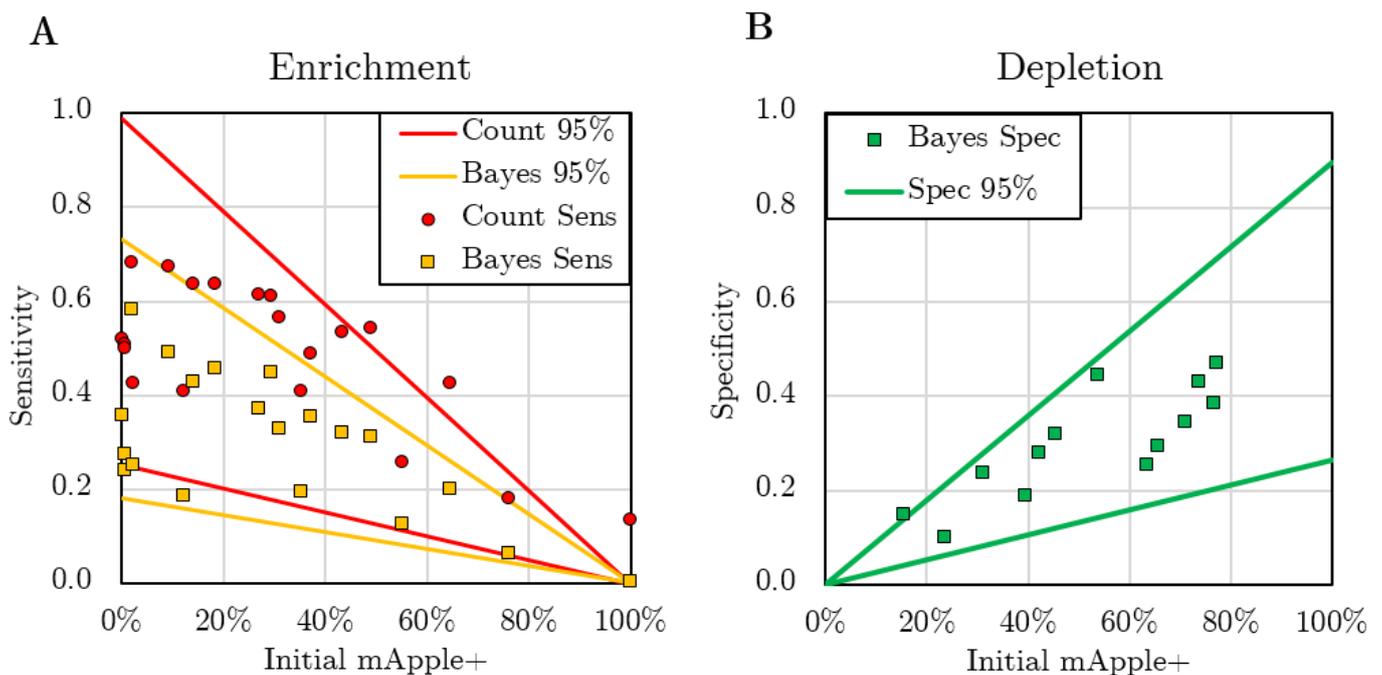
Sensitivity and specificity for enrichment and depletion protocols were determined by cell count (frequentist) or Bayesian methods for each trial. Values shown are the mean scores as measured across all %Ip. Confidence intervals were calculated using the beta distribution with shape parameters α and β as shown, without taking potential %Ip dependence into account.

		Cell Count			Bayesian		
Protocol	Factor	Mean (95% CI)	α	β	Mean (95% CI)	α	β
Enrichment	Sensitivity	.525 (.314-.731)	10.74	9.73	.330 (.126-.576)	4.99	10.14
	Specificity	.925 (.837-.981)	44.07	3.55	.966 (.918-.993)	81.44	2.91
Depletion	Sensitivity	.999 (.998-1.00)	2204.63	0.96	.977 (.884-1.00)	20.30	0.48
	Specificity	.011 (.001-.018)	10.49	988.24	.300 (.111-.536)	4.90	11.41

However, this did not account for the linear correlation between $\%Ip$ and sensitivity in enrichment or between $\%Ip$ and specificity in depletion, instead assuming one flat value for all $\%Ip$. It was possible to account for this by dividing sensitivity by $\%In$ for enrichment and dividing specificity by $\%Ip$ for depletion. A new confidence interval could then be calculated, and linear bounds of this non-constant confidence interval could be determined by multiplying these bounds by $\%In$ or $\%Ip$ respectively (Fig. 3.10).

Figure 3.10: Adjusted 95% confidence intervals by $\%Ip$.

- (a) Enrichment sensitivity based on cell count was found to lie between 25.4% and 99.1% of $\%In$. Using Bayesian calculations, the same was found to be between 18.2% and 73.3% of $\%In$. Thus, sensitivity improved but was also more variable at low $\%Ip$.
- (b) Depletion specificity using Bayesian calculations was found to lie between 26.4% and 89.6% of $\%Ip$. Specificity by cell count was deflated by low cell density in the negative fraction and thus could not be measured accurately.



3.6 Modelling enrichment and depletion

3.6.1 Receiver operating characteristic (ROC)

By plotting the true positive rate (se) as a function of the false positive rate ($1-sp$), a sense of the quality of selection can be determined via the receiver operating characteristic (Fawcett, 2006; Hajian-Tilaki, 2013). Any test falling along the diagonal is no better than chance, while the upper left corner represents perfect sorting, correctly assigning every cell into its respective fraction. It is evident that bead selection is better than chance but is also imperfect, showing a tradeoff between sensitivity and specificity at varying $\%Ip$ and using different test parameters (Fig. 3.11). Optimization steps effectively separated enrichment and depletion in ROC space, but both tests observed similar trends: samples with high $\%Ip$ tended to be in the bottom left of their respective populations, and as $\%Ip$ decreased from 100%, samples tend to move up and to the right from $(0, 0)$ to $(1, 1)$ along a similar curve. Enrichment can thus be thought of as a process to move samples towards the bottom left, while depletion moves samples towards the upper right.

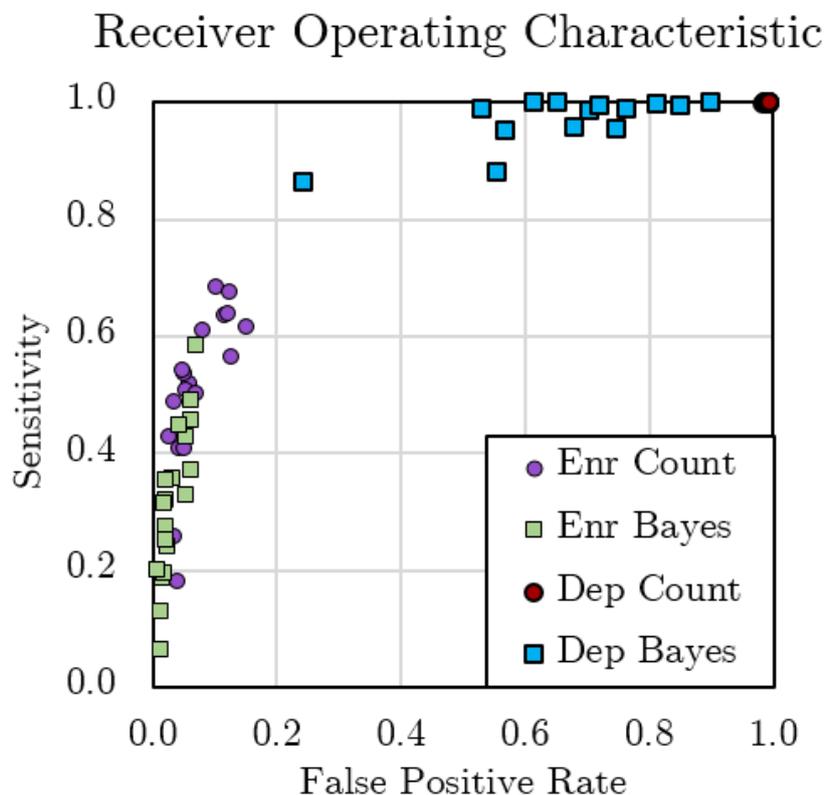
Enrichment is designed to be a strict process, admitting cells into the positive fraction only after rigorous testing. Cells must bind to beads quickly in a competitive environment, and this binding must be strong enough to survive at least three washes. The ideal for enrichment is along the left wall of the ROC plot, with no false positives, and anything that improves the number of true positives is a bonus. In other words, the cost of false positives is considered high, while the cost of false negatives is considered negligible. Enrichment samples are broadly along the left wall as desired, only giving up some false positives when $\%Ip$ is very low.

Depletion by contrast weights false negatives as more costly, and thus it was designed to be a permissive test, admitting as many positive cells into the positive fraction as possible. In doing so, it is leaving behind in the negative fraction only those that are

most likely to be negative. In terms of ROC space, depletion has its ideal along the ceiling, where all positives are assigned to the positive fraction. In some respects, optimization was too successful, admitting >99% of all cells into the positive fraction and leaving so few in the negative as to cluster nearly all samples as calculated by cell counts in the upper right corner. The Bayesian calculations were a bit more forgiving, running tightly along the top as %Ip increases, reflective of the maintenance of high sensitivity for all %Ip.

Figure 3.11: ROC of enrichment & depletion.

The false positive rate and sensitivity were plotted for all trials of the optimized enrichment (Enr) and depletion (Dep) protocols. Samples tended to fall along a curve clustered by protocol, with high %Ip closer to (0,0) and low %Ip closer to (1,1). Maximum balanced accuracy was achieved at approximately (0.2, 0.8), but optimization to minimize the FPR for enrichment or maximize sensitivity for depletion separated the protocols. Using trapezoidal estimation, the area under the curve was calculated as approximately 0.871.



3.6.2 Area under the ROC curve

While it was unclear if the same curve applied to both enrichment and depletion given the differences in parameters and separation in ROC space, the lack of overlap between the two samples suggests that a single curve could be used to model the overall area under the curve, and thus provide insight into the quality of bead selection more generally. The area was estimated with the trapezoidal column method, in which all points were sorted by FPR, and the points (0, 0) and (1, 1) are appended as FPR_0 and FPR_n respectively. The plot was then split into columns of width $FPR_k - FPR_{k-1}$ for all $0 < k \leq n$, this width was multiplied by the average height $(se_k + se_{k-1})/2$ for each k , and the sum of all areas was calculated. The area under the curve was thereby estimated to be 0.871, with a lower bound of 0.839 and an upper bound of 0.903 using the minimum and maximum se values for each column respectively. Because random chance is represented by a diagonal line, any area greater than 0.5 corresponds to an increased probability of correct fraction assignment, with greater area suggesting better sorting. The area calculated here was thus indicative of strong sorting for both enrichment and depletion.

There are two ways to improve selection: to move along the curve by adjusting the variables, or to improve selectivity itself, pulling the entire curve closer to (0, 1). None of the optimization experiments tested were shown to be capable of changing the curve dynamics directly; rather, they served as methods of prioritizing the purity of either fraction. Improvements to the SBP- Δ LNGFR tag, beads, or magnet may improve the curve, and the data shown here may be useful for benchmarking updated protocols. It may be further possible to design experimental parameters to act as functions of $\%Ip$ rather than as fixed values. Such a program would recognize the correlation between initial purity and sensitivity (in enrichment) or specificity (in depletion), and could seek to adjust the protocol to make up for differences in efficiency and make results more predictable.

3.6.3 Modelling selection as a function of %Ip

Another interesting result of determining the efficiency of the optimized protocols was that the data were found to be highly regular, so I sought to make a mathematical model to quantify the efficacy of bead selection and predict the results of future experiments. Such a model would make it easier to compare this protocol with other selection methods or to adjust additional variables. It would also facilitate experimental planning; for example, an HDR-mediated insertion with a given integration efficiency could be used to determine the number of rounds of selection needed to make this population at least as pure as desired. However, the data do not follow a clear linear, logistic, or exponential trend, so simple regression methods were not effective and a different model was sought.

First, several specifications for such a model were determined. Because %p is defined as a proportion of the total population ($\#p/\#t$), data for %Ip, %Pp, and %Np lie continuously on the range [0, 1]. While alterations to the protocol, including differences in the insert plasmid or resultant expression of SBP- Δ LNGFR, would be expected to change the efficiency of the protocol, an ideal model would require only minor alterations to its parameters to account for these changes (and not alteration of the model as a whole). It should also be solely dependent on the %Ip so long as all other variables remain equal (in other words, the model should be univariate), such that all other parameters are represented as constants. Additionally, a pure initial population would remain pure regardless of bead selection, so the model should intersect $y=x$ at (0, 0) and (1, 1). Both protocols also feature a rapid change in purity near the opposite extreme (very low %Ip for enrichment and high %Ip for depletion), followed by a strong ceiling/floor effect suggesting asymptotic limiting.

The beta cumulative distribution function (CDF) was identified because it fulfils the stated conditions; while it is used above for calculating confidence intervals, it is also commonly used to model relationships between probabilities, for instance in

survivorship studies and population genetics (Balding & Nichols, 1995; R. W. Johnson, 2013). In this context, the shape parameters α and β are not as readily interpretable as expected successes and failures, but rather are related to the rate of enrichment or depletion because the input variable is $\%Ip$ and the output is either $\%Pp$ or $\%Np$. The percent of positive cells is again used for depletion for consistency; the variables are easily transformed by the relationship $\%p = (1-\%n)$, so the $\%n$ may be visualized by simply rotating the $\%p$ graphs by 180 degrees.

While the beta is extremely versatile, the fact that the data underlying each point can also be expressed with its own beta is suggestive of a deeper relationship. The input variable of this model, $\%Ip$, is also the probability parameter for each individual binomial distribution used to calculate significance and confidence intervals previously. As a conjugate prior distribution of the binomial, the beta distribution is thus ideally suited to model the expected posterior probability of success ($\%Pp$ or $\%Np$) based on the prior probability of success ($\%Ip$).

To calculate the shape parameters, the same mean and variance method was applied as used above for determining confidence intervals of sensitivity and specificity. For the mean, the $\%Ip$ for which the $\%Pp$ or $\%Np$ is expected to be 50%, termed I_{50} , was estimated for both enrichment (16%) and depletion (80%) based on extrapolation from nearby data points. Next, the variance, termed V , of enrichment (2.25%) and depletion (1.69%) were calculated as the square of the standard deviation of the total enrichment or depletion. Using these values, (α, β) was calculated to be (0.8, 4.18) for enrichment and (6.77, 1.69) for depletion. A null model can also be made as a function that follows the diagonal, which can be obtained with $(\alpha, \beta) = (1, 1)$ using an $I_{50} = 50\%$ and $V = 8.35\%$. While V in this case is not referring to a change in $\%p$, it is reflective of a wider dispersion of values than in the selected fractions, which tend to change rapidly but then level off near purity.

Plotting these functions alongside the data shows a high degree of correspondence (Fig. 3.12a-b). Comparing the observed $\%Pp$ and $\%Np$ to the model's predicted values demonstrates strong linear correlation and low overall error (Pearson's $R = 0.993$ and mean squared error = 0.003 for enrichment; Pearson's $R = 0.753$ and mean squared error = 0.019 for depletion; Fig. 3.12c). While high variability and a paucity of high $\%Ip$ samples reduces the confidence in the depletion model, the ability of the same method to accurately predict values in enrichment suggests that additional replicates should be expected to fall near the model line. The nature of the beta model as a conjugate prior also admits the addition of future replicates to refine this model further.

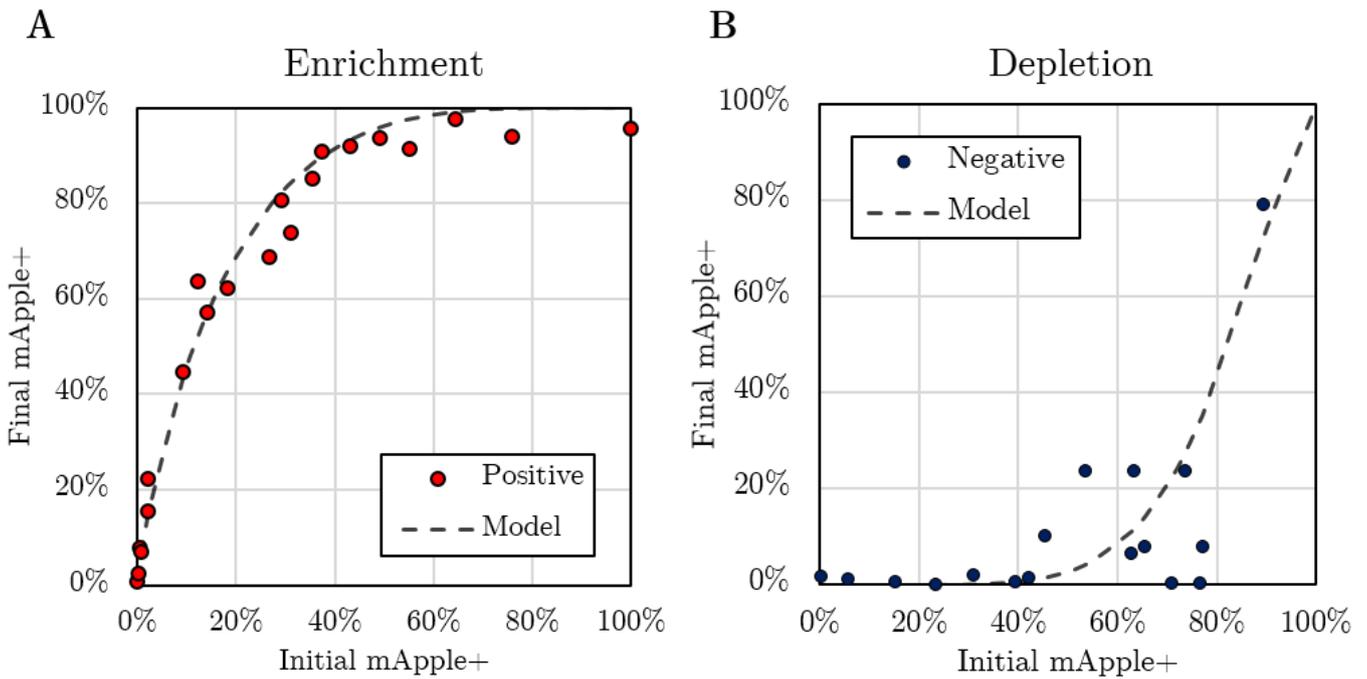
3.6.4 Area under the beta model curve

The cumulative beta distribution can be integrated from 0 to 1 to directly yield the area under the curve. Because the diagonal represents no change to $\%p$ and thus no selection, the difference of the calculated area from 0.5 can be used as a measure of the quality of enrichment (if increased) or depletion (if decreased). The enrichment model yielded an area of 0.839, while depletion yielded an area of 0.200 (Fig. 3.12c). Using the trapezoidal column method on the points also provided consistent values (0.808 for enrichment and 0.099 for depletion). Together, and alongside data shown previously, these calculations serve to validate that enrichment and depletion are useful and effective protocols.

Figure 3.12: Modeling enrichment and depletion.

Selection data were modeled with cumulative beta distributions. The approximate initial fraction such that the final % mApple+ is 50% (I_{50}) was used as the mean and the approximate variance of the difference between initial and selected fractions (variance of enrichment/depletion, V) was used as the variance in order to calculate the parameters (α, β) of the distributions.

- (a) Enrichment data for the positive fractions (Positive) and path of the model.
- (b) Depletion data for the negative fractions (Negative) and path of the model.
- (c) Parameters used to calculate the above distributions as well as a null model along the diagonal are shown. Pearson's correlation coefficient (R) and mean squared error (MSE) are calculated between the model and data for all points, and area under the curve (AUC) is calculated by integrating the beta cumulative distribution function from 0 to 1.



C

Model	I_{50}	V	α	β	R	MSE	AUC
Enrichment	16%	2.25%	.796	4.18	.993	.003	.839
Depletion	80%	1.69%	6.77	1.69	.753	.019	.200
Null	50%	8.35%	1	1	1	0	.500

3.7 Discussion

3.7.1 Summary of findings

The data presented here establishes SBP- Δ LNGFR as an effective and versatile selection gene for transgenic iPSCs, capable of both positive and negative selection with magnetic streptavidin-coated beads. Regardless of initial purity, nearly-pure positive or negative populations of non-clonal cells could be achieved in at most three rounds of enrichment or depletion, and the entire protocol could be completed in under one hour. Unlike FACS, this process can be done without expensive, specialized equipment, and unlike antibiotic and TK selection, it does not require extensive tuning for each application. Since this work was completed and originally published, the area of bead-based cell sorting has continued to grow (Gao et al., 2018a; Sutermaister & Darling, 2019). While new bead products have been developed, they are primarily marketed for antibody-conjugated surface markers and for fractionating blood cell types, and there is not a one-size-fits-all solution. This work will hopefully encourage and facilitate the development of further products explicitly targeted towards use with iPSCs in order to ensure sterility and to improve recovery and survival.

3.7.2 Enrichment and depletion in practice

Even before the protocol was optimized, a single round of enrichment from a moderately mixed population yielded a positive fraction with high purity and good survival after direct plating without detaching or removing the beads. It is likely that biotin present in the media helped to passively detach beads once cells were returned to culture, but nonetheless it was surprising that there was no appreciable toxicity or spontaneous differentiation while culturing cells directly with beads.

While no head-to-head comparison of the efficacy of bead selection, FACS, and antibiotic selection was performed here, in my experience FACS is often capable of purifying a mixed population in a single sort (as in Chapter 5.4.1), and a few days of puromycin selection usually results in a nearly pure population. Compared with these established selection methods, the bead selection presented here was generally more permissive; enrichment had low sensitivity, while depletion had low specificity. In contrast, FACS can recover iPSCs within a positive (or negative) gate more effectively, albeit typically with low survival. Antibiotic selection also has no true recovery step and thus is limited only in its capacity to kill negative cells while allowing positive cells to proliferate (which further increases apparent efficacy). Despite this, I contend that the flexibility, reliability, and ease of the bead sorting protocols, coupled with the additional functionality of the SBP- Δ LNGFR tag, makes it a viable alternative to (and synergistic partner of) fluorescent proteins and antibiotic resistance genes.

While the Mag-hNIL plasmid was used extensively in this chapter for protocol development, I later found that the function of rtTA3G was disrupted, as shown in detail in Chapter 4. I hypothesized that this loss of function was due to the same promoter being used to also express the selection cassette, so trafficking of the SBP- Δ LNGFR protein to the extracellular membrane may have decreased cytoplasmic availability of rtTA3G. However, mApple was evident throughout the cell, so future work could seek to better characterize the trafficking and translation of mRNAs including SBP- Δ LNGFR and to assess whether IRES linkers permit better bioavailability of downstream genes than 2A cleavage peptides.

To counteract the loss of rtTA3G activity, I constructed and evaluated Mag2-hNIL, which dedicated the CAG promoter to only express rtTA3G while introducing the EF-1 α promoter for the selection genes, as described in Chapter 2. I found this plasmid to be functional for differentiation while still permitting enrichment and depletion with equivalent efficacy, as shown in Chapter 4. Mag2-hNIL is available from Addgene (105842) for use directly in the i³LMN differentiation system or for excision of the

SBP- Δ LNGFR tag for the purpose of inserting it into other constructs. However, it must be stressed that the SBP- Δ LNGFR tag should only be used with a dedicated constitutive promoter and not be included in polycistronic cassettes.

3.7.3 Comparison of analysis approaches

In this chapter, I showed how the frequentist and Bayesian perspectives can both provide valid measures of sensitivity, specificity, and other derived test characteristics. Moreover, I strove to emphasize how both approaches can work together to improve confidence in results and provide clarity for analyses. Consequently, I recommend taking both approaches forward to future work in this area since they are both relatively simple to calculate and each augments the shortcomings of the other. With this in mind, the remainder of this section will summarize the notable differences between the approaches and postulate instances in which one is preferable to the other.

The frequentist approach is perhaps the more intuitive method, since cells can be more easily conceived as individual, physical objects than as a conceptual random variable which applies to the whole population. Likewise, once the cells are assayed, it is simpler to think of each of them as assigned “positive” or “negative” for a fluorescent marker than it is to imagine the underlying probability that the next cell is positive. Additionally, the frequentist approach does not require the separate sampling of the initial population, since calculating sensitivity and specificity using cell counts relies on the calculated prevalence as the “true” total percent of positive cells without regard to the initial fraction, while the Bayesian method only uses the measured $\%Ip$. Where these two measurements differ is where differences in the two methods of calculating se and sp arise. In an experiment where measuring the initial population is impractical or would drastically increase the number of conditions, the frequentist may be the approach of choice.

When it is possible to measure each population, however, I contend that the Bayesian has an edge. While frequentist calculations require that the relative volume sampled from both the positive and negative populations be equal, leading to corrective measures such as the volume factor v , these considerations are irrelevant to the Bayesian approach. The Bayesian is therefore ideal for situations in which population frequencies are known but relative count numbers are not, or if there is reason to doubt that sampling volumes have been properly normalized. While both approaches are sensitive to stochastic sampling discrepancies, the Bayesian was also shown to be more permissive of lopsided sampling numbers and is therefore more appropriate for comparing populations of different sizes.

Using the Bayesian method, sensitivity and specificity are not necessarily bounded by 0 and 1 and thus can give “impossible” results. If both $\%Pp$ and $\%Np$ are measured as greater than $\%Ip$, se is negative and sp is greater than 1, while the opposite is true if both $\%Pp$ and $\%Np$ are measured as less than $\%Ip$. This is not to suggest that any cells are being created or destroyed out of whole cloth, but rather that the Bayesian method can be readily used as a diagnostic tool for the test itself (i.e., to identify stochastic issues with sampling). The only samples that returned impossible values for sensitivity and specificity in this work were the 0% and 100% pure control samples, while the same samples gave highly variable, but not technically erroneous, values by the frequentist method.

A potential caveat to the Bayesian is that it cannot support actually pure populations, since a value of 0% positive or negative in any of the fractions results in dividing by zero during calculation. The frequentist method is not affected, simply returning the suspect (but not technically erroneous) value of 1 or 0 for the respective characteristic. As with the impossible values above, this should be considered a feature of this method rather than a bug, as it more readily highlights suspect samples, for instance with poor recovery, in which cells of only a single type are sampled.

In general, this chapter used probabilistic modelling to understand the processes underlying selection. This is not an approach with a direct comparison in the literature, as most of the analyses used here were instituted by logical extension of the concepts rather than by direct application of existing methods. For instance, the beta distribution was ideally suited to model the data in Chapter 3.6.3, but I am not aware of it being used previously to assess the efficacy of a selection process. A larger meta-analysis of the methods used to assess selection processes in the literature may therefore be useful to highlight discrepancies and recommend best practices.

3.7.4 Future directions

This work demonstrated that iPSCs expressing SBP- Δ LNGFR can be labeled with fluorophore-conjugated streptavidin, enabling this tag to be substituted in any technique for which fluorescent protein marker genes are normally used, such as for imaging to estimate purity, marking clones for picking, or for selection via FACS. Labeling in this way avoids the disadvantages of fluorescent proteins by being transient, thereby maintaining downstream fluorescence channels, and by being extracellular, thereby minimally affecting internal cell biology. While the tag does limit the use of biotinylation-based downstream applications such as proximity labeling, it can be removed via Cre expression and negative selection, and the depletion protocol facilitates this process.

A potential disadvantage that cannot be so easily dismissed is the digestibility of SBP- Δ LNGFR with common dissociating enzymes, which may limit its use in more strongly-adherent cell lines such as HeLas that are not as easily singularized with only EDTA treatment. Matheson et al. (2014) report the use of a non-enzymatic dissociation buffer in HEK 293T cells, but this also required filtering to remove cell clumps. In these instances, further work would need to be done to either alter the SBP- Δ LNGFR

construct to be resistant to these enzymes or to develop better non-enzymatic cell singularization techniques.

Finally, the analyses developed and applied here facilitate common measures of benchmarking against other selection methods or alternative protocols in order to assess the performance of each in various situations. Calculating and comparing the sensitivity and specificity of both positive and negative selection strategies via both frequentist and Bayesian interpretations of the data provided insight into the mechanisms affecting selection processes in general and serve to identify target areas for improvement. For instance, it is likely worthwhile to sacrifice sensitivity in depletion in exchange for increased cell density in the negative fraction. ROC analysis revealed that enrichment and depletion are well-optimized on opposite ends of a continuum that differentially prioritizes specificity and sensitivity, respectively, and modelling of the selection processes with cumulative beta functions helped to predict results of future trials. Both approaches to modelling also serve to underscore the high efficacy of the selection protocols via their respective areas under the curve.

The modelling performed here takes a broadly top-down perspective, analyzing the efficacy of selection as a function on $%Ip$. A potential area for future research is to apply a bottom-up approach to ascertain the binding affinity of individual beads to SBP- Δ LNGFR constructs as well as to investigate the characteristics of bead interactions generally. Such an approach could apply dynamical systems modelling or methods used in biochemistry to model molecular interactions, such as the Hill-Langmuir equation which is designed to model the affinity between a receptor and ligand. This would enable better comparisons between different types of beads; those with high affinity for SBP- Δ LNGFR but low affinity for other surface proteins would be preferable. Such an approach would also be useful for developing new generations of SBP- Δ LNGFR constructs, perhaps designed to increase streptavidin affinity or bioavailability or to resist protease digestion.

Chapter 4: Derivation of AR polyQ allelic series to model SBMA

4.1 Introduction

4.1.1 Spinal and bulbar muscular atrophy

Spinal and bulbar muscular atrophy (SBMA) is a disease that causes progressive skeletal muscle weakness and degeneration of lower motor neurons with typical onset in mid-adulthood (Fischbeck, 2016). As mentioned in Chapter 1, SBMA is caused by a polyglutamine expansion in the androgen receptor (AR) protein, and clinical manifestations also include signs of androgen insensitivity, such as gynecomastia and reduced fertility, suggesting that the normal function of AR is impaired (La Spada et al., 1991). Notably however, a full genetic knockout of AR does not cause neurodegeneration, indicating that SBMA is caused by a toxic gain-of-function effect rather than a loss of AR signaling. While disease may resemble ALS, SBMA progresses more slowly, typically leading to decreased ambulation, dysarthria, and dysphagia over many years which may necessitate long-term care (Guber et al., 2018).

As a steroid receptor, AR is activated by binding to an androgen ligand, particularly the testosterone derivative 5 α -dihydrotestosterone (DHT). While inactive AR is restricted to the cytoplasm, activated AR translocates to the nucleus, where it binds DNA and functions as a transcription factor (Farla et al., 2004; Tyagi et al., 2000; van Royen et al., 2012). Full manifestation of SBMA occurs only in males, due in part to the location of the *AR* gene on the X chromosome and to higher levels of androgens; female carriers are thus protected both by a second copy of the chromosome and lower relative amounts of androgens, and this is recapitulated in animal models of the disease (Katsuno et al., 2003; Takeyama et al., 2002).

Neurotoxicity in SBMA therefore appears to require AR activation; a clinical trial with leuprorelin, which inhibits androgen production, was successful at reducing polyQ accumulation, but it did not significantly affect key clinical outcome measures (Katsuno et al., 2010). Longitudinal analysis of leuprorelin treatment suggests that it slows disease progression; however, this study was done with a small cohort and without randomization or placebo controls (Hashizume et al., 2017). These findings suggest that restricting androgens, especially after symptom onset, provides marginal benefit at best.

The polyQ tract is encoded by a CAG repeat region in the first exon of the *AR* gene. Repeats are normally between CAG₁₃ and CAG₃₁, while longer lengths are correlated with increasing SBMA severity and earlier disease onset (Kim et al., 2017; Palazzolo et al., 2008). The location of the polyQ repeat in the transactivation domain of the AR protein, rather than the ligand- or DNA-binding domains, is consistent with a preservation of androgen-induced nuclear localization; however, once in the nucleus, polyQ-expanded AR is likely impaired in its ability to bind coregulators and to form its transcription complex (Heinlein & Chang, 2002; Jänne et al., 2000; Shang et al., 2002). Mutant AR has been shown to accumulate in nuclei and to bind DNA promiscuously, potentially contributing to its toxic gain-of-function effects (Adachi, 2005; Belikov et al., 2015; Pourshafie et al., 2020). Preventing nuclear translocation of mutant AR reduces neurotoxicity in transgenic mice and cells, suggesting that nuclear translocation is a key step in SBMA pathogenesis (Montie et al., 2009).

4.1.2 Limitations of existing models

As a disease to model, SBMA presents several advantages. The role of the AR protein is well-established in the literature, particularly due to its stated importance in development and in the context of its role in prostate cancer (Formaggio et al., 2021).

Genetic modelling is likewise simplified by a known monogenic cause with a single allele in males. AR is expressed in many tissue types, and SBMA patients present with other manifestations that could be attributed to either an AR loss-of-function or gain-of-function—for example, reduced development of male sexual characteristics is a hallmark of both SBMA and androgen insensitivity syndrome (Oakes et al., 2008). SBMA patients were also found to be at increased risk of nonalcoholic fatty liver disease, but androgen treatment was found to reduce fatty liver in rats, possibly indicating that its comorbidity with SBMA is an effect of AR loss-of-function (Guber et al., 2017; Zhang et al., 2013a). The main problem in SBMA is progressive weakness due to degeneration of lower motor neurons and muscle; it is therefore critical for understanding the pathophysiology of SBMA to understand the effects of polyQ-expanded AR specifically in these cells and tissues.

Numerous animal models have been developed for studying SBMA. While *Drosophila* lacks an ortholog to AR and does not produce androgens, much of the nuclear receptor cofactor machinery is conserved (King-Jones & Thummel, 2005). Thus, transgenic expression of AR in *Drosophila* has been shown to recapitulate nuclear translocation and DNA binding in response to DHT treatment; moreover, expression of polyQ-expanded AR has been shown to induce neurotoxicity in flies in a manner dependent on androgen availability, nuclear translocation, and transcription cofactor binding (Nedelsky et al., 2010). However, exogenous expression of wildtype AR has been shown to predispose flies to neurotoxicity as well (Scaramuzzino et al., 2015). Overexpression of AR is therefore toxic even in the absence of a polyglutamine expansion, suggesting that transcriptional control is an important part of normal AR biology and possibly that a toxic gain of normal function is a component of SBMA.

Mouse models have been widely used to study SBMA pathophysiology due to interspecies similarity of steroid receptor pathways. Mice notably lack a pure CAG repeat region in the orthologous *Ar* gene, instead featuring a 60 bp region of mixed CAG, CAA, and CAC codons that code for a mixture of sixteen glutamines and four

histidine residues, including an 8-glutamine repeat (GRCm39 reference genome). SBMA-like phenotypes with motor dysfunction and decreased survival have been shown in mice with transgenic overexpression of human polyQ-expanded AR using constitutive promoters (Katsuno et al., 2002), but overexpression of wildtype 22Q AR in skeletal muscle is also sufficient to cause these phenotypes (Monks et al., 2007).

Commonly-used mouse models express human polyQ-expanded AR via the addition of a bacterial artificial chromosome (Cortes et al., 2014a) or via replacement of the endogenous mouse *Ar* exon 1 with the human *AR* exon 1 (Yu et al., 2006). Such models recapitulate SBMA phenotypes only in CAG-expanded genotypes; however, they rely on AR with 100Q or more, longer than what has been observed in SBMA patients (Arnold & Merry, 2019), and similar endogenous gene replacement using the clinically-relevant 48Q shows a slight loss-of-function but no SBMA-like phenotypes (Albertelli et al., 2006). Mouse models are thus well-suited to study the pathophysiology of SBMA and to assess effects of candidate treatments to prevent neuromuscular weakness. However, they are limited in their ability to identify molecular phenotypes by their tissue complexity, artificial polyQ lengths, and non-human genetic background.

Early SBMA cell models also relied on overexpressing polyQ-expanded AR in mouse and human cell lines (Pennuto & Basso, 2016). MN-1 and NSC-34 cells were both derived via the fusion of embryonic mouse LMNs with the N18TG2 mouse neuroblastoma cell line (Hounoum et al., 2016; Salazar-Grueso et al., 1991). PC12 cells are derived from a rat tumor and can be differentiated to a neuron-like state by treatment with nerve growth factor (Walcott & Merry, 2002). HEK 293 cells are of human origin, but as mentioned in Chapter 1, the line has an unstable genome and irregular gene expression patterns (Lin et al., 2014). These models provide a tractable platform for experimentation and have helped to understand the molecular function of the polyQ-expanded AR, but they are likewise limited by a reliance on overexpression, a non-human origin in most cases, and an unstable genome (Mandrusiak, 2003).

At the time of undertaking this work, there were only a few studies that had used iPSC-derived neurons as a cell model of SBMA (Cortes et al., 2014b; Grunseich et al., 2014b). In these studies, differentiation was performed with small molecule-based techniques, which leads to cell type heterogeneity. Multiple patient-derived and healthy control lines were also compared; while it is important to determine whether potential disease mechanisms generalize in multiple genetic backgrounds, it is difficult to directly compare effects seen in lines with different genetic backgrounds without isogenic controls, since extraneous factors could be responsible.

4.1.3 Aims & hypotheses

Clearly delineating the loss- and gain-of-function effects remains an outstanding challenge for understanding and potentially treating SBMA. AR knockouts can recapitulate the loss-of-function effects, and polyQ-expanded AR overexpression combines the loss- and gain-of-function. However, at the time, no systematic effort to separate these effects on a transcriptomic, proteomic, or epigenomic scale had been achieved, and the overall aim of this project was to rigorously address this question.

Working together, Xia Feng, Jill Hakim, and I therefore began this project aiming to derive an improved cell model of SBMA to overcome the limitations of existing models. First, we posited that an isogenic series of cell lines derived by CRISPR-editing the endogenous *AR* gene to express a range of polyQ lengths would provide the best system to assess the effects of the polyQ expansion as well as provide strong controls for the genetic background. Furthermore, we hypothesized that performing these edits in iPSCs would take advantage of their genomic stability and ability to differentiate to recapitulate an LMN-like phenotype.

To overcome the heterogeneity of chemical differentiation, I aimed to further edit these lines to incorporate doxycycline-inducible expression of human *NGN2*, *ISL1*, and *LHX3*

(hNIL) to promote i³LMN differentiation. I hypothesized that the loss- and gain-of-function effects of polyQ-expanded AR could be separated by identifying transcriptional differences in the i³LMNs that correlated with polyQ length, that were not present in AR KO, and that replicated in the two genetic backgrounds.

4.2 Editing the endogenous *AR* locus

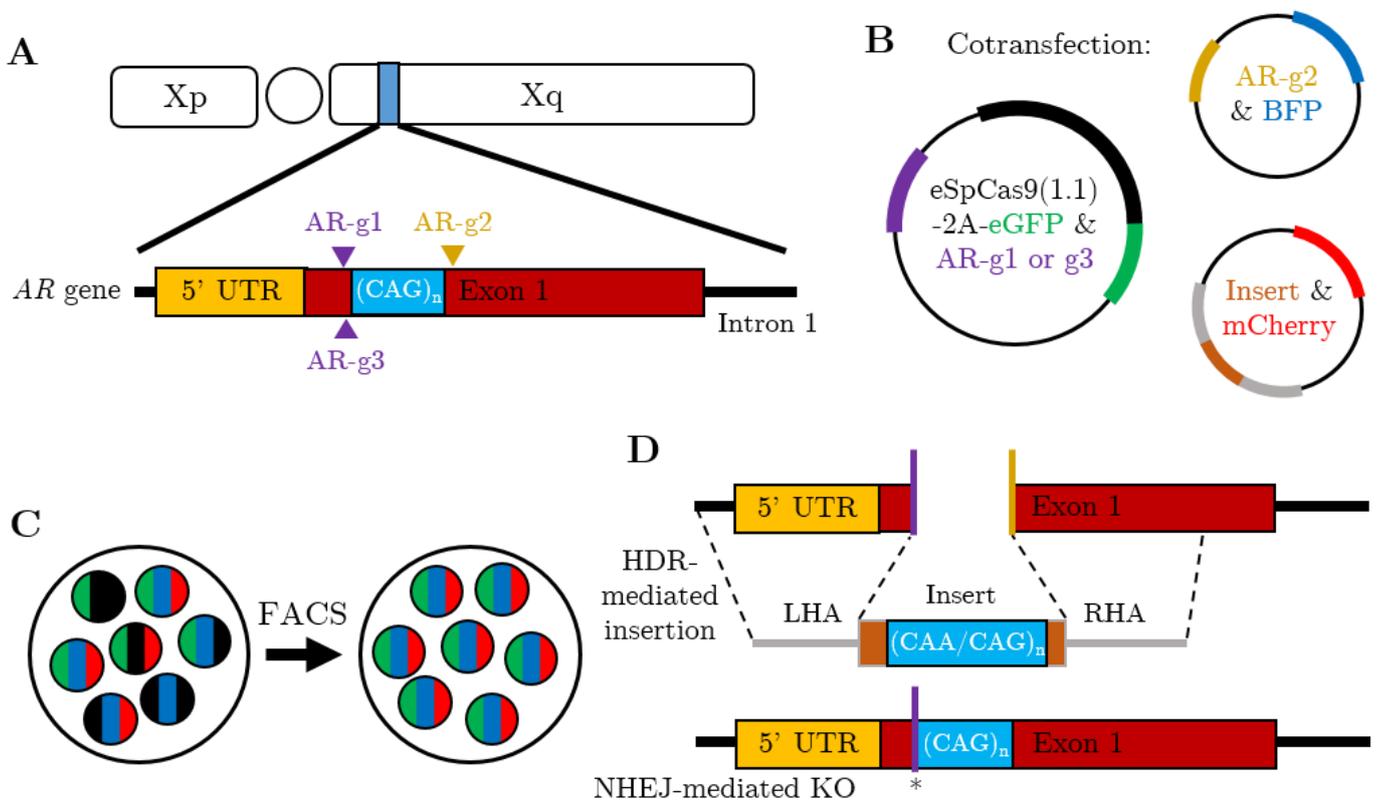
4.2.1 Gene editing strategy & plasmid construction

Two iPSC lines had been previously derived and characterized, and thus served as parental lineages: a wildtype line with 23Q in AR designated HT180 and a patient-derived line with 68Q designated SB18MP2, referred to as H23 and S68 respectively hereafter (Grunseich et al., 2014a). These polyQ lengths represent a typical healthy length and the longest SBMA repeat clinically observed at the time, respectively. Intermediate lengths of 40Q and 54Q were also chosen to model milder forms of disease, and AR knockout lines were included in order to isolate the loss-of-function effects. Thus, ten distinct lines were planned: two parental lines and four edited lines arising from each.

Xia Feng, Jill Hakim, and I used a CRISPR-based strategy to excise the endogenous CAG repeat region of *AR* in H23 and S68 and replace it with inserts of desired length (Fig. 4.1a). Plasmids and gRNAs were designed and constructed as described in Section 2.2.2. AR-g1 was on the same plasmid as eSpCas9(1.1)-T2A-GFP, and AR-g2 was on the same plasmid as a BFP reporter, so co-transfection of AR-g1 and AR-g2 plasmids led to the co-expression of GFP and BFP. The target for AR-g1 is upstream of the *AR* CAG repeat, and the target for AR-g2 is downstream, so expression of both gRNAs and the eSpCas9(1.1) nuclease can lead to the excision of the entire repeat, amounting to 215 bp in H23 and 281 bp in S68 (Fig. 4.1a-b).

Figure 4.1: AR editing strategy.

- Diagram of the 5' UTR and exon 1 of the *AR* gene on the X chromosome. AR-g1 and AR-g2 were co-expressed for all insertion lines, while AR-g3 was used for both knockout lines.
- Diagram of the three plasmids constructed for editing *AR*. Insert lines were co-transfected with all three, while KO lines were only transfected with the plasmid encoding eSpCas9(1.1)-2A-eGFP and AR-g3.
- One day after transfection, cells were purified by FACS to select for only those cells expressing all relevant fluorescent proteins (i.e., GFP, BFP, and mCherry for insertion lines, and GFP alone for KO lines)
- After cutting by both gRNA1 and gRNA2, the insert may be used as a template for HDR, yielding a successfully-edited cell. Meanwhile, the KO lines relied on indels formed by non-homologous end joining (NHEJ) repair, resulting in a premature stop codon (*)



Insert plasmids were also designed and constructed as described in Section 2.2.2. Codon optimization of the insert retained the wildtype protein sequence but mutated the gRNA target sites to prevent re-cutting by the nuclease; it also permitted specific PCR validation that the insert was present at the correct locus as described in more detail below. Mainly CAA codons were used to code for glutamine in order to facilitate insert production and promote polyQ length stability, but this did not capture potential mechanisms of RNA-induced toxicity which have been shown to play a role in other polyQ diseases (Lee et al., 2019; Nalavade et al., 2013). Instead, molecular mechanisms caused only by the polyQ-expanded AR were recapitulated in the edited cells, while the S68 parental line would exhibit phenotypes based on both the polyQ-expanded protein and CAG-expanded RNA.

4.2.2 Clonal isolation & validation

Xia Feng and Jill Hakim derived the H23 AR knockout (HKO) and S68 AR knockout (SKO) lines previously, as described in Pourshafie et al., 2020. In this section, transfection and most of the PCR and agarose gel electrophoresis was performed by Xia Feng and Jill Hakim, and FACS was performed by the NHGRI Flow Core. I performed all cell culture, including picking, splitting, gDNA collection, freezing stocks, and preparing flasks for karyotyping. I also performed a subset of the PCR and agarose gel electrophoresis. Xia Feng and I generated the western blot image together, and WiCell Cytogenetics performed karyotyping and fluorescence in-situ hybridization (FISH).

First, iPSCs of the H23 and S68 lines were co-transfected with the desired insert plasmid and the two CRISPR plasmids (eSpCas9(1.1)-GFP+AR-g1 & BFP+AR-g2); the two AR KO lines were derived by only transfecting the plasmid with eSpCas9(1.1)-GFP and AR-g3 (Fig. 4.1b). One day after transfection, FACS was performed to select

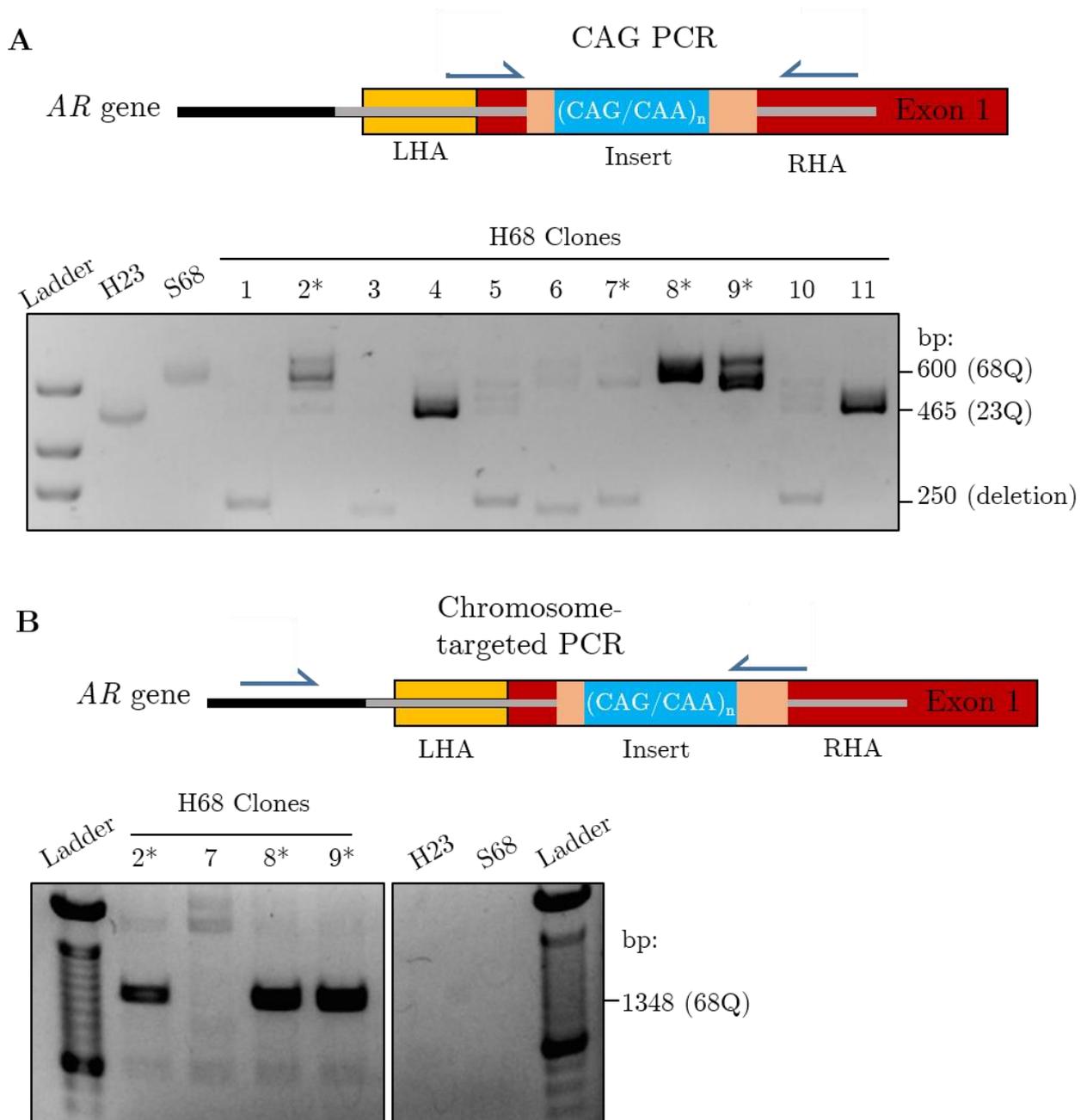
for those cells that co-expressed all of the transient fluorescent markers: GFP in all lines, and additionally BFP and mCherry in the insert lines (Fig. 4.1c). In this way, only cells that successfully received all necessary plasmids for proper editing were carried forward for clonal isolation and validation. Fluorescence was lost by four days after transfection, indicating that the plasmids were degraded instead of being integrated into the genome. However, due to the lack of a permanent selectable marker, we subsequently isolated and genotyped clones without further enrichment.

Cells were then split with accutase onto new 6-well plates at approximately 1,000 cells per well, a density low enough to permit the outgrowth of single-cell clones. After about one week, clones were manually picked onto 24-well plates. Those that survived and presented normal morphology after outgrowth were expanded, and gDNA was collected with QuickExtract; those that presented abnormal growth rates or high levels of spontaneous differentiation were discarded. The relative complexity of the insertion (requiring nearly simultaneous cutting of the two gRNA target loci along with HDR rather than NHEJ in both cases) led to the expectation that the desired edited genotype would be quite rare in the population (Fig. 4.1d). In all, between 70 and 120 clones from each insertion line were picked, expanded, and gDNA collected.

Each clone was then genotyped by PCR and agarose gel electrophoresis. First, the length of the CAG repeat was assessed using primers outside of the mutated region and spanning the repeat, resulting in amplicons of 465 bp (23Q), 516 bp (40Q), 558 bp (54Q), or 600 bp (68Q; Fig. 4.2a; Table 2.4). In some cases, an amplicon of approximately 250 bp was observed that corresponded with a deletion of the region between gRNAs. In this way, the size of the repeat region was estimated, and clones without a band at the desired size or with more than one band were discarded.

Figure 4.2: Genotyping strategy.

- (a) After gDNA extraction, a PCR was performed across the polyQ repeat region. Clones with a band at the correct length (indicated by *) were taken forward.
- (b) PCR was then used to ensure that insertion occurred at the correct locus. Clones with a band (indicated by *) were sequenced; all three sequenced as expected.



The PCR product was intentionally small in order to emphasize small differences in the length of this region, but primers were outside of the insert so that amplification would occur in the parental and KO lines with the same primers. In addition, this PCR was able to distinguish the purity of clones, since multiple bands would be present if unedited cells and edited cells were mixed. Since the KO lines did not have an insert template, variable mutagenesis was expected, so all KO clones were sequenced. Ultimately, an HKO and an SKO clone with matching deletions of 14 bp that introduced a premature stop codon were chosen.

Clones that successfully amplified a band at only the expected length were further tested by a chromosome-targeted PCR, in which one primer upstream of the left homology arm was coupled with a primer unique to the insert that amplified only if the insert was present at the endogenous *AR* locus (Table 2.4). This resulted in amplicons ranging in size between 1213 bp for 23Q and 1348 bp for 68Q. In this way, any clones that failed to present a band potentially indicated an off-target insertion and were discarded. Those clones that amplified were again PCR amplified, and the fragments were purified and Sanger sequenced to confirm the length of the repeat and that no undesired mutations were introduced at the insertion site.

By comparing the number of clones that genotyped correctly with the total number of clones picked, a rough estimate of cutting and integration efficiency could be calculated, while variability of transfection efficiency was accounted for by the aforementioned FACS (Table 4.1). It was immediately clear that cutting and indel efficiency for AR-g3 is high; HKO and SKO showed mutagenesis in 50% of screened clones (16/32 between both lines). AR-g1 and AR-g2 were predicted to have similar efficiencies when designed *in silico*, but integration efficiency in the insertion lines (H40, H54, H68, S23, S40, and S54) was typically in the low single-digit percentages. In fact, none of the clones screened for H54 and S54 in the first attempt sequenced correctly, so the parental lines were transfected again and additional clones were picked in order to find correctly-edited cells. Low efficiency was expected, since NHEJ is the

most active repair mechanism for double-stranded breaks in iPSCs, while HDR is less common (Miyaoka et al., 2016). Interestingly, the length of the insert did not appear to have an effect on integration efficiency, although more clones would need to be screened for this to be analyzed with rigor.

Table 4.1: Summary of isogenic series derivation.

Two series of five isogenic cell lines were derived with the indicated *AR* genotype. Editing efficiencies were calculated by the number of clones that sequenced as expected divided by the total number of clones for which gDNA was collected and CAG length assessed by PCR.

Human iPSC Series	Cell Line Name *parental line	AR PolyQ Length	Editing Efficiency Sequence-confirmed/ total clones
H-Series	HKO	--	10/16 (62.5%)
	H23*	23	--
	H40	40	1/70 (1.4%)
	H54	54	5/116 (4.3%)
	H68	68	4/120 (3.3%)
S-Series	SKO	--	6/16 (37.5%)
	S23	23	5/91 (5.5%)
	S40	40	1/96 (1.0%)
	S54	54	3/78 (3.8%)
	S68*	68	--

Stocks of those clones that sequenced with the desired genotype were then expanded and frozen, and bulk protein samples were collected after treatment for 48 h with EtOH or 10 nM DHT. It was expected that AR protein would be expressed at the molecular weight corresponding to the desired polyQ length or, in the case of the KO lines, have no detectable expression. It was additionally expected that AR expression would be increased in DHT-treated conditions. Some clones failed to express *AR*, possibly due to epigenetic silencing, and these were discarded; most clones were found to express as expected. A spectrum of AR sizes in each series was thus achieved (Fig. 4.3a).

Ronald Wang and Xia Feng additionally tested for off-target mutagenesis by promiscuous cutting of the eSpCas9(1.1) nuclease. The three gRNA targets were analyzed using the crispr.mit.edu tool, which ranked alternative genomic loci with an adjacent PAM motif by sequence similarity, location of mismatches, and whether alternative targets are within a coding region (Hsu et al., 2013). While this tool is no longer available, the same data is incorporated into the off-target score analysis of the Benchling CRISPR design tool. The top results that were within known protein-coding regions were identified, since mutations in exons are usually the most disruptive to the biology of the cell. These loci were amplified in the relevant lines by PCR using primers approximately 500 bp upstream and downstream of each site, and agarose gel electrophoresis was performed with a subset of the reaction to confirm amplification. Each locus was then Sanger sequenced in each line, and no off-target mutations were found at any of these loci in any line, confirming that the gRNAs and the nuclease demonstrated tight specificity (Table 4.2; data by Ronald Wang & Xia Feng).

A flask of one clone from each line that had passed all previous validation tests was then karyotyped (Fig. 4.3b). Only the S54 clone was found to have a chromosomal abnormality; about 15% of these cells carried a duplication in chromosome 20 as shown by FISH. Fortunately, an additional validated clone from this line karyotyped normally. Together, the tests performed on these lines effectively demonstrated that the desired genotypes were induced in all eight edited lines, that the clones were pure

and expressed a functional and edited protein, the most likely off-target loci were unaffected, and no major chromosomal abnormalities were introduced. While there remained a chance that other off-target loci could have been mutated, it was deemed unnecessary to sequence the full genome of these lines at the time.

Figure 4.3: Validation of AR protein expression and genomic integrity.

- (a) Clones that sequenced correctly were expanded and treated with either EtOH (-) or 10 nM DHT (+) for 48 h prior to lysis. AR expression at the correct size and with a relative increase in the presence of DHT was confirmed.
- (b) One successful clone from each line was karyotyped. Example of a normal karyotype generated from the S23 derived model line shown.

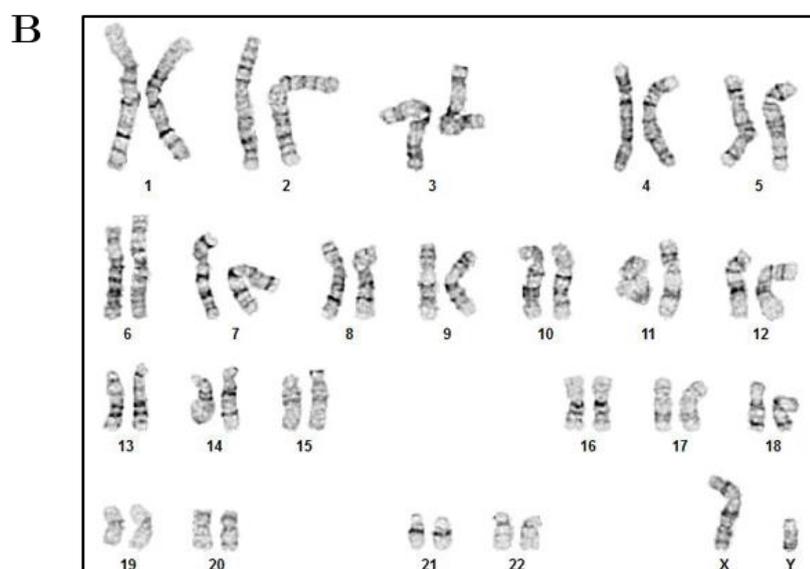
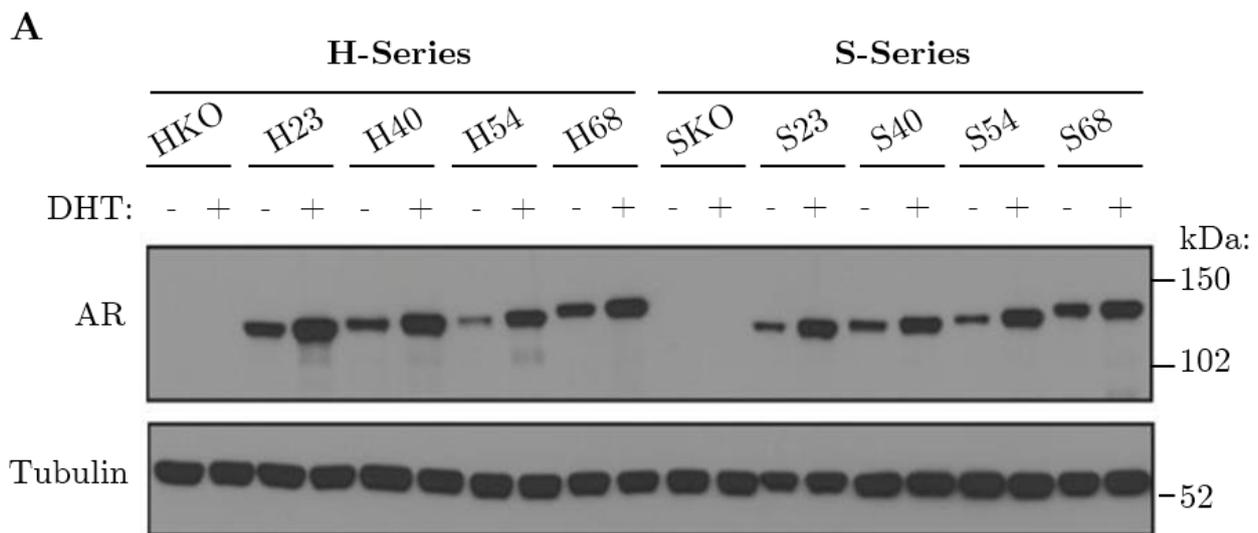


Table 4.2: Top exonic potential off-target loci & PCR primers.

By Ronald Wang and Xia Feng.

AR gRNA	Gene	Mismatches (bp)	Primer Pair (5'-3')	
1	JAK3	4	GCTTCGGGCTACGCAAGGATT	
			GAAGAAGTGCTGGGAGTCCGT	
2	VIM	3	CGTCCTCCTACCGCAGGATGTT	
			CAGGTTCTTGGCAGCCACACT	
	MYO18B	4	GTGGGTCTGGAAGCAGTTCATGA	
			GGAACACCTTCCTCTTGATATCGCTGTA	
	CDIP1	4	GTCTACCCGCCTCCTTGTGACA	
			GCATGTAGGTGCCATCTGCACT	
	STOML1	4	GCCTGCTACCAGTTC AATGTCGT	
			GTCAGGCTGTTTCTGTGGGTCCT	
	SCML4	4	CATTCTGTAAGACTGGCACCAGCA	
			GGAGGACATAGCCGATGCTGT	
	3	DEDD	3	GTGGACTCATCCGAAATGGACGTGA
				GCTGGCAGTATTCAGCCCGA
VPS16		4	GGACTGCTACACGGCGAACT	
			GGTAGGTGAAGGAGACAGCCATCT	
SCAND2P		4	CCATGGCTGCTTGTTATCCGGT	
			GCACCTGGCCAGACTTGACT	
MYOM3		4	GACTGAGATCGCCTTCCTGCAGA	
			CAAGCATGACAGGGAGAAGGGT	
CCNL1		4	GGCTAAAGGAAGAGGCCGGAA	
			CCAGGATGCGTAAGTCCGTCTCA	
ADAM19		4	GCCTGGATGGACAAGAGGAAGTGA	
			GGAGTGCTCGAACCCACAGT	
ZC3H4		4	CCACCCTTCCTAGGGTGGCATT	
			GTTCTCACTGTGGAGGCTTCTGCT	
KRT80		4	GGCTGGCACTATCTCCAAGGTGA	
			GCAATGGCTCTTGACAGAGAGGA	
MYCBP2		4	GAACTGTGAAGCTGAGACTCTGCGA	
			GGTTGAGCCCTTCATCCACCA	

4.3 Mag-hNIL does not induce i³LMN differentiation

4.3.1 Insertion & genotyping of Mag-hNIL

As described in Chapter 1, expression of developmental transcription factors has been shown to be more rapid and uniform in inducing neuronal differentiation than chemical induction, and expression of the hNIL factors effectively induces differentiation to cells that are transcriptionally and morphologically similar to lower motor neurons (Mazzoni et al., 2013).

As described in Chapters 2 & 3, the Mag-hNIL plasmid contained the hNIL factors in a polycistronic cassette separated by 2A cleavage peptides under a TRE3G promoter (Fig. 2.1). Adjacent to this cassette was a constitutive CAG promoter driving the expression of SBP- Δ LNGFR, mApple, and the third-generation reverse tetracycline transactivator (rtTA3G), also separated by 2A cleavage peptides, and SBP- Δ LNGFR and mApple were flanked by loxP sites, enabling excision with Cre. Mag-hNIL contained homology arms targeting the second intron of the *CLYBL* gene (Cerbini et al., 2015). Similar constructs had been used previously to differentiate iPSCs into i³LMNs using a simple 3-day dox induction followed by extended culture to promote maturation (Fig. 2.3; Fernandopulle, Prestil et al., 2018).

All ten isogenic lines were edited to insert Mag-hNIL using a pair of TALENs targeting the *CLYBL* locus as described in Chapter 3. With the aid of magnetic bead enrichment enabled by SBP- Δ LNGFR (Chapter 3) and live visualization of mApple, several putative Mag-hNIL clones from each of the ten lines were isolated. As with the establishment of the isogenic series, gDNA was extracted from each clone for genotyping. First, PCR was done using primers within the homology arms, so a band of 790 bp indicated a wildtype allele, but alleles with Mag-hNIL insertion were too long to amplify efficiently (Fig. 4.4a). Additionally, PCR was done using a forward primer upstream of the homology arm and a reverse primer unique to the insert, such

that the presence of a band of 1465 bp indicated insertion at the correct locus (Fig. 4.4b).

Thus, the presence of both an insert band and a wildtype band indicated heterozygosity, while an insert band but no wildtype band suggested either homozygosity or a large indel at the cut site (Fig. 4.4c). In this case, heterozygous clones were preferred so as to standardize expression levels and retain an unedited CLYBL insert locus should further editing be desired. Clones for which both bands were observed were Sanger sequenced to confirm that insertion was effective and that the wildtype allele was not subject to small indel mutations.

When it came to functionally testing the Mag-hNIL clones, however, it was found that the cells did not respond to dox treatment. No morphological changes were evident, and cells continued to divide as before; these were sufficient criteria to conclude that the cells were not neuronal, so no additional expression markers were assessed. WTC11-hNIL, a line with the pUCM-CLYBL-hNIL insert and known to differentiate in response to dox treatment, did induce such changes, indicating that the dox used was effective (Fig. 4.5).

Because the main difference between the insert in WTC11-hNIL and Mag-hNIL cells was the selection genes, it was hypothesized that the presence of the SBP- Δ LNGFR tag in the same cassette as rtTA3G interfered with the activity of the latter and thereby prevented induction of the hNIL genes. Despite its widespread use, no straightforward method for assessing the functionality of rtTA3G was readily found, so it was undertaken to develop such an assay to aid troubleshooting in this case and in dox-inducible systems generally.

Figure 4.4: Genotyping strategy of Mag-hNIL insertion.

- (a) Diagram of the wildtype *CLYBL* locus. A pair of TALENs targeting intron 2 were co-transfected with Mag-hNIL. PCR genotyping included primers within the homology arm sequence (gray bar); after insertion, this amplicon was too long to efficiently amplify, so only WT alleles produce a band.
- (b) Additional primers were used for insert-specific PCR, with one primer upstream of the left homology arm paired with another primer within the insert.
- (c) PCR products were run on an agarose gel; putative heterozygous clones (*) were sequenced to validate that WT alleles were unmutated and insertion alleles were correct.

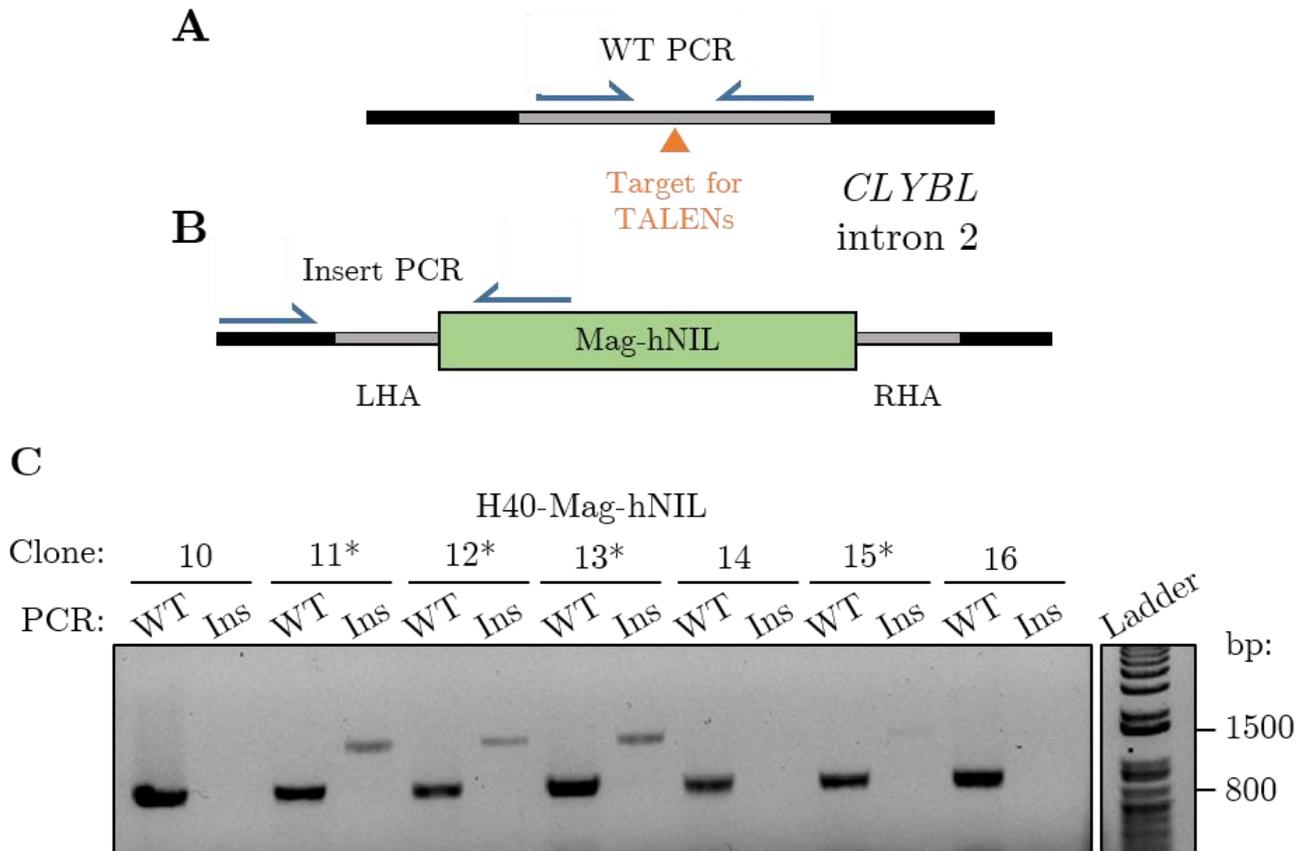
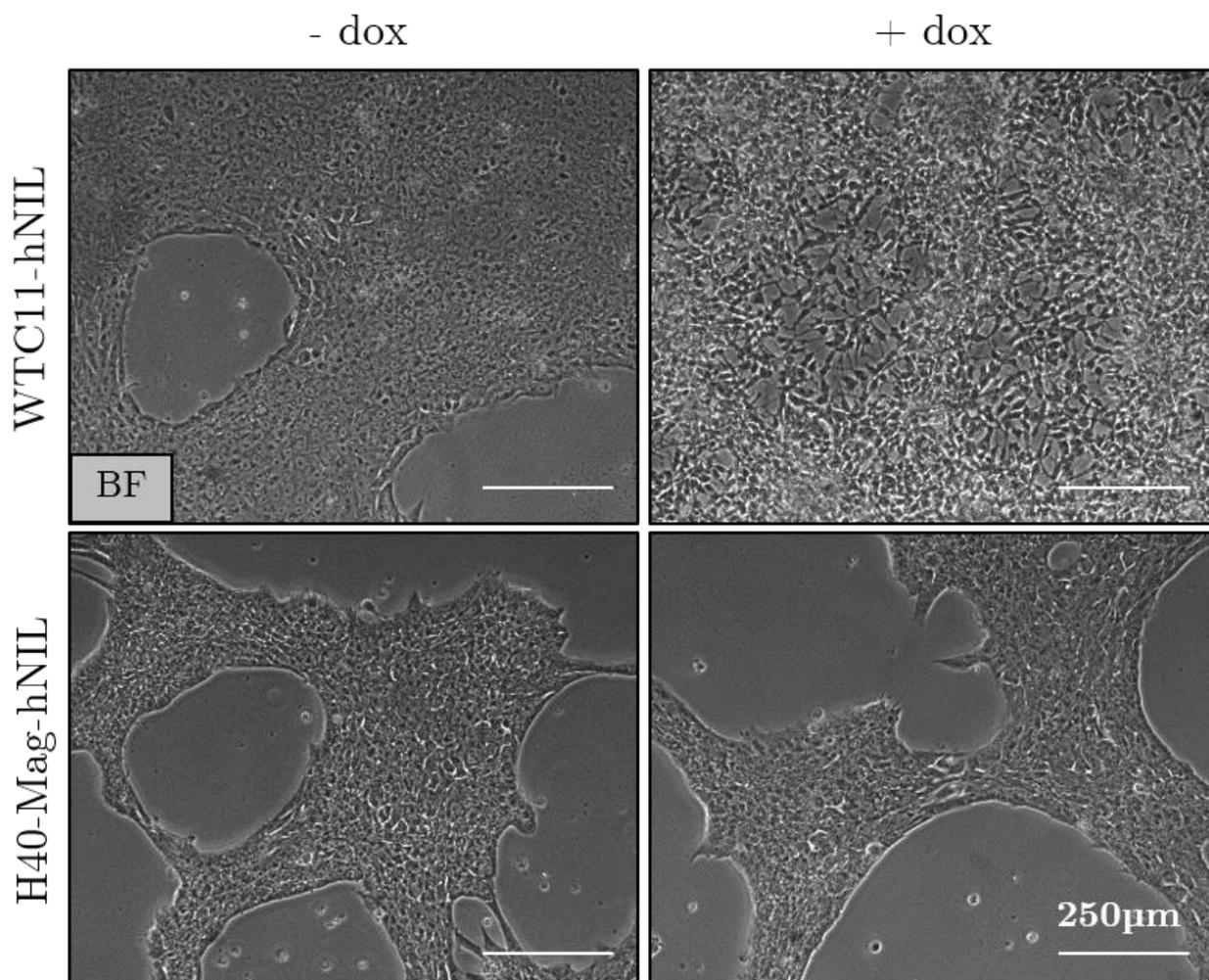


Figure 4.5: Mag-hNIL cells lack dox sensitivity.

WTC11-hNIL and Mag-hNIL cells were untreated (-) or treated with 2 $\mu\text{g}/\text{mL}$ dox (+) for 3 days. WTC11-hNIL responded to dox treatment as expected, but no change to morphology was evident in Mag-hNIL cells (H40 line shown).



4.3.2 A simple assay to monitor rtTA3G activity

In designing an assay for functional testing of the dox response, it was desired that the protocol be modular; in other words, that key reagents may be swapped and improved without needing to redesign the experiment as a whole. Furthermore, it was desired that the assay be rapid, so no genetic edits or additional cloning would be necessary. Thus, a primary readout of fluorescent protein expression from a transiently

transfected plasmid was chosen as a way to qualitatively assess the presence of active rtTA3G in the population, and an account of this protocol was published as Support Protocol 6 in Fernandopulle, Prestil et al., 2018.

For each cell line, four wells of a 6-well plate were prepared to at least 30% confluency, and two of these wells were transiently transfected with an equal amount of pBI-MCS-EGFP (TRE-eGFP) plasmid. In addition, the medium of one transfected and one non-transfected well was supplemented with 2 $\mu\text{g}/\text{mL}$ dox on the same day as transfection, while the other two wells were untreated. The following day, each well was washed with PBS to remove debris and imaged live with a fluorescent microscope. Results of such an experiment are shown in Fig. 4.6.

To simplify this assay, no enrichment was performed; as a result, a variable number of plasmid copies was expected to be transfected into each cell, and some did not receive any plasmid. Expression levels therefore varied greatly between individual cells. However, by comparing the overall intensity of eGFP in dox-treated and untreated transfected wells of each cell line, a dramatic increase could be noted for cells with functional rtTA3G, indicating that the plasmid was expressed in response to dox treatment. Meanwhile, no change to eGFP intensity was observed in non-transfected wells or in cells lacking rtTA3G regardless of dox treatment. The non-transfected wells served as negative controls to calibrate background fluorescence in the GFP channel, and transfected wells without dox accounted for TRE promoter leakiness.

The TRE-eGFP plasmid used here was identified for protocol development as a simple and widely available construct, but alternative constructs may be substituted. In particular, the TRE promoter used here was found to have detectible levels of nonspecific expression in the absence of dox, suggesting that later generations of TRE promoters with tighter expression specificities would improve the assay. Furthermore, while eGFP was ideal for this context, the modularity of the protocol enables the use of any easily-detectible reporter gene (e.g., luciferase or another fluorescent protein), so long as it is under a tetracycline-responsive promoter.

Fluorescent microscopy was used here for rapid, qualitative validation of expression, but flow cytometry could be used for a quantitative readout by measuring median eGFP intensity or the percentage of cells expressing eGFP above a baseline threshold. Heterogeneity of transfection efficiency could be accounted for by introducing a different fluorescent protein under a constitutive promoter on the same plasmid to streamline gating out of non-transfected cells and enable normalization of eGFP intensity to this internal transfection loading control. In this way, the relative level of rtTA3G activity could be compared between different cell lines or treatments. Overall, this assay is a simple, flexible, and qualitative method for determining whether rtTA3G is active, and it can be further sophisticated with modular changes.

4.3.3 Mag-hNIL cells are insensitive to dox treatment

Using the assay described in the preceding subsection, rtTA3G was indeed found to be inactive in the Mag-hNIL lines, but it was active in WTC11-hNIL, a line known to be effective at stimulating differentiation under dox treatment and therefore used as a positive control (Fig. 4.6).

Figure 4.6: Mag-hNIL cells lack functional rtTA3G.

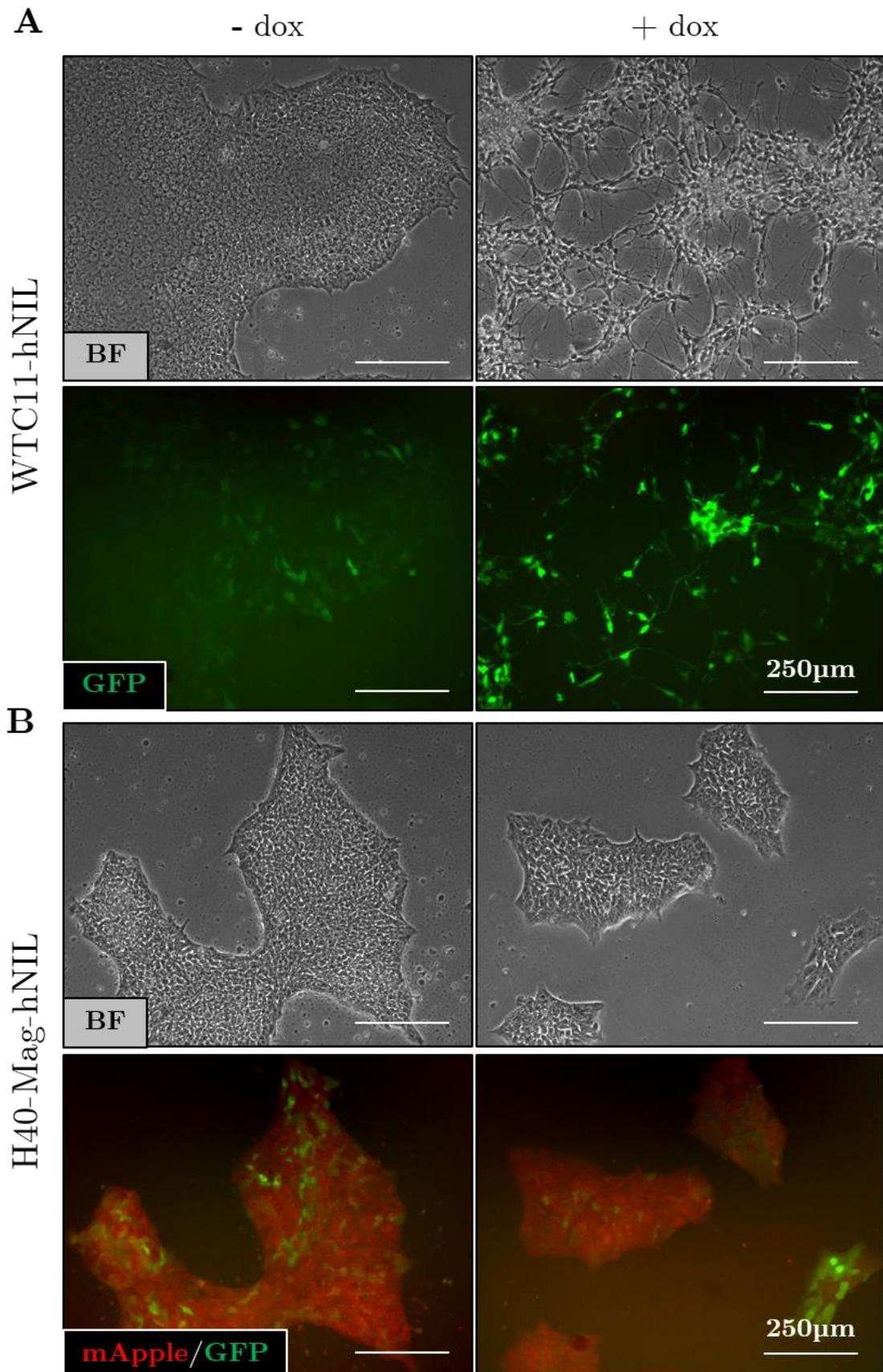
WTC11-hNIL and Mag-hNIL cells were transfected with TRE-GFP and either untreated (-) or treated with 2 $\mu\text{g}/\text{mL}$ dox (+) for 3 days.

(a) WTC11-hNIL cells showed a strong increase to GFP expression in transfected cells after dox treatment, indicating active rtTA3G.

(b) Mag-hNIL cells (H40 shown) showed negligible GFP expression both with and without dox treatment, indicating negligible rtTA3G activity.

(continued)

Figure 4.6 (*continued*): Mag-hNIL cells lack functional rtTA3G.



Since SBP- Δ LNGFR was efficiently localized to the cell membrane, it was possible that the mRNA transcript for the cassette was processed by ribosomes on the endoplasmic reticulum and trafficked in vesicles, reducing cytoplasmic/nuclear availability of rtTA3G. This was supported by decreased relative mApple intensity as compared to cassettes without SBP- Δ LNGFR. However, the presence of T2A cleavage peptides should allow mApple and rtTA3G to be trafficked independently of SBP- Δ LNGFR, and indeed mApple does not co-localize with SBP- Δ LNGFR, nor does it appear trapped in vesicles (Fig. 3.2). Further experiments would therefore be required to determine whether the observed lack of rtTA3G functionality was due to protein mislocalization, a decrease in translation efficiency, an increase in degradation, a lack of 2A-mediated ribosomal skipping, or another explanation.

4.3.4 Cre excision of selection genes does not restore differentiability

It was hypothesized that if the selection genes were responsible for preventing rtTA3G activity, then excision of these genes by Cre recombinase could restore this function. Several of the Mag-hNIL lines were therefore transiently transfected with pCAG-Cre, which resulted in a loss of mApple expression in a subset of cells, indicating that the selection genes were excised. Using the rtTA activity assay described above, an increase in expression of TRE-eGFP was noted following dox treatment only in cells that did not express mApple, suggesting that rtTA function was indeed restored after selection gene excision (Fig. 4.7a).

Using the depletion protocol described in Chapter 3, mApple-negative cells were purified, and single-cell subclones were then obtained by picking. Genomic DNA was then collected from each clone for genotyping. The Mag-hNIL insert was PCR amplified using primers flanking the selection genes, resulting in an amplicon of 3196 bp without excision and 1300 bp with excision. Clones with only the smaller band

visible on agarose gel electrophoresis were Sanger sequenced to confirm that excision was successful and did not alter the promoter or rtTA3G sequence.

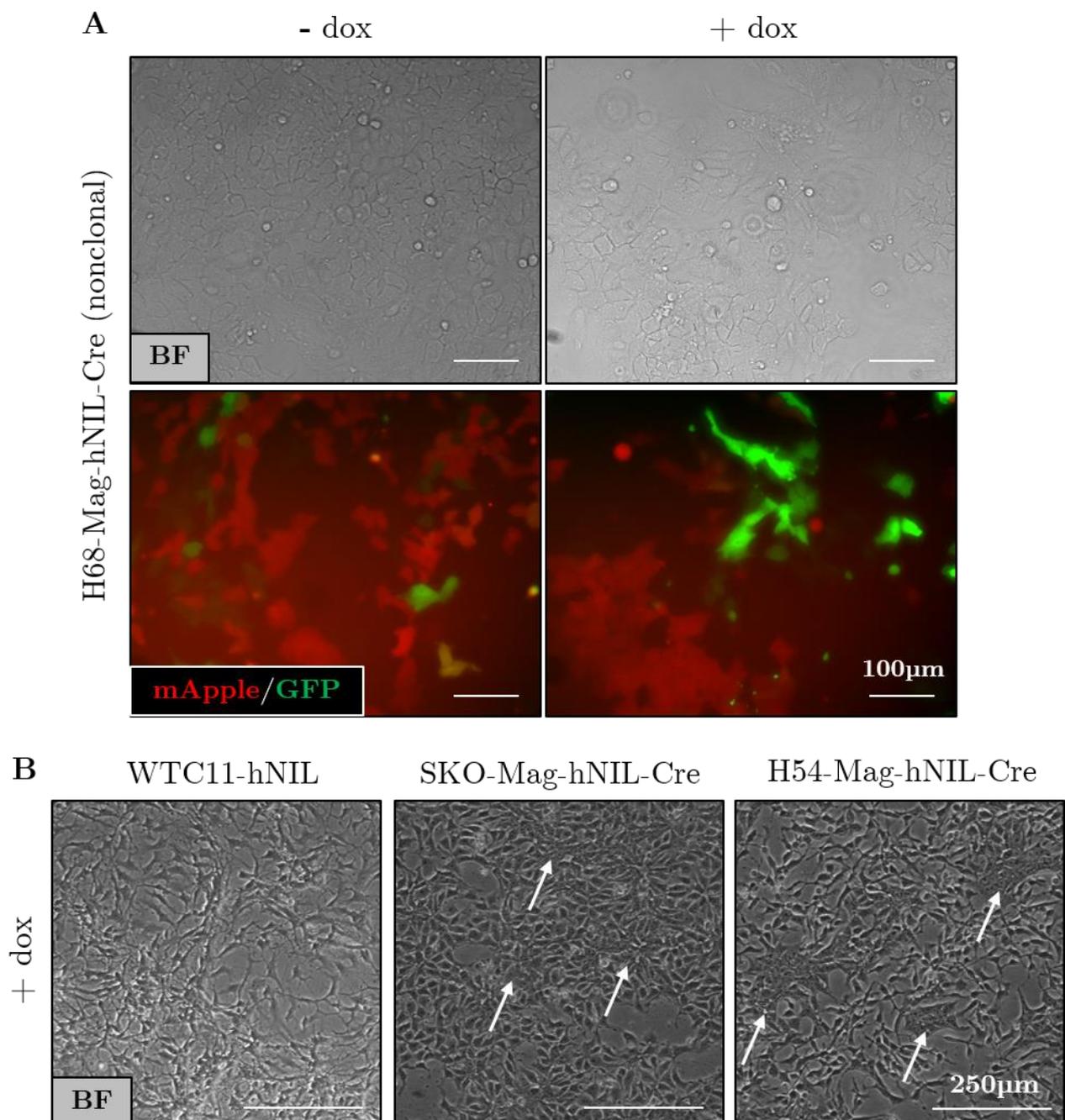
Several clones from each line were identified as having scarless excision of the selection genes, and subsequent dox treatment did induce some morphological changes in these cells. However, many dense colonies that resembled undifferentiated iPSCs remained despite dox treatment, and no extended processes were formed, indicating incomplete differentiation (Fig. 4.7b). As before, a lack of neuronal morphology and continued mitotic activity was used to exclude the neuronal cell state, and further cell state markers were not assessed.

The finding that Mag-hNIL did not functionally induce differentiation was unexpected, since polycistronic cassettes had been used previously for this purpose. Furthermore, while excision of the selection genes appeared to enable rtTA3G activity, the continued failure of the cells to differentiate fully and equally was a major roadblock to progress. However, it should be noted that the development and use of magnetic bead-based cell sorting substantially helped to facilitate the numerous insertions and excisions performed, highlighting its utility despite the finding that SBP- Δ LNNGFR presumably interfered with rtTA3G function and thus it was not amenable to polycistronic expression. I therefore redesigned the hNIL induction plasmid.

Figure 4.7: Cre excision of does not restore differentiability.

(a) H68-Mag-hNIL cells were transiently transfected with pCAG-Cre, and the rtTA activity assay was performed. Bright-field and combined mApple & GFP expression is shown. Only in cells without mApple, rtTA3G activity in response to dox treatment is restored.

(b) Following Mag-hNIL insertion and Cre excision, dox treatment for 3 days still fails to induce complete differentiation. WTC11-hNIL line included as a positive control; SKO and H54 lines shown for reference. Compact colonies suggestive of undifferentiated cells are noted (white arrows).



4.4 **Mag2-hNIL induces differentiation but silences *AR***

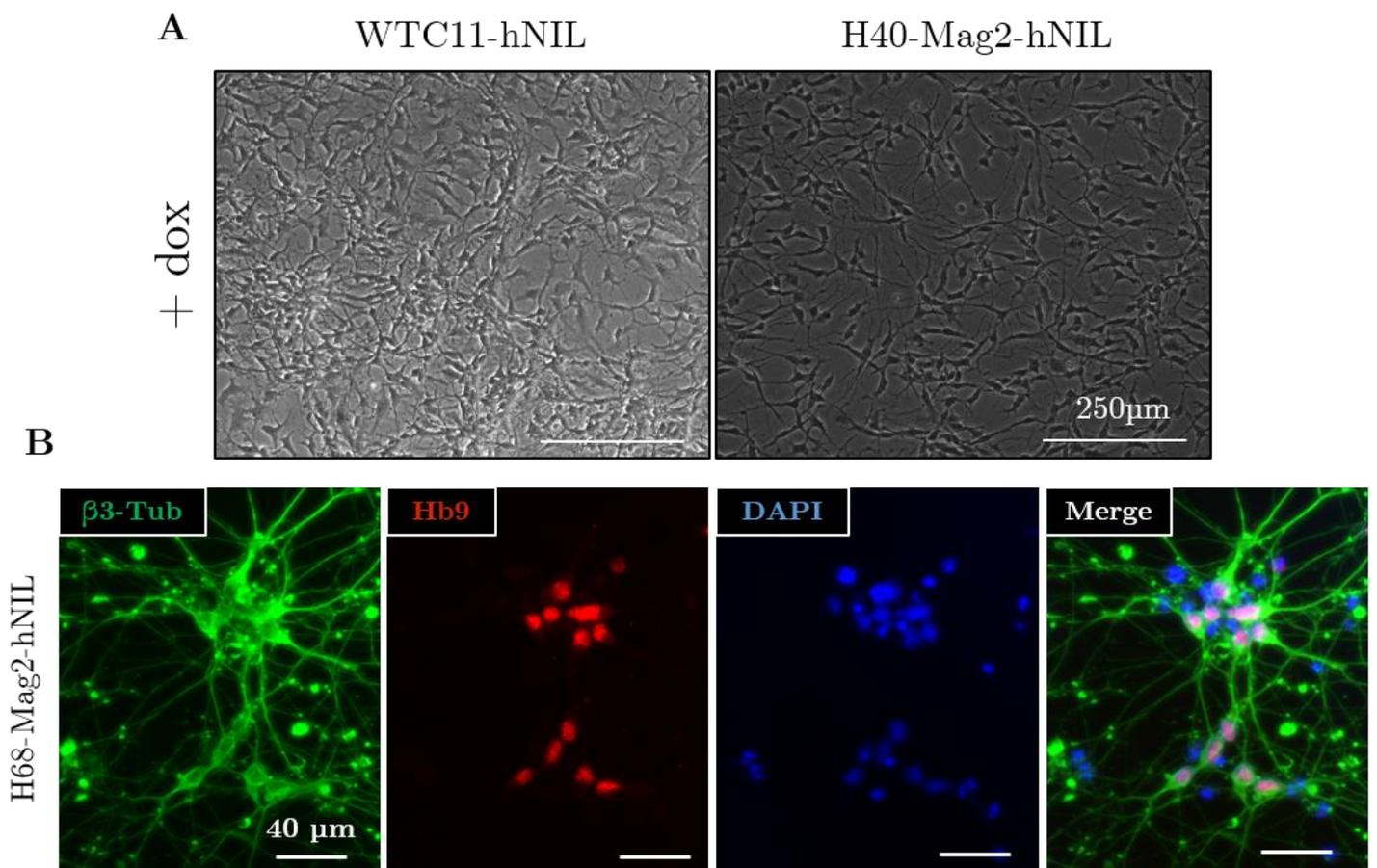
4.4.1 **Insertion & validation of Mag2-hNIL differentiation**

The hNIL insert was reconstructed to constitute Mag2-hNIL as described in Chapter 2. This plasmid dedicated the CAG promoter to only the rtTA3G, and it introduced a separate EF-1 α promoter to drive expression of the selection genes SBP- Δ LNGFR and mApple, and all of these selection genes were flanked by loxP sites to permit Cre excision (Fig. 2.1). The Mag2-hNIL plasmid was inserted into cells with CLYBL-targeting TALENs in the same manner as Mag-hNIL. Non-clonal populations were then enriched with magnetic bead selection, clones were picked, and gDNA was collected. Genotyping also proceeded in the same manner as Mag-hNIL, with the exception of a different insert-specific primer (Table 2.4), and pure clones of each SBMA model line with heterozygous insertion of Mag2-hNIL at the CLYBL locus were purified.

It was found that Mag2-hNIL clones were able to efficiently induce differentiation following dox treatment, becoming morphologically similar to other i³LMNs after three days of dox treatment (Fig. 4.8a). After the full 14-day i³LMN differentiation protocol, Mag2-hNIL cells were found to co-express Hb9 and β 3-tubulin, indicative of a lower motor neuron cell state (Fig. 4.8b). These cells were thus considered i³LMNs.

Figure 4.8: Characterization of Mag2-hNIL differentiation.

- (a) Dox treatment efficiently induced morphological changes and neurite outgrowth in Mag2-hNIL clones that was equivalent to WTC11-hNIL cells.
- (b) Following dox treatment and maturation to day 14, cells co-expressed the general neuronal marker β 3-Tubulin and lower motor neuron marker Hb9.



4.4.2 Differentiated Mag2-hNIL i³LMNs lack AR expression

In the process of validating the SBMA model lines as iPSCs, AR was shown to be expressed at the predicted size, and the protein was found to increase expression and translocate to the nucleus in response to DHT treatment, indicating proper AR functionality. However, after differentiation with Mag2-hNIL, it was found that the SBMA model lines did not have detectable AR protein by immunofluorescence or western blot. It was therefore hypothesized that the expression of AR was silenced during the differentiation process.

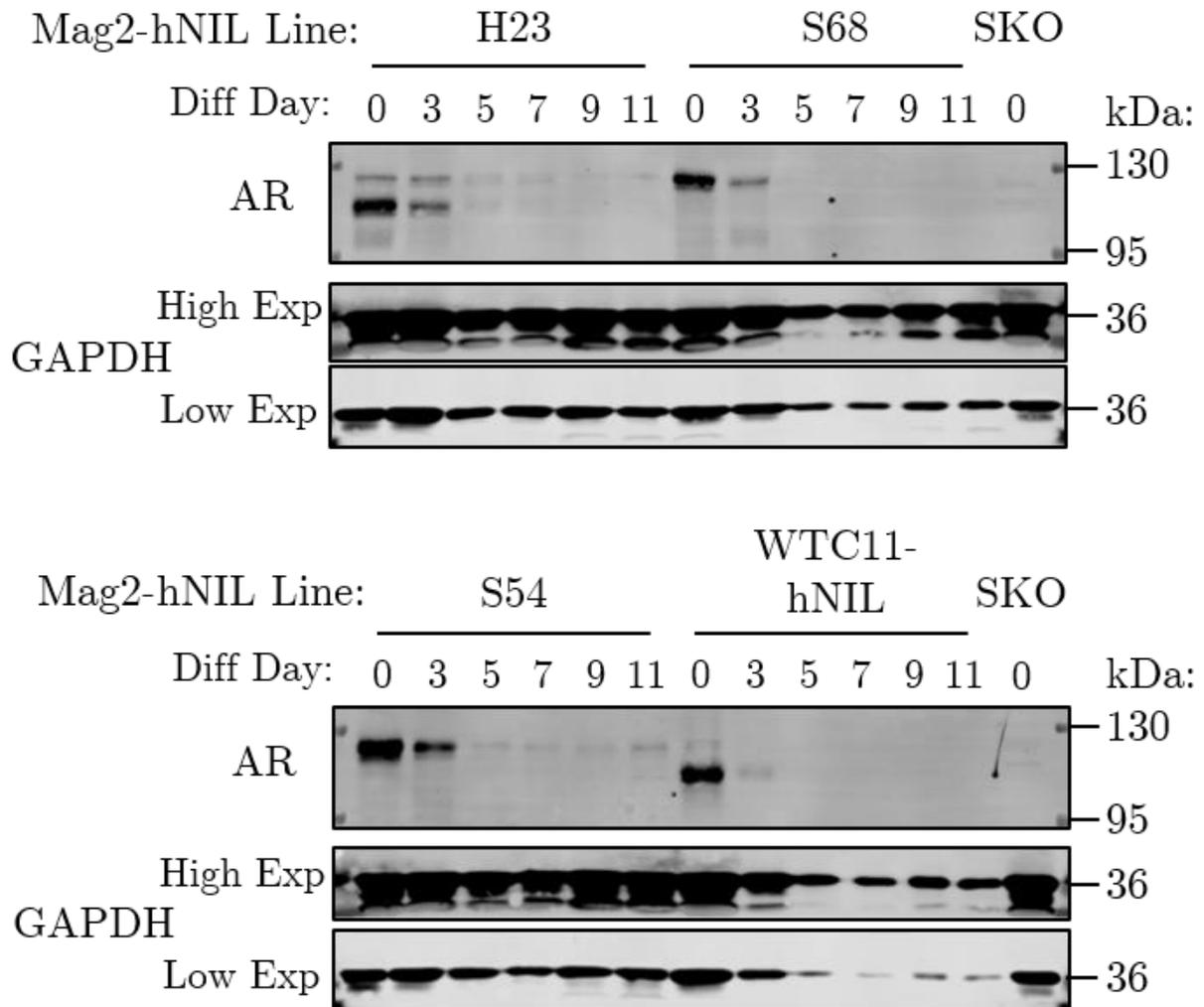
To assess this issue, multiple wells of each cell line were differentiated simultaneously, and protein samples were collected throughout differentiation. It was found that the levels of AR dropped dramatically early on in differentiation and did not recover after the cessation of dox treatment (Fig. 4.9). Even after accounting for decreased loading control signal indicating a decrease in overall protein concentration, all lines were indistinguishable from the AR KO iPSCs within five days. This remained the case regardless of genetic background (shown for H23, S68, and WTC11), polyQ length (shown for cells with 23Q, 54Q, and 68Q), and whether the AR gene was edited (i.e., in both parental and derived model cell lines). Furthermore, the WTC11-hNIL line used the pUCM-CLYBL-hNIL insert rather than Mag2-hNIL; this insert lacks SBP- Δ LNGFR, so SBP- Δ LNGFR was not responsible for AR silencing.

While some AR was detected at three days post-induction, the cells were not fully differentiated at this time, so it was deemed not possible to perform neuron-specific experiments with these cells as initially planned.

Figure 4.9: Differentiation timecourse of AR expression.

Protein samples were collected from iPSCs (day 0) and on the indicated day after initiating dox induction, following the i³LMN protocol. While AR was expressed in iPSCs at the expected size, it was substantially reduced after three days of dox induction. After another two days of maturation in the absence of dox, AR levels were indistinguishable from the KO line, even after taking into account reduced GAPDH indicative of decreased protein concentration.

This pattern was found to hold regardless of whether the *AR* gene was edited, whether the polyQ repeat was expanded, or whether SBP- Δ LNNGFR was present; the WTC11-hNIL line used a different hNIL insert that lacked SBP- Δ LNNGFR.



4.4.3 Transcriptional analysis of SBMA model lines as iPSCs & i³LMNs

While AR protein was not detected, it was hypothesized that the *AR* gene may be transcribed but that additional mechanisms altered splicing, prevented translation, or destabilized the protein. Furthermore, I desired to determine whether polyglutamine expansion affected the erasure of pluripotency or the acquisition of the i³LMN cell state. I thus performed RNAseq on the parental H23-Mag2-hNIL and S68-Mag2-hNIL lines as iPSCs and i³LMNs in order to determine whether any *AR* transcript was present in differentiated i³LMNs and to profile the cell state of each line. Additionally, each cell line and cell type was treated with either 10 nM DHT or EtOH for 48 h prior to RNA extraction.

Three biological replicates were prepared for each condition, and total RNA was collected, prepared, and sequenced as described in Chapter 2. RIN values were between 9.2 and 9.8 for iPSCs, indicating very little RNA degradation; samples from i³LMNs had more variable RIN values between 6.5 and 8.0, indicating more RNA degradation, but this was within the accepted range of quality for total transcriptome RNAseq (Chen et al., 2014). No notable bias was found in any sample for GC content, base contribution, or read quality distribution, indicating good sequencing quality, and of 173,113 total annotated transcripts in the reference genome, 155,698 were detected in at least one sample, indicating good coverage.

After normalizing and comparing DHT and EtOH-treated samples from the same line and cell type, no transcripts were found to be significantly changed following FDR correction. The same stock of DHT successfully induced AR nuclear translocation in iPSCs in previous and subsequent experiments, suggesting that the reagent was functional; however, this may indicate that 48 h of DHT treatment was not sufficient to alter gene expression in these iPSCs, despite prior data to the contrary. On the other hand, no changes to expression would be expected in the i³LMNs if AR was not

present. As a result of this finding, subsequent analyses used only EtOH-treated samples, although additional analyses were performed using only DHT-treated samples, and nearly identical results were noted.

Twenty-five genes of particular interest were selected and analyzed as described in Chapter 2. Principally, *AR* was assessed, along with *GAPDH*, *ACTB*, and *TUBB* as positive controls. Next, established lineage marker genes were identified for pluripotent cells (*NANOG*, *POU5F1* [aka *OCT-3/4*], and *SOX2*), neural progenitors and glia (*NOTCH1*, *PAX6*, *GFAP*, and *OLIG2*), and neurons (Dai et al., 2019; Hu et al., 2010; Redwine & Evans, 2002). Neuronal markers were further broken down into general markers of a neuronal lineage (*TUBB3*, *SYP*, *MAP2*, *DCX*, and *NGN2*) as well as specific markers for cholinergic lower motor neurons (*ISL1*, *LHX3*, *MNX1* [aka *Hb9*], *CHAT*, and *ACHE*) and other neuronal subtypes; specifically, glutamatergic (*VGLUT2* [aka *SLC17A6*] & *GRIN1*), dopaminergic (*TH*), and GABAergic (*GABBR1*) neurons (Arber et al., 1999; Chambers et al., 2012; Mazzoni et al., 2011; Zhang et al., 2013b). Combined and scaled TPM values from all annotated transcripts of these genes are presented in Fig. 4.10.

The positive control genes *GAPDH*, *ACTB*, and *TUBB* demonstrated high and similar expression across all samples. While all three were significantly increased in i³LMNs compared with undifferentiated iPSCs of the same line, the relative differences between these were small, with $\log_2\text{FC} < 0.75$. Because quantile normalization was performed across all transcripts for these samples, the distribution of expression was equalized and thus no normalization by these reference genes was required. A slight bias towards i³LMNs in highly-expressed genes was acknowledged, and only differences of $\log_2\text{FC} > 0.75$ were considered notable. Furthermore, while no noise filter was applied to this data, previous noise modelling using all annotated transcripts (as described in Chapter 2) suggested that the coefficient of variance for genes with $\Sigma\text{TPM} < 6$, equivalent to $\log_2(\Sigma\text{TPM}+1) < 2.8$, was too high to confidently distinguish signal from noise. Thus, genes measured above this threshold may be considered to be expressed with high

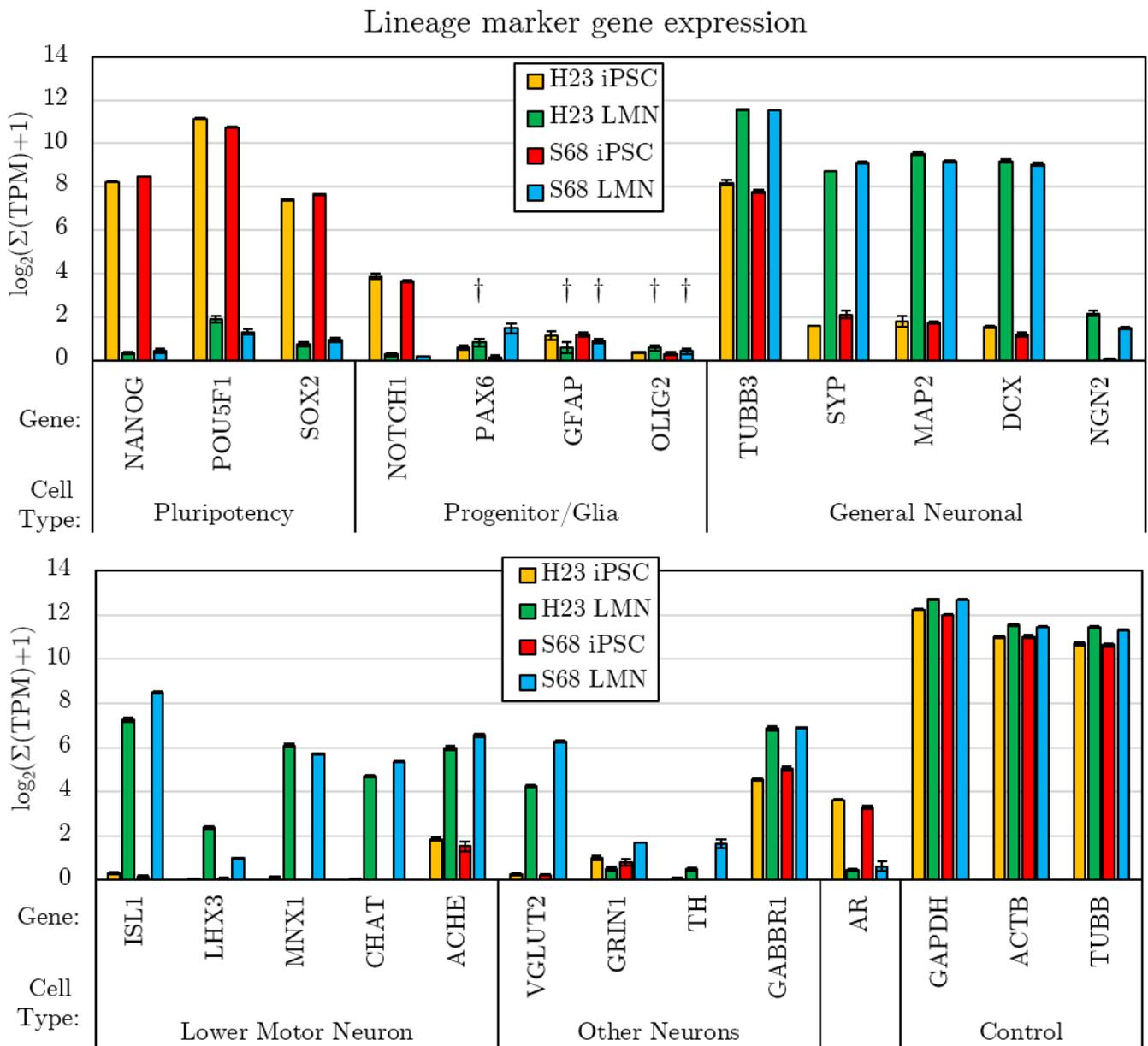
confidence, while those below this threshold may instead be due to noise, so confidence in their expression was reduced.

While iPSC lines expressed low but detectable levels of *AR* (mainly the full-length AR-201 transcript, between approximately 4 and 8 TPM), no i³LMN sample expressed any known *AR* transcript greater than 0.46 TPM, well below the noise threshold. Moreover, *AR* had a much higher coefficient of variance in i³LMNs (H23 = 19.3%; S68 = 63.6%) than in iPSCs (H23 = 0.80%; S68 = 3.74%), suggesting that what little *AR* transcript was detected in the i³LMNs was likely due to noise. The detection of *AR* transcripts in iPSCs, but not in the same cell lines as i³LMNs, provides strong evidence that *AR* was silenced during i³LMN differentiation.

Pluripotency marker genes were found to be strongly expressed in iPSCs, while all of these were strongly downregulated in i³LMNs. In addition, while *NOTCH1* was expressed in iPSCs, no neural progenitor or glial marker genes were upregulated in the i³LMNs. In fact, *PAX6*, *GFAP*, and *OLIG2* were the only genes for which t-tests between iPSC and i³LMN expressions yielded $p > 0.05$ after FDR correction. Additionally, general neuronal marker genes were nearly all highly-expressed in i³LMNs; *TUBB3* (which codes for β 3-tubulin) was also found to be expressed strongly in iPSC lines, but all were strongly upregulated in differentiated cells. Interestingly, *NGN2* expression was relatively low in i³LMNs; since these cells had ceased dox treatment eleven days prior to lysis, any expression of *NGN2* should only be due to endogenous transcription rather than induction of the hNIL insert. This also tracks with a reduction to *NGN2* expression during neuronal maturation (Hindley et al., 2012). Furthermore, a lack of detectable *NGN2* (as well as *ISL1* and *LHX3*) in the iPSC state showed that expression of hNIL is inactive in the absence of dox.

Figure 4.10: Expression of lineage marker genes in parental iPSCs and i³LMNs.

The sum of quantile normalized TPM data for all detected transcripts of each gene was rescaled, yielding $\log_2(\Sigma(\text{TPM})+1)$. Significance was assessed by two-tailed unpaired t-tests, followed by FDR correction for 25 tests. No significant differences were found between DHT and EtOH-treated samples; only data from EtOH-treated samples is shown for each line & cell type (n=3). DHT-treated samples demonstrate highly similar trends. All differences between iPSC and i³LMN samples were found to be significant except those denoted by † (indicating $p > 0.05$). No direct comparisons between H23 and S68 lines were performed due to their different genetic backgrounds.



Next, specific LMN lineage markers were assessed. While expression of *ISL1*, *MNX1*, *CHAT*, and *ACHE* were all highly upregulated in the i³LMNs, *LHX3* expression was low. While *LHX3* expression is characteristic of embryonic lower motor neurons that extend ventrally from the developing spinal cord, expression is downregulated in most LMNs as they mature, while *ISL1* remains expressed more widely (Sharma et al., 1998). This, along with reduced expression of *NGN2*, may suggest that these i³LMNs are more mature than a strictly embryonic state, although comparison of the full transcriptome with *in vivo* LMNs would be necessary to characterize these cells.

Furthermore, strong upregulation of *VGLUT2* suggests that the i³LMNs may be in part glutamatergic as well as cholinergic. *VGLUT2* in particular is known to be expressed in LMNs, and glutamate is used by LMNs as a neurotransmitter for dendritic feedback to interneurons (Herzog et al., 2004; Nishimaru et al., 2005). A relative lack of *GRIN1* and *TH*, meanwhile, showed that i³LMNs neither have the NMDA-type glutamate receptor nor produce dopamine, respectively. Finally, *GABBR1* expression was found to be slightly upregulated in i³LMNs, although its relatively high expression in iPSCs was notable as well. LMNs are known to be receptive to GABA, in particular from inhibitory interneurons, and expression of *GABBR1* is known to be widespread in non-neural tissues (Uhlén et al., 2015).

Also of note, expression of both *VGLUT2* and *GABBR1* were substantially less than that observed in i³Neurons that were differentiated for the same length of time. For both H23 and S68 i³LMNs, $VGLUT2 \log_2FC > 5.2$; $GABBR1 \log_2FC > 3.0$ (C. Ludwig & Kampmann, 2017). Thus, i³LMNs effectively acquired a neuronal phenotype that was distinct from that acquired by overexpression of *NGN2* alone, and, together with the other marker genes, support an identity similar to LMNs. Furthermore, changes to the expression of all these genes correspond well between the H23 and S68 cell lines despite different genetic backgrounds, suggesting that the presence of polyglutamine-expanded AR in the S68 iPSCs was not sufficient to prevent either the erasure of pluripotency or the later acquisition of a lower motor neuron phenotype.

On the other hand, these data confirmed that *AR* is effectively silenced in i³LMNs, suggesting that differentiation by hNIL induction was not amenable to studying the biology of *AR*, and thus it was not a viable method of differentiating the SBMA model lines.

4.5 Chemical differentiation produces DHT-insensitive LMN-like cells

As an alternative approach, an earlier passage of the SBMA model lines before addition of Mag2-hNIL was differentiated to an LMN-like state using a small molecule inhibitor approach (Chapter 2; Hall et al., 2017). After completion of this protocol, cells from all lines were found to downregulate NANOG; while iPSCs contained high levels of nuclear staining; only occasional non-nuclear puncta were observed after differentiation (Fig. 4.11a). In addition, mitotic cells were common in iPSCs, and nuclei were large, round, and stained diffusely with DAPI. After differentiation, no mitotic cells were observed, and nuclei were smaller and stained more intensely with DAPI, indicative of a differentiated state.

Differentiated cells were also found to often co-express nuclear Hb9 and β 3-tubulin, suggestive of a lower motor neuron cell state. However, it was also found that the resulting cell cultures contained a subset of cells that did not express these markers (Fig. 4.11b). This suggests that the differentiated cultures featured some degree of cell type heterogeneity, limiting their usefulness for biochemical studies. It was also noted that cells expressed variable levels of ChAT, and *AR* was detected, albeit at low intensity (Fig. 4.11c).

Figure 4.11: Characterization of chemically differentiated cells.

Undifferentiated iPSCs and cells after chemical differentiation were stained for the indicated markers, imaged, and presented as a single confocal slice through the nucleus. No notable differences were observed between cell lines or treatments, so the H68 line with EtOH treatment is shown as a representative example.

- (a) NANOG was observed to be highly expressed in the nuclei of iPSCs. Differentiated cells (diff) were imaged with a higher gain to account for low intensity and contained only non-nuclear puncta (e.g., white arrow). Nuclear characteristics were also suggestive of undifferentiated and differentiated cells, respectively.
 - (b) Most differentiated cells co-expressed Hb9 and β 3-Tubulin, but a subset of cells were noted to stain less brightly for Hb9 and appeared to lack β 3-Tubulin (e.g., white arrow).
 - (c) Most differentiated cells co-expressed ChAT and cytoplasmic AR. However, intensity of ChAT staining was variable, and intensity of AR was low.
- (continued)*

Figure 4.12: Chemical differentiation disrupts AR nuclear translocation.

After chemical differentiation, cells were treated with either EtOH or 10 nM DHT for 48 h and imaged by confocal microscopy.

- (a) H23 cells after EtOH treatment. No nuclear AR staining was observed, but cells with a non-neuronal morphology were observed to grow flat against the bottom of the well (white arrow). These cells had diffuse ChAT staining and larger cell bodies without long, thin processes similar to neurites.
 - (b) H23 cells after DHT treatment. Nuclear AR staining was only observed in cells with a non-neuronal morphology (white arrows).
 - (c) H68 cells after DHT treatment. Nuclear AR was more pronounced in the polyQ-expanded cell line, but this also was observed only in cells with a non-neuronal morphology (white arrows).
 - (d) HKO cells with DHT treatment were used to calibrate background fluorescence. Staining intensity was only slightly higher than HKO for non-KO LMN-like cells, indicating AR expression was very low in these cells.
- (continued)*

Figure 4.11 (*continued*): Characterization of chemically differentiated cells.

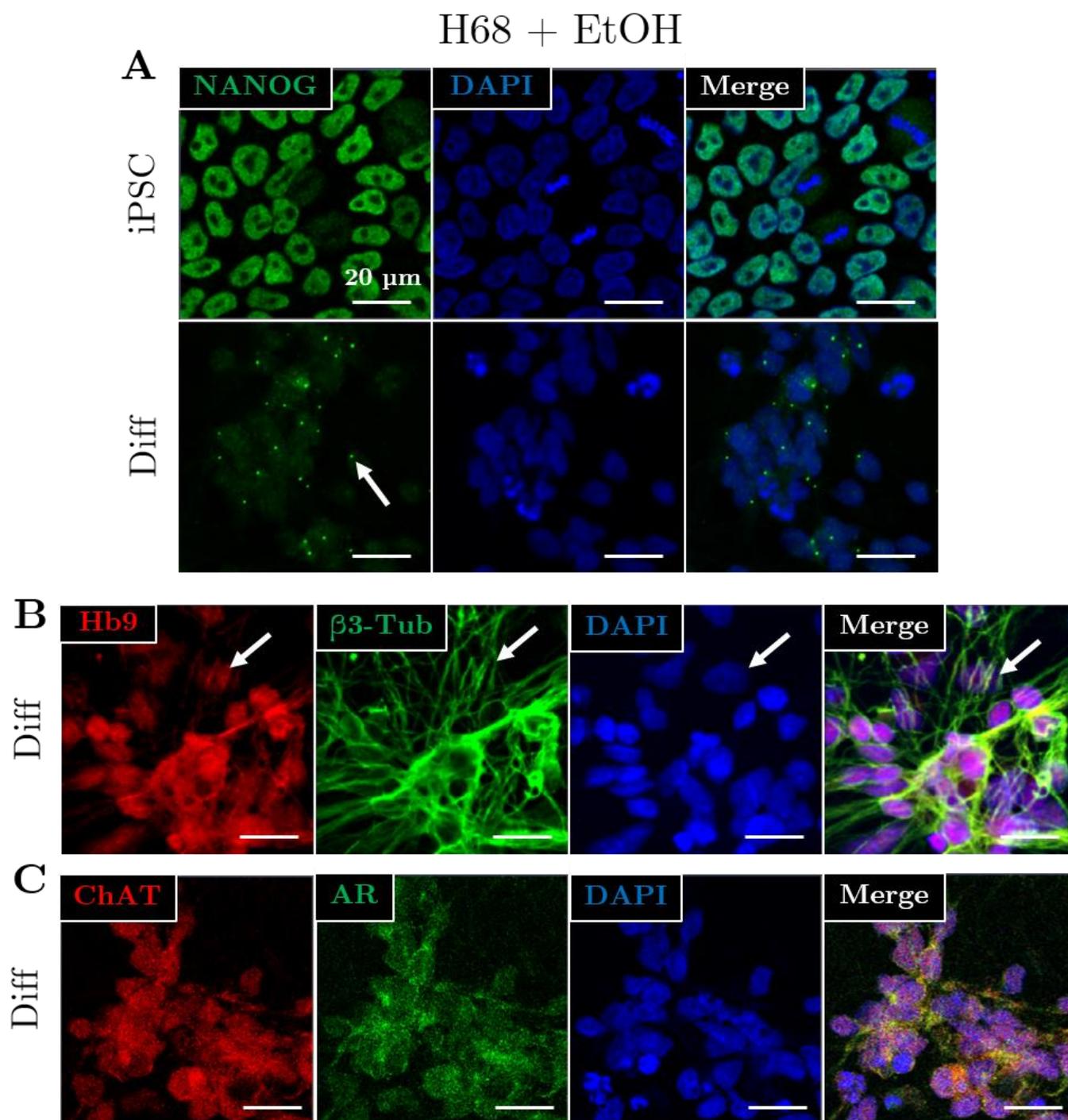
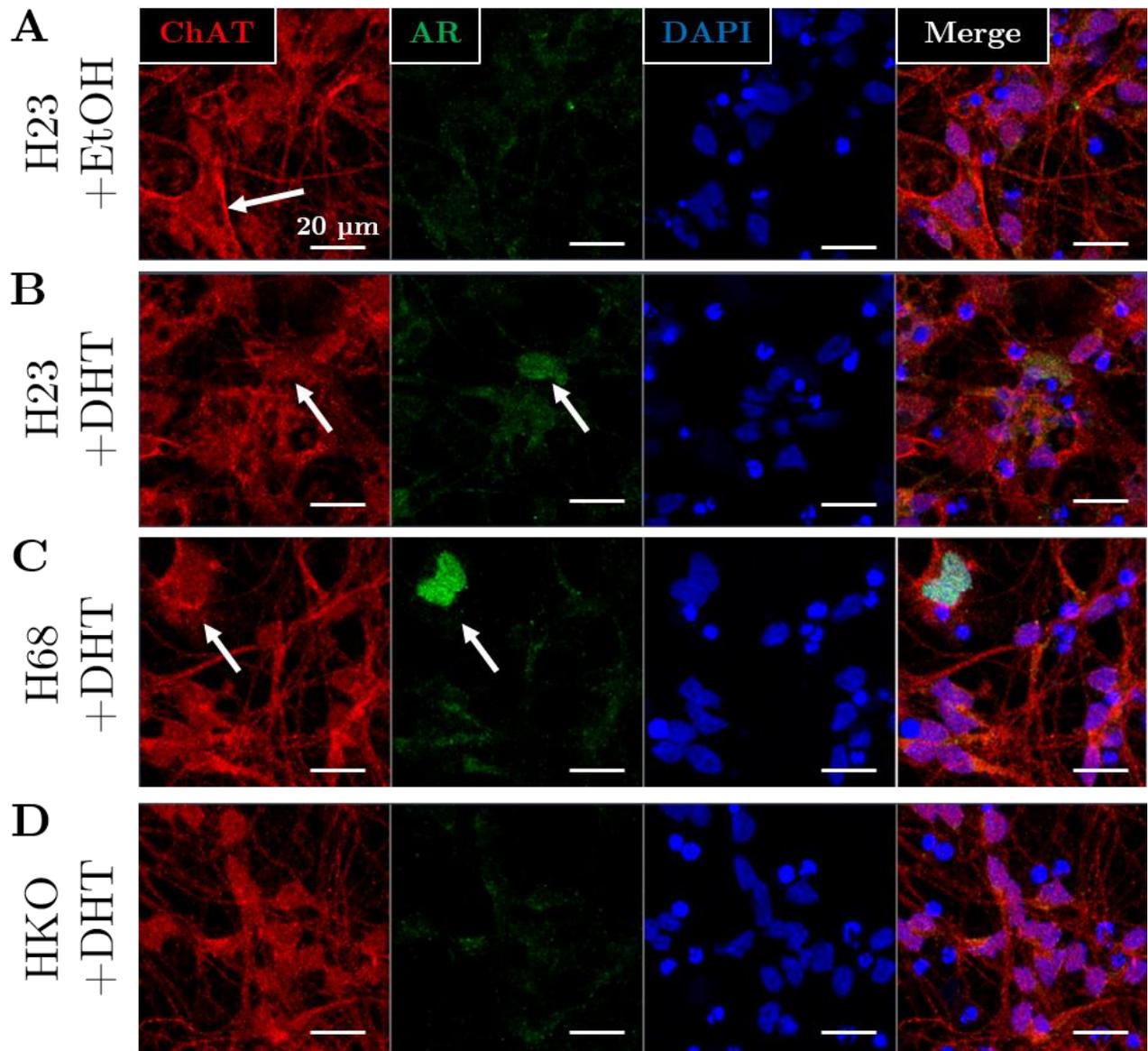


Figure 4.12 (*continued*): AR nuclear translocation is disrupted.



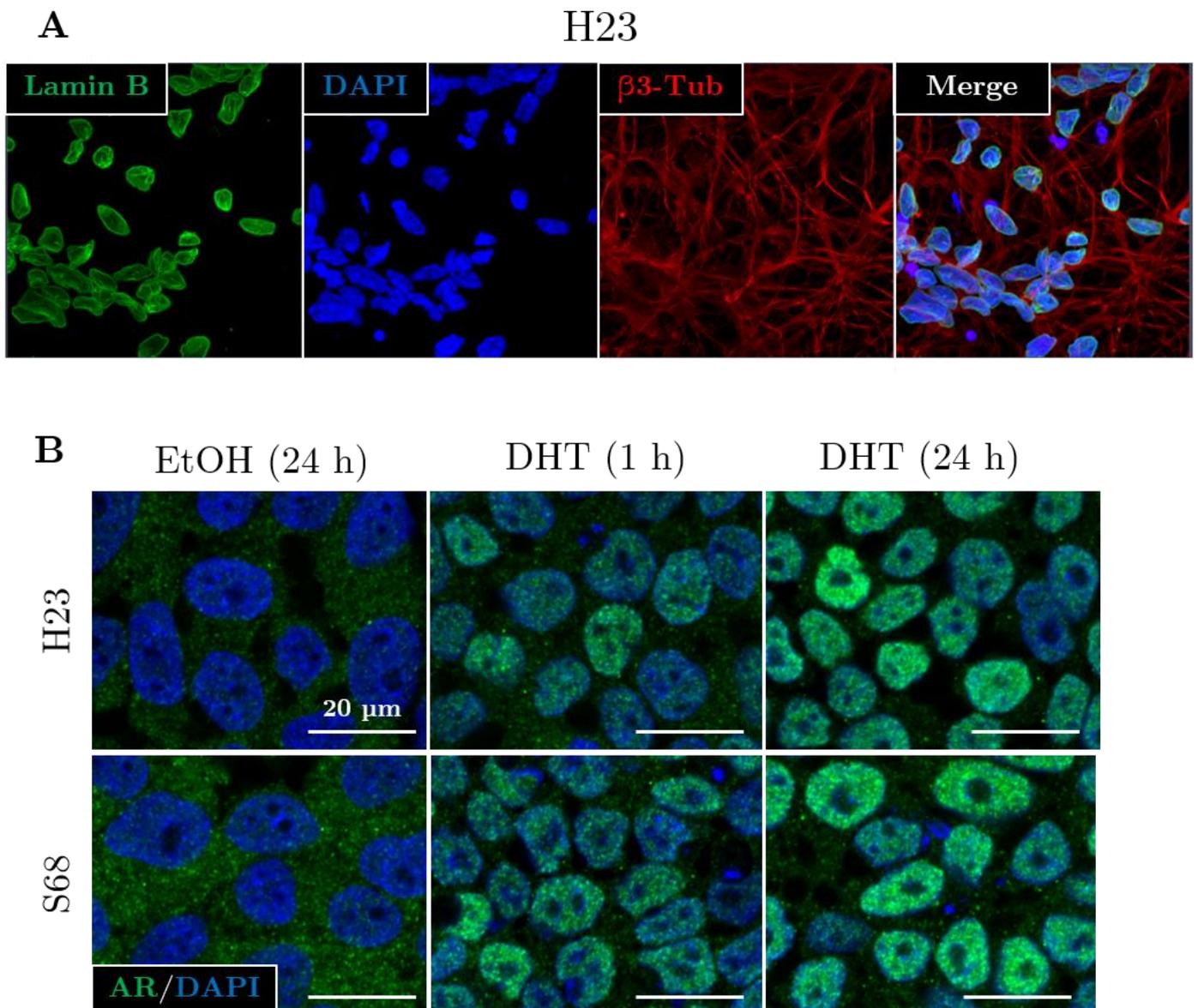
Troublingly, while AR was detectible at low levels in the LMN-like cells by immunocytochemistry, treatment with DHT did not induce nuclear translocation in these cells. Instead, the only cells that showed AR nuclear translocation were larger, flatter cells that lacked neurite outgrowth and did not stain for motor neuron markers. These cells were noted to be primarily tight against the culture substrate and underneath the network of neurites and neural cell bodies, as evidenced by confocal imaging of the same (x,y) location at different z coordinates (Fig. 4.12).

It was recognized that this effect may have been due to an inability for AR antibodies to penetrate the nuclei of the LMN-like cells or to a degraded DHT stock resulting in only partial AR activation. Unfortunately, heterogeneity of DHT responsiveness within the culture prevented the use of bulk nuclear fractionation as another method of determining AR localization. Immunofluorescence staining of Lamin B showed that antibodies were able to penetrate the nuclei of these cells (Fig. 4.13a), and undifferentiated iPSCs cultured alongside these neurons had strong nuclear translocation of AR upon treatment with the same stock of DHT, demonstrating its efficacy (Fig. 4.13b). It was therefore concluded that any AR present in these LMN-like cells was not functionally sensitive to DHT treatment.

As mentioned, previous research has established that the toxic gain-of-function effects in SBMA are dependent on AR polyQ expansion, DHT activation, and nuclear translocation, so a lack of a clear response to DHT in the putative LMN-like cells was deemed sufficient exclusionary criteria to end the project.

Figure 4.13: Lack of nuclear AR staining was not due to technical issues.

- (a) Nuclei of chemically differentiated neurons were permeable to Lamin B antibodies using the same protocol as previous AR staining (4% PFA fixation and treatment with 0.2% Triton X-100 in PBS). This indicated that a lack of nuclear permeabilization was not the cause of the lack of nuclear AR staining.
- (b) Undifferentiated iPSCs exhibited efficient AR nuclear translocation following DHT treatment. This indicated that 10 nM DHT was sufficient to induce AR translocation and that the chemical stock was potent.



4.6 Discussion

4.6.1 Summary of findings

Derivation of the eight novel SBMA model lines and the numerous insertions and excisions of Mag-hNIL and Mag2-hNIL described in this chapter represent a large investment of time, effort, and research funds. It is only since the relatively recent development and widespread use of both iPSCs and CRISPR/Cas9 that such disease modelling has been possible, and this project was undertaken as an ambitious combination of both technologies.

CRISPR-editing of the endogenous *AR* locus was performed and validated in iPSCs, and these lines hold promise for future research. The subsequent work described in this chapter consisted mainly of obstacles encountered and efforts to overcome them. None of these obstacles were necessarily foreseeable, although they may be explainable in hindsight. The structure of this chapter was therefore intended to highlight shortcomings of iPSC-based disease modelling, particularly for discovery-based research, and to promote an understanding of the risks and rewards of such a venture.

It is my hope that this chapter will help other researchers to avoid the same problems. These studies led directly to improved protocol recommendations, as with SBP- Δ LNGFR not being amenable to polycistronic expression (Fernandopulle, Prestil et al., 2018), and indirectly through collaborative feedback to the development and proliferation of new tools to facilitate experimental planning, as with the iNeuron RNA-Seq web app (Ludwig & Kampmann, 2017; Tian et al., 2019). In particular, the data presented through this app suggests that *AR* mRNA is also strongly reduced in i³Neurons. Interestingly, this data notes an increase in *AR* expression at later time points and with BrainPhys rather than Neurobasal as a basal medium, suggesting that a longer period of maturation and an adjusted media formulation may be sufficient for *AR* reactivation.

Before discovering that *AR* was silenced, it was my intention for this project to directly compare the transcriptional profiles of all of the lines in the two allelic series. Transcripts that correlated expression with the length of the polyglutamine tract, but not with *AR* KO, would be strong candidates for specifically mediating a toxic gain-of-function and would thus merit further study. It would be worthwhile to test whether *AR* can be reactivated by protocol adjustments in order to pursue these aims. As explored below, additional treatment may also be able to restore *AR* functionality in chemically differentiated LMN-like cells, and it may be relevant to better understanding and treating SBMA to discern mechanisms by which *AR* can be restricted to the cytoplasm in such cells.

It may also be important to critically assess the DHT-sensitivity of LMNs *in vivo*, particularly as an approach to address another outstanding question: whether neurotoxicity in SBMA is due to cell-autonomous mechanisms or whether it requires interactions with other cell types. For example, it has been posited that primary muscle degeneration may cause motor neuron death due to a lack of trophic feedback, rather than the traditional model of motor neuron degeneration leading to muscle denervation and atrophy (Gromova & La Spada, 2020). The data in this chapter tend to support the muscle-first theory, since low *AR* expression and responsiveness to DHT in LMNs would make it unlikely for these cells to be the primary source of toxicity. Future work may seek to differentiate these lines into a skeletal muscle phenotype, for instance via induction of *MYOD* (Uchimura et al., 2017).

The reason why rtTA3G was inactive in Mag-hNIL was shown to be interference by the selection genes, particularly SBP- Δ LNGFR, in a polycistronic cassette. This obstacle was overcome, albeit at a substantial loss of time due to my attempt to restore rtTA3G activity via Cre excision and the eventual necessity of redesigning and reintegrating the Mag2-hNIL plasmid. However, it was crucial to the development of SBP- Δ LNGFR as a marker gene in iPSCs that its use be shown to be compatible with the dox-inducible expression system, so this process proved to be worthwhile. The

resultant Mag2-hNIL cells were found to be both selectable with magnetic streptavidin-coated beads and differentiable by dox induction of the hNIL factors. The remainder of this chapter will explore the two remaining obstacles that were not overcome: the silencing of AR in i³LMNs and the lack of observed AR nuclear translocation in response to DHT in chemically differentiated LMN-like cells.

4.6.2 Retrospective on AR silencing in i³LMNs

It is unclear why AR was silenced following hNIL and hNGN2 differentiation, although several possibilities arise. Overexpression of hNGN2 (with or without additional factors) has been shown to accelerate differentiation past the progenitor state and directly to neurogenesis, suggesting that the epigenetic reprogramming that occurs during i³LMN differentiation may not fully reflect the changes that occur during motor neuron differentiation *in vivo*. Future work may therefore seek to specifically assess epigenetic modifications to the endogenous *AR* promoter before and after differentiation as well as the expression of the *AR* promoter via a reporter clone.

The idea that AR silencing in i³LMNs represents a deviation from normal development rests on the assumption that AR is actually expressed in lower motor neurons *in vivo*. Several studies have reported that AR protein is present (and accumulates) in SBMA patient motor neurons (Adachi, 2005; Mei Li et al., 1998), and radiolabeled DHT was shown to localize to motor neuron nuclei in rats (Sar & Stumpf, 1977). On the other hand, two large-scale human proteomics databases report no detectible AR protein in the healthy spinal cord (Kim et al., 2014; Wilhelm et al., 2014), and mouse spinal cord *in-situ* hybridization does not detect *AR* mRNA in lower motor neurons of either embryonic or adult mice (Allen Institute for Brain Science, 2008). At best, human RNAseq databases report *AR* mRNA TPM values in the low single-digits from spinal cord samples, similar to those seen here and which also fall below the noise threshold

used in this work (Uhlén et al., 2015). The conclusion that may be drawn is that *AR* is normally expressed at low levels in lower motor neurons; as explored below, nuclear translocation and polyQ expansion may further protect AR from degradation. In the context of differentiation, a reduction of *AR* expression to below detectible levels may actually be more reflective of physiological LMNs than previously thought.

Another study from our lab used a similar approach, with the same wildtype and AR KO cell lines and i³LMN differentiation using the pUCM-CLYBL-hNIL insert, a precursor for Mag2-hNIL (Pourshafie et al., 2020). This study detected AR protein after hNIL induction, and moreover AR ChIPseq was done, which would not be possible without the presence of AR.

Differences in the techniques used to differentiate cells may account for these different results and reveal additional aspects of AR biology during differentiation. The cells used by Pourshafie et al. (2020) were differentiated for either four or six days (depending on the assay) and treated throughout differentiation with DHT, while cells in this work were differentiated for fourteen days and only introduced DHT treatment for the final two days. Two possibilities thus arise to explain the discrepancy between these works: *AR* mRNA may be transcribed differently, or AR protein may be degraded differently.

Notably, raw *AR* TPM values from RNAseq that are reported by Pourshafie et al. (2020)—approximately 1.8 TPM for both WT and SBMA lines—are similar to those reported here, and data from both studies fall below the threshold of noise applied to the data in this work, which required at least 6.0 TPM to be considered detectable. In addition, undifferentiated iPSC samples measured here had *AR* expression of approximately 16.0 TPM, nearly 9-fold higher than those seen in i³LMNs in both studies. More nuanced noise filtering may provide better sensitivity, but at best, *AR* mRNA was expressed at a very low level in both studies, and both are consistent with the physiological data noted above. It is therefore unlikely that *AR* transcription was markedly different between these studies.

Inactive AR is retained in the cytoplasm and thus may be more easily degraded, but early and continuous DHT treatment by Pourshafie et al. (2020) likely activated AR and induced its nuclear localization throughout the differentiation process. AR in the nucleus has been shown to be less susceptible to degradation in prostate cancer cells (Gong et al., 2012), which may translate to a longer AR half-life in iPSCs and during differentiation, particularly since mitosis is arrested within three to five days in hNGN2 and hNIL-driven differentiation (Fernandopulle, Prestil et al., 2018). Additionally, DHT treatment was shown to increase levels of AR protein expression in iPSCs, so AR activation with DHT before *AR* silencing may support *AR* expression for longer into differentiation, while a lack of AR activation may lead to earlier *AR* silencing.

Dynamic stable isotope labeling with amino acids in cell culture (SILAC) proteomics was developed to determine the rate of protein turnover by conditioning cell cultures in media containing a specific isotopic weight of the essential amino acids arginine and lysine, and then switching to a different (“heavy”) isotope at multiple time points. Proteomics is then used to measure the relative abundance of the different isotopes in each peptide and thereby calculate the rate of protein degradation (Doherty et al., 2009). The median half-life of proteins in mouse primary cultures of mature hippocampal neurons and glia was shown to be 5.4 days (Dörrbaum et al., 2018). In contrast, mouse embryonic neurons were shown to have much faster protein turnover, with a median half-life of just 1.9 days (Mathieson et al., 2018).

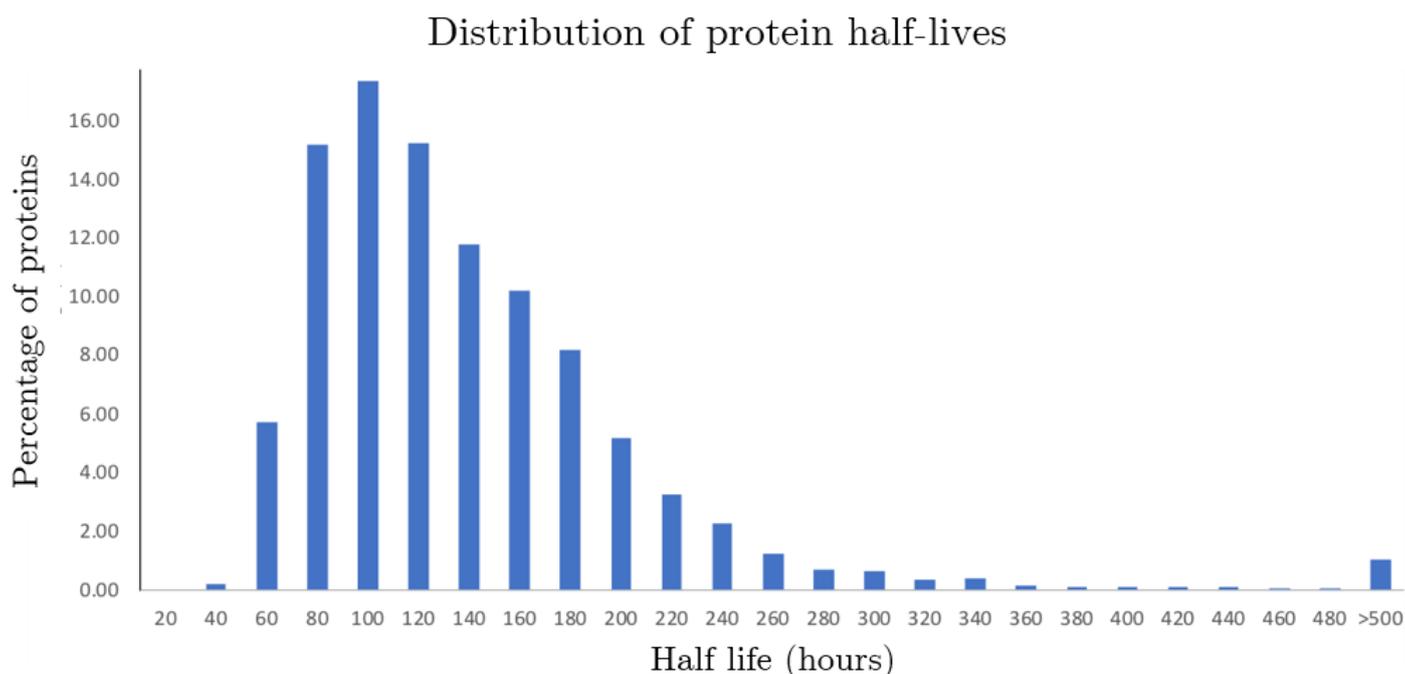
As part of a collaboration with Ling Hao and Michael Ward, I adapted dynamic SILAC to label iPSCs and ³H neurons during differentiation, and Ling Hao performed proteomics and analysis in order to calculate the half-lives of proteins. Many long-lived proteins were identified, and the overall median protein half-life after 14 days of differentiation was found to be 4.8 days (Fig. 4.14; Hao et al., unpublished). It is likely that ³H LMNs feature similarly slow protein dynamics, and therefore the AR seen in the cells of Pourshafie et al. (2020) may be carried over from the iPSC state and retained

during differentiation while transcription and translation of new AR is lost. Additional isotopic labelling experiments may be performed to investigate this possibility.

While a shorter duration of differentiation may enable the detection of AR protein and avoid cellular toxicity noted in SBMA cells at later time points, it sacrifices maturation, as the neurite network continues to grow and gene expression continues to change past day 6 in i^3 Neurons and i^3 LMNs (Tian et al., 2019; Zhang et al., 2013b). The hNIL differentiation process allows cells to attain some aspects of LMN morphology and gene expression within four days, but, as mentioned in Chapter 1, differences in the duration of dox induction and maturation can have significant effects on cell identity.

Figure 4.14: Protein half-lives in i^3 Neurons.

Dynamic SILAC proteomics was performed in WT i^3 Neurons that had been labeled at multiple time points during differentiation in order to calculate the rate of degradation of each protein. The half-life of 2800 proteins was calculated and binned into 20-hour segments. The median half-life was found to be 115.2 hours (4.8 days). From Hao et al., unpublished.



In general, the data reported here and by Pourshafie et al. (2020) are compatible. However, an abbreviated differentiation protocol and constitutive treatment with DHT necessitates further investigation to validate potential disease mechanisms. It may be worthwhile to replicate these findings using the Mag2-hNIL SBMA cells either as iPSCs or by partially-differentiating the cells in a similar manner such that AR is not fully degraded by the time of lysis. The lines derived here have the added benefits of isogenicity and a range of polyQ lengths, so the sufficiency and necessity of polyQ expansion for the observed effects could be tested rigorously.

4.6.3 Retrospective on DHT insensitivity in chemical differentiation

Chemical differentiation was pursued as an alternative after the inability to resolve AR silencing during i³LMN differentiation. At the time, it was considered likely that this process would produce viable LMN-like cells in the SBMA model lines because a previous study from the our lab had successfully used a similar protocol to differentiate the same parental lines (Grunseich et al., 2014b). Here, however, AR did not translocate to the nucleus in response to DHT treatment after chemical differentiation in LMN-like cells, while non-neuronal cells did respond to DHT.

Again, no conclusive explanation was forthcoming, although these observations correspond with generally low expression of AR in LMNs, while the cells that escaped neuronal commitment may retain higher AR expression. Single-cell RNAseq could be used to elucidate the lineage identities represented in a mixed culture setting and could test whether there is a negative correlation between expression of motor neuron marker genes and AR.

As mentioned in Chapter 2, the chemical differentiation protocol used in this work was adapted from Hall et al. (2017), which reported >85% cell type purity. This is in part due to differentiation being performed entirely on adherent cells, which promotes

equitable exposure to the small molecules in the media and thereby reduces cell type heterogeneity. The differentiation protocol used by Grunseich et al. (2014b) was based on Amoroso et al. (2013). These protocols used largely the same small molecules in the same order; the only differences for induction were that I added CHIR99021 and did not use recombinant FGF, and for patterning I used only purmorphamine rather than smoothed agonist and purmorphamine. The main difference was that Grunseich et al. (2014b) cultured cells in suspension as embryoid bodies for the first two weeks, leading to greater cell type heterogeneity (Bauwens et al., 2008). Grunseich et al. (2014b) likewise reported a relatively low percentage of cells as LMN-like, based on expression of either ISL1 or Hb9.

Grunseich et al. (2014b) noted that the cells shown to feature the highest intensity of nuclear AR staining in response to DHT treatment were the cells that lacked appreciable Isl1/Hb9 staining and that low levels of nuclear AR staining were observed in all cells in the absence of DHT treatment. In contrast, I found that AR was not detected in the nuclei of LMN-like cells regardless of treatment. These studies used different antibodies to detect AR, which may account for some variation; ectopic expression of a tagged AR could improve the quality and intensity of staining to better determine whether or not AR is present in the nucleus.

Because the differentiation protocol in both cases extends over four to six weeks, it is unlikely that any AR protein translated in undifferentiated iPSCs remained in the differentiated cells, so the AR detected by immunofluorescence and western blot can be attributed to the differentiated cells. However, AR staining suggests that much of the AR detected in pooled protein samples may be attributed to alternative cell types rather than the LMN-like cells. Grunseich et al. (2014b) also used a longer DHT treatment period than me, of approximately ten days. Unlike Pourshafie et al. (2020), this treatment was only started after neuronal plating, reducing potential effects on differentiation itself.

One possibility considered by Grunseich et al. (2014b) was that differences in DHT sensitivities could be due to interactions of AR with CRM1, a cofactor that has been shown to promote AR nuclear export and which increases this function when GSK-3 β is inhibited in prostate cancer cell lines (Schütz et al., 2010). Grunseich et al. (2014b) reported that GSK-3 β phosphorylated at serine 9 (a marker of inactivation; Fang et al., 2000) was unchanged between WT and SBMA patient lines; however, more rigorous analysis is warranted. Furthermore, Grunseich et al. (2014b) noted that the cells that stain most strongly for p-GSK-3 β were also those that stained for Isl1/Hb9, suggesting that this may be a mechanism by which AR is excluded from the nucleus in specifically LMN-like cells despite DHT treatment.

In unpublished work, the authors of Grunseich et al. (2014b) tested the effect of CRM1 inhibition using leptomycin B; however, this was not sufficient to reinstate DHT sensitivity in the LMN-like cells. Additional experiments would need to be performed to determine whether phospho-inactivation of GSK-3 β and CRM1-mediated export is in fact responsible for AR insensitivity to DHT treatment, particularly since the GSK-3 β inhibitor CHIR99021 was used to activate WNT for the first eight days of differentiation. By the endpoint, it had not been used for twenty days, but this initial period may result in long-lasting GSK-3 β inactivation.

Grunseich et al. (2014b) were able to validate their findings in pooled protein samples by immunocytochemistry and in other SBMA cell models, lending confidence to their conclusions despite a heterogeneous mixture of cell types in culture. However, the goal of this work was to identify biological differences in LMN-like cells that may contribute to disease pathology. Had the motor neurons shown a robust response to DHT, the heterogeneity of cell types present may have been accounted for, but without such a response, there was no way to determine whether any effects seen in culture-wide protein or RNA samples were due to the influence of non-neuronal cells.

4.6.4 Lessons learned

With the benefit of hindsight, there are several aspects of this study that I would do differently were I to undertake this project today. First, the reliance on a majority of CAA codons in the *AR* insert likely does not capture RNA-mediated toxicity. While this was required for GeneArt synthesis, I would instead construct the insert plasmids with pure CAG repeats, possibly by using synthesis of long iterative polynucleotides (SLIP; Figura et al., 2015). Low *AR* integration efficiency was also noted, so I would use more efficient methods for insertion, such as Cas9-CtIP to promote HDR (Charpentier et al., 2018) or ribonucleoproteins with insert sequences conjugated onto the Cas9 or gRNA (Carlson-Stevermer et al., 2017; Ling et al., 2020).

The obstacles faced in this study also could have been mitigated with more and earlier proof-of-concept; for instance, I could have tested differentiability after Mag-hNIL addition or Cre excision in a single cell line rather than undertaking these edits in all ten lines in parallel. Faster recognition that AR was not expressed after Mag2-hNIL differentiation would also have permitted more troubleshooting, for example by trying longer maturation and using BrainPhys as a basal medium. Because my data showing loss of AR appeared to conflict with reports of others in our lab and did not have a basis in the literature at the time, it was not until several replicates of orthogonal approaches supported this hypothesis that I took it seriously as a biological effect. In hindsight, I should have focused on *AR* silencing at once rather than performing experiments to characterize the i³LMN cell type and assess possible SBMA phenotypes, the results of which were not useful without *AR* expression.

The chemically differentiated LMN-like cells were not an ideal model system due to cell type heterogeneity, so their usefulness would have been limited even if DHT-induced nuclear translocation was observed. In hindsight, the protocol could have been modified to remove CHIR99021 to prevent GSK-3 β inhibition at any point during differentiation, and I could have tried treating LMN-like cells with a CRM1 inhibitor

such as leptomycin B. I also could have assessed GSK-3 β and CRM1 activity in these cells directly, both between LMN-like and non-neuronal cells in the same culture and between WT and SBMA lines.

In the end, I did not meet the main aim of this project, which was to separate the loss- and gain-of-function effects of the polyQ expansion in SBMA. Because this aim relied on extensive technical development prior to data acquisition, the delays in deriving and characterizing the model lines compounded, and despite taking several approaches to differentiation, I eventually faced an insurmountable obstacle on each one. While the main aim was not met, the aims of deriving the edited lines and performing the differentiations were achieved, which allow me to conclude that the issues arose from biological, rather than technical, sources.

Were I to redesign the project from the point of recognizing AR silencing in i³LMNs, I would instead pivot to differentiation of skeletal muscle rather than performing chemical differentiation to an LMN-like state. I initiated a collaboration to pursue this approach, but I could have explored this avenue myself when this did not move ahead. Previous studies suggest that AR is expressed at high levels in skeletal muscle cell lines (Choi et al., 2020; Ting & Chang, 2008; Wannenes et al., 2008), and Choi et al. (2020) specifically predicts an increase in *AR* expression in myotube-like cells after chemical differentiation from iPSCs, although AR expression was reported to be reduced in intermediate progenitors (available at myogenesis.net). The role of polyQ-expanded AR in muscle cells is an important question for understanding SBMA, and the isogenic series cell lines may provide valuable insight.

Chapter 5: *L1CAM* is a novel regulator of autophagy

5.1 Introduction

5.1.1 L1CAM in health & disease

The L1 cell adhesion molecule (L1CAM) is a protein that is important for cell adherence, growth, and motility. It is highly expressed in neurons, and it has been shown to play a critical role in axonal extension, neuronal migration, myelination, and synaptogenesis (Kallunki et al., 1997; Kamiguchi et al., 1998; Maretzky et al., 2005). Expression of L1CAM is activated early in development, although its importance continues well into adulthood (Grońska-Peński et al., 2020). Its conformation and binding properties are disrupted by ethanol binding, and this property may be a major contributor to the pathology of fetal alcohol spectrum disorders (Bearer, 2001; Dou et al., 2018; Ramanathan et al., 1996).

Mutations in the *L1CAM* gene are known to cause a variety of neurological disorders, collectively termed L1 syndrome (Stumpel & Vos, 2004). The patients commonly manifest a combination of corpus callosum hypoplasia, intellectual disability, adducted thumbs, spastic paraplegia, and hydrocephalus, leading the acronym CRASH syndrome. The severity of symptoms and prognosis have been found to correlate with the mutation characteristics; full gene deletions and nonsense mutations typically cause the most severe disease, with over half of those born with truncating mutations in one study dying within three years (Chidsey et al., 2014; Vos et al., 2010). Milder forms of disease have been associated with point mutations, and over 220 disease-causing genetic variants have been reported to date (Bousquet et al., 2021; Vos & Hofstra, 2010).

While a loss of L1CAM is often fatal, its overexpression in non-neuronal tissues can be an indicator of tumorigenesis (Chen et al., 2013). It is well-documented that L1CAM expression in a variety of cancers correlates with the development of metastases and chemoresistance and promotes angiogenesis, all of which are associated with poor prognosis (Angiolini et al., 2019; Maten et al., 2019). Targeting *L1CAM* via lentiviral-encoded shRNAs has been shown to attenuate tumor growth in mice (Hung et al., 2010). As a surface protein, the possibility of targeting L1CAM with immunotherapies has also been explored, with mixed results (Doberstein et al., 2015). A recently-completed clinical trial found that L1CAM levels in serum were significantly elevated in endometrial cancer (NCT04603599; Sertel et al., 2019), but similar studies in other cancers have reported conflicting data (Chu et al., 2020). There is also an ongoing phase 1 clinical trial of L1CAM-targeting CAR T cells for treatment of neuroblastoma, although as of this writing no results have been posted (NCT02311621).

5.1.2 Biology of L1CAM

L1CAM is a member of a large class of evolutionarily-conserved cell adhesion molecules, with three paralogs in humans and orthologs with high sequence and structural homology as distant as *Drosophila* and *C. elegans* (Godenschwege et al., 2006; Hortsch, 2000). The L1CAM protein is situated on the cell membrane, with a majority of the protein projecting out into the extracellular matrix. The first 1120 amino acids from the N-terminus comprise the extracellular domain (ECD) consisting of six immunoglobulin-like domains followed by four fibronectin type III domains; these structures enable interactions with numerous proteins, including itself, to form cell-cell junctions and transduce signals into the cell (Maten et al., 2019). This is followed by a 23 amino acid single-pass transmembrane domain (TMD) that connects to the remaining 114 amino acid intracellular domain (ICD). The ICD contains protein-

binding and phosphorylation sites that mediate interactions with several signaling pathways (Herron et al., 2009).

The *L1CAM* gene is located on the X chromosome and consists of 29 exons. The most upstream exon in the longest transcript variant is designated here as exon 1 in accordance with RefSeq and Ensembl nomenclature (Howe et al., 2021), but this exon is sometimes referred to as exon 0, exon A, or simply ignored in the literature because it consists of only the 5' UTR, is located 10kb upstream of exon 2, and is not included in all transcripts (Pfeifer et al., 2010). A map of the *L1CAM* gene and target sites for the various genetic manipulations used in this work is presented in Fig. 5.1.

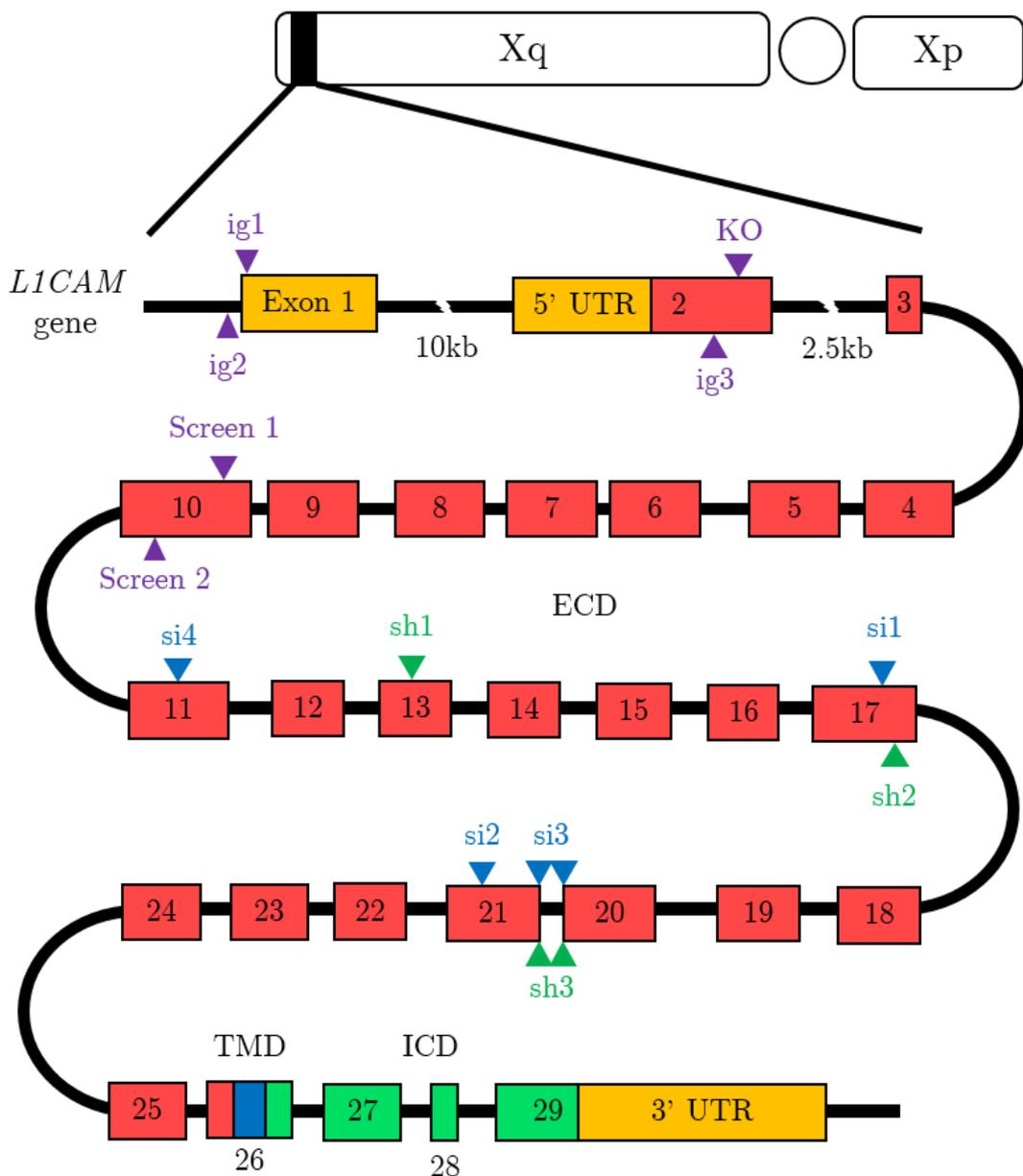
While neurons have been shown to preferentially express the full-length isoform, many additional splice variants have been documented, including transcripts lacking exons 3, 26, and 28 (Reid & Hemperly, 1992; Takeda et al., 1996). Exon 3 contains just 15bp, but it encodes an important binding motif such that both homophilic and heterophilic binding are reduced when it is lost (De Angelis et al., 2001). The transmembrane domain is located entirely within exon 26, and skipping this exon results in a soluble form of L1CAM that is released into the extracellular environment as a signaling molecule that promotes angiogenesis in tumors (Angiolini et al., 2019). Exon 28 is another short sequence of just 12bp, but it encodes a key regulatory motif that enables clathrin-mediated endocytosis of L1CAM (Kamiguchi et al., 1998). Targeted deployment and reuptake of adhesion molecules is important for neurite outgrowth and cell motility.

L1CAM is known to have several cleavage products; plasmin, trypsin, and MBP can digest L1CAM in the fibronectin domains, releasing the Ig-like domains to the extracellular matrix (Kleene et al., 2021). Several metalloproteases can also cleave the ECD near the membrane, releasing almost the entire ECD and leaving only the TMD and ICD attached to the membrane (Kiefel et al., 2012a).

Figure 5.1: Map of *L1CAM* features and RNA targets.

L1CAM is located near the end of the long arm of the X chromosome. Lengths of features are not to scale. Exons are labeled with their respective numbers; orange segments are 5' and 3' UTRs, red segments encode the extracellular domain (ECD), the blue segment is the transmembrane domain (TMD), and green segments encode the intracellular domain (ICD). Introns 1 and 2 are particularly long (10kb and 2.5kb, respectively); others are less than 1kb. Exons 3 and 28 are short (15 and 12bp, respectively); these and exons 1 and 26 are not included in some transcripts.

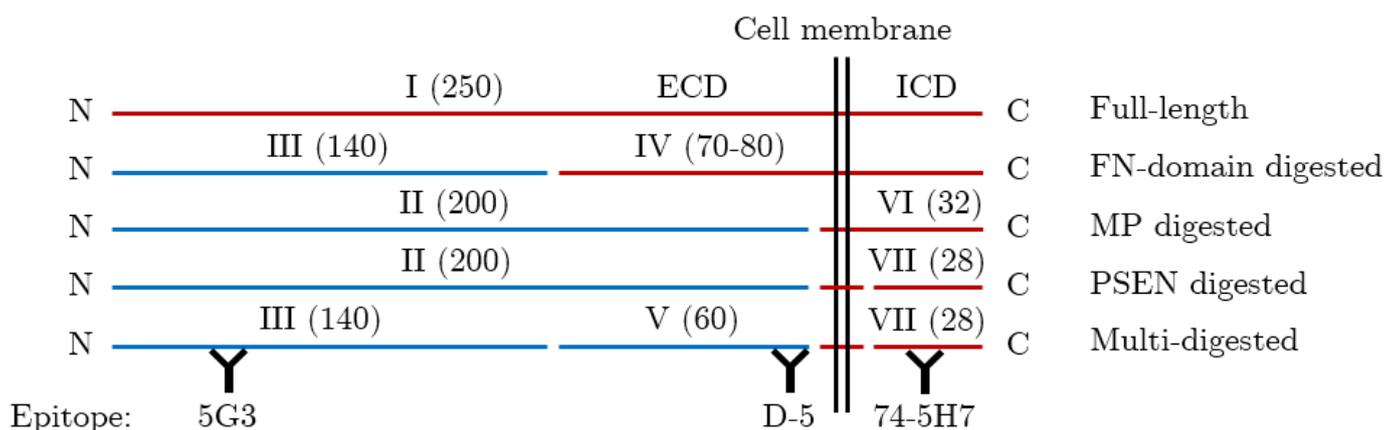
CRISPR gRNA targets are in purple, including targets for CRISPRi (ig1-3), the gRNA used for inducing knockout clones (KO), and for the initial screen (Screen 1 & 2). siRNA targets are in blue (si1-4), and shRNA targets are in green (sh1-3).



The ICD can then be released into the cytoplasm by subsequent digestion with the γ -secretase complex, mediated by presenilin (Riedle et al., 2009). The free cytoplasmic ICD has been shown to localize to the nucleus, where it acts as a transcription factor to influence gene expression (Gast et al., 2008). Some products are shed into the culture media, while others are retained by the cell; in the context of this work, cleavage activity resulted in a variety of different sizes of L1CAM peptides that differed in epitope availability and resulted in a range of bands by western blot analysis (Fig. 5.2). In general though, the full-length protein was the most abundant form and was thus the band primarily quantified.

Figure 5.2: Diagram of known L1CAM cleavage products.

Abbreviations: fibronectin (FN), metalloprotease (MP), presenilin (PSEN). Epitopes and approximate binding sites shown for the three monoclonal L1CAM antibodies used in this work. Fragments are not to scale; red is retained by the cell, and blue is released (but may bind to other cells). Adapted from Konar et al. (2018).



The final piece of L1CAM biology that must be established is its relationship to signaling pathways. While much of the signaling data is derived from cancer cells, thus the specific relevance to healthy neurons is potentially suspect, L1CAM has been shown to bind integrins in its ECD which then induce ezrin and FAK binding to the ICD, leading to activation of, in succession: Src, PI3K, AKT, and NF κ B (Guo & Giancotti, 2004; Kiefel et al., 2010; Kiefel, et al., 2012b; Madrid et al., 2001). The same process has also been shown to activate YAP (Er et al., 2018; Yamamoto-Fukuda et al., 2020), and MAPK is activated by L1CAM binding to other growth factor receptors and by its endocytosis (Poplawski et al., 2012; Schaefer et al., 1999). Finally, L1CAM binds the cytoskeleton via a complex that includes MAP2 and ankyrins B and G (Gil et al., 2003). Formation of this complex appears to limit L1CAM mobility and likely helps to fix neurites in place (Nagaraj & Hortsch, 2006).

5.1.3 Identification of *L1CAM* as a gene of interest

5.1.3.1 Flow cytometry screening for autophagy regulators

An arrayed screen to identify potential modulators of autophagy was performed by Eleanna Stamatakou: HeLa cells stably expressing A53T- α -synuclein-eGFP (A53T-eGFP) and Cas9 were transduced with CRISPR gRNAs using a lentiviral library (Metzakopian et al., 2017). This screen aimed to find genes that, when mutated, impaired or enhanced the accumulation of eGFP-A53T. A53T α -synuclein overexpression is a common model of Parkinson's disease, and the protein has been shown to aggregate in cells (Narhi et al., 1999). Clearance of A53T α -synuclein is thought to depend in large part on macroautophagy (Webb et al., 2003), and thus perturbations to eGFP fluorescence intensity occur when macroautophagy is enhanced (decreased eGFP) or impaired (increased eGFP).

To perform this screen, an empty control vector and two non-targeting gRNAs (NTg1 and NTg2; Table 2.1) were transduced individually as negative controls. Two gRNAs targeting each experimental gene were used, and data from both were combined. Thus, greater confidence could be afforded that genes with corresponding effects for both gRNAs were not due to off-target effects. Positive control gRNAs targeting the core autophagy gene *ATG7* were also included (Komatsu et al., 2005; Tanida et al., 2001).

Clearance was assessed by flow cytometry, and the median intensity of eGFP fluorescence in BFP-expressing cells (indicating integration and expression of the gRNA) was normalized by dividing by the median eGFP intensity in the BFP-negative cells for each sample. This provided an internal control to account for, e.g., differences in cell density leading to different relative intensities between conditions. All values were then normalized by the average intensity of the empty vector for analysis.

Two-tailed unpaired t-tests were performed on each condition against the empty vector control. None of the negative controls were significantly changed from each other, although both of the NT gRNAs were slightly elevated compared to the empty vector. The positive control *ATG7* was significantly increased when compared to the empty vector, but not when compared to the NT gRNAs.

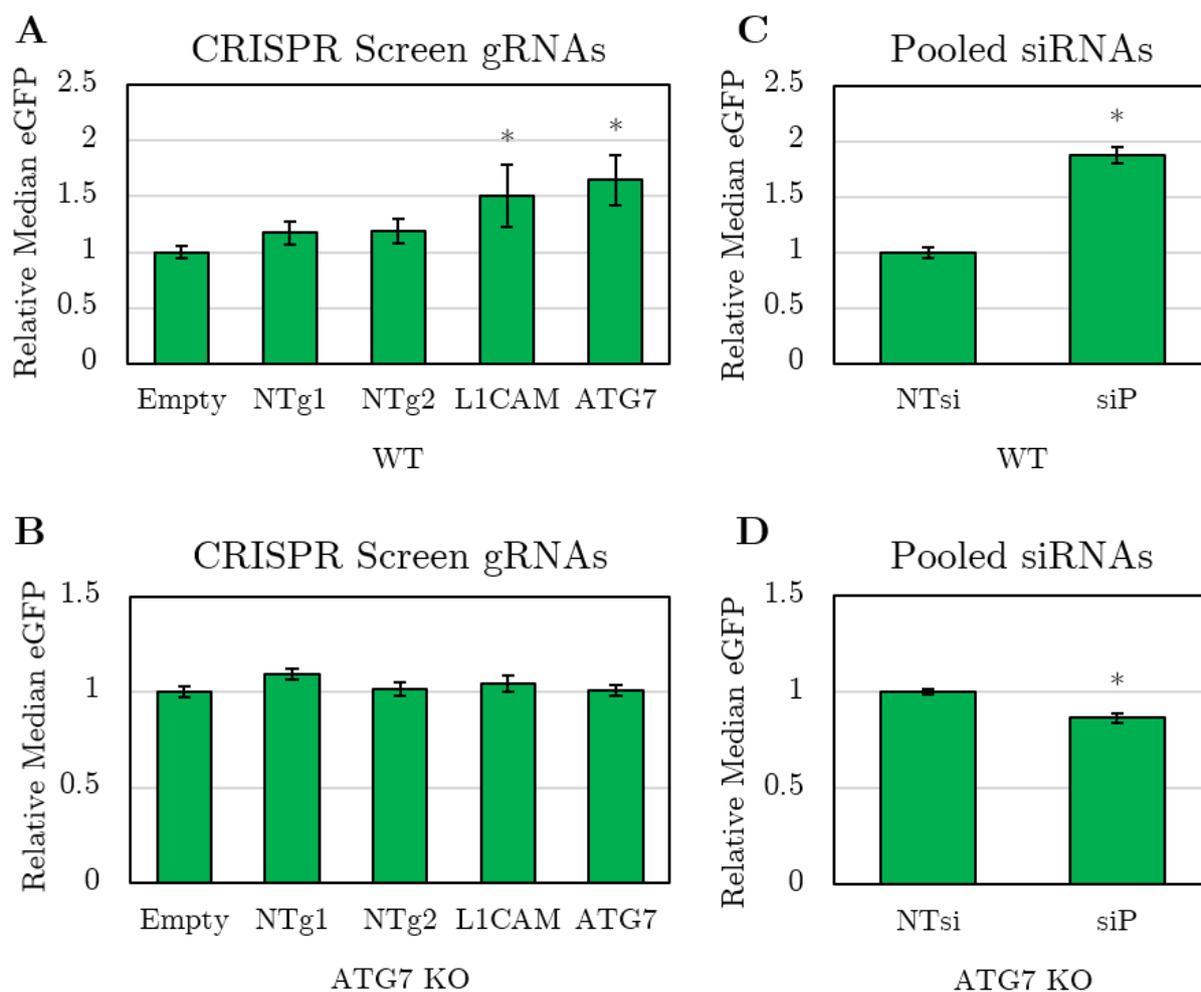
5.1.3.2 Clearance of A53T-eGFP is impaired by *L1CAM* mutagenesis

L1CAM-targeting gRNAs (Table 2.1) significantly increased eGFP fluorescence compared to the empty vector, similar in magnitude to the *ATG7* gRNAs (Fig. 5.3a). This suggests that mutagenesis of *L1CAM* impaired clearance of eGFP-A53T. Qualitatively, several samples also had poor cell density following transduction with *L1CAM* gRNAs, which may suggest that *L1CAM* mutagenesis impaired cell survival.

Figure 5.3: Identification of L1CAM as a potential autophagy modulator.

Data by Eleanna Stamatakou.

- (a) Cas9 eGFP-A53T HeLa cells were transduced with an empty vector (Empty), one of two nontargeting gRNAs (NTg1 & NTg2) or two gRNAs targeting an array of genes. *L1CAM* gRNAs caused a significant increase of median eGFP fluorescence intensity, indicating reduced clearance in a manner similar to the *ATG7*-targeting positive control. Neither NTg were significantly changed.
- (b) The experiment in (a) was performed in *ATG7* KO Cas9 eGFP-A53T HeLa cells, and no significant change was seen with any gRNAs.
- (c) Wildtype eGFP-A53T HeLa cells were transfected with either a nontargeting siRNA pool (NTsi) or *L1CAM*-targeting siRNA pool (siP). Median eGFP intensity was significantly increased in cells transfected with siP.
- (d) The experiment in (c) was performed in *ATG7* KO eGFP-A53T HeLa cells, and median eGFP was significantly decreased, albeit only slightly.



However, it must be noted that performing FDR correction via the Benjamini-Hochberg procedure to account for the 42 genes tested in this experiment resulted in *L1CAM* yielding a corrected $p=0.0504$; in addition, neither of the nontargeting gRNAs demonstrated significant differences from *L1CAM* when compared directly via additional two-tailed unpaired t-tests. Further validation was thus required in order to provide stronger evidence that a loss of L1CAM affected clearance of eGFP-A53T.

The experiment was repeated in ATG7 KO HeLa cells that express eGFP-A53T and Cas9. ATG7 KO cells are macroautophagy-deficient and must rely on other pathways for protein degradation, so if changes to mutant protein clearance do not extend to these cells, then effects likely depend on macroautophagy. In ATG7 KO Cas9 eGFP-A53T HeLa cells, no change was observed to eGFP fluorescence for either *L1CAM* or *ATG7* gRNAs, suggesting that the effect was related to macroautophagy and not alternative clearance pathways (Fig. 5.3b)

5.1.3.3 Assessment of the screen

Overall, this screen aimed to take a broad look across many genes to identify potential novel regulators of autophagy that could then be validated with further analyses; as such, it should not be faulted for relatively low statistical power. The eGFP fluorescence intensity of each cell was widely variable in the population, with cells typically distributed over a range of three \log_{10} values in basal conditions, possibly due to variability in the rates of transgene expression and degradation between cells. The median eGFP intensity measure used here thus has statistical inertia, requiring a high proportion of cells to exhibit a high magnitude of effect on eGFP for a gene to be identified. The screen is thereby insensitive to small fluctuations in levels of eGFP-A53T and to changes in only a small fraction of cells. Using this method, the false positive rate of the assay may therefore be reduced at the expense of also increasing the false negative rate.

However, this assay features several caveats that may impair reproducibility, highlighting the need for follow-up validation. Foremost, it did not control for rates of CRISPR cutting, nor did it have a readout to ensure that transcription or translation of the target gene was reduced in practice. HeLa cells have been shown to have high rates of CRISPR-mediated indel formation, but the genetic instability of HeLa cells also makes it likely that cells in the population have a variable number of target alleles and may have acquired mutations that affect the gRNA target sites. While it would reduce throughput considerably, the sample could be split in half when preparing cells for endpoint flow cytometry to assess the rate of mutagenesis (e.g., using Surveyor or TIDE). A gRNA directly targeting GFP could also have been used to determine the percent of cells that lose GFP expression—thereby determining the efficiency of gene knockout by this method.

The screen also used non-clonal cells, which are better able to compare between conditions than clonal lines but suffer from heterogeneity of lentiviral integration loci and effects of the gRNAs. This approach was justifiable as two gRNAs were used for each target gene and both negative and positive controls were included for calibration, but a chance still exists that off-target effects could be responsible for differences in clearance of the eGFP-A53T. To overcome this, it would have been preferable to use a non-integrating vector and to use additional gRNAs for each gene.

Observed effects may additionally be due to alterations to the rate of eGFP-A53T production; a gRNA that targets a transcriptional or translational repressor of the transgene would also show an increase to eGFP without affecting its degradation. While it may be possible to include conditions in which protein production or degradation are blocked for a time to account for this, the intention of the screen was to rapidly identify candidates and inform follow-up experimentation rather than provide a definitive answer at once, so any of the possible modifications mentioned here must be balanced with the overall throughput of the screen.

5.1.3.4 L1CAM knockdown with pooled siRNAs reduces clearance of mutant α -synuclein in HeLa cells

To assess whether the observed effect on mutant protein clearance replicated with an orthogonal method of protein reduction, Eleanna Stamatakou transfected HeLa cells expressing eGFP-A53T (without Cas9) with a pool of four non-targeting siRNAs (NTsi) or with a pool of four siRNAs targeting *L1CAM* (siP). The *L1CAM* siP was made up of four distinct targeting sequences, named L1-si1 through L1-si4 (Table 2.2). These map to *L1CAM* exons 17, 21, 20-21 (cross-junction), and 11 respectively, all of which are included in all known *L1CAM* transcripts (Fig. 5.1).

Cells transfected with the *L1CAM* siP showed a significant increase in eGFP fluorescence as compared to the NTsi controls (Fig. 5.3c). Following the same siRNA treatments in ATG7 KO eGFP-A53T HeLa cells (also without Cas9), the siP-treated cells displayed slightly, but significantly, less eGFP than the NTsi control (Fig. 5.3d). This did not imply that *L1CAM* knockdown was necessarily enhancing autophagy in ATG7 KO HeLa cells; the raw intensity of eGFP was substantially changed from a wide distribution in HeLa cells with ATG7 to a narrow peak at high intensity in ATG7 KO cells. It was possible that cells with higher intensities were less fit and did not survive, or *L1CAM* knockdown may have activated compensatory degradation pathways such as the ubiquitin-proteasome system. In any case, the observed reduction was relatively small and thus did not represent a major alteration to clearance.

The correspondence of the effects of CRISPR mutagenesis and transfected siRNA knockdown of *L1CAM*—namely, impaired clearance of eGFP-A53T in wildtype cells but not in ATG7 knockout cells—provides greater confidence that L1CAM is involved in regulating macroautophagy.

5.1.4 Aims & Hypotheses

To validate the observation that L1CAM may be involved in the clearance of mutant proteins via macroautophagy, I first aimed to assess the level of macroautophagy in L1CAM knockdown conditions. Because L1CAM influences several signaling pathways that are known to impact macroautophagy initiation, I hypothesized that reducing L1CAM protein impaired macroautophagy induction through reduced activation of these pathways upstream of autophagosome biogenesis. I therefore posited that L1CAM knockdown would reduce the relative abundance of LC3B-I and LC3B-II in both basal conditions and after blocking lysosomal degradation with bafilomycin A1 (baf) treatment (Klionsky et al., 2021).

Next, I aimed to demonstrate that the effects of a loss of L1CAM extended to iPSC-derived neurons due to the importance of L1CAM to neural development. I therefore sought to knockdown or knockout (KO) L1CAM in iPSCs using the established G3 cell line, which includes a dox-inducible *NGN2* insert at the *AAVS1* locus enabling i³Neuron differentiation. I hypothesized that both of these approaches would recapitulate macroautophagy impairment, and that overexpressing L1CAM from a plasmid transfection driven by a constitutive promoter in L1CAM knockdown or KO cells would restore or increase macroautophagy levels compared with wildtype levels.

5.2 LC3B is reduced by individual siRNAs but not pooled siRNAs

The experiments in this section were performed in collaboration with So Yeong Cheon. In particular, we each independently performed a subset of experimental replicates, and these data were combined to comprise the full results described here.

5.2.1 Pooled siRNAs reduce L1CAM but not LC3B-II in HeLa cells

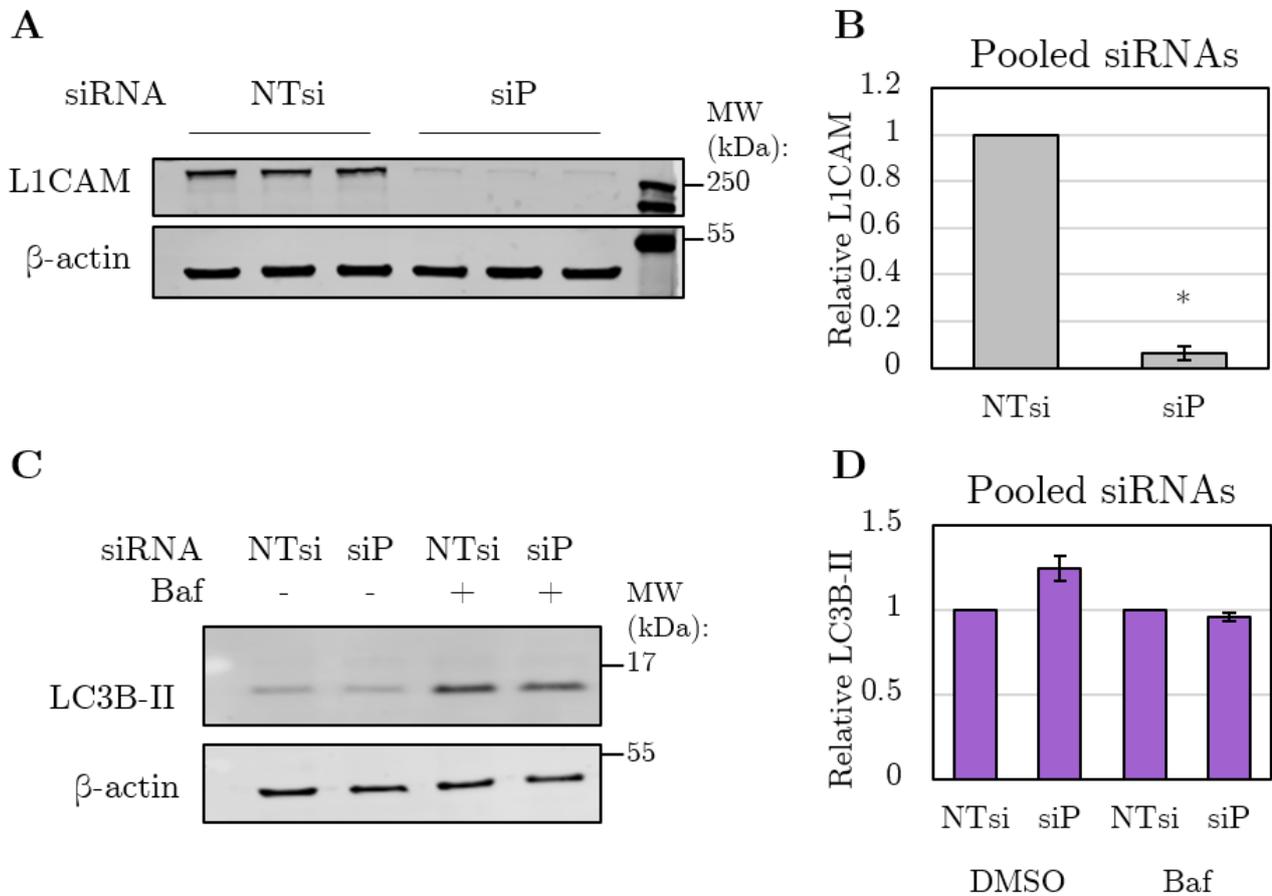
Wildtype HeLa cells were transfected with either the NTsi pool or the *L1CAM*-targeting siP. Protein lysate was collected and analyzed by western blotting to assess the strength of L1CAM knockdown. L1CAM was significantly reduced for siP-transfected cells; on average, less than 7% of the protein remained as compared to the NTsi condition (Fig. 5.4a-b).

Next, levels of LC3B-II were assessed as a proxy for the amount of autophagosomes present in the cells. Because a decrease in LC3B-II could be caused by either a decrease in autophagosome synthesis or an increase in autophagosome degradation, cells were treated with DMSO as a negative control or with 400 nM Bafilomycin (Baf) for 4 h prior to lysis to inhibit lysosomal acidification and thereby disrupt autophagosome degradation (Streeter et al., 2016). In this way, the total amount of macroautophagy induction during the time of treatment can be measured. In both conditions, however, no significant differences were evident for LC3B-II (Fig. 5.4c-d).

Figure 5.4: Pooled siRNAs reduce L1CAM but not LC3B-II.

Blot images by So Yeong Cheon.

- (a) Transfection of WT HeLas with *L1CAM* pooled siRNAs (siP) effectively reduced L1CAM as compared to a pool of nontargeting siRNAs (NTsi). The 5G3 antibody was used.
- (b) Quantification of n=4 experiments as in (a). L1CAM was significantly reduced in siP-transfected cells.
- (c) Transfection of siP did not alter LC3B-II levels after 4 h treatment with DMSO (-) or 400 nM baf (+).
- (d) Quantification of n=4 experiments as in (c), each with three separate wells as technical replicates. Baf-treated samples were normalized to the NTsi+baf condition; no differences are significant.



It was surprising that siP effectively reduced L1CAM and mediated a macroautophagy-dependent effect to the clearance of eGFP-A53T mutant α -synuclein, and yet it did not significantly affect LC3B-II. Pooled siRNAs benefit from a multiplicity of mRNA targeting sites, but they suffer from an effective loss of concentration of the more potent siRNAs. Reducing the relative concentration of each individual siRNA by pooling may reduce the associated off-target effects, but it also combines these effects, while variability between individual siRNAs is more evident (Petri & Meister, 2013). Additionally, pooled siRNAs have a high false negative rate and a comparatively low false positive rate in genome-wide screens, indicating that effect sizes may be muted with siRNA pools (Hao et al., 2013). We therefore decided to test the four constitutive siRNAs that made up the siP individually in order to gain clarity.

5.2.2 Individual siRNAs reduce L1CAM and LC3B in HeLa cells

Each siRNA was transfected at the same concentration that siP was previously transfected as a whole. After culture, DMSO or baf treatment, and lysis, total protein was blotted for L1CAM and LC3B. All four siRNAs were found to significantly decrease the amount of L1CAM protein detected; however, si4 was found to provide a weaker knockdown and more variable results and was thus dropped from further experiments (Fig. 5.5a-b).

Next, levels of LC3B were measured in the same samples. LC3B-I was quantified in order to assess possible effects upstream of lipidation, and it was significantly decreased in cells transfected with either si1 or si2 for both DMSO- and Bafilomycin-treated conditions (Fig. 5.5c-e). Meanwhile, LC3B-II was decreased only after transfection with si2 for both treatment conditions (Fig. 5.5c-e).

A decrease of LC3B-I in the absence of an increase to LC3B-II as seen for si1 and si2 was suggestive of a downregulation of *LC3B* transcription or translation (Klionsky et al., 2021). Likewise, a decrease to LC3B-II in both the DMSO- and bafilomycin-treated conditions as seen in si2 was suggestive of a downregulation of LC3B-II conjugation or autophagosome formation, which was also consistent with a transcriptional decrease (Streeter et al., 2016). However, the latter effect was only evident in si2 despite all three demonstrating strong knockdown of L1CAM protein. Because there was no overlap in the potential off-target effects of the target sequences (Table 2.2), it was unlikely that the observed LC3B effects in si1 and si2 were both due to promiscuity. On the other hand, since the binding sites of si2 and si3 were separated by just 26bp, it was surprising that they showed such varied effects. It was possible that the lack of an effect on LC3B in si3 was related to its binding site overlapping a junction between two exons, and this is explored further in the discussion.

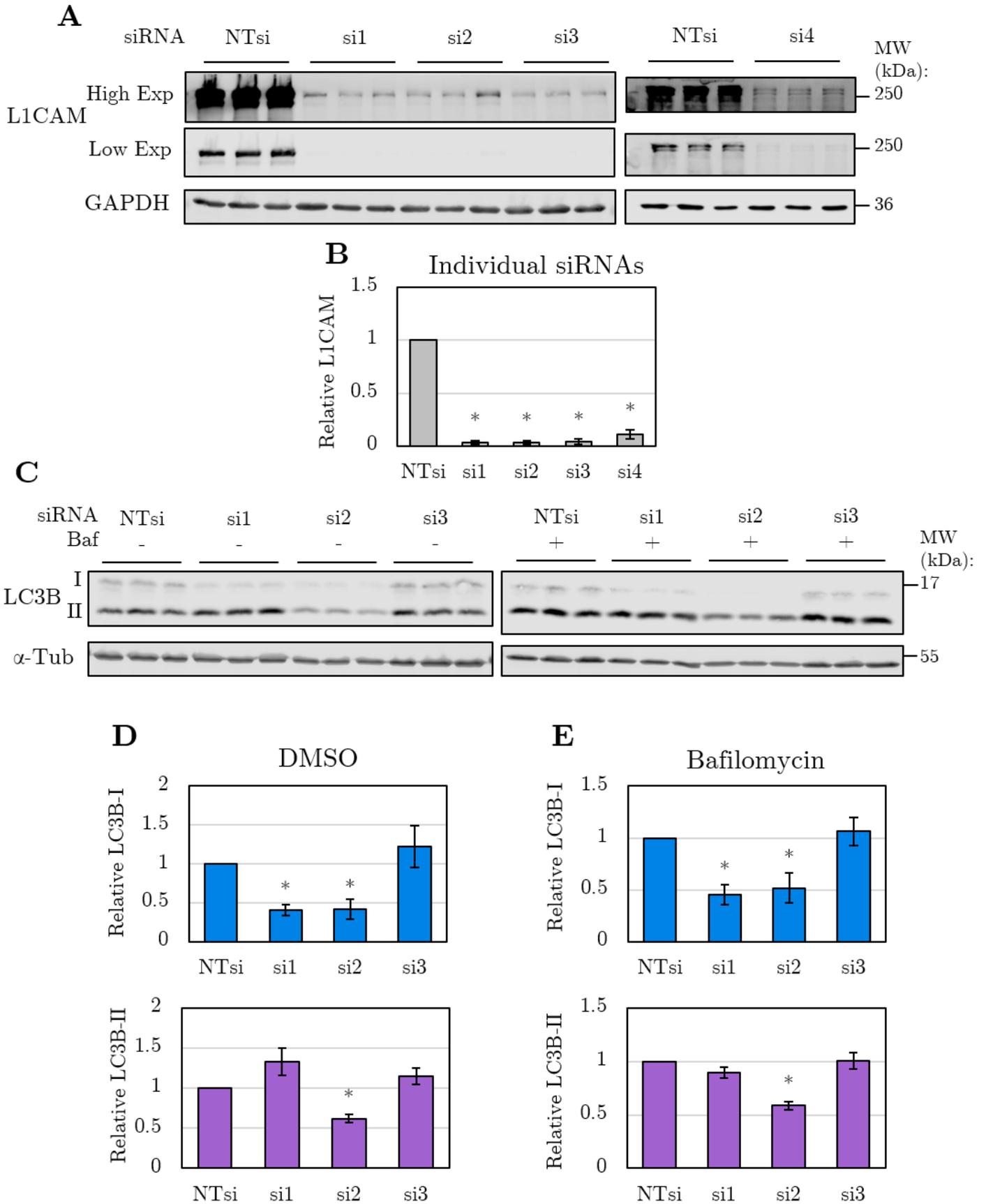
Figure 5.5: Deconvolution of L1CAM siRNAs.

The individual siRNAs that composed the siP were each transfected into wildtype HeLa cells, and L1CAM, LC3B-I, and LC3B-II protein levels were assessed.

- (a) All siRNAs dramatically decreased L1CAM protein, but si4 performed the worst and was thus dropped from further analysis. The D-5 antibody was used.
- (b) Quantification of n=5 experiments as in (a), each with three separate wells as technical replicates.
- (c) Transfection of si1 and si2 reduced LC3B-I, and si2 reduced LC3B-II, both in the absence and presence of bafilomycin. Transfection with si3 did not affect either form of LC3B.
- (d-e) Quantification of n=5 experiments as in (c), each with three separate wells as technical replicates. Baf-treated samples were normalized to the NTsi+baf condition.

(continued)

Figure 5.5 (*continued*): Deconvolution of L1CAM siRNAs.



5.3 Individual siRNAs impair autophagic clearance

5.3.1 Mutant protein degradation is reduced in HeLa cells

Reduction of LC3B by individual siRNAs validated that macroautophagy was likely affected by L1CAM knockdown. Therefore, additional methods of assessing the practical clearance of mutant aggregating proteins were sought. As si1 and si2 were both shown to be highly effective at L1CAM reduction as well as at reducing LC3B, these two were chosen for further functional analysis.

First, the NTsi pool or individual siRNAs were transfected into WT and ATG7 KO lines of eGFP-A53T HeLa cells (without Cas9). Flow cytometry revealed that median eGFP intensity was significantly increased by both si1 and si2 in HeLa cells with ATG7 intact, with more than a sevenfold increase in si2-transfected cells (Fig. 5.6a). However, eGFP intensity was significantly reduced in ATG7 KO HeLa cells after si2 transfection, albeit only slightly (Fig. 5.6b). These findings were consistent with the effects of the *L1CAM*-targeting siRNA pool shown previously; as discussed, the decrease in ATG7 KO HeLa cells was likely due to a ceiling effect or activation of compensatory degradation pathways. Interestingly, si1 showed a nonsignificant increase to eGFP in ATG7 KO cells, possibly suggesting that si1 may impair clearance of eGFP-A53T via additional mechanisms than canonical macroautophagy, or it may affect the rate of transgene transcription or translation to effectively increase the ceiling of expression.

Next, a plasmid encoding eGFP-A53T was transfected alongside a plasmid encoding only eGFP. The second plasmid served as a loading control to normalize for differences in the efficiency of transfection and plasmid expression. While autophagy constitutes an important degradation route for eGFP-A53T, eGFP alone is readily degraded by the proteasome and thus was not subject to obvious buildup if autophagy was impaired. Furthermore, the use of eGFP as a loading control enabled the same GFP antibody to be used for both proteins together on the same blot, minimizing the effects

of different antibody binding affinities. To assess the dependency of any changes to clearance on the presence of core macroautophagy machinery, a HeLa clone that had been CRISPR-edited to knock out ATG16L1 was also used.

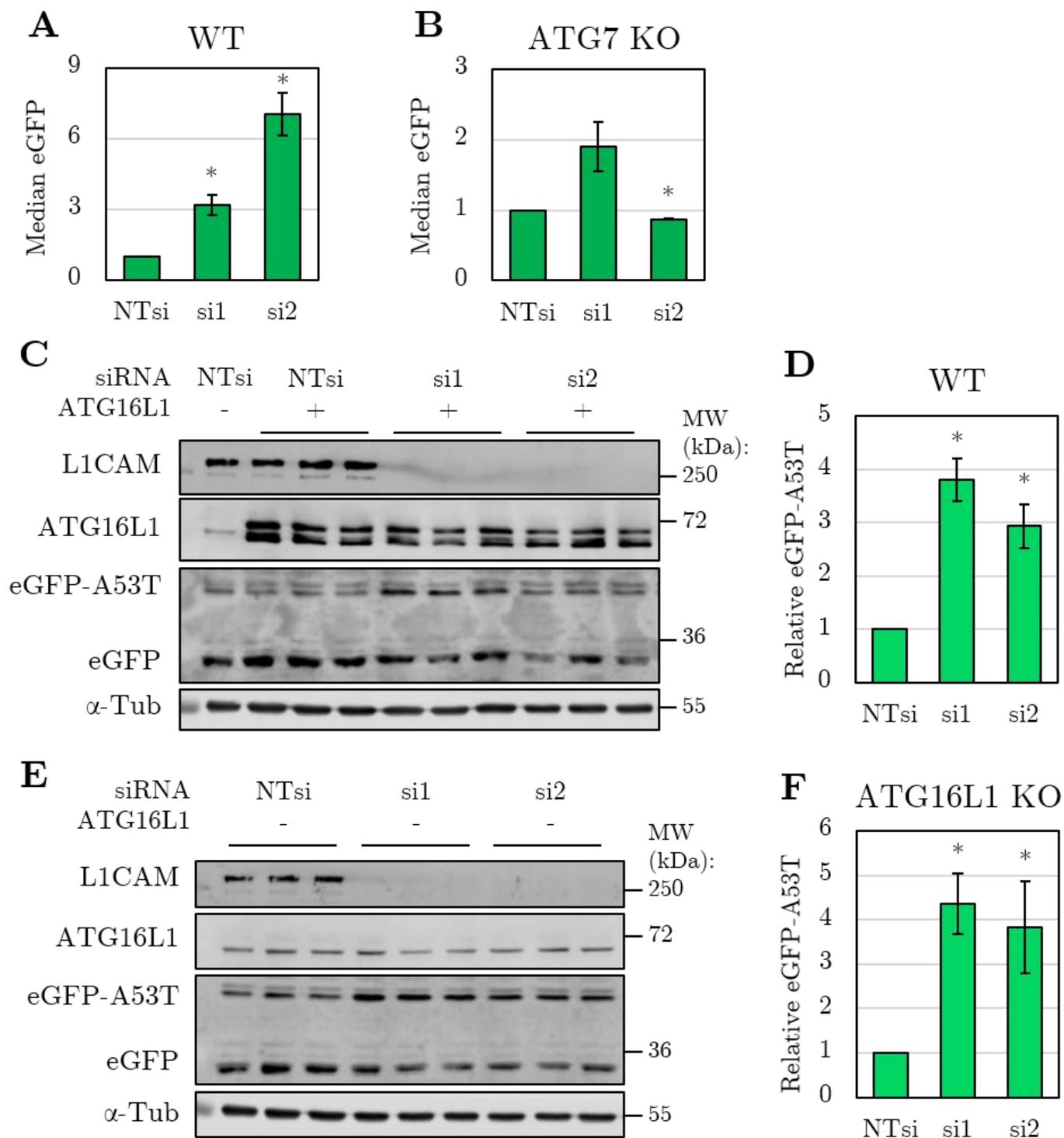
ATG16L1 expresses two distinct isoforms in HeLa cells: α at a full length of 607 amino acids, and β composed of 588 amino acids as a result of exon skipping (Jiang et al., 2013). These have been shown to have similar structures and functions, and both were observed in the wildtype HeLa cells. The ATG16L1 KO line lacked both, although a nonspecific band was present between the two isoform bands (Fig. 5.6c). These cells were thus used with confidence that ATG16L1 was fully ablated. Additionally, L1CAM protein was strongly reduced by both si1 and si2, consistent with previous experiments.

Figure 5.6: Individual siRNAs impair eGFP-A53T clearance.

- (a) eGFP-A53T HeLa cells were transfected with NTsi, si1, or si2. Both si1 and si2 significantly increased median eGFP fluorescence intensity, as measured by flow cytometry (n=3 experiments, each with 3 technical replicates).
- (b) The experiment in (a) was performed in ATG7 KO eGFP-A53T HeLa cells, and si2 was found to significantly decrease eGFP while si1 non-significantly increased eGFP (n=3 experiments, each with 3 technical replicates).
- (c) Wildtype (+) or ATG16L1 KO (-) HeLas were transiently co-transfected with plasmids expressing eGFP-A53T and eGFP as well as the indicated siRNA. A nonspecific band was evident between the α and β isoforms of ATG16L1, but both isoforms were ablated in KO cells.
- (d) Quantification of n=3 experiments as in (c), each with three separate wells as technical replicates. Measured band intensity of eGFP-A53T was normalized by dividing by eGFP band intensity as a transfection loading control. Both si1 and si2 significantly increased eGFP-A53T compared to NTsi.
- (e) The experiment in (c) was performed in ATG16L1 KO HeLa cells.
- (f) Quantification of n=3 experiments as in (e), each with three separate wells as technical replicates, by the same method as in (d). Both si1 and si2 significantly increased eGFP-A53T compared to NTsi.

(continued)

Figure 5.6 (*continued*): Individual siRNAs impair eGFP-A53T clearance.



The density of the eGFP-A53T band was then measured and normalized with the eGFP band as indicated in Fig. 5.6c. Both si1- and si2-treated cells demonstrated a significant increase to eGFP-A53T protein levels (Fig. 5.6d). In ATG16L1 KO HeLa cells, similar trends were observed, and both si1 and si2 transfection again resulted in significant increases to eGFP-A53T (Fig. 5.6e-f).

In addition, the same assay was performed using a plasmid encoding eGFP fused to the first exon of HTT with a 74Q repeat (eGFP-HTT74) alongside an eGFP plasmid, used again as a transfection loading control. Cells treated with si1 and si2 similarly demonstrated a significant increase to eGFP-HTT74 compared to the NTsi control, again for both WT and ATG16L1 KO HeLa cells (Fig. 5.7).

Together, these experiments indicated that L1CAM knockdown correlated with increased levels of two mutant proteins that are associated with neurodegenerative diseases and that are substrates of autophagy. While this effect was probably due to impaired degradation, the extension of this increase to ATG16L1 KO cells was unexpected. Clearance of eGFP-A53T was not further impaired in ATG7 KO HeLa cells, suggesting that L1CAM knockdown either requires the presence of ATG7 to mediate its effect or that it affects the same pathways as ATG7 KO. Meanwhile, these data suggest that L1CAM knockdown impaired clearance beyond the effect of a loss of ATG16L1. The specific roles of ATG7 and ATG16L1 may thus provide insight into the mechanism by which L1CAM knockdown affected mutant protein clearance, as explored in the discussion.

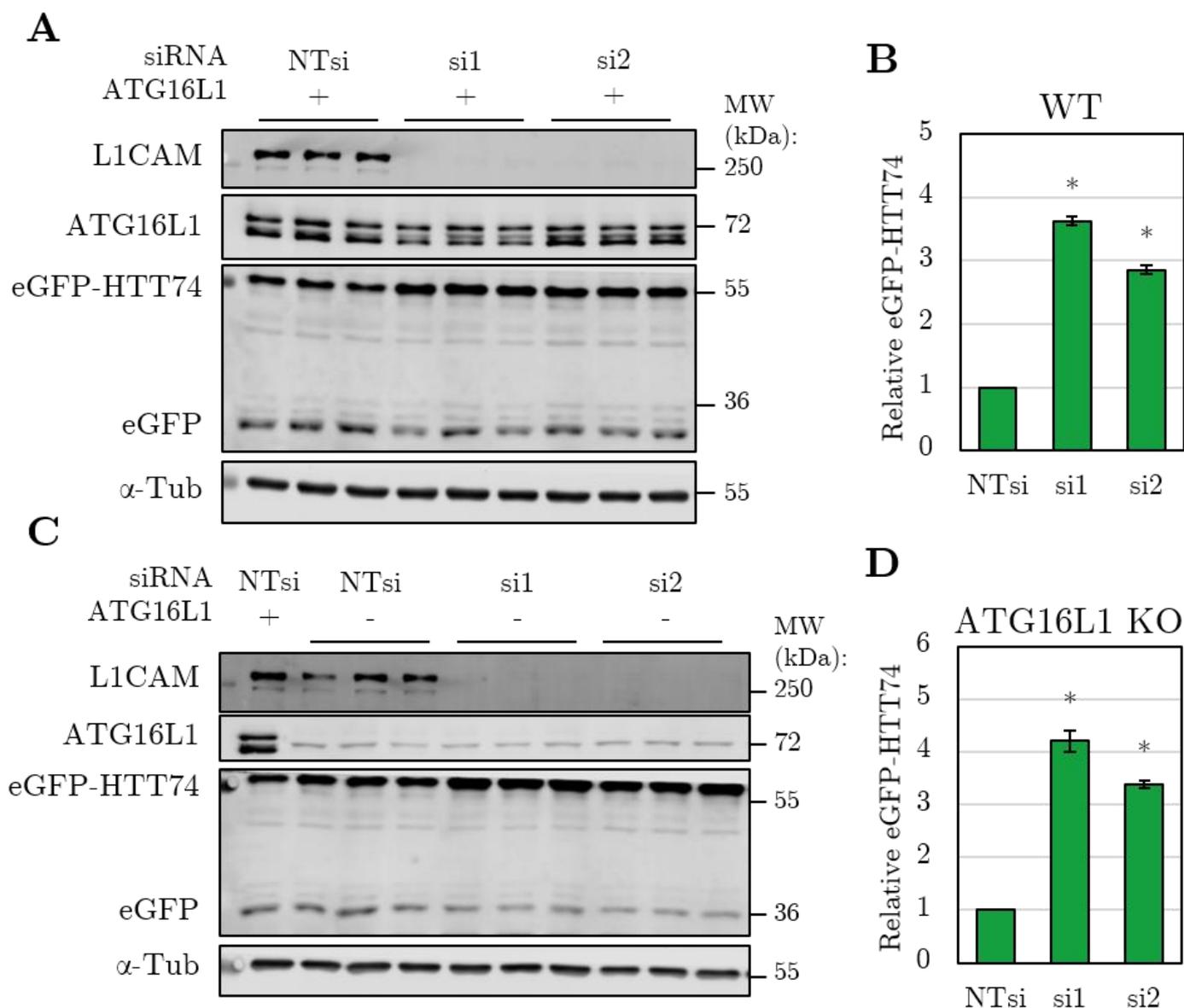
Figure 5.7: Individual siRNAs impair eGFP-HTT74 clearance.

(a) Wildtype (+) or ATG16L1 KO (-) HeLa cells were transiently transfected with plasmids expressing eGFP-HTT74 and eGFP as well as the indicated siRNA.

(b) Quantification of n=3 experiments as in (a), each with three separate wells as technical replicates. Measured band intensity of eGFP-HTT74 was normalized by dividing by eGFP band intensity as a transfection loading control. Both si1 and si2 significantly increased eGFP-HTT74 compared to NTsi.

(c) The experiment in (a) was performed in ATG16L1 KO HeLa cells.

(d) Quantification of n=3 experiments as in (c), each with three separate wells as technical replicates, by the same method as in (b). Both si1 and si2 significantly increased eGFP-HTT74 compared to NTsi.



5.3.2 The proteasome is not impaired by L1CAM reduction

It was also desired to ascertain if proteasomal degradation was affected by L1CAM knockdown. A reduction to autophagy has been shown to upregulate the ubiquitin-proteasome system as a compensatory mechanism (Kocaturk & Gozuacik, 2018), so if proteasomal activity was upregulated or unaffected after siRNA knockdown, such a finding could increase confidence that the observed impairment of mutant protein degradation was specifically due to autophagic dysfunction.

HeLa cells were used that constitutively expressed GFP fused to the CL1 degron (GFP-dg; Dantuma et al., 2000). The degron tag causes GFP-dg to be rapidly polyubiquitinated and degraded by the proteasome, so very little GFP was present in cells in basal conditions. When proteasomal clearance was impaired, such as after treatment with the proteasomal inhibitor MG-132, GFP was able to accumulate and thus fluorescence intensity increased (Greussing et al., 2012). Thus, this assay could detect changes to proteasomal activity; if L1CAM knockdown impaired the proteasome, then GFP fluorescence was expected to increase.

GFP-dg HeLa cells were transfected with NTsi, si1, or si2. After three days in culture, cells were treated with either DMSO or 10 μ M MG-132 for 6 h before being fixed and imaged or dissociated for flow cytometry (Fig. 5.8a-b).

Figure 5.8: Individual siRNAs do not impair proteasomal clearance.

GFP-dg HeLa cells were transfected with the indicated siRNA, and three days later cells were treated for 6 h with either DMSO or 10 μ M MG-132.

- (a) Representative fluorescent microscopy images of GFP in each condition. DMSO-treated cells lacked visually-detectible GFP fluorescence, and treatment with MG-132 increased GFP intensity to varying degrees.

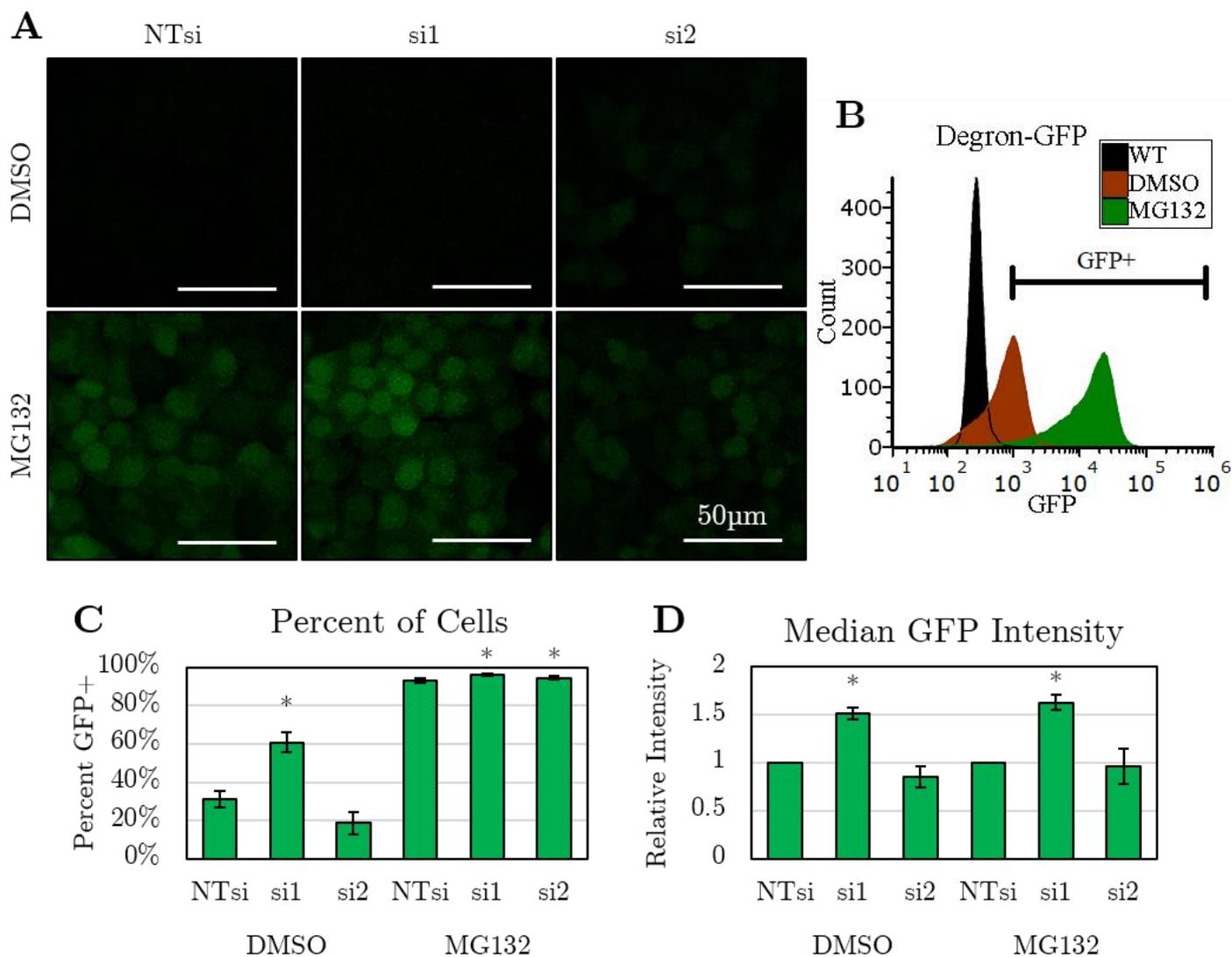
(continued)

Figure 5.8 (*continued*): Individual siRNAs do not impair proteasomal clearance.

(b) WT HeLa cells were used to set a threshold of detectable GFP (GFP+), and the percent of GFP+ single cells and the median GFP intensity for all single cells were measured for each sample by flow cytometry. Technical triplicates of each condition were averaged for each of n=3 independent experiments. Median GFP was normalized to NTsi of each treatment condition. Significance was assessed by two-tailed paired t-tests.

(c) The percent of GFP+ cells was significantly increased for si1 and non-significantly decreased for si2 after DMSO treatment; it was significantly increased for both si1 and si2 after MG-132 treatment, but all MG-132 treated conditions were >90% GFP+.

(d) In both treatment conditions, median GFP intensity was significantly increased for si1 but not significantly changed for si2.



Two measures of GFP expression were assessed: first, WT HeLa cells were used to set a threshold for detectable GFP expression, and the percentage of single cells that expressed GFP above this level was recorded for each sample. Using this metric, si1 demonstrated a significant increase in DMSO treated conditions, and both si1 and si2 were significantly increased after MG-132 treatment (Fig. 5.8c). However, all MG-132-treated samples were between 91-98% GFP+, so the magnitude of this change was small. Second, the median value for GFP fluorescence was measured for all single cells, and values were normalized to the associated NTsi condition. For both DMSO and MG-132, si1 was significantly increased, but si2 was unchanged from NTsi (Fig. 5.8d).

If the proteasome was inhibited, both measures were expected to increase in DMSO conditions only. MG-132 is a potent proteasome inhibitor, so treatment should block GFP-dg degradation by the proteasome and thus should equalize GFP fluorescence to the relative total expression of GFP-dg. The observed increase to the percent of GFP+ cells and median GFP fluorescence for sh1 in both DMSO and MG-132 treatment conditions may have been due to an increase of GFP-dg synthesis and cannot be fully attributed to inhibition of the proteasome. Furthermore, neither measure was significantly changed for si2 in DMSO conditions despite also demonstrating strong impairment of mutant protein clearance and LC3B protein levels previously. These data suggested that the impairment of clearance to mutant proteins noted above was likely due to inhibition of autophagy rather than the proteasome.

5.4 Modelling loss of L1CAM in iPSC-derived neurons

5.4.1 Cell line derivation

Because variability between the effects of individual siRNAs in HeLa cells was noted, and because affected pathways in a cancer cell line may be markedly different than those affected in neurons, the effects of L1CAM loss in iPSC-derived neurons were assessed. *L1CAM* was not expressed in iPSCs, but it was highly expressed in differentiated i³Neurons (Fig. 5.9a), so the G3 iPSC line was used for subsequent experiments because it contains an integrated doxycycline-inducible *NGN2* construct facilitating cortical neuron differentiation. All genetic manipulations were performed and validated in iPSCs, and analysis of resultant expression was done after *NGN2* induction with dox for three days followed by eleven days in culture without dox for stabilization and maturation, totaling two weeks.

Three parallel strategies for reducing or removing L1CAM were pursued: transcriptional silencing via CRISPRi, mRNA degradation via shRNA expression, and genetic ablation via CRISPR cutting. Because each involved the expression of an RNA (either a gRNA or shRNA), the same lentiviral backbone was used for each featuring a U6 promoter driving RNA expression. The backbone also contained the selection markers PuroR and NLS-BFP driven by a constitutive EF-1 α promoter (Fig. 5.9b). BFP was effectively localized to the nucleus for easy identification of expression (Fig. 5.9c). Details of vector design and construction are expanded in Chapter 2.

CRISPR KO of *L1CAM* was performed with only transient transfection of the gRNA and Cas9 plasmids, and details of the cloning procedure are provided below. However, both CRISPRi and shRNA knockdown feature genetic integration via lentiviral preparation in HEK 293T cells followed by transduction of iPSCs to create stable lines. Following transduction and recovery of the iPSCs, each line was FACS purified using the same sorting parameters in order to standardize expression across all lines. Single

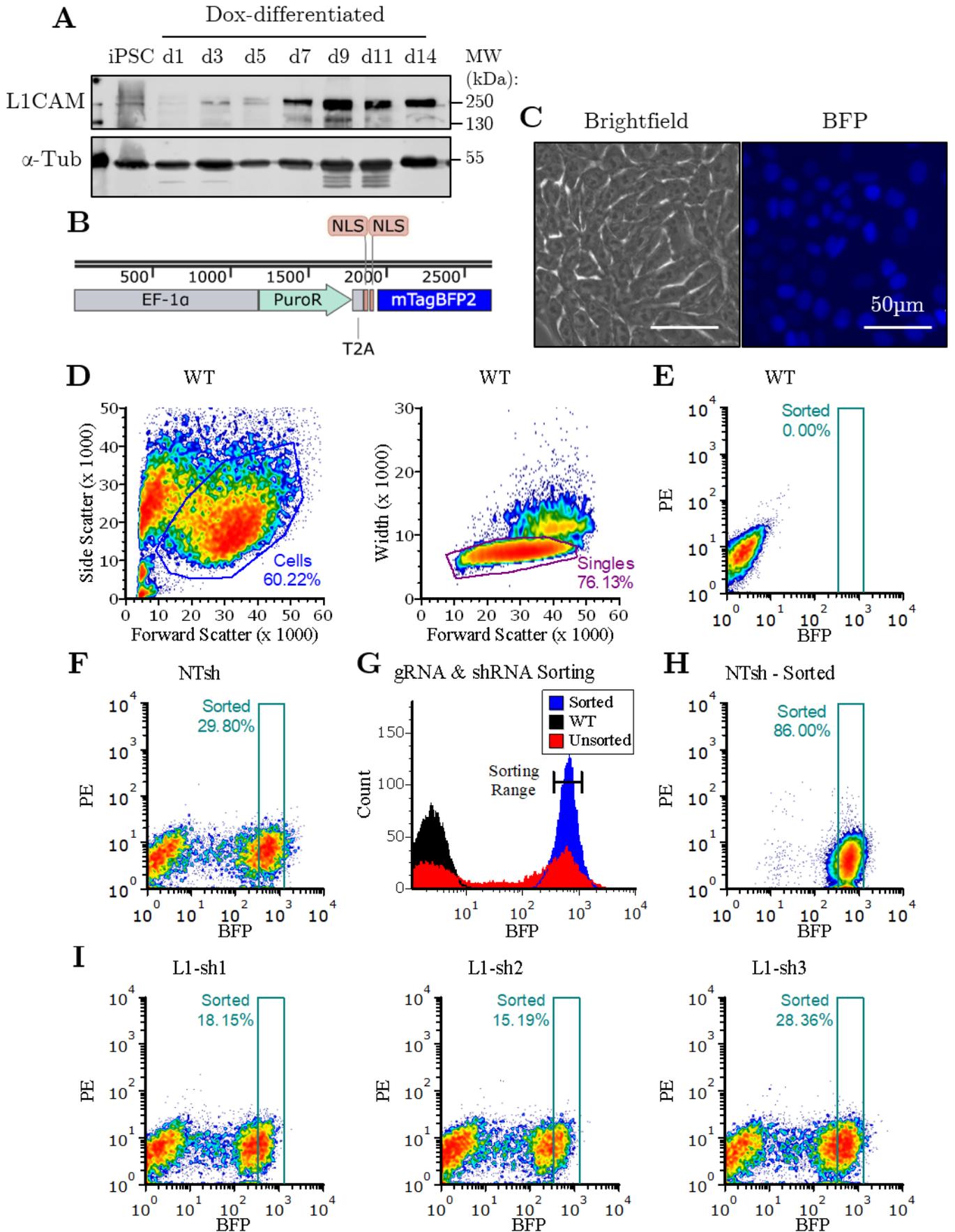
cells were identified with forward scatter, side scatter, and trigger pulse width (Fig. 5.9d), and BFP expression in these cells was assessed using a 405 nm laser with 460/50 nm bandpass emission filter (Fig. 5.9e). Sorting parameters were calibrated by first setting a threshold of BFP expression using a non-transduced WT negative control, and only a narrow band of BFP+ cells was sorted approximately $\pm 0.25 \times \log_{10}(\text{mode of BFP+ population})$ calculated using the transduced control (non-targeting) line (Fig. 5.9f). This resulted in a narrow distribution of BFP expression that remained stable over at least one month in culture (Fig. 5.9g-h). The same sorting parameters were also applied for each experimental line (Fig. 5.9i).

Figure 5.9: Stable CRISPRi and shRNA line derivation.

- (a) L1CAM was not expressed in iPSCs, but it was turned on during differentiation.
- (b) The RNA expression backbone featured a selection cassette consisting of the EF-1 α promoter driving expression of PuroR-T2A-2 \times NLS-BFP.
- (c) 2 \times NLS-BFP was readily visible in cell nuclei after transfection or transduction.
- (d) Non-clonal stable lines were generated by FACS. Single cells were identified by forward scatter, side scatter, and trigger pulse width (width).
- (e) A non-transduced negative control (WT) was used to identify the background level of BFP fluorescence. Phycoerythrin (PE) on the y-axis was used as a negative control fluorophore.
- (f) Unsorted nontargeting shRNA (NTsh) cells were used to set sorting parameters in a narrow band around the mode of the BFP+ population. CRISPRi lines used the same process but sorting parameters were based on the NTig line.
- (g) Unsorted NTsh cells (red) showed a wide distribution of BFP expression, including a population that resembled non-transduced WT cells (black). After sorting and culture for one month (blue), cells remained tightly clustered within the sorting parameters.
- (h) Sorted and cultured NTsh cells as in (f) demonstrated a tight distribution of BFP expression that was maintained over approximately one month in culture.
- (i) The same sorting parameters were applied to each of the L1CAM-targeting shRNA lines.

(continued)

Figure 5.9 (*continued*): Stable CRISPRi and shRNA line derivation.



By sorting for this narrow band, cell lines were effectively normalized for potential differences in viral titer, integration copy number, and expression intensity. In other words, similar amounts of BFP between lines indicated that similar amounts of each gRNA or shRNA were likely being produced as well. These lines intentionally remained non-clonal because variability in integration sites would render any clonal lines unable to be properly compared to a control NTsh clone with different integration sites. Instead, non-clonal populations relied on cell numbers to smooth over potential extraneous effects of any individual cell. Furthermore, virally-inserted transgenes have a propensity to be silenced over time, so clonal cells would have little utility.

Finally, gDNA was collected from these lines, PCR amplified using primers within the insert (Table 2.4), and Sanger sequenced to confirm that the correct RNA sequence was integrated in each case. Once these lines were purified, they were expanded and frozen to provide a common source for subsequent experimentation, and cells were thawed and cultured as iPSCs for one to two weeks prior to differentiation. The process of lentiviral construction, transduction, and sorting was not repeated; rather, each experimental replicate shown below represents an independent differentiation performed together for each cell line compared, and at least three different wells from each line and treatment condition were included as technical replicates.

5.4.2 CRISPRi-mediated transcriptional silencing

CRISPRi relies on constitutive expression of a nuclease-dead Cas9 (dCas9) that is fused to the KRAB repressor domain. By providing the dCas9 with a gRNA that targets the transcription start site of a gene, transcription initiation may be prevented and thus the gene may be silenced (McTague et al., 2021). *L1CAM* is known to have two distinct promoter regions: upstream of exon 1, which consists entirely of the 5' untranslated region (UTR), and upstream of exon 2, which contains additional 5'UTR and the beginning of the protein coding sequence (Pfeifer et al., 2010). Because these

are separated in the genome by approximately 10kb, it was unlikely that a single gRNA would be capable of silencing both transcripts, and it was unclear which promoter was preferentially used in i³Neurons. Two gRNAs were chosen that target the first site (ig1 and ig2), and one was chosen that was within exon 2 (ig3) to account for variable transcription initiation at the downstream promoter (Kohl et al., 1992). Details of gRNA design are provided in Chapter 2, and target locations are displayed in Fig. 5.1.

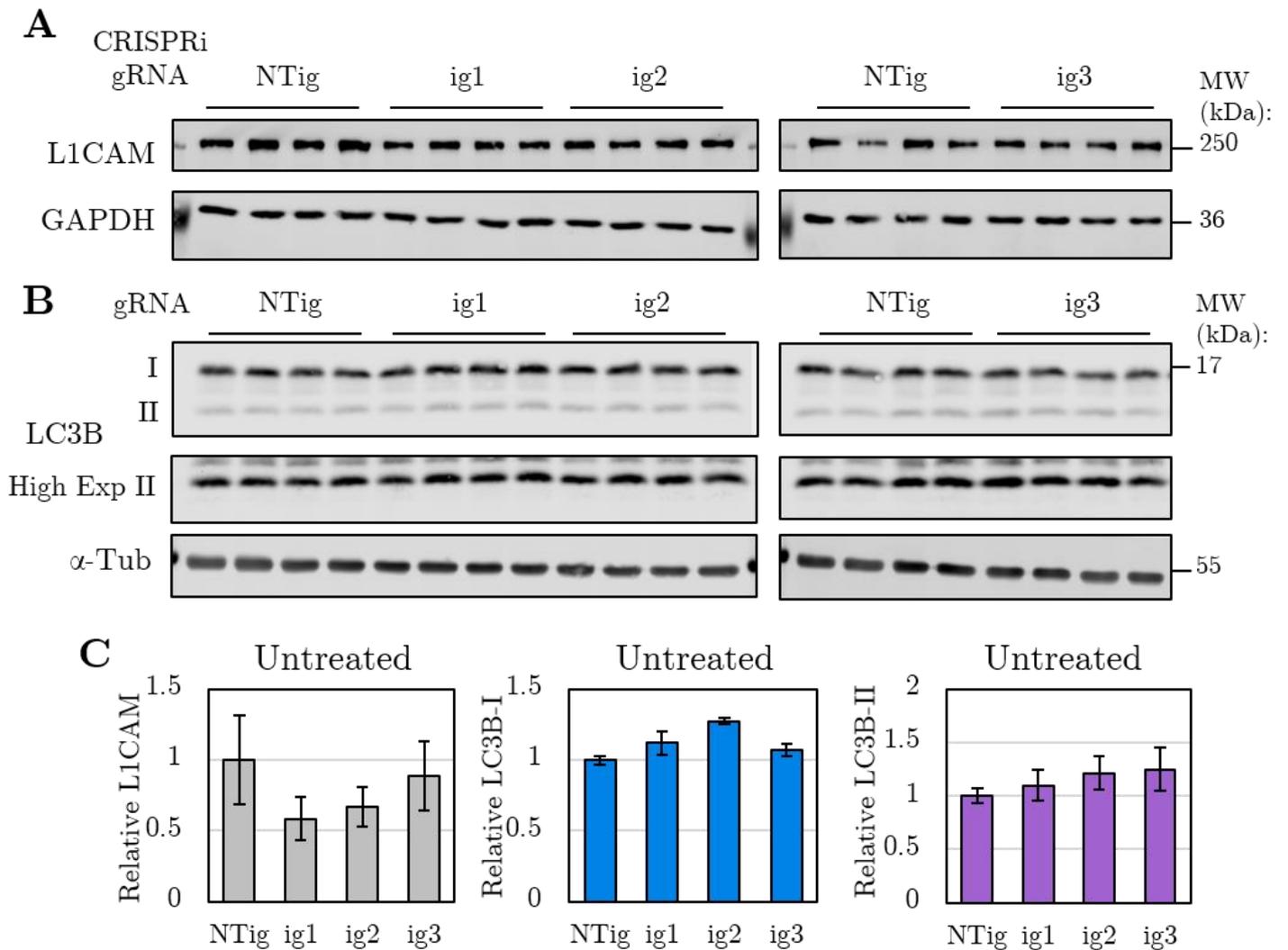
G3-dCas9 cells were transduced with the nontargeting gRNA (NTig) or one of the three *L1CAM*-targeting gRNAs and purified by FACS as described above. Following differentiation, expression of L1CAM protein was reduced slightly but not significantly, and LC3B was not reduced in untreated d14 i³Neurons (Fig. 5.10). In fact, LCB-I was consistently increased over the NTig control in L1-ig2 cells, although the magnitude of this change was small.

These results implied that a moderate loss of L1CAM was unable to affect macroautophagy in the same way as a near-complete loss. It further suggested that a mixture of transcripts may be present, or that transcription initiation may be adjusted to compensate for dCas9 inhibition of one site. While it was hypothesized that a combination of gRNAs may be able to silence *L1CAM* expression more effectively, more promising results from the other strategies resulted in CRISPRi to be left as a potential approach for future studies.

Figure 5.10: CRISPRi fails to silence *L1CAM* in *i*³Neurons.

G3-dCas9 cells were transduced with a nontargeting gRNA (NTig) or with gRNAs targeting either exon 1 (ig1 & ig2) or exon 2 (ig3) *L1CAM* transcription start sites. Lines were FACS purified and differentiated for 14 days prior to lysis.

- (a) L1CAM was not obviously reduced in untreated CRISPRi *i*³Neurons. The L1CAM 74-5H7 antibody was used.
- (b) LC3B-I was the predominant form of LC3B, indicative of a low rate of autophagosome synthesis in *i*³Neurons. A high exposure image of LC3B-II is also shown. No differences between gRNA conditions were noted.
- (c) Quantification of n=1 differentiation with 4 separate wells as technical replicates (error bars \pm SD of technical replicates). L1CAM protein levels were slightly reduced, but LC3B-I and LC3B-II levels were not reduced. Because of poor efficacy, additional replicates were not performed.



5.4.3 shRNA-mediated degradation of *L1CAM* mRNA

Genetically encoded shRNAs function similarly to siRNAs to silence gene expression by binding to mRNA and inducing its degradation via the RISC complex (Moore et al., 2010). As a negative control, one non-targeting shRNA (NTsh) was designed with the same sequence as the non-targeting CRISPRi gRNA above, and three shRNAs were designed to target all known *L1CAM* transcripts. L1-sh1 targeted exon 13; L1-sh2 targeted exon 17 and partially overlapped with the target for si1; and L1-sh3 targeted the exon 20-21 junction, partially overlapping the target for si3 (Table 2.2). These overlaps were by design in an effort to test whether the varying effects of the siRNAs were replicable by binding to similar regions of the mRNA while featuring different collections of potential off-target effects.

Plasmids containing each shRNA expression cassette were constructed as described in Chapter 2, and iPSCs were transduced with the shRNAs and purified by FACS as above. After differentiation, L1CAM protein was significantly reduced in all shRNA cells as compared to the NTsh cells (Fig. 5.11a-b). This reduction was not equal, however, as sh1 and sh2 typically reduced L1CAM to <1% of NTsh, while sh3 reduced L1CAM to approximately 20-30% of the control. Also, NTsh cells demonstrated comparable expression of L1CAM and LC3B with non-transduced G3 wildtype cells differentiated at the same time (Fig. 5.11c). This indicated that neither the process of viral transduction and FACS purification nor the expression of the selection genes and shRNAs altered the expression of L1CAM or LC3B, so NTsh was a viable control for the *L1CAM*-targeting shRNAs.

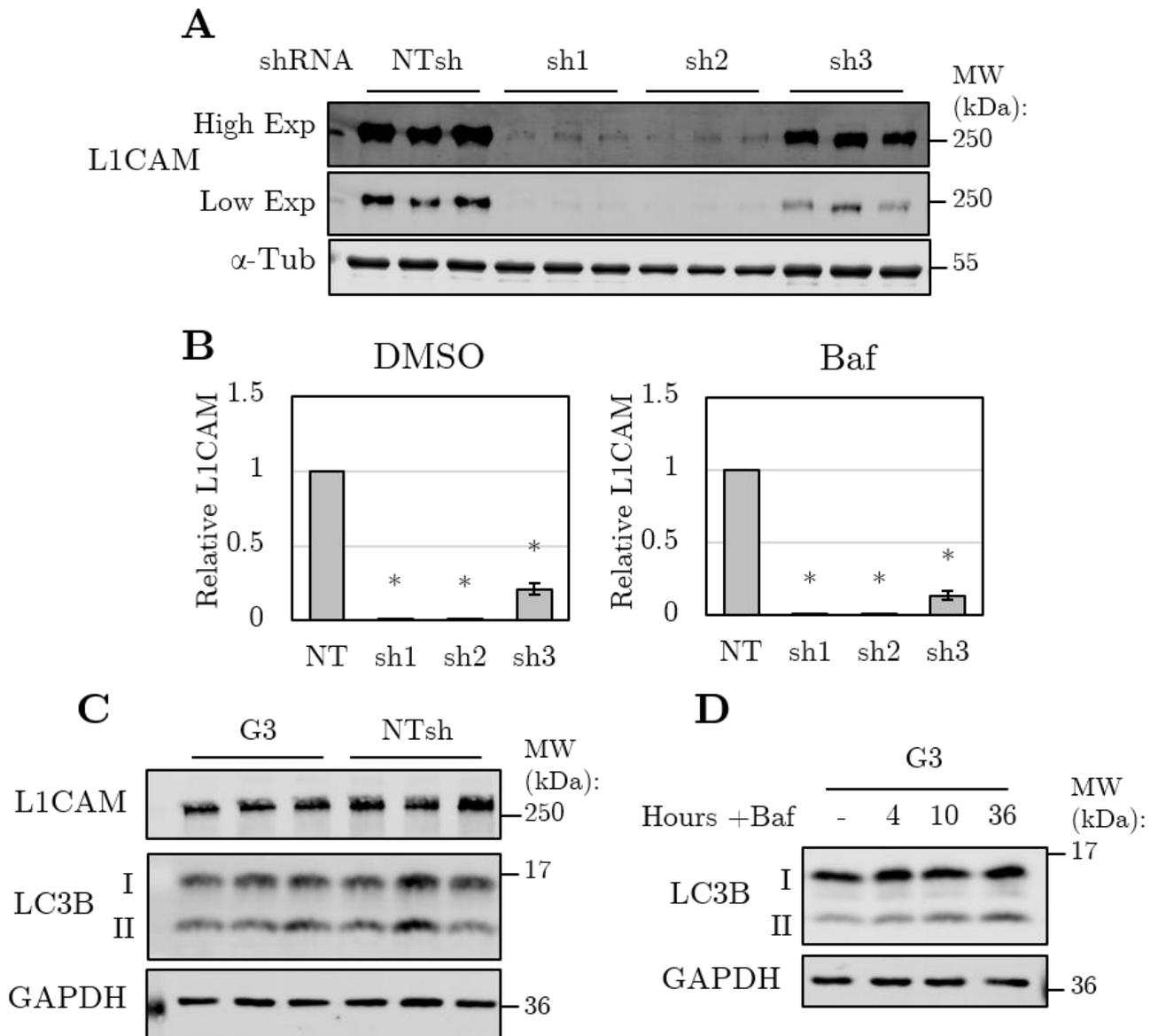
Figure 5.11: shRNAs reduce L1CAM in i³Neurons.

Wildtype G3 iPSCs were transduced with a nontargeting shRNA (NTsh) or with shRNAs targeting all known L1CAM transcripts. Lines were FACS purified and differentiated for 14 days prior to lysis and western blotting for analysis.

(continued)

Figure 5.11 (*continued*): shRNAs reduce L1CAM in i³Neurons.

- (a) Blots of L1CAM in untreated differentiated i³Neurons expressing the indicated shRNA. The L1CAM 74-5H7 antibody is shown.
- (b) Quantification of n=5 independent differentiations, each with 3 separate wells as technical replicates. L1CAM protein levels were significantly reduced in all three shRNA lines as compared to the NTsh line regardless of treatment.
- (c) Similar expression was noted for both L1CAM and LC3B in DMSO-treated i³Neurons from the G3 parental and NTsh lines that were differentiated at the same time. The L1CAM D-5 antibody was used.
- (d) WT i³Neurons were treated with 400nM baf for the indicated duration immediately before lysis, and levels of LC3B were assessed.



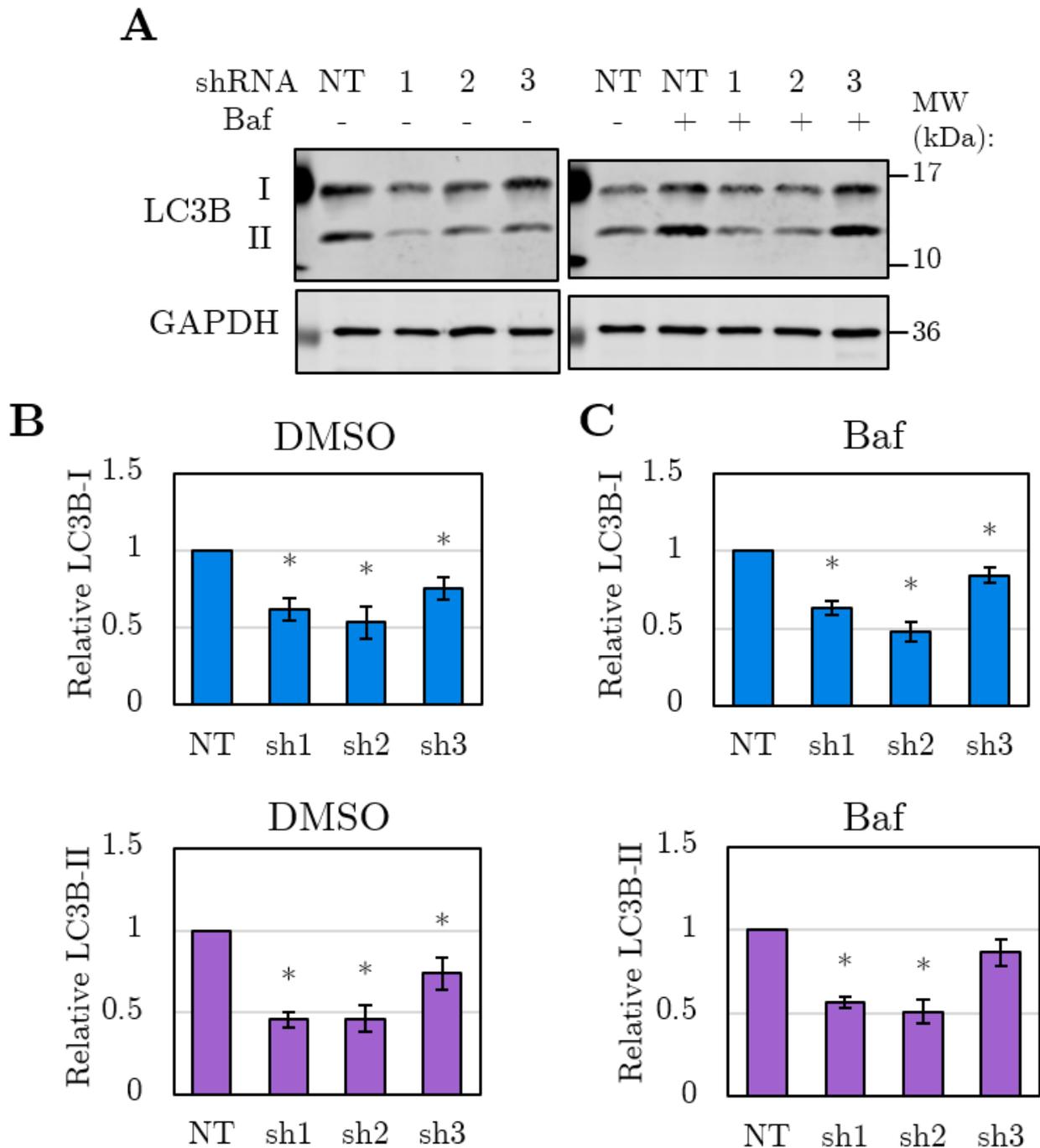
In conjunction with the observation that LC3B-I was the predominant form of the protein in i³Neurons, it was found that 4 h of 400 nM baf treatment, as typically used in HeLa cells, did not increase LC3B-II. This indicated that these cells had a slow rate of macroautophagy flux, consistent with a low rate of protein turnover in general. A long average half-life of proteins has likewise been noted via SILAC proteomics in mature mouse primary neurons (Dörrbaum et al., 2018) and in other i³Neurons (Fig. 4.14; Hao et al., unpublished).

Extending the duration of baf treatment was found to increase the level of LC3B-II with time (Fig. 5.11d). However, after 36 h baf treatment, i³Neurons exhibited increased debris in culture medium, so subsequent experiments used baf at 400 nM for 24 h. While LC3B-I remained the predominant form, LC3B-II was sufficiently increased to adequately indicate whether effects to LC3B were due to a change to lysosomal degradation (if there was a “smoothing over” of differences in conditions as compared to the DMSO condition) or to upstream synthesis (if there was a continuation or widening of differences).

Next, the effects of the shRNAs on LC3B were assessed. In DMSO-treated samples, both LC3B-I and LC3B-II were significantly decreased for all L1CAM shRNAs, and it was noted that the magnitude of this change was greater in sh1 and sh2, corresponding to the greater level of knockdown of L1CAM with these shRNAs (Fig. 5.12a-b). After 24 h baf treatment, LC3B-I was significantly decreased for all L1CAM shRNAs, but LC3B-II was only significantly decreased for sh1 and sh2 (Fig. 5.12a & c). The magnitude of the decrease for sh3 was again notably less than sh1 or sh2, which parallels its reduced effect on L1CAM reduction.

Figure 5.12: shRNA knockdown reduces LC3B-I & LC3B-II.

- (a) Representative blots of the effects of L1CAM-targeting shRNAs on LC3B-I and LC3B-II in *i*³Neurons after 24 h treatment with DMSO (-) or 400nM baf (+).
- (b-c) Quantification of n=5 independent differentiations as in (a), each with 3 separate wells as technical replicates, after treatment with DMSO (b) or baf (c). LC3B-I was significantly reduced for all shRNAs for both treatments; LC3B-II was significantly reduced for all shRNAs after DMSO treatment and for sh1 and sh2 after baf treatment.



Together, these data supported the hypothesis that a strong knockdown of L1CAM (i.e., in sh1 and sh2 i³Neurons) decreases macroautophagy induction. The continuation of these differences after baf treatment further indicated that the defect was not due to degradation; rather, it suggests that decreased LC3B-I provided less substrate to be conjugated to autophagosome membranes as LC3B-II. These changes appeared to be contingent on the strength of L1CAM knockdown, as sh3 demonstrated significant but muted decreases to LC3B-I in both treatment conditions.

5.4.4 CRISPR-mediated knockout

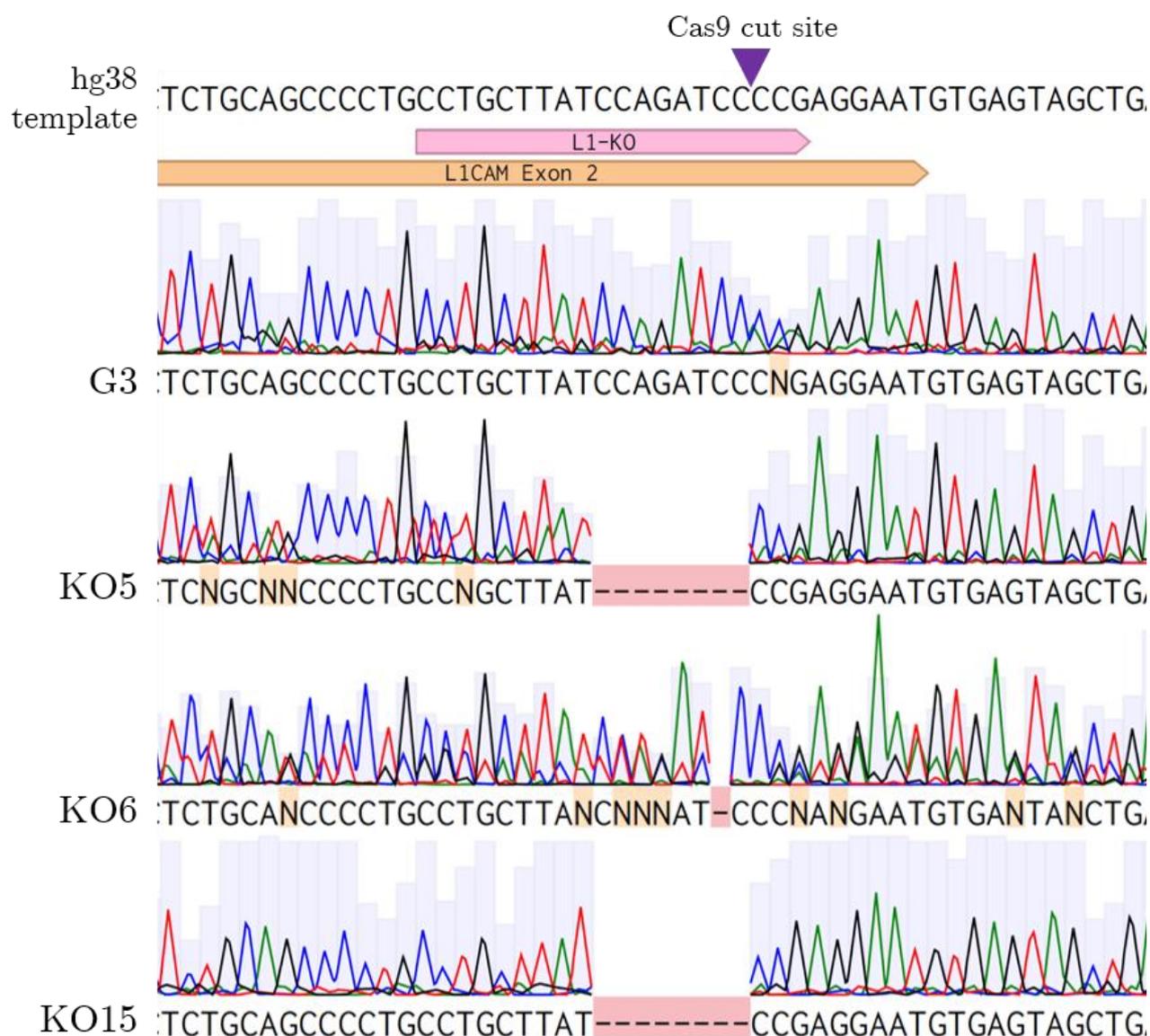
Genetic ablation of *L1CAM* was achieved by transient co-transfection of iPSCs with plasmids expressing Cas9 and L1-KO gRNA. This gRNA was designed to target exon 2 near the beginning of the translated region in order to induce an early frameshift mutation and thereby prevent the expression of functional L1CAM protein.

One week after transfection, no BFP fluorescence was noted, indicative of plasmid degradation without integration. These cells were plated at low density, clones were picked, and genomic DNA was collected. Primers flanking the target site were used to PCR amplify the target region, and this fragment was Sanger sequenced for each clone. Because *L1CAM* is on the X chromosome and the iPSCs were derived from a male donor, only one copy of the gene was present, simplifying the genotyping process.

Three clones were identified that featured frameshift mutations at the gRNA target site consisting of a deletion of either 8 bp for KO clones 5 and 15, or 1 bp for KO clone 6 (Fig. 5.13). Sequence quality was mediocre, particularly for KO6; while this could have been due to mosaicism, some low amplitude traces continued longer than the predicted amplicon, suggesting rather that off-target amplification occurred or exogenous DNA was introduced after PCR purification. This was therefore discounted.

Figure 5.13: Sequence traces of L1CAM KO clones.

G3 cells were transiently co-transfected with plasmids encoding Cas9 and a gRNA targeting exon 2 (L1-KO), with a cut site 67bp after the beginning of translation. Clones were picked and validated by PCR and Sanger sequencing (Genewiz). KO clones 5 and 15 featured 8bp deletions; KO clone 6 featured a 1bp deletion, although poor sequence quality was noted. Alignment & image by Benchling. Bars behind the trace represent base quality.



Using cDNA positions, the specific mutations observed were c.60_67del in KO clones 5 and 15 and c.66del in KO clone 6 (den Dunnen et al., 2016). It was possible that KO clones 5 and 15 originated from the same parental cell because they featured the same mutation, but they were subsequently used as separate clones since they were picked and sequenced independently.

Both deletions resulted in a frameshift in the protein coding sequence, but they were each in different frames (i.e., +1 and -1). While the KO5 & KO15 deletion maintained Ile20, it was followed by a termination sequence after four codons. If exon 3 was skipped, termination instead occurred after 24 codons. Meanwhile, the KO6 deletion maintained Ile22 and was followed by a termination sequence after nine codons with (or four codons without) the inclusion of exon 3. A diagram of the effect of the KO clone mutations is provided in Fig. 5.14a.

In addition to the non-transfected parental G3 line, five clones that had been transfected, picked, and sequenced without any mutation at the cut site were identified as wildtype control clones (WT clones 2, 8, 10, 12, 14). Other clones that demonstrated unclear or impure sequencing results were excluded from subsequent experiments.

Figure 5.14: CRISPR-mediated KO of L1CAM.

WT clones 2, 8, 10, 12, and 14 were transfected and picked alongside the KO clones and sequenced with no mutations. Lines were differentiated for 14 days prior to lysis and western blotting for analysis.

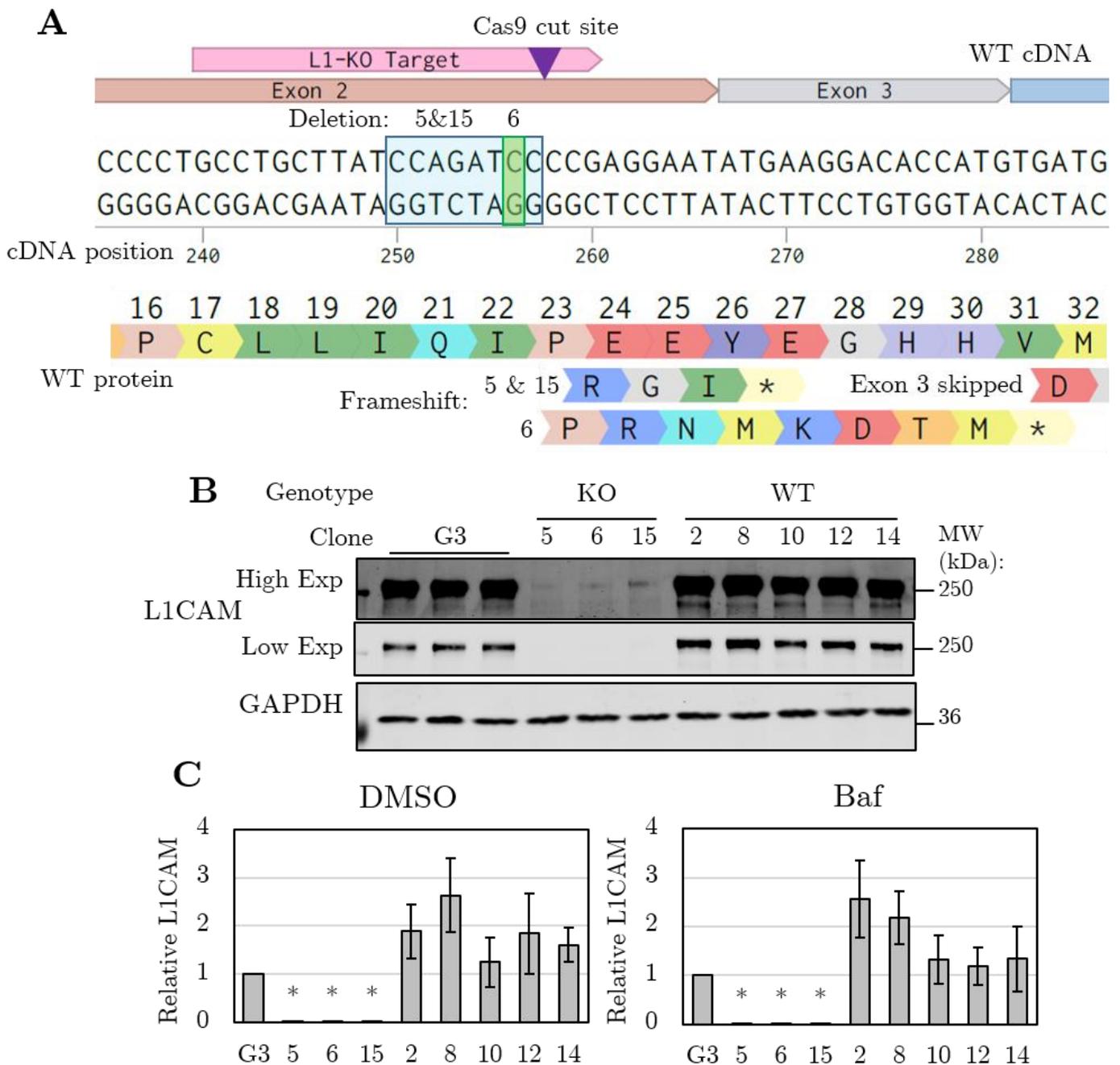
(a) Diagram of the wildtype allele and KO clones with frameshift deletions. If exon 3 was skipped, another 21 codons were translated for KO5 & KO15, and five fewer codons were translated for KO6. Image adapted from Benchling.

(b) Blots of L1CAM levels in the non-transfected WT G3 parental line and the indicated clones of DMSO-treated i³Neurons differentiated at the same time. The L1CAM 74-5H7 antibody was used.

(continued)

Figure 5.14 (continued): CRISPR-mediated KO of L1CAM.

(c) Quantification of n=4 independent differentiations for G3 and KO clones, each with 3 separate wells as technical replicates, and n=1 differentiation for the WT clones, with 4 separate wells as technical replicates (WT clone error bars \pm SD of technical replicates). WT clones were variable, but all expressed at least as much L1CAM protein as the G3 parental line, so G3 was deemed a conservative control. Significance was only tested for the KO clones against G3; L1CAM protein levels were significantly reduced in all three KO clones as compared to the G3 parental line after both DMSO and baf treatment.



After differentiation, L1CAM was not detectable in any of the knockout clones, while the wildtype clones demonstrated L1CAM expression that was equal to or greater than the parental G3 line (Fig. 5.14b-c). A panel of three antibodies were tested against full-length blots to confirm that no alternative gene products were expressed; while the cleavage product at 80kDa was prevalent in the WT lines, little to no protein was observed in any of the KO clones for any of the antibodies tested (Fig. 5.15).

Surprisingly, then, the amount of LC3B-I and LC3B-II did not significantly change in any of the KO clones as compared to the G3 parental line after DMSO treatment (Fig. 5.16), and the KO clones demonstrated a slight increase to both LC3B-I and LC3B-II after 24 h baf treatment, with LC3B-I significantly increased for KO5 and KO6 (Fig. 5.17). These cells were all seeded at the same density, and no obvious differences to survival or growth were observed. This finding effectively negated the hypothesis that L1CAM protein is the main actor in mediating the effect to LC3B as seen in the siRNA-treated HeLa cells and shRNA-expressing i³Neurons above.

While such a finding may further suggest that the reduction to LC3B and autophagic clearance seen in siRNAs and shRNAs were due to off-target effects, this was unlikely because these effects were consistent across two cells types and several reagents that lacked shared off-target binding loci. Alternatively, since the CRISPR gRNAs used in the HeLa screen targeted exon 10 and demonstrated a decrease to mutant protein clearance, mutagenesis further downstream in the gene may affect L1CAM differently than near the N-terminus. This possibility is explored in more detail in the discussion.

Figure 5.15: L1CAM KO reduces all cleavage products.

Three L1CAM-targeting monoclonal antibodies with affinity for different epitopes (Fig. 5.2) were blotted for all molecular weights in the indicated i^3 Neuron line. Blots were co-stained for β 3-Tubulin as a loading control, and this signal can be seen to bleed through to the 800 nm fluorescence channel at 55 kDa and 100 kDa. Cleavage products (as in Fig. 5.2) are marked on the left.

- (a) The 5G3 antibody binds in the Ig-like domains within 230 amino acids of the N terminus. It was therefore specific for the full-length protein (I) since other products were shed from the cells.
- (b) The D-5 antibody binds near amino acids 925-1120, the region of the ECD abutting the TMD. It recognized the full length protein (I) as well as multiple cleavage products (II, IV, V). Particularly prevalent was IV near 80 kDa. At most, the KO clones presented a dim signal near the full-length protein. A nonspecific band was also present at 90 kDa.
- (c) The 74-5H7 antibody binds near amino acid 1172 in the ICD. It was thus capable of detecting the full-length protein (I) and any retained fragment (IV, VI) as well as the cytoplasmic ICD (VII) at 28-32 kDa.

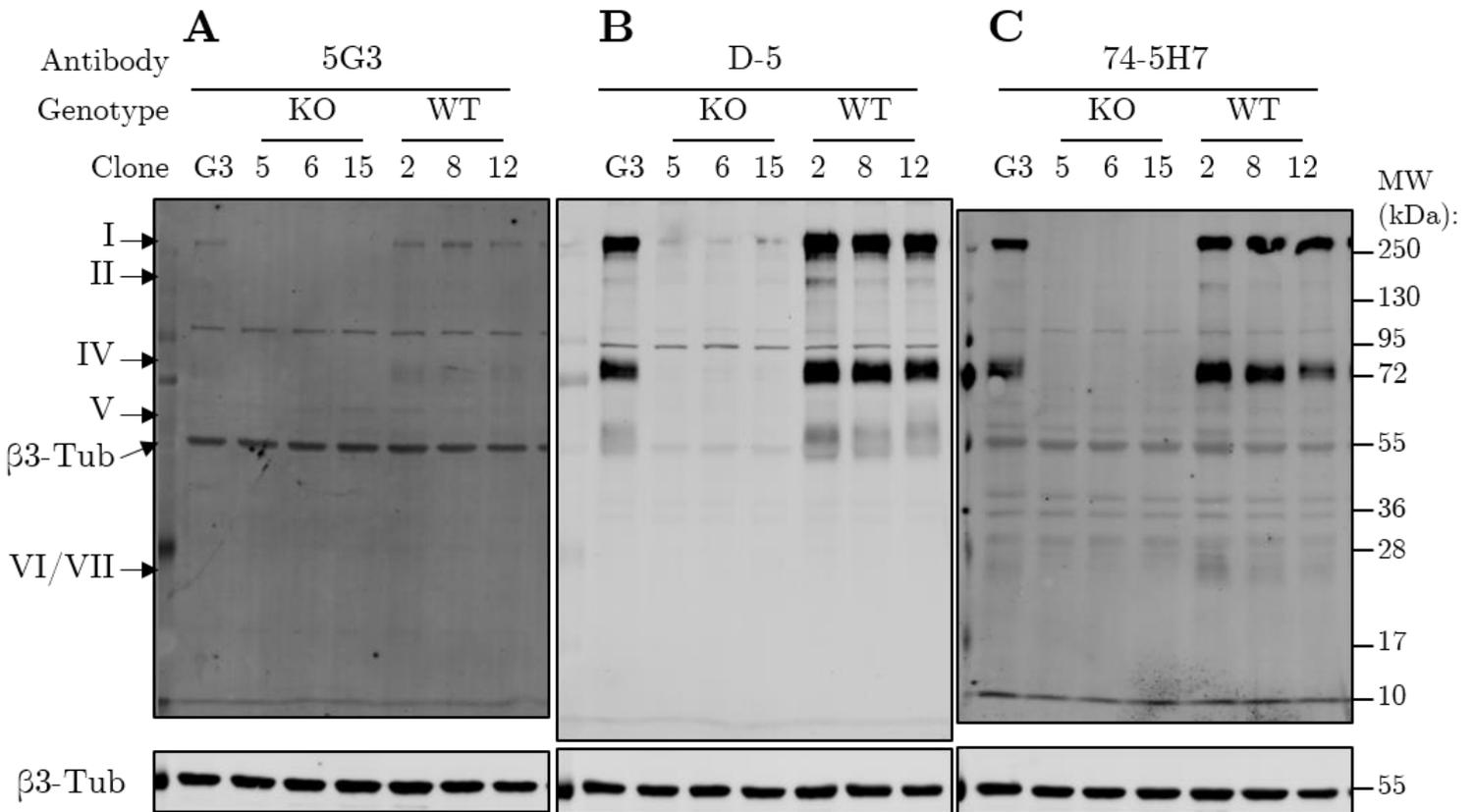


Figure 5.16: L1CAM KO & DMSO treatment does not affect LC3B.

- (a) Representative blots of LC3B-I and LC3B-II in the parental G3 line or the indicated clones after 24h treatment with DMSO (-).
- (b) Quantification of n=4 independent differentiations for G3 and KO clones, each with 3 separate wells as technical replicates, and n=1 differentiation for the WT clones, each with 4 separate wells as technical replicates (WT clone error bars \pm SD of technical replicates). No significant changes were observed.

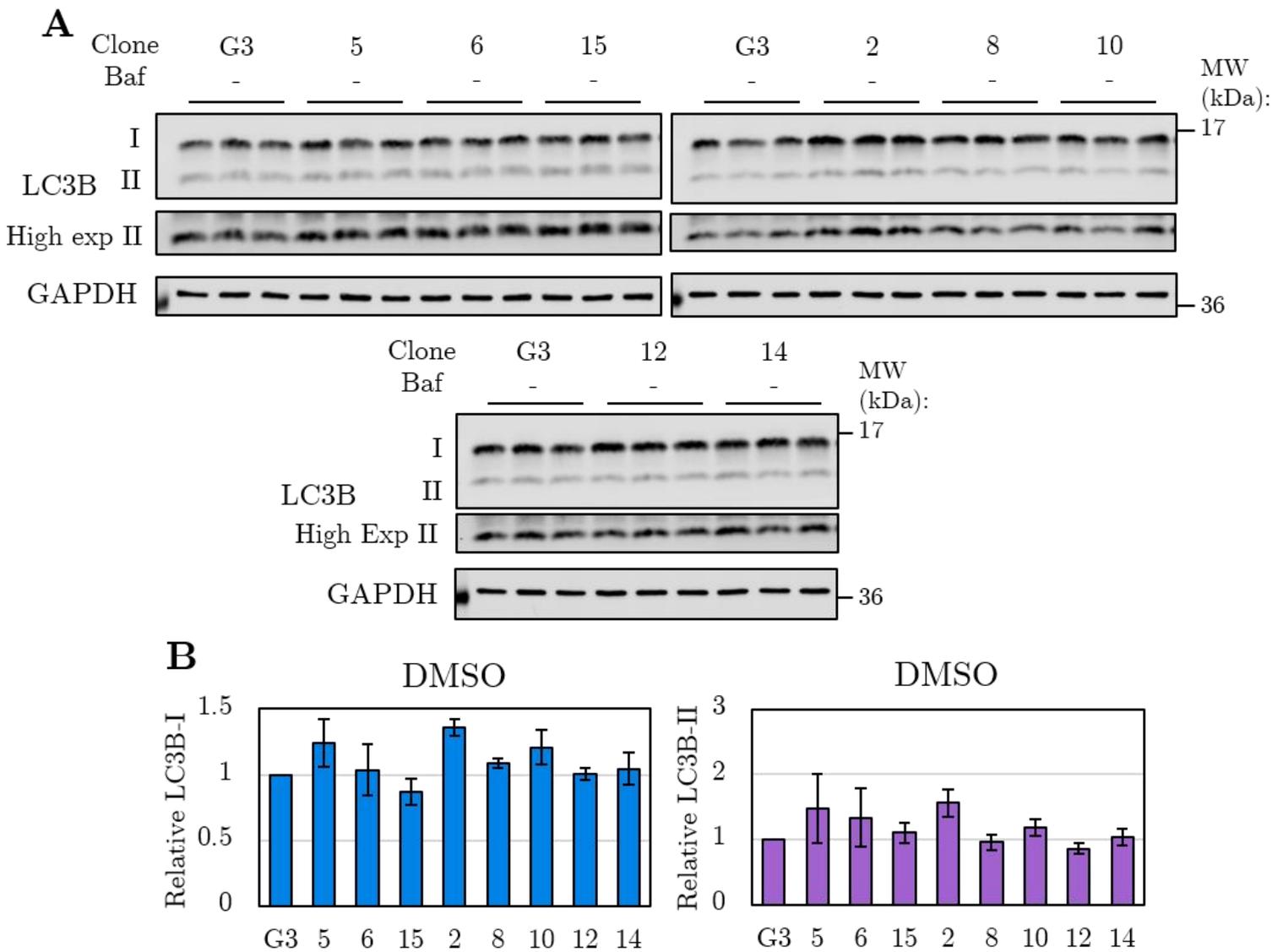
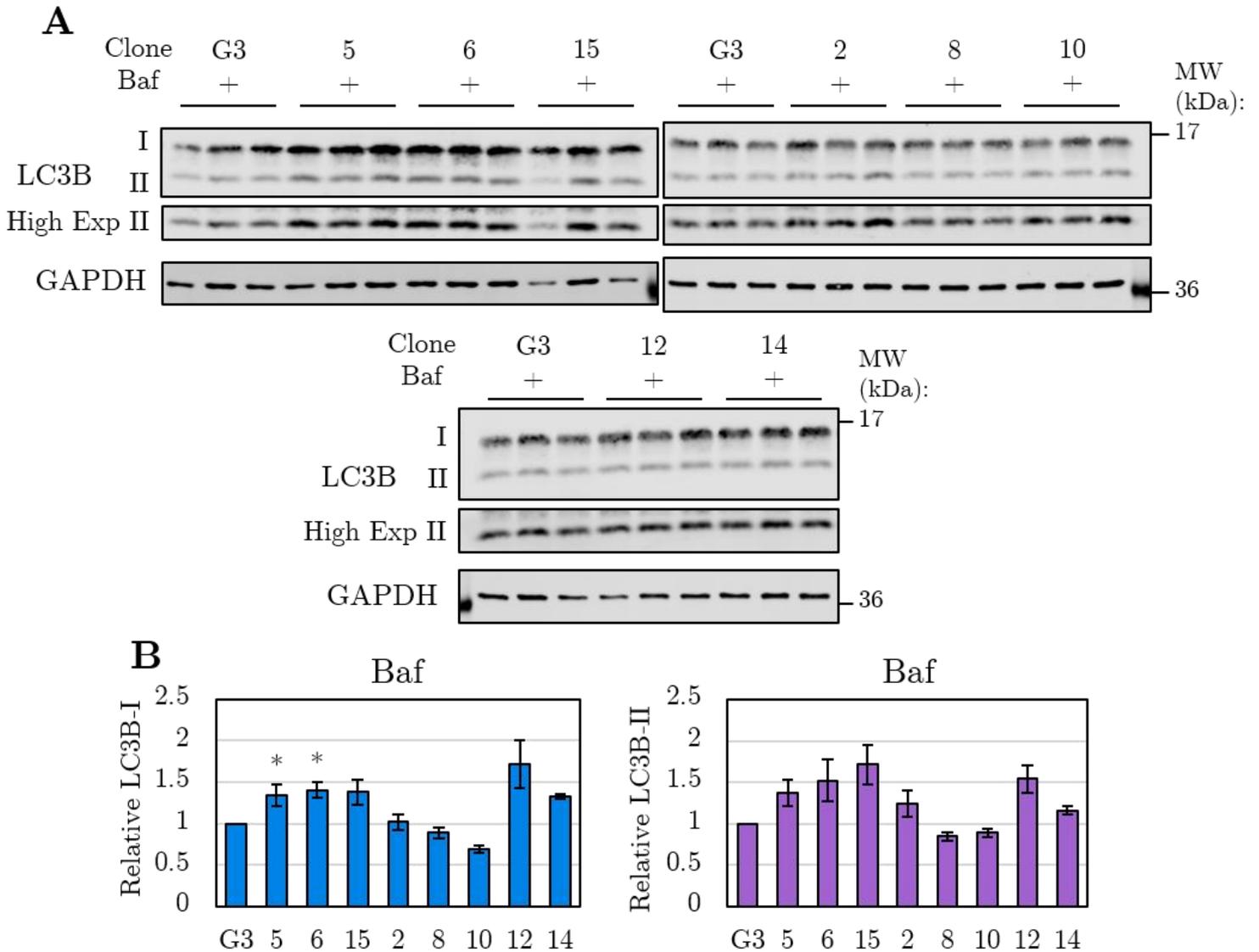


Figure 5.17: L1CAM KO & baf treatment does not reduce LC3B.

- (a) Representative blots of LC3B-I and LC3B-II in the parental G3 line or the indicated clones after 24h treatment with baf (+).
- (b) Quantification of n=4 independent differentiations for G3 and KO clones, each with 3 separate wells as technical replicates, and n=1 differentiation for the WT clones, each with 4 separate wells as technical replicates (WT clone error bars \pm SD of technical replicates). Significance was only tested for the KO clones against G3; LC3B-I was significantly increased in KO5 and KO6, but the magnitude of this change was relatively small.



5.5 Characterizing i³Neuron model lines

5.5.1 LC3B is unaffected in iPSCs

As was shown previously, L1CAM was not expressed in undifferentiated iPSCs. However, if the shRNAs were affecting LC3B expression through an off-target effect, it stands to reason that a deficit in macroautophagy may be present in the iPSCs regardless of L1CAM expression—unless the expression of such an off-target gene was also restricted to differentiated i³Neurons.

Indeed, L1CAM was not detected in any iPSCs, and both LC3B-I and LC3B-II were stable across all shRNA lines (Fig. 5.18a & c-d). This suggested that, in the absence of L1CAM expression, the presence of any of the shRNAs alone was not sufficient to reduce LC3B, and it limited the potential for the effects seen in shRNA-expressing i³Neurons to be due to promiscuous mRNA silencing. This also supported the use of independent differentiations as the point of replication since the biological action of the shRNAs is restricted to differentiation. Furthermore, no changes to LC3B were noted in any of the KO clones as compared to the G3 parental line, consistent with the findings in differentiated i³Neurons (Fig. 5.18b & c-d).

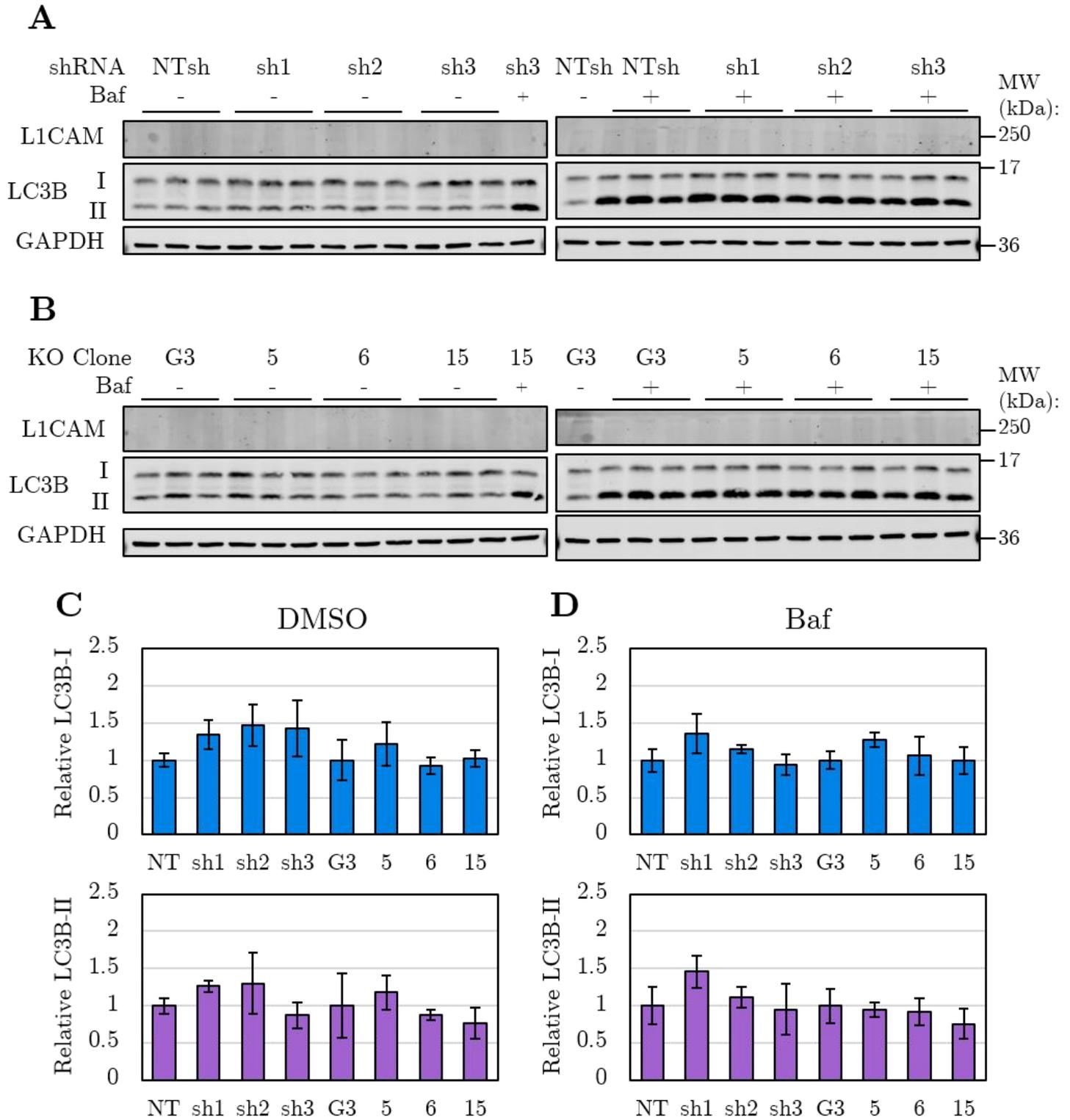
Figure 5.18: iPSCs lack L1CAM & exhibit no LC3B effects.

(a-b) Blots of L1CAM and LC3B in the indicated shRNA line (a) and G3 or KO clone (b) as undifferentiated iPSCs after 4 h treatment with DMSO (-) or 400 nM baf (+). The L1CAM 74-5H7 antibody was used.

(c-d) Quantification of n=1 experiment for each line with 3 separate wells as technical replicates, normalized to the mean of the associated control line (error bars \pm SD of technical replicates). No changes were notable.

(continued)

Figure 5.18 (*continued*): iPSCs lack L1CAM & exhibit no LC3B effects.



5.5.2 Pluripotency erasure is not affected by reduction of L1CAM

Because the shRNAs were constitutively expressed, *L1CAM* mRNA was actively degraded in these lines as its transcription was activated during differentiation. Additionally, KO clones lacked functional L1CAM protein altogether due to the introduction of nonsense frameshift mutations. It was thus important to determine if the differentiation process and resulting i³Neuron cell state was affected.

While all iPSC lines expressed high levels of NANOG in the nucleus, this expression was lost after differentiation (Fig. 5.19a & c). Interestingly, some puncta were noted to be stained by the NANOG antibody in only the L1CAM-targeting shRNA and KO lines. These were not associated with nuclei and thus were not considered to be actively influencing gene expression, but their presence may indicate that a full loss of pluripotency proceeded more slowly in these lines than in the control lines (NTsh and G3). Furthermore, all lines were found to exhibit changes to nuclei indicative of a loss of pluripotency, as described in Chapter 2 (Fig. 5.19b & d). A loss of L1CAM was therefore not sufficient to prevent the erasure of pluripotency.

Figure 5.19: L1CAM reduction does not prevent pluripotency erasure.

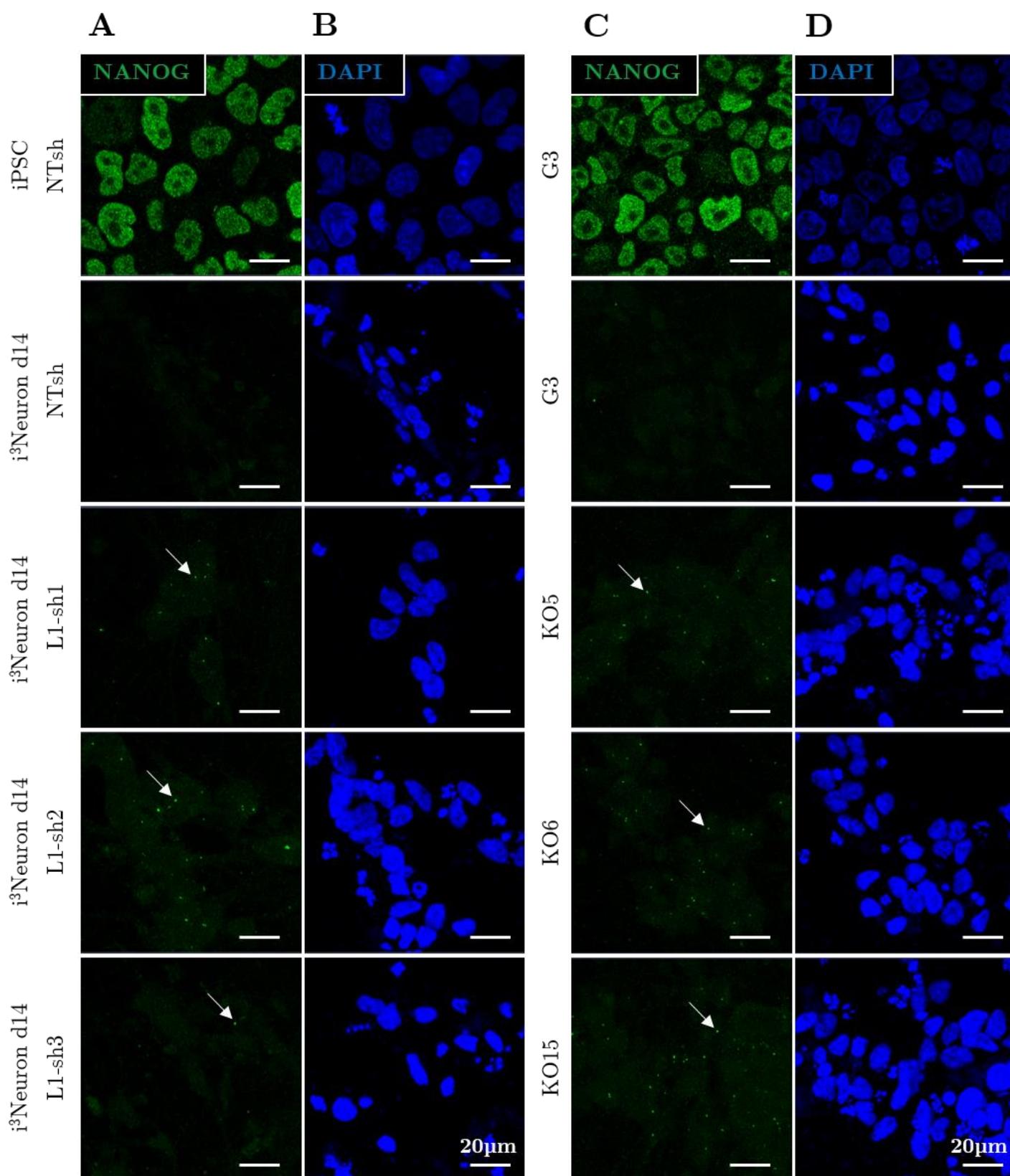
iPSCs were imaged and presented as a single confocal slice, while neurons were imaged as a z-stack and presented as maximum intensity projections.

(a&c) NANOG was used as a marker of pluripotency. While iPSCs expressed it strongly, none was detected in nuclei after 14 days of differentiation to i³Neurons in shRNA lines (a) or in the G3 parental line and KO clones (c). However, scattered puncta were observable in only the L1CAM-reduced lines that did not associate with nuclei or debris; one such punctum is indicated for each line with a white arrow.

(b&d) DAPI was used to stain nuclei, and morphological changes indicative of differentiation were evident in i³Neurons (b & d).

(continued)

Figure 5.19 (*continued*): L1CAM reduction does not prevent pluripotency erasure.



5.5.3 Neuronal acquisition is not affected by reduction of L1CAM

It was next sought to image L1CAM directly in order to further validate the efficacy of knockdown and knockout and to assess potential changes to protein localization. Mutations to and inhibition of L1CAM are known to affect neurodevelopment, so it was also incumbent to determine whether the shRNA and KO cells were able to properly acquire a neuronal phenotype and whether alternative differentiation pathways were active.

The L1CAM 5G3 antibody was found to provide the best immunofluorescence staining of the antibodies used in this work, possibly because it binds to the furthest N-terminal region of the protein and demonstrated the most specificity by western blot. L1CAM was undetectable in any lines as iPSCs, consistent with expected expression patterns (Fig. 5.20a). After differentiation to i³Neurons, L1CAM was expressed and localized principally on the plasma membrane of control lines (NTsh and G3), while it was substantially decreased in shRNA lines and undetectable in KO clones, consistent with western blots of total protein (Fig. 5.21a).

Interestingly, a small subset of cells in the shRNA lines had L1CAM staining at levels similar to the NTsh and G3 lines, possibly indicating that these cells lost shRNA expression. However, these cells were rare, and no additional morphological differences were noted in these cells compared to those without detectable L1CAM staining. This suggested that a loss of shRNA expression was not heavily selected for as a survival characteristic and that a loss of L1CAM did not prevent neuronal differentiation. It further indicated that shRNA knockdown may have been more effective than previously thought; the small fraction of cells expressing L1CAM may account for a substantial portion of the L1CAM protein detected in the western blots of total protein.

Figure 5.20: L1CAM & GFAP immunofluorescence in iPSCs.

L1CAM (a) and GFAP (b) were co-stained in the indicated line of undifferentiated iPSCs. DAPI was used to stain nuclei (c), and an overlay of the three channels is shown (d). The L1CAM 5G3 antibody was used. iPSCs were imaged and presented as a single confocal slice using the same parameters for L1CAM & GFAP but increased DAPI brightness compared to Fig. 5.21.

(a-b) No staining was detectible for L1CAM or GFAP.

(c) DAPI staining was indicative of an undifferentiated cell state, including large, round nuclei and frequent instances of mitosis.

(continued)

Figure 5.21: L1CAM & GFAP immunofluorescence in i³Neurons.

L1CAM (a) and GFAP (b) were co-stained in the indicated line of i³Neurons differentiated for 14 days. DAPI was used to stain nuclei (c), and an overlay of the three channels is shown (d). The L1CAM 5G3 antibody was used. i³Neurons were imaged as a z-stack and presented as maximum intensity projections using the same parameters for L1CAM & GFAP but decreased DAPI brightness compared to Fig. 5.20.

(a) L1CAM was prevalent on the membrane of the NTsh and G3 lines; while most cells in the sh1, sh2, and sh3 lines were not detectible, a small subset featured staining similar to NTsh, possibly indicating a loss of shRNA expression in these cells (white arrow). No L1CAM was detectible in the KO clones.

(b) No appreciable GFAP staining was present for any line, although it was detectible in all lines at very low intensity.

(c) DAPI staining was indicative of a differentiated cell state, with small, dense nuclei and no mitosis. Frequent instances of fragmented nuclei were observed.

(continued)

Figure 5.20 (*continued*): L1CAM & GFAP immunofluorescence in iPSCs.

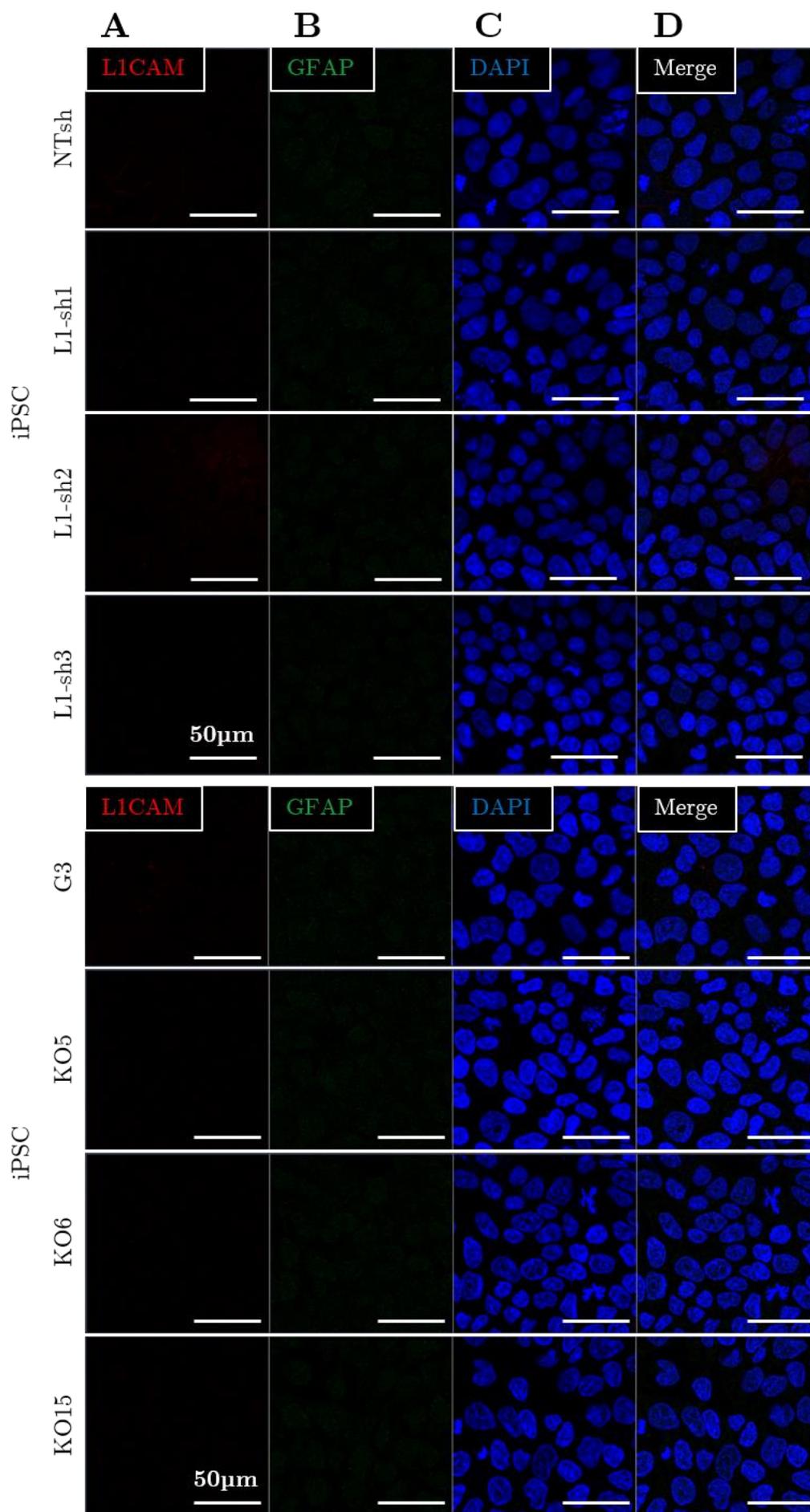
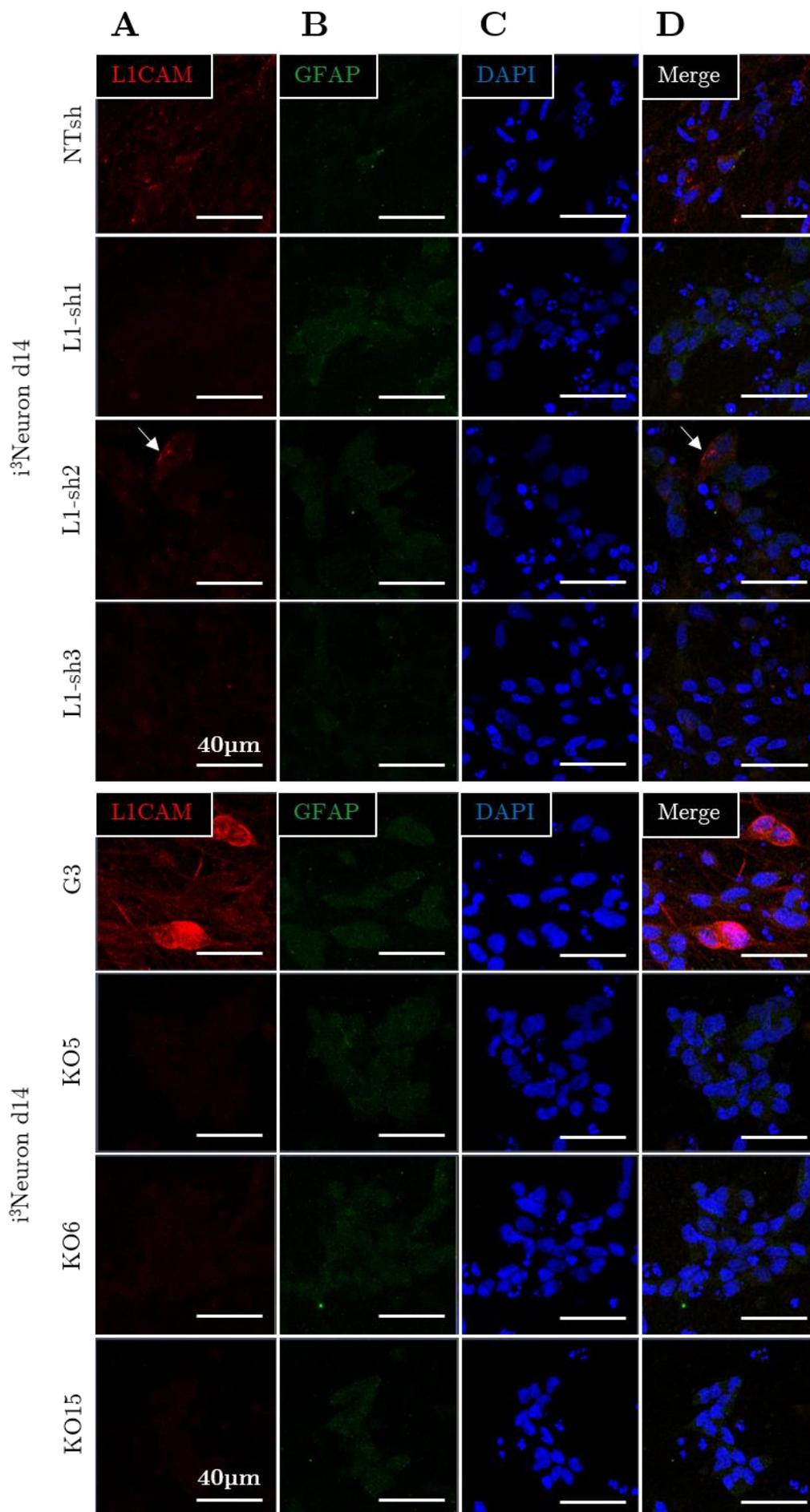


Figure 5.21 (*continued*): L1CAM & GFAP immunofluorescence in i^3 Neurons.



No detectable levels of GFAP staining were observed in any lines as iPSCs (Fig. 5.20b), and very low levels were detected in i³Neurons (Fig. 5.21b). This may indicate that the i³Neurons were not fully mature neurons, but it also provided evidence that the population was pure and that no differences arose after L1CAM reduction. As with the experiment demonstrating erasure, nuclear characteristics were indicative of an undifferentiated cell state in iPSCs (Fig. 5.20c) and of a differentiated cell state in i³Neurons (Fig. 5.21c).

Because all lines were found to share similar expression patterns for other lineage markers, representative images of sh2 and KO6 lines are shown in the following two figures since these particularly demonstrate that cells with L1CAM knockdown and KO expressed these markers as expected. The oligodendrocyte marker MBP was found to be expressed throughout all lines as iPSCs, but it was absent in all of the lines after differentiation to i³Neurons (Fig. 5.22a). Furthermore, all lines were found to express high levels of β 3-Tubulin after differentiation to i³Neurons, while staining was minimal in iPSCs (Fig. 5.22b). The extent of the network of long cellular processes suggestive of neurites is also evident. Co-staining with NeuN and MAP2 demonstrated that both neuronal markers were widely expressed in all lines of i³Neurons but not detectable in any lines of iPSCs (Fig. 5.23).

These data presented strong evidence that a loss of L1CAM due either to shRNA expression or genetic knockout did not impede the acquisition of a neuronal phenotype. While there still may be differences in the maturation state not tested here, i³Neurons from all lines were found to express markers indicative of neurons and to not express markers indicative of pluripotent, progenitor, or glial cell states. Further study may seek to better characterize the neural subtype(s) present, the level of maturation, and the functionality of these neurons, but these data granted sufficient confidence that the cell lines used in this work were neuron-like and reasonably similar to one other.

Figure 5.22: MBP & β 3-Tubulin immunofluorescence in i^3 Neurons.

No changes were observed between any cell lines, so representative images of iPSCs and i^3 Neurons are shown for sh2 and KO6 lines. The oligodendrocyte marker MBP (a) and neuronal marker β 3-tubulin (b) were co-stained, and DAPI was used to stain nuclei (c). An overlay of the three channels is shown (d). iPSCs were imaged and presented as a single confocal slice, while i^3 Neuron were imaged as a z-stack and presented as maximum intensity projections.

(a) MBP was prevalent in iPSCs but not in i^3 Neurons.

(b) β 3-tubulin was present only in i^3 Neurons.

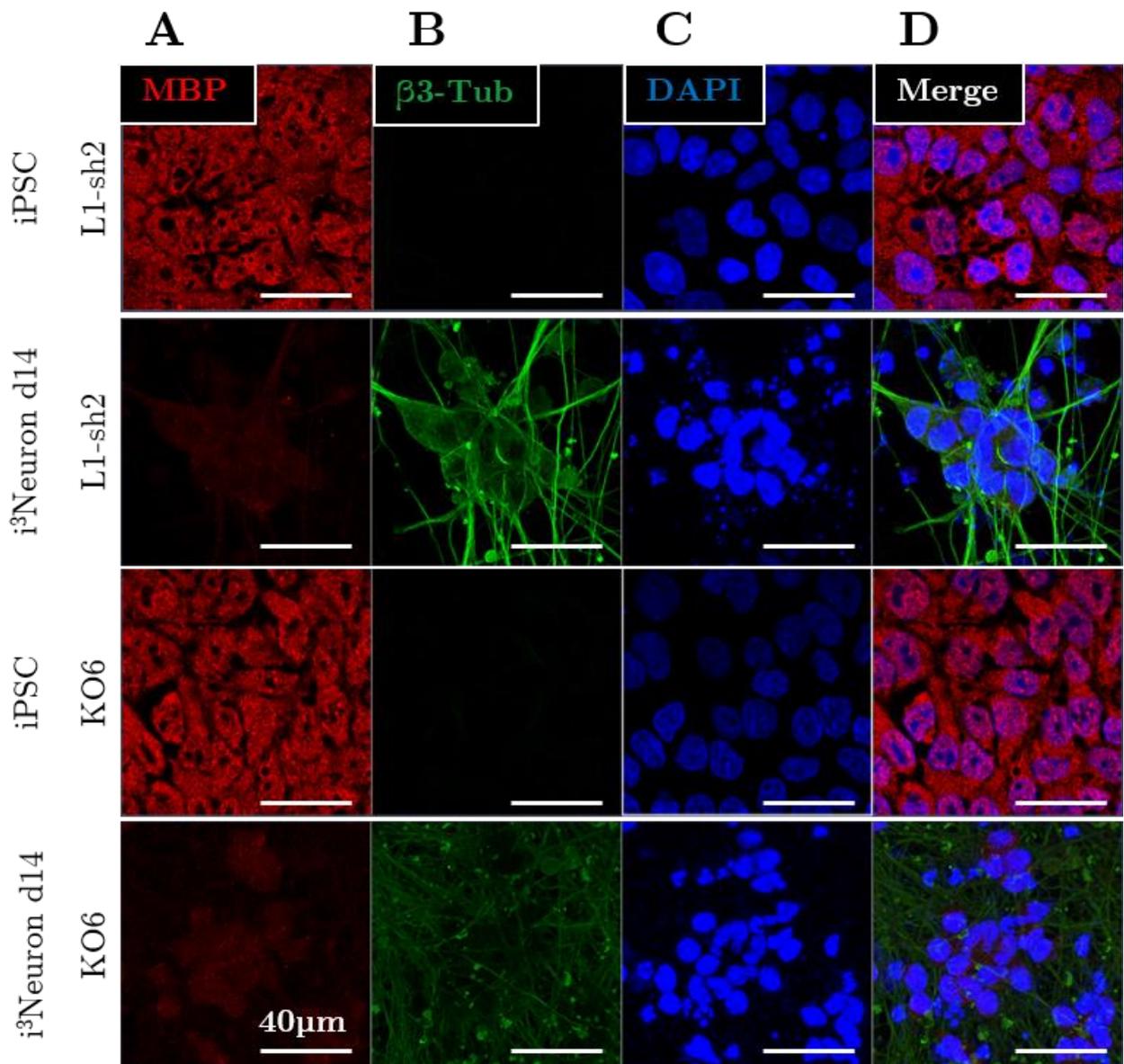
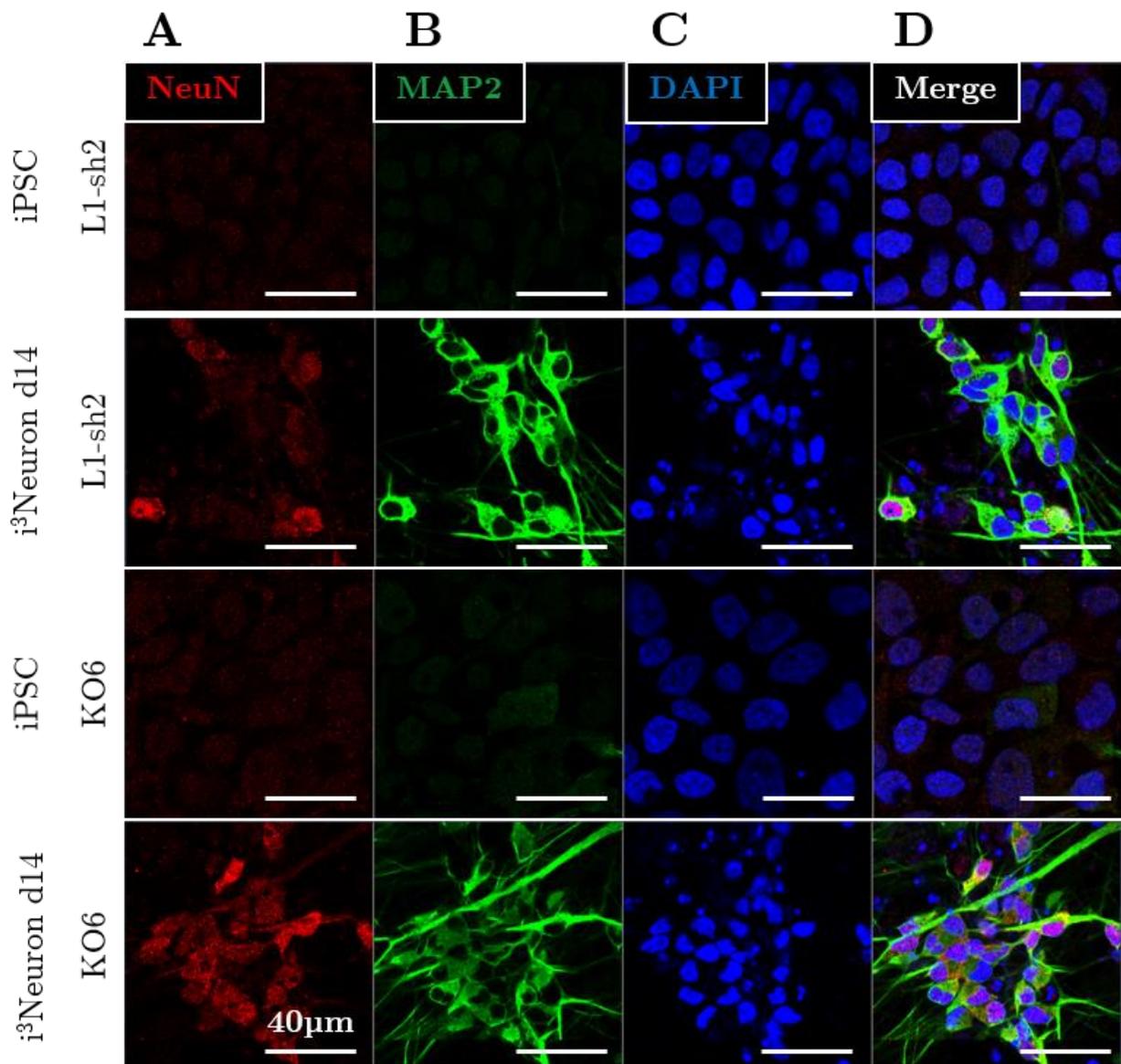


Figure 5.23: NeuN & MAP2 immunofluorescence in i^3 Neurons.

As in Fig. 5.22, representative images of iPSCs and i^3 Neurons are shown for sh2 and KO6 lines. The neuronal markers NeuN (a) and MAP2 (b) were co-stained, and DAPI was used to stain nuclei (c). An overlay of the three channels is shown (d). iPSCs were imaged and presented as a single confocal slice, while i^3 Neuron were imaged as a z-stack and presented as maximum intensity projections. The same experimental parameters for NeuN and DAPI were used for all lines and cell types, but MAP2 brightness was reduced in i^3 Neuron images.

- (a) NeuN was detected at very low levels in iPSCs, but it was present in the nuclei of many i^3 Neurons.
- (b) MAP2 was not detected in iPSCs; it was expressed highly in i^3 Neurons and localized to cell bodies and a subset of projections.



5.6 L1CAM shRNAs reduce transcription of the *ATG8*s

Because the shRNA-expressing i^3 Neurons and siRNA-transfected HeLa cells demonstrated a decrease to LC3B-I, it was hypothesized that this effect was mediated by a reduction in *LC3B* transcription. As noted, the transcriptional pathways that regulate *LC3B* are similar to those of the other *ATG8* family genes, so it was further hypothesized that the other *ATG8*s may be affected. To test these hypotheses, the following set of experiments were performed.

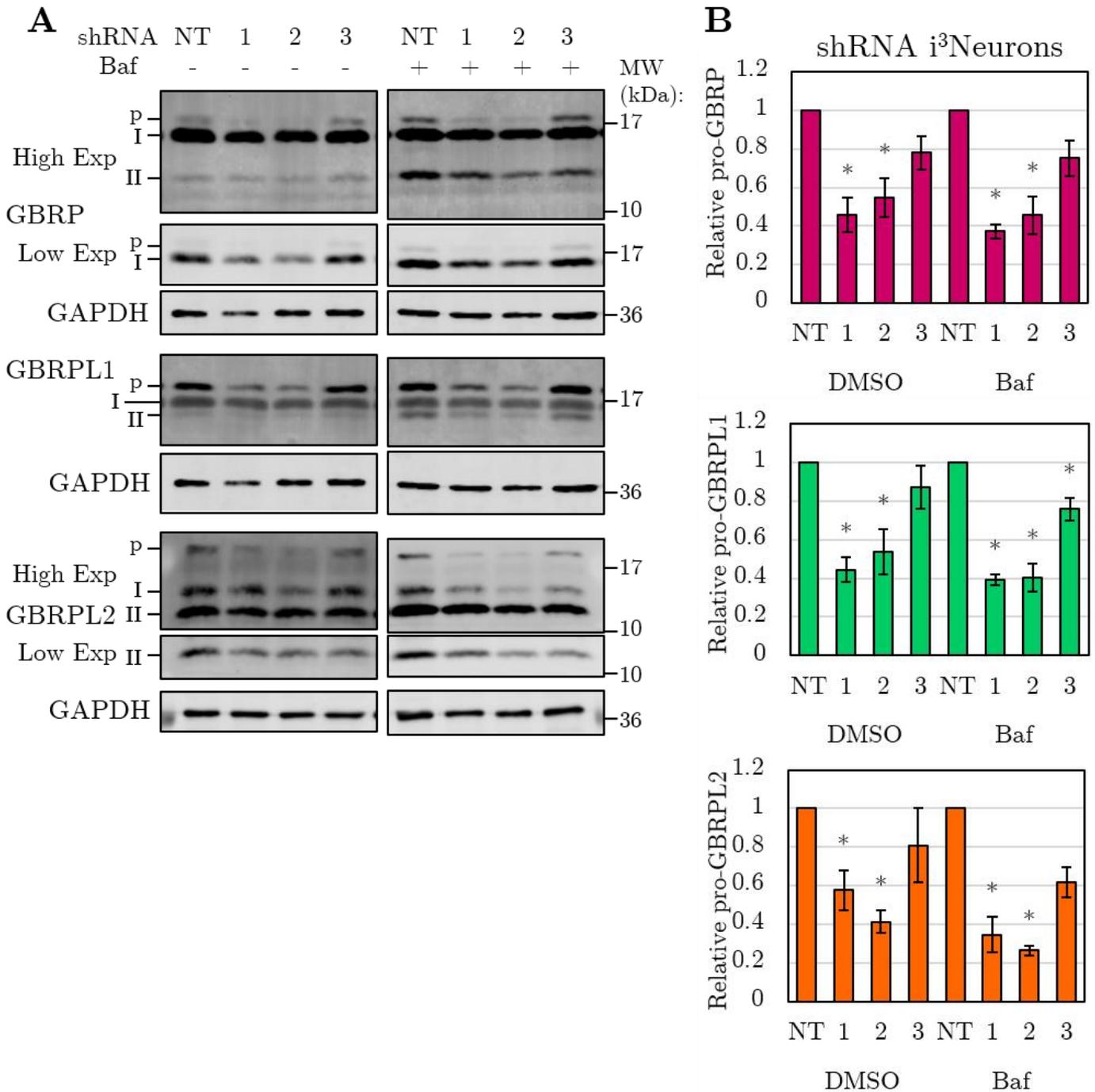
5.6.1 L1CAM shRNAs reduce levels of GABARAP propeptides

First, the three GABARAP proteins in shRNA-expressing i^3 Neurons were assessed by western blot, and several distinct bands were observed. All three GABARAPs (shortened to GBRP in figures) showed the same trends as LC3B; namely, all forms appeared to decrease in sh1 and sh2 i^3 Neurons, there was a muted effect in sh3 cells, and these differences were maintained following baf treatment (Fig. 5.24a). As mentioned in Chapter 1, the *ATG8*s are translated as propeptides, undergo C-terminal cleavage by an *ATG4*, and are lipidated via *ATG7/3/5-12-16*. In order to better assess differences in the synthesis of GABARAPs rather than downstream processing, specifically the propeptide was quantified.

While pro-LC3B migrates at a similar rate as LC3B-II during PAGE separation, making its identification challenging by western blot, the pro-GABARAPs have been shown to migrate more slowly than either modified form (Agrotis et al., 2019; Kabeya et al., 2004). The highest bands evident by western blotting were therefore attributed to the unmodified propeptide forms, while bands that migrated faster were attributed to the cleaved and lipidated forms, respectively. The lipidated form was also identified as the band that increased in relative intensity after baf treatment.

Figure 5.24: pro-GABARAPs are reduced in shRNA *i*³Neurons.

- (a) Representative blots of GABARAP (GBRP), GABARAPL1 (GBRPL1), and GABARAPL2 (GBRPL2) in the indicated shRNA lines after 14 days differentiation ending with a 24 h treatment with DMSO (-) or 400 nM baf (+). Identification of bands: propeptide (p), cleaved by ATG4 (I), lipidated (II).
- (b) Quantification of n=4 independent differentiations, each with 3 separate wells as technical replicates, using the propeptide band of each protein. All three were significantly decreased for sh1 and sh2 in both treatment conditions.



All three pro-GABARAPs were significantly decreased in sh1 and sh2 *i*³Neurons in both DMSO and Baf-treated conditions (Fig. 5.24b). While sh3 was also generally decreased, this difference was only significant for pro-GABARAPL1 in baf-treated cells. The same trends held for HeLa cells treated with siRNAs (Fig. 5.25), and, as with LC3B, L1CAM KO *i*³Neurons did not show any notable changes for any of the GABARAPs compared to WT clones (Fig. 5.26). However, these were not the main focus of this experiment and thus were not performed with full statistical rigor.

It must also be recognized that assigning band identities by size alone was not conclusive, and different cell types can have very different expression patterns, as was previously shown for LC3B, for which HeLa cells primarily contained LC3B-II but *i*³Neurons primarily contained LC3B-I. No distinct pro-GABARAPL1 band was visible in HeLa cells or in the KO/WT *i*³Neuron clones, so this band in shRNA *i*³Neurons may be GABARAPL1-I instead. While band assignment was consistent with previous studies, treatment of the protein lysate with lipases may help to determine the lipidated band more definitively, and additional experiments might be done to inhibit cleavage by the ATG4s to assign the identity of propeptide and ATG8-I bands with more confidence.

Figure 5.25: pro-GABARAPs appear reduced in siRNA HeLas.

- (a) Representative blots of GABARAP (GBRP), GABARAPL1 (GBRPL1), and GABARAPL2 (GBRPL2) in HeLa cells transfected with the indicated siRNA and treated for 4 h with either DMSO (-) or 400 nM baf (+). Identification of bands: propeptide (p), cleaved by ATG4 (I), lipidated (II).
- (b) Quantification of n=1 experiment with 6 separate wells as technical replicates using the indicated band of each protein. Error bars are \pm SD of technical replicates; no statistical tests were performed.

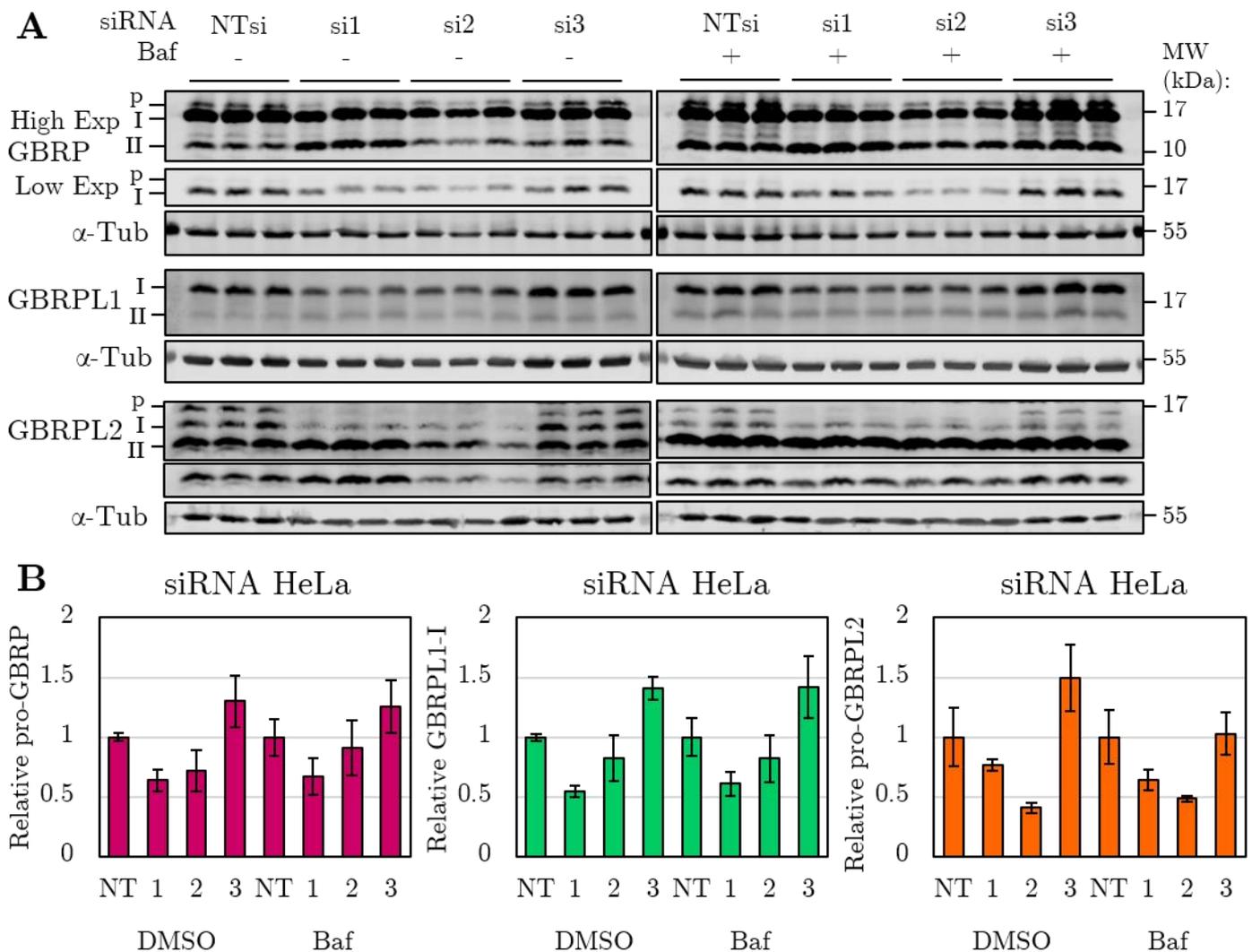
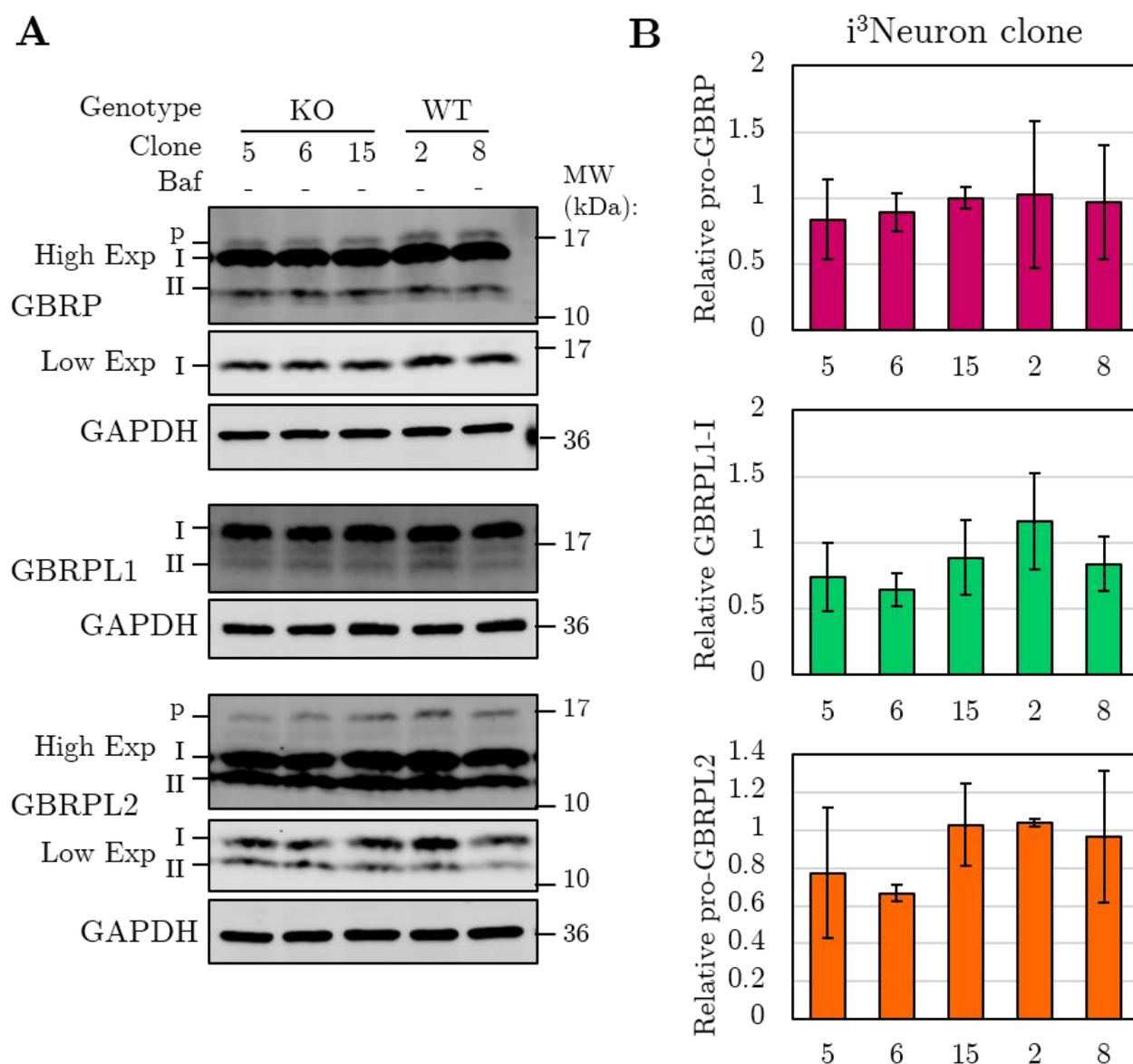


Figure 5.26: pro-GABARAPs appear similar in KO and WT i^3 Neurons.

- (a) Representative blots of GABARAP (GBRP), GABARAPL1 (GBRPL1), and GABARAPL2 (GBRPL2) in the indicated i^3 Neuron clone after 14 days differentiation and treated for 24 h with either DMSO (-) or 400 nM baf (+). Identification of bands: propeptide (p), cleaved by ATG4 (I), lipidated (II).
- (b) Quantification of n=1 differentiation with 2 separate wells as technical replicates using the indicated band of each protein. Values are normalized to the average of the wildtype clones 2 and 8. Error bars are \pm SD of technical replicates; no statistical tests were performed.



5.6.2 Transcriptional analysis of *LICAM* and *ATG8* genes

To support that the observed effects to ATG8 proteins was transcriptional and not due to a difference in the rate of processing or clearance, RT-qPCR was performed on total RNA purified from d14 differentiated i³Neurons for 41 cycles using the primer pairs in Table 2.9 for amplification and SYBR Green for detection. Both *GAPDH* and *ACTB* were included as loading controls, and all reactions were designed to span introns in order to be selective for processed mRNA.

Three reactions targeting the *LICAM* transcript were designed to assess whether different regions of the transcript reacted differently to the shRNAs or KO mutagenesis. It was expected that the KOs would exhibit nonsense-mediated decay resulting in little to no transcript (Baker & Parker, 2004). The forward primer in the N-terminal reaction overlapped the KO cut site, so only cells with an unedited genetic locus were expected to amplify. The Middle reaction overlapped the target site for sh2, and the ICD reaction was entirely within the intracellular domain near the 3' end of the gene. Finally, primers were designed for the five *ATG8* genes that were predicted to be expressed in i³Neurons: *LC3A*, *LC3B*, *GABARAP*, *GABARAPL1*, and *GABARAPL2*. *LC3B2* and *LC3C* were not predicted to be expressed in i³Neurons, so these were not included (Ludwig & Kampmann, 2017).

Three biological replicates were tested of the NTsh line against the three *LICAM*-targeting shRNA lines and of the G3 parental line against the three knockout clones, and each reaction was performed in technical triplicate. Controls were included that lacked any template for each reaction in order to assess purity of the reaction mix and of the plate itself, and no amplification was detected until at least nine cycles after all samples with template had passed the threshold of detection. In addition, cDNA conversion was performed without RT in one of each biological replicate to test for the presence and amplification of genomic DNA. No amplification was detected until at least five cycles after detection of the same sample that underwent cDNA synthesis

with RT. Taken together, these controls show that the role of exogenous sample template and genomic DNA contamination was within commonly accepted bounds and had a negligible effect on the data (Taylor et al., 2019). Additionally, melt curve analysis was performed after completion of the final cycle of qPCR. All reactions had a single peak with a variation of at most 0.5°C between sample maxima, indicating that each reaction was specific for only one amplicon.

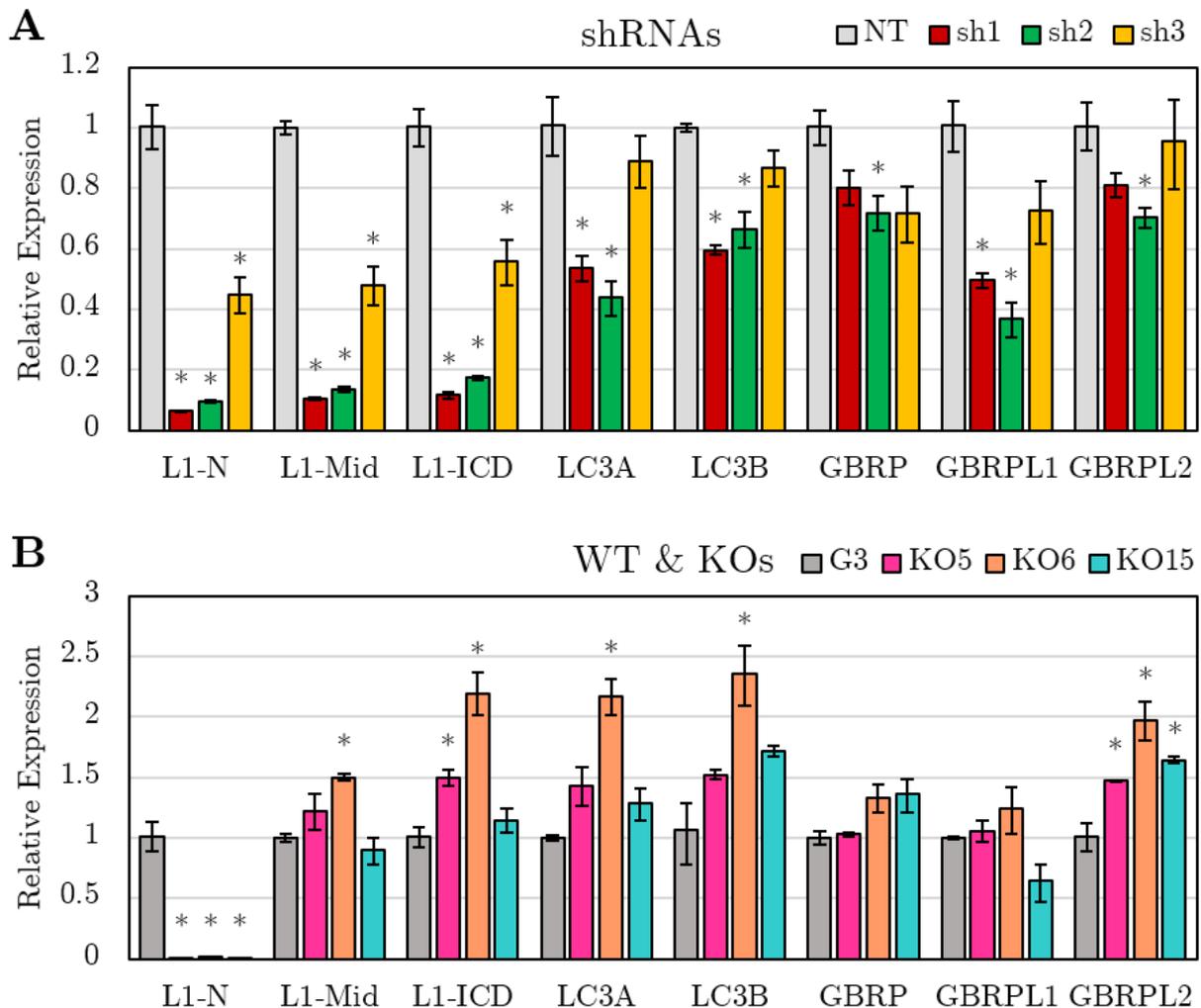
The coefficient of variance for each set of technical triplicates was low (maximum of 5% for all samples, with a mean CV of 1.3%), so the mean value of these technical triplicate reactions was used in subsequent analysis. Normalizing with only *GAPDH* as a loading control resulted in ΔC_T values that highly correlated with the same samples normalized using only *ACTB* (Pearson's $R=0.995$ for shRNA samples and 0.999 for KO samples). These findings increased confidence that variation from reaction efficiency and sample loading was low and that neither reference gene was influenced by cell line conditions. The mean of *GAPDH* and *ACTB* C_T values was thus subtracted from the C_T value of each reaction for the associated sample to calculate the ΔC_T .

Significance of differences was assessed by comparing ΔC_T values with a two-tailed unpaired t-test; shRNA lines were tested against the NTsh control, and KO clones were tested against the G3 parental line. Because each set of samples consists of 24 distinct tests (8 reactions \times 3 experimental lines), a Benjamini-Hochberg false discovery rate correction with a denominator of 24 was applied. Significance testing of the ΔC_T values was more statistically sound than testing the relative expression values directly ($2^{-\Delta\Delta C_T}$; Taylor et al., 2019). It was also strictly more stringent; all significant differences found by testing ΔC_T were also significant by performing the same tests and FDR correction on relative expression, but not the inverse. Next, the mean ΔC_T for the control lines (NTsh or G3) was subtracted from each sample for the associated reaction target to normalize expression relative to the control line as $\Delta\Delta C_T$. Standard error was calculated as the standard error of these $\Delta\Delta C_T$ values, and error bars shown in Fig. 5.27 represent $2^{-(\text{Average}(\Delta\Delta C_T) \pm \text{SEM})}$.

Figure 5.27: RT-qPCR analysis of i³Neurons.

RT-qPCR was performed on 3 independent samples for each line, and 3 technical replicates of each reaction were run. *L1CAM* reactions amplified fragments of exons 2-4 (L1-N), 17-18 (L1-Mid), or 27-29 (L1-ICD). Significance was assessed via two-tailed unpaired t-tests of ΔC_T for each KD or KO line against the associated control for each reaction followed by FDR correction. *GABARAP(/L1/L2)* is abbreviated GBRP(/L1/L2) in chart labels.

- (a) In shRNA i³Neurons, all *L1CAM* transcript reactions were significantly decreased at approximately the same amount, and all *ATG8* genes were significantly decreased in sh2 cells; *LC3A*, *LC3B*, and *GABARAPL1* were also significantly decreased in sh1 cells.
- (b) In KO i³Neurons, the *L1CAM* N-terminal reaction was significantly decreased in all KO clones; however, the Mid and ICD transcript templates were at least as prevalent as in the G3 parental line. *LC3A* and *LC3B* were significantly increased in KO6, and *GABARAPL2* was significantly increased in all KO lines.



5.6.3 shRNA knockdown reduces mRNA of *L1CAM* and *ATG8s*

Data from the shRNA lines is presented in Fig. 5.27a. All three reactions targeting *L1CAM* showed significant decreases for all three shRNAs, with sh1 and sh2 reducing *L1CAM* mRNA by about 85%. However, sh3 was markedly less effective, only reducing *L1CAM* mRNA to about 50% relative to the NTsh line, and this tracked with the less effective knockdown to L1CAM protein seen previously for sh3. These data strongly suggested that total *L1CAM* mRNA was effectively reduced by the shRNAs.

Furthermore, all *ATG8* genes tested were significantly decreased in sh2 relative to the NTsh line, and *LC3A*, *LC3B*, and *GABARAPL1* were also significantly decreased in sh1. While none were significantly decreased for sh3, even these and the nonsignificant values for sh1 were consistently below the NTsh control. It was therefore likely that the shRNAs reduced levels of the ATG8 family of proteins via a reduction in transcription.

5.6.4 L1CAM KO does not induce nonsense-mediated decay

Data from the KO lines is presented in Fig. 5.27b. The N-terminal reaction was significantly decreased for all KO lines; in fact, no amplification was observed even after 41 cycles for 8/9 total technical replicates for both KO5 and KO15. KO6, meanwhile, consistently amplified seven cycles after G3, equivalent to <1% of relative transcript abundance. However, the Middle and ICD reactions demonstrated that the abundance of *L1CAM* mRNA was at least as high as the control in all lines; in fact, KO6 was significantly increased for both reactions and KO5 was significantly increased for the ICD reaction. It was thus clear that, while little to no L1CAM protein was present, mRNA was still produced, and it was not degraded by nonsense-mediated decay as expected. These data also tentatively suggested that L1CAM protein may

reduce its own transcription, since *L1CAM* mRNA increased in some conditions in which the L1CAM protein was not present.

Since both KO5 and KO15 had 8 bp deletions at the mutation site, the lack of any amplification for the N-terminal reaction may be due to a loss of primer affinity despite mRNA being present. KO6 had only a 1 bp deletion, which may explain why KO6 samples amplified this reaction, albeit at a low rate. Additionally, these data were strong evidence that these clones were in fact pure, since any cells expressing wildtype transcript would be expected to amplify more quickly. No significant decreases were seen in any *ATG8* genes; rather, KO6 showed a significant increase in *LC3A* and *LC3B*, and *GABARAPL2* was significantly increased for all KO clones.

5.7 *L1CAM* transcript, not protein, regulates *LC3B*

5.7.1 shRNA expression in KO i³Neurons reduces LC3B protein

The data above established that the targeting shRNAs reduced mRNA levels of both *L1CAM* and the *ATG8*s, while the KO clones contained *L1CAM* mRNA despite lacking L1CAM protein. It was therefore hypothesized that, if the *L1CAM* transcript was mediating the reduction to ATG8 proteins, expression of the shRNAs would be sufficient to reduce LC3B even in the absence of L1CAM protein. To this end, KO clone 5 was transduced with each shRNA, and each line was FACS purified in the same manner as the shRNA lines previously. These lines were collectively termed 5+sh, and each individual line was termed using the same sh name as the shRNA lines previously (i.e., 5+NTsh for the KO5 clone transduced with non-targeting shRNA).

After differentiation and western blotting, it was noted that low-intensity smearing was detected by only the L1CAM D-5 antibody at a molecular weight just below the full-length band in both KO5 and 5+NTsh cells (Fig. 5.28a). This signal was initially

seen in all three KO clones, as noted in Fig. 5.15, but it was discounted as background noise because it was consistently <0.1% of the signal intensity of the full-length band in the unedited parental line. It was also diffuse rather than a distinct band and presented at a different molecular weight than the known full-length protein or cleavage products. The D-5 antibody was typically the most sensitive but least specific of the three antibodies used in this work, resulting in higher signal intensity but also higher levels of background and off-target banding than the other L1CAM antibodies when the same samples were blotted.

With the finding that *L1CAM* mRNA remained expressed in the KO clones, it was recognized that this smearing may be due to alternative translation initiation or splicing that recovered the translation frame. Indeed, while undocumented in the literature, there are in-frame ATG sequences in exon 4 (aa32), exon 5 (aa 124 & 131), and beyond that could theoretically be recognized by ribosomes to start translation downstream of the frameshift deletion. In addition, skipping both exons 3 and 4 would recover the frame for KO5 while preventing early termination, reducing the size of L1CAM by 43 amino acids. The latter strategy would not work for KO6, however, and the observed smearing of multiple protein sizes lends greater credence to the possibility of alternative translation initiation.

This smearing was maintained in 5+NTsh *i*³Neurons, but it was reduced in all three targeting 5+sh lines (Fig. 5.28b). No staining was observed at the full-length band size for any of the 5+sh lines as well. The presence of *L1CAM* mRNA was therefore required for the presence of low-intensity smearing, which implied that the *L1CAM* transcript may have been translated at low levels from alternative initiation sites. However, the intensity of this staining relative to the G3 line full-length L1CAM band was lower in the 5+sh lines than in the lines expressing *L1CAM*-targeting shRNAs alone, so this finding did not imply that knockout of the L1CAM protein was ineffective, nor did it suggest that a reduction in L1CAM smearing could on its own be responsible for reinstating a reduction of LC3B. Rather, it is notable that the *L1CAM* transcript was

both resilient to nonsense-mediated decay and possibly capable of alternative translation despite a frameshift in the first coding exon.

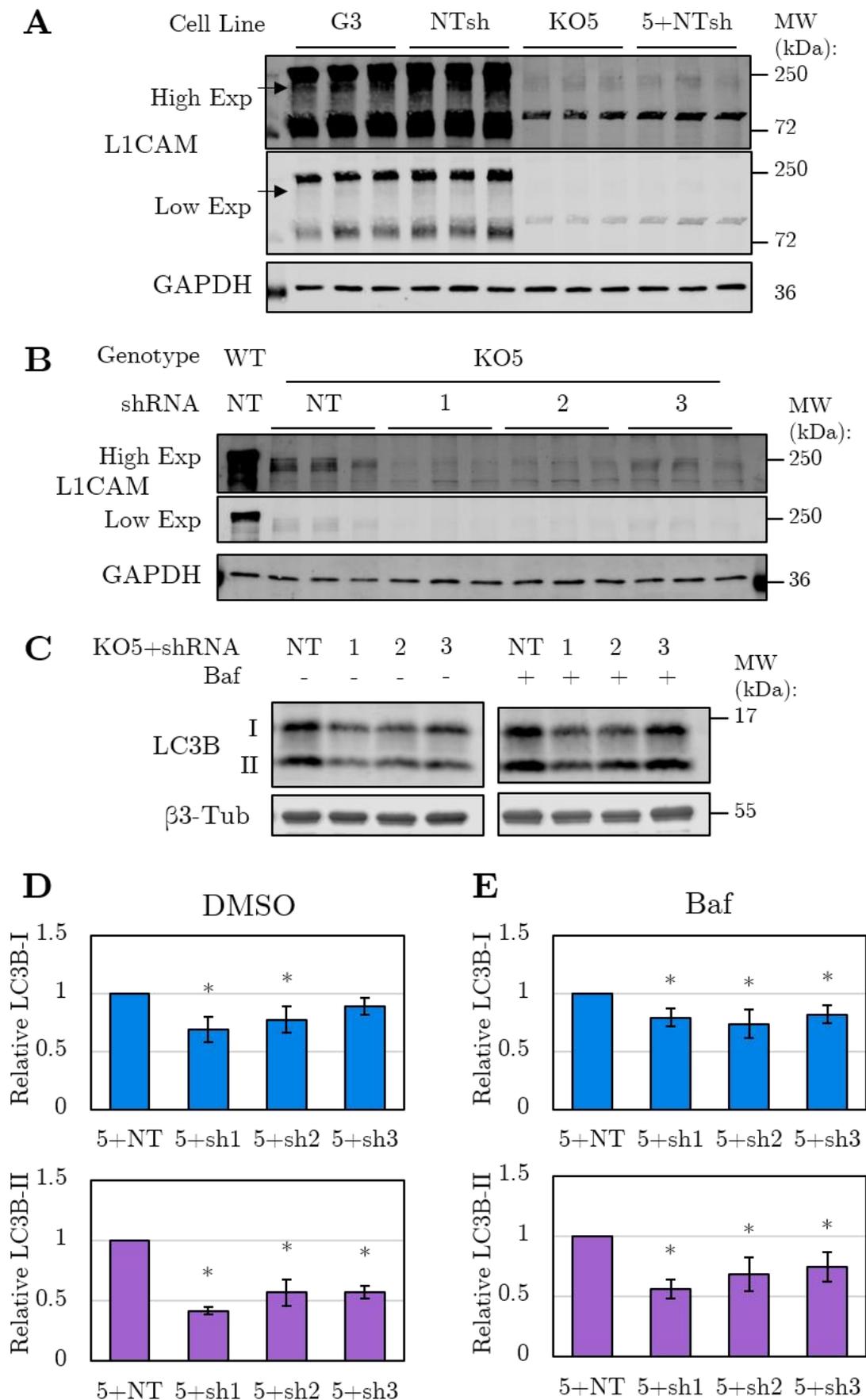
Furthermore, LC3B-I and LC3B-II were significantly decreased in *L1CAM*-targeting 5+sh lines compared to 5+NTsh after both DMSO and Baf treatment (Fig. 5.28c-e). Only DMSO-treated 5+sh3 cells did not reach the threshold of significance for LC3B-I. Since the shRNAs were shown to reduce *L1CAM* mRNA previously, the *L1CAM* transcript that was expressed in KO5 cells was presumably reduced in the *L1CAM*-targeting 5+sh lines, and these shRNAs caused a reduction of LC3B despite little to no L1CAM protein being present in any of the 5+sh lines.

Figure 5.28: LC3B is reduced in KO clone 5+shRNAs.

- (a) A dim smear at a molecular weight just below the full-length band was noted in KO5 and 5+NTsh cells (black arrow). Also evident was a nonspecific band at 90 kDa and cleavage product IV at 80 kDa. The images shown are extensions of the blot shown in Fig. 5.11c; the D-5 antibody was used.
- (b) Transduction & FACS of KO5 with *L1CAM*-targeting shRNAs reduced this smearing as shown on overexposed L1CAM blots. The D-5 antibody was used.
- (c) Representative blots of LC3B-I and LC3B-II in differentiated KO5+shRNA i³Neurons after 24 h treatment with DMSO (-) or 400 nM baf (+).
- (d-e) Quantification of n=3 independent differentiations, each with 3 separate wells as technical replicates, after treatment with DMSO (d) or baf (e). LC3B-I in DMSO-treated 5+sh3 cells was the only condition not significantly reduced.

(continued)

Figure 5.28 (*continued*): LC3B is reduced in KO clone 5+shRNAs.



5.7.2 L1CAM overexpression does not rescue LC3B

If L1CAM protein was the primary agent regulating LC3B and, by extension, the other ATG8s, then it would stand to reason that a return of L1CAM protein to levels similar to basal conditions would rescue LC3B as well. Conversely, if *L1CAM* transcript was instead responsible for regulating *LC3B* transcription as hypothesized, then no change to LC3B should be expected no matter the amount of L1CAM protein present.

To test these possibilities, HeLa cells were transfected with either a constitutive expression plasmid containing full-length *L1CAM* cDNA (L1-OE) or an empty vector control, along with either the NTsi pool or *L1CAM*-targeting siRNA 2. As shown previously, si2 was the siRNA that most effectively reduced L1CAM along with both LC3B-I and LC3B-II, so si2 was used because a restoration of these levels (or a failure to do so) would be the most evident. For clarity throughout the rest of this section, cell conditions are referred first by the siRNA with which they were transfected (either NTsi or si2), followed by a slash and the plasmid that they received (- for empty or + for L1-OE). For example, cells that were transfected with the NTsi pool and the L1-OE plasmid are “NTsi/+”.

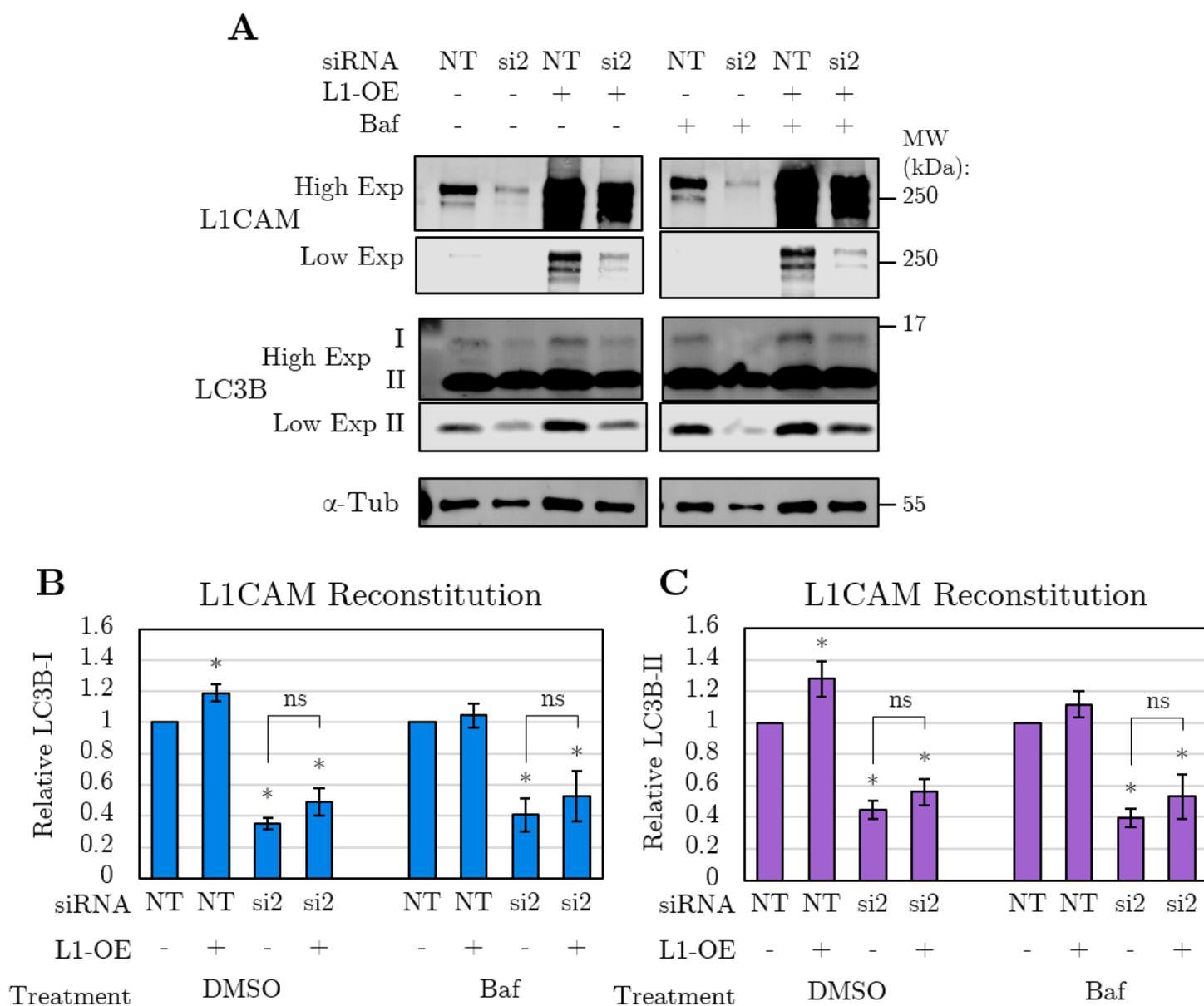
After western blotting, NTsi/+ cells showed increased L1CAM compared to NTsi/- cells; band intensity was typically >50-fold higher for NTsi/+ (Fig. 5.29a). Meanwhile, si2/- cells showed comparable reduction to L1CAM as without plasmid transfection, and si2/+ cells demonstrated consistent reconstitution of L1CAM protein to levels at or above control NTsi/- cells. In addition, L1-OE conditions appeared to show a slight increase to both LC3B-I and LC3B-II compared to their associated empty vector counterparts. However, both forms of LC3B were still reduced in si2/+ compared to NTsi/- cells despite high relative levels of L1CAM protein.

Figure 5.29: LC3B is not rescued by L1CAM reconstitution.

HeLa cells were transiently transfected with the indicated siRNA and either an empty vector (-) or a plasmid encoding the full-length L1CAM cDNA driven by a constitutive CAG promoter (L1-OE; +).

(a) After transfection and treatment, cells that received the L1-OE plasmid exhibited consistently strong L1CAM staining but little change to LC3B.

(b-c) Quantification of n=4 independent experiments, each with 3 separate wells as technical replicates, for LC3B-I (b) and LC3B-II (c). Significance was assessed by two-tailed paired t-tests. Asterisks denote significant changes from the associated NT/- condition; while si2/+ cells were consistently above si2/- cells, they were not significantly changed from each other (indicated by “ns”).



Indeed, these trends were found to replicate across four experiments, each with independent transfection of all components. L1CAM was consistently reconstituted to at least basal levels in si2/+, but despite this, neither LC3B-I nor LC3B-II protein levels were significantly changed in si2/+ compared to si2/-. While both forms of LC3B were typically higher in si2/+ than in si2/-, this change was too slight and too variable to reach the threshold of significance, and even accounting for variability in the effects of si2 by re-normalizing with si2/- did not bring any changes to si2/+ into statistical significance. Additionally, regardless of plasmid transfection, si2 conditions were all significantly decreased compared to the NTsi/- control (Fig. 5.29b-c).

While LC3B was significantly increased in the DMSO condition for NTsi/+ compared to NTsi/-, this difference did not extend to baf-treated cells (Fig. 5.29b-c). This may suggest that L1CAM overexpression acted as a blocker of autophagic degradation but did not directly induce the expression of LC3B (Streeter et al., 2016). It is possible that this reduction in clearance was caused indirectly simply due to the large amount of protein produced rather than by any direct regulatory mechanisms.

5.7.3 L1CAM overexpression does not significantly affect LC3B

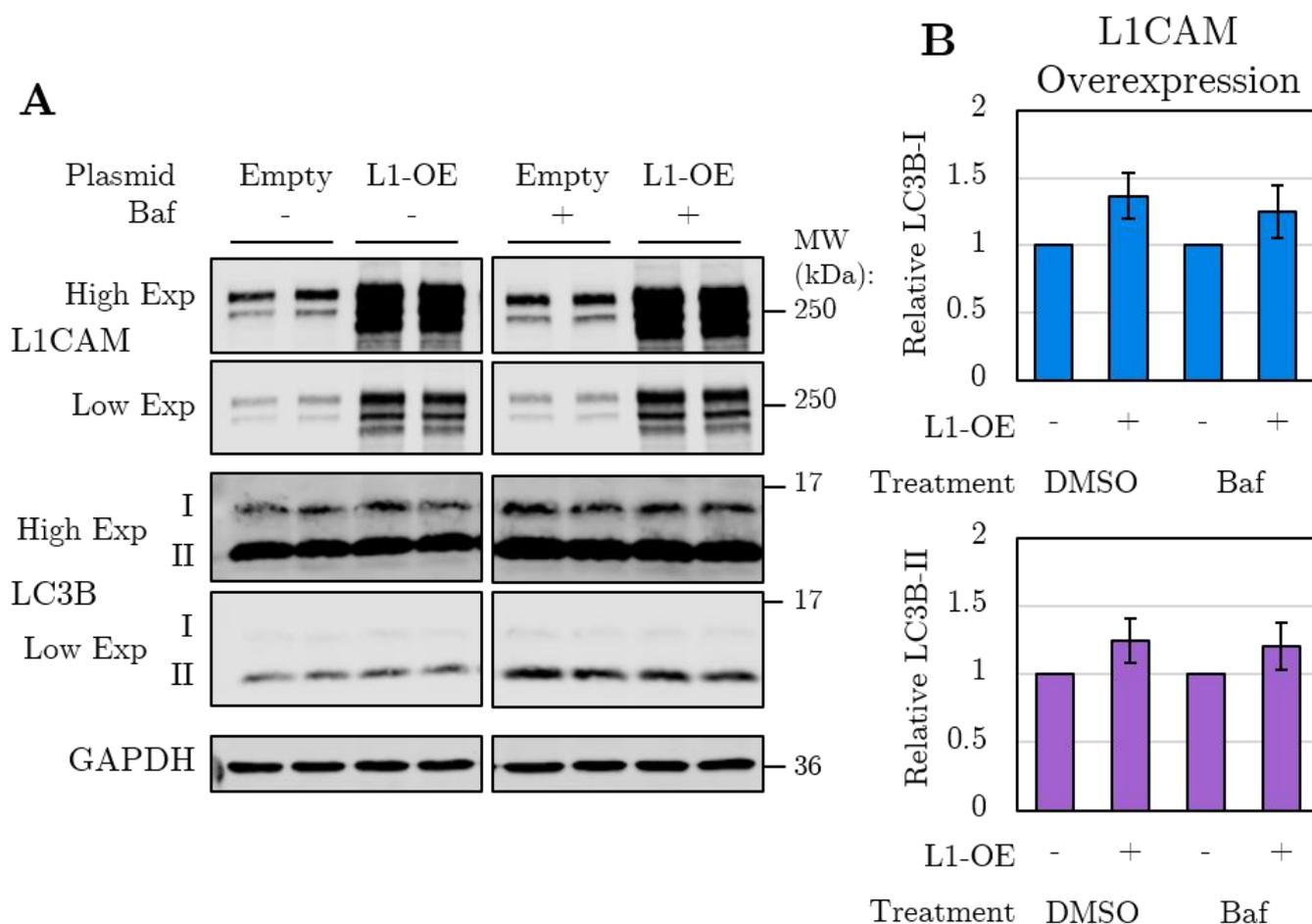
To assess the capacity of L1CAM to affect LC3B expression without the extraneous influence of siRNAs, HeLa cells were transfected only with the empty vector control or the L1-OE plasmid. As observed in the NTsi/+ cells previously, the level of L1CAM protein was increased by an average of 50-fold in cells transfected with L1-OE compared to the empty vector. While the observed band density for both LC3B-I and LC3B-II tended to be slightly higher in cells transfected with L1-OE compared to the empty vector, no changes were statistically significant (Fig. 5.30).

Figure 5.30: LC3B is not affected by L1CAM overexpression.

HeLa cells were transiently transfected with either an empty vector (-) or the L1-OE plasmid (+).

(a) After transfection and treatment, cells that received the L1-OE plasmid exhibited consistently strong L1CAM staining but little change to LC3B.

(b) Quantification of n=4 independent experiments as in (a), each with 3 separate wells as technical replicates. While LC3B-I and LC3B-II were generally higher in L1-OE transfection, no conditions were significantly changed from the empty vector.



Reconstitution of L1CAM by plasmid overexpression was thus unable to protect LC3B from reduction, despite L1CAM protein levels exceeding basal conditions. In addition, overexpressing L1CAM on its own did not significantly affect either form of LC3B. These findings corresponded with additional L1CAM overexpression experiments performed by So Yeong Cheon; however, the *L1CAM* cDNA expressed in those experiments lacked exon 28, so these data were not used here.

On the other hand, both L1-OE and the siRNAs were transiently expressed, so variability in the efficiency of either transfection may have confounded results. These data may be further supported with immunocytochemistry of both L1CAM and LC3B in a cell-by-cell manner. If these findings were accurate, the relative intensities of both proteins should positively correlate in the empty vector conditions, but no correlation should occur in the L1CAM overexpression conditions. It would thus be worthwhile to assess transfection efficiency in additional replicates.

Perhaps most convincingly, LC3B was reduced in the KO i³Neurons only after transduction with *L1CAM*-targeting shRNAs. Since KO clones were shown to express *L1CAM* mRNA but not protein, and the shRNAs were shown to reduce mRNA effectively, this provided compelling evidence against the centrality of the L1CAM protein as a regulator of macroautophagy. Instead, the hypothesis that the *L1CAM* transcript itself is the regulator of *ATG8* transcription was supported.

Thus, reduction of *L1CAM* transcript consistently reduced macroautophagy induction even in the absence of L1CAM protein. Since the reconstitution experiments involved transcription of the cDNA, the data also imply, but are not robust enough to demonstrate, that the autophagy-regulating element may lie in a noncoding sequence (e.g., an intron), as explored below.

5.8 Discussion

5.8.1 Summary of findings

The data presented here help to establish *L1CAM* as a novel regulator of autophagy. Reduction of L1CAM by siRNAs in HeLa cells was shown to impair clearance of mutant proteins in an ATG7 but not ATG16L1-dependent manner, to reduce LC3B-I and LC3B-II protein levels, and not to affect proteasomal degradation.

Reduction of L1CAM by shRNAs in i³Neurons was then shown to reduce LC3B-I and LC3B-II protein levels and moreover to reduce transcription of the entire *ATG8* gene family. Critically, these effects were shown to be dependent on a reduction of the *L1CAM* transcript but not on a reduction of the L1CAM protein.

Since L1CAM protein interacts with many signaling pathways that feed into known macroautophagy transcriptional regulatory pathways (e.g., AKT, MAPK, JNK, NF κ B), it was surprising that the protein was dispensable. The remainder of this chapter delineates outstanding questions, proposes avenues for further study, and explores additional implications of the data.

5.8.2 Key outstanding questions

This work revealed that the *L1CAM* transcript, when it is expressed, regulates transcription of the *ATG8*s; in order for this to be the case, some molecular mechanism must exist that connects the *L1CAM* transcript with the *ATG8* genes. Such a mechanism must also account for the finding that conditions in which *L1CAM* is not expressed—as in iPSCs—the *ATG8*s are still transcribed normally, but when epigenetic conditions are such that *L1CAM* should be expressed—as in differentiated i³Neurons and HeLa cells—a subsequent loss of *L1CAM* mRNA is sufficient to reduce *ATG8*

transcription. Critical to proving the existence of and understanding this mechanism will be to identify the specific RNA regulatory element involved as well as to identify any cofactors affected by the loss or mutation of this element.

5.8.2.1 What is the identity of the *ATG8*-regulating element?

The *L1CAM* transcript was shown to influence *ATG8* transcription, but this work did not determine the specific identity of the element responsible for regulating the *ATG8*s. It is unlikely that an independently-transcribed long non-coding RNA (lncRNA) was responsible, since no known transcripts other than *L1CAM* itself overlap the target sites of any of the siRNAs or shRNAs used in this work (Ponting et al., 2009). However, the multiplicity of *L1CAM* spliceforms results in numerous known retained introns and transcripts with poorly defined ORFs, suggesting there is more to *L1CAM* transcription than is currently understood. The major possibilities for this element are thus *L1CAM* mRNA, pre-mRNA, and a retained intronic sequence that is released after pre-mRNA processing. Any of these may act in a similar manner to a lncRNA to influence transcription of other genes (Statello et al., 2021). While no direct sequence homology exists between *L1CAM* and any of the *ATG8* genes, stronger computational tools may be used to predict secondary structures and complementarity which could help predict the region of the transcript critical for its regulatory function.

The specific targets of the siRNAs and shRNAs used here may provide additional clues. While si1, si2, sh1, and sh2 all target individual exons, si3 and sh3 both target the junction of exons 20-21. Nuclear RNAi activity has been observed, including translocation of AGO2 loaded with RNA and additional components of the RISC, suggesting that the siRNAs and shRNAs used here may be loaded cytoplasmically and imported to the nucleus (Gagnon et al., 2014). Thus, si1, si2, sh1, and sh2 are each capable of binding to and mediating the degradation of pre-mRNA, while sh3 and si3 can only recognize mRNA after intron 20 has been spliced out.

If the processed mRNA is not the *ATG8*-regulating element, degradation of the *L1CAM* mRNA only after splicing would not be expected to affect the *ATG8*s. This was supported by findings that the sh3 line featured a less potent knockdown of both L1CAM protein and mRNA compared to sh1 and sh2; likewise, sh1 and sh2 were more effective at reducing *ATG8* transcription. Further, si3-treated HeLa cells showed a reduction of L1CAM but lacked an effect on LC3B. Meanwhile, the degradation of nascent pre-mRNA by si1, si2, sh1, or sh2 would be expected to reduce transcription of the *ATG8*s regardless of whether the *ATG8*-regulating element is the pre-mRNA as a whole or one of the introns that is only active after splicing.

Because the L1-OE plasmid expressed *L1CAM* cDNA, it was transcribed as processed mRNA directly without introns or splicing. The inability of the mRNA alone to induce LC3B expression may further suggest that the regulatory component of the *L1CAM* transcript is not the processed mRNA. However, there also may be a self-limiting mechanism to prevent LC3B expression from exceeding basal levels, consistent with a consistently slight but nonsignificant increase to LC3B after L1-OE transfection. In addition, si2 was able to bind and degrade the L1-OE transcript as it was being produced, so the L1CAM reconstitution experiment is not able to specifically assess the role of the mRNA. Rather, it mainly provides support for the hypothesis that L1CAM protein is not directly involved in LC3B regulation, while qPCR could be done in the same conditions to test relative *L1CAM* mRNA levels.

Future work may seek to clarify the identity of the suspected regulatory element. The processed mRNA may be shown to be dispensable by mutagenizing *L1CAM* cDNA such that it is insensitive to siRNA degradation and transfecting it alongside the siRNA. If LC3B expression is still not restored (or, better, if immunocytochemistry is used to show a correlation between L1CAM and LC3B without mutagenized cDNA, but no correlation after cDNA transfection) it is likely that pre-mRNA plays a greater role. Then, expressing fragments of the L1CAM genetic sequence may narrow down the location of a regulator.

5.8.2.2 How is macroautophagy affected?

Most RNA transcriptional regulatory elements mediate their effects via protein cofactors, and in this case the transcript may either support activation of a macroautophagy-inducing factor or inhibit activity of a macroautophagy-repressing factor. Pathways that influence transcription of multiple *ATG8s* were therefore specifically delineated in Chapter 1. Since each is known to regulate transcription of multiple *ATG8s*, the most likely inducers are the TFE and FOXO families, PPAR α , and CREB, while the most likely repressors are ZKSCAN3 and FXR.

Preliminary data from additional experiments did not show a change to TFEB activity or localization as a result of *L1CAM* knockdown in HeLa cells, but this may be assessed in more detail. In i³Neurons, *TFE3* and *FOXO3* are the most highly-expressed members of each family, while *TFEB*, *TFEC*, and *FOXO1* are not expressed (Ludwig & Kampmann, 2017). *CREB1*, *PPARA*, and *ZKSCAN3* were also detected, but *NR1H4* (aka *FXR*) was not expressed. The most likely candidates for *L1CAM* transcript interaction are thus *TFE3*, *FOXO3*, *CREB1*, *PPARA*, and *ZKSCAN3*. Future work may seek to assess the activity of and effect of modulating these transcription factors in *L1CAM* knockdown conditions.

An interesting finding was that *L1CAM*-targeting siRNAs did not impair clearance in ATG7 KO cells (particularly for si2), but they did impair clearance in ATG16L1 KO cells. Since ATG7 primes all ATG8s for conjugation to PE, a reduction of *L1CAM* leading to a reduction of ATG8s may act on a similar level to a loss of ATG7, so a combination of the two may not increase this effect further. The specific roles of the other ATG8s in human cells have not been fully explored, and the impairment of clearance in ATG16L1 KO cells after siRNA treatment may therefore point to the involvement of the ATG8s in alternative clearance pathways that are also affected by *L1CAM* reduction and ATG7 KO but not by a loss of ATG16L1. This may also imply

that *L1CAM* reduction affects other components of macroautophagy not explored in this study.

Decreased availability of the ATG8s is highly disruptive to proteostasis, and these proteins are important for phagocytosis, degradative endocytosis, and vesicular trafficking beyond autophagosomes, as noted in Chapter 1. Thus, a reduction to ATG8 protein production could impair these processes in addition to macroautophagy. A recent study that published RNAseq data from mouse astrocytes and microglia reported very little expression of *L1cam* in either cell type (Pan et al., 2020), so neurons are likely the main cell type affected according to the proposed mechanism of action.

For a wider perspective, RNAseq may be used to separate the effects of reduction of the L1CAM protein and transcript (i.e., changes compared to basal conditions that occur in knockdown but not KO cells) and to identify affected transcriptional pathways. This would also provide evidence of whether transcription of other autophagy-related genes is affected, as preliminary data from additional experiments suggest (i.e., components of the PI3K complex I). Furthermore, specific changes to transcription factor binding of the *ATG8* promoters as a result of *L1CAM* reduction may be identified with proteomics of isolated chromatin segments (Kan et al., 2017).

5.8.2.3 A proposed mechanism of action

As established above, the most plausible regulatory element is an intronic sequence of the *L1CAM* transcript, and there are several transcription factors that could be cofactors of this element to enact the effects demonstrated in this work. A possible molecular mechanism to explain the data presented in this chapter was therefore hypothesized and diagrammed (Fig. 5.31).

In this model, the regulatory element of the *L1CAM* transcript (most likely a retained intron) normally binds with cofactors as a scaffold to promote transcription

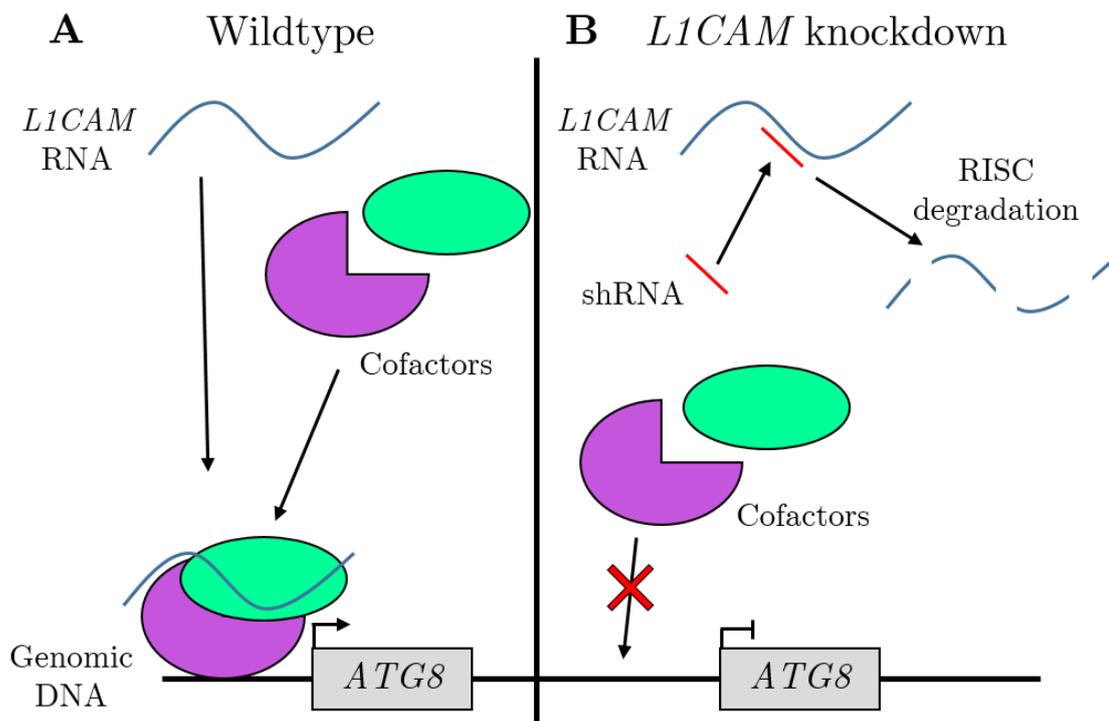
of the *ATG8* genes (Fig. 5.31a). This accounts for the lack of homology between the *L1CAM* transcript and the *ATG8* genes, which suggests that regulation does not occur via direct RNA-mediated guiding of such factors to the promoter, as a decoy for an *ATG8* transcriptional repressor, or via lncRNA-like binding to *ATG8* pre-mRNA to influence splicing. While it is also possible that the *L1CAM* transcript influences chromatin remodeling, the ability of iPSCs to express the *ATG8*s without *L1CAM* rather implies that the chromatin remains open regardless and that iPSCs instead have different transcriptional regulation of the *ATG8*s.

In neurons and HeLa cells, then, the expression of *L1CAM* enables this transcriptional complex to take over, resulting in the *ATG8*s to become dependent on *L1CAM*. Thus, when the *L1CAM* transcript is degraded, the transcriptional cofactors are unable to bind, and *ATG8* transcription is impaired (Fig. 5.31b). Even if one of the cofactors is able to bind the *ATG8* promoter, lacking the scaffold and additional cofactors that form the complex would prevent transcription initiation. While alternative transcriptional pathways exist, as evidenced by the presence of any LC3B despite *L1CAM* knockdown and in iPSCs, this *L1CAM*-containing complex appears to be dominant in the cell types shown in this work. It may be investigated whether ectopically expressing *L1CAM* in iPSCs similarly makes the cells dependent on the continued expression of *L1CAM* for maintaining *ATG8* expression or whether there is a secondary mechanism by which the cell type determines the supremacy of different transcriptional programs.

This model may be supported and refined by performing the experiments proposed in the preceding subsections; namely, to determine the sequence and structure of the exact regulatory element in the *L1CAM* transcript and to determine its binding partners. Iterative shortening and mutation of the transcript could effectively narrow down the former, while co-immunoprecipitation is the most straightforward approach for the latter.

Figure 5.31: Hypothesized mechanism of *ATG8* transcriptional regulation.

- (a) In the healthy wildtype state, the *L1CAM* transcript is present and enables cofactors to form a complex at the *ATG8* promoter to drive transcription.
- (b) When the *L1CAM* transcript is absent (i.e., due to RISC degradation), cofactor binding is blocked, leading to decreased transcription of the *ATG8* genes.



5.8.3 Implications of *ATG8* reduction in disease

As noted, L1CAM acts to promote axonal outgrowth and neuronal migration in the developing brain. While the L1CAM protein is thought to principally mediate these functions via direct adherence and signal transduction, if a mutation impairs *L1CAM* transcription and thus reduces the ATG8s *in vivo*, the disease manifestations of patients with L1 syndrome may be partially caused by a reduction of neuronal macroautophagy. Variable effects of different L1CAM mutations on macroautophagy may account for some of the variability in clinical manifestations and may suggest a wider combination of insults to neuronal health underlying L1CAM-linked diseases

than previously appreciated. Impairment of neuronal autophagy has been linked to other forms of neurodevelopmental diseases, including other forms of hereditary spastic paraplegia that have features similar to L1 syndrome (Marchesi et al., 2021), so it would be useful to directly assess whether neuronal macroautophagy is affected in L1 syndrome patients.

Data presented here also suggest that impairment of *L1CAM* expression may predispose neurons to a buildup of aggregating proteins, which has been associated with the pathogenesis of numerous late-onset neurodegenerative diseases as described in Chapter 1. If *L1CAM* expression in neurons is reduced by aging, this may sensitize neurons to degeneration, and a restoration of *L1CAM* transcription may present an opportunity to restore autophagic function. *L1CAM* silent mutations may also serve as risk factors for other neurodegenerative diseases.

As noted in the context of cancer, aberrant activation of *L1CAM* transcription plays a key role in metastasis, and tumor growth in mice was reduced following immunotherapy using anti-L1CAM antibodies as well as treatment with *L1CAM*-targeting siRNAs and shRNAs (Ganesh et al., 2020). Data presented here suggest that therapies that target the *L1CAM* transcript may additionally decrease macroautophagy, an effect that has not been previously noted. Autophagy inhibition has been explored directly as a potential cancer treatment; most notably, blocking lysosomal degradation with chloroquine sensitizes tumors to chemo- and radiotherapies (Kocaturk et al., 2019). However, the effects of autophagy inhibition in cancer are highly context-dependent, with some reports suggesting that autophagy is critical for early immunogenic recognition and apoptosis of cancer cells (Pérez-Hernández et al., 2019). This suggests that autophagy inhibition is most beneficial in late-stage cancers, which also corresponds with *L1CAM* activation. Clarification of the mechanism(s) linking *L1CAM* to the regulation of autophagy would therefore enhance the understanding of a variety of disease conditions and may present novel opportunities for therapy.

Chapter 6: Discussion

6.1 Overall summary & future directions

In this work, I aimed to advance the use of iPSC-derived neuron models of neurodegenerative diseases through methodological improvement and biological discovery, and despite numerous challenges, I had success in achieving this goal. Much of the implications of the data and potential future directions were described in the discussion sections of each chapter, so this chapter will focus on connecting these by summarizing key findings and discussing overall strengths and weaknesses of iPSC-derived neurons.

In the work described in Chapter 3, I used an analytical approach to develop, optimize, and assess the efficacy of SBP- Δ LNGFR as a positive and negative selection gene for iPSCs, building on the foundational work by Matheson et al. (2014). SBP- Δ LNGFR was useful for transgenic cell selection, and its utility should generalize to arbitrary transgene insertions. I hope that the refinement and characterization described here, along with the availability of Mag2-hNIL on Addgene, will lead to broader adoption of SBP- Δ LNGFR, and the polycistronic interference described in Chapter 4 was an important caveat to inform future plasmid design. Future work could also seek to generalize the analyses developed in this work, including binomial significance testing, distinguishing between frequentist and Bayesian methods of calculating test characteristics, and modelling with the beta distribution.

Initially, the aim of the work described in Chapter 4 was to establish the two isogenic series of i³LMNs in order to investigate transcriptional effects of the AR polyQ expansion in SBMA. While I did not accomplish this aim, the challenges I faced provided insights into the regulation of AR and the general weaknesses of iPSC-derived neurons. Like all model systems, these cells do not perfectly recapitulate human

physiology, and this project would have benefited had I been more skeptical and performed earlier proof-of-concept experiments. Along with others who wished to predict gene expression in these cells for their own studies, my feedback about AR silencing was crucial in developing the iNeuron RNA-Seq web app (Michael Ward, personal communication, 29 June 2021; Ludwig & Kampmann, 2017). This app is freely available online (at: https://ineuronrnaseq.shinyapps.io/rnaseq_app/) and provides gene expression data from wildtype iPSCs and i³Neurons after 14, 21, 28, and 35 days of differentiation using either Neurobasal or BrainPhys media. I used this app extensively to plan experiments described in Chapter 5 and to advise colleagues about whether particular genes of interest are predicted to be expressed (or not) in i³Neurons. In this way, I hope that the challenges I faced in this work can help to prevent others from making the same mistakes.

While the data presented here makes it unlikely that these lines can be used for full neuronal differentiation, I remain hopeful that collaborators with whom these lines have been shared will be able to successfully differentiate skeletal muscle, which would enable transcriptomic and proteomic analyses similar to those planned for this project, followed by the identification of possible disease-modifying targets. In addition, Pourshafie et al. (2020) show that iPSCs and partially-differentiated neurons can also yield valuable biological insight, and these lines may be used to provide more granularity and additional controls.

Finally, in Chapter 5, I demonstrated the strengths of iPSC-derived neurons as a model system for hypothesis-testing research. Because *L1CAM* was highly expressed in i³Neurons, manipulations to the gene and transcript could be performed robustly, and the macroautophagy deficiency as a result of *L1CAM* knockdown identified in HeLa cells could be replicated with greater consistency. Differences between the effects of genetic knockout and shRNA-mediated knockdown led to the identification of a possible mechanism of action in which the *L1CAM* transcript or a retained intron acts as a novel noncoding RNA regulator of the transcription of multiple core

macroautophagy genes in neurons and cancer. It is therefore crucial for this work to be carried forward that the questions outlined in Chapter 5.8 be addressed in order to provide additional support for the mechanism as a whole and to provide clarity as to the identity of the *L1CAM*-related regulatory element, its cofactors, and the extent of its effects.

6.2 Limitations of iPSC-derived neuron models

6.2.1 Model line derivation requires high up-front investment

The derivation of the SBMA model lines was ambitious, with a dual cut and insertion without an integrated selection gene for six different lines. Even after accounting for transfection efficiency with FACS, an overall integration efficiency of 3.3% was observed (19 of 571 insert clones genotyped, Fig. 4.1). Further integration of the hNIL inserts was done after the base line derivation in order to preserve lines with only an *AR* edit, but this in turn required integration, cloning, and validation of ten distinct lines.

Gene editing at specific loci is preferable to undirected insertion of lentiviral vectors for the establishment of model lines, but this required a considerable investment of time and effort to properly clone and validate these 36 edited lines (six SBMA insertion lines, ten Mag-hNIL insertions, ten Cre excisions, and ten Mag2-hNIL insertions). The hNIL insertions and Cre excisions were facilitated by the use of SBP- Δ LNGFR for enrichment and depletion, respectively, so these edits were performed in parallel, but it would have been more efficient to test the differentiability of one line first.

6.2.2 Unexpected gene expression patterns

Once the issue of differentiability was addressed in Chapter 4, the obstacle of *AR* silencing posed the next challenge. As discussed, *in vivo* data suggests that *AR* should be expressed at least at low levels in LMNs, so the lack of expression in i³LMNs highlights the possibility for unexpected gene expression patterns to arise during differentiation. On the other hand, understanding the mechanisms underlying *AR* silencing may be useful for identifying therapeutic targets to modulate *AR* expression, which could be beneficial for SBMA and prostate cancer.

In general, transcription factor-mediated differentiation provides rapid, scalable, and pure cultures at the expense of skipping intermediate cell states which may reduce their likeness to physiological cell types. Conversely, chemical differentiation mirrors normal developmental signals, which may permit stepwise changes to gene expression that more accurately reflect physiological cells, but these cultures are limited in scale and more often feature cell type heterogeneity.

Both of these strategies have their origins in developmental biology, so a better understanding of how gene expression is affected by differentiation (both *in vivo* and *in vitro*) may lead to improved techniques, and it is important to recognize that iPSC-derived neurons are not the same as cells that differentiate during normal development, so there is no certainty that any particular gene of interest will be expressed after a cell state transition. Differentiation can cause widespread epigenetic changes, so good culture practices including careful optimization and standardization of protocols are vital for reproducibility.

Tools to predict expression of specific genes (e.g., iNeuron RNA-Seq and myogenesis.net) greatly facilitate experimental planning, but they are more beneficial for hypothesis-testing rather than discovery-based research. A major advantage to transcription factor-mediated differentiation is that high cell type purity facilitates genome-wide screens and multi-omics studies, but the possibility for false negatives

and false positives to arise due to epigenetic differences between these cells and cell types *in vivo* must be recognized. Before commencing a study, it is therefore important to assess whether the potential rewards are worth the risk of unforeseen complications, and any findings should be validated in other cell lines, primary neuron cultures, or animal models.

6.2.3 Difficulties manipulating and maintaining differentiated neurons

While exogenous DNA can generally be introduced into undifferentiated iPSCs, differentiated neurons typically have poor transfection efficiency and may not survive lentiviral transduction. This weakness was not directly addressed in this work, but I was unable to assess clearance of eGFP-A53T or eGFP-HTT74 in i³Neurons in Chapter 5 for this reason. In addition, iPSC-derived neurons may react differently to drug treatments than mitotic cells, as was shown with 24 h baf treatment that would be expected to cause widespread death in HeLa cells. On the other hand, iPSC-derived neurons are more sensitive in other respects, and a high degree of cell debris is common in both transcription factor-mediated and chemical differentiation. Improved substrates and media conditions may help to maintain cell health and promote maturation in extended culture (Kamaraj et al., 2020).

In addition, adherent neuronal culture platforms promote cell immobility, but these substrates are far harder than *in vivo* conditions, which can further influence gene expression (Macrì-Pellizzeri et al., 2015). Differentiated cells are also easily detached from the plate, which makes full media changes and the washes necessary for immunocytochemistry more challenging. Entire wells can slough off as a sheet of cells, so improved culture substrates would be beneficial.

Finally, both iPSCs and differentiated neurons are cultured in media with recombinant growth factors. This has the advantage of being chemically-defined and thus more reproducible, but these reagents can be prohibitively expensive and degrade quickly, possibly leading to variation. Additionally, the often finicky and lengthy techniques required for iPSC and differentiated neuron culture present a barrier to entry. My coauthors and I attempted to address this issue by publishing detailed, accessible protocols for each step of the process (Fernandopulle, Prestil et al., 2018).

6.3 Advantages of iPSC-derived neuron models

6.3.1 Genetic stability produces better controls

In cell models with unstable genomes such as HeLa cells, mosaicism within a population complicates the direct comparison of clonal lines. In addition, controls for transient manipulations must arise from the same passage of cells, and every step from a base wildtype state must be repeated for each replicate. This is because the genetic makeup of the culture can change within a single passage since each passage acts as a genetic bottleneck.

While not perfect, the genome of iPSCs is far more stable when cultured correctly. Single nucleotide variations can arise from iPSC reprogramming, but only a handful of these were observed in coding regions even after prolonged culture (Cheng et al., 2012). Larger-scale chromosomal variations are also rare (Popp et al., 2018). While usually presented in the literature as a concern for the use of iPSCs in cell therapies, these rates of mutagenesis are small compared with tumor-derived cell lines. In this study, I PCR amplified and Sanger sequenced across an intron of *L1CAM*; in HeLa cells, the exons sequenced as expected, but the trace of the intron was too variable to be

interpreted, while iPSCs had no such issue. Thus, iPSC control conditions can be more accurately compared with those subjected to experimental manipulations.

6.3.2 Enhanced reproducibility

Hypotheses in Chapter 5 were initially informed and tested by experiments performed in HeLa cells. As was shown, several siRNAs were effective at reducing L1CAM, but each affected LC3B in different ways. Therefore, it could not be determined whether the observed macroautophagy impairment was due to a reduction of L1CAM or to off-target effects of the siRNAs. The development and use of shRNA and L1CAM KO i³Neurons was crucial to resolving this issue.

In i³Neurons, LC3B turnover was slow, and LC3B-I was the predominant form, which was ideal for a manipulation that reduced LC3B transcription. In HeLa cells, LC3B was rapidly lipidated and degraded, so these changes were more difficult to parse. Furthermore, the extension of LC3B reduction to different shRNA targets (with different sets of potential off-targets) increased the probability that L1CAM knockdown was responsible for mediating this effect. The effects of the shRNAs in i³Neurons were found to be more consistent between targets than the siRNAs in HeLa cells; sh1 and sh2 were each highly effective at reducing both L1CAM and LC3B, while sh3 was moderately effective.

The discovery that *L1CAM* mRNA remained expressed in the KO clones despite a loss of the protein and that shRNA transduction in a KO clone induced LC3B reduction recontextualized the puzzling results from HeLa cells and led to the model presented here. In the absence of the i³Neurons, I would have instead sought to knock out L1CAM in HeLa cells. A lack of effect on LC3B in such a case would have likely caused the siRNA effect to instead be attributed to off-target effects, so no such model would have been developed.

6.3.3 Recapitulation of neuron-specific phenotypes

The biology of neurons differs substantially from that of other cell types. As was shown, protein turnover in neurons (both *in vivo* and iPSC-derived) is much slower than in HeLa cells and iPSCs, and the mitotic inactivity of neurons adds additional weight to the importance of nuclear translocation in SBMA cell models since the nuclear envelope is not refreshed. Cells differentiated in this work were shown to consistently express neuronal markers, including key developmental transcription factors, and the RNAseq data in Chapter 4 specifically assessed markers of neurotransmitter metabolism and synaptic activity.

Cellular features that involve neuron-specific functionality, such as axonal trafficking or electrophysiological activity, can only be studied in a neuronal model system. In this case, L1CAM could only be identified in the initial CRISPR screen because it was expressed in HeLa cells, but this expression was more likely related to the role of L1CAM in cancer than its role in neurodevelopment. However, i³Neurons were well-suited for studying the latter role, and experiments done in these cells permitted me to conclude that *L1CAM* regulates autophagy in neurons as well as cancer.

This study illustrates how different cell models can work together for greater effect. The ease of use of HeLa cells facilitated preliminary experiments and worked well to show effects on a constitutive system (i.e., autophagy). This provided the proof-of-concept needed to justify the investment of generating iPSC-derived neuron models, which were more relevant to the biology of *L1CAM*. Findings in HeLa cells that generalized to iPSCs or iPSC-derived cells were also more trustworthy than effects seen in one cell type alone, and HeLa cells were used for testing additional hypotheses, as with the lack of ability for L1CAM overexpression to increase LC3B.

6.4 Tools of the trade

Medical research studies human biology for the explicit purpose of understanding and treating human diseases. Recent advances have made gene editing and culture of human neuron-like cells commonplace, but it should be recognized that the ability to accurately model human biology *in vitro* is still a new and developing practice. As with any model system, iPSCs and the cells derived from them have strengths and weaknesses that need to be acknowledged and addressed. This work typified both aspects and, in the end, makes a case for the utility of iPSC-derived neurons for understanding the biology underpinning neurological diseases.

A common theme throughout this work was refining, using, and candidly assessing the strengths and weaknesses of various tools. I sought to address the bottleneck of transgenic selection with SBP- Δ LNGFR, I show its efficacy in detail in Chapter 3 and as a means to the end of developing the SBMA model lines in Chapter 4, and I show how I came to recognize its limitations. I sought to derive an improved model of SBMA, but I found that AR was not expressed or was not functional in the cell types of interest to this study; while this was not ideal, it was far better to accept the biological effects than to alter the protocol in an effort to artificially permit experimentation. Finally, I assessed the biological impact of a loss of *L1CAM*, which led me to recognize the importance of the transcript in maintaining expression of core components of macroautophagy. In so doing, I showed how iPSC-derived neurons can be used together with traditional cell models to gain further insight than either could alone.

I am hopeful that my work can be carried forward and be an example to others, both to expand on my successes and to avoid my mistakes. Beyond any innovation or discovery, the most rewarding aspect of this work was the opportunity to share the tools and knowledge I developed with the research community. Whether by teaching cell culture techniques to colleagues, publishing extensive protocols, or sharing cell lines and plasmids with collaborators, enabling others is the most satisfying use of expertise.

References

- Adachi, H. (2005). Widespread nuclear and cytoplasmic accumulation of mutant androgen receptor in SBMA patients. *Brain*, *128*(3), 659–670. <https://doi.org/10.1093/brain/awh381>
- Adey, A., Burton, J. N., Kitzman, J. O., Hiatt, J. B., Lewis, A. P., Martin, B. K., Qiu, R., Lee, C., & Shendure, J. (2013). The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature*, *500*(7461), 207–211. <https://doi.org/10.1038/nature12064>
- Agrotis, A., Pengo, N., Burden, J. J., & Ketteler, R. (2019). Redundancy of human ATG4 protease isoforms in autophagy and LC3/GABARAP processing revealed in cells. *Autophagy*, *15*(6), 976–997. <https://doi.org/10.1080/15548627.2019.1569925>
- Ahmed, A. I., Shtaya, A. B., Zaben, M. J., Owens, E. V., Kiecker, C., & Gray, W. P. (2012). Endogenous GFAP-positive neural stem/progenitor cells in the postnatal mouse cortex are activated following traumatic brain injury. *Journal of Neurotrauma*, *29*(5), 828–842. <https://doi.org/10.1089/neu.2011.1923>
- Albertelli, M. A., Scheller, A., Brogley, M., & Robins, D. M. (2006). Replacing the mouse androgen receptor with human alleles demonstrates glutamine tract length-dependent effects on physiology and tumorigenesis in mice. *Molecular Endocrinology*, *20*(6), 1248–1260. <https://doi.org/10.1210/me.2006-0021>
- Allen Institute for Brain Science. (2008). *Allen Spinal Cord Atlas*. <https://mousespinal.brain-map.org/>
- Amoroso, M. W., Croft, G. F., Williams, D. J., O’Keeffe, S., Carrasco, M. A., Davis, A. R., Roybon, L., Oakley, D. H., Maniatis, T., Henderson, C. E., & Wichterle, H. (2013). Accelerated high-yield generation of limb-innervating motor neurons from human stem cells. *Journal of Neuroscience*, *33*(2), 574–586. <https://doi.org/10.1523/JNEUROSCI.0906-12.2013>
- Angiolini, F., Belloni, E., Giordano, M., Campioni, M., Forneris, F., Paronetto, M. P., Lupia, M., Brandas, C., Pradella, D., Di Matteo, A., Giampietro, C., Jodice, G., Luise, C., Bertalot, G., Freddi, S., Malinverno, M., Irimia, M., Moulton, J. D., Summerton, J., ... Ghigna, C. (2019). A novel L1CAM isoform with angiogenic activity generated by NOVA2-mediated alternative splicing. *ELife*, *8*. <https://doi.org/10.7554/eLife.44305>
- Arakawa, S., Honda, S., Yamaguchi, H., & Shimizu, S. (2017). Molecular mechanisms and physiological roles of Atg5/Atg7-independent alternative autophagy. *Proceedings of the Japan Academy. Series B, Physical and*

- Biological Sciences*, 93(6), 378–385. <https://doi.org/10.2183/pjab.93.023>
- Araki, K., Imaizumi, T., Okuyama, K., Oike, Y., & Yamamura, K. I. (1997). Efficiency of recombination by Cre transient expression in embryonic stem cells: Comparison of various promoters. *Journal of Biochemistry*, 122(5), 977–982. <https://doi.org/10.1093/oxfordjournals.jbchem.a021860>
- Arber, S., Han, B., Mendelsohn, M., Smith, M., Jessell, T. M., & Sockanathan, S. (1999). Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. *Neuron*, 23(4), 659–674. [https://doi.org/10.1016/S0896-6273\(01\)80026-X](https://doi.org/10.1016/S0896-6273(01)80026-X)
- Arendt, D., Bertucci, P. Y., Achim, K., & Musser, J. M. (2019). Evolution of neuronal types and families. *Current Opinion in Neurobiology*, 56, 144–152. <https://doi.org/10.1016/j.conb.2019.01.022>
- Arnold, F. J., & Merry, D. E. (2019). Molecular Mechanisms and Therapeutics for SBMA/Kennedy’s Disease. *Neurotherapeutics*, 16(4), 928–947. <https://doi.org/10.1007/s13311-019-00790-9>
- Ashkenazi, A., Bento, C. F., Ricketts, T., Vicinanza, M., Siddiqi, F., Pavel, M., Squitieri, F., Hardenberg, M. C., Imarisio, S., Menzies, F. M., & Rubinsztein, D. C. (2017). Polyglutamine tracts regulate beclin 1-dependent autophagy. *Nature*, 545(7652), 108–111. <https://doi.org/10.1038/nature22078>
- Baker, K. E., & Parker, R. (2004). Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Current Opinion in Cell Biology*, 16(3), 293–299. <https://doi.org/10.1016/j.ceb.2004.03.003>
- Balding, D. J., & Nichols, R. A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica*, 96(1–2), 3–12. <https://doi.org/10.1007/BF01441146>
- Barbarese, E., Barry, C., Chou, C. J., Goldstein, D. J., Nakos, G. A., Hyde-DeRuyscher, R., Scheld, K., & Carson, J. H. (1988). Expression and Localization of Myelin Basic Protein in Oligodendrocytes and Transfected Fibroblasts. *Journal of Neurochemistry*, 51(6), 1737–1745. <https://doi.org/10.1111/j.1471-4159.1988.tb01153.x>
- Barrangou, R., & Doudna, J. A. (2016). Applications of CRISPR technologies in research and beyond. *Nature Biotechnology*, 34(9), 933–941. <https://doi.org/10.1038/nbt.3659>
- Bauwens, C. L., Peerani, R., Niebruegge, S., Woodhouse, K. A., Kumacheva, E., Husain, M., & Zandstra, P. W. (2008). Control of Human Embryonic Stem Cell Colony and Aggregate Size Heterogeneity Influences Differentiation Trajectories. *Stem Cells*, 26(9), 2300–2310. <https://doi.org/10.1634/stemcells.2008-0183>

- Bearer, C. F. (2001). Developmental neurotoxicity: Illustration of principles. *Pediatric Clinics of North America*, *48*(5), 1199–1213. [https://doi.org/10.1016/S0031-3955\(05\)70369-2](https://doi.org/10.1016/S0031-3955(05)70369-2)
- Belikov, S., Bott, L. C., Fischbeck, K. H., & Wrangé, Ö. (2015). The polyglutamine-expanded androgen receptor has increased DNA binding and reduced transcriptional activity. *Biochemistry and Biophysics Reports*, *3*, 134–139. <https://doi.org/10.1016/j.bbrep.2015.07.014>
- Bence, N. F., Sampat, R. M., & Kopito, R. R. (2001). Impairment of the ubiquitin-proteasome system by protein aggregation. *Science*, *292*(5521), 1552–1555. <https://doi.org/10.1126/science.292.5521.1552>
- Benjamini, Y., & 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. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bento, C. F., Ashkenazi, A., Jimenez-Sanchez, M., & Rubinsztein, D. C. (2016). The Parkinson's disease-associated genes ATP13A2 and SYT11 regulate autophagy via a common pathway. *Nature Communications*, *7*(1), 1–16. <https://doi.org/10.1038/ncomms11803>
- Bibikova, M., Golic, M., Golic, K. G., & Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, *161*(3), 1169–1175. <https://doi.org/10.1093/genetics/161.3.1169>
- Binder, D. K., & Scharfman, H. E. (2004). Brain-derived Neurotrophic Factor. *Growth Factors*, *22*(3), 123–131. <https://doi.org/10.1080/08977190410001723308>
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, *326*(5959), 1509–1512. <https://doi.org/10.1126/science.1178811>
- Bousquet, I., Bozon, M., Castellani, V., Touraine, R., Piton, A., Gérard, B., Guibaud, L., Sanlaville, D., Edery, P., Saugier-veber, P., & Putoux, A. (2021). X-linked partial corpus callosum agenesis with mild intellectual disability: identification of a novel L1CAM pathogenic variant. *Neurogenetics*, *22*(1), 43–51. <https://doi.org/10.1007/s10048-020-00629-y>
- Branscum, A. J., Gardner, I. A., & Johnson, W. O. (2005). Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Preventive Veterinary Medicine*, *68*(2–4), 145–163. <https://doi.org/10.1016/j.prevetmed.2004.12.005>
- Brinster, R. L., Chen, H. Y., Trumbauer, M. E., Yagle, M. K., & Palmiter, R. D. (1985). Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. *Proceedings of the National Academy of Sciences*, *82*(13),

4438–4442. <https://doi.org/10.1073/pnas.82.13.4438>

- Cai, J., Hong, Y., Weng, C., Tan, C., Imperato-McGinley, J., & Zhu, Y. S. (2011). Androgen stimulates endothelial cell proliferation via an androgen receptor/VEGF/cyclin A-mediated mechanism. *American Journal of Physiology - Heart and Circulatory Physiology*, *300*(4), H1210. <https://doi.org/10.1152/ajpheart.01210.2010>
- Cantone, I., & Fisher, A. G. (2013). Epigenetic programming and reprogramming during development. *Nature Structural & Molecular Biology*, *20*(3), 282–289. <https://doi.org/10.1038/nsmb.2489>
- Carlson-Stevermer, J., Abdeen, A. A., Kohlenberg, L., Goedland, M., Molugu, K., Lou, M., & Saha, K. (2017). Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nature Communications*, *8*(1), 1–13. <https://doi.org/10.1038/s41467-017-01875-9>
- Carlson-Stevermer, J., Goedland, M., Steyer, B., Movaghar, A., Lou, M., Kohlenberg, L., Prestil, R., & Saha, K. (2016). High-Content Analysis of CRISPR-Cas9 Gene-Edited Human Embryonic Stem Cells. *Stem Cell Reports*, *6*(1), 109–120.
- Cerbini, T., Funahashi, R., Luo, Y., Liu, C., Park, K., Rao, M., Malik, N., & Zou, J. (2015). Transcription activator-like effector nuclease (TALEN)-mediated CLYBL targeting enables enhanced transgene expression and one-step generation of dual reporter human induced pluripotent stem cell (iPSC) and neural stem cell (NSC) lines. *PloS One*, *10*(1), e0116032. <https://doi.org/10.1371/journal.pone.0116032>
- Chambers, S. M., Fasano, C. A., Papapetrou, E. P., Tomishima, M., Sadelain, M., & Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology*, *27*(3), 275–280. <https://doi.org/10.1038/nbt.1529>
- Chambers, S. M., Qi, Y., Mica, Y., Lee, G., Zhang, X. J., Niu, L., Bilsland, J., Cao, L., Stevens, E., Whiting, P., Shi, S. H., & Studer, L. (2012). Combined small-molecule inhibition accelerates developmental timing and converts human pluripotent stem cells into nociceptors. *Nature Biotechnology*, *30*(7), 715–720. <https://doi.org/10.1038/nbt.2249>
- Charpentier, M., Khedher, A. H. Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., Boix, C., Perrouault, L., Tesson, L., Geny, S., De Cian, A., Itier, J. M., Anegon, I., Lopez, B., Giovannangeli, C., & Concordet, J. P. (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nature Communications*, *9*(1), 1–11. <https://doi.org/10.1038/s41467-018-03475-7>
- Chauhan, S., Goodwin, J. G., Chauhan, S., Manyam, G., Wang, J., Kamat, A. M., & Boyd, D. D. (2013). ZKSCAN3 Is a Master Transcriptional Repressor of

- Autophagy. *Molecular Cell*, 50(1), 16–28.
<https://doi.org/10.1016/j.molcel.2013.01.024>
- Chen, D., Zeng, Z., Yang, J., Ren, C., Wang, D., Wu, W., & Xu, R. (2013). L1cam promotes tumor progression and metastasis and is an independent unfavorable prognostic factor in gastric cancer. *Journal of Hematology & Oncology*, 6(1), 43.
<https://doi.org/10.1186/1756-8722-6-43>
- Chen, E. A., Souaiaia, T., Herstein, J. S., Evgrafov, O. V., Spitsyna, V. N., Rebolini, D. F., & Knowles, J. A. (2014). Effect of RNA integrity on uniquely mapped reads in RNA-Seq. *BMC Research Notes*, 7(1), 753.
<https://doi.org/10.1186/1756-0500-7-753>
- Chen, G., Gulbranson, D. R., Hou, Z., Bolin, J. M., Ruotti, V., Probasco, M. D., Smuga-Otto, K., Howden, S. E., Diol, N. R., Propson, N. E., Wagner, R., Lee, G. O., Antosiewicz-Bourget, J., Teng, J. M. C., & Thomson, J. A. (2011). Chemically defined conditions for human iPSC derivation and culture. *Nature Methods*, 8(5), 424–429. <https://doi.org/10.1038/nmeth.1593>
- Cheng, L., Hansen, N. F., Zhao, L., Du, Y., Zou, C., Donovan, F. X., Chou, B. K., Zhou, G., Li, S., Dowey, S. N., Ye, Z., Chandrasekharappa, S. C., Yang, H., Mullikin, J. C., & Liu, P. P. (2012). Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell*, 10(3), 337–344.
<https://doi.org/10.1016/j.stem.2012.01.005>
- Chia, N. Y., Chan, Y. S., Feng, B., Lu, X., Orlov, Y. L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M. S., Huss, M., Soh, B. S., Kraus, P., Li, P., Lufkin, T., Lim, B., Clarke, N. D., Bard, F., & Ng, H. H. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*, 468(7321), 316–320. <https://doi.org/10.1038/nature09531>
- Chidsey, B. A., Baldwin, E. E., Toydemir, R., Ahles, L., Hanson, H., & Stevenson, D. A. (2014). L1CAM whole gene deletion in a child with L1 syndrome. *American Journal of Medical Genetics, Part A*, 164(6), 1555–1558.
<https://doi.org/10.1002/ajmg.a.36474>
- Choi, I. Y., Lim, H., Cho, H. J., Oh, Y., Chou, B. K., Bai, H., Cheng, L., Kim, Y. J., Hyun, S., Kim, H., Shin, J. H., & Lee, G. (2020). Transcriptional landscape of myogenesis from human pluripotent stem cells reveals a key role of TWIST1 in maintenance of skeletal muscle progenitors. *ELife*, 9.
<https://doi.org/10.7554/eLife.46981>
- Chu, L. Y., Guo, D. M., Chen, J. T., Fang, W. K., Xie, J. J., Peng, Y. H., Xu, Y. W., & Li, X. X. (2020). The Diagnostic Value of Serum L1CAM in Patients With Colorectal Cancer. *Technology in Cancer Research & Treatment*, 19.
<https://doi.org/10.1177/1533033820920971>

- Cipriani, G., Dolciotti, C., Picchi, L., & Bonuccelli, U. (2011). Alzheimer and his disease: A brief history. *Neurological Sciences*, *32*(2), 275–279. <https://doi.org/10.1007/s10072-010-0454-7>
- Cordie, T., Harkness, T., Jing, X., Carlson-Stevermer, J., Mi, H. Y., Turng, L. S., & Saha, K. (2014). Nanofibrous electrospun polymers for reprogramming human cells. *Cellular and Molecular Bioengineering*, *7*(3), 379–393. <https://doi.org/10.1007/s12195-014-0341-z>
- Cortes, C. J., Ling, S.-C., Guo, L. T., Hung, G., Tsunemi, T., Ly, L., Tokunaga, S., Lopez, E., Sopher, B. L., Bennett, C. F., Shelton, G. D., Cleveland, D. W., & La Spada, A. R. (2014a). Muscle Expression of Mutant Androgen Receptor Accounts for Systemic and Motor Neuron Disease Phenotypes in Spinal and Bulbar Muscular Atrophy. *Neuron*, *82*(2), 295–307. <https://doi.org/10.1016/j.neuron.2014.03.001>
- Cortes, C. J., Miranda, H. C., Frankowski, H., Batlevi, Y., Young, J. E., Le, A., Ivanov, N., Sopher, B. L., Carroneu, C., Muotri, A. R., Garden, G. A., & La Spada, A. R. (2014b). Polyglutamine-expanded androgen receptor interferes with TFEB to elicit autophagy defects in SBMA. *Nature Neuroscience*, *17*(9), 1180–1189. <https://doi.org/10.1038/nn.3787>
- Cuervo, A. M., & Wong, E. (2014). Chaperone-mediated autophagy: roles in disease and aging. *Cell Research*, *24*(1), 92–104. <https://doi.org/10.1038/cr.2013.153>
- Cyranoski, D., & Ledford, H. (2018). Genome-edited baby claim provokes international outcry. *Nature*, *563*(7733), 607–608. <https://doi.org/10.1038/d41586-018-07545-0>
- Dai, R., Chen, Y., Jiao, C., Dai, J., Chen, C., & Liu, C. (2019). Evaluating brain cell marker genes based on differential gene expression and co-expression. *BioRxiv*, 554626. <https://doi.org/10.1101/554626>
- Daly, A. C., Davidson, M. D., & Burdick, J. A. (2021). 3D bioprinting of high cell-density heterogeneous tissue models through spheroid fusion within self-healing hydrogels. *Nature Communications*, *12*(1). <https://doi.org/10.1038/s41467-021-21029-2>
- Dantuma, N. P., Lindsten, K., Glas, R., Jellne, M., & Masucci, M. G. (2000). Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. *Nature Biotechnology*, *18*(5), 538–543. <https://doi.org/10.1038/75406>
- Das, A. T., Tenenbaum, L., & Berkhout, B. (2016). Tet-On Systems For Doxycycline-inducible Gene Expression. *Current Gene Therapy*, *16*(3), 156–167. <https://doi.org/10.2174/1566523216666160524144041>
- Davis, A. J., & Chen, D. J. (2013). DNA double strand break repair via non-

- homologous end-joining. *Translational Cancer Research*, 2(3), 130–143.
<https://doi.org/10.3978/j.issn.2218-676X.2013.04.02>
- Davis, R. L., Weintraub, H., & Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, 51(6), 987–1000.
[https://doi.org/10.1016/0092-8674\(87\)90585-X](https://doi.org/10.1016/0092-8674(87)90585-X)
- De Angelis, E., Brümmendorf, T., Cheng, L., Lemmon, V., & Kenwrick, S. (2001). Alternative Use of a Mini Exon of the L1 Gene Affects L1 Binding to Neural Ligands. *Journal of Biological Chemistry*, 276(35), 32738–32742.
<https://doi.org/10.1074/jbc.M105156200>
- Debnath, M., Prasad, G. B. K. S., & Bisen, P. S. (2010). Reporter Gene. In *Molecular Diagnostics: Promises and Possibilities* (pp. 71–84). Springer Netherlands. https://doi.org/10.1007/978-90-481-3261-4_5
- Delsing, L., Herland, A., Falk, A., Hicks, R., Synnergren, J., & Zetterberg, H. (2020). Models of the blood-brain barrier using iPSC-derived cells. *Molecular and Cellular Neuroscience*, 107, 103533. <https://doi.org/10.1016/j.mcn.2020.103533>
- den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., Roux, A. F., Smith, T., Antonarakis, S. E., & Taschner, P. E. M. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation*, 37(6), 564–569.
<https://doi.org/10.1002/humu.22981>
- Deng, Z., Purtell, K., Lachance, V., Wold, M. S., Chen, S., & Yue, Z. (2017). Autophagy Receptors and Neurodegenerative Diseases. *Trends in Cell Biology*, 27(7), 491–504. <https://doi.org/10.1016/j.tcb.2017.01.001>
- Di Malta, C., Cinque, L., & Settembre, C. (2019). Transcriptional regulation of autophagy: Mechanisms and diseases. *Frontiers in Cell and Developmental Biology*, 7, 114. <https://doi.org/10.3389/fcell.2019.00114>
- Dikic, I., & Elazar, Z. (2018). Mechanism and medical implications of mammalian autophagy. *Nature Reviews Molecular Cell Biology*, 19(6), 349–364.
<https://doi.org/10.1038/s41580-018-0003-4>
- Dimos, J. T., Rodolfa, K. T., Niakan, K. K., Weisenthal, L. M., Mitsumoto, H., Chung, W., Croft, G. F., Saphier, G., Leibel, R., Golland, R., Wichterle, H., Henderson, C. E., & Eggan, K. (2008). Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science*, 321(5893), 1218–1221. <https://doi.org/10.1126/science.1158799>
- Djajadikerta, A., Keshri, S., Pavel, M., Prestil, R., Ryan, L., & Rubinsztein, D. C. (2020). Autophagy Induction as a Therapeutic Strategy for Neurodegenerative Diseases. *Journal of Molecular Biology*, 432(8), 2799–2821.
<https://doi.org/10.1016/j.jmb.2019.12.035>

- Doberstein, K., Harter, P. N., Haberkorn, U., Bretz, N. P., Arnold, B., Carretero, R., Moldenhauer, G., Mittelbronn, M., & Altevogt, P. (2015). Antibody therapy to human L1CAM in a transgenic mouse model blocks local tumor growth but induces EMT. *International Journal of Cancer*, *136*(5), E326–E339. <https://doi.org/10.1002/ijc.29222>
- Doherty, M. K., Hammond, D. E., Clague, M. J., Gaskell, S. J., & Beynon, R. J. (2009). Turnover of the human proteome: Determination of protein intracellular stability by dynamic SILAC. *Journal of Proteome Research*, *8*(1), 104–112. <https://doi.org/10.1021/pr800641v>
- Dooley, H. C., Razi, M., Polson, H. E. J., Girardin, S. E., Wilson, M. I., & Tooze, S. A. (2014). WIPI2 Links LC3 Conjugation with PI3P, Autophagosome Formation, and Pathogen Clearance by Recruiting Atg12-5-16L1. *Molecular Cell*, *55*(2), 238–252. <https://doi.org/10.1016/j.molcel.2014.05.021>
- Dörrbaum, A. R., Kochen, L., Langer, J. D., & Schuman, E. M. (2018). Local and global influences on protein turnover in neurons and glia. *ELife*, *7*. <https://doi.org/10.7554/eLife.34202>
- Dou, X., Menkari, C., Mitsuyama, R., Foroud, T., Wetherill, L., Hammond, P., Suttie, M., Chen, X., Chen, S.-Y. Y., & Charness, M. E. (2018). L1 coupling to ankyrin and the spectrin-actin cytoskeleton modulates ethanol inhibition of L1 adhesion and ethanol teratogenesis. *FASEB Journal*, *32*(3), 1364–1374. <https://doi.org/10.1096/fj.201700970>
- Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213), 1258096. <https://doi.org/10.1126/science.1258096>
- Drewe, J. A., Tomlinson, A. J., Walker, N. J., & Delahay, R. J. (2010). Diagnostic accuracy and optimal use of three tests for tuberculosis in live badgers. *PLoS ONE*, *5*(6), e11196. <https://doi.org/10.1371/journal.pone.0011196>
- Dugger, B. N., & Dickson, D. W. (2017). Pathology of Neurodegenerative Diseases. *Cold Spring Harbor Perspectives in Biology*, *9*(7), a028035. <https://doi.org/10.1101/cshperspect.a028035>
- Elegheert, J., Behiels, E., Bishop, B., Scott, S., Woolley, R. E., Griffiths, S. C., Byrne, E. F. X., Chang, V. T., Stuart, D. I., Jones, E. Y., Siebold, C., & Aricescu, A. R. (2018). Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nature Protocols*, *13*(12), 2991–3017. <https://doi.org/10.1038/s41596-018-0075-9>
- Er, E. E., Valiente, M., Ganesh, K., Zou, Y., Agrawal, S., Hu, J., Griscom, B., Rosenblum, M., Boire, A., Brogi, E., Giancotti, F. G., Schachner, M., Malladi, S., & Massagué, J. (2018). Pericyte-like spreading by disseminated cancer cells activates YAP and MRTF for metastatic colonization. *Nature Cell Biology*,

20(8), 966–978. <https://doi.org/10.1038/s41556-018-0138-8>

- Ericson, J., Morton, S., Kawakami, A., Roelink, H., & Jessell, T. M. (1996). Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell*, 87(4), 661–673. [https://doi.org/10.1016/S0092-8674\(00\)81386-0](https://doi.org/10.1016/S0092-8674(00)81386-0)
- Fang, X., Yu, S. X., Lu, Y., Bast, R. C., Woodgett, J. R., & Mills, G. B. (2000). Phosphorylation and inactivation of glycogen synthase kinase 3 by protein kinase A. *Proceedings of the National Academy of Sciences*, 97(22), 11960–11965. <https://doi.org/10.1073/pnas.220413597>
- Farla, P., Hersmus, R., Geverts, B., Mari, P. O., Nigg, A. L., Dubbink, H. J., Trapman, J., & Houtsmuller, A. B. (2004). The androgen receptor ligand-binding domain stabilizes DNA binding in living cells. *Journal of Structural Biology*, 147(1), 50–61. <https://doi.org/10.1016/j.jsb.2004.01.002>
- Farzaneh, M. (2020). Concise Review; Effects of Antibiotics and Antimycotics on the Biological Properties of Human Pluripotent and Multipotent Stem Cells. *Current Stem Cell Research & Therapy*, 16(4), 400–405. <https://doi.org/10.2174/1574888x16999201203214425>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fernandopulle, M. S., Prestil, R., Grunseich, C., Wang, C., Gan, L., & Ward, M. E. (2018). Transcription Factor–Mediated Differentiation of Human iPSCs into Neurons. *Current Protocols in Cell Biology*, 79(1). <https://doi.org/10.1002/cpcb.51>
- Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815. <https://doi.org/10.1080/0266476042000214501>
- Figura, G., Koscianska, E., & Krzyzosiak, W. J. (2015). In Vitro expansion of CAG, CAA, and mixed CAG/CAA repeats. *International Journal of Molecular Sciences*, 16(8), 18741–18751. <https://doi.org/10.3390/ijms160818741>
- Fischbeck, K. H. (2016). Spinal and Bulbar Muscular Atrophy Overview. *Journal of Molecular Neuroscience*, 58(3), 317–320. <https://doi.org/10.1007/s12031-015-0674-7>
- Florey, O., Kim, S. E., Sandoval, C. P., Haynes, C. M., & Overholtzer, M. (2011). Autophagy machinery mediates macroendocytic processing and entotic cell death by targeting single membranes. *Nature Cell Biology*, 13(11), 1335–1343. <https://doi.org/10.1038/ncb2363>
- Formaggio, N., Rubin, M. A., & Theurillat, J.-P. (2021). Loss and revival of androgen receptor signaling in advanced prostate cancer. *Oncogene*, 40(7), 1205–

1216. <https://doi.org/10.1038/s41388-020-01598-0>

- Forman, M. S., Trojanowski, J. Q., & Lee, V. M. Y. (2004). Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs. *Nature Medicine*, *10*(10), 1055–1063. <https://doi.org/10.1038/nm1113>
- Füllgrabe, J., Ghislat, G., Cho, D. H., & Rubinsztein, D. C. (2016). Transcriptional regulation of mammalian autophagy at a glance. *Journal of Cell Science*, *129*(16), 3059–3066. <https://doi.org/10.1242/jcs.188920>
- Funderburk, S. F., Wang, Q. J., & Yue, Z. (2010). The Beclin 1–VPS34 complex – at the crossroads of autophagy and beyond. *Trends in Cell Biology*, *20*(6), 355–362. <https://doi.org/10.1016/j.tcb.2010.03.002>
- Furlong, R. A., Narain, Y., Rankin, J., Wyttenbach, A., & Rubinsztein, D. C. (2000). α -Synuclein overexpression promotes aggregation of mutant huntingtin. *Biochemical Journal*, *346*(3), 577–581. <https://doi.org/10.1042/0264-6021:3460577>
- Gagnon, K. T., Li, L., Chu, Y., Janowski, B. A., & Corey, D. R. (2014). RNAi factors are present and active in human cell nuclei. *Cell Reports*, *6*(1), 211–221. <https://doi.org/10.1016/j.celrep.2013.12.013>
- Gaj, T., Gersbach, C. A., & Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405. <https://doi.org/10.1016/j.tibtech.2013.04.004>
- Ganesh, K., Basnet, H., Kaygusuz, Y., Laughney, A. M., He, L., Sharma, R., O'Rourke, K. P., Reuter, V. P., Huang, Y. H., Turkecul, M., Er, E. E., Masilionis, I., Manova-Todorova, K., Weiser, M. R., Saltz, L. B., Garcia-Aguilar, J., Koche, R., Lowe, S. W., Pe'er, D., ... Massagué, J. (2020). L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer. *Nature Cancer*, *1*(1), 28–45. <https://doi.org/10.1038/s43018-019-0006-x>
- Gao, X., Sprando, R. L., & Yourick, J. J. (2018a). A Rapid and Highly Efficient Method for the Isolation, Purification, and Passaging of Human-Induced Pluripotent Stem Cells. *Cellular Reprogramming*, *20*(5), 282–288. <https://doi.org/10.1089/cell.2018.0022>
- Gao, Z., Herrera-Carrillo, E., & Berkhout, B. (2018b). Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. *Molecular Therapy - Nucleic Acids*, *10*, 36–44. <https://doi.org/10.1016/j.omtn.2017.11.006>
- Gast, D., Riedle, S., Issa, Y., Pfeifer, M., Beckhove, P., Sanderson, M. P., Arlt, M., Moldenhauer, G., Fogel, M., Krüger, A., & Altevogt, P. (2008). The cytoplasmic part of L1-CAM controls growth and gene expression in human tumors that is reversed by therapeutic antibodies. *Oncogene*, *27*(9), 1281–1289. <https://doi.org/10.1038/sj.onc.1210747>

- Gil, O. D., Sakurai, T., Bradley, A. E., Fink, M. Y., Cassella, M. R., Kuo, J. A., & Felsenfeld, D. P. (2003). Ankyrin binding mediates L1CAM interactions with static components of the cytoskeleton and inhibits retrograde movement of L1CAM on the cell surface. *Journal of Cell Biology*, *162*(4), 719–730. <https://doi.org/10.1083/jcb.200211011>
- Godenschwege, T. A., Kristiansen, L. V., Uthaman, S. B., Hortsch, M., & Murphey, R. K. (2006). A conserved role for Drosophila Neuroglian and human L1-CAM in central-synapse formation. *Current Biology*, *16*(1), 12–23. <https://doi.org/10.1016/j.cub.2005.11.062>
- Gong, Y., Wang, D., Dar, J. A., Singh, P., Graham, L., Liu, W., Ai, J., Xin, Z., Guo, Y., & Wang, Z. (2012). Nuclear export signal of androgen receptor (NESAR) regulation of androgen receptor level in human prostate cell lines via ubiquitination and proteasome-dependent degradation. *Endocrinology*, *153*(12), 5716–5725. <https://doi.org/10.1210/en.2012-1841>
- Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A., & Ruddle, F. H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. *Proceedings of the National Academy of Sciences*, *77*(12), 7380–7384. <https://doi.org/10.1073/pnas.77.12.7380>
- Greussing, R., Unterluggauer, H., Koziel, R., Maier, A. B., & Jansen-Dürr, P. (2012). Monitoring of ubiquitin-proteasome activity in living cells using a degron (dgn)-destabilized green fluorescent protein (GFP)-based reporter protein. *Journal of Visualized Experiments*, *69*, 3327. <https://doi.org/10.3791/3327>
- Gromova, A., & La Spada, A. R. (2020). Harmony Lost: Cell–Cell Communication at the Neuromuscular Junction in Motor Neuron Disease. *Trends in Neurosciences*, *43*(9), 709–724. <https://doi.org/10.1016/j.tins.2020.07.002>
- Grońska-Pełski, M., Schachner, M., & Hébert, J. M. (2020). L1cam curbs the differentiation of adult-born hippocampal neurons. *Stem Cell Research*, *48*. <https://doi.org/10.1016/j.scr.2020.101999>
- Grunseich, C., Kats, I. R., Bott, L. C., Rinaldi, C., Kokkinis, A., Fox, D., Chen, K. lian, Schindler, A. B., Mankodi, A. K., Shrader, J. A., Schwartz, D. P., Lehky, T. J., Liu, C. Y., & Fischbeck, K. H. (2014a). Early onset and novel features in a spinal and bulbar muscular atrophy patient with a 68 CAG repeat. *Neuromuscular Disorders*, *24*(11), 978–981. <https://doi.org/10.1016/j.nmd.2014.06.441>
- Grunseich, C., Zukosky, K., Kats, I. R., Ghosh, L., Harmison, G. G., Bott, L. C., Rinaldi, C., Chen, K. lian, Chen, G., Boehm, M., & Fischbeck, K. H. (2014b). Stem cell-derived motor neurons from spinal and bulbar muscular atrophy patients. *Neurobiology of Disease*, *70*, 12–20. <https://doi.org/10.1016/j.nbd.2014.05.038>

- Guber, R. D., Kokkinis, A. D., Schindler, A. B., Bendixen, R. M., Heatwole, C. R., Fischbeck, K. H., & Grunseich, C. (2018). Patient-identified impact of symptoms in spinal and bulbar muscular atrophy. *Muscle & Nerve*, *57*(1), 40–44. <https://doi.org/10.1002/mus.25957>
- Guber, R. D., Takyar, V., Kokkinis, A., Fox, D. A., Alao, H., Kats, I., Bakar, D., Remaley, A. T., Hewitt, S. M., Kleiner, D. E., Liu, C.-Y., Hadigan, C., Fischbeck, K. H., Rotman, Y., & Grunseich, C. (2017). Nonalcoholic fatty liver disease in spinal and bulbar muscular atrophy. *Neurology*, *89*(24), 2481–2490. <https://doi.org/10.1212/WNL.0000000000004748>
- Gulati, G. S., Sikandar, S. S., Wesche, D. J., Manjunath, A., Bharadwaj, A., Berger, M. J., Ilagan, F., Kuo, A. H., Hsieh, R. W., Cai, S., Zabala, M., Scheeren, F. A., Lobo, N. A., Qian, D., Yu, F. B., Dirbas, F. M., Clarke, M. F., & Newman, A. M. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. *Science*, *367*(6476), 405–411. <https://doi.org/10.1126/science.aax0249>
- Guo, W., & Giancotti, F. G. (2004). Integrin signalling during tumour progression. *Nature Reviews Molecular Cell Biology*, *5*(10), 816–826. <https://doi.org/10.1038/nrm1490>
- Gurdon, J. B., & Graham, C. F. (1967). Nuclear changes during cell differentiation. *Science Progress*, *55*(218), 259–277. <https://www.jstor.org/stable/43419644>
- Gurtan, A. M., Lu, V., Bhutkar, A., & Sharp, P. A. (2012). In vivo structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. *RNA*, *18*(6), 1116–1122. <https://doi.org/10.1261/rna.032680.112>
- Gusel'nikova, V. V., & Korzhevskiy, D. E. (2015). NeuN As a Neuronal Nuclear Antigen and Neuron Differentiation Marker. *Acta Naturae*, *7*(2), 42–47. <https://doi.org/10.32607/20758251-2015-7-2-42-47>
- Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, *4*(2), 627–635. [/pmc/articles/PMC3755824/](https://pubmed.ncbi.nlm.nih.gov/24111111/)
- Hall, C. E., Yao, Z., Choi, M., Tyzack, G. E., Serio, A., Luisier, R., Harley, J., Preza, E., Arber, C., Crisp, S. J., Watson, P. M. D., Kullmann, D. M., Abramov, A. Y., Wray, S., Burley, R., Loh, S. H. Y., Martins, L. M., Stevens, M. M., Luscombe, N. M., ... Patani, R. (2017). Progressive Motor Neuron Pathology and the Role of Astrocytes in a Human Stem Cell Model of VCP-Related ALS. *Cell Reports*, *19*(9), 1739–1749. <https://doi.org/10.1016/j.celrep.2017.05.024>
- Hanna, J., Saha, K., Pando, B., Van Zon, J., Lengner, C. J., Creighton, M. P., Van Oudenaarden, A., & Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature*, *462*(7273), 595–601. <https://doi.org/10.1038/nature08592>

- Hao, L., He, Q., Wang, Z., Craven, M., Newton, M. A., & Ahlquist, P. (2013). Limited Agreement of Independent RNAi Screens for Virus-Required Host Genes Owes More to False-Negative than False-Positive Factors. *PLoS Computational Biology*, *9*(9), 1003235. <https://doi.org/10.1371/journal.pcbi.1003235>
- Hashizume, A., Katsuno, M., Suzuki, K., Hirakawa, A., Hijikata, Y., Yamada, S., Inagaki, T., Banno, H., & Sobue, G. (2017). Long-term treatment with leuprorelin for spinal and bulbar muscular atrophy. *Journal of the Neurological Sciences*, *381*, 205–206. <https://doi.org/10.1016/j.jns.2017.08.588>
- Hayashi, Y., Hsiao, E. C., Sami, S., Lancero, M., Schlieve, C. R., Nguyen, T., Yano, K., Nagahashi, A., Ikeya, M., Matsumoto, Y., Nishimura, K., Fukuda, A., Hisatake, K., Tomoda, K., Asaka, I., Toguchida, J., Conklin, B. R., & Yamanaka, S. (2016). BMP-SMAD-ID promotes reprogramming to pluripotency by inhibiting p16/INK4A-dependent senescence. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(46), 13057–13062. <https://doi.org/10.1073/pnas.1603668113>
- Heckmann, B. L., Teubner, B. J. W., Tummers, B., Boada-Romero, E., Harris, L., Yang, M., Guy, C. S., Zakharenko, S. S., & Green, D. R. (2019). LC3-Associated Endocytosis Facilitates β -Amyloid Clearance and Mitigates Neurodegeneration in Murine Alzheimer's Disease. *Cell*, *178*(3), 536-551.e14. <https://doi.org/10.1016/j.cell.2019.05.056>
- Heinlein, C. A., & Chang, C. (2002). Androgen Receptor (AR) Coregulators: An Overview. *Endocrine Reviews*, *23*(2), 175–200. <https://academic.oup.com/edrv/article-abstract/23/2/175/2424160>
- Herron, L. R., Hill, M., Davey, F., & Gunn-Moore, F. J. (2009). The intracellular interactions of the L1 family of cell adhesion molecules. *Biochemical Journal*, *419*(3), 519–531. <https://doi.org/10.1042/BJ20082284>
- Herzog, E., Landry, M., Buhler, E., Bouali-Benazzouz, R., Legay, C., Henderson, C. E., Nagy, F., Dreyfus, P., Giros, B., & El Mestikawy, S. (2004). Expression of vesicular glutamate transporters, VGLUT1 and VGLUT2, in cholinergic spinal motoneurons. *European Journal of Neuroscience*, *20*(7), 1752–1760. <https://doi.org/10.1111/j.1460-9568.2004.03628.x>
- Hester, M. E., Murtha, M. J., Song, S., Rao, M., Miranda, C. J., Meyer, K., Tian, J., Boulting, G., Schaffer, D. V, Zhu, M. X., Pfaff, S. L., Gage, F. H., & Kaspar, B. K. (2011). Rapid and Efficient Generation of Functional Motor Neurons From Human Pluripotent Stem Cells Using Gene Delivered Transcription Factor Codes. *Molecular Therapy*, *19*(10), 1905–1912. <https://doi.org/10.1038/mt.2011.135>
- Heurtier, V., Owens, N., Gonzalez, I., Mueller, F., Proux, C., Mornico, D., Clerc, P., Dubois, A., & Navarro, P. (2019). The molecular logic of Nanog-induced self-

- renewal in mouse embryonic stem cells. *Nature Communications*, *10*(1), 1–15. <https://doi.org/10.1038/s41467-019-09041-z>
- Hindley, C., Ali, F., McDowell, G., Cheng, K., Jones, A., Guillemot, F., & Philpott, A. (2012). Post-translational modification of Ngn2 differentially affects transcription of distinct targets to regulate the balance between progenitor maintenance and differentiation. *Development*, *139*(10), 1718–1723. <https://doi.org/10.1242/dev.077552>
- Hlavin, M. L., & Lemmon, V. (1991). Molecular structure and functional testing of human L1CAM: An interspecies comparison. *Genomics*, *11*(2), 416–423. [https://doi.org/10.1016/0888-7543\(91\)90150-D](https://doi.org/10.1016/0888-7543(91)90150-D)
- Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., Cost, G. J., Zhang, L., Santiago, Y., Miller, J. C., Zeitler, B., Cherone, J. M., Meng, X., Hinkley, S. J., Rebar, E. J., Gregory, P. D., Urnov, F. D., & Jaenisch, R. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nature Biotechnology*, *29*(8), 731–734. <https://doi.org/10.1038/nbt.1927>
- Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M., & Weissman, J. S. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *ELife*, *5*(September2016). <https://doi.org/10.7554/eLife.19760>
- Hortsch, M. (2000). Structural and Functional Evolution of the L1 Family: Are Four Adhesion Molecules Better Than One? *Molecular and Cellular Neuroscience*, *15*(1), 1–10. <https://doi.org/10.1006/mcne.1999.0809>
- Hounoum, B. M., Vourc'h, P., Felix, R., Corcia, P., Patin, F., Guéguinou, M., Potier-Cartereau, M., Vandier, C., Raoul, C., Andres, C. R., Mavel, S., & Blasco, H. (2016). NSC-34 motor neuron-like cells are unsuitable as experimental model for glutamate-mediated excitotoxicity. *Frontiers in Cellular Neuroscience*, *10*(MAY), 118. <https://doi.org/10.3389/fncel.2016.00118>
- Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Ridwan Amode, M., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., ... Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, *49*(D1), D884–D891. <https://doi.org/10.1093/nar/gkaa942>
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), 827–832. <https://doi.org/10.1038/nbt.2647>
- Hu, B.-Y., Weick, J. P., Yu, J., Ma, L.-X., Zhang, X.-Q., Thomson, J. A., & Zhang,

- S.-C. (2010). Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proceedings of the National Academy of Sciences*, *107*(9), 4335–4340.
<https://doi.org/10.1073/pnas.0910012107>
- Huang, R., Xu, Y., Wan, W., Shou, X., Qian, J., You, Z., Liu, B., Chang, C., Zhou, T., Lippincott-Schwartz, J., & Liu, W. (2015). Deacetylation of nuclear LC3 drives autophagy initiation under starvation. *Molecular Cell*, *57*(3), 456–466.
<https://doi.org/10.1016/j.molcel.2014.12.013>
- Huang, W.-P., & Klionsky, D. J. (2002). Autophagy in Yeast: A Review of the Molecular Machinery. *Cell Structure and Function*, *27*(6), 409–420.
<https://doi.org/10.1247/csf.27.409>
- Hung, S. C., Wu, I. H., Hsue, S. S., Liao, C. H., Wang, H. C., Chuang, P. H., Sung, S. Y., & Hsieh, C. L. (2010). Targeting L1 cell adhesion molecule using lentivirus-mediated short hairpin RNA interference reverses aggressiveness of oral squamous cell carcinoma. *Molecular Pharmaceutics*, *7*(6), 2312–2323.
<https://doi.org/10.1021/mp1002834>
- Itakura, E., Kishi, C., Inoue, K., & Mizushima, N. (2008). Beclin 1 forms two distinct phosphatidylinositol 3-kinase complexes with mammalian Atg14 and UVRAG. *Molecular Biology of the Cell*, *19*(12), 5360–5372.
<https://doi.org/10.1091/mbc.E08-01-0080>
- Jänne, O. A., Moilanen, A.-M., Poukka, H., Rouleau, N., Karvonen, U., Kotaja, N., Häkli, M., & Palvimo, J. J. (2000). Androgen-receptor-interacting nuclear proteins. *Biochemical Society Transactions*, *28*(4), 401.
<https://doi.org/10.1042/0300-5127:0280401>
- Jasanoff, S., Hurlbut, J. B., & Saha, K. (2019). Democratic Governance of Human Germline Genome Editing. *The CRISPR Journal*, *2*(5), 266–271.
<https://doi.org/10.1089/crispr.2019.0047>
- Jiang, T., Qin, B., He, J., Lin, S., & Ding, S. (2013). Three isoforms of the Atg16L1 protein contribute different autophagic properties. *Molecular and Cellular Biochemistry*, *378*(1–2), 257–266. <https://doi.org/10.1007/s11010-013-1616-8>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821.
<https://doi.org/10.1126/science.1225829>
- Johansen, T., & Lamark, T. (2020). Selective Autophagy: ATG8 Family Proteins, LIR Motifs and Cargo Receptors. *Journal of Molecular Biology*, *432*(1), 80–103.
<https://doi.org/10.1016/j.jmb.2019.07.016>
- Johnson, R. W. (2013). Applications of the Beta Distribution Part 1: Transformation

- Group Approach. *ArXiv*. <http://arxiv.org/abs/1307.6437>
- Johnson, W. O., Gastwirth, J. L., & Pearson, L. M. (2001). Screening without a “gold standard”: The Hui-Walter paradigm revisited. *American Journal of Epidemiology*, *153*(9), 921–924. <https://doi.org/10.1093/aje/153.9.921>
- Joseph, L., Gyorkos, T. W., & Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology*, *141*(3), 263–272. <https://doi.org/10.1093/oxfordjournals.aje.a117428>
- Kabeya, Y., Mizushima, N., Yamamoto, A., Oshitani-Okamoto, S., Ohsumi, Y., & Yoshimori, T. (2004). LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *Journal of Cell Science*, *117*(13), 2805–2812. <https://doi.org/10.1242/jcs.01131>
- Kallunki, P., Edelman, G. M., & Jones, F. S. (1997). Tissue-specific expression of the L1 cell adhesion molecule is modulated by the neural restrictive silencer element. *Journal of Cell Biology*, *138*(6), 1343–1354. <https://doi.org/10.1083/jcb.138.6.1343>
- Kamaraj, U. S., Chen, J., Katwadi, K., Ouyang, J. F., Yang Sun, Y. B., Lim, Y. M., Liu, X., Handoko, L., Polo, J. M., Petretto, E., & Rackham, O. J. L. (2020). EpiMogrify Models H3K4me3 Data to Identify Signaling Molecules that Improve Cell Fate Control and Maintenance. *Cell Systems*, *11*(5), 509-522.e10. <https://doi.org/10.1016/j.cels.2020.09.004>
- Kamiguchi, H., Hlavin, M. L., & Lemmon, V. (1998). Role of L1 in neural development: What the knockouts tell us. In *Molecular and Cellular Neurosciences* (Vol. 12, Issues 1–2, pp. 48–55). Academic Press Inc. <https://doi.org/10.1006/mcne.1998.0702>
- Kamiguchi, H., Long, K. E., Pendergast, M., Schaefer, A. W., Rapoport, I., Kirchhausen, T., & Lemmon, V. (1998). The neural cell adhesion molecule L1 interacts with the AP-2 adaptor and is endocytosed via the clathrin-mediated pathway. *Journal of Neuroscience*, *18*(14), 5311–5321. <https://doi.org/10.1523/jneurosci.18-14-05311.1998>
- Kan, S. L., Saksouk, N., & Déjardin, J. (2017). Proteome characterization of a chromatin locus using the proteomics of isolated chromatin segments approach. In L. Comai (Ed.), *Methods in Molecular Biology* (Vol. 1550, pp. 19–32). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-6747-6_3
- Katsuno, M., Adachi, H., Inukai, A., & Sobue, G. (2003). Transgenic mouse models of spinal and bulbar muscular atrophy (SBMA). *Cytogenetic and Genome Research*, *100*(1–4), 243–251. <https://doi.org/10.1159/000072860>
- Katsuno, M., Adachi, H., Kume, A., Li, M., Nakagomi, Y., Niwa, H., Sang, C.,

- Kobayashi, Y., Doyu, M., & Sobue, G. (2002). Testosterone Reduction Prevents Phenotypic Expression in a Transgenic Mouse Model of Spinal and Bulbar Muscular Atrophy. *Neuron*, *35*(5), 843–854. [https://doi.org/10.1016/S0896-6273\(02\)00834-6](https://doi.org/10.1016/S0896-6273(02)00834-6)
- Katsuno, M., Banno, H., Suzuki, K., Takeuchi, Y., Kawashima, M., Yabe, I., Sasaki, H., Aoki, M., Morita, M., Nakano, I., Kanai, K., Ito, S., Ishikawa, K., Mizusawa, H., Yamamoto, T., Tsuji, S., Hasegawa, K., Shimohata, T., Nishizawa, M., ... Sobue, G. (2010). Efficacy and safety of leuprorelin in patients with spinal and bulbar muscular atrophy (JASMITT study): a multicentre, randomised, double-blind, placebo-controlled trial. *The Lancet Neurology*, *9*(9), 875–884. [https://doi.org/10.1016/S1474-4422\(10\)70182-4](https://doi.org/10.1016/S1474-4422(10)70182-4)
- Kiefel, H., Bondong, S., Erbe-Hoffmann, N., Hazin, J., Riedle, S., Wolf, J., Pfeifer, M., Arlt, A., Schäfer, H., Mürköster, S. S., & Altevogt, P. (2010). L1CAM–integrin interaction induces constitutive NF- κ B activation in pancreatic adenocarcinoma cells by enhancing IL-1 β expression. *Oncogene*, *29*(34), 4766–4778. <https://doi.org/10.1038/onc.2010.230>
- Kiefel, H., Bondong, S., Hazin, J., Ridinger, J., Schirmer, U., Riedle, S., & Altevogt, P. (2012a). L1CAM: A major driver for tumor cell invasion and motility. *Cell Adhesion and Migration*, *6*(4), 374–384. <https://doi.org/10.4161/cam.20832>
- Kiefel, H., Bondong, S., Pfeifer, M., Schirmer, U., Erbe-Hoffmann, N., Schäfer, H., Sebens, S., & Altevogt, P. (2012b). EMT-associated up-regulation of L1CAM provides insights into L1CAM-mediated integrin signalling and NF- κ B activation. *Carcinogenesis*, *33*(10), 1919–1929. <https://doi.org/10.1093/carcin/bgs220>
- Kim, H., Lim, Y., Lee, E., Oh, Y. J., & Kim, K. (2018). Correlation between the CAG repeat size and electrophysiological findings in patients with spinal and bulbar muscular atrophy. *Muscle & Nerve*, *57*(4), 683–686. <https://doi.org/10.1002/mus.25977>
- Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabudde, N. A., Balakrishnan, L., Advani, J., George, B., Renuse, S., ... Pandey, A. (2014). A draft map of the human proteome. *Nature*, *509*(7502), 575–581. <https://doi.org/10.1038/nature13302>
- King-Jones, K., & Thummel, C. S. (2005). Nuclear receptors — a perspective from Drosophila. *Nature Reviews Genetics*, *6*(4), 311–323. <https://doi.org/10.1038/nrg1581>
- Kirkeby, A., & Parmar, M. (2012). Building authentic midbrain dopaminergic neurons from stem cells - lessons from development. *Translational Neuroscience*,

3(4), 314–319. <https://doi.org/10.2478/s13380-012-0041-x>

- Kleeman, B., Olsson, A., Newkold, T., Kofron, M., DeLay, M., Hildeman, D., & Grimes, H. L. (2018). A guide to choosing fluorescent protein combinations for flow cytometric analysis based on spectral overlap. *Cytometry Part A*, *93*(5), 556–562. <https://doi.org/10.1002/cyto.a.23360>
- Kleene, R., Lutz, D., Loers, G., Bork, U., Borgmeyer, U., Hermans-Borgmeyer, I., & Schachner, M. (2021). Revisiting the proteolytic processing of cell adhesion molecule L1. *Journal of Neurochemistry*, *157*(4), 1102–1117. <https://doi.org/10.1111/jnc.15201>
- Klionsky, D. J., Abdel-Aziz, A. K., Abdelfatah, S., Abdellatif, M., Abdoli, A., Abel, S., Abeliovich, H., Abildgaard, M. H., Abudu, Y. P., Acevedo-Arozena, A., Adamopoulos, I. E., Adeli, K., Adolph, T. E., Adornetto, A., Aflaki, E., Agam, G., Agarwal, A., Aggarwal, B. B., Agnello, M., ... Tong, C. K. (2021). Guidelines for the use and interpretation of assays for monitoring autophagy (4th edition). *Autophagy*, *17*(1). <https://doi.org/10.1080/15548627.2020.1797280>
- Klionsky, D. J., Cregg, J. M., Dunn, W. A., Emr, S. D., Sakai, Y., Sandoval, I. V., Sibirny, A., Subramani, S., Thumm, M., Veenhuis, M., & Ohsumi, Y. (2003). A Unified Nomenclature for Yeast Autophagy-Related Genes. *Developmental Cell*, *5*(4), 539–545. [https://doi.org/10.1016/S1534-5807\(03\)00296-X](https://doi.org/10.1016/S1534-5807(03)00296-X)
- Kocaturk, N. M., Akkoc, Y., Kig, C., Bayraktar, O., Gozuacik, D., & Kutlu, O. (2019). Autophagy as a molecular target for cancer treatment. *European Journal of Pharmaceutical Sciences*, *134*, 116–137. <https://doi.org/10.1016/j.ejps.2019.04.011>
- Kocaturk, N. M., & Gozuacik, D. (2018). Crosstalk Between Mammalian Autophagy and the Ubiquitin-Proteasome System. *Frontiers in Cell and Developmental Biology*, *6*, 128. <https://doi.org/10.3389/fcell.2018.00128>
- Kohl, A., Giese, K. P., Mohajeri, M. H., Montag, D., Moos, M., & Schachner, M. (1992). Analysis of promoter activity and 5' genomic structure of the neural cell adhesion molecule L1. *Journal of Neuroscience Research*, *32*(2), 167–177. <https://doi.org/10.1002/jnr.490320206>
- Komatsu, M., Waguri, S., Ueno, T., Iwata, J., Murata, S., Tanida, I., Ezaki, J., Mizushima, N., Ohsumi, Y., Uchiyama, Y., Kominami, E., Tanaka, K., & Chiba, T. (2005). Impairment of starvation-induced and constitutive autophagy in Atg7-deficient mice. *Journal of Cell Biology*, *169*(3), 425–434. <https://doi.org/10.1083/jcb.200412022>
- Konar, A., Kumar, A., Maloney, B., Lahiri, D. K., & Thakur, M. K. (2018). A serine protease KLK8 emerges as a regulator of regulators in memory: Microtubule protein dependent neuronal morphology and PKA-CREB signaling. *Scientific Reports*, *8*(1), 1–9. <https://doi.org/10.1038/s41598-018-27640-6>

- Kondo, T., Imamura, K., Funayama, M., Tsukita, K., Miyake, M., Ohta, A., Woltjen, K., Nakagawa, M., Asada, T., Arai, T., Kawakatsu, S., Izumi, Y., Kaji, R., Iwata, N., & Inoue, H. (2017). iPSC-Based Compound Screening and In Vitro Trials Identify a Synergistic Anti-amyloid β Combination for Alzheimer's Disease. *Cell Reports*, *21*(8), 2304–2312. <https://doi.org/10.1016/j.celrep.2017.10.109>
- Kreitzer, F. R., Salomonis, N., Sheehan, A., Huang, M., Park, J. S., Spindler, M. J., Lizarraga, P., Weiss, W. A., So, P. L., & Conklin, B. R. (2013). A robust method to derive functional neural crest cells from human pluripotent stem cells. *American Journal of Stem Cells*, *2*(2), 119–131. [/pmc/articles/PMC3708511/](https://doi.org/10.1002/ajsc.12001)
- Lacoste, S., Wiechec, E., dos Santos Silva, A. G., Guffei, A., Williams, G., Lowbeer, M., Benedek, K., Henriksson, M., Klein, G., & Mai, S. (2010). Chromosomal rearrangements after ex vivo Epstein–Barr virus (EBV) infection of human B cells. *Oncogene*, *29*(4), 503–515. <https://doi.org/10.1038/onc.2009.359>
- Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain*, *8*(6), 221–223. <https://doi.org/10.1093/bjaceaccp/mkn041>
- Lapierre, L. R., Kumsta, C., Sandri, M., Ballabio, A., & Hansen, M. (2015). Transcriptional and epigenetic regulation of autophagy in aging. *Autophagy*, *11*(6), 867–880. <https://doi.org/10.1080/15548627.2015.1034410>
- Lasky, J. L., & Wu, H. (2005). Notch signaling, brain development, and human disease. *Pediatric Research*, *57*(5 Pt 2), 104R-109R. <https://doi.org/10.1203/01.PDR.0000159632.70510.3D>
- Ledford, H. (2020). CRISPR treatment inserted directly into the body for first time. *Nature*, *579*(7798), 185–185. <https://doi.org/10.1038/d41586-020-00655-8>
- Lee, J.-M., Correia, K., Loupe, J., Kim, K.-H., Barker, D., Hong, E. P., Chao, M. J., Long, J. D., Lucente, D., Vonsattel, J. P. G., Pinto, R. M., Abu Elneel, K., Ramos, E. M., Mysore, J. S., Gillis, T., Wheeler, V. C., MacDonald, M. E., Gusella, J. F., McAllister, B., ... Myers, R. H. (2019). CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell*, *178*(4), 887-900.e14. <https://doi.org/10.1016/j.cell.2019.06.036>
- Lee, J. M., Wagner, M., Xiao, R., Kim, K. H., Feng, D., Lazar, M. A., & Moore, D. D. (2014). Nutrient-sensing nuclear receptors coordinate autophagy. *Nature*, *516*(729), 112–115. <https://doi.org/10.1038/nature13961>
- Lee, M. K., Tuttle, J. B., Rebhun, L. I., Cleveland, D. W., & Frankfurter, A. (1990). The expression and posttranslational modification of a neuron-specific β -tubulin isotype during chick embryogenesis. *Cell Motility and the Cytoskeleton*, *17*(2), 118–132. <https://doi.org/10.1002/cm.970170207>

- Leil, T. A. (2004). GABAA Receptor-Associated Protein Traffics GABAA Receptors to the Plasma Membrane in Neurons. *Journal of Neuroscience*, *24*(50), 11429–11438. <https://doi.org/10.1523/JNEUROSCI.3355-04.2004>
- Levine, B., & Kroemer, G. (2019). Biological Functions of Autophagy Genes: A Disease Perspective. *Cell*, *176*(1–2), 11–42. <https://doi.org/10.1016/j.cell.2018.09.048>
- Li, L., Liu, H., Dong, P., Li, D., Legant, W. R., Grimm, J. B., Lavis, L. D., Betzig, E., Tjian, R., & Liu, Z. (2016). Real-time imaging of Huntingtin aggregates diverting target search and gene transcription. *ELife*, *5*. <http://dx.doi.org/10.7554/elife.17056>
- Li, Mei, Miwa, S., Kobayashi, Y., Merry, D. E., Yamamoto, M., Tanaka, F., Doyu, M., Hashizume, Y., Fischbeck, K. H., & Sobue, G. (1998). Nuclear inclusions of the androgen receptor protein in spinal and bulbar muscular atrophy. *Annals of Neurology*, *44*(2), 249–254. <https://doi.org/10.1002/ana.410440216>
- Li, Min, Hou, Y., Wang, J., Chen, X., Shao, Z. M., & Yin, X. M. (2011). Kinetics comparisons of mammalian Atg4 homologues indicate selective preferences toward diverse Atg8 substrates. *Journal of Biological Chemistry*, *286*(9), 7327–7338. <https://doi.org/10.1074/jbc.M110.199059>
- Li, X.-J., Hu, B.-Y., Jones, S. A., Zhang, Y.-S., LaVaute, T., Du, Z.-W., & Zhang, S.-C. (2008). Directed Differentiation of Ventral Spinal Progenitors and Motor Neurons from Human Embryonic Stem Cells by Small Molecules. *Stem Cells*, *26*(4), 886–893. <https://doi.org/10.1634/stemcells.2007-0620>
- Liao, Y.-C., Fernandopulle, M. S., Wang, G., Choi, H., Hao, L., Drerup, C. M., Patel, R., Qamar, S., Nixon-Abell, J., Shen, Y., Meadows, W., Vendruscolo, M., Kowles, T. P. J., Nelson, M., Czekalska, M. A., Musteikyte, G., Gachechiladze, M. A., Stephens, C. A., Pasolli, H. A., Forrest, L. R., St George-Hyslop, P., Lippincott-Schwartz, J., Ward, M. E. (2019). RNA Granules Hitchhike on Lysosomes for Long-Distance Transport, Using Annexin A11 as a Molecular Tether. *Cell*, *179*(1): 147-164.e20. <https://doi.org/10.1016/j.cell.2019.08.050>
- Lin, C.-Y., Yoshida, M., Li, L.-T., Ikenaka, A., Oshima, S., Nakagawa, K., Sakurai, H., Matsui, E., Nakahata, T., & Saito, M. K. (2019). iPSC-derived functional human neuromuscular junctions model the pathophysiology of neuromuscular diseases. *JCI Insight*, *4*(18). <https://doi.org/10.1172/jci.insight.124299>
- Lin, H.-C., He, Z., Ebert, S., Schörnig, M., Santel, M., Weigert, A., Hevers, W., Kasri, N. N., Taverna, E., Camp, J. G., & Treutlein, B. (2020). Ngn2 induces diverse neuronal lineages from human pluripotency. *BioRxiv*, 2020.11.19.389445. <https://doi.org/10.1101/2020.11.19.389445>
- Lin, X. X., Sen, I., Janssens, G. E., Zhou, X., Fonslow, B. R., Edgar, D., Stroustrup, N., Swoboda, P., Yates, J. R., Ruvkun, G., & Riedel, C. G. (2018). DAF-

- 16/FOXO and HLH-30/TFEB function as combinatorial transcription factors to promote stress resistance and longevity. *Nature Communications*, 9(1), 1–15.
<https://doi.org/10.1038/s41467-018-06624-0>
- Lin, Y. C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van De Peer, Y., Tavernier, J., & Callewaert, N. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications*, 5(1), 12.
<https://doi.org/10.1038/ncomms5767>
- Ling, X., Xie, B., Gao, X., Chang, L., Zheng, W., Chen, H., Huang, Y., Tan, L., Li, M., & Liu, T. (2020). Improving the efficiency of precise genome editing with site-specific Cas9-oligonucleotide conjugates. *Science Advances*, 6(15), eaaz0051.
<https://doi.org/10.1126/sciadv.aaz0051>
- Liscic, R. M., Alberici, A., Cairns, N. J., Romano, M., & Buratti, E. (2020). From basic research to the clinic: innovative therapies for ALS and FTD in the pipeline. *Molecular Neurodegeneration* 2020 15:1, 15(1), 1–17.
<https://doi.org/10.1186/S13024-020-00373-9>
- Liu, Zhen, Lu, Z., Yang, G., Huang, S., Li, G., Feng, S., Liu, Y., Li, J., Yu, W., Zhang, Y., Chen, J., Sun, Q., & Huang, X. (2018). Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-04768-7>
- Liu, Ziqing, Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., & Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports*, 7(1), 2193.
<https://doi.org/10.1038/s41598-017-02460-2>
- Ludwig, C., & Kampmann, M. (2017). *iNeuron RNA-Seq*.
<https://kampmannlab.ucsf.edu/ineuron-rna-seq>
- Ludwig, T., & Thomson, J. (2007). Defined, Feeder-Independent Medium for Human Embryonic Stem Cell Culture. *Current Protocols in Stem Cell Biology*, 2(1), 1C.2.1-1C.2.16. <https://doi.org/10.1002/9780470151808.sc01c02s2>
- Maceneaney, P. M., & Malone, D. E. (2000). The meaning of diagnostic test results: A spreadsheet for swift data analysis. *Clinical Radiology*, 55(3), 227–235.
<https://doi.org/10.1053/crad.1999.0444>
- Macrí-Pellizzeri, L., Pelacho, B., Sancho, A., Iglesias-García, O., Simón-Yarza, A. M., Soriano-Navarro, M., González-Granero, S., García-Verdugo, J. M., De-Juan-Pardo, E. M., & Prosper, F. (2015). Substrate Stiffness and Composition Specifically Direct Differentiation of Induced Pluripotent Stem Cells. *Tissue Engineering Part A*, 21(9–10), 1633–1641.
<https://doi.org/10.1089/ten.tea.2014.0251>

- Maden, M. (2002). Retinoid signalling in the development of the central nervous system. *Nature Reviews Neuroscience*, *3*(11), 843–853. <https://doi.org/10.1038/nrn963>
- Madrid, L. V., Mayo, M. W., Reuther, J. Y., & Baldwin, A. S. (2001). Akt Stimulates the Transactivation Potential of the RelA/p65 Subunit of NF- κ B through Utilization of the I κ B Kinase and Activation of the Mitogen-activated Protein Kinase p38. *Journal of Biological Chemistry*, *276*(22), 18934–18940. <https://doi.org/10.1074/jbc.M101103200>
- Mammucari, C., Milan, G., Romanello, V., Masiero, E., Rudolf, R., Del Piccolo, P., Burden, S. J., Di Lisi, R., Sandri, C., Zhao, J., Goldberg, A. L., Schiaffino, S., & Sandri, M. (2007). FoxO3 Controls Autophagy in Skeletal Muscle In Vivo. *Cell Metabolism*, *6*(6), 458–471. <https://doi.org/10.1016/j.cmet.2007.11.001>
- Mandrusiak, L. M. (2003). Transglutaminase potentiates ligand-dependent proteasome dysfunction induced by polyglutamine-expanded androgen receptor. *Human Molecular Genetics*, *12*(13), 1497–1506. <https://doi.org/10.1093/hmg/ddg161>
- Mansuy, V., Boireau, W., Fraichard, A., Schlick, J.-L., Jouvenot, M., & Delage-Mourroux, R. (2004). GEC1, a protein related to GABARAP, interacts with tubulin and GABAA receptor. *Biochemical and Biophysical Research Communications*, *325*(2), 639–648. <https://doi.org/10.1016/j.bbrc.2004.10.072>
- Maretzky, T., Schulte, M., Ludwig, A., Rose-John, S., Blobel, C., Hartmann, D., Altevogt, P., Saftig, P., & Reiss, K. (2005). L1 Is Sequentially Processed by Two Differently Activated Metalloproteases and Presenilin/ -Secretase and Regulates Neural Cell Adhesion, Cell Migration, and Neurite Outgrowth. *Molecular and Cellular Biology*, *25*(20), 9040–9053. <https://doi.org/10.1128/mcb.25.20.9040-9053.2005>
- Maten, M. van der, Reijnen, C., Pijnenborg, J. M. A., & Zegers, M. M. (2019). L1 cell adhesion molecule in cancer, a systematic review on domain-specific functions. In *International Journal of Molecular Sciences* (Vol. 20, Issue 17, p. 4180). MDPI AG. <https://doi.org/10.3390/ijms20174180>
- Matheson, N. J., Peden, A. A., Lehner, P. J., Kiefer, T., & Monie, D. (2014). Antibody-Free Magnetic Cell Sorting of Genetically Modified Primary Human CD4⁺ T Cells by One-Step Streptavidin Affinity Purification. *PLoS ONE*, *9*(10), e111437. <https://doi.org/10.1371/journal.pone.0111437>
- Mathieson, T., Franken, H., Kosinski, J., Kurzawa, N., Zinn, N., Sweetman, G., Poeckel, D., Ratnu, V. S., Schramm, M., Becher, I., Steidel, M., Noh, K. M., Bergamini, G., Beck, M., Bantscheff, M., & Savitski, M. M. (2018). Systematic analysis of protein turnover in primary cells. *Nature Communications*, *9*(1), 1–10. <https://doi.org/10.1038/s41467-018-03106-1>

- Matsuda, T., & Cepko, C. L. (2007). Controlled expression of transgenes introduced by in vivo electroporation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(3), 1027–1032.
<https://doi.org/10.1073/pnas.0610155104>
- Mazzoni, E. O., Mahony, S., Closser, M., Morrison, C. A., Nedelec, S., Williams, D. J., An, D., Gifford, D. K., & Wichterle, H. (2013). Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nature Neuroscience*, *16*(9), 1219–1229.
<https://www.nature.com/neuro/journal/v16/n9/pdf/nn.3467.pdf>
- Mazzoni, E. O., Mahony, S., Iacovino, M., Morrison, C. A., Mountoufaris, G., Closser, M., Whyte, W. A., Young, R. A., Kyba, M., Gifford, D. K., & Wichterle, H. (2011). Embryonic stem cell-based mapping of developmental transcriptional programs. *Nature Methods*, *8*(12), 1056–1058.
<https://doi.org/10.1038/nmeth.1775>
- McTague, A., Rossignoli, G., Ferrini, A., Barral, S., & Kurian, M. A. (2021). Genome Editing in iPSC-Based Neural Systems: From Disease Models to Future Therapeutic Strategies. *Frontiers in Genome Editing*, *3*, 630600.
<https://doi.org/10.3389/fgeed.2021.630600>
- Menzies, F. M., Fleming, A., & Rubinsztein, D. C. (2015). Compromised autophagy and neurodegenerative diseases. *Nature Reviews Neuroscience*, *16*(6), 345–357.
<https://doi.org/10.1038/nrn3961>
- Metzakopian, E., Strong, A., Iyer, V., Hodgkins, A., Tzelepis, K., Antunes, L., Friedrich, M. J., Kang, Q., Davidson, T., Lamberth, J., Hoffmann, C., Davis, G. D., Vassiliou, G. S., Skarnes, W. C., & Bradley, A. (2017). Enhancing the genome editing toolbox: Genome wide CRISPR arrayed libraries. *Scientific Reports*, *7*(1), 1–9. <https://doi.org/10.1038/s41598-017-01766-5>
- Mikkelsen, T. S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B. E., Jaenisch, R., Lander, E. S., & Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. *Nature*, *454*(7200), 49–55. <https://doi.org/10.1038/nature07056>
- Milone, M. C., & O’Doherty, U. (2018). Clinical use of lentiviral vectors. *Leukemia*, *32*(7), 1529–1541. <https://doi.org/10.1038/s41375-018-0106-0>
- Miltenyi, S., & Schmitz, J. (2000). High Gradient Magnetic Cell Sorting. In *Flow Cytometry and Cell Sorting* (pp. 218–247). Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-662-04129-1_20
- Miyaoka, Y., Berman, J. R., Cooper, S. B., Mayerl, S. J., Chan, A. H., Zhang, B., Karlin-Neumann, G. A., & Conklin, B. R. (2016). Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Scientific Reports*, *6*(1), 1–12. <https://doi.org/10.1038/srep23549>

- Mizushima, N. (2007). Autophagy: process and function. In *Genes and Development* (Vol. 21, Issue 22, pp. 2861–2873). Cold Spring Harbor Laboratory Press.
<https://doi.org/10.1101/gad.1599207>
- Mizushima, N., Sugita, H., Yoshimori, T., & Ohsumi, Y. (1998). A new protein conjugation system in human: The counterpart of the yeast Apg12p conjugation system essential for autophagy. *Journal of Biological Chemistry*, *273*(51), 33889–33892. <https://doi.org/10.1074/jbc.273.51.33889>
- Molugu, K., Harkness, T. T., Carlson-Stevermer, J., Prestil, R., Piscopo, N. J. N. J., Seymour, S. K. S. K., Knight, G. T. G. T., Ashton, R. S. R. S., & Saha, K. (2020). Tracking and Predicting Human Somatic Cell Reprogramming Using Nuclear Characteristics. *Biophysical Journal*, *118*(9), 2086–2102.
<https://doi.org/10.1016/j.bpj.2019.10.014>
- Monks, D. A., Johansen, J. A., Mo, K., Rao, P., Eagleson, B., Yu, Z., Lieberman, A. P., Breedlove, S. M., & Jordan, C. L. (2007). Overexpression of wild-type androgen receptor in muscle recapitulates polyglutamine disease. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 18259–18264. <https://doi.org/10.1073/pnas.0705501104>
- Montie, H. L. I., Cho, M. S., Holder, L., Liu, Y., Tsvetkov, A. S., Finkbeiner, S., & Merry, D. E. (2009). Cytoplasmic retention of polyglutamine-expanded androgen receptor ameliorates disease via autophagy in a mouse model of spinal and bulbar muscular atrophy. *Human Molecular Genetics*, *18*(11), 1937–1950.
<https://doi.org/10.1093/hmg/ddp115>
- Moore, C. B., Guthrie, E. H., Huang, M. T. H., & Taxman, D. J. (2010). Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown. *Methods in Molecular Biology (Clifton, N.J.)*, *629*, 141–158.
https://doi.org/10.1007/978-1-60761-657-3_10
- Mossman, D., & Berger, J. O. (2001). Intervals for Posttest Probabilities: A Comparison of 5 Methods. *Medical Decision Making*, *21*(6), 498–507.
<https://doi.org/10.1177/0272989x0102100608>
- Mullard, A. (2021). Landmark Alzheimer’s drug approval confounds research community. *Nature*, *594*(7863), 309–310. <https://doi.org/10.1038/d41586-021-01546-2>
- Muñoz-Sanjuán, I., & Brivanlou, A. H. (2002). Neural induction, the default model and embryonic stem cells. *Nature Reviews Neuroscience*, *3*(4), 271–280.
<https://doi.org/10.1038/nrn786>
- Nagaraj, K., & Hortsch, M. (2006). Phosphorylation of L1-type cell-adhesion molecules – ankyrins away! *Trends in Biochemical Sciences*, *31*(10), 544–546.
<https://doi.org/10.1016/j.tibs.2006.07.010>

- Nakatogawa, H. (2020). Mechanisms governing autophagosome biogenesis. *Nature Reviews Molecular Cell Biology*, 21(8), 439–458. <https://doi.org/10.1038/s41580-020-0241-0>
- Nalavade, R., Griesche, N., Ryan, D. P., Hildebrand, S., & Krauß, S. (2013). Mechanisms of RNA-induced toxicity in CAG repeat disorders. *Cell Death & Disease*, 4(8), e752–e752. <https://doi.org/10.1038/cddis.2013.276>
- Narain, Y., Wyttenbach, A., Rankin, J., Furlong, R. A., & Rubinsztein, D. C. (1999). A molecular investigation of true dominance in Huntington's disease. *Journal of Medical Genetics*, 36(10), 739–746. <https://doi.org/10.1136/jmg.36.10.739>
- Narhi, L., Wood, S. J., Steavenson, S., Jiang, Y., Wu, G. M., Anafi, D., Kaufman, S. A., Martin, F., Sitney, K., Denis, P., Louis, J. C., Wypych, J., Biere, A. L., & Citron, M. (1999). Both familial Parkinson's disease mutations accelerate α -synuclein aggregation. *Journal of Biological Chemistry*, 274(14), 9843–9846. <https://doi.org/10.1074/jbc.274.14.9843>
- Nedelsky, N. B., Pennuto, M., Smith, R. B., Palazzolo, I., Moore, J., Nie, Z., Neale, G., & Taylor, J. P. (2010). Native functions of the androgen receptor are essential to pathogenesis in a drosophila model of spinobulbar muscular atrophy. *Neuron*, 67(6), 936–952. <https://doi.org/10.1016/j.neuron.2010.08.034>
- Nguyen, T. N., Padman, B. S., Usher, J., Oorschot, V., Ramm, G., & Lazarou, M. (2016). Atg8 family LC3/GABARAP proteins are crucial for autophagosome–lysosome fusion but not autophagosome formation during PINK1/Parkin mitophagy and starvation. *Journal of Cell Biology*, 215(6), 857–874. <https://doi.org/10.1083/jcb.201607039>
- Nieto-Torres, J. L., Leidal, A. M., Debnath, J., & Hansen, M. (2021a). Beyond Autophagy: The Expanding Roles of ATG8 Proteins. *Trends in Biochemical Sciences*, 46(8), 673–686. <https://doi.org/10.1016/j.tibs.2021.01.004>
- Nieto-Torres, J. L., Shanahan, S.-L., Chassefeyre, R., Chaiamarit, T., Zaretski, S., Landeras-Bueno, S., Verhelle, A., Encalada, S. E., & Hansen, M. (2021b). LC3B phosphorylation regulates FYCO1 binding and directional transport of autophagosomes. *Current Biology*. <https://doi.org/10.1016/j.cub.2021.05.052>
- Nishida, Y., Arakawa, S., Fujitani, K., Yamaguchi, H., Mizuta, T., Kanaseki, T., Komatsu, M., Otsu, K., Tsujimoto, Y., & Shimizu, S. (2009). Discovery of Atg5/Atg7-independent alternative macroautophagy. *Nature*, 461(7264), 654–658. <https://doi.org/10.1038/nature08455>
- Nishimaru, H., Restrepo, C. E., Ryge, J., Yanagawa, Y., & Kiehn, O. (2005). Mammalian motor neurons corelease glutamate and acetylcholine at central synapses. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14), 5245–5249. <https://doi.org/10.1073/pnas.0501331102>

- Oakes, M. B., Eyvazzadeh, A. D., Quint, E., & Smith, Y. R. (2008). Complete Androgen Insensitivity Syndrome—A Review. *Journal of Pediatric and Adolescent Gynecology*, *21*(6), 305–310. <https://doi.org/10.1016/j.jpag.2007.09.006>
- Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K. I., Shibata, T., Kunisada, T., Takahashi, M., Takahashi, J., Saji, H., & Yamanaka, S. (2011). A more efficient method to generate integration-free human iPS cells. *Nature Methods*, *8*(5), 409–412. <https://doi.org/10.1038/nmeth.1591>
- Oku, M., & Sakai, Y. (2018). Three Distinct Types of Microautophagy Based on Membrane Dynamics and Molecular Machineries. *BioEssays*, *40*(6), 1800008. <https://doi.org/10.1002/bies.201800008>
- Orr, H. T., & Zoghbi, H. Y. (2007). Trinucleotide Repeat Disorders. *Annual Review of Neuroscience*, *30*(1), 575–621. <https://doi.org/10.1146/annurev.neuro.29.051605.113042>
- Palazzolo, I., Gliozzi, A., Rusmini, P., Sau, D., Crippa, V., Simonini, F., Onesto, E., Bolzoni, E., & Poletti, A. (2008). The role of the polyglutamine tract in androgen receptor. *Journal of Steroid Biochemistry and Molecular Biology*, *108*(3–5), 245–253. <https://doi.org/10.1016/j.jsbmb.2007.09.016>
- Pan, J., Ma, N., Yu, B., Zhang, W., & Wan, J. (2020). Transcriptomic profiling of microglia and astrocytes throughout aging. *Journal of Neuroinflammation*, *17*(1), 1–19. <https://doi.org/10.1186/s12974-020-01774-9>
- Pang, Z. P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D. R., Yang, T. Q., Citri, A., Sebastiano, V., Marro, S., Südhof, T. C., & Wernig, M. (2011). Induction of human neuronal cells by defined transcription factors. *Nature*, *476*(7359), 220–223. <https://doi.org/10.1038/nature10202>
- Pardo, B., Gómez-González, B., & Aguilera, A. (2009). DNA Repair in Mammalian Cells. *Cellular and Molecular Life Sciences*, *66*(6), 1039–1056. <https://doi.org/10.1007/s00018-009-8740-3>
- Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, *56*(1), 45–50. <https://doi.org/10.4103/0301-4738.37595>
- Park, I. H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M. W., Cowan, C., Hochedlinger, K., & Daley, G. Q. (2008a). Disease-Specific Induced Pluripotent Stem Cells. *Cell*, *134*(5), 877–886. <https://doi.org/10.1016/j.cell.2008.07.041>
- Park, I. H., Zhao, R., West, J. A., Yabuuchi, A., Huo, H., Ince, T. A., Lerou, P. H.,

- Lensch, M. W., & Daley, G. Q. (2008b). Reprogramming of human somatic cells to pluripotency with defined factors. *Nature*, *451*(7175), 141–146. <https://doi.org/10.1038/nature06534>
- Patani, R. (2016). Generating Diverse Spinal Motor Neuron Subtypes from Human Pluripotent Stem Cells. *Stem Cells International*, *2016*, 1–13. <https://doi.org/10.1155/2016/1036974>
- Pengo, N., Agrotis, A., Prak, K., Jones, J., & Ketteler, R. (2017). A reversible phospho-switch mediated by ULK1 regulates the activity of autophagy protease ATG4B. *Nature Communications*, *8*(1), 294. <https://doi.org/10.1038/s41467-017-00303-2>
- Pennuto, M., & Basso, M. (2016). In Vitro and In Vivo Modeling of Spinal and Bulbar Muscular Atrophy. *Journal of Molecular Neuroscience*, *58*(3), 365–373. <https://doi.org/10.1007/s12031-015-0677-4>
- Pérez-Hernández, M., Arias, A., Martínez-García, D., Pérez-Tomás, R., Quesada, R., & Soto-Cerrato, V. (2019). Targeting Autophagy for Cancer Treatment and Tumor Chemosensitization. *Cancers*, *11*(10), 1599. <https://doi.org/10.3390/cancers11101599>
- Perez, M. K., Paulson, H. L., Pendse, S. J., Saionz, S. J., Bonini, N. M., & Pittman, R. N. (1998). Recruitment and the Role of Nuclear Localization in Polyglutamine-mediated Aggregation. *Journal of Cell Biology*, *143*(6), 1457–1470. <https://doi.org/10.1083/jcb.143.6.1457>
- Petri, S., & Meister, G. (2013). siRNA design principles and off-target effects. *Methods in Molecular Biology*, *986*, 59–71. https://doi.org/10.1007/978-1-62703-311-4_4
- Petros, T. J., Tyson, J. A., & Anderson, S. A. (2011). Pluripotent Stem Cells for the Study of CNS Development. *Frontiers in Molecular Neuroscience*, *4*, 30. <https://doi.org/10.3389/fnmol.2011.00030>
- Pezzi, H. M., Niles, D. J., Schehr, J. L., Beebe, D. J., & Lang, J. M. (2018). Integration of Magnetic Bead-Based Cell Selection into Complex Isolations. *ACS Omega*, *3*(4), 3908–3917. <https://doi.org/10.1021/acsomega.7b01427>
- Pfeifer, M., Schirmer, U., Geismann, C., Schäfer, H., Sebens, S., & Altevogt, P. (2010). L1CAM expression in endometrial carcinomas is regulated by usage of two different promoter regions. *BMC Molecular Biology*, *11*(1), 64. <https://doi.org/10.1186/1471-2199-11-64>
- Pierfelice, T. J., Schreck, K. C., Eberhart, C. G., & Gaiano, N. (2008). Notch, Neural Stem Cells, and Brain Tumors. *Cold Spring Harbor Symposia on Quantitative Biology*, *73*, 367–375. <https://doi.org/10.1101/sqb.2008.73.013>
- Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and Functions of Long

- Noncoding RNAs. *Cell*, 136(4), 629–641.
<https://doi.org/10.1016/j.cell.2009.02.006>
- Poplawski, G. H. D., Tranziska, A. K., Leshchyns'ka, I., Meier, I. D., Streichert, T., Sytnyk, V., & Schachner, M. (2012). L1CAM increases MAP2 expression via the MAPK pathway to promote neurite outgrowth. *Molecular and Cellular Neuroscience*, 50(2), 169–178. <https://doi.org/10.1016/j.mcn.2012.03.010>
- Popp, B., Krumbiegel, M., Grosch, J., Sommer, A., Uebe, S., Kohl, Z., Plötz, S., Farrell, M., Trautmann, U., Kraus, C., Ekici, A. B., Asadollahi, R., Regensburger, M., Günther, K., Rauch, A., Edenhofer, F., Winkler, J., Winner, B., & Reis, A. (2018). Need for high-resolution Genetic Analysis in iPSC: Results and Lessons from the ForIPS Consortium. *Scientific Reports*, 8(1), 1–14. <https://doi.org/10.1038/s41598-018-35506-0>
- Pourshafie, N., Masati, E., Bunker, E., Nickolls, A. R., Thepmankorn, P., Johnson, K., Feng, X., Ekins, T., Grunseich, C., & Fischbeck, K. H. (2020). Linking epigenetic dysregulation, mitochondrial impairment, and metabolic dysfunction in SBMA motor neurons. *JCI Insight*, 5(13).
<https://doi.org/10.1172/jci.insight.136539>
- Prasad, A., Manivannan, J., Loong, D. T. B., Chua, S. M., Gharibani, P. M., & All, A. H. (2016). A review of induced pluripotent stem cell, direct conversion by trans-differentiation, direct reprogramming and oligodendrocyte differentiation. *Regenerative Medicine*, 11(2), 181–191. <https://doi.org/10.2217/rme.16.5>
- Prestil, R. (2020). Human germline gene editing needs global regulation. *Cambridge Journal of Science and Policy*, 1(2). <https://doi.org/10.17863/CAM.51719>
- Rackham, O. J. L., Firas, J., Fang, H., Oates, M. E., Holmes, M. L., Knaupp, A. S., Suzuki, H., Nefzger, C. M., Daub, C. O., Shin, J. W., Petretto, E., Forrest, A. R. R., Hayashizaki, Y., Polo, J. M., & Gough, J. (2016). A predictive computational framework for direct reprogramming between human cell types. *Nature Genetics*, 48(3), 331–335. <https://doi.org/10.1038/ng.3487>
- Ramanathan, R., Wilkemeyer, M. F., Mittal, B., Perides, G., & Charness, M. E. (1996). Alcohol inhibits cell-cell adhesion mediated by human L1. *Journal of Cell Biology*, 133(2), 381–390. <https://doi.org/10.1083/jcb.133.2.381>
- Ramos, D. M., Skarnes, W. C., Singleton, A. B., Cookson, M. R., & Ward, M. E. (2021). Tackling neurodegenerative diseases with genomic engineering: A new stem cell initiative from the NIH. *Neuron*, 109(7), 1080–1083.
<https://doi.org/10.1016/j.neuron.2021.03.022>
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), 2281–2308. <https://doi.org/10.1038/nprot.2013.143>

- Ravikumar, B., Vacher, C., Berger, Z., Davies, J. E., Luo, S., Oroz, L. G., Scaravilli, F., Easton, D. F., Duden, R., O’Kane, C. J., & Rubinsztein, D. C. (2004). Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nature Genetics*, *36*(6), 585–595. <https://doi.org/10.1038/ng1362>
- Redwine, J. M., & Evans, C. F. (2002). Markers of Central Nervous System Glia and Neurons In Vivo During Normal and Pathological Conditions. In *Current Topics in Microbiology and Immunology* (Vol. 265, pp. 119–140). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-09525-6_6
- Reid, R. A., & Hemperly, J. J. (1992). Variants of human L1 cell adhesion molecule arise through alternate splicing of RNA. *Journal of Molecular Neuroscience*, *3*(3), 127–135. <https://doi.org/10.1007/BF02919404>
- Riedle, S., Kiefel, H., Gast, D., Bondong, S., Wolterink, S., Gutwein, P., & Altevogt, P. (2009). Nuclear translocation and signalling of L1-CAM in human carcinoma cells requires ADAM10 and presenilin/ γ -secretase activity. *Biochemical Journal*, *420*(3), 391–402. <https://doi.org/10.1042/BJ20081625>
- Roelink, H., Porter, J. ., Chiang, C., Tanabe, Y., Chang, D. ., Beachy, P. ., & Jessell, T. . (1995). Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. *Cell*, *81*(3), 445–455. [https://doi.org/10.1016/0092-8674\(95\)90397-6](https://doi.org/10.1016/0092-8674(95)90397-6)
- Rogov, V., Dötsch, V., Johansen, T., & Kirkin, V. (2014). Interactions between Autophagy Receptors and Ubiquitin-like Proteins Form the Molecular Basis for Selective Autophagy. *Molecular Cell*, *53*(2), 167–178. <https://doi.org/10.1016/j.molcel.2013.12.014>
- Rossi, G., Manfrin, A., & Lutolf, M. P. (2018). Progress and potential in organoid research. *Nature Reviews Genetics*, *19*(11), 671–687. <https://doi.org/10.1038/s41576-018-0051-9>
- Rothstein, J., Berry, J., Svendsen, C., Thompson, L., Finkbeiner, S., Eyk, J. Van, Fraenkel, E., Cudkowicz, M., Maragakis, N., Sareen, D., Norel, R., Dardov, V., Coyne, A., Frank, A., Matlock, A., Pandey, R., Vibhav, V., Lima, L., Wu, J., ... ALS, A. (2020). Answer ALS: A Large-Scale Resource for Sporadic and Familial ALS Combining Clinical Data with Multi-Omics Data from Induced Pluripotent Cell Lines. *Research Square*. <https://doi.org/10.21203/rs.3.rs-96858/v1>
- Rouet, P., Smih, F., & Jasin, M. (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(13), 6064–6068. <https://doi.org/10.1073/pnas.91.13.6064>
- Rowe, R. G., & Daley, G. Q. (2019). Induced pluripotent stem cells in disease

- modelling and drug discovery. *Nature Reviews Genetics*, 20(7), 377–388.
<https://doi.org/10.1038/s41576-019-0100-z>
- Rui, Y. N., Xu, Z., Patel, B., Chen, Z., Chen, D., Tito, A., David, G., Sun, Y., Stimming, E. F., Bellen, H. J., Cuervo, A. M., & Zhang, S. (2015). Huntingtin functions as a scaffold for selective macroautophagy. *Nature Cell Biology*, 17(3), 262–275. <https://doi.org/10.1038/ncb3101>
- Ryan, M. D., King, A. M. Q., & Thomas, G. P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *Journal of General Virology*, 72(11), 2727–2732.
<https://doi.org/10.1099/0022-1317-72-11-2727>
- Sagiv, Y., Legesse-Miller, A., Porat, A., & Elazar, Z. (2000). GATE-16, a membrane transport modulator, interacts with NSF and the Golgi v-SNARE GOS-28. *The EMBO Journal*, 19(7), 1494–1504. <https://doi.org/10.1093/EMBOJ/19.7.1494>
- Saha, K., & Jaenisch, R. (2009). Technical Challenges in Using Human Induced Pluripotent Stem Cells to Model Disease. *Cell Stem Cell*, 5(6), 584–595.
<https://doi.org/10.1016/j.stem.2009.11.009>
- Salazar-Grueso, E. F., Kim, S., & Kim, H. (1991). Embryonic mouse spinal cord motor neuron hybrid cells. *NeuroReport*, 2(9), 505–508.
<https://doi.org/10.1097/00001756-199109000-00002>
- Salick, M. R., Lubeck, E., Riesselman, A., & Kaykas, A. (2021). The future of cerebral organoids in drug discovery. *Seminars in Cell & Developmental Biology*, 111, 67–73. <https://doi.org/10.1016/j.semcd.2020.05.024>
- Sambataro, F., & Pennuto, M. (2017). Post-translational Modifications and Protein Quality Control in Motor Neuron and Polyglutamine Diseases. *Frontiers in Molecular Neuroscience*, 10, 82. <https://doi.org/10.3389/fnmol.2017.00082>
- Sar, M., & Stumpf, W. E. (1977). Androgen concentration in motor neurons of cranial nerves and spinal cord. *Science*, 197(4298), 77–79.
<https://doi.org/10.1126/science.867053>
- Scaramuzzino, C., Casci, I., Parodi, S., Lievens, P. M. J., Polanco, M. J., Milioto, C., Chivet, M., Monaghan, J., Mishra, A., Badders, N., Aggarwal, T., Grunseich, C., Sambataro, F., Basso, M., Fackelmayer, F. O., Taylor, J. P., Pandey, U. B., & Pennuto, M. (2015). Protein Arginine Methyltransferase 6 Enhances Polyglutamine-Expanded Androgen Receptor Function and Toxicity in Spinal and Bulbar Muscular Atrophy. *Neuron*, 85(1), 88–100.
<https://doi.org/10.1016/j.neuron.2014.12.031>
- Schaefer, A. W., Kamiguchi, H., Wong, E. V., Beach, C. M., Landreth, G., & Lemmon, V. (1999). Activation of the MAPK signal cascade by the neural cell adhesion molecule L1 requires L1 internalization. *Journal of Biological*

- Chemistry*, 274(53), 37965–37973. <https://doi.org/10.1074/jbc.274.53.37965>
- Schütz, S. V., Cronauer, M. V., & Rinnab, L. (2010). Inhibition of glycogen synthase kinase-3 β promotes nuclear export of the androgen receptor through a CRM1-dependent mechanism in prostate cancer cell lines. *Journal of Cellular Biochemistry*, 109(6), 1192–1200. <https://doi.org/10.1002/jcb.22500>
- Scotter, E. L., Vance, C., Nishimura, A. L., Lee, Y.-B., Chen, H.-J., Urwin, H., Sardone, V., Mitchell, J. C., Rogelj, B., Rubinsztein, D. C., & Shaw, C. E. (2014). Differential roles of the ubiquitin proteasome system and autophagy in the clearance of soluble and aggregated TDP-43 species. *Journal of Cell Science*, 127(6), 1263–1278. <https://doi.org/10.1242/jcs.140087>
- Sengupta, A., Molkentin, J. D., & Yutzey, K. E. (2009). FoxO transcription factors promote autophagy in cardiomyocytes. *Journal of Biological Chemistry*, 284(41), 28319–28331. <https://doi.org/10.1074/jbc.M109.024406>
- Seok, S., Fu, T., Choi, S. E., Li, Y., Zhu, R., Kumar, S., Sun, X., Yoon, G., Kang, Y., Zhong, W., Ma, J., Kemper, B., & Kemper, J. K. (2014). Transcriptional regulation of autophagy by an FXR-CREB axis. *Nature*, 516(729), 108–111. <https://doi.org/10.1038/nature13949>
- Sertel, E., Gezer, Ş., Pular, A., Çorakçı, A., & Yücesoy, İ. (2019). EP500 The diagnostic and prognostic value of serum L1CAM levels in endometrial cancer. *EPoster*, 29(Suppl 4), A312.1-A312. <https://doi.org/10.1136/ijgc-2019-ESGO.558>
- Settembre, C., Di Malta, C., Polito, V. A., Arencibia, M. G., Vetrini, F., Erdin, S., Erdin, S. U., Huynh, T., Medina, D., Colella, P., Sardiello, M., Rubinsztein, D. C., & Ballabio, A. (2011). TFEB Links Autophagy to Lysosomal Biogenesis. *Science*, 332(6036), 1429–1433. <https://doi.org/10.1126/science.1204592>
- Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., & Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, 343(6166), 84–87. <https://doi.org/10.1126/science.1247005>
- Shang, Y., Myers, M., & Brown, M. (2002). Formation of the androgen receptor transcription complex. *Molecular Cell*, 9(3), 601–610. [https://doi.org/10.1016/S1097-2765\(02\)00471-9](https://doi.org/10.1016/S1097-2765(02)00471-9)
- Sharma, K., Sheng, H. Z., Lettieri, K., Li, H., Karavanov, A., Potter, S., Westphal, H., & Pfaff, S. L. (1998). LIM homeodomain factors Lhx3 and Lhx4 assign subtype identities for motor neurons. *Cell*, 95(6), 817–828. [https://doi.org/10.1016/S0092-8674\(00\)81704-3](https://doi.org/10.1016/S0092-8674(00)81704-3)
- Shpilka, T., Weidberg, H., Pietrokovski, S., & Elazar, Z. (2011). Atg8: An autophagy-related ubiquitin-like protein family. *Genome Biology*, 12(7), 1–11.

<https://doi.org/10.1186/gb-2011-12-7-226>

- Slymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science (New York, N.Y.)*, *351*(6268), 84–88. <https://doi.org/10.1126/science.aad5227>
- Smith, K. (2001). Theoretical mechanisms in targeted and random integration of transgene DNA. *Reproduction Nutrition Development*, *41*(6), 465–485. <https://doi.org/10.1051/rnd:2001102>
- Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A., & Kucherlapati, R. S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*, *317*(6034), 230–234. <https://doi.org/10.1038/317230a0>
- Soltani, M. H., Pichardo, R., Song, Z., Sangha, N., Camacho, F., Satyamoorthy, K., Sanguenza, O. P., & Setaluri, V. (2005). Microtubule-associated protein 2, a marker of neuronal differentiation, induces mitotic defects, inhibits growth of melanoma cells, and predicts metastatic potential of cutaneous melanoma. *American Journal of Pathology*, *166*(6), 1841–1850. [https://doi.org/10.1016/S0002-9440\(10\)62493-5](https://doi.org/10.1016/S0002-9440(10)62493-5)
- Song, Michael, Yang, X., Ren, X., Maliskova, L., Li, B., Jones, I. R., Wang, C., Jacob, F., Wu, K., Traglia, M., Tam, T. W., Jamieson, K., Lu, S.-Y., Ming, G.-L., Li, Y., Yao, J., Weiss, L. A., Dixon, J. R., Judge, L. M., ... Shen, Y. (2019). Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nature Genetics*, *51*(8), 1252–1262. <https://doi.org/10.1038/s41588-019-0472-1>
- Song, Minjung, Kim, Y.-H., Kim, J.-S., & Kim, H. (2014). Genome Engineering in Human Cells. In *Methods in Enzymology* (Vol. 546, Issue C, pp. 93–118). Academic Press Inc. <https://doi.org/10.1016/B978-0-12-801185-0.00005-2>
- Sorek, R., Kunin, V., & Hugenholtz, P. (2008). CRISPR - A widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology*, *6*(3), 181–186. <https://doi.org/10.1038/nrmicro1793>
- Spada, A. R. La, Wilson, E. M., Lubahn, D. B., Harding, A. E., & Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, *352*(6330), 77–79. <https://doi.org/10.1038/352077a0>
- Srinivasan, P., Westover, M. B., & Bianchi, M. T. (2012). Propagation of Uncertainty in Bayesian Diagnostic Test Interpretation. *Southern Medical Journal*, *105*(9), 452–459. <https://doi.org/10.1097/SMJ.0b013e3182621a2c>
- Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*, *22*(2), 96–118. <https://doi.org/10.1038/s41580-020-00315-9>

- Strauss, W. L., Kemper, R. R., Jayakar, P., Kong, C. F., Hersh, L. B., Hilt, D. C., & Rabin, M. (1991). Human choline acetyltransferase gene maps to region 10q11–q22.2 by in situ hybridization. *Genomics*, *9*(2), 396–398.
[https://doi.org/10.1016/0888-7543\(91\)90273-H](https://doi.org/10.1016/0888-7543(91)90273-H)
- Streeter, A., Menzies, F. M., & Rubinsztein, D. C. (2016). *LC3-II Tagging and Western Blotting for Monitoring Autophagic Activity in Mammalian Cells* (pp. 161–170). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-2627-5_8
- Stumpel, C., & Vos, Y. (2004). L1 Syndrome. In M. Adam, H. Ardinger, & R. Pagon (Eds.), *GeneReviews*. University of Washington, Seattle.
<https://www.ncbi.nlm.nih.gov/books/NBK1484/>
- Sun, J., & Roy, S. (2021). Gene-based therapies for neurodegenerative diseases. *Nature Neuroscience*, *24*(3), 297–311. <https://doi.org/10.1038/s41593-020-00778-1>
- Sutermaster, B. A., & Darling, E. M. (2019). Considerations for high-yield, high-throughput cell enrichment: fluorescence versus magnetic sorting. *Scientific Reports*, *9*(1), 1–9. <https://doi.org/10.1038/s41598-018-36698-1>
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, *131*(5), 861–872.
<https://doi.org/10.1016/j.cell.2007.11.019>
- Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *126*(4), 663–676. <https://doi.org/10.1016/j.cell.2006.07.024>
- Takeda, Y., Asou, H., Murakami, Y., Miura, M., Kobayashi, M., & Uyemura, K. (1996). A Nonneuronal Isoform of Cell Adhesion Molecule L1: Tissue-Specific Expression and Functional Analysis. *Journal of Neurochemistry*, *66*(6), 2338–2349. <https://doi.org/10.1046/j.1471-4159.1996.66062338.x>
- Takeyama, K. I., Ito, S., Yamamoto, A., Tanimoto, H., Furutani, T., Kanuka, H., Miura, M., Tabata, T., & Kato, S. (2002). Androgen-dependent neurodegeneration by polyglutamine-expanded human androgen receptor in *Drosophila*. In *Neuron* (Vol. 35, Issue 5, pp. 855–864).
[https://doi.org/10.1016/S0896-6273\(02\)00875-9](https://doi.org/10.1016/S0896-6273(02)00875-9)
- Tanida, I. (2011). Autophagy basics. In *Microbiology and Immunology* (Vol. 55, Issue 1, pp. 1–11). <https://doi.org/10.1111/j.1348-0421.2010.00271.x>
- Tanida, I., Tanida-Miyake, E., Ueno, T., & Kominami, E. (2001). The human homolog of *Saccharomyces cerevisiae* Apg7p is a protein-activating enzyme for multiple substrates including human Apg12p, GATE-16, GABARAP, and MAP-

- LC3. *Journal of Biological Chemistry*, 276(3), 1701–1706.
<https://doi.org/10.1074/jbc.C000752200>
- Tanida, I., Ueno, T., & Kominami, E. (2004). LC3 conjugation system in mammalian autophagy. *The International Journal of Biochemistry & Cell Biology*, 36(12), 2503–2518. <https://doi.org/10.1016/j.biocel.2004.05.009>
- Taylor, J. P., Hardy, J., & Fischbeck, K. H. (2002). Toxic proteins in neurodegenerative disease. *Science*, 296(5575), 1991–1995.
<https://doi.org/10.1126/science.1067122>
- Taylor, S. C., Nadeau, K., Abbasi, M., Lachance, C., Nguyen, M., & Fenrich, J. (2019). The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. *Trends in Biotechnology*, 37(7), 761–774.
<https://doi.org/10.1016/j.tibtech.2018.12.002>
- Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Thoma, E. C., Wischmeyer, E., Offen, N., Maurus, K., Sirén, A. L., Scharl, M., & Wagner, T. U. (2012). Ectopic expression of neurogenin 2 alone is sufficient to induce differentiation of embryonic stem cells into mature neurons. *PLoS ONE*, 7(6). <https://doi.org/10.1371/journal.pone.0038651>
- Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science*, 282(5391), 1145–1147.
<https://doi.org/10.1126/science.282.5391.1145>
- Thor, S., & Thomas, J. B. (2002). Motor neuron specification in worms, flies and mice: conserved and “lost” mechanisms. *Current Opinion in Genetics & Development*, 12(5), 558–564. [https://doi.org/10.1016/S0959-437X\(02\)00340-4](https://doi.org/10.1016/S0959-437X(02)00340-4)
- Tian, R., Gachechiladze, M. A., Ludwig, C. H., Laurie, M. T., Hong, J. Y., Nathaniel, D., Prabhu, A. V., Fernandopulle, M. S., Patel, R., Abshari, M., Ward, M. E., & Kampmann, M. (2019). CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron*, 104(2), 239–255.e12. <https://doi.org/10.1016/j.neuron.2019.07.014>
- Ting, H.-J., & Chang, C. (2008). Actin associated proteins function as androgen receptor coregulators: An implication of androgen receptor’s roles in skeletal muscle. *The Journal of Steroid Biochemistry and Molecular Biology*, 111(3–5), 157–163. <https://doi.org/10.1016/j.jsbmb.2008.06.001>
- Torrance, C. J., Agrawal, V., Vogelstein, B., & Kinzler, K. W. (2001). Use of isogenic human cancer cells for high-throughput screening and drug discovery. *Nature Biotechnology*, 19(10), 940–945. <https://doi.org/10.1038/nbt1001-940>
- Toupenet Marchesi, L., Leblanc, M., & Stevanin, G. (2021). Current Knowledge of

- Endolysosomal and Autophagy Defects in Hereditary Spastic Paraplegia. *Cells*, 10(7), 1678. <https://doi.org/10.3390/cells10071678>
- Tsai, Y.-L., & Manley, J. L. (2021). Multiple ways to a dead end: diverse mechanisms by which ALS mutant genes induce cell death. *Cell Cycle*, 20(7), 631–646. <https://doi.org/10.1080/15384101.2021.1886661>
- Tsukada, M., & Ohsumi, Y. (1993). Isolation and characterization of autophagy-defective mutants of *Saccharomyces cerevisiae*. *FEBS Letters*, 333(1–2), 169–174. [https://doi.org/10.1016/0014-5793\(93\)80398-E](https://doi.org/10.1016/0014-5793(93)80398-E)
- Turner, D. P., Deng, H., & Houle, T. T. (2020). Bayesian Approaches to Statistical Inferences. *Headache: The Journal of Head and Face Pain*, 60(9), 1879–1885. <https://doi.org/10.1111/head.13952>
- Tyagi, R. K., Lavrovsky, Y., Ahn, S. C., Song, C. S., Chatterjee, B., & Roy, A. K. (2000). Dynamics of intracellular movement and nucleocytoplasmic recycling of the ligand-activated androgen receptor in living cells. *Molecular Endocrinology*, 14(8), 1162–1174. <https://doi.org/10.1210/mend.14.8.0497>
- Uchimura, T., Otomo, J., Sato, M., & Sakurai, H. (2017). A human iPS cell myogenic differentiation system permitting high-throughput drug screening. *Stem Cell Research*, 25, 98–106. <https://doi.org/10.1016/j.scr.2017.10.023>
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I. M., Edlund, K., Lundberg, E., Navani, S., Szigyanto, C. A. K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., ... Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220). <https://doi.org/10.1126/science.1260419>
- Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., & Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, 11(9), 636–646. <https://doi.org/10.1038/nrg2842>
- Valadez-Barba, V., Cota-Coronado, A., Hernández-Pérez, O. R., Lugo-Fabres, P. H., Padilla-Camberos, E., Díaz, N. F., & Díaz-Martínez, N. E. (2020). iPSC for modeling neurodegenerative disorders. *Regenerative Therapy*, 15, 332–339. <https://doi.org/10.1016/j.reth.2020.11.006>
- van Royen, M. E., van Cappellen, W. a., de Vos, C., Houtsmuller, A. B., & Trapman, J. (2012). Stepwise androgen receptor dimerization. *Journal of Cell Science*, 125(Pt 8), 1970–1979. <https://doi.org/10.1242/jcs.096792>
- Vasquez, K. M., Marburger, K., Intody, Z., & Wilson, J. H. (2001). Manipulating the mammalian genome by homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), 8403–8410. <https://doi.org/10.1073/pnas.111009698>
- Vergara, H. M., Bertucci, P. Y., Hantz, P., Tosches, M. A., Achim, K., Vopalensky,

- P., & Arendt, D. (2017). Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of *Platynereis dumerilii*. *Proceedings of the National Academy of Sciences*, *114*(23), 5878–5885. <https://doi.org/10.1073/pnas.1610602114>
- Vierbuchen, T., Ostermeier, A., Pang, Z. P., Kokubu, Y., Südhof, T. C., & Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature*, *463*(7284), 1035–1041. <https://doi.org/10.1038/nature08797>
- Vilar, M. J., Ranta, J., Virtanen, S., & Korkeala, H. (2015). Bayesian Estimation of the True Prevalence and of the Diagnostic Test Sensitivity and Specificity of Enteropathogenic *Yersinia* in Finnish Pig Serum Samples. *BioMed Research International*, *2015*. <https://doi.org/10.1155/2015/931542>
- von Muhlinen, N., Akutsu, M., Ravenhill, B. J., Foeglein, Á., Bloor, S., Rutherford, T. J., Freund, S. M. V., Komander, D., & Randow, F. (2012). LC3C, Bound Selectively by a Noncanonical LIR Motif in NDP52, Is Required for Antibacterial Autophagy. *Molecular Cell*, *48*(3), 329–342. <https://doi.org/10.1016/j.molcel.2012.08.024>
- Vos, Y. J., De Walle, H. E. K., Bos, K. K., Stegeman, J. A., Ten Berge, A. M., Bruining, M., Van Maarle, M. C., Elting, M. W., Den Hollander, N. S., Hamel, B., Fortuna, A. M., Sunde, L. E. M., Stolte-Dijkstra, I., Schrandt-Stumpel, C. T. R. M., & Hofstra, R. M. W. (2010). Genotype - phenotype correlations in L1 syndrome: A guide for genetic counselling and mutation analysis. *Journal of Medical Genetics*, *47*(3), 169–175. <https://doi.org/10.1136/jmg.2009.071688>
- Vos, Y. J., & Hofstra, R. M. W. (2010). An updated and upgraded L1CAM mutation database. *Human Mutation*, *31*(1), E1102–E1109. <https://doi.org/10.1002/humu.21172>
- Waddington, C. H. (1966). *Principles of Development and Differentiation*. The Macmillan Co. <https://wellcomecollection.org/works/zd7nft5n>
- Walcott, J. L., & Merry, D. E. (2002). Ligand promotes intranuclear inclusions in a novel cell model of spinal and bulbar muscular atrophy. *Journal of Biological Chemistry*, *277*(52), 50855–50859. <https://doi.org/10.1074/jbc.M209466200>
- Wang, C., Ward, M. E., Chen, R., Liu, K., Tracy, T. E., Chen, X., Xie, M., Sohn, P. D., Ludwig, C., Meyer-Franke, A., Karch, C. M., Ding, S., & Gan, L. (2017). Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. *Stem Cell Reports*, *9*(4), 1221–1233. <https://doi.org/10.1016/j.stemcr.2017.08.019>
- Wang, J., Zhang, K., Xu, L., & Wang, E. (2011). Quantifying the Waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences*, *108*(20), 8257–8262. <https://doi.org/10.1073/pnas.1017017108>

- Wang, Jing, Lu, X. X., Chen, D. Z., Li, S. F., & Zhang, L. S. (2004). Herpes simplex virus thymidine kinase and ganciclovir suicide gene therapy for human pancreatic cancer. *World Journal of Gastroenterology*, *10*(3), 400–403. <https://doi.org/10.3748/wjg.v10.i3.400>
- Wannenes, F., Caprio, M., Gatta, L., Fabbri, A., Bonini, S., & Moretti, C. (2008). Androgen receptor expression during C2C12 skeletal muscle cell line differentiation. *Molecular and Cellular Endocrinology*, *292*(1–2), 11–19. <https://doi.org/10.1016/j.mce.2008.05.018>
- Webb, J. L., Ravikumar, B., Atkins, J., Skepper, J. N., & Rubinsztein, D. C. (2003). α -synuclein Is Degraded by Both Autophagy and the Proteasome. *Journal of Biological Chemistry*, *278*(27), 25009–25013. <https://doi.org/10.1074/jbc.M300227200>
- Weidberg, H., Shpilka, T., Shvets, E., Abada, A., Shimron, F., & Elazar, Z. (2011). LC3 and GATE-16 N Termini Mediate Membrane Fusion Processes Required for Autophagosome Biogenesis. *Developmental Cell*, *20*(4), 444–454. <https://doi.org/10.1016/j.devcel.2011.02.006>
- Weidberg, H., Shvets, E., Shpilka, T., Shimron, F., Shinder, V., & Elazar, Z. (2010). LC3 and GATE-16/GABARAP subfamilies are both essential yet act differently in autophagosome biogenesis. *EMBO Journal*, *29*(11), 1792–1802. <https://doi.org/10.1038/emboj.2010.74>
- Weil, B. D., Jenkins, M. J., Uddin, S., Bracewell, D. G., Wellings, D., Farid, S. S., & Veraitch, F. (2017). An integrated experimental and economic evaluation of cell therapy affinity purification technologies. *Regenerative Medicine*, *12*(4), 397–417. <https://doi.org/10.2217/rme-2016-0156>
- Wesch, N., Kirkin, V., & Rogov, V. V. (2020). Atg8-Family Proteins—Structural Features and Molecular Interactions in Autophagy and Beyond. *Cells*, *9*(9). <https://doi.org/10.3390/CELLS9092008>
- Whitman, J. D., Hiatt, J., Mowery, C. T., Shy, B. R., Yu, R., Yamamoto, T. N., Rathore, U., Goldgof, G. M., Whitty, C., Woo, J. M., Gallman, A. E., Miller, T. E., Levine, A. G., Nguyen, D. N., Bapat, S. P., Balcerek, J., Bylsma, S. A., Lyons, A. M., Li, S., ... Marson, A. (2020). Evaluation of SARS-CoV-2 serology assays reveals a range of test performance. *Nature Biotechnology*, *38*(10), 1174–1183. <https://doi.org/10.1038/s41587-020-0659-0>
- Wichterle, H., Lieberam, I., Porter, J. A., & Jessell, T. M. (2002). Directed differentiation of embryonic stem cells into motor neurons. *Cell*, *110*(3), 385–397. [https://doi.org/10.1016/S0092-8674\(02\)00835-8](https://doi.org/10.1016/S0092-8674(02)00835-8)
- Wiedenmann, J., Oswald, F., & Nienhaus, G. U. (2009). Fluorescent proteins for live cell imaging: Opportunities, limitations, and challenges. *IUBMB Life*, *61*(11), 1029–1042. <https://doi.org/10.1002/iub.256>

- Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J. H., Bantscheff, M., ... Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, *509*(7502), 582–587. <https://doi.org/10.1038/nature13319>
- Wilson, M. I., Dooley, H. C., & Tooze, S. A. (2014). WIPI2b and Atg16L1: setting the stage for autophagosome formation. *Biochemical Society Transactions*, *42*(5), 1327–1334. <https://doi.org/10.1042/BST20140177>
- Wirth, M., Zhang, W., Razi, M., Nyoni, L., Joshi, D., O'Reilly, N., Johansen, T., Tooze, S. A., & Mouilleron, S. (2019). Molecular determinants regulating selective binding of autophagy adapters and receptors to ATG8 proteins. *Nature Communications*, *10*(1), 1–18. <https://doi.org/10.1038/s41467-019-10059-6>
- Würtele, H., Little, K. C. E., & Chartrand, P. (2003). Illegitimate DNA integration in mammalian cells. *Gene Therapy*, *10*(21), 1791–1799. <https://doi.org/10.1038/sj.gt.3302074>
- Xie, Y. Z., & Zhang, R. X. (2015). Neurodegenerative diseases in a dish: the promise of iPSC technology in disease modeling and therapeutic discovery. *Neurological Sciences*, *36*(1), 21–27. <https://doi.org/10.1007/s10072-014-1989-9>
- Yamamoto-Fukuda, T., Akiyama, N., & Kojima, H. (2020). L1CAM–ILK–YAP Mechanotransduction Drives Proliferative Activity of Epithelial Cells in Middle Ear Cholesteatoma. *American Journal of Pathology*, *190*(8), 1667–1679. <https://doi.org/10.1016/j.ajpath.2020.04.007>
- Yamamoto, M., Sobue, G., Mutoh, T., Li, M., Doyu, M., Mitsuma, T., & Kimata, K. (1993). Gene expression of high- (p140trk) and low-affinity nerve growth factor receptor (LNGFR) in the adult and aged human peripheral nervous system. *Neuroscience Letters*, *158*(1), 39–43. [https://doi.org/10.1016/0304-3940\(93\)90607-M](https://doi.org/10.1016/0304-3940(93)90607-M)
- Yang, W., Liu, Y., Slovik, K. J., Wu, J. C., Duncan, S. A., Rader, D. J., & Morrissey, E. E. (2015). Generation of iPSCs as a pooled culture using magnetic activated cell sorting of newly reprogrammed cells. *PLoS ONE*, *10*(8), 134995. <https://doi.org/10.1371/journal.pone.0134995>
- Yu, Jian, Zhang, L., Hwang, P. M., Rago, C., Kinzler, K. W., & Vogelstein, B. (1999). Identification and classification of p53-regulated genes. *Proceedings of the National Academy of Sciences*, *96*(25), 14517–14522. <https://doi.org/10.1073/pnas.96.25.14517>
- Yu, Junying, Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., & Thomson, J. A. (2007). Induced pluripotent stem cell lines derived from human

- somatic cells. *Science*, *318*(5858), 1917–1920.
<https://doi.org/10.1126/science.1151526>
- Yu, Z., Dadgar, N., Albertelli, M., Gruis, K., Jordan, C., Robins, D. M., & Lieberman, A. P. (2006). Androgen-dependent pathology demonstrates myopathic contribution to the Kennedy disease phenotype in a mouse knock-in model. *Journal of Clinical Investigation*, *116*(10), 2663–2672.
<https://doi.org/10.1172/JCI28773>
- Zakrzewski, W., Dobrzyński, M., Szymonowicz, M., & Rybak, Z. (2019). Stem cells: past, present, and future. *Stem Cell Research & Therapy*, *10*(1), 68.
<https://doi.org/10.1186/s13287-019-1165-5>
- Zhang, H., Liu, Y., Wang, L., Li, Z., Zhang, H., Wu, J., Rahman, N., Guo, Y., Li, D., Li, N., Huhtaniemi, I., Tsang, S. Y., Gao, G. F., & Li, X. (2013a). Differential effects of estrogen/androgen on the prevention of nonalcoholic fatty liver disease in the male rat. *Journal of Lipid Research*, *54*(2), 345–357.
<https://doi.org/10.1194/jlr.M028969>
- Zhang, S.-C., Wernig, M., Duncan, I. D., Brüstle, O., & Thomson, J. A. (2001). In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nature Biotechnology*, *19*(12), 1129–1133.
<https://doi.org/10.1038/nbt1201-1129>
- Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., Xu, W., Yang, N., Danko, T., Chen, L., Wernig, M., & Südhof, T. C. (2013b). Rapid Single-Step Induction of Functional Neurons from Human Pluripotent Stem Cells. *Neuron*, *78*(5), 785–798.
<https://doi.org/10.1016/j.neuron.2013.05.029>
- Zhao, J., Brault, J. J., Schild, A., Cao, P., Sandri, M., Schiaffino, S., Lecker, S. H., & Goldberg, A. L. (2007). FoxO3 Coordinately Activates Protein Degradation by the Autophagic/Lysosomal and Proteasomal Pathways in Atrophying Muscle Cells. *Cell Metabolism*, *6*(6), 472–483.
<https://doi.org/10.1016/j.cmet.2007.11.004>
- Zwaka, T. P., & Thomson, J. A. (2009). Homologous Recombination in Human Embryonic Stem Cells. In *Essentials of Stem Cell Biology* (pp. 417–422). Elsevier. <https://doi.org/10.1016/B978-0-12-374729-7.00046-9>

Appendix

A.1: Index of Abbreviations

The first six abbreviations are also established in Chapter 2.11.1 and relate to the primary output measures of magnetic bead cell sorting and flow cytometry analysis.

#I(<i>n</i> , <i>p</i> , <i>t</i>)	The number of cells in the initial population of a certain type (<i>n</i> for mCherry negative, <i>p</i> for mCherry positive, and <i>t</i> for total)
#N(<i>n</i> , <i>p</i> , <i>t</i>)	The number of cells in the sorted negative population of a certain type (<i>n</i> for mCherry negative, <i>p</i> for mCherry positive, and <i>t</i> for total)
#P(<i>n</i> , <i>p</i> , <i>t</i>)	The number of cells in the sorted positive population of a certain type (<i>n</i> for mCherry negative, <i>p</i> for mCherry positive, and <i>t</i> for total)
%I(<i>n</i> , <i>p</i>)	Percent of the initial population that are mCherry negative (<i>n</i>) or positive (<i>p</i>), equal to $\#In/\#It$ or $\#Ip/\#It$ respectively
%N(<i>n</i> , <i>p</i>)	Percent of the sorted negative population that are mCherry negative (<i>n</i>) or positive (<i>p</i>), equal to $\#Nn/\#Nt$ or $\#Np/\#Nt$ respectively
%P(<i>n</i> , <i>p</i>)	Percent of the sorted positive population that are mCherry negative (<i>n</i>) or positive (<i>p</i>), equal to $\#Pn/\#Pt$ or $\#Pp/\#Pt$ respectively
2A	A “self-cleaving” peptide sequence, commonly used for polycistronic expression. Includes E2A, F2A, P2A, and T2A sequences
<i>AAVS1</i>	Adeno-associated virus integration site 1
<i>acc</i>	Accuracy
AD	Alzheimer’s disease
AF	Alexa Fluor
ALS	Amyotrophic lateral sclerosis
AR	Androgen Receptor
ATG	Autophagy-related; specifically, mammalian homologues of the Atg genes/proteins as enumerated in yeast (e.g., ATG5, ATG8 family)

Baf	Bafilomycin A1
BDNF	Brain-derived neurotrophic factor
BF	Bright-field
BFP	Blue fluorescent protein
BM	Basal Medium, specifically in chemically differentiated motor neurons
BrdU	5-bromo-2-deoxyuridine
BSA	Bovine serum albumin
Ca ²⁺	Calcium
CAG	Chimeric promoter composed of the cytomegalovirus early enhancer, chicken β -actin promoter, and rabbit β -globin splice acceptor. To distinguished from the CAG trinucleotide repeat, it is always followed by “promoter” or used in the name of a plasmid.
CAG _(n)	Repeats of cytosine, adenine, and guanine nucleotides of length n (e.g., CAG ₂₃ is a repeat of 23 CAG triplets, or 69 nucleotides)
Cas9	CRISPR-associated protein 9
cDNA	Complementary DNA
CLYBL	Citrate lyase beta-like gene; a locus in intron 2 is used for transgene insertion
CM	Cortical neuron culture medium, specifically in i ³ Neurons
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR interference
C _T	Threshold cycle, as used in reference to the cycle in which a qPCR amplification crosses a fixed threshold.
d(n)	Day n, refers to days during an extended protocol, as during differentiation (i.e., d0 is the day that the protocol begins, d3 is three days later)
DAPI	4',6-diamidino-2-phenylindole
dCas9	Nuclease-dead CRISPR-associated protein 9

DHT	Dihydrotestosterone, the androgen ligand responsible for activating AR
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
Dox	Doxycycline Hyclate
DRPLA	Dentatorubral-pallidolusian atrophy
E8	Essential 8 Medium
ECD	Extracellular Domain
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
EtOH	Ethanol
F12	Ham's F-12 Nutrient Mixture (DMEM/F12 is a 1:1 mixture of DMEM and F12)
FACS	Fluorescence-activated cell sorting
FC	Fold change, as in \log_2FC
FDA	United States Food & Drug Administration
FDR	False discovery rate
FISH	Fluorescence in situ hybridization
FNR	False negative rate
FOXO	Forkhead box class O transcription factor family
FPR	False positive rate
FTD	Frontotemporal dementia
G3	WTC11 cells to which was added a dox-inducible hNGN2 cassette
GABARAP	γ -aminobutyric acid receptor-associated protein. Also abbreviated GBRP
GABARAPL1	γ -aminobutyric acid receptor-associated protein-like 1. Also abbreviated GBRPL1

GABARAPL2	γ -aminobutyric acid receptor-associated protein-like 2. Commonly known as Golgi-associated ATPase enhancer of 16kDa (GATE-16), also abbreviated GBRPL2
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
gDNA	Genomic DNA
gRNA	Guide RNA
H(n)	Cell line series derived from a healthy (non-SBMA) donor. The n refers to the polyQ repeat length in AR; H23 is the parental line, while 40, 54, 68, and KO are isogenic clones with altered polyQ domains.
HD	Huntington's disease
HDR	Homology-directed repair
HEK 293	Human embryonic kidney cells
HeLa	Human epithelial cervical cancer cells
hESC	Human embryonic stem cell
hNIL	A polycistronic cassette containing human NGN2, ISL1, and LHX3
HRP	Horseradish peroxidase
HTT	Huntingtin
i ³	Integrated, inducible, and isogenic; used in reference to the neural differentiation systems utilizing polycistronic cassettes <i>integrated</i> into the genome that drive doxycycline- <i>inducible</i> expression of transcription factors, all in a shared <i>isogenic</i> background
ICD	Intracellular domain
IM	Induction Medium, specifically in i ³ Neurons and i ³ LMNs
iPrOH	Isopropanol
iPSC	Induced Pluripotent Stem Cell
ISL1	Insulin gene enhancer protein ISL-1, or Islet-1
KO	Knockout, as in mutagenesis intended to induce a frameshift
L1/L1CAM	L1 Cell Adhesion Molecule

LB	Lysogeny broth
LC3A	Microtubule-associated protein 1 light chain 3A (MAP1LC3A)
LC3B	Microtubule-associated protein 1 light chain 3B (MAP1LC3B)
LC3C	Microtubule-associated protein 1 light chain 3C (MAP1LC3C)
LHX3	LIM homeobox 3
LIR	LC3-interacting region
LMN	Lower motor neuron
lncRNA	long non-coding RNA
LNGFR	Low-affinity nerve growth factor receptor
Mag	Shorthand for the synthetic marker gene SBP- Δ LNGFR (truncated LNGFR lacking its ICD), enabling magnetic bead cell sorting
MAPK	Mitogen-activated protein kinase
MASA	A type of hereditary spastic paraplegia associated with mutations to L1CAM, specifically referring to the canonical clinical presentation of mental retardation, aphasia, shuffling gait, and adducted thumbs
MBP	Myelin basic protein
MeOH	Methanol
Mg ²⁺	Magnesium
MM	Motor neuron culture medium, specifically in i ³ LMNs
MSE	Mean squared error
mTOR	Mechanistic target of Rapamycin
NCM	Neural culture medium, specifically in chemically differentiated motor neurons
NEAA	Non-essential amino acids
NGN2	Neurogenin 2
NHEJ	Non-homologous end joining
NIM	Neural Induction Medium, specifically in chemically differentiated motor neurons

NLS	Nuclear localization sequence
NPC	Neuroepithelial progenitor cell
NT-3	Neurotrophin-3
NT	Nontargeting, as in an siRNA, shRNA or gRNA designed to have no matching sequences in the genome
p-	Phospho, as in an antibody that specifically recognizes a phosphorylated protein
p62	Sequestrome 1 (SQSTM1)
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBST	PBS + 0.1% w/v Tween-20
PCR	Polymerase chain reaction
PD	Parkinson's disease
PE	Phosphatidylethanolamine
PFA	paraformaldehyde
PI	phosphatidylinositol
PI3K	PI 3-kinase
PI3P	PI-3-phosphate
PLO	Poly-L-ornithine
PM	Patterning medium, specifically in chemically differentiated motor neurons
PolyQ _(n)	Repeats of glutamine of length n (e.g., polyQ ₂₃ is a repeat of 23 glutamines). Also used in the form (n)Q, as in 23Q
qPCR	Quantitative (real-time) PCR
RI	ROCK inhibitor, specifically Y-27632
RIN	RNA integrity number
ROCK	Rho-associated protein kinase
RO	Reverse osmosis

ROC	Receiver operating characteristic
RT	Reverse transcription, or room temperature when preceded by “at”
rtTA3G	Third-generation reverse tetracycline transactivator
S(n)	Cell line series derived from a donor with SBMA. The n refers to the polyQ repeat length in AR; S68 is the parental line, while 23, 40, 54, and KO are isogenic clones with altered polyQ domains
SBMA	Spinal and bulbar muscular atrophy
SBP	Streptavidin-binding peptide
SCA	Spinocerebellar ataxia
SDS	Sodium dodecyl sulfate
<i>se</i>	Sensitivity
shRNA	Short hairpin RNA
SILAC	Stable isotope labeling with amino acids in cell culture
siRNA	Short interfering RNA
SMA	Spinal muscular atrophy
<i>sp</i>	Specificity
TALENs	Transcription activator-like effector nucleases
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFE	Transcription factor E family
TM	Terminal Medium, specifically in chemically differentiated motor neurons
TMD	Transmembrane domain
TORC1	target of rapamycin complex 1
TPM	Transcripts per million
TRE3G	Third-generation tetracycline response element
Tub	Tubulin (as in α -Tubulin and β 3-Tubulin)
UPS	Ubiquitin-proteasome system
v/v	Volume of solute per volume of total solution, in mL/mL

w/v	Weight of solute per volume of total solution, in g/mL
WT	Wildtype, referring to a basal genetic background without relevant edits/manipulations
WTC11	An apparently healthy human male iPSC line donated to the Coriell Institute

A.2: Reagents

Reagent	Supplier	Product Number
20× Borate Buffer	Thermo	28341
5-Bromo-2'-deoxyuridine	Sigma	B9285
5× siRNA Buffer	Horizon	B-002000-UB-100
Acrylamide Bis-Acrylamide 30% Stock Solution	Severn Biotech	20-2100-10
Ammonium Persulfate (APS)	Acros Organics	327081000
Ampicillin	Sigma	A9518
AutoMACS Rinsing Solution	Miltenyi	130-091-222
B27 Supplement	Thermo	17504044
Bafilomycin A1	Enzo	BML-CM110-0100
BDNF, Recombinant human	PeproTech	450-02
BlpI	NEB	R0585
Bromophenol Blue	VWR	200152E
BSA	Sigma	A7906
BSA (IgG & Protease-free)	Jackson Labs	001-000-161
BstXI	NEB	R0113
β-mercaptoethanol	Sigma	M6250
CHIR99021	Sigma	SML1046
cOmplete Protease Inhibitor Cocktail	Sigma	11836170001
Compound E (γ-secretase inhibitor XXI)	Calbiochem	565790
Countess Cell Counting Chamber Slides	Thermo	C10228
CultureOne Supplement	Thermo	A3320201
DAPI	Thermo	62248
Direct-zol RNA Miniprep Kit	Zymo	R2051
Dispase	Sigma	D4693
DMEM, high glucose, without L-glutamine	Sigma	D6546
DMEM/F12, HEPES	Thermo	11330032
DMSO	Sigma	D8418
DNA Ligation Kit, Mighty Mix	Takara	6023

Reagent	Supplier	Product Number
Doxycycline hyclate	Sigma	D9891
Dulbecco's PBS, without Ca ²⁺ /Mg ²⁺	Thermo	14190094
Dulbecco's PBS, without Ca ²⁺ /Mg ²⁺	Sigma	D8537
Dynabeads MyOne Streptavidin C1	Thermo	65001
DynaMag2 Magnetic Rack	Thermo	12321D
ECL Prime	Cytiva	RPN2232
EDTA (0.5M)	Thermo	AM9260G
Essential 8 Medium	Thermo	A1517001
Fetal Bovine Serum (FBS)	Sigma	F7524
Geneticin (G418)	Thermo	1181031
Gibson Assembly Master Mix	NEB	E2611
Glass Pasteur pipettes	Fisher	10026041
GlutaMAX	Thermo	35050061
Glycerol	Thermo	17904
HEPES	Sigma	H0887
Immobilion-FL PVDF membrane	Millipore	IPFL00010
Insulin	Sigma	I9278
Kanamycin	VWR	1.05177.0005
L-Glutamine	Sigma	G7513
L1CAM siRNA 1	Horizon	J-011069-05
L1CAM siRNA 2	Horizon	J-011069-06
L1CAM siRNA 3	Horizon	J-011069-07
L1CAM siRNA 4	Horizon	J-011069-08
Laminin Mouse Protein	Thermo	23017015
LDN193189	Sigma	SML0559
Lipofectamine 2000	Thermo	11668019
Lipofectamine RNAiMAX	Thermo	13778150
Lipofectamine Stem	Thermo	STEM00001
MACS BSA Stock Solution	Miltenyi	130-091-376
Matrigel, hESC-Qualified	Corning	11573560

Reagent	Supplier	Product Number
MEM NEAA	Thermo	11140035
MG-132 (Z-Leu-Leu-Leu-al)	Sigma	C2211
N,N,N',N'-Tetramethylethylenediamine (TEMED)	Sigma	T9281
N2 Supplement	Thermo	17502048
NE-PER Nuclear and Cytoplasmic Extraction Reagents	Thermo	78833
NEBuffer 2.1	NEB	B7202
NEBuilder HiFi DNA Assembly Master Mix	NEB	E2621
Neurobasal medium	Thermo	21103049
NT-3, Recombinant human	PeptoTech	450-03
NuPAGE Bis-Tris gels	Thermo	NP0322
ON-TARGETplus Non-targeting Pool siRNAs	Horizon	D-001810-10-50
Opti-MEM	Thermo	31985047
pAdVantage plasmid	Promega	E1711
PageRuler Plus Prestained Protein Ladder	Thermo	26620
Penicillin-Streptomycin	Sigma	P0781
pFLAG CMV 5a	Sigma	E6908
PfuUltra High-Fidelity DNA Polymerase	Agilent	600385
Platinum SuperFi PCR Master Mix	Thermo	12358010
Poly-L-Ornithine (PLO)	Sigma	P3655
PowerUp SYBR Green 2× Master Mix	Thermo	A25742
PureLink HiPure Plasmid Filter Maxiprep Kit	Thermo	K210017
Purmorphamine	Tocris	4551
Puromycin dihydrochloride	Sigma	P8833
Q5 High-Fidelity 2× Master Mix	NEB	M0492
QIAquick PCR Purification kit	Qiagen	28104
QIAprep Spin Mini prep kit	Qiagen	27104
QIAquick Gel Extraction Kit	Qiagen	28704
QuickExtract DNA Extraction Solution	Epicentre	QE0905T

Reagent	Supplier	Product Number
QuickLoad Purple 1kb Plus DNA Ladder	NEB	N0550
Retinoic acid	Sigma	R2625
RNase-free Water	Horizon	B-003000-WB-100
RNaseZAP	Sigma	R2020
SB431542	Sigma	S4317
StemPro Accutase	Thermo	A1110501
Streptavidin Alexa Fluor 488	Thermo	S11223
SuperScript III First-Strand Synthesis System	Thermo	18080051
T4 DNA Ligase	NEB	M0202
TransIT-2020 Transfection Reagent	Mirus	MIR 5400
Triton X-100	Fisher	10254583
TRIzol LS	Thermo	10296028
Trypan blue (0.4%)	Thermo	T10282
Trypsin-EDTA	Sigma	T3924
Tween-20	VWR	663684B
UltraPure Water	Thermo	10977035
XL10-Gold Ultracompetent Cells	Agilent	200315
Y-27632 dihydrochloride	Tocris	1254
μ -slide 8-well chamber slides	ibidi	IB-80826

A.3: Media Formulations

A.3.1: HeLa & HEK 293T maintenance culture medium

Component	Amount per 500 mL	Final Concentration
DMEM, high glucose, without L-glutamine	445 mL	
Fetal Bovine Serum (FBS)	50 mL	10% v/v
L-Glutamine	5 mL	1% v/v
Penicillin-Streptomycin*	5 mL	1% v/v

*Optional/as needed

A.3.2: iPSC maintenance culture medium

	Component	Solvent	Amount	Final
			per 500 mL	Concentration
Essential 8 (E8)	DMEM/F12 with HEPES		500 mL	
	L-Ascorbic acid 2-phosphate sesquimagnesium salt hydrate		32 mg	221 μ M
	Sodium selenite (0.1 mg/mL)	PBS	70 μ L	81.0 nM
	Sodium bicarbonate		271.5 mg	6.46 mM
	Sodium chloride	As needed to adjust osmolarity to 340 mOsm		
	Sodium hydroxide & Hydrochloric acid	As needed to adjust pH to 7.4		
	Insulin (10 mg/mL)		250 μ L	5 μ g/mL
	TGF- β 1 (2 μ g/mL)	PBS	500 μ L	2 ng/mL
	FGF- β (100 μ g/mL)	PBS	500 μ L	100 ng/mL
	Holo-transferrin (10.7 mg/mL)	PBS	500 μ L	10.7 μ g/mL

Aliquot & add fresh to each bottle:

(As developed by G. Chen et al., 2011)

A.3.3: Transcription factor-mediated differentiation media

	Component	Solvent	Amount	Final
			per 500 mL	Concentration
Induction Medium (IM)	DMEM/F12 with HEPES		485 mL	
	GlutaMAX (100×)		5 mL	1% v/v
	NEAA (100×)		5 mL	1% v/v
	N2 (100×)		5 mL	1% v/v

			Amount per 50 mL	Final Concentration
Add fresh immediately before use:	ROCK inhibitor Y-27632 (d0 only; 10 mM)	PBS	50 μ L	10 μ M
	Doxycycline (2 mg/mL)	PBS	50 μ L	2 μ g/mL
i ³ LMNs only:	Compound E (γ -secretase inhibitor XXI)	DMSO	50 μ L	200 nM
	BrdU (d3 only; 40 mM)	H ₂ O	50 μ L	40 μ M

	Component	Solvent	Amount per 50 mL	Final Concentration	
Culture Medium (CM)	Neurobasal medium		47.5 mL		
	GlutaMAX (100×)		500 μ L	1% v/v	
	NEAA (100×)		500 μ L	1% v/v	
	N2 (100×)		500 μ L	1% v/v	
	B27 (50×)		1 mL	2% v/v	
	For both i ³ Neurons & i ³ LMNs:	BDNF (10 μ g/mL)	PBS+0.1% BSA	50 μ L	10 ng/mL
		NT-3 (10 μ g/mL)	PBS+0.1% BSA	50 μ L	10 ng/mL
		Laminin (1 mg/mL)		50 μ L	1 μ g/mL
i ³ LMNs only:	CultureOne (100×)		5 mL	1% v/v	

A.3.4: Chemical differentiation media

	Component	Solvent	Amount	Final
			per 500 mL	Concentration
Basal Medium (BM)	DMEM/F12		235 mL	
	Neurobasal medium		235 mL	
	GlutaMAX (100×)		5 mL	1% v/v
	NEAA (100×)		5 mL	1% v/v
	N2 (100×)		5 mL	1% v/v
	B27 (50×)		10 mL	2% v/v
	Insulin (10 mg/mL)		250 μ L	5 μ g/mL
	β -mercaptoethanol (50 mM)	PBS	500 μ L	50 μ M

To BM, add the following fresh:

	Component	Solvent	Amount	Final
			per 50 mL	Concentration
Neural Induction Medium (NIM)	SB431542 (20 mM)	DMSO	5 μ L	2 μ M
	CHIR99021 (30 mM)	DMSO	5 μ L	3 μ M
	LDN193189 (3 mM)	H ₂ O	5 μ L	300 nM
Patterning Medium (PM)	Retinoic acid (5 mM)	DMSO	5 μ L	500 nM
	Purmorphamine (10 mM)	DMSO	5 μ L	1 μ M
Terminal Medium (TM)	Purmorphamine (1 mM)	DMSO	5 μ L	100 nM
Neural Culture Medium (NCM)	Compound E (2 mM)	DMSO	5 μ L	200 nM
	CultureOne (100×)		5 mL	1% v/v