The identification of novel factors required for the development of primordial germ cells

Gonçalo Pizarro
Madureira Salgado de Oliveira

Homerton College, University of Cambridge
MRC Laboratory of Molecular Biology
Cambridge, United Kingdom

This dissertation is submitted for the degree of Doctor of Philosophy

July 2019

Preface

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 that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This work has not been submitted either as a whole or in part for a degree or qualification at the University of Cambridge or any other institution of higher degree. This dissertation does not exceed the length limit specified by the Degree Committee for the Faculty of Biology.

Gonçalo Oliveira

Acknowledgements

First of all I would like to thank my supervisor, Gerry Crossan, for providing me with the opportunity to do very exciting research in his lab and for all the guidance throughout my studies. I am very thankful for all the time and effort dedicated to pushing me to achieve my maximum potential.

Secondly, I would like to thank the organisers of the GABBA programme, which funded my studies, for providing me with the absolute freedom to pursue the PhD project of my own choice.

I also have to thank my PhD colleague Ross Hill with whom I shared the majority of experiences in the lab. I am particularly thankful for the support, advice and the good times both inside and outside the lab. I also have to thank Ross for the extensive help doing mouse work and for carefully reading my PhD thesis.

I would like to also acknowledge Jonathan Ribeiro for the helpful discussions during the last year of my studies and all of the other current and past members of the Crossan lab for the good times in the lab. These include: Ben Butt, Jun-Yu Ma, Vlad Sacalean, Pranay Shah and Nazareno Bona.

I want to extend my acknowledgments to Maria Daly, Fan Zhang and Martyn Balmont for the constant help with cell sorting. Also I want to thank Patrycja Kozik and Aydan Szeto for the input and helpful discussions about CRISPR screening and Alastair Crisp for extensively supporting me analysing the CRISPR screen results.

I also want to acknowledge Sofia Venceslau and Inês Cardoso Pereira with whom I first experienced science and who strongly encouraged my further studies.

Finally, I would like to thank my family, my parents and my friends for their constant support.

Abstract

DNA is constantly exposed to both exogenous and endogenous sources of DNA damage. Therefore, organisms evolved a plethora of DNA repair pathways to repair these distinct lesions and prevent mutations from arising. Mutations that occur in somatic cells can lead to disease, which may be detrimental to that single individual. However, germ line mutations can be passed on to the next generation. These mutations not only drive genome evolution but can also impair the fitness of the offspring.

It has been shown that primordial germ cells (PGCs), the germ cells specified during embryonic development, have the highest mutation frequency compared to germ cells belonging to other stages of gametogenesis. Furthermore, mutations in PGCs are particularly important because mutations in these cells can affect multiple offspring. Evidence shows that during PGC development, which entails extensive genomic and epigenomic transactions, there is the generation of DNA damage. However, the DNA repair pathways used in PGCs and the nature of DNA damage encountered by these cells which can provide mechanistic insight into how mutagenesis is suppressed in the germ line remains largely unexplored.

Here we describe the use of an *in vitro* system that generates PGC-like cells (PGCLCs) from ESCs to discover novel factors required for PGC development. Using this system, we have firstly shown that we can recapitulate the genetic requirement for DNA repair in PGCLCs *in vitro*. Secondly, we coupled this system with CRISPR/Cas9 screening technology, which permitted us to identify novel factors required during the development of PGCLCs. We identified the translesion synthesis (TLS) factor *Rev1* as a key factor involved in PGCLC development. Furthermore, we validated this requirement *in vivo* and showed that absence of *Rev1* causes a numeric reduction in PGCs during the migratory period of PGC development. We also identified additional factors, namely involved in homologous recombination, chromosome cohesion which will be investigated in the future.

This screen permitted the identification for the first time of the requirement for TLS in PGCs suggesting that the germ cell lineage may employ this mutagenic pathway to promote genome evolution.

Table of Contents

List of abbreviations

BER - base excision repair

AP - apurinic/apyrimidinic

NER - nucleotide excision repair

CPD - cyclobutane–pyrimidine dimer

6–4PP - 6–4 pyrimidine–pyrimidone photoproducts

NER- nucleotide excision repair

GG-NER - global-genomic nucleotide excision repair

TC-NER - transcription-coupled nucleotide excision repair

RNA Pol II - RNA polymerase II

insertion/deletion - indel

MMR - mismatch repair

ICL - inter-strand crosslink

FA - Fanconi anemia

DSB - double-stranded break

TLS - translesion synthesis

HR - homologous recombination

DPC - DNA-protein crosslink

SSB - single-stranded break

SSBR - single-stranded break repair

IR - ionizing radiation

NHEJ - nonhomologous end-joining

SSA - single-strand annealing

SDSA – synthesis dependent strand annealing

BIR – break induced replication

ssDNA – single-stranded DNA

HJ – Holliday junction

MMEJ - microhomology-mediated end joining

UV - ultra-violet

TC - transcription-coupled

SSC - spermatogonial stem cell

PGC - primordial germ cell

SSC – spermatogonial stem cell

SCNT - somatic cell nuclear transfer

iPSC - induced pluripotent stem cell

ICM - inner cell mass

5mC - 5-methylcytosine

5hmC - 5-hydroxymethylcytosine

5fC - 5-formylcytosine

5caC - 5-carboxylcytosine

MEF – mouse embryonic fibroblast

TKO – triple knockout

ND - not determined

TE - transposable elements

IAP - intracisternal A particle

LINE1 - long interspersed nuclear element 1

BAC – bacterial artificial chromosome

ESC - embryonic stem cell

PGCLC - primordial germ cell-like cell

EpiLC - epiblast-like cells

FGFb - fibroblast growth factor b

KSR - Knockout serum replacement

SCF - stem cell factor

LIF - leukemia inhibitory factor

EGF - epidermal growth factor

FACS - fluorescence-activated cell sorting

RT-qPCR - reverse-transcription quantitative PCR

sgRNA - single guide RNA

LR-PCR - long-range PCR

KO – knockout

NGS - next generation sequencing

MOI - multiplicity of infection

NTC - non-targeting control

iCas9 - Cas9 inducible

ssODN - single stranded oligonucleotide

TU/ml - transducing units per ml

List of tables

List of figures

Chapter 1 - Introduction

1.1 - DNA repair is essential to all life forms

All of the genetic information required to generate an entirely new organism is encoded in the DNA molecule. The structure of this molecule allows genetic information to be stored, replicated and also transcribed. However, in order to propagate life this genetic information has to be faithfully passed between generations. Therefore, the chemical stability and integrity of the genome is essential to life. Additionally, the processes of DNA replication and transcription pose significant threats to the integrity of the genome (Maya-Mendoza et al. 2018; Herman and Dworkin 1971). However, like all biological macromolecules, DNA itself is chemically not inert (Lindahl 1993) and damage can occur not only to the bases but also the sugar-phosphate backbone. In contrast, DNA can undergo spontaneous damage due to the intrinsic chemical instability of the DNA molecule (Lindahl 1993). These lesions can also result from chemical insults from a wide variety of sources. Genotoxins can be generated as toxic byproducts of metabolism or exposure to exogenous or environmental agents. It is estimated that each cell in our bodies experiences up to 100,000 spontaneous DNA lesions per day (Hoeijmakers 2009) (Lindahl and Barnes 2000). Therefore, in order to maintain the sequence and structural integrity of the genome, organisms have evolved several DNA damage repair pathways with specialised mechanisms to detect, signal and repair DNA lesions **(Figure 1)**. As these lesions are diverse in nature, they require different enzymatic activities in order to repair them. The majority of DNA repair transactions resolve the lesion in an error-free manner. However, errors in the sequence or the structure of the DNA, or DNA mutations, can still emerge either from illegitimate or error-prone repair of DNA lesions. The introduction of mutations in the genome can have catastrophic consequences. Whilst the acquisition of mutations can have deleterious consequences in somatic tissue, such as tumourigenesis, *de novo* mutagenesis in the germ line can have not only a deleterious effect, such as loss of fitness in the progeny, but also can drive genome evolution.

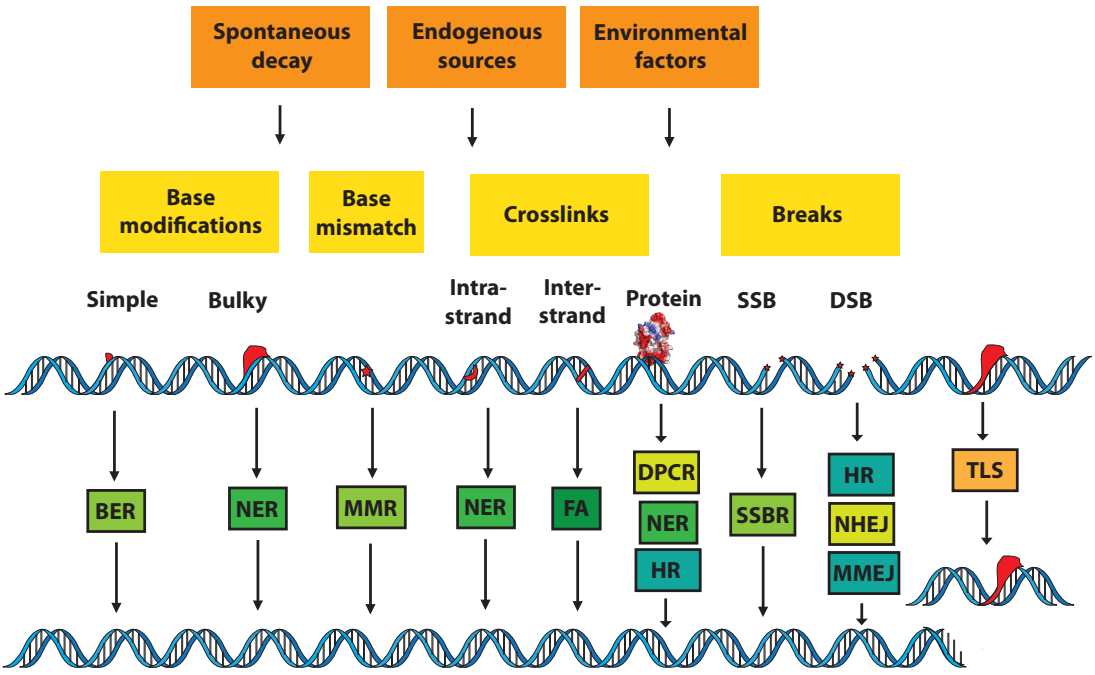


Figure 1. A plethora of DNA repair pathways have evolved to deal with defined lesions in the genome. DNA is continually exposed to a series of insults that cause a range of lesions. These lesions can occur not only at the base or sugar backbone level but can also include breaks in one or both of the DNA strands. Therefore, the choice of repair mechanism is largely defined by the type of lesion, but also influenced by factors such as the stage in the cell cycle. Both the types of lesion and the major pathways thought to repair or tolerate that kind of damage are depicted.

1.2 - The emergence of diverse DNA repair pathways

1.2.1 – Base excision repair

Damage to the base, the fundamental unit of the genetic code, can occur due to the intrinsic chemical nature of the bases - their ability to form Watson-Crick base pairs (guanine-cytosine and adenine-thymidine). An example of this is the spontaneous deamination of cytosine to uracil (Lindahl and Nyberg 1974). This poses a threat to genome stability as uracil pairs with adenine, therefore when the genome is replicated a C to T mutation can be introduced. Alternatively, the base of the DNA molecule can react with an exogenous agent adducting the base, in general by oxidation, alkylation or hydrolysis (Lindahl 1993). Common sources of base adducts are not only the spontaneous decay of DNA (Lindahl 1993) but also environmental agents (such as chemicals in cigarette smoking (Phillips et al. 1988), radiation, or cancer chemotherapy treatment drugs). These lesions pose a threat to the genome integrity since they may lead to incorrect base pairing with the potential for mutagenesis. Alternatively, they may pose an obstacle to either the DNA replication or transcription machinery. In general, base adducts lead to small distortions in the topology of the DNA helix. One route to repair this class of lesions is known as base excision repair (BER) (Lindahl 1976). Initially, a DNA glycosylase recognizes the adducted base and then removes the base by cleaving the covalent bond between the base and the sugar in the DNA (glycosidic bond). This leaves the sugar-phosphate backbone intact creating an apurinic/apyrimidinic (AP) site (Lindahl 1976). At least 12 distinct mammalian DNA glycosylases have been described, each of them recognizing a subset of lesions, frequently with some overlap in specificities (Krokan and Bjoras 2013). The resultant AP site is further processed by short-patch repair or long-patch repair that largely uses different proteins to complete repair (Fortini et al. 1998). Both routes of repair require cleavage of the sugar-phosphate backbone and then a polymerase fills in the gap using the other strand as a template (Lindahl 1976). Therefore, BER specifically removes the lesion and uses the undamaged strand as the template ensuring that the genomic information is not corrupted during the repair transaction.

1.2.2 – Nucleotide excision repair

In addition to small base adducts in DNA, bulky adducts which are more distorting to the DNA helix also occur necessitating repair. The archetypal lesions of this class are cyclobutane-pyrimidine dimers (CPDs) and 6–4 pyrimidine–pyrimidone photoproducts (6–4PPs) in which covalent bonds are formed between consecutive bases along the nucleotide chain (Ravanat, Douki, and Cadet 2001). As these lesions are much bulkier than simple base adducts, they cannot fit into the active site of DNA glycosylases and subsequently be removed from the helix. Instead, these lesions are repaired by the nucleotide excision repair (NER) pathway. There are two main branches of NER: global-genomic NER (GG-NER) and transcription-coupled NER (TC-NER). In GG-NER the entire genome is continually examined by proteins that recognize distortion of the DNA helix (Sugasawa et al. 1998; Sugasawa et al. 2001), whereas TC-NER is activated by stalling of the RNA polymerase II (RNA Pol II) at lesions within the template strand of transcribed regions (Fousteri et al. 2006). Although these two sub-pathways of NER rely on different mechanisms to detect helix-distorting lesions, they both converge on common downstream apparatus to resolve the damage. In contrast to BER in which the glycosidic bond is cleaved, NER employs nucleases to cleave bonds in the sugar phosphate backbone (Hu et al. 2013). After binding to the lesion and melting the duplex to form a bubble, NER factors recruit two structure specific endonucleases: XPF (Xeroderma pigmentosum complementation group f) together with ERCC1 (Excision repair cross-complementation group 1) and XPG (Xeroderma pigmentosum complementation group g) (Evans et al. 1997; Bardwell et al. 1994). These nucleases incise the strand containing the lesions resulting in a short patch of single stranded DNA, which is filled in by the replicative DNA polymerases (Shivji et al. 1995).

1.2.3 – Mismatch repair

In addition to base lesions, the DNA helix may become distorted due to base mispairing (*i.e.* when a base is introduced which cannot form Watson-Crick pairing). These lesions often occur during DNA replication when the polymerase introduced the wrong nucleotide resulting in DNA mismatches or when the polymerase slips which leads to an insertion/deletion (indel) loop (Shcherbakova et al. 2003; Kroutil et al. 1996). The frequency of these errors is significantly reduced by the proofreading activity of the replicative DNA polymerases (Tran et al. 1997; Shcherbakova et al. 2003). Mismatches can also occur when a base is damaged altering its ability to form normal Watson-Crick base pairs (Mu et al. 1997). However, when this type of lesion persists, a repair pathway named mismatch repair (MMR) ensures that these lesions are corrected. MMR entails an enzymatic machinery that detects the mismatched base in the correct strand (Lahue, Au, and Modrich 1989) and initiates repair by activating an exonuclease-mediated degradation of DNA from a nick that is distant to the mismatch (Cooper, Lahue, and Modrich 1993). Then, DNA polymerase δ activity re-fills this gap by inserting the correct nucleotide and therefore repairing the mismatch (Lahue, Au, and Modrich 1989; Schaaper 1988).

1.2.4 – Interstrand crosslink repair

Another consequence of chemical modifications of DNA bases is the formation of interstrand crosslinks (ICLs). These crosslinks covalently bind the opposite strands of DNA together and pose extremely toxic lesions to cells as they can block both DNA replication and also transcription. These lesions are caused by agents with independently reactive groups within the same molecule such as reactive aldehydes and cancer chemotherapy treatment drugs including mitomycin C or cisplatin (Stone et al. 2008; Sasaki and Tonomura 1973; Poll et al. 1985). ICLs can be repaired by the Fanconi anemia (FA) repair pathway (Sasaki 1975; Auerbach 1988; German et al. 1987), which comprises 21 gene products (*Fanca* to *Fancv*) that act together with other DNA damage response pathways including NER but also double-strand break (DSB) repair and translesion sysnthesis (TLS), which both will be described next. The FA pathway operates mainly during the S phase of the cell cycle and requires converging replication

forks (Zhang et al. 2015) to orchestrate the various steps that are necessary for ICL repair: lesion recognition, DNA incision, lesion bypass and lesion repair. Lesion recognition is mediated by the protein FANCM that binds the ICL and serves as a platform for the FA core complex, which comprises 14 proteins (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM, FANCT, FAAP100, MHF1, MHF2, FAAP20 and FAAP24) (Xue et al. 2008; Kim et al. 2008). The FA core complex functions as a large ubiquitin ligase for two other FA proteins, FANCD2 and FANCI. After loading on to chromatin, ubiquitylated FANCD2-I controls the nucleotytic incision at the converging replication forks, which will ultimately release the ICL from one of the two parental DNA strands. This is achieved by FANCD2-Ub recruitment of SLX4 (Structure-specific Endonuclease Subunit 4 also known ans FANCP) which in turn recruits and activates several structure-specific endonucleases such as the XPF (also known and FANCQ)-ERCC1 heterodimer, MUS81 (MUS81 Structure-Specific Endonuclease Subunit)-EME1 (Essential meiotic structure-specific endonuclease 1) and SLX1 (Structure-specific endonuclease subunit 1) (Knipscheer et al. 2009). Once the ICL has been released from one of the DNA strands, low-fidelity TLS polymerases that have larger binding pockets compared to replicative polymerases, bypass the lesion by incorporating nucleotides opposite the ICL and extending the nascent strand (Sarkar et al. 2006). The restoration of one intact DNA duplex permits its use as a template for homologous recombination (HR), a subset of DSB repair that permits the repair of DSBs by invading an undamaged DNA molecule and resynthesizing the missing strand using the other strand as template. Following strand invasion and repair synthesis, resolution and ligation give rise to an intact double stranded DNA duplex, thereby completing FA pathway mediated ICL repair. Interestingly, in addition to the structure-specific nucleases XPF-ERCC1, MUS81-EME1 and SLX1, other nucleases including FAN1 (Fanconi-associated nuclease 1) and three genes from the SNM1 (Sensitive to nitrogen mustard 1) family have been implicated in ICL repair. FAN1 has been shown to interact directly with mono-ubiquitylated FANCD2–I via a ubiquitin-binding zinc finger domain, however this domain was shown to be dispensable for ICL recruitment and resistance to crosslinking agents (Liu et al. 2010; Smogorzewska et al. 2010; Thongthip et al. 2016). Also, epistasis studies suggest that FAN1 has ICL repair activities that are independent of the FA pathway (Yoshikiyo et al. 2010; Zhou et al. 2012). The *Snm1* family of genes in vertebrates include SNM1A, APOLLO/SNM1B, ARTEMIS/SNM1C, ELAC2 and

CPSF73 of which SNM1A, APOLLO/SNM1B, ARTEMIS/SNM1C are reported to confer cellular resistance to ICLs (Cattell, Sengerova, and McHugh 2010). One hypothesis for the presence of additional structure-specific endonucleases may be that outside S-phase cells have additional factors that repair ICLs independently of the FA pathway. Alternatively, the different nucleases may have different substrates specificities to repair the diverse lesions caused by crosslinking agents.

1.2.5 – DNA-protein crosslink repair

In addition to crosslinking to DNA itself, DNA can also be crosslinked to proteins forming DNA-protein crosslinks (DPCs) (Stingele, Bellelli, and Boulton 2017). DPCs can be repaired by direct crosslink hydrolysis as exemplified by TDP1 (Tyrosyl-DNA phosphodiesterase 1), which removes TOP1 (DNA Topoisomerase I) covalently bound to DNA (Yang et al. 1996; Pouliot et al. 1999). Furthermore, NER had been shown to be responsible for removing small protein adducts on DNA (up to 10 kDa) (Baker et al. 2007; Nakano et al. 2007), whereas HR provided tolerance to DPC causing agents (de Graaf, Clore, and McCullough 2009) likely via promoting repair of replication forks that are stalled or collapsed in the presence of DPCs. Recently, a novel DPC repair pathway was identified which involves DPC removal via proteolytic activity of the protease SPRTN together with the proteasome in a replication-dependent manner (Duxin et al. 2014; Larsen et al. 2019).

1.2.6 – Single-strand break repair

DNA lesions can occur at the base and nucleotide level but the DNA duplex can also be broken in one or both strands. If a break occurs in one of the DNA strands, a single-stranded DNA break (SSB) occurs. This type of lesions can also arise via direct sugar damage caused by reactive oxygen species (ROS) (Li, Trush, and Yager 1994) and from erroneous or abortive activity of cellular enzymes such as TOP1, which generates a nick in DNA to relax DNA during transcription and DNA replication (Pommier et al. 2003). The repair of this lesion requires a different DNA repair pathway named single-stranded break repair (SSBR) which has a partial overlap with the BER

in the enzymatic machinery required to process the DNA ends, fill and ligate the DNA gap (Caldecott 2008).

1.2.7 – Double strand break repair

If the DNA duplex is broken on both DNA strands, a DNA DSB is formed. DSBs are highly toxic lesions which can cause genomic instability via chromosomal rearrangements and can trigger cell cycle arrest or cell death via apoptosis (Stephens et al. 2011; Lips and Kaina 2001; Di Leonardo et al. 1994). DSBs arise following exposure to exogenous clastogens, such as ionising radiation (IR) or ROS. Also, DSBs can also be generated endogenously when the replication fork encounters unrepaired DNA lesions causing fork collapse or when DNA interacting enzymes are not removed from DNA causing the DNA replication or transcription machinery to be blocked and ultimately released by endonucleases. DSBs can be repaired generally via two distinct repair pathways: homologous recombination (HR) or nonhomologous end-joining (NHEJ). The choice of DSB repair pathway is dependent on chromatin context and phase of the cell cycle. HR is the pathway normally used by cells in the G2 phase of the cell cycle when sister chromatids are available for serving as templates for repair. There are several subpathways of HR including single-strand annealing (SSA), synthesis dependent strand annealing (SDSA) or break induced replication (BIR) and studies have shown the complexity and mechanistic details of this repair pathway that are beyond the scope of this thesis.

1.2.7.1 – Homologous recombination

In order for cells to repair a DSB via HR, several steps involving different machinery are required. HR mediated repair starts by processing of the DSB by nucleolytic resection. This resection generates a 3′ single-strand DNA (ssDNA) overhang, which becomes the substrate for the HR protein machinery to execute strand invasion of a partner chromosome. RAD51 proteins coat ssDNA forming a filament that guides strand invasion into a homologous sequence and a D-loop intermediate is formed. Then the replicative DNA polymerase extends from the 3′ end of the invading strand permitting then the capture of the second DSB end by annealing to the extended

D-loop. Finally, two crossed strands or Holliday junctions (HJs) are formed. HJs can be “dissolved” by the branch migration and topoisomerase activity of the BLM (Bloom syndrome RecQ like helicase)/TOP3a (DNA topoisomerase III alpha)/RMI1 (RecQ mediated genome instability 1) complex (Wu and Hickson 2003) and intermediates that escape this complex can be resolved by several different resolvases, including MUS81/EME1, GEN1 (GEN1 Holliday junction 5' flap endonuclease), and SLX1/SLX4 (Ho et al. 2010; Wechsler, Newman, and West 2011; De Muyt et al. 2012; Zakharyevich et al. 2012), the choice of which may be cell-cycle regulated (Matos et al. 2011). In addition to these resolvases, the nuclease XPF-ERCC1 complex has been shown to play a role in a subset of HR named SSA (Sargent et al. 2000). In SSA, the DSB occurs in two repeated sequences that after resection allow for the 3' overhangs to align and anneal to each other, restoring the DNA as a continuous duplex. The XPF-ERCC1 nuclease complex has been suggested to promote the removal of the nonhomologous 3’ ssDNA tails to complete repair of the DSB by SSA (Sargent et al. 2000). In addition to this, the XPF-ERCC1 nuclease complex has been proposed to play a role later in HR by nicking the D-loop during HJ formation or removing 5’ flaps during SDSA (Al-Minawi, Saleh-Gohari, and Helleday 2008).

In various specialised contexts, DSBs are programmed by the cell. For example, during meiosis, which entails a type of cell division specific to germ cells, the generation of DSBs and DSB repair is essential for correct chromosome pairing and segregation at the first meiotic division and also for gamete generation (Neale and Keeney 2006). This physiological process generates gametes with allele combinations distinct from the parental germ line therefore increasing genetic variability. Also DSB-induced rearrangements at immunoglobulin genes are also critical for the multiplicity of antigen receptor diversity over limited numbers of loci (Dudley et al. 2005).

1.2.7.1 – End joining

NHEJ repairs DSBs via direct ligation of the broken ends of DNA without a homology template (Chang et al. 2017) and, in contrast to HR, NHEJ occurs in the G1 phase of the cell cycle. After the DSB is generated, a protein complex slides onto the DNA end and translocates inward (Chang et al. 2017). The DNA end is then processed by removing damaged or mismatched nucleotides by nucleases and resynthesized by DNA polymerase activity (Chang et al. 2017). After end processing, the ends are then ligated via DNA ligase activity therefore rejoining the broken strands and repairing the DSB. Unlike HR in which a DNA molecule is used as template for repair, NHEJ joins together two broken helices, therefore being considered a more error-prone DSB repair pathway.

In addition to NHEJ another repair pathway named microhomology-mediated end joining (MMEJ) or alternative nonhomologous end-joining can repair DSBs. MMEJ differs from HR and NHEJ by the use of 5–25 base pair micro-homologous sequences to align the broken strands before joining. This results in frequent deletions and occasionally insertions which are much larger than those produced by NHEJ. MMEJ is completely independent from NHEJ and does not rely on NHEJ core factors (Simsek and Jasin 2010).

1.2.8 – Translesion synthesis

Finally, all the above repair mechanisms act by detecting and repairing lesions in the genome. However, cells possess another important mechanism that permits to tolerate genomic insults. This DNA damage tolerance mechanism does not repair lesions but allows the cell to temporarily bypass the lesion. This provides a mechanism by which a cell can complete replication of the genome even when that genome contains damage, which is named translesion synthesis (TLS) (Yang and Gao 2018). Whenever absent, TLS causes cells to face the risk of replication fork collapse, translocations, chromosome aberrations, and cell death. Although this mechanism allows cells to survive, the cost is mutagenesis. This is because TLS makes use of a class of specialized DNA polymerases which can use damaged DNA as templates and insert nucleotides opposite to lesions, despite the conformational constraints that many modified bases

may impose (Prakash, Johnson, and Prakash 2005). TLS polymerases differ from the replicative polymerases by being able to accommodate modified bases, therefore allowing the insertion of a base opposite the lesion. However, this very property causes TLS polymerases to become intrinsically more error-prone than the replicative polymerases. The low fidelity of TLS is therefore due partly to bypassing past lesions, the reduced fidelity of TLS polymerases and the fact that the TLS polymerases do not have proofreading activity (Cleaver et al. 1999; Limoli et al. 2002; Limoli, Laposa, and Cleaver 2002). On the other hand, this tolerance pathway can be advantageous in specific physiological scenarios as demonstrated in somatic hypermutation of immunoglobulin genes by increasing desired genetic variability. A critical player in the TLS pathway is the deoxycytidyl transferase REV1 (UV-induced reversion 1), an enzyme that inserts deoxycytidine (dC) across from DNA lesions (Nelson, Lawrence, and Hinkle 1996a). The importance of this factor in DNA damage tolerance is clear as absence of *Rev1* in cells causes hypersensisitivity to DNA damaging agents (Lemontt 1971; Simpson and Sale 2003). However, cells expressing a catalytic dead version of REV1 do not cause hypersensitivity to DNA damaging agents, suggesting an alternative function of *Rev1* is required for genome stability (Ross, Simpson, and Sale 2005). Indeed, the extreme carboxyl terminus of vertebrate REV1 has been shown to interact with TLS DNA polymerases (Guo et al. 2003; Ohashi et al. 2004; Tissier et al. 2004) and with DNA polymerase ζ via its REV7 (UV-induced reversion 7) subunit (Murakumo et al. 2001). Interestingly, the deletion of this domain causes no sensitivity to DNA damaging agents suggesting that the recruitment of TLS enzymes via REV1 permits to bypass lesions caused by genotoxic agents (Ross, Simpson, and Sale 2005). In addition to *Rev1*, post-translational modifications of PCNA (proliferating cell nuclear antigen) provide a scaffold to which the various TLS polymerases bind to gain access to the replicative ensemble stalled at the lesion site and to execute their roles in lesion bypass (Hoege et al. 2002; Stelter and Ulrich 2003; Kannouche, Wing, and Lehmann 2004; Watanabe et al. 2004; Bi et al. 2006; Garg and Burgers 2005).

Overall, the evolution of highly specialised pathways permitted cells to repair the diverse set of DNA lesions they are exposed to thereby preventing cell death. Furthermore, another critical role for DNA repair is to suppress mutations from arising.

1.3 - DNA repair and mutagenesis

DNA mutations exist in different forms: there are base substitutions or point mutations, in which a single nucleotide is changed; insertions and deletions, in which a sequence of one or more nucleotides is inserted or deleted from the DNA sequence, respectively; and also chromosomal abnormalities, which can change the number and the structure of chromosomes in a cell. These mutations arise via five main mechanisms which include errors in DNA replication (generated by TLS or by errors made by high fidelity DNA polymerases), genotoxin exposure (endogenous or environmental mutagens), DNA enzymatic activity (*e.g.* cytidine deaminase enzymes which introduces C to T mutations) and errors in DNA repair. DNA repair causes mutagenesis via misrepair of DNA lesions, which occurs namely when DNA lesions are not accurately repaired and hence accumulate mutations at the previously damaged site. In addition, mutations may also arise in the absence of DNA repair factors as exemplified by the increase in mutation rate whenever the BER (Sobol et al. 2002) or MMR pathways are depleted (Kato and Nakano 1981). Furthermore, HR repair generates genomic rearrangements involving repeated DNA elements with identical (homologous) or near-identical (homeologous) sequences which may also generate genomic mutations (Putnam, Hayes, and Kolodner 2009). The importance of DNA repair in regulating mutagenesis is evident as failure to correctly repair DNA damage results in mutations that at a cellular level cause neoplastic transformation or death/dysfunction (Nakane et al. 1995; Kaina et al. 1997). At a physiological level there is loss of tissue homeostasis or an increased cancer predisposition (Reitmair et al. 1995; Prolla et al. 1998). In the past few years, large-scale analyses have revealed many mutational signatures across the spectrum of human cancer types (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al. 2013; Alexandrov, Nik-Zainal, Wedge, Campbell, et al. 2013; Nik-Zainal et al. 2012; Alexandrov and Stratton 2014). As examples of how failure to undergo DNA repair promotes mutagenesis is the emergence of mutational signature 3, which occurs in tumours lacking HR. This signature is associated with increased burden of large indels (up to 50 nucleotides) with overlapping microhomology at the breakpoints (Alexandrov, Nik-Zainal, Wedge, Campbell, et al. 2013). In such tumours, DSBs in DNA are repaired by the imprecise repair mechanisms of NHEJ or MMEJ instead of via the high fidelity HR repair. Also, signature 6, which is observed in tumours defective of MMR exhibiting micro-satellite instability, also

features enrichment of 1 base pair indels in nucleotide repeat regions (Alexandrov, Nik-Zainal, Wedge, Campbell, et al. 2013). Besides leading to cancer predisposition, the absence of DNA repair causes various syndromes in human patients with tissue specific phenotypes ranging from immunodeficiency, neurodegeneration, progeria and/or infertility.

1.4 - Tissue specificity in the DNA repair deficiency phenotypes

One example of the tissue specificity observed in DNA repair deficiency phenotypes is the absence of the repair factor *BRCA1* (Breast cancer 1) which is involved in repairing DNA DSBs via HR. Women carrying a germ line mutation in *BRCA1* have an increased risk of developing breast cancer from 12% to 72% and from 1.3% to 40% to develop ovarian cancer (Kuchenbaecker et al. 2017). The underlying cause of oncogenesis in these defined tissues in female patients is not entirely clear. However, there is accumulating evidence for deregulated hormonal signalling in *BRCA1* mutation carriers to drive oncogenesis. This is corroborated by the increase in cancer development during pregnancy, when there are high circulating levels of steroid hormones (Narod 2001) and a sharp decline in cancer risk after menopause (Antoniou et al. 2003). Evidence suggests that steroid hormones produced specifically in luminal progenitor cells promote cellular proliferation which therefore increases the chances of mutations to occur due to faulty DNA repair and/or replication leading to a biallelic mutation that promotes tumourigenesis (Nolan, Lindeman, and Visvader 2017). The development of cancer in defined tissues in the absence of DNA repair reflects a differential requirement for DNA repair mechanisms in different tissues. This tissue specificity for DNA repair has three main explanations: Firstly, some cells in our bodies are more exposed to certain sources of damage. To illustrate this, Xeroderma pigmentosum patients, who lack a form of DNA repair responsible to deal with damage caused by ultra-violet (UV) radiation, exhibit abnormalities in the skin tissue if exposed to sunlight due to increased mutagenesis in this tissue (de Vries et al. 1995). Secondly, cells of different tissues can use different repair mechanisms to repair the same lesion *e.g.* in Cockayne syndrome, patients lack the ability to perform transcription-coupled (TC) DNA repair of bulky DNA adducts, which causes these patients to develop neurodegeneration. One explanation for the specific requirement of TC repair in neurons may be the fact that neurons are post-mitotic cells and therefore cannot detect damage during replication. This may render neurons more dependent on TC repair than other cells, as this pathway may be the main or only mechanism to detect forms of damage in this cell type (Cockayne 1936). Thirdly, different cell types may have different DNA damage checkpoints that will respond to an insult in a differential manner. The DNA damage checkpoint is normally mediated by *Trp53* (Transformation-related protein p53), the so called guardian of the genome, which was

first described to be active upon DNA damage induced by UV irradiation (Kapoor and Lozano 1998). Furthermore, deletion of this gene confers cellular resistance to DNA damage which therefore avoid apoptosis and progress through the cell cycle (Lee and Bernstein 1993). Several DNA repair phenotypes in mice can be suppressed by co-depleting *Trp53*. To illustrate this, *Lig4* (DNA ligase 4) deficient mice display embryonic lethality due to neuronal apoptosis (Barnes et al. 1998). These mice also show defects in lymphocyte development and *Lig4*-deficient cells exhibit defective proliferation, senescence and sensitivity to ionising radiation (Barnes et al. 1998). It has been shown that co-depletion of *Trp53* together with *Lig4* can suppress the neuronal apopotosis phenotype rescuing the embryonic lethality. However, it does not rescue the lymphocyte development and radiosensitivity phenotypes observed in the single-knockout mice (Frank et al. 2000). In addition to *Trp53,* two other genes from the same family exist: *Trp63* and *Trp73* (Belyi et al. 2010). Evidence suggests that both genes can also mediate DNA damage checkpoints (Bolcun-Filas et al. 2014; Urist et al. 2004). Interestingly, *Trp63* but not *Trp53* can rescue the oocyte numerical defect in ovaries treated with ionising radiation (Suh et al. 2006). This therefore suggests that different cell types may have different DNA damage checkpoints further highlighting the importance of the tissue specificity in the DNA damage response.

1.5 - The importance of DNA repair in the germ line

The tissue specificity observed in DNA repair suggests that different tissues may also have different rates of mutagenesis. Moreover, the tissue or cell type in which a mutation occurs is of critical importance. Whenever a mutation occurs in a somatic cell, the organism can develop a particular pathology linked to that mutation. However, the consequence of that mutation will be confined to a single organism and have no further consequences for the progeny. On the other hand, if a mutation occurs in the germ line, this mutation can be passed on to the progeny. If this mutation is deleterious it may negatively impact the fitness of the next generations but on the other hand, if a neutral or beneficial mutation occurs, it can drive the evolution of that species. Therefore, from an evolutionary standpoint, it is beneficial for a species to maintain the integrity of the genome in the germ line in order to protect the fitness of the future generations and the survival of the species but at the same time providing sufficient genetic diversity to permit potential advantageous phenotypes in changing selective pressures. In the "theory of the disposable soma" proposed in 1977 by T.B. Kirkwood he postulated that high-fidelity quality control that reduces errors made during somatic maintenance is costly and that these costs compete with the costs of reproduction. Because resources are limited, investing into error-proof somatic maintenance is wasteful and not an evolutionarily stable strategy. In this sense, the soma is disposable and investment into somatic maintenance has to be optimised to allow error-prone repair in order to invest the rest of the limited resources into reproduction (Kirkwood 1977). This postulate suggests that organisms invest and have evolved mechanisms to maintain the integrity of the genome of germ cells pristine for the subsequent generations at the expenditure of more energy and in a differential manner compared to how the genome is protected in the disposable somatic cells. To corroborate this theory, evidence exists showing that mammalian germ cells have a reduced mutation frequency compared to the somatic counterparts (Kohler et al. 1991; Dycaico et al. 1994; Hill et al. 2005). *De novo* mutations arise in the germ line at a rate of ~1-1.5 x 10^8 for point mutations in humans, not only determining our genome evolution but also causing 20% of genetic disorders present in live-born human offspring (Michaelson et al. 2012). *De novo* mutations in the germ line include point mutations but also numerical and structural chromosomal abnormalities, changes in gene dosage due to duplications or deletions or remodelling of chromosomes by translocations or

inversions. These mutations arise during germ cell development and exhibit sexual dimorphism. On the one hand, the rate of chromosomal abnormalities transmitted by the female is greater than the rate passed by the male germ line (Hassold and Hunt 2001). The majority of these abnormalities results from defects during meiosis. Meiosis is a significant threat to the stability of the genome as this process necessitates programmed DSB formation. The initiation of these breaks and the resolution of these very toxic lesions are essential to the meiotic process, however pose a significant threat to genome integrity and therefore necessitate the DNA repair machinery. Failure to correctly repair meiotic intermediates is an important source of structural abnormalities in the germ line. Even though meiosis occurs in both males and females it is the female germ line that is responsible for passing the majority of structural abnormalities to the progeny (Hassold and Hunt 2001). This is largely explained by the fact that all oocytes initiate meiosis during embryonic development but arrest immediately after synapses of homologous chromosomes and initiation of recombination. Oocytes, already committed to meiosis, remain arrested in prophase I until completion of meiosis, which occurs years later in the ovary of sexually mature females, prior to ovulation. Since this period can exceed 40 years in humans, it greatly increases the risk of abnormalities in the meiotic spindle, which have been further shown to accumulate with age (Battaglia et al. 1996). This explains the increased prevalence of structural abnormalities in the female germ line. On the other hand, it has been known for more than 60 years that the male germ line is responsible for the generation of the majority of *de novo* mutations (Haldane 1947). These are more frequently point mutations and the increased burden is largely due to the increased the number of mitotic divisions that the spermatogonial stem cells (SSCs) undergo during self-renewal and proliferation (Drost and Lee 1995). In stark contrast to the female germ cells (estimated to undergo 31 rounds of division in humans), male germ cells undergo ~401 rounds of divisions, which increases the chance of mutations occurring during DNA replication, thereby explaining the increased rate of *de novo* mutagenesis in the male germ line (Drost and Lee 1995). Furthermore, ageing also contributes to the accumulation of point mutations in the male sperm, with an estimated rate of 2 new mutations per year in humans (Kong et al. 2012). Moreover, the age of the father at the time of conception correlates with the incidence of polygenic disorders like schizophrenia and autism in progeny (Sipos et al. 2004; Hultman et al. 2011). This suggests that the increased burden of mutagenesis can lead to severe phenotypic consequences in the progeny.

In addition to mature germ cells, primordial germ cells (PGCs), *i.e.* the precursors of egg and sperm that are specified early during embryonic development (**Figure 2**), may also be prone to mutagenesis. Indeed, a recent study showed that the highest mutation rate per cell division in the germ line of both males and females was observed during PGC development (Rahbari et al. 2016).

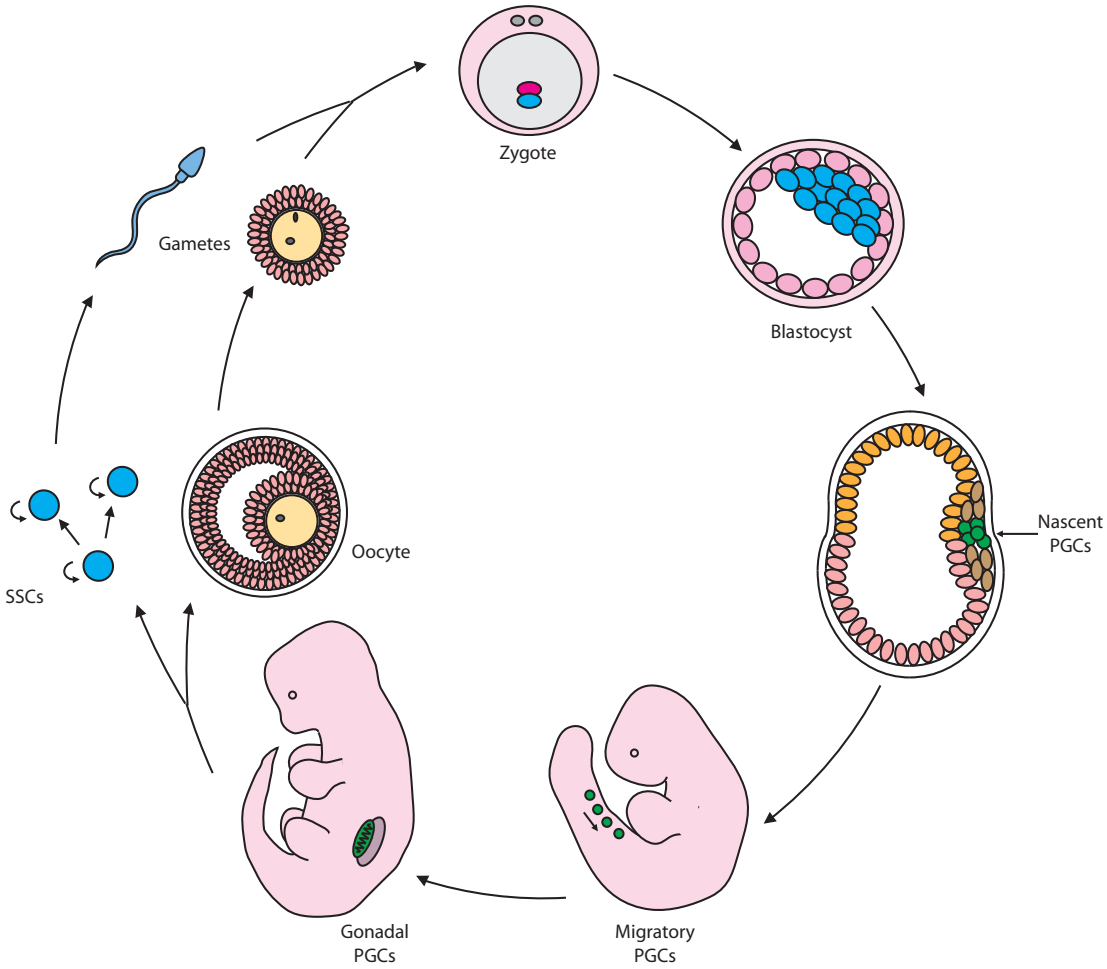


Figure 2. The mammalian germ cell lineage. The germ line is responsible for passing genetic information from one generation to the next. The germ line emerges as PGCs around E7.25 near the extra-embryonic ectoderm, at the base of the allantois. These cells migrate to the developing midgut at around E7.75, into the mesentery at E9.5 and colonise the genital ridges at ~E10.5. During migration, PGCs undergo an epigenetic reprogramming, most notably genome-wide DNA demethylation, including imprint erasure in both sexes (Saitou, Kagiwada, and Kurimoto 2012). In the fetal gonads a striking sexual dimorphism occurs in germ cell maturation. In females, PGCs proliferate until E13.5 in which 25,000 cells (oogonia) start prophase I of meiosis These cells are then named oocytes and arrest at the diplotene stage of prophase I (Borum 1967). Upon hormonal stimulation, meiosis progresses with the extrusion of the first polar body and finally, during fertilisation, the second polar body is extruded finishing the second meiotic division. Subsequent formation of the zygote and embryo progression reinitiates the germ cell cycle. In males, PGCs undergo a mitotic arrest around E13.5 remaining quiescent in G_0/G_1 (gonocytes) until postpartum day 3-7 (P3-7) (Hilscher et al. 1974; Western et al. 2008). At P3-7 gonocytes resume proliferation becoming spermatogonial stem cells (SSC). SSC proliferate and self-renew producing spermatocytes that meiotically divide to form spermatids. During spermiogenesis, spermatids differentiate into mature and functional gametes, spermatozoa, which can fertilise the egg and further reinitiate the cycle.

1.6 - Primordial germ cell development

In order to study the importance of DNA repair in PGCs and how germ line mutagenesis is suppressed in this particular cell population, one must first understand the unique biology of PGCs. PGCs are the precursors of egg and sperm which generate a new organism upon fertilisation. The specification of germ cells in animals occurs via two forms: in model organisms like *D. melanogaster*, *C. elegans*, *D. rerio*, and *X. laevis* the specification occurs via the inheritance of preformed germ plasm in a process named preformation or determinative mode of germ line establishment (Whittle and Extavour 2017). However, in mammals and axolotls, PGCs are specified by epigenesis which occurs via induction among equipotent cells in the embryo by instructive signals (Whittle and Extavour 2017). In mice, PGCs originate from the early postimplantation epiblast cells, which also gives rise to all somatic cells in response to signals from the extraembryonic tissues (Lawson and Hage 1994). In order to achieve totipotency, PGCs undergo extensive epigenetic reprogramming which permits the repression of the somatic transcriptional programme and activation of pluripotency genes such as *Nanog, Oct4* (Octamer-binding protein 4)*, cMyc* (Avian Myelocytomatosis Viral Oncogene Homolog) and *Klf4* (Kruppel like factor 4) (Ohinata et al. 2005; Yabuta et al. 2006; Yamaguchi et al. 2005). Additionally, this epigenetic reset leads to the erasure of imprints, causes the reactivation of the X-chromosome and is believed to turn on specific genes required for the progression towards gametogenesis (Monk and McLaren 1981; Hill et al. 2018). The maternal and paternal imprints are erased in PGCs to allow new imprints to be established (Lee et al. 2002). These imprints are not of parental origin but are determined by the sex of the developing embryo. This process is not only essential to make functional germ cells, competent of generating a zygote capable of normal development, but also prevents parthenogenesis (Surani, Barton, and Norris 1984). In female somatic cells, one X-chromosome is inactivated, which balances X-linked gene expression with that of the male XY cells (Lyon 1961). This process must be reversed in the germ line so that both X chromosomes are active following meiosis. The epigenetic reprogramming taking place in gonadal PGCs has also been recently shown to promote the activation of a subset of genes, named germ line reprogramming-

responsive genes which are involved mainly in gamete generation and meiosis, therefore permitting sexual reproduction (Hill et al. 2018).

1.6.1 - Epigenetic reprogramming

In more detail, epigenetic reprogramming entails the erasure and remodelling of epigenetic marks, including DNA methylation but also covalent modifications of histones in order to permit the acquisition of pluripotency. This occurs during mammalian development and can also be induced in tissue culture. Epigenetic reprogramming *in vitro* can be recapitulated to some extent in three different systems: somatic cell nuclear transfer (SCNT) (Gurdon 1962), induced pluripotent stem cell (iPSC) generation (Takahashi and Yamanaka 2006) and cell fusion (Tada et al. 2001). The key feature outlining *in vivo* reprogramming processes are also observed *in vitro* suggesting that similar molecular pathways are involved. However, the efficiency in generating pluripotent cells *in vitro* is very low and time consuming in comparison to reprogramming observed *in vivo. In vivo* epigenetic reprogramming occurs in two distinct stages: during early mammalian pre-implantation development and also during embryonic germ cell development. During early mammalian pre-implantation development there is firstly the erasure of DNA methylation in the paternal pronucleus of the zygote a few hours after the fertilisation (Mayer et al. 2000); secondly, a gradual (passive) loss of DNA methylation during the cleavage stages (Rougier et al. 1998); and thirdly, reshaping of heterochromatin based silencing in the pluripotent cells of inner cell mass (ICM)/early epiblast connected with the re-activation of inactive X chromosome in female embryos (Okamoto et al. 2004). Epigenetic reprogramming in early pre-implantation embryos culminates with the appearance of pluripotent cells of ICM in the blastocyst. The second time in development in which epigenetic reprogramming occurs is during the development of the germ line. The developing germ cell lineage has the unique property of being able to reacquire totipotency as evidenced by the ability to generate gametes which will give rise to a new organism. This is achieved after specification of the germ line into PGCs and happens in two distinct phases of DNA demethylation: Phase I, which occurs in pre-gonadal PGCs from ~E8.0 to E10.5 in migratory PGCs and Phase II, which occurs in gonadal PGCs from E10.5 to E13.5 (**Figure 3**) (Hargan-Calvopina et al. 2016; Seisenberger et al. 2012). Epigenetic reprogramming in PGCs has been suggested to occur in a DNA repair-dependent manner (Hajkova et al. 2010). This therefore may contribute for mutagenesis control in the germ line. In order to understand how epigenetic reprogramming occurs in PGCs and how this process affects germ line mutagenesis it

is essential to understand how these cells are formed and develop before becoming mature germ cells.

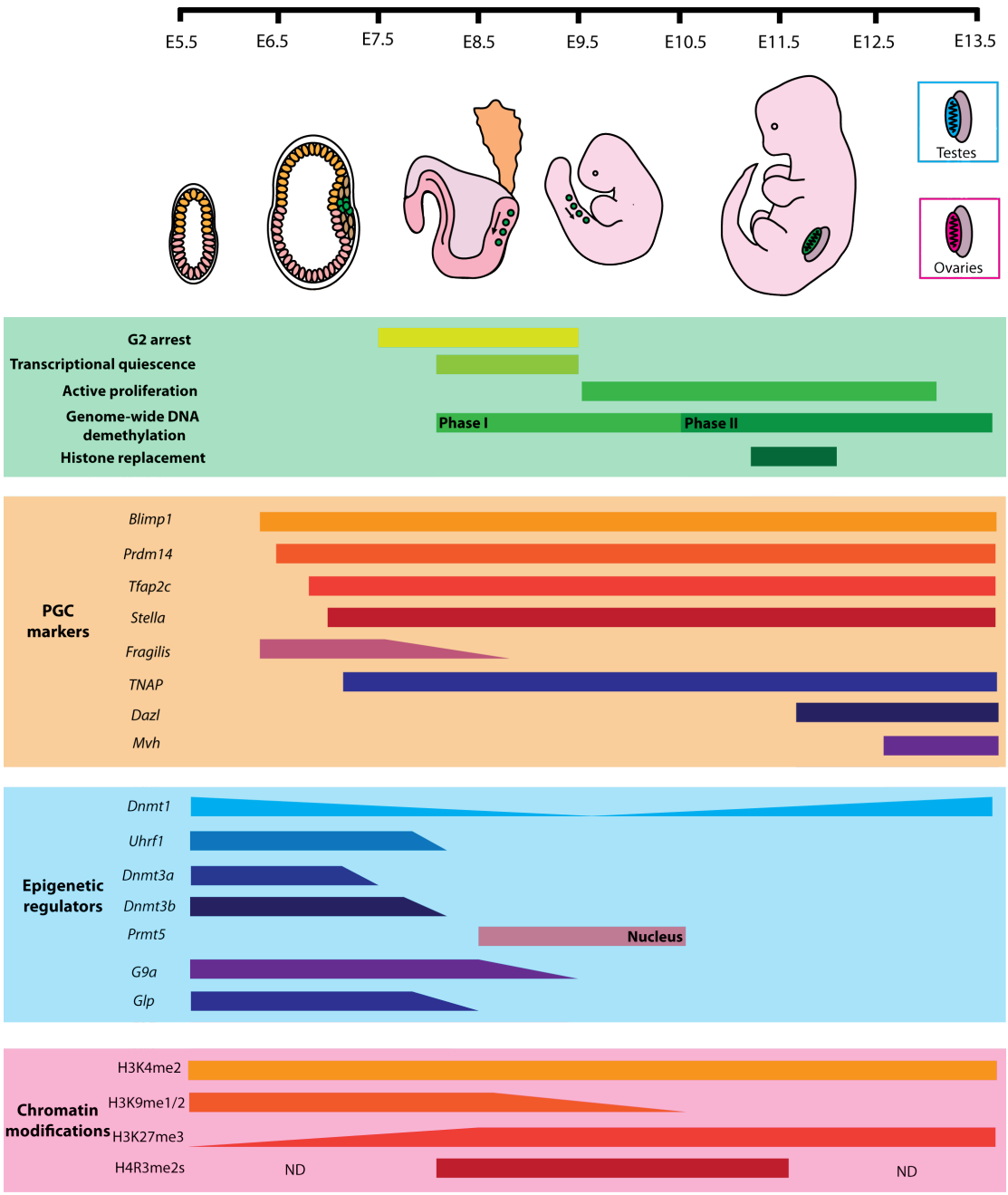


Figure 3 - A summary of mouse PGC development. A summary of the main genomic events occurring in PGCs from E5.5 to E13.5: temporal expression patterns of key proteins involved in PGC specification and epigenetic regulators responsible for epigenetic reprogramming; and chromatin modifications occurring from PGC specification to gonocyte formation. ND - not determined.

1.6.2 - Specification

In mice, the progenitors of PGCs are located in the most proximal region of the postimplantation epiblast, close to the extraembryonic tissues. These PGC precursors are detected in the epiblast at embryonic day (E) 6.0-6.5, and eventually generate a founding population of ~40 PGCs at the base of the incipient allantois at ~E7.25 (Chiquoine 1954; Ginsburg, Snow, and McLaren 1990; Lawson and Hage 1994). The instructive signals for PGC specification are the bone morphogenetic proteins BMP4 (Lawson et al. 1999; Fujiwara, Dunn, and Hogan 2001; Ohinata et al. 2009) and BMP8B (Ohinata et al. 2009; Ying et al. 2000) from the posterior extra-embryonic ectoderm and BMP2 (Ying and Zhao 2001) from the visceral endoderm. These cytokines bind to membrane receptors that phosphorylate SMAD1 and SMAD5, which dimerise then with SMAD4, translocate to the nucleus and induce transcriptional regulators of PGC fate (Hayashi et al. 2002; Lawson et al. 1999; Tam and Snow 1981; Ying, Qi, and Zhao 2001). These include *Blimp1* (B-lymphocyte-induced maturation protein 1) initially expressed at ~E6.25 (Ohinata et al. 2005), *Prdm14* (PR domain-containing protein 14) at ~E6.5 (Yamaji et al. 2008) and *Tcfap2c* (Transcription factor AP-2 gamma) at ~E6.75 (Kurimoto et al. 2008; Auman et al. 2002). *Blimp1* is a transcriptional repressor that is expressed in a few most proximal epiblast cells at ~E6.25, marking the onset of PGC specification (Lawson and Hage 1994). *Blimp1* deficiency in mice results in a small cluster of aberrant PGC-like cells at ~E8.5 that resemble the neighbouring somatic cells due to failure to repress somatic gene expression and to induce PGC-specific genes (Ohinata et al. 2005). A comparison between Wild type and *Blimp1*-deficient PGCs was subsequently critical for the identification of *Prdm14* (Grabole et al. 2013; Kurimoto et al. 2008; Yamaji et al. 2008). *Prdm14* is also a transcriptional repressor and its depletion from mice leads to the formation of aberrant PGCs that are lost by ~E11.5. These cells exhibit a defective epigenetic programme, based on the failure of the characteristic global erasure of H3K9me2 histone modification at ~E8.5, partly because the *Ehmt1-Ehmt2* (Euchromatic histone lysine methyltransferase) mediated H3K9 methylasae activity was not repressed. *Prdm14*-deficient PGCs also fail to show genome-wide induction of the EZH2 (Enhancer of zeste 2 Polycomb repressive complex 2 subunit) mediated H3K27me3. This suggests that *Prdm14* is important at least for the epigenetic programme in early germ cells (Hajkova et al. 2008; Hajkova et al. 2010; Yamaji et al.

2008; Seki et al. 2007). Furthermore, *Blimp1* expression is sustained by *Prdm14*, which might explain the induction of somatic genes in mutant cells (Grabole et al. 2013). *Prdm14* also induces the expression of *Stella* and *Sox2* (Sex determining region Y-Box 2) at ~E8.5. Thus, neither PGC-specific gene expression nor the re-initiation of pluripotency gene expression occurs in *Prdm14*-null PGCs (Grabole et al. 2013; Yamaji et al. 2008). Finally, *Tcfap2c*, which encodes AP2γ (a direct target of BLIMP1), is a basic helix-span-helix domain transcription factor also crucial for PGC specification (Kurimoto et al. 2008; Magnusdottir et al. 2013; Weber et al. 2010). Depletion of *Tcfap2c* in mice causes a PGC numerical defect around E8.0 (Weber et al. 2010). Although these cells remain to be fully characterised, it is likely that AP2γ might also be involved in the repression of somatic genes, including the early mesodermal marker *Hoxb1* (Homeobox b1)(Weber et al. 2010). *Blimp1*, *Prdm14* and *Tcfap2c* together constitute a mutually interdependent transcriptional network that promotes PGC specification (Magnusdottir et al. 2013) via repressing somatic genes and promoting the totipotent state. In addition to these germ line specifiers, single-cell analysis of gene expression of the founder population of PGCs at E7.25 identified two genes, namely *Fragilis* and *Stella*, which are highly and specifically, respectively, expressed in PGCs (Saitou, Barton, and Surani 2002). *Fragilis* (also known as *Ifitm3* - Interferon induced transmembrane protein 3) (Tanaka and Matsui 2002) is a member of the interferon-inducible transmembrane proteins. This gene begins to exhibit expression around the most proximal epiblast cells at E6.25–E6.5, and its expression intensifies in the posterior extra-embryonic mesoderm, where alkaline phosphatase-positive PGCs arise at E7.0–E7.25. *Stella* (also known as *Dppa3* - Developmental pluripotency associated 3) is a small highly basic nucleo–cytoplasmic shuttling protein (Sato et al. 2002). *Stella* begins to be expressed specifically in *Fragilis*-expressing cells in the extra-embryonic mesoderm at E7.0–E7.25 and continues to be expressed in migrating PGCs. *Stella*-positive cells show high expression of *Tnap* (tissue nonspecific alkaline phosphatase), a gene for AP activity of PGCs (MacGregor, Zambrowicz, and Soriano 1995). *Stella*-positive cells and high levels of *Fragilis* repress the expression of Homeobox genes such as *Hoxb1* and *Hoxa1*, whereas *Fragilis*-positive but *Stella*-negative cells retain *Hox* expression (Saitou, Barton, and Surani 2002). Thus, it has been proposed that the *Stella*-positive and *Hox*-negative cells are the established PGCs (Saitou, Barton, and Surani 2002). However, knocking out both *Fragillis* and *Stella* in mice shows that neither is essential for PGC specification (Payer et al. 2003; Lange et al. 2008). Instead,

Stella has been found to be a critical factor to protect the maternal genome and paternally imprinted genes from genome-wide DNA demethylation that occurs in the zygotes (Nakamura et al. 2007; Huang et al. 2017; Li, Zhang, et al. 2018), whereas *Fragilis* has a critical function in restricting the replication of multiple pathogenic viruses including influenza (Brass et al. 2009). After specification, PGCs start DNA demethylation, which permits the acquisition of the totipotent state.

1.6.3 - Migratory PGCs (E7.75 to E10.5)

Once specified, PGCs migrate to the developing midgut at around E7.75, into the mesentery at E9.5 and colonise the genital ridges at ~E10.5. During this migratory period, PGCs initiate a progressive cell-by-cell genome-wide DNA demethylation process which is associated with extensive chromatin remodelling, transcriptional quiescence and an arrest in the G2 phase of the cell cycle (Seki et al. 2007). It has been demonstrated that nascent PGCs downregulate *Glp*, a cofactor of EHMT2 H3K9 histone methyltransferase, leading to a genome wide loss of H3K9me2 specifically in early PGCs (Hajkova et al. 2008; Seki et al. 2007). The combination of a lack of the common cross-talk between H3K9me2 and DNA methylation (Cedar and Bergman 2009), and the additional observed downregulation of *de novo* DNA methyltransferases *Dnmt3a* and *Dnmt3b* in nascent PGCs (Kurimoto et al. 2008), may explain the reduction in DNA methylation after PGC specification. Considering these observations, a model of passive DNA demethylation has been proposed for migratory PGCs (Seisenberger et al. 2012). However, the complete lack of both maintenance and *de novo* DNA methylation would lead to much faster and extensive DNA demethylation than is observed in migratory PGCs (Seisenberger et al. 2012; Yamaguchi et al. 2013; Kagiwada et al. 2013). Additionally, some of the genomic regions, such as genomic imprints, seem to maintain DNA methylation levels throughout this period of PGC development (Hajkova et al. 2010; Seisenberger et al. 2012; Hajkova et al. 2008; Guibert, Forne, and Weber 2012). This suggests that the mechanism implicated in the loss of 5mC in pre-gonadal PGCs is likely to be more complex. In addition, during migration pre-gonadal PGCs undergo a period in which RNA polymerase II is quiescent (Seki et al. 2007). On the one hand this is believed to impede the toxicity of transposable elements that become active during DNA demethylation and on the other hand to restrict the transcriptional programme of PGCs impeding these cells to differentiate to somatic cells. This repression can be accomplished by the Polycomb family proteins through promoter inhibition and consequent chromatin remodelling, which leads to inactivation of transcription. Another level of RNA polymerase II-depended transcriptional repression in migratory PGCs is chromatin-based repression through post-translational modifications of histones and increased chromatin compaction. A prominent member of this class of proteins is HP1 (Heterochromatin protein 1), which is represented by several variants in Metazoa. The absence of *Cbx3* (Chromobox 3) also

known as HP1γ (Heterochromatin protein 1 gamma) in mice causes failure to produce normal numbers of PGCs at E.7.25 which emphasises the importance of HP1 during PGC development (Abe et al. 2011). Nevertheless, there is no link between the RNA pol II quiescence and chromatin modification and Seki *et al.* observed no alterations in H3K4me2 and H3K4me3 that were previously associated with transcriptionally permissive/active states compared to the somatic counterparts. This therefore suggests that the RNA pol II quiescence may be independent of the chromatin status. Interestingly, a majority of migrating PGCs (~60%) from around E8.0 to around E9.0, are shown to be in the G2 phase of the cell cycle. This has also been shown to occur in PGCs collected from pig embryos and in mouse PGCs induced *in vitro* from pluripotent stem cells (Hyldig et al. 2011; Hayashi et al. 2011). *Blimp1* may be a key factor for the G2 arrest, since it is involved in the repression of S-phase promoting factors *e.g. Ccne1* and *Ccnd1* (Cyclin E1 and D1), *Cdc25a* and *Cdc6* (Cell division cycle 25A and 6), *Pold2* (DNA Polymerase delta 2), *E2F3* (E2F transcription factor 3), and *Myc* (Kurimoto et al. 2008) These observations indicate that PGC specification creates a unique cellular state for the epigenetic reprogramming: PGCs repress key active epigenetic enzymes, arrest at the G2 phase of the cell cycle, and pause RNA Pol II-based transcription. However, the interplay between these three processes in migratory PGCs remains largely unexplored.

1.6.4 - Gonadal PGCs (E10.5 to E13.5)

The last major event that characterises early germ line development is the rapid and global epigenetic reprogramming that occurs in PGCs upon colonisation of the gonads where the germ cells will reside throughout the rest of embryonic and adult life. This process, also named as phase II of reprogramming, includes DNA demethylation which involves the erasure of parental imprints, demethylation of both single-copy genes and transposable/repetitive elements and reactivation of inactive X chromosome in females (Guibert, Forne, and Weber 2012; Monk, Boubelik, and Lehnert 1987; Kafri et al. 1992). Following gonadal epigenetic reprogramming, the genome of PGCs shows the lowest level of genome-wide DNA methylation seen at any point of development. This global hypomethylated state is long lived and lasts until postnatal development in female germ cells that start to reacquire DNA methylation only during the oocyte growth in postnatal ovaries (Kelsey and Feil 2013); on the contrary, male germ cells

start to accumulate DNA methylation relatively quickly only a few days following the global reprogramming (Kelsey and Feil 2013).

1.7 - Molecular mechanisms of DNA demethylation and link to DNA repair

The molecular mechanism underlying the active genome-wide removal of DNA methylation in PGCs remains subject of intensive study. Conceptually, the mechanism of DNA demethylation can be separated by their dependence on DNA replication: DNA methylation patterns are typically maintained during DNA replication in a faithful manner due to the activity of the *Dnmt1* (*de novo* DNA methyltransferase 1) which is tethered by *Uhrf1* (Ubiquitin like with PHD and ring finger domains 1) to the replication fork and hemi-methylated DNA (Sharif et al. 2007; Bostick et al. 2007). Loss of this maintenance methylation activity results in passive DNA demethylation, a gradual loss of DNA methylation which has been demonstrated in systems lacking *Dnmt1* or *Uhrf1*. This system explains the demethylation of the maternal pronucleus in the early pre-implantation embryo but fails to explain the rapid replication-independent paternal pronucleus demethylation and DNA demethylation in gonadal PGCs. Therefore, as an alternative mechanism, active DNA demethylation would lead to the removal of 5mC in a replication independent manner. In flowering plants the 5mC specific glycosylases Dme (DEMETER) and Ros1 (Repressor of transcriptional gene silencing 1) are used to actively remove 5mC from DNA (Morales-Ruiz et al. 2006). However, the family of 5mC specific DNA glycosylases have evolved independently in flowering plants with no direct sequence homology in Metazoa. Since it was observed that *Tet1* (Ten-eleven translocation 1) is upregulated in PGCs (Hajkova et al. 2010; Hackett et al. 2013) a mechanism of Tet-dependent active demethylation in PGCs with TET enzyme-dependent modification of 5mC to 5hmC subsequently being removed via BER was proposed. This could occur via different mechanisms. There is evidence that the cytosine deaminases AID (Activation induced cytidine deaminase and APOBEC1 (Apolipoprotein B mRNA editing enzyme catalytic subunit 1) can convert 5mC to thymidine (T) by deamination, creating thereby T:G mismatches, which then become targets of glycosylases such as TDG or MBD4 (Methyl-CpG binding domain 4). *Tdg* knockout mice die during embryonic development (Cortazar et al. 2011; Cortellino et al. 2011). Furthermore, *Tdg* has been shown to maintain the unmethylated state of CpG islands in promoters of developmentally regulated genes (Cortazar et al. 2011) and in promoters of Wild type cells, TDG forms a complex with BER pathway

components, including XRCC1 (X-ray repair cross complementing 1), APE1 (Apurinic-apyrimidinic endonuclease 1) and PARP1 (Poly(ADP-ribose) polymerase 1), and with the transcription-activating histone acetyltransferase CBP/p300 and the H3K4-specific Lysine methyltransferase 2A (*Kmt2a*). Moreover, TDG-deficient mouse embryonic fibroblasts (MEFs) show an iPSC conversion defect similar to Tet triple knockout (TKO) MEFs, supporting the model of DNA demethylation through TDG driven excision of higher 5mC oxidative derivatives (Hu et al. 2014). However, to counterpoint this, depletion of TDG in early zygotes does not affect active paternal pronucleus DNA demethylation (Guo et al. 2014). *Tdg* interacts with *Gadd45a* (Growth arrest and DNA damage inducible alpha) which is another factor reported to play roles in DNA demethylation (Li et al. 2015) and *Aid. Aid*-deficient male and female PGCs at E13.5 show median methylation levels of ~22% and 20%, respectively, which are both higher than the methylation levels of Wild type male and female PGCs (16.3% and 7.8%, respectively), indicating that AID functions in genome-wide DNA demethylation in PGCs (Popp et al. 2010). Importantly, *Aid* deficiency does not impact genome-wide methylation levels in cells/tissues other than PGCs (Popp et al. 2010). However, AID is expressed at a very low level in PGCs (Hajkova et al. 2008; Muramatsu et al. 2000) and *Aid*-deficient mice are fertile showing that *Aid* does not play an essential role in germ cell development (Muramatsu et al. 2000). As the methylation levels of AID-deficient PGCs are still lower compared to those of earlier Wild type PGCs, the demethylation events occur even without AID, possibly owing to compensation by other deaminases, including APOBEC1/2/3. Furthermore, it was reported that BER actively triggers DNA demethylation as evidence shows components of the BER pathway such as XRCC1, APE1 and PARP1 specifically concentrated in the nuclei of gonadal PGCs (Hajkova et al. 2010). To corroborate this, it has been recently shown that absence of the BER factor *Xrcc1* in mouse zygotes results in DNA lesions in the paternal pronucleus in a *Tet3*-dependent manner (Ladstatter and Tachibana-Konwalski 2016). This further suggests that there may be a requirement for BER to process DNA lesions generated during active DNA demethylation. Whether this process is intimately linked to active DNA demethylation in zygotes and PGCs still remains to be clarified.

1.8 - Evidence of genomic instability in developing PGCs

Even though it still remains unclear how DNA demethylation occurs in PGCs and how this is dependent on DNA repair, it is established that during PGC development there is a specific requirement for DNA repair to ensure the correct formation of this cell population. This is demonstrated by the sterility phenotypes observed in DNA repair deficient mice due to PGC numerical defects (**Table 1**).

Gene	Ontology	Time of defect	PGC phenotype	Reference
Rev7	TLS	E9.5	ND	Pirouz, M., *et al.* (2013)
Espl1	Cell division	E12.5	Proliferation	Huang, X., *et al.* (2008)
Pds5b	Cohesion	≤ E12.5	Proliferation	Zhang, B., *et al.* (2007)
Fanca	FA	E11.5	Survival/proliferation	Wong, J.C., *et al.* (2003)
Fancb	FA	E9.5	Survival/proliferation	Kato, Y., *et al.*(2015)
Fancl	FA	E9.5	Survival/proliferation	Agoulnik, A.I., *et al.* (2014)
Fancm	FA	E12.5	Proliferation	Luo, Y., *et al.* (2014)
Ercc1	FA/NER/HR	E11.5	DNA damage/apoptosis	Hill and Crossan (2019)
Mcm9	HR/replication/MMR	E11.5	Proliferation	Luo, Y., *et al.* (2015)
Rad54	HR	≤ E11.5	ND	Messiaen, S., *et al.* (2013)
Prmt5	Arginine methyltransferase	≤ E11.5	TE activation	Kim, S., *et al.* (2014)

Table 1 - **Representation of genome stability genes involved in PGC development.** During PGC development, genes involved in DNA repair, cell division and chromatin modifications are required at different stages to ensure normal numbers of PGCs in the developing embryo. Depletion of any of these genes results in a temporally defined numerical defect with consequences for the integrity of the reproductive system. ND - not determined, TE - transposable elements

Furthermore, this suggests that during PGC development unique genomic transactions occur that render DNA repair essential for PGC development. During early development, PGCs undergo a unique G2 arrest that temporally overlaps with the phase I of DNA demethylation. It has been shown that mice deficient in *Rev7* display a PGC defect that occurs during this period of PGC development (Pirouz, Pilarski, and Kessel 2013; Watanabe et al. 2013). REV7 functions as an adaptor between the deoxycytidyl transferase REV1 and the catalytic subunit of Pol ζ, REV3 (UV-induced reversion 3), which is involved in replicating damaged DNA by TLS. Furthermore, *Rev7* plays other roles in cell cycle regulation, pathway choice in DSB repair and chromatin remodelling (Listovsky and Sale 2013; Boersma et al. 2015; Rahjouei et al. 2017; Xu et al. 2015).

Therefore, it remains to be ascertained if TLS, namely polymerase ζ and its unique function of extending beyond the lesion, is indeed required for early PGC development. If proven true this suggests a unique requirement for TLS in PGCs and a potential source of damage occurring at this stage of development that should be avoided for correct formation of this cell population. Furthermore, the requirement for TLS in PGCs may present a strategy employed by the organism to promote genetic diversity by increasing the mutation rate. In addition to *Rev7*, there is emerging evidence showing that after the G2 arrest, inhibitory phosphorylation of separase plays a critical role in the maintenance of sister chromatid cohesion and genome stability in developing PGCs. Activity of separase, a cysteine protease that cleaves sister chromatid cohesin at the onset of anaphase, is tightly regulated via two mechanisms: inhibition by securin and phosphorylation on serine 1121. Point mutant mouse embryos unable to undergo this post-translational modification of separase show a PGC numerical defect at E12.5 (Huang et al. 2008). Furthermore, mice lacking the sister chromatid cohesion factor *Pds5b* (Precocious dissociation of sisters 5b) also show PGC defects starting at least at E12.5 (Zhang et al. 2007). This indicates that during this proliferative expansion period, cells require cohesion for proper cell division to ensure the stability of the genome is maintained. Another repair pathway that has been shown to be involved in the proliferation of PGCs is the FA repair pathway. In female FA patients the reduced fertility manifests as primary ovarian insufficiency (POI) (Fouquet et al. 2017) whereas in males there is a severe decrease in spermatogenesis, often accompanied with diagnoses of non-obstructive azoospermia and Sertoli cell-only syndrome (Krausz et al. 2019). The requirement for the FA repair pathway for the correct development of PGCs and fertility has been shown in different mouse models of the disease (Kato et al. 2015; Luo et al. 2014; Tsui and Crismani 2019; Wong et al. 2003; Agoulnik et al. 2002). The numerical defect in PGCs observed in the different mouse models seems to start around E11.5 and culminates with sub-fertility in both sexes (Luo et al. 2014). This defect therefore suggests that during gonadal PGC development there is a source of DNA damage that potentially crosslinks DNA, causing the requirement for the FA repair pathway. Notably, histone demethylation may cause the release of alkylating or crosslinking agents (namely formaldehyde) in close proximity to DNA, being a potential source of DNA damage encountered by PGCs at this stage of development (Shi et al. 2004). Furthermore, during PGC development at ~E11.5 there are chromatin changes that may cause replication stress to developing PGCs and further exacerbate

the requirement for this repair pathway in PGC development (Hajkova et al. 2008). In addition, the *Mcm9* (Minichromosome maintenance complex component 9) gene has been shown to play roles in HR, MMR and DNA replication and whenever absent from mice causes a PGC defect at E9.5 (Luo and Schimenti 2015). Interestingly, depletion of both *Mcm9* and *Fancm* (FA complementation group M) exhibits an additive effect suggesting these are not acting in the same pathway. This may suggest that PGCs during proliferation require HR or alternatively that something specific about replication in PGCs is causing a numerical reduction of these cells in the absence of the *Mcm9* factor. To corroborate the first hypothesis, the *Rad54* (X-ray sensitive 54) deficient mice also show a reduction in the number of PGCs as early as E11.5 (Messiaen et al. 2013), however this decrease in PGC numbers does not cause sterility in both sexes. It remains to be determined if the role of *Mcm9* in PGCs is linked to its function in HR and also if there is a particular requirement for HR for the development of PGCs and to maintain the integrity of the germ line genome with high fidelity. Finally, it has been shown that during PGC development the arginine methyl transferase 5 (PRMT5) translocates from the cytoplasm to the nucleus between E8.5 to E10.5 playing a role in repressing histones H2A and H4R3me2. PGCs deficient in *Prmt5* exhibit sterility in both sexes preceded by increased transcription of two transposable elements - IAP (intracisternal A particle) and LINE1 (Long interspersed nuclear element 1) which were shown to cause sufficient DNA damage to clear these cells from the embryonic gonads (Kim et al. 2014). The unique cellular and genomic transactions of PGCs make the genome of these cells prone to DNA damage. Overall, we can conclude that there are several steps in the development of PGCs in which there is a genetic requirement for DNA repair. However, the nature of the DNA damage encountered by PGCs during the different phases of development and the factors which are required for these processes still remain largely unexplored. To understand which DNA repair factors are required for PGC development will permit to gain insight into how germ line mutations are suppressed.

Summary

DNA repair is required to suppress mutations. Furthermore, experimental evidence shows tissue specificity in DNA repair activity. The germ line has a lower mutation rate compared to the soma as mutations in the germ line can affect the progeny and fitness of subsequent generations. PGCs have the highest mutation frequency and the study of mutagenesis in PGCs is particularly important because mutations in these cells can affect multiple offspring. Genetic evidence shows that during PGC development, which entails extensive genomic and epigenomic transactions, there is the generation of DNA damage. Therefore, this makes DNA repair essential to ensure normal numbers of PGCs and control mutagenesis in the germ line. In order to understand the nature of DNA damage encountered by PGCs, it is critical to understand which DNA repair pathways are used during PGC development. To achieve this, it is crucial to investigate which genetic factors are required for the repair of DNA lesions generated during PGC development. In this thesis we developed a platform to screen for DNA repair factors required in PGCs using an *in vitro* system that generates PGCs from mouse embryonic stem cells. This screen permitted the discovery of novel factors that shed light into which repair pathways are required during PGC development.

Chapter 2 - *In vitro* system recapitulates the requirement for DNA repair *in vivo*

2.1 - *Ercc1*-deficient mice show a numerical reduction in the PGC population

In order to study the genetic requirement for DNA repair during PGC development our lab chose to study the gene *Ercc1*. We chose this gene firstly because *Ercc1*-deficient mice have impaired fertility (Hsia et al. 2003) and secondly because *Ercc1* is involved in several DNA repair pathways, including NER, crosslink repair and SSA. *ERCC1* was the first human DNA repair gene to be cloned (Westerveld et al. 1984) and ERCC1 is known to interact with XPF and form a structure-specific endonuclease that not only generates incisions at the site of UV-damaged DNA (Sijbers et al. 1996) but also causes incisions at ICL-generated lesions (Hodskinson et al. 2014). ERCC1–deficiency in cells causes hypersensitivity to UV irradiation and crosslinking agents such as mitomycin C (Melton et al. 1998). In mice, absence of *Ercc1* causes severe runting at birth and death before weaning due to liver failure (McWhir et al. 1993; Weeda et al. 1997). Two lines of *Ercc1*-deficient mice were independently generated and showed polyploidy in perinatal liver with elevated TRP53 levels, which were also detected in the kidney and brain. This observation led the authors to propose a model in which DNA damage generated endogenously via crosslinks would activate TRP53, which in turn induces *p21*, causing cell cycle arrest and nuclear ploidy (McWhir et al. 1993; Weeda et al. 1997). In addition to the two knockout mice described, an *Ercc1*-deficient mouse expressing an *Ercc1* transgene under the control of a liver specific promoter, which allowed the mice to survive until adulthood, was generated therefore permitting the study of other phenotypes (Selfridge et al. 2001). Interestingly these mice were shown to display sterility in both sexes. Furthermore,

another study showed that these mice display an increased number of Sertoli cell-only tubules and increased levels of apoptosis in germ cells which culminate in the absence of sperm in the epididymis (Paul et al. 2007). However the nature and cellular origin of the sterility phenotype remained to be understood. Studies in flies suggested that *Ercc1* could play a role in meiotic recombination as the authors observed an absence of normal levels of crossovers in *Ercc1*-deficient flies. However, only recently our lab has shown that absence of *Ercc1* causes sterility in mice due to a severe reduction in the PGC population and that the few remaining germ cells that are not lost can progress through meiosis (Hill, R.J. and Crossan, G.P 2019). This therefore shows that the fertility phenotype is not caused by failure to undergo meiosis. To quantify the number of PGCs we used a reporter mouse line named *Stella-GFP*. This line was generated by integration of a bacterial artificial chromosome (BAC) with all four exons of *Stella* (a marker of PGCs expressed at E7.25) fused in frame with EGFP (Payer et al. 2006). Using *Stella-GFP* reporter mice, we observed by epifluorescence microscopy a clear reduction in the number of PGCs in *Ercc1*-deficient embryonic gonads compared to Wild type controls (**Figure 1a**). Also, by staining cells from embryonic gonads of *Stella-GFP* embryos with SSEA1 (Stage-specific embryonic antigen 1), a surface marker that labels PGCs (Kudo et al. 2004), we could stringently quantify the PGC population using flow cytometry. As depicted in **Figure 1b** and **c** there is a 30-fold reduction in the number of PGCs present in Wild type versus mutant embryonic gonads, which likely causes the sterility observed in adult mice (Hsia et al. 2003). These results suggest that PGC development necessitates repair by *Ercc1* and that loss of *Ercc1* leads to the accumulation of unrepaired damage ultimately leading to PGC loss. This loss of germ cells is sufficient to explain the sterility phenotype observed in these mice. However, the nature and the source of the damage encountered by PGCs during development remain unknown.

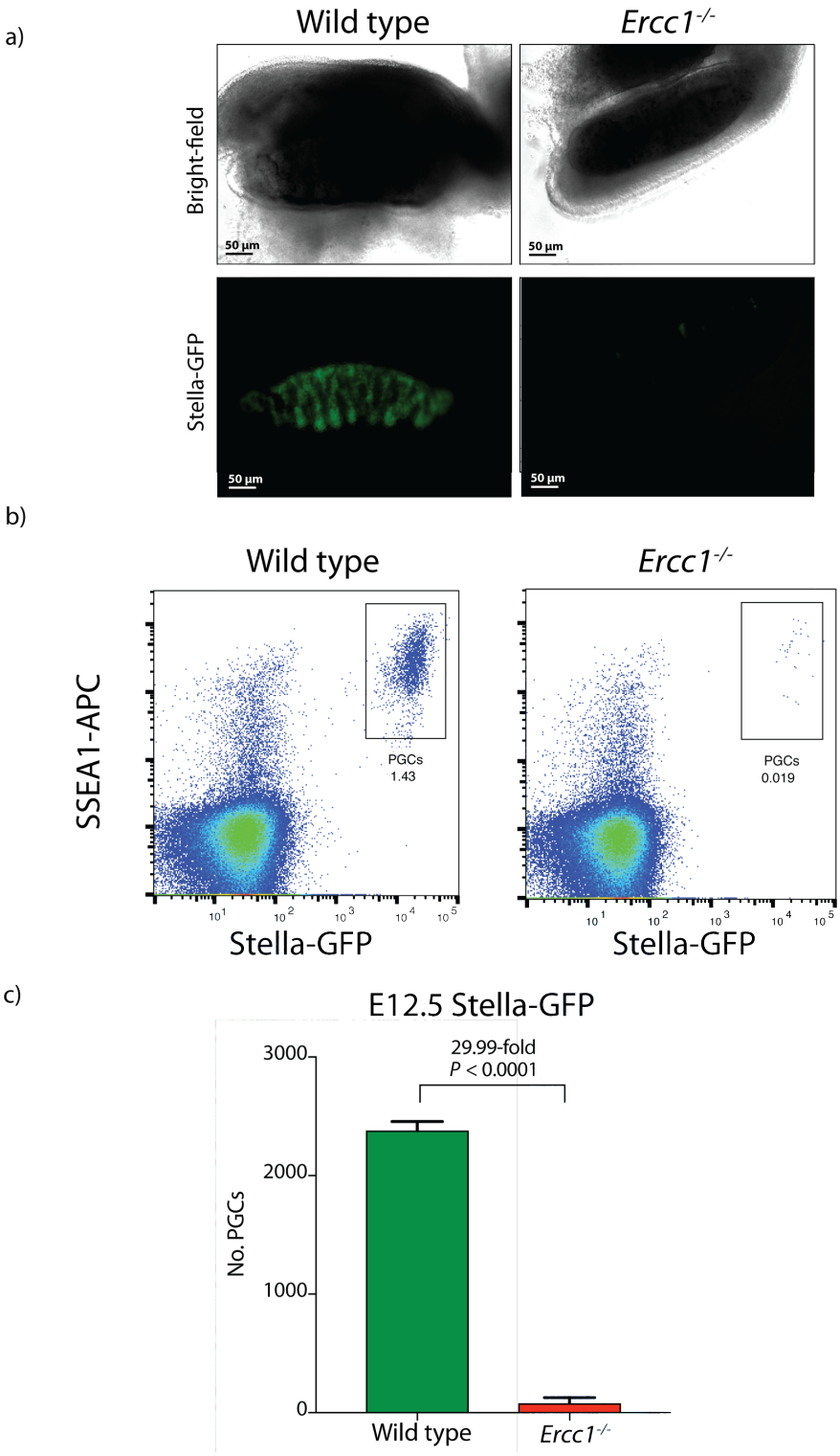


Figure 1 – *Ercc1*-deficient mice show a numerical reduction in the PGC population which culminates in sterility in both sexes.

a) Bright-field and epifluorescence microscopy images of E12.5 embryonic gonads from Wild type and *Ercc1*$^{-/-}$ embryos.

b) Representative flow cytometric profile of embryonic gonads after SSEA1 staining. The frequency of PGCs refers to the frequency of GFP and SSEA1 double positive population. The data represent the mean of three independent experiments per genotype.

c) Quantification of the number of SSEA1 and GFP positive cells (PGCs) from Wild type and *Ercc1*$^{-/-}$ embryos. The data represent the mean of three independent experiments. The statistical analysis performed was an unpaired *t*-test.

2.2 – Studying PGC development *in vitro*

In order to understand what genomic challenges PGCs encounter during development that render DNA repair essential to these cells, we set out to employ an *in vitro* system previously developed by Hayashi *et al.* (Hayashi et al. 2011). This system not only provides us with abundant experimental material which is not achievable by collecting PGCs from mouse embryos but most importantly makes the study of PGCs genetically tractable *in vitro*. This system is based on the differentiation of mouse embryonic stem cells (ESCs) into primordial germ cell-like cells (PGCLCs) by culturing the ESCs with a defined combination of cytokines. The differentiated PGCLCs are characterised not only by their ability to recapitulate the epigenetic reprogramming and cellular events observed during PGC specification but also for their spermatogenic potential (Hayashi et al. 2011; Miyoshi et al. 2016). Most notably, these cells undergo upregulation of genes involved in PGC specification such as *Blimp1*, *Prdm14*, *Ap2γ* and *Stella*. Despite the close resemblance that these cells pose to PGCs *in vivo*, there are several key differences. Firstly, whilst epigenetic reprogramming is initiated, it does not reach completion *in vitro*. Secondly, the embryonic stage recapitulated in this system is still under debate, varying from E9.5 to E12.5 according to different research groups (Hayashi et al. 2011; Miyoshi et al. 2016). And finally, since this system aims to recapitulate development *in vitro,* the physiological context and the interaction of PGCs with their niche cannot be fully recapitulated.

We used a reporter mouse ESC line named Stella-BAC9, which was derived from *Stella-GFP* reporter mice and the number 9 refers to the transgenic line used. In order to maintain ESCs in pluripotent state, referred to as a ground state, ESCs were cultured under serum and feeder free-conditions in a media containing MEK (mitogen-associated protein kinase) inhibitors and GSK3 (glycogen synthase kinase 3) inhibitors (2i), which together lead to a decrease in the levels of global DNA methylation (Sim et al. 2017), and also mLIF (mouse leukaemia inhibitory factor) which promotes embryonic stem cell self-renewal (Smith et al. 1988). ESCs cultured in 2i + mLIF media are said to be in a ground state, resembling the inner cell mass/preimplantation epiblast (E4.5) (Ying et al. 2008), maintaining therefore the capacity to give rise to all germ lineages upon differentiation.

In order to differentiate ESCs into PGCLCs two steps are required (**Figure 2a**). In the first differentiation step, ESCs cultured in 2i + mLIF are differentiated into

epiblast-like cells (EpiLCs) by culturing the ESCs in N2B27 medium supplemented with activinA, FGFb (basic fibroblast growth factor b) and KSR (Knockout serum replacement) for 42 hours in tissue culture 12 well tissue culture plates. The EpiLCs display a transcriptional profile which resembles the cells present in pre-gastrulating epiblasts (Hayashi et al. 2011). Then, in the second step of differentiation, the EpiLCs are induced to differentiate into PGCLCs in Lipidure-Coat 96 well-plates in the presence of GK15 medium with BMP4 (Bone morphogenetic protein 4), SCF (Stem cell factor), mLIF and EGF (Epidermal growth factor) for 4 days, during which time they form cellular aggregates on top of which PGCLCs are formed in small colonies. These cytokines provided in the differentiation media are the same cytokines responsible for the specification of the germ cell lineage *in vivo*. We employed Stella BAC9 ESCs, which allows us to detect the formation of PGCLCs using the *Stella-GFP* reporter (**Figure 2b**). In order to quantify PGCLCs, we combined the *Stella-GFP* reporter with an anti-SSEA1 antibody, therefore permitting us to specifically and more stringently and quantify PGCLCs by flow cytometry (**Figure 2c**).

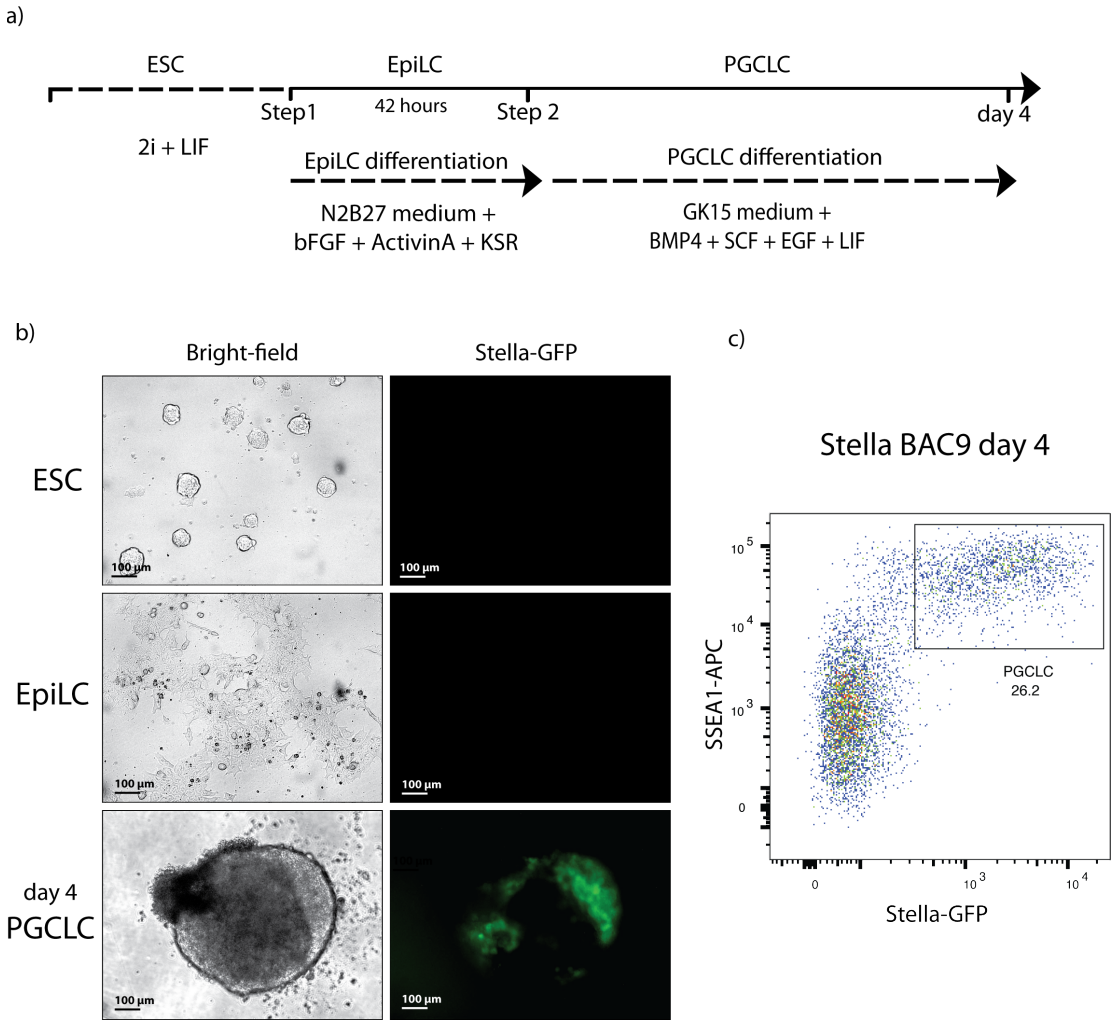


Figure 2 – PGCLC induction from ESC in culture.
a) Schema outlining the steps involved in the induction of PGCLCs from ESCs. Firstly, ESCs are cultured in 2i + mLIF to be maintained in a ground state. Then, these cells are differentiated into EpiLCs by being cultured in the presence of N2B27 medium supplemented with bFGF, ActivinA and KSR for 42 hours. After this period the cells are differentiated into PGCLCs by being cultured in the presence of GK15 media in the presence of BMP4, SCF, EGF and mLIF for 4 days.
b) Bright-field and epifluorescence microcopy images ESCs, EpiLCs after 42 hours of differentiation and day 4 PGCLCs.
c) Representative flow cytometric profile of PGCLCs induced from Stella BAC9 Wild type ESCs after SSEA1 staining. The frequency of PGCLCs refers to the frequency of GFP and SSEA1 double positive population.

To further understand the requirement for *Ercc1* in PGCs it is essential to assess which stage of PGC development our PGCLCs generated *in vitro* resemble. This would permit us to understand if the stage of differentiation obtained *in vitro* is comparable to the developmental stage observed in *Ercc1*-deficient embryos when there is a PGC numerical reduction. To achieve this goal we fluorescence-activated cell sorting (FACS)-purified E10.5 and E12.5 PGCs from mice and also ESCs, EpiLCs and PGCLCs following 2, 4 and 6 days of *in vitro* differentiation. Using these samples we determined the levels of expression of PGC specific genes using reverse-transcription quantitative PCR (RT-qPCR) (**Figure 3a**). We selected *Ap2γ* (Kurimoto et al. 2008) and *Prdm14* (Yabuta et al. 2006; Yamaji et al. 2008) as these genes are expressed early in PGC development (around E6.25), *Dazl* (Deleted in azoospermia-like) as it is starts to be expressed at E11.5 (Seligman and Page 1998) and *Mvh* (Mouse vasa homolog) since this gene is expressed later in PGC development (starting at E12.5) (Toyooka et al. 2000). In **Figure 3b** it is depicted that day 6 PGCLCs have comparable levels of expression to E12.5 PGCs for *Ap2γ* and *Prdm14* expression but this is not observed for later germ cell markers, namely *Mvh* or *Dazl*. Also, day 4 PGCLCs seem to transcriptionally resemble E10.5 PGCs for all markers tested. These results suggest that, at day 4 of differentiation, the PGCLCs we generated in our lab transcriptionally resemble E10.5 PGCs whilst at day 6 the PGCLCs resemble PGCs between E10.5 and E12.5.

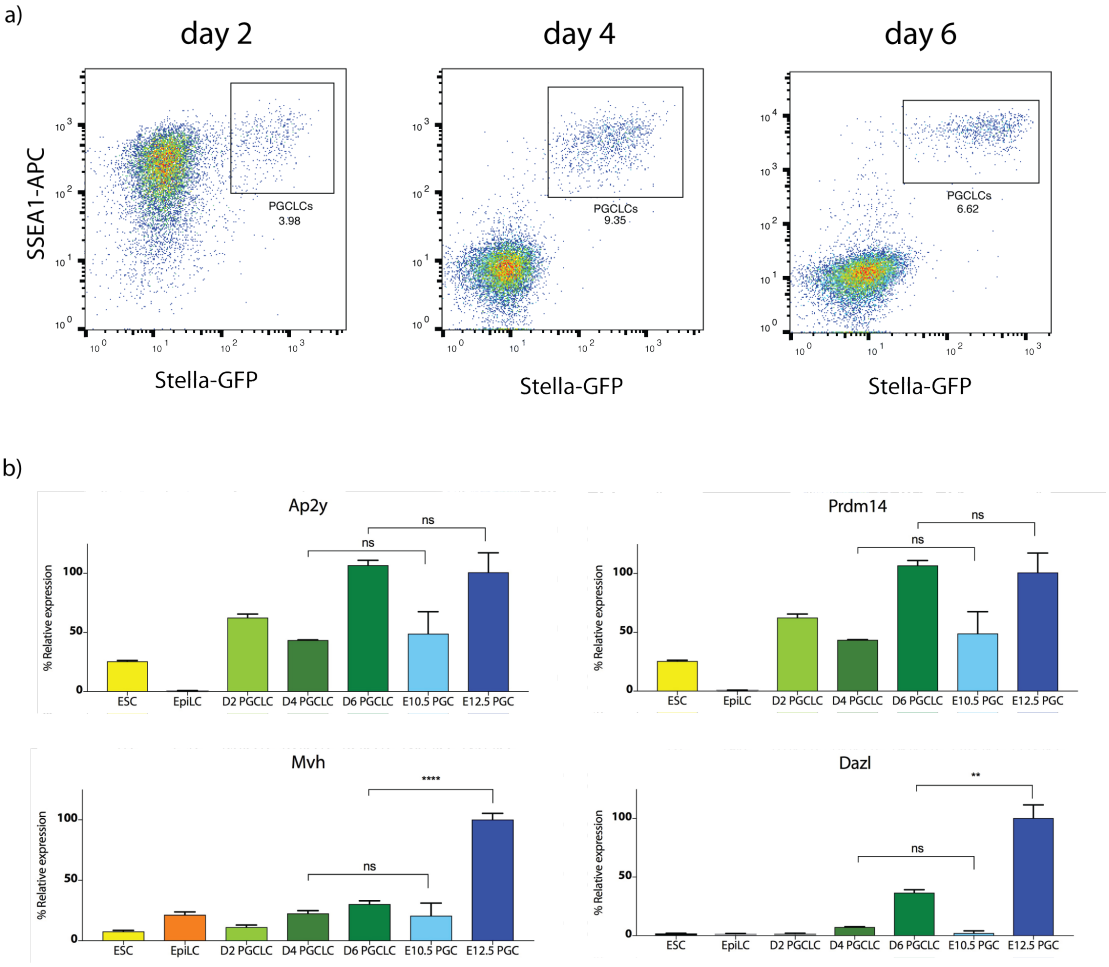


Figure 3 – Transcriptional characterisation of PGCLCs.
a) Representative flow cytometric profile of PGCLCs induced for 2, 4 and 6 days from Stella BAC9 Wild type ESCs after SSEA1 staining. The frequency of PGCLCs refers to the frequency of GFP and SSEA1 double positive population.
b) Expression analysis of PGC markers in day 2, 4 and 6 PGCLCs, E10.5 and E12.5 PGCs and ESCs and EpiLCs by qPCR. The data represent the mean of three independent experiments from one biological replicate and three technical replicates. The statistical analysis performed was an unpaired *t*-test. On all occasions the data were normalized to the expression of *Gapdh* and made relative to the expression in E12.5 PGCs.

2.3 - Generation of a DNA repair-deficient ESC line

As previously shown, *Ercc1*-deficient mice have a significant reduction in the number of PGCs at E12.5. In order to assess if we can exploit the PGCLC *in vitro* system to study DNA repair requirements during PGC development we set out to generate *Ercc1*-deficient PGCLCs. Firstly, we performed targeted disruption of the *Ercc1* gene in ESCs. In order to disrupt this gene, we designed a targeting vector containing a neomycin resistance cassette disrupting the third exon flanked by two 1.5 kb homology arms (**Figure 4a**). The third exon of *Ercc1* encodes for part of the central domain which is structurally homologous to the XPF nuclease domain (Tsodikov et al. 2005; Nishino et al. 2003). This targeting vector was co-transfected with a plasmid encoding for the Cas9 enzyme and a gene-specific single guide RNA (sgRNA) targeting the third exon of *Ercc1*. Two different sgRNAs were used to target *Ercc1*. Using this strategy, we stimulated HR by inducing a site-specific DSB and providing a template to change the outcome of repair by HR. Upon transfection, neomycin-resistant clones were selected and a long-range PCR (LR-PCR) permitted the detection of clones successfully disrupted in the *Ercc1* locus. Only when the targeting vector recombined correctly did we obtain a 2.6 kb product. (**Figure 4b**). To confirm that the correct recombination of the targeting vector produced a homozygous mutation, we performed an immunoblot analysis using a mouse monoclonal anti-ERCC1 antibody. This analysis revealed that the targeted clone did not express *Ercc1* (**Figure 4c**). Finally, since it has been reported in the literature that *Ercc1*-deficient cells are hypersensitive to the DNA crosslinking agent mitomycin C (Weeda et al. 1997), we performed a colonogenic survival assay using this agent to functionally test the cell line we generated. We found that the *Ercc1*-deficient ESCs we generated were hypersensitive to mitomycin C (**Figure 4d**). Taken together these data demonstrated that we have generated an *Ercc1*-deficient ESC line.

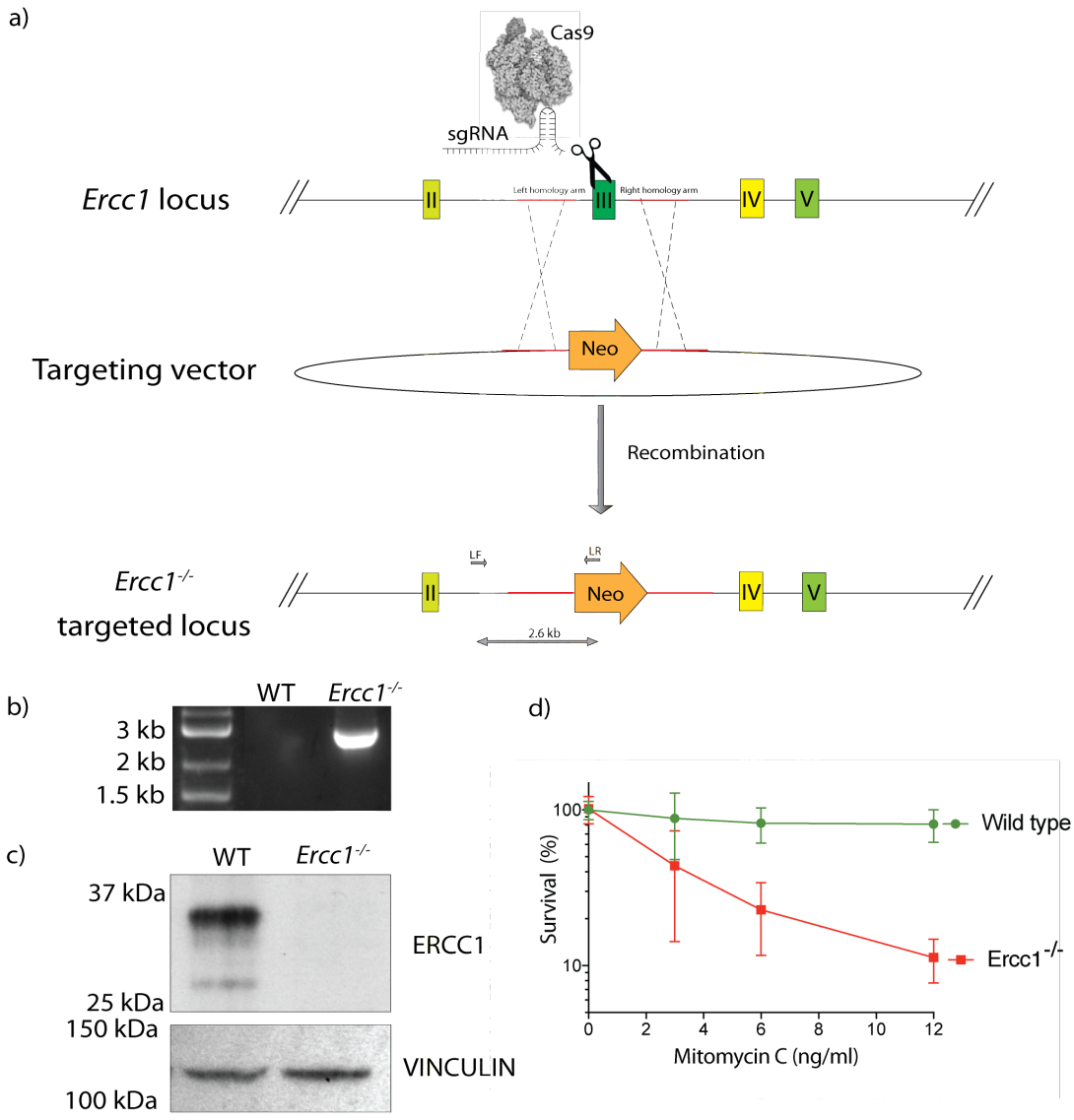


Figure 4 – Disruption of *Ercc1* in ESCs.
a) Schema representing the *Ercc1* locus, the targeting vector and the targeted locus. Coding exons are represented as boxes and the oligos used for the LR-PCR are indicated.
b) LR-PCR amplification from genomic DNA of a Wild type BAC9 Stella and *Ercc1*-deficient ESC line.
c) Western blot analysis of whole-cell extracts of Wild type BAC9 Stella and *Ercc1*$^{-/-}$ ESC lines, incubated with 1:100 dilution of anti-ERCC1 antibody and 1:2000 dilution of anti-VINCULIN antibody which served as a loading control.
d) Colonogenic survival assay with mitomycin C. ESCs were incubated with increasing doses of mitomycin C and the number of colonies observed were counted following 6 days. The data represent the mean of three independent experiments each carried out in duplicate and made relative to the untreated of each cell line.

2.4 - Differentiation of *Ercc1*$^{-/-}$ ESCs into PGCLCs

We first assessed the ability of these *Ercc1*$^{-/-}$ ESCs to form EpiLCs *in vitro*. This would permit us to ascertain whether the absence of *Ercc1* affects the differentiation potential of ESCs. In order to achieve this, we induced parental Stella BAC9 Wild type and *Ercc1*$^{-/-}$ ESCs to form EpiLCs (**Figure 5a**). After 42 hours of differentiation, we could observe by microscopy no clear difference in the morphology of *Ercc1*$^{-/-}$ EpiLCs compared to Wild type. Moreover, we set out to analyse the differentiation status of the ESCs by using c-Kit (or CD117), which is a surface marker expressed on undifferentiated stem cells (Motro et al. 1991). This permits us to accurately quantify the percentage of ESCs that have fully differentiated into EpiLCs and therefore lost c-Kit expression. The quantification of ESCs and EpiLCs by flow cytometry demonstrated that there is no significant difference between the ability of *Ercc1*-deficient ESCs to differentiate into EpiLCs compared to Wild type ESCs after 42 hours of differentiation (**Figure 5b and c**). In conclusion, *Ercc1* is not essential for the differentiation of ESCs to EpiLCs.

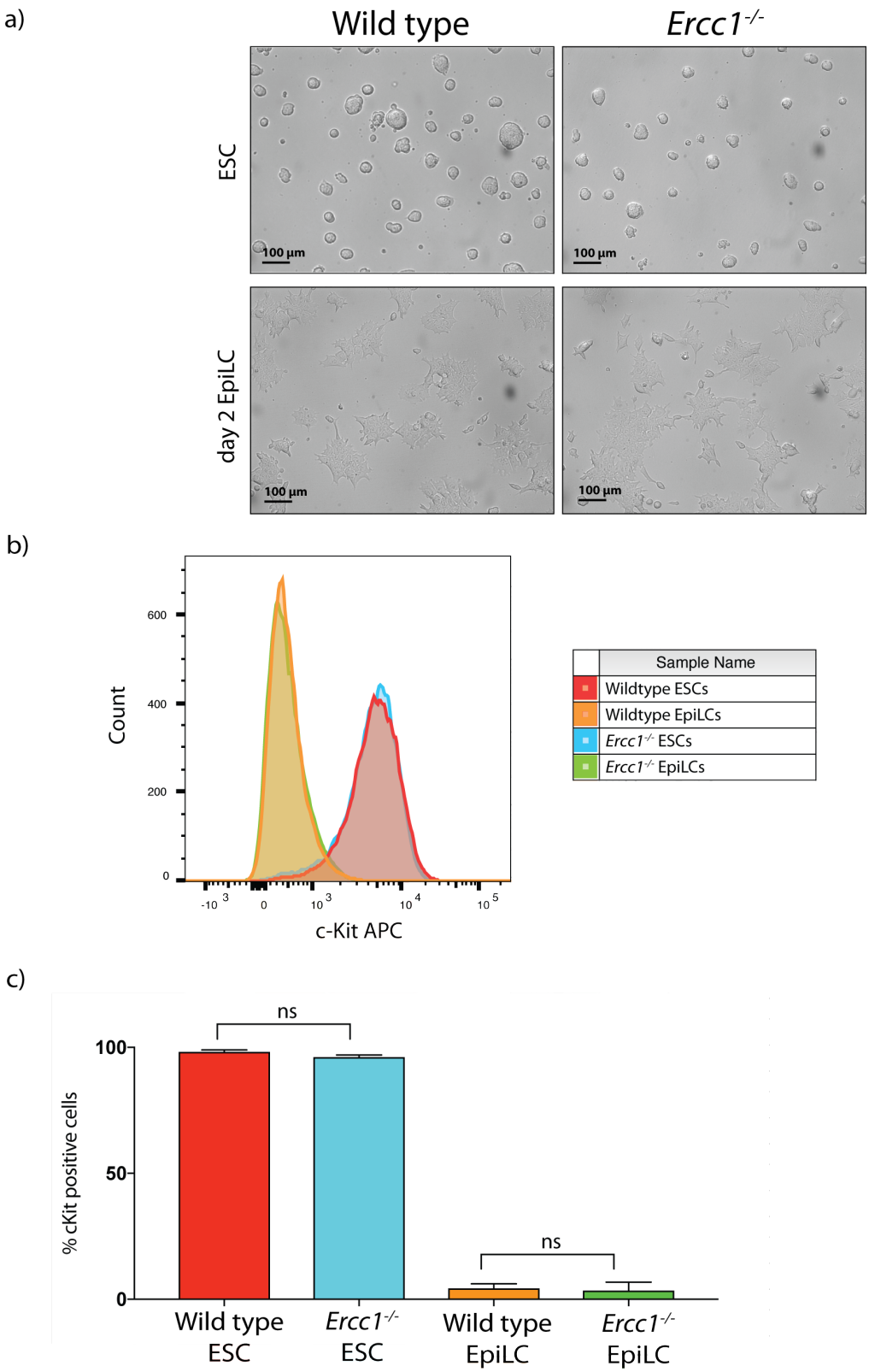


Figure 5 – *Ercc1*-deficient ESCs do not show impaired differentiation potential to EpiLCs.
a) Bright-field microscopy images of Stella BAC9 Wild type and *Ercc1*$^{-/-}$ ESCs and day 2 EpiLCs.
b) Representative flow cytometry profiles of Stella BAC9 Wild type and *Ercc1*$^{-/-}$ ESCs and day 2 EpiLCs after c-Kit (CD117) APC stain.
c) Quantification of the percentage of c-Kit (CD117) APC positive cells in Stella BAC9 Wild type and *Ercc1*$^{-/-}$ ESCs and day 2 EpiLCs. The data represent the mean of three independent experiments. The statistical analysis performed was an unpaired *t*-test.

Then, in order to determine if the requirement for *Ercc1* for PGC development observed *in vivo* can be recapitulated using the *in vitro* differentiation assay, we generated PGCLCs from both Wild type and *Ercc1*-deficient *Stella GFP* ESCs. We chose to generate *Ercc1* PGCLCs after 6 days of differentiation since day 6 PGCLCs resemble transcriptionally PGCs between E10.5 and E12.5 and the PGC defect in *Ercc1* embryos starts around E10.5, therefore matching temporally with the day 6 PGCLCs (Hill, R.J. and Crossan, G.P. 2019). After 6 days of differentiation, we could see by microscopy a clear reduction in the number of *Stella-GFP* cells in the absence of *Ercc1* (**Figure 6a**). The frequency of GFP and SSEA1 double positive cells was determined by flow cytometry (**Figure 6b**). *Ercc1*$^{-/-}$ ESCs have a 3.3 fold-reduction in the frequency of SSEA1^{+} GFP^{+} cells per total number of cells relative to the parental Wild type cells (**Figure 6c**) and a numerical defect was also observed when we quantified PGCs previously in *Ercc1*-deficient embryos at E12.5 (**Figure 6d**). Overall we can conclude that our *in vitro* system recapitulates the requirement for *Ercc1* in PGCs *in vivo.* Therefore we have generated a tool that allows us to study DNA repair in a system more close to physiology. This tool will permit us to dissect the genetic requirements for DNA repair in the mammalian germ line and facilitates investigation aimed at understanding the underlying mechanisms. Furthermore, using this tool we can aim to find novel genetic factors required for the development of PGCs.

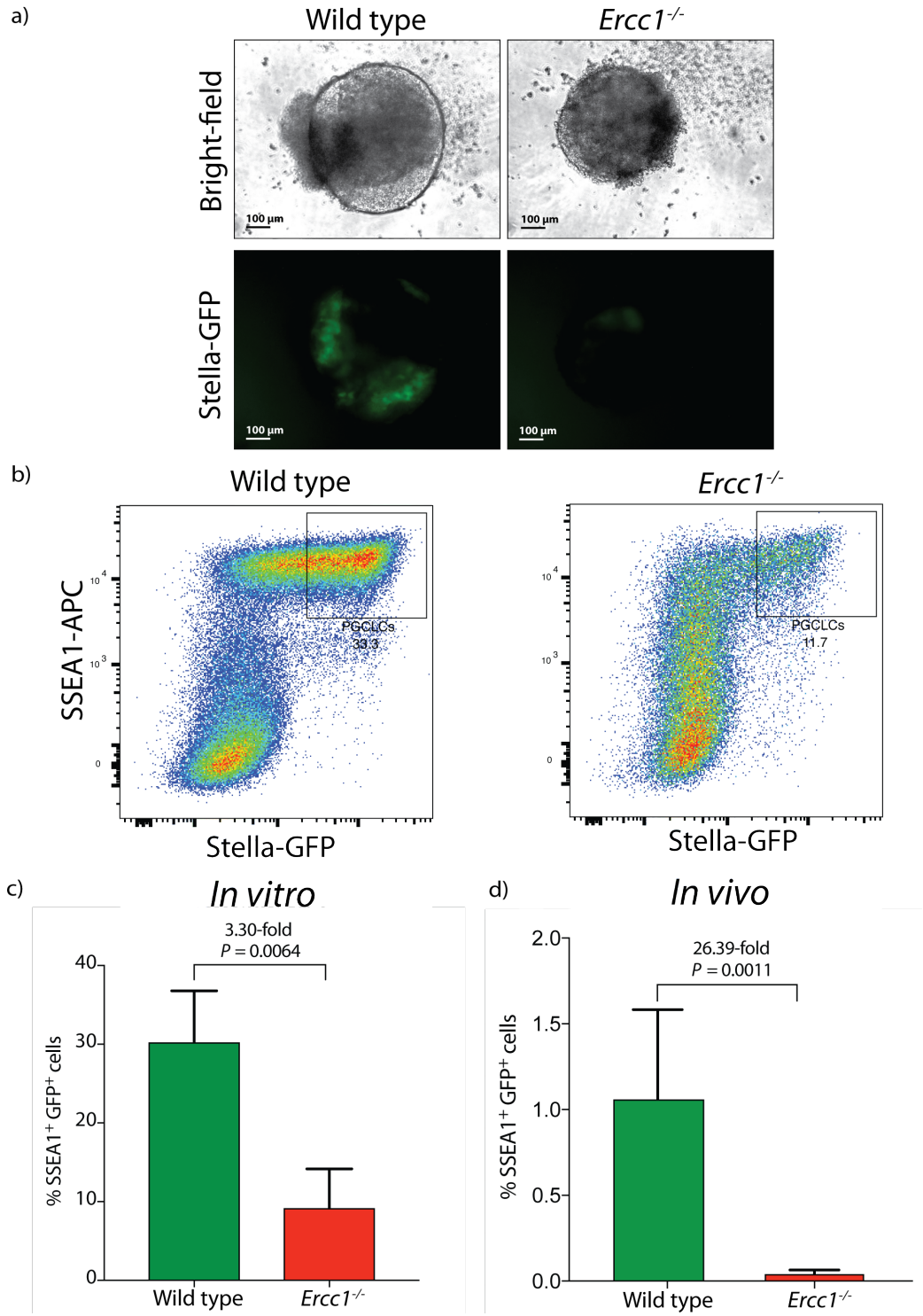


Figure 6 – Differentiation of *Ercc1*$^{-/-}$ ESCs into PGCLCs.
a) Bright-field and epifluorescence microscopy images of day 6 Stella BAC9 Wild type and *Ercc1*$^{-/-}$ PGCLCs.
b) Representative flow cytometry profiles of PGCLCs induced from Stella BAC9 Wild type and *Ercc1*$^{-/-}$ ESCs after SSEA1 stain. The percentage of PGCLC refers to the frequency of GFP and SSEA1 double positive population. The data represent the mean of three independent experiments.
c) Quantification of the frequency of SSEA1 and GFP positive cells (PGCLCs) in Stella BAC9 Wild type and *Ercc1*$^{-/-}$ PGCLCs after SSEA1 stain. The data represent the mean of three independent experiments. The statistical analysis performed was an unpaired *t*-test.
d) Quantification of the frequency of SSEA1 and GFP positive cells (PGCs) present at E12.5 in Wild type and *Ercc1*-deficient embryos. The statistical analysis performed was an unpaired *t*-test.

2.5 - Generation of a panel of CRISPR-mediated DNA repair ESC mutants

Having shown that the *in vitro* system recapitulates the requirement for *Ercc1* in PGC development, we aimed to further test the *in vitro* system so we could ultimately use it for finding novel DNA repair factors. To do this we set out to knockout other factors that similarly to *Ercc1* are required for repairing ICLs. In order to generate knockouts (KOs) in ESCs in our reporter cell line in a fast manner we employed the strategy depicted in **Figure 7**.

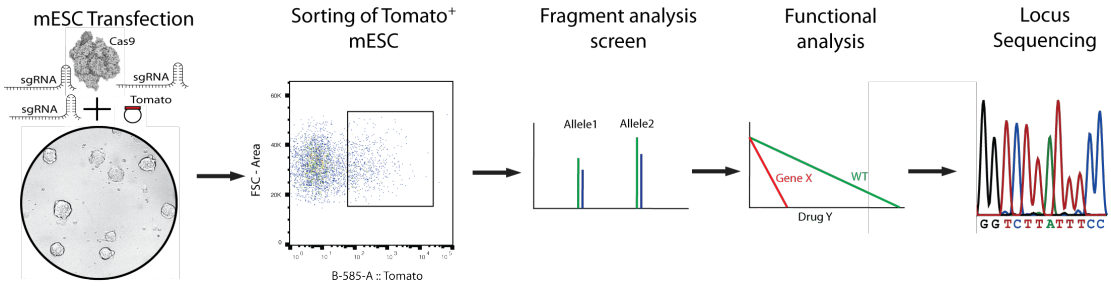


Figure 7 – Generation of a panel of CRISPR-mediated DNA repair ESC mutants.
Diagram representing the strategy employed to generate the panel of mutants. First, we designed three sgRNAs per gene, which would either specifically target an essential exon whose disruption would compromise the function of the gene or would target the 5' end of the gene therefore likely causing a frameshift that would disrupt the majority of the coding sequence. Then, we co-transfected ESCs with the px458 plasmid, which encodes for the sgRNA aforementioned, the Cas9 enzyme and GFP, with a second plasmid that leads to expression of Tomato. The use of the Tomato expressing plasmid is necessary since the mouse ESCs express basal levels of *Stella* when in a naive state, therefore not allowing us to separate the transfected cells with the sgRNAs based only on GFP fluorescence. Therefore, the Tomato positive cells were FACS sorted, which permitted us to enrich for only cells which received both the Tomato and px458 plasmids by transfection. These cells were plated at 1000 cells per dish in 10 cm dishes so that ESC colonies emerge after 7-9 days of ESC culture in 2i + LIF media from single cells in the putative KOs ESCs. These colonies were picked, expanded and then screened at the genomic level. In order to screen for KO clones at the genomic level we performed DNA fragment analysis, which permits us to identify indels with single base-pair resolution. DNA fragment analysis consists of first PCR amplifying the targeted locus with fluorescent primers. The fluorescent PCR product is then separated on a capillary electrophoresis system coupled with a fluorescent beam and detector which permits to determine the size of the PCR amplicon at a single base pair resolution. This permits us to identify clones in which indels are generated in the targeted loci and also to select clones in which the indel causes a frameshift therefore either disrupting the protein coding sequence or causing the emergence of a premature stop codon. Subsequently, for the putative clones we performed functional analysis, in which we treated Wild type and the putative clones with DNA damaging agents known to cause hypersensitivity in cells deficient in the gene of interest. To determine the nature of the genetic perturbation induced, the targeted genomic regions were amplified by PCR, cloned into a TOPO Zero Blunt vector and 10 colonies of each PCR product were used for Sanger sequencing.

We decided to target three genes that, like *Ercc1,* encode for nucleases required to confer cellular resistance to DNA interstrand crosslinking agents. With this strategy we could interrogate if the requirement for nuclease activity for repairing DNA crosslinks during PGC development. In order to do this, we chose the nuclease genes *Fan1*, *Mus81* and *Snm1b*. Furthermore, we decided to target another gene which is involved in TLS and was recently included in the FA repair pathway named *Rev7* (Bluteau et al. 2016). We chose both Fan1 and Mus81 since these genes not only encode for nucleases involved in ICL repair but also have been shown to be dispensable for fertility (Thongthip et al. 2016; McPherson et al. 2004). These factors would then be expected to generate normal frequency of PGCLCs. Then we decided to assess if other components known to act downstream of the Fanconi pathway in repairing crosslinks would be required for PGC development. This would also validate our approach since Rev7-deficient mouse embryos have been shown to have a reduction in PGC numbers (Pirouz, Pilarski, and Kessel 2013; Watanabe et al. 2013). (Pirouz, Pilarski, and Kessel 2013; Watanabe et al. 2013). Finally to assess if this system can make predictions about the role of DNA repair factors in PGCs where KO mice have not been generated we decided to target Snm1b, a nuclease involved in ICL repair that when disrupted in mice causes embryonic lethality, therefore making the study of PGCs in vivo not possible (Akhter et al 2010). This approach would allow us to observe if genes expected to have no phenotype do display a frequency of PGCLCs similar to the one observed in the parental Wild type cell line. Also, this would further test if other factors, such as Rev7 in this case, also recapitulate the reduction in PGC numbers observed in mice. Finally, we decided to interrogate if this system could make predictions of what is the genetic requirement for DNA repair factors in PGCs for genes that have not yet been studied in mice or cannot be studied in mice due to embryonic lethality.

The first gene we targeted was *Fan1*. *Fan1* encodes a protein with 5' flap endonuclease and 5'-3' exonuclease activity (Smogorzewska et al. 2010). We succeeded in deleting *Fan1* by targeting the third exon as it is located upstream of the catalytic site of this nuclease (**Figure 8a and b**). Absence of *Fan1* in cells is known to cause hypersensitivity to DNA crosslinking agents as we observed in our *Fan1* disrupted ESCs (**Figure 8c**) and in mice causes karyomegaly and liver dysfunction but nevertheless these mice are fertile (Thongthip et al. 2016).

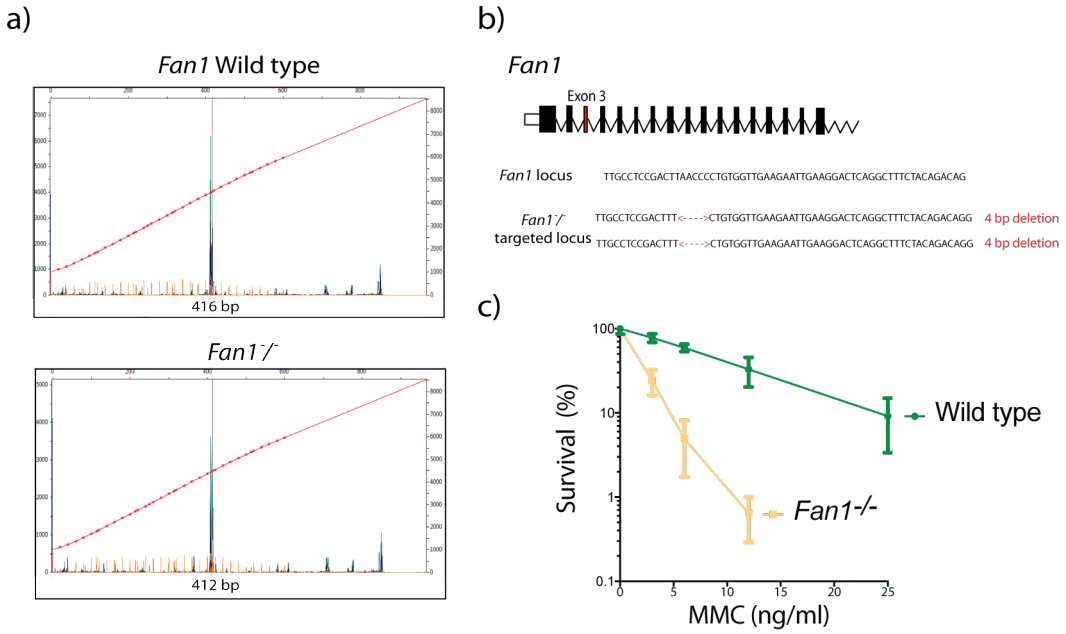


Figure 8 – Disruption of *Fan1*$^{-/-}$ in ESCs.
a) DNA fragment analysis of Wild type and *Fan1*-deficient ESCs. The y-axis shows the fluorescent signal output and the x-axis shows the size of the PCR product.
b) Schema representing the *Fan1* locus with the targeted exon indicated in red locus and the sequence of disrupted alleles after CRISPR-mediated disruption. Coding exons are represented as black boxes.
c) Colonogenic survival assay with mitomycin C. ESCs were incubated with increasing doses of mitomycin C and the number of colonies observed were counted following 6 days. The data represent the mean of three independent experiments each carried out in duplicate and made relative to the untreated of each cell line.

The next gene we targeted was *Mus81* which encodes an endonuclease that plays a critical role in the resolution of recombination intermediates, replication fork collapse, and DNA DSBs (Boddy et al. 2001). For *Mus81* we targeted the ninth exon, where the nuclease catalytic domain is encoded (**Figure 9a and b**). The absence of *Mus81* in cells causes hypersensitivity to DNA crosslinking agents, as we observed in the knockout cells we generated (**Figure 9c**) (Weeda et al. 1997). In mice *Mus81* was initially reported to be dispensable for fertility (McPherson et al. 2004) however a more detailed analysis revealed a meiotic defect (Holloway et al. 2008). Nevertheless, this phenotype does not render the mice infertile (Dendouga et al. 2005; Holloway et al. 2008).

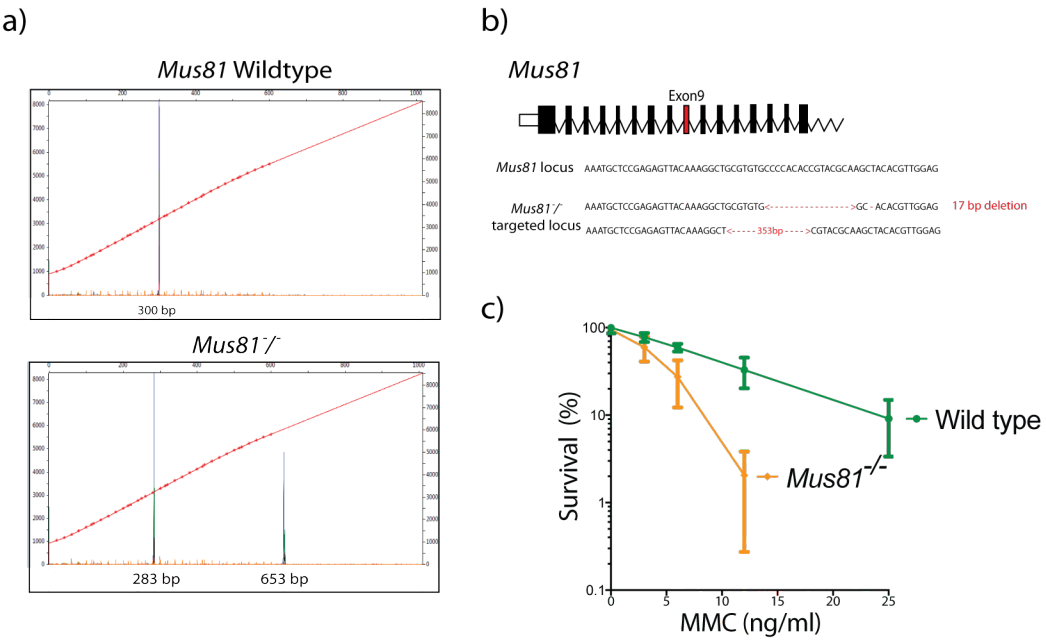


Figure 9 – Disruption of *Mus81*$^{-/-}$ in ESCs.
a) DNA fragment analysis of Wild type and *Mus81*-deficient ESCs. The y-axis shows the fluorescent signal output and the x-axis shows the size of the PCR product.
b) Schema representing the *Mus81* locus with the targeted exon indicated in red locus and the sequence of disrupted alleles after CRISPR-mediated disruption. Coding exons are represented as black boxes.
c) Colonogenic survival assay with mitomycin C. ESCs were incubated with increasing doses of mitomycin C and the number of colonies observed were counted following 6 days. The data represent the mean of three independent experiments each carried out in duplicate and made relative to the untreated of each cell line.

We also disrupted *Snm1b* which encodes for an endonuclease with 5'-3' exonuclease activity that plays a central role in telomere maintenance (Lenain et al. 2006; van Overbeek and de Lange 2006). SNM1B has been shown to physically interact with MUS81, FANCD2 and MRE11 therefore potentially maintaining or recruiting these factors to damage sites and/or stalled forks (Bae et al. 2008). In order disrupt this gene, we targeted the second exon as it disrupts most of the gene, including the catalytic domain (**Figure 10a and b**). When disrupted in cells, *Snm1b* causes a moderate increase in sensitivity to crosslinking agents as we observed (**Figure 10c**) (Demuth, Digweed, and Concannon 2004) and in mice, absence of *Snm1b* causes embryonic lethality therefore impeding the study of fertility (Akhter et al. 2010).

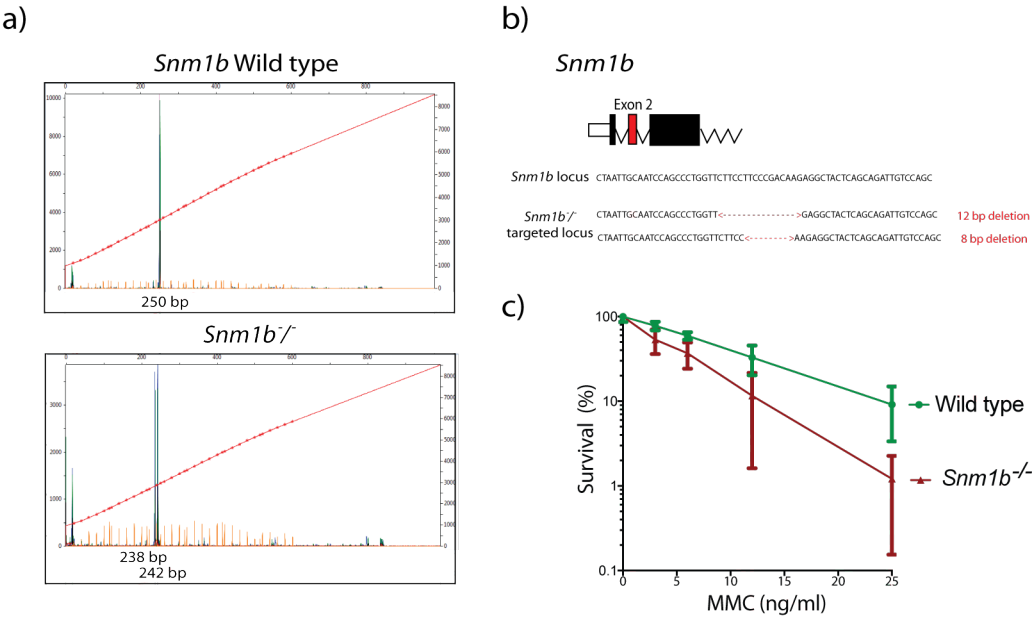


Figure 10 – Disruption of *Snm1b*$^{-/-}$ in ESCs.
a) DNA fragment analysis of Wild type and *Snm1b*-deficient ESCs. The y-axis shows the fluorescent signal output and the x-axis shows the size of the PCR product.
b) Schema representing the *Snm1b* locus with the targeted exon indicated in red locus and the sequence of disrupted alleles after CRISPR-mediated disruption. Coding exons are represented as black boxes.
c) Colonogenic survival assay with mitomycin C. ESCs were incubated with increasing doses of mitomycin C and the number of colonies observed were counted following 6 days. The data represent the mean of three independent experiments each carried out in duplicate and made relative to the untreated of each cell line.

Finally, we decided to knockout *Rev7* (also known as *Mad2l2* for Mitotic arrest deficient 2 like 2 or *Fancv)* to ask if factors acting downstream of the FA repair pathway in dealing with ICLs are required for PGC development (**Figure 11a**). In order to do this, we targeted the second exon of *Rev7* aiming to cause a frameshift that would disrupt the majority of the coding sequence since we did not know which function of this factor could be required for its function in PGCs. *Rev7* encodes for a protein that forms an essential component of the heterodimeric polymerase zeta (Polζ) together with *Rev3* (Nelson, Lawrence, and Hinkle 1996b). Polζ is a critical component in TLS as it extends beyond the site of translesion bypass to the next replication fragment. Loss of Polζ significantly reduces the ability of the cell to perform a wide range of TLS transactions rendering them sensitive to crosslinking agents as observed in the cell line we generated (**Figure 11b**) (Okada et al. 2005). REV7 is also thought to participate in pathway choice in DSB repair (Boersma et al. 2015; Xu et al. 2015) and fine tuning anaphase promoting complex/cyclosome (APC/C) activation by sequestering FZR1 (Fizzy and cell division cycle 20 related 1, also known as CDH1) (Listovsky and Sale 2013). *Rev7*$^{-/-}$ mice show growth retardation and infertility in both sexes, associated with a PGC defect after E8.5 (Pirouz, Pilarski, and Kessel 2013; Watanabe et al. 2013).

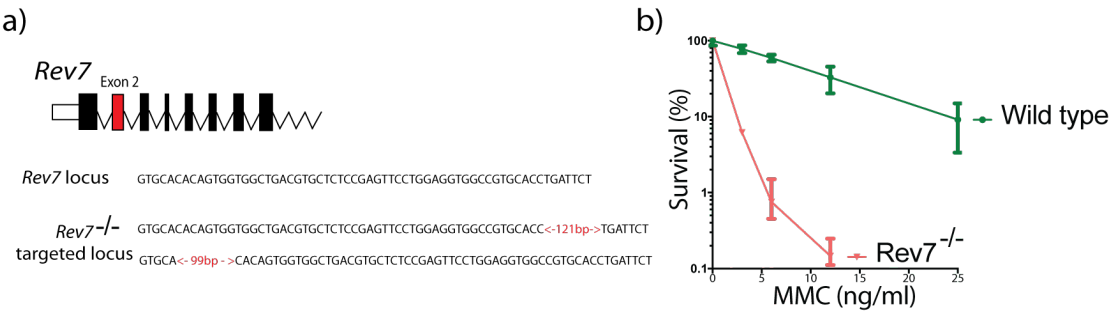


Figure 11 – Disruption of *Rev7*$^{-/-}$ in ESCs.
a) Schema representing the *Rev7* locus with the targeted exon indicated in red locus and the sequence of disrupted alleles after CRISPR-mediated disruption. Coding exons are represented as black boxes.
b) Colonogenic survival assay with mitomycin C. ESCs were incubated with increasing doses of mitomycin C and the number of colonies observed were counted following 6 days. The data represent the mean of three independent experiments each carried out in duplicate and made relative to the untreated of each cell line.

2.6 – ICL repair nuclease genes apart from *Ercc1* are dispensable for PGCLC formation but not TLS factor *Rev7*

In order to find out whether *Fan1, Mus81, Snm1b* and *Rev7* are required for PGCLC development *in vitro,* we used the previously generated cell lines and induced them to form PGCLCs. Interestingly, flow cytometry analysis did not reveal a requirement for PGCLC differentiation for any of the nucleases except for *Ercc1* (**Figure 12a and b**). This therefore suggests that *Ercc1* is the sole endonuclease required to deal with the damage encountered by PGCs during development. This result was expected in the case of *Fan1* and *Mus81* since the respective KO mice are fertile and should not display a perturbation in PGC numbers. However, it is interesting to observe that the *Snm1b* mutant does not also display a PGC numerical reduction as the KO mouse is embryonically lethal and the fertility phenotype cannot therefore be assessed.

Overall these results are important since they show that our system recapitulates the different requirements for DNA repair factors in PGCs as observed *in vivo.* Also, this permits us to conclude that not all DNA repair mutants have impaired ability to form PGCLCs *in vitro* and that the phenotypes we observed *in vitro* reproduce the phenotype of DNA repair deficient PGCs in a mouse.

In contrast to the nucleases tested, *Rev7*-deficient ESCs show a 3.4-fold reduction in the ability to differentiate into PGCLCs (**Figure 12a and b**). Since it is known that *Rev7* is required for PGC development *in vivo*, this result further validates our *in vitro* system. We could speculate that *Rev7* is required for PGC development due to its role in TLS. Furthermore, *Rev7* acts together with the FA pathway to maintain cellular resistance to crosslinking agents (Bluteau et al. 2016). However, *Rev7* also plays roles in cell cycle regulation, pathway choice in DSB repair and chromatin remodelling which could explain its requirement in PGCs. Therefore, the function of *Rev7* in promoting the correct development of PGCs remains to be ascertained. Overall we can conclude that we have established and validated the PGCLC differentiation system for the study of DNA repair in PGCs which will permit us to find novel DNA repair factors required for PGC development.

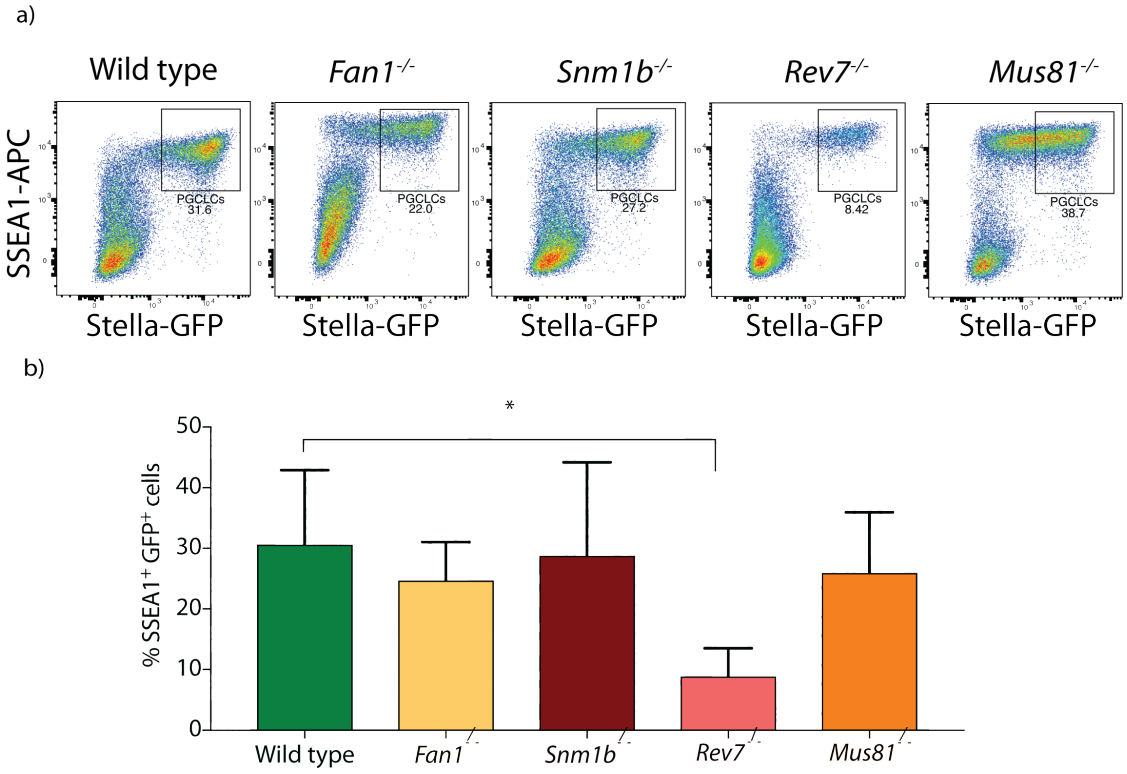


Figure 12 – Panel of DNA repair deficient ESCs induction to PGCLCs.
a) Representative flow cytometry profiles of PGCLC induced from Stella BAC9 Wild type, *Fan1*$^{-/-}$, Snm1b$^{-/-}$, *Rev7*$^{-/-}$ and *Mus81*$^{-/-}$ PGCLCs after SSEA1 stain. The percentage of PGCLC refers to the frequency of GFP and SSEA1 double positive population.
b) Quantification of the frequency of SSEA1 and GFP positive cells (PGCLCs) in the different cell lines shown. The data represent the mean of three independent experiments. The statistical analysis performed was an unpaired *t*-test.

Summary

The data presented here show that an *in vitro* system can recapitulate the requirement for DNA repair in the development of PGCs. Disruption of *Ercc1* in ESCs showed a reduction in the number of PGCLCs similar in magnitude to what is observed *in vivo.* We further tested if there was a general requirement for genes encoding nucleases responsible for conferring resistance to crosslinking agents during PGC development. To achieve this we knocked out *Fan1, Mus81* and *Snm1b* in ESC and found that the absence of these genes, contrary to *Ercc1*, is dispensable for PGCLC generation. This result recapitulates what happens in knockout mice in the case of *Fan1* and *Mus81*, in which the respective knockout mice are fertile with no known reduction in PGC numbers, but not for *Snm1b*, since the knockout mice die during embryonic development and the fertility cannot be assessed. In addition, we disrupted *Rev7* in ESCs in order to assess if a gene acting downstream of the FA pathway could be required for PGCLC formation. We found *Rev7* to be required for PGCLC development as observed in *Rev7*-deficient mouse embryos. Overall, these results allowed us to conclude that the *in vitro* system recapitulates the phenotypes observed in PGCs in mouse embryos, therefore permitting us to use of this system to perform a genetic screen and find novel DNA repair factors in PGCs.

Chapter 3 - A screen to identify novel DNA repair factors required for PGC development

3.1 - Introduction

Having established a system that recapitulates the requirement for *Ercc1* in PGCs *in vitro*, we decided to develop a platform to perform a CRISPR/Cas9 dropout screen to find novel DNA repair factors required for PGC development. The discovery of the factors required for PGC development will elucidate which repair pathways are used during this developmental process and therefore give insight into the type and nature of DNA damage encountered by PGCs. This is particularly relevant since mutations emerging during this stage of development will affect not only the individual but also the next generations. Therefore, the discovery of factors acting specifically in the germ cell lineage may elucidate how the germ line suppresses the emergence of novel mutations early during embryonic development.

CRISPR/Cas9 based screens led to discoveries in many fields of molecular biology since the first screens performed in 2014 (Shalem et al. 2014; Wang et al. 2014). Using the high efficiency of CRISPR, we can generate gene deletions in both alleles of mammalian cells thereby permitting the discovery of new genes or novel functions of previously known genes for a certain phenotypic readout. The general principle of CRISPR/Cas9 knockout screening is to generate a pooled library of sgRNAs, which is subsequently virally delivered into a Cas9 expressing cell line. Then, the cells containing a random integration of the sgRNA in the genome will be selected and expanded. Provided that the sgRNAs generate DSBs efficiently and therefore causing indels to disrupt the function of the gene, a pool of KO cells is generated. This pool of KO cells then undergoes a selective pressure, which can be either resistance to

a drug or FACS sorting based on a reporter signal. The selected pool of KO cells before and after selective pressure are then collected, genomic DNA is extracted and PCR is used to amplify the sequence of the sgRNA that integrated in the genome. These PCR products are analysed by Next-Generation Sequencing (NGS) so that the abundance of each sgRNA is determined for the samples with and without selective pressure. By comparing the sgRNAs present in the treated vs untreated samples, it is possible to identify genes that are required for the selective pressure used.

In order to perform a robust screen in PGCLCs, we had to optimise, upscale and adapt the screening conditions to correctly address our question (**Figure 1**). Firstly, the *in vitro* system relied on the use of low adherence 96-well plates to form the aggregates on top of which the PGCLCs emerge. Each of these plates yielded only a few thousand PGCLCs, therefore causing the cell number to be a potential limitation for screening while using this system. Since using a genome-wide sgRNA library would not provide sufficient coverage of the whole genome having the number of cells as a limitation, we designed a sgRNA library targeting only mouse genes involved in DNA metabolism and genome stability. To obtain the list of genes to target in the library we searched for mouse genes with gene ontology of DNA metabolism, repair, replication, cell cycle and genome stability using the AmiGO 2 tool. In order to keep our library small, we decided to use sgRNAs obtained from the Brie library. This library has successfully permitted the discovery of novel genes whilst using only 4 sgRNAs per gene as opposed to other libraries that normally use 10 sgRNAs to target one gene (Doench et al. 2016). This therefore permits us to screen for DNA repair genes still maintaining a reasonable coverage of the sgRNAs used. Secondly, we aimed at expressing Cas9 only in the ESC population. This would prevent Cas9 to generate DNA breaks during the differentiation protocol to PGCLCs in our DNA repair deficient cells. The generation of DSBs in DNA repair deficient cells could generate false-positives as a consequence of these cells being unable to survive the damage caused by the Cas9 activity rather than being unable to differentiate into PGCLCs. Therefore, we generated an inducible Cas9 Stella BAC9 cell line. In addition to this, we chose to keep a 2000-fold representation of each sgRNA throughout the screen. On the one hand, at low representation (*e.g.* 50-fold representation) the library becomes more spread out, leading to erroneous interpretation of dropout or enrichment but on the other hand, having a library representation beyond 1000 or 2000-fold does not improve the interpretation of the results. Therefore, aiming to keep the library at 2000-fold would minimise the costs of potential cell loss in any

of the steps of the screen. Finally, in order to be able to have a 2000-fold representation of each sgRNAs per cell population we had to sort 6.6 million PGCLCs. This number of cells is achieved by multiplying the number of sgRNAs present in the library (3.300) by the fold-representation aimed at keeping throughout the screen (2.000).

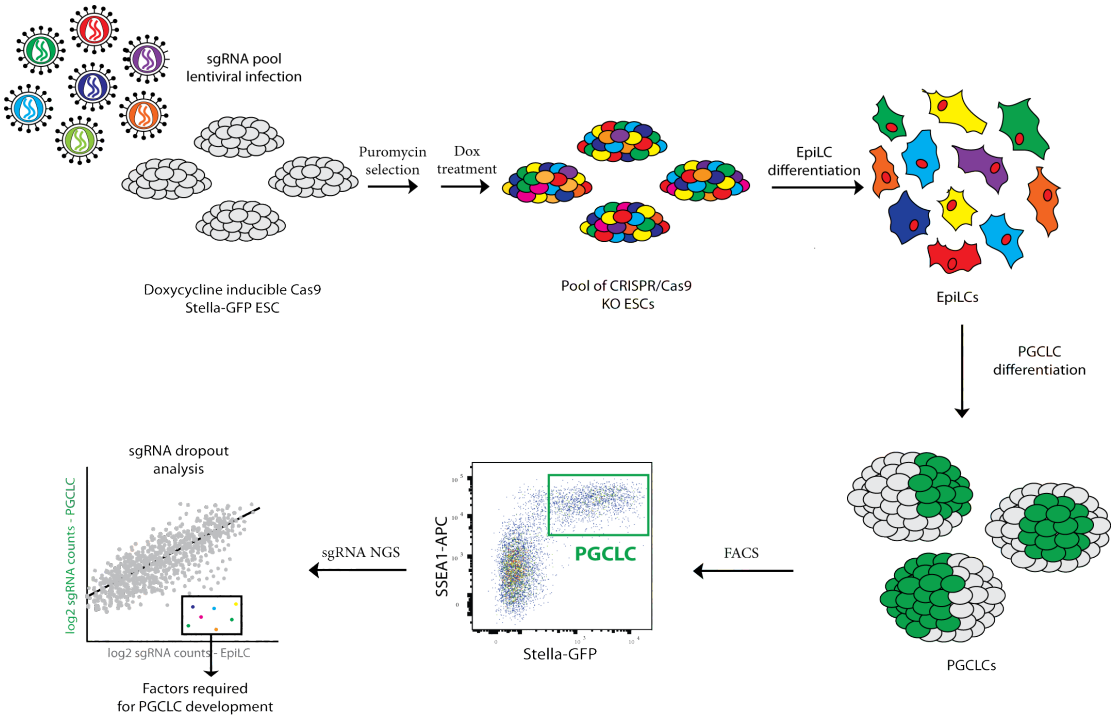


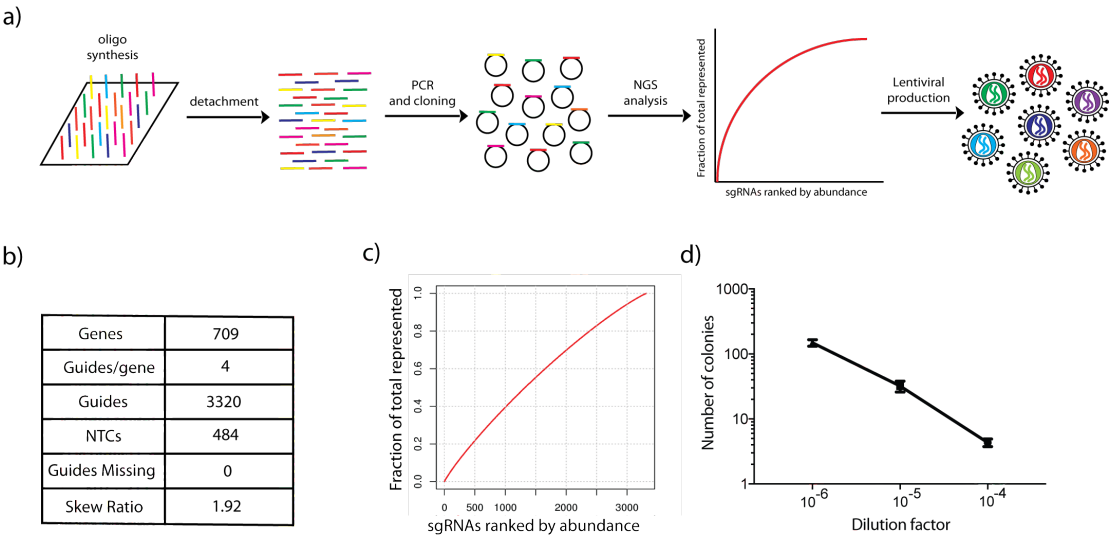
Figure 1 – A CRISPR/Cas9 based screen to discover novel DNA repair factors required for PGC development.

a) Schema representing the approach employed to screen for novel DNA repair factors required for PGC development. Firstly, we will prepare a library of sgRNAs targeting genes involved in genome stability/replication, from which we will produce lentiviruses. Then, we will generate Stella BAC9 ESCs expressing a doxycycline inducible Cas9. These cells will then be transduced with lentiviruses containing the sgRNA library with a low multiplicity of infection (MOI). Then, we will culture these cells in the presence of doxycycline so that Cas9 expression is induced. After the pool of KOs is generated, the cells will be differentiated into EpiLCs. The representation of sgRNAs present in the EpiLC population will be determined by sequencing the pool of EpiLCs. These cells will then be induced to differentiate into PGCLCs. The population of GFP^+ $SSEA1^+$ cells (PGCLCs) will be sorted and subjected to NGS to determine the frequency of each sgRNAs in this population. Finally, by analysing the sgRNAs present in the EpiLCs (which represent all the genes that do not impair EpiLC formation) and comparing them to the sgRNAs depleted in PGCLCs, we will be able to obtain the genes required for PGC development.

3.2 - Generation of a sgRNA library targeting mouse genome stability genes

In order to perform the CRISPR screen, we cloned a library of sgRNAs. Since we could only generate a limited number of PGCLCs, we decided to clone a small library instead of performing a genome-wide screen so that the coverage of the sgRNAs per PGCLC is sufficient to perform robust statistical analysis. For this we selected the genes we aimed to target based on gene ontology. We prepared a list of 709 genes involved in cell cycle, DNA repair and DNA metabolism from which we designed four sgRNAs per gene using the sgRNAs in the Brie library (Doench et al. 2016). Also, 484 non-targeting controls (NTCs) were included in the pool to use as a negative control. The oligos corresponding to the sgRNAs were designed and cloned into the pLentiguide Puro backbone (**Figure 2a**) (Sanjana, Shalem, and Zhang 2014). In order to ensure that all sgRNAs were present in the pool in a similar distribution, we performed NGS on a MiSeq sequencer (**Figure 2b)**. We obtained around 486000 sequencing reads, which permitted us to observe that all sgRNAs were present in the library. Also, this analysis permitted us to test the distribution of the different sgRNAs in the pool. By calculating the skew ratio of the sgRNAs in the library, which is obtained by dividing the sgRNAs present in the percentile 90 by the sgRNAs present in the percentile 10, we observed that in our library this value is below 10 and close to 0 as expected in a evenly distributed library (**Figure 2b**). This is clearer in **Figure 2c** in which plotting the sgRNAs ranked by abundance vs the fraction of total sgRNAs represented displays a linear regression showing that there is equal distribution of the different sgRNAs in the library pool. Having cloned the library, we generated lentiviruses which we titrated by counting the number of puromycin resistant colonies, while using the same cell type we would use in the screen, in a 10-fold serial dilution (**Figure 2d**). This permitted us to obtain the number of transducing units per ml (5×10^5 transducing units (TU)/ml) in the library we generated. In order to obtain a single sgRNA per cell, the pool of lentiviruses was transduced at an MOI of 0.3. An MOI of 0.3 is based on the Poisson distribution which dictates that at MOI of 0.3, 70% of the cells remain untransduced (uninfected). This means that at an MOI 0.3 or less, greater than 95% of infected cells are predicted to have a single integration. In order to obtain a fold-representation of 2000 (which means each sgRNA of the library is present in an average of 2000 cells at

all stages of the screen) and transducing the cells at an MOI of 0.3, we would need to transduce at least 30 x 10^6 ESCs. This means that the number of particles we would need to use to transduce is 9 x 10^6 (0.3 corresponding to the value of MOI aimed for multiplied by the total number of cells, 30 x 10^6). Finally, to determine the volume of lentiviruses to transduce 30 x 10^6 cells we divided the number of particles to use (9 x 10^6) by the titre in TU/ml (5.0 x 10^5) obtained before. Therefore, in order to transduce 30 x 10^6 ESCs at an MOI of 0.3 we would need 17.8 ml of lentiviruses.



Genes	709
Guides/gene	4
Guides	3320
NTCs	484
Guides Missing	0
Skew Ratio	1.92

Figure 2 – Generation of a pooled library of sgRNAs targeting mouse genes involved in maintaining genome stability.
a) Schema representing the generation of a pooled library of sgRNAs to target mouse genes involved in maintaining genome stability. Firstly, a library of oligos containing the sequence of the sgRNAs was synthesised. Subsequently, these pooled oligos were detached and cloned by PCR and Gibson assembly into the pLentiguide Puro vector. After cloning, the pool of plasmids was run on the MiSeq to assess the representation of each sgRNAs in the pool. Finally, these pooled plasmids were assembled into lentiviral particles in order to be delivered to the cells used for the screen.
b) Table depicting the number of genes, guides/gene, number of guides, NTCs, guides missing and the skew ration of the sgRNA library to be used in the screen.
c) Graph depicting the distribution of sgRNAs in the pooled library after MiSeq analysis.
d) Graph depicting the number of colonies of mouse ESCs that emerged after transduction with the lentiviral particles corresponding to the sgRNA pool in serial dilutions and after treatment with puromycin in order to determine the titre of viruses to then use in the screen.

3.3 - Generation of a Cas9 inducible cell line

Having generated the lentiviruses containing the pool of sgRNAs, the next step would be to generate an inducible Cas9 cell line. We achieved this by employing a Tetracycline-controlled transcriptional activation system using doxycycline in a Tet-ON 3G system. This system would allow Cas9 to be expressed only before the differentiation protocol, during which doxycycline would not be used and Cas9 would not be expressed. To generate this cell line, we swapped the puromycin resistance cassette present in the pCWCas9 Puro plasmid with a hygromycin resistance cassette from the pCW57-MCS1-P2A-MCS2 (Hygro) plasmid by restriction cloning. This plasmid encodes for Cas9 under the expression of a Tet Response Element immediately followed by a minimal cytomegalovirus promoter and also the reverse tetracycline-controlled transactivator required for the Tet-ON system. Additionally, having a hygromycin resistance cassette instead of puromycin permits us to select both the sgRNAs and the cell line with the integration of the inducible Cas9 construct. After transfecting cells and selecting for hygromycin, we picked and screened clones based on Cas9 expression after 24 hours in the presence or absence of 1 μg/ml doxycycline induction. We chose the clone with both the highest Cas9 expression after induction and no expression in the absence of doxycycline, which was the clone number 8 and therefore we named the cell line iCas9 8. In order to show this cell line was expressing Cas9 in a doxycycline inducible and dose dependent manner, we treated the cell line with increasing doses of doxycycline for 24 hours and assessed the level of Cas9 expression by immunoblot. **Figure 3a** clearly shows that increasing the concentration of doxycycline caused an incremental increase in the expression of Cas9 up to 10 μg/ml of doxycycline after which we observed toxicity. Having generated a Cas9 inducible cell line, we asked if this cell line could still differentiate into PGCLCs both before and after being treated with doxycycline. For this we generated PGCLCs with the Wild type and the iCas9 8 cell line both in the presence and absence of 1 μg/ml of doxycycline induction for 48 hours. This showed that the cell line we generated is suitable to use for screening.

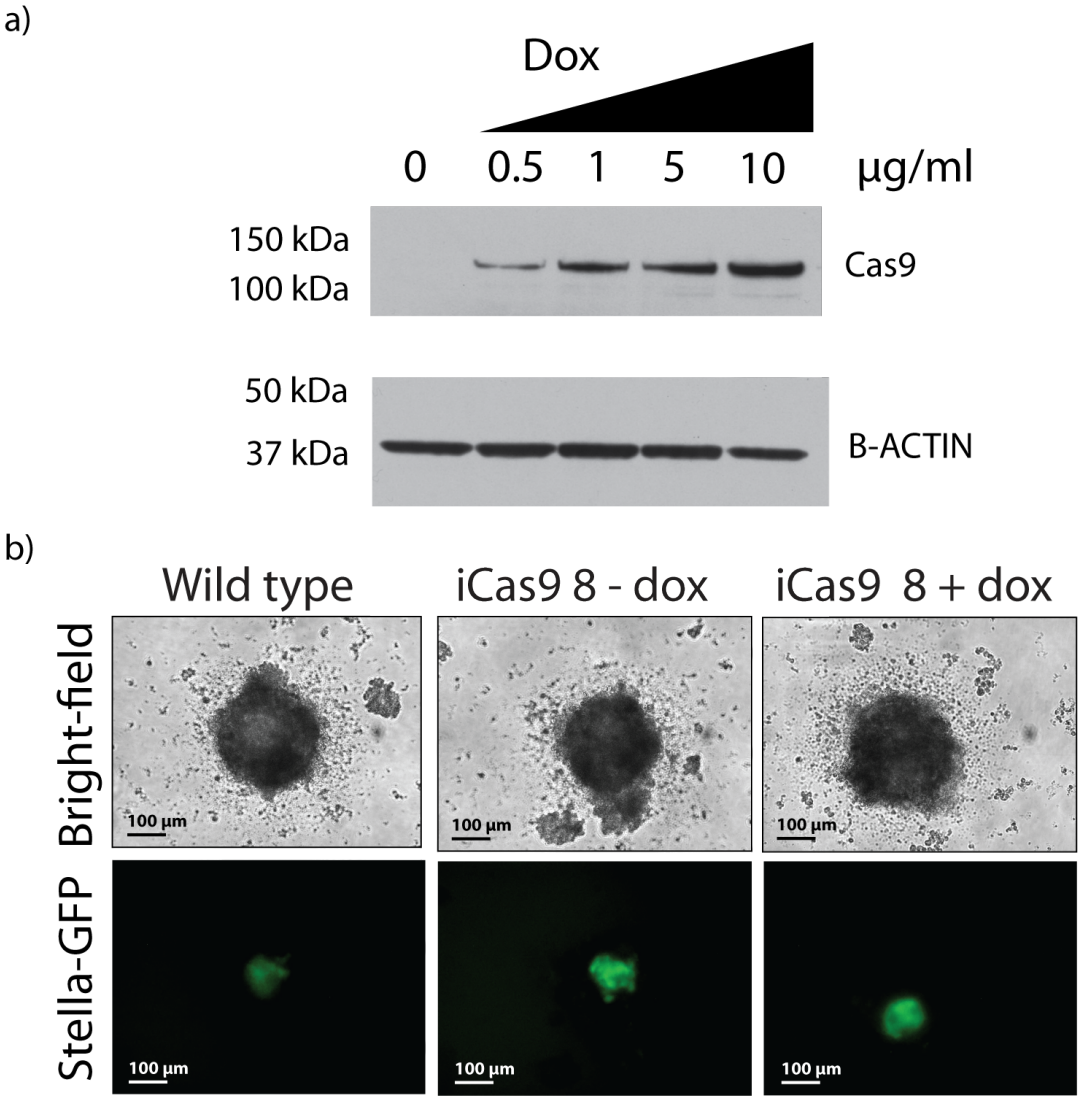


Figure 3 – Generation of a doxycycline inducible Cas9 *Stella-GFP* ESC line.
a) Western blot analysis of whole-cell extracts of Stella BAC9 Wild Type cells containing a Cas9 inducible (iCas9) construct treated with increasing doses of doxycycline for 24 hours. The lysates were incubated with anti-Cas9 antibody and anti-B-ACTIN antibody, which served as a loading control.
b) Bright-field and epifluorescence microcopy images of day 4 PGCLC induced from Stella BAC9 Wild type and iCas9 clone 8 before and after 24 hour treatment with 1 µg/ml of doxycycline.

3.4 - Validation of the cell line to use in the CRISPR/Cas9 Screen

Then we aimed at finding if the Cas9 inducible cell line we generated would produce gene deletions efficiently in the presence of doxycycline. If this were to be the case, this experiment would also permit us to determine the optimal concentration and duration of the doxycycline treatment. To achieve this, we transduced cells with a sgRNA targeting a surface marker present on mouse ESCs and treated these cells with increasing doses of doxycycline. We decided to target *c-Kit* (CD117) since it is an ESC surface marker that permits us to easily follow protein expression using flow cytometry. We could then determine the proportion of ESCs in culture that contribute to normal levels of *c-Kit* expression (CD117 positive cells) and compare it to those in which the expression was lost due to disruption of the *c-Kit* locus (CD117 negative cells) (**Figure 4a**). In order to ensure that the disruption of *c-Kit* was caused directly by the sgRNA targeting the *c-Kit* locus and to exclude potential effects of the doxycycline, Cas9, sgRNA expression or other variable to cause non-specific reduction of *c-Kit* expression, we decided to transduce cells using a different sgRNA targeting the Rosa26 locus. Using this strategy we observed that targeting the Rosa26 locus did not have any effect on the CD117 expression up until 5 µg/ml of doxycycline excluding the possibility that the decrease in CD117 was due to differentiation caused by the doxycycline treatment (**Figure 4b**). As observed in **Figure 4c**, treating the cells for 18 days caused up to 80% depletion of cells expressing CD117. This allowed us to conclude that the cell line we have generated permits the generation of CRISPR deletions in a highly efficient manner. These results also permitted us to select the optimal concentration and duration of doxycyclin treatment. We had observed loss of CD117 expression with 5 and 7.5 µg/ml of doxycycline whilst targeting the Rosa26 locus, which was likely due to differentiation of the ESCs. Since we require the ESCs to be in a naive state so that we can differentiate them into PGCLCs with high efficiency, we decided to use 2.5 µg/ml of doxycycline throughout the screen. Furthermore, we decided to induce Cas9 expression for 21 days to maximise the chances of efficient gene disruption for different sgRNAs targeting different genes.

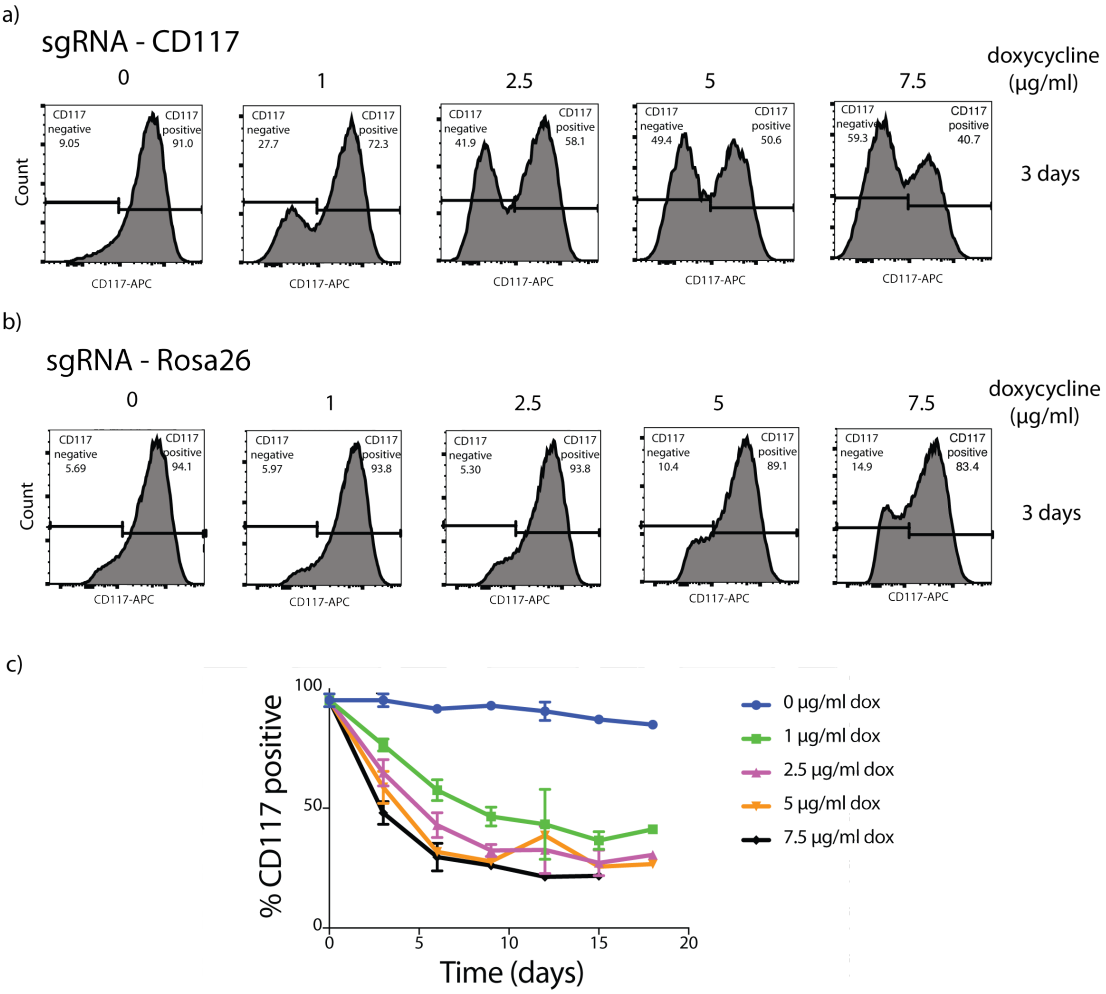


Figure 4 – Efficiency of gene deletion in a doxycycline dose and time dependent manner using the iCas9 8 cell line.

a) Representative flow cytometric profile of iCas9 8 ESCs transduced with a sgRNA targeting the CD117 gene at a MOI inferior to 0.3 treated with increasing doses of doxycycline for 3 days after CD117 staining.

b) Representative flow cytometric profile of iCas9 8 ESCs transduced with a sgRNA targeting the Rosa26 locus at a MOI inferior to 0.3 treated with increasing doses of doxycycline for 3 days after CD117 staining.

c) Quantification of the percentage of CD117 positive iCas9 8 ESCs transduced with a sgRNA targeting the CD117 gene at a MOI inferior to 0.3 treated with increasing doses of doxycycline for 18 days after CD117 staining. The data represent the mean of three independent experiments.

3.5 - Identification of DNA repair factors required for the proliferation of mouse embryonic stem cells

Having cloned the sgRNA library and generated and validated the Cas9 inducible cell line, we decided to perform a CRISPR screen using only the ESCs we developed and assaying the effect of proliferation in the pool of CRISPR KOs. This screen would permit us to test if the screening platform is sufficiently efficient at droping out sgRNAs targeting essenital genes and also if it permits us to identify known and novel factors required for ESC proliferation. In order to do this, we performed a CRISPR screen collecting 6.6 million ESCs after 21 days of doxycycline induction, thereby covering the sgRNA library at a representation of 2000-fold. Then we compared the sgRNAs in the sgRNA pool plasmid DNA before transduction into ESCs with the sgRNAs present in the ESC population after doxycycline treatment. sgRNAs which are enriched in the ESC population should target genes which when disrupted in ESCs cause a growth advantage; sgRNAs which are depleted in the ESC population should correspond to sgRNAs targeting essential genes or genes that when disrupted in ESCs cause a proliferative disadvantage; sgRNA without any effect in ESC proliferation should target genes which when disrupted from ESCs do not affect cell growth. In order to achieve this, we first ranked the sgRNAs of the libraries according to the number of counts in the ESC population when compared to the plasmid pool of sgRNAs. As expected we observed that at least the 11 most underrepresented sgRNAs in this analysis correspondeded to sgRNAs targeting essential genes (**Figure 5a**). Then we chose four genes that we knew were essential and also were present in our library, namely *Mcm2* (Minichromosome maintenance complex component 2), *Pcna* (Proliferating cell nuclear antigen), *Rpa2* (Replication protein a2) and *Tuba1c* (Tubulin alpha 1c)) and observed that the 4 sgRNAs targeting each of these genes were depleted from the ESC population in comparison to the plasmid pool of sgRNAs (**Figure 5b**). Furthermore, sgRNAs targeting the additional subunits of the RPA complex (*Rpa1* and *Rpa3*) and the genes responsible for encoding the *Mcm2-7* complex were also depleted. Also we observed that the NTC sgRNAs were generally distributed equally in the ESC population with equivalent numbers of reads as in sgRNAs targeting genes which are known to cause no alteration in cell growth (for example *Xpa*). This distribution was also observed in all of the other cell populations.

a)

Rank	sgRNA counts	Gene	Function
1	97	*Rrm2*	DNA metabolism
2	102	*Cct5*	Chaperone
3	123	*Rpa1*	DNA replication/repair
4	127	*Parg*	DNA repair
5	157	*Hinfp*	Transcription
6	160	*Gtf2h3*	Transcription
7	175	*Rpl27*	Translation
8	177	*Cdc25a*	Cell cycle
9	183	*Pcna*	DNA replication
10	198	*Rps3*	Translation
11	209	*Cdc16*	Cell cycle
Median	8492	*Xpa*	DNA repair

b)

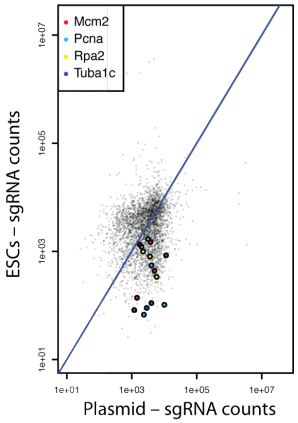


Figure 5 – Disruption of essential genes in the ESC population causes reduction of sgRNA counts.
a) Table representing the rank of the sgRNAs with the fewest counts in the ESC population with the respective gene and gene function.
b) Scatter plot representing the count of the four sgRNAs targeting essential genes (*Mcm2*, *Pcna*, *Rpa2* and *Tuba1c*) in the ESC population compared to the plasmid pool of sgRNAs.

Second, by analysing the distribution of sgRNAs in the ESC population whilst comparing to the plasmid pool of sgRNAs, we observed the four sgRNAs targeting *Trp53* to the most strongly enriched (**Figure 6a**). sgRNAs targeting *Trp53* have been previously reported to be enriched in the ESC population whenever depleted in other screens (Hackett et al. 2018; Li, Yu, et al. 2018) which shows that absence of this gene provides a proliferative advantage to ESCs. *Trp53* has been shown to be required for the suppression of proliferation (Sabapathy et al. 1997) and differentiation of ESCs by repressing the pluripotency-associated transcription factor *Nanog* to induce differentiation and to be a barrier for reprogramming (Lin et al. 2005). The second most enriched set of sgRNAs in the ESC population corresponded to *Setd2*. *Setd2* is a histone methyltransferase that is specific for lysine-36 of histone H3, and methylation of this residue is associated with active chromatin (**Figure 6b**). However no phenotype has been studied regarding its proliferation in ESCs. Furthermore we found *Chek2* (Checkpoint kinase 2) and *Usp28* (Ubiquitin specific peptidase 28) to confer a proliferative advantage to ESCs. *Chek2* is a DNA damage response cell cycle checkpoint regulator and a putative tumour suppressor that stabilises *Trp53* (**Figure 6c**). *Usp28* is a deubiquitinase involved in the DNA damage response checkpoint, which recently has been shown to stimulate *Trp53* DNA-binding activity (Cuella-Martin et al. 2016). The discovery of sgRNAs targeting *Trp53* and genes involved in the same DNA damage response axis and the fact that sgRNAs which are depleted in the ESC population correspond to sgRNAs targeting essential genes gives us

confidence that the targeting of the initial ESC population using the Cas9 inducible system occurred in an efficient manner.

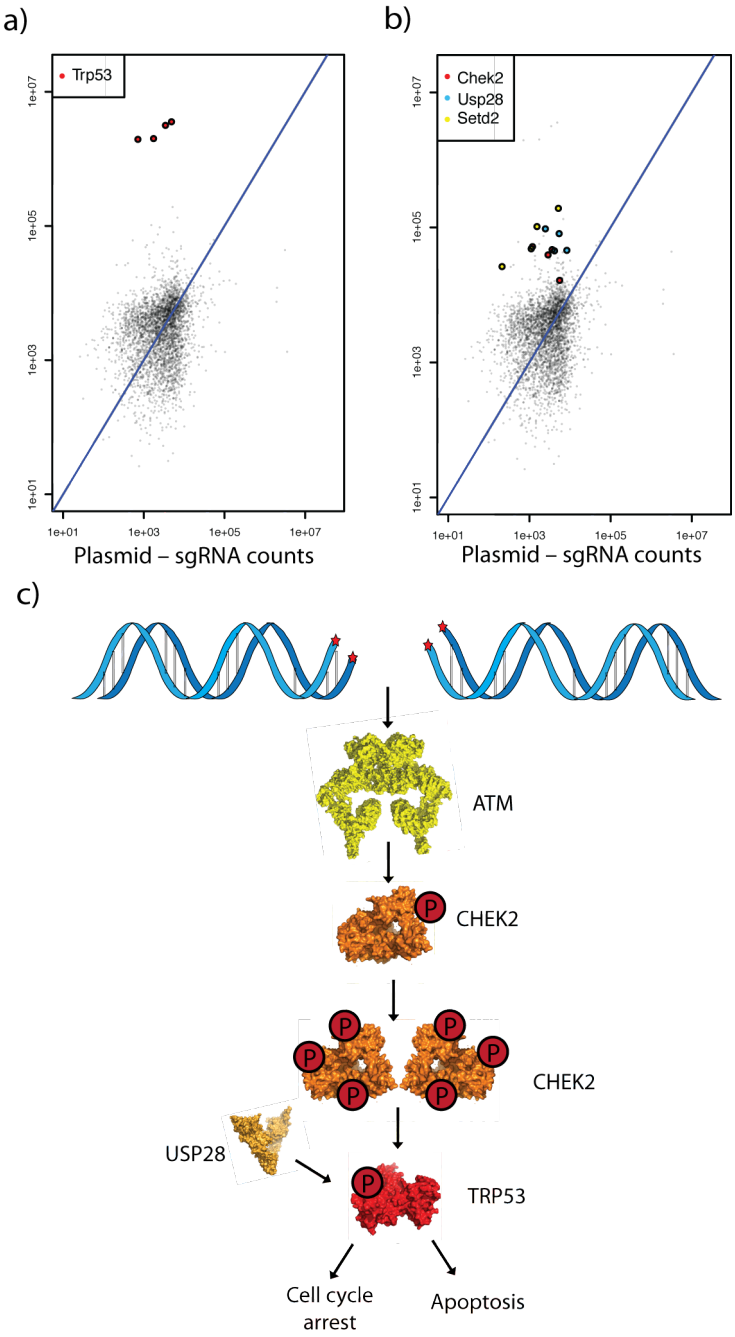


Figure 6 – A CRISPR Screen identifies that disruption of *Trp53*-axis factors confer a proliferative advantage to ESCs.

a) Scatter plot representing the count of the four sgRNAs targeting the *Trp53* in the ESC population compared to the plasmid pool of sgRNAs.

b) Scatter plot representing the count of the four sgRNAs targeting the *Chek2, Usp28* and *Setd2* in the ESC population compared to the plasmid pool of sgRNAs.

c) Schema representing the factors found to be required for ESC proliferation that act on the *Trp53*-axis. DNA DSBs activate the ATM kinase, which then activates CHEK2 by phosphorylation. The phosphorylarion of CHEK2 causes this enzyme to homodimerise, subsequently permitting the phosphorylation of other substrates including TRP53. TRP53 when phosphorylated becomes stabilized and can therefore activate the transcription of genes involved in cell cycle arrest or apoptosis, ultimately mediating the response to DNA damage. USP28 has also been shown to form a complex with 53BP1 and to stimulate TRP53 DNA binding activity.

3.6 - Upscaling of EpiLC differentiation to perform a genetic screen

In order for each sgRNA of the plasmid pool to be present in 2000 cells on average in the PGCLC population (fold-representation of 2000), we aimed to FACS sort 6.6 million cells. This posed a technical limitation since from a 96 well-plate containing one aggregate per well we could only generate a few thousand PGCLCs. Therefore, as a first step to reach the 6.6 million PGCLCs, we had to first be able to generate millions of EpiLCs to then differentiate into PGCLCs and so we could keep the fold-representation of 2000 throughout the screen. Previously, we were culturing the EpiLCs in 12 well-plates as described by Hayashi *et al.* (Hayashi et al. 2011). In order to more easily generate millions of EpiLCs, we aimed at culturing these cells in larger plates so we could more easily culture, wash and tryspinise these cells before differentiation into PGCLCs. In order to test if changing the culture dish had an effect on the differentiation, we cultured the iCas8 8 cell line to generate EpiLCs in 12 well-plates as before but also in 10 cm dishes. Then we followed the protocol of differentiation using the adjusted volumes for the respective dishes and differentiated the EpiLCs grown in the two different ways into PGCLCs. Using flow cytometry we observed no difference in the profiles and percentages of SSEA1 and GFP positive cells (**Figure 7a and b)**.

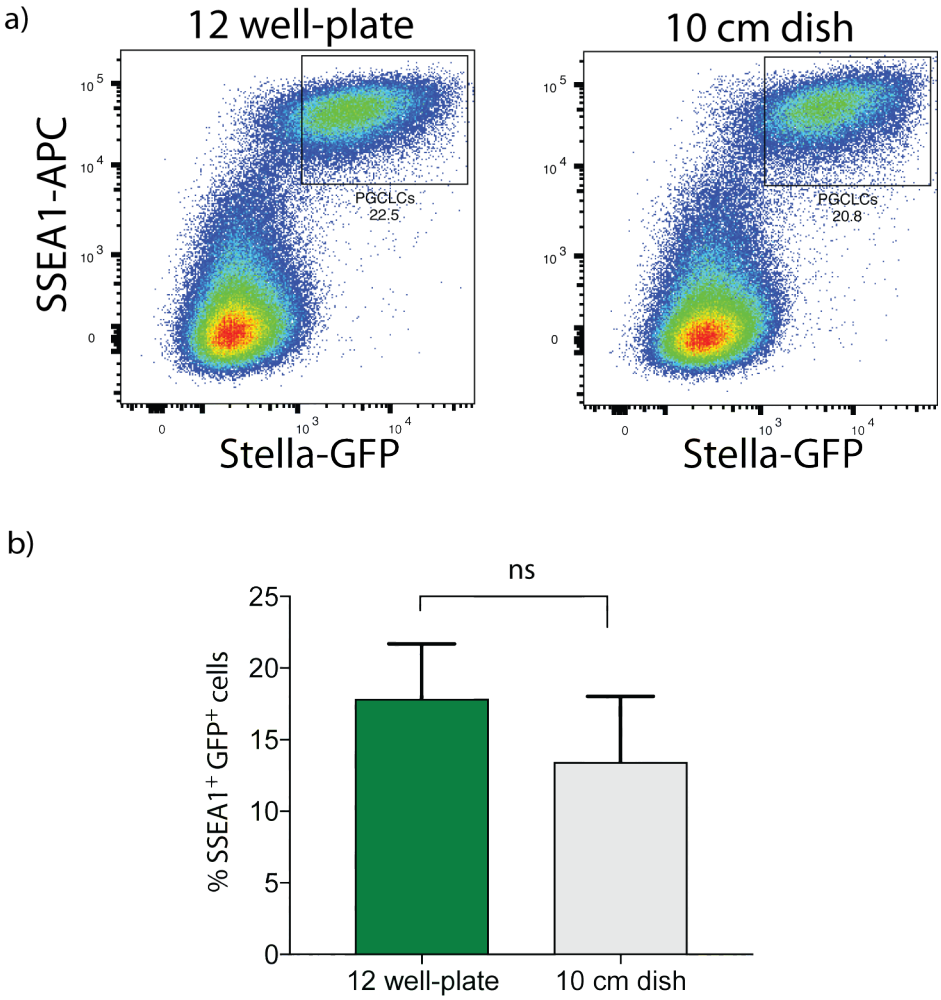


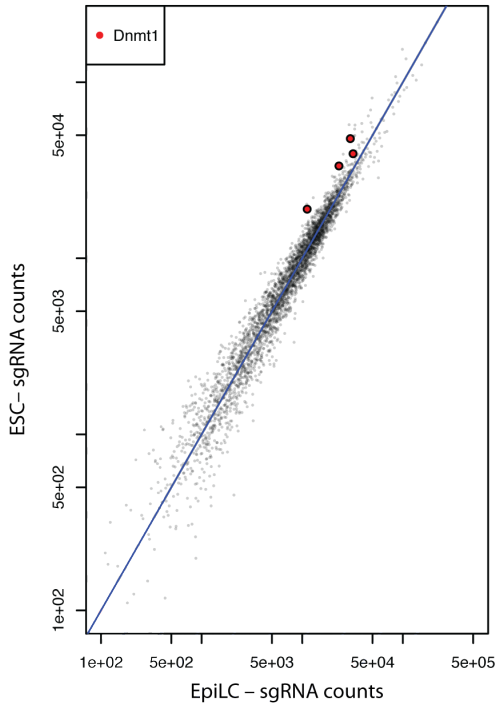
Figure 7 – Upscaling production of EpiLCs using 10 cm dishes.
a) Representative flow cytometric profile of PGCLCs induced from EpiLCs grown in 12 well-plates and 10 cm dishes. The frequency of PGCLCs refers to the frequency of GFP and SSEA1 double positive population.
b) Quantification of the percentage of iCas9 8 day 4 PGCLCs induced from EpiLCs grown in 12 well-plates and 10 cm dishes from iCas9 8 ESCs. The data represent the mean of three independent experiments. The statistical analysis performed was an unpaired *t*-test.

3.7 - Identification of DNA repair factors required for the exit of mouse embryonic stem cells from pluripotency

Having developed a system that permits us to generate millions of EpiLCs and having shown that the sgRNA library we cloned coupled to the Cas9 inducible ESC line permits the discovery of novel factors in ESCs, we aimed to perform a screen in which we would compare the ESCs with the EpiLCs. This screen would permit to test if the screening platform we developed together with the differentiation system permits the identification of factors required for ESC differentiation to EpiLC. To achieve this we collected 6.6 million ESCs and EpiLCs after doxycycline induction and compared the sgRNAs in both population to find factors required for exit to pluripotency in ESCs. We found no NTC sgRNAs in the top rank of gene dropouts, suggesting that the genes we found are very likely to play a role in the differentitation of ESCs to EpiLCs. The gene *Dnmt1* (DNA methyltransferase 1) was in the top of the list of dropouts in the EpiLC population (**Figure 8a**). *Dnmt1* is a maintenance DNA methyltransferase enzyme that transfers methyl groups to cytosine bases in genomic DNA. This protein is the principal enzyme responsible for maintaining methylation patterns after DNA replication with a preference for hemi-methylated DNA. It has been previously reported that *Dnmt1* is required for ESC differentiation (Panning and Jaenisch 1996) and a CRISPR/Cas9 screen using the PGCLC system also identified *Dnmt1* to be required for the transition from ESCs to EpiLCs (Hackett et al. 2018). This result suggests that inability to maintain methylation patterns in the genome impedes differentiation in ESCs. However, which genes are required for the transition from ESC to EpiLC and how they are regulated via *Dnmt1* remains to be explored. In addition to *Dnmt1*, we found other factors in the top of the list of most depleted genes in EpiLCs, such as *Usp28, Rnaseh2a* (Ribonuclease H2 subunit a) and *Cdh1* (Cadherin 1*)* (**Figure 8b**). *Usp28* was previoulsy observed to be required for ESC proliferation and we hypothesised this gene would function together with *Trp53* regulating proliferation and differentiation. Since *Usp28*-deficient mice are viable, which suggests absence of this gene still permits differentiation *in vivo*, this putative hit would have to be first validated *in vitro* before studying its role in ESC differentiation. We also found sgRNAs targeting *Rnaseh2a* to be depleted in EpiLCs in our screen. *Rnaseh2a* is a component of the heterotrimeric ribonuclease H enzyme (RNASEH2) (Jeong et al. 2004), an enzyme

responsible for removing ribonucleotides from genomic DNA (Hiller et al. 2012). Since *Rnaseh2a*-deficient mice die during embryonic development one can hypothesise that during differentiation an increase of ribonucleotides accummulates in the genome which, if not removed by RNASEH2A, cause lethality and impair differentiation. In addition to *Rnaseh2a*, the genes encoding the remaining subunits of the trimeric RNASEH2 complex (*Rnaseh2b* and *Rnaseh2c*), were not observed to be depleted in the EpiLC population since sgRNAs targeting these genes were not present in the initial library of sgRNAs. We also found *Cdh1*, which encodes for a calcium-dependent cell-cell adhesion protein which has been shown to be required for ESC differentiation (Mohamet, Lea, and Ward 2010). In addition, *Cdh1*-deficient mice fail to survive embryonic development (Riethmacher, Brinkmann, and Birchmeier 1995) suggesting this gene may be required for differentiation. However, the requirement for *Cdh1* during exit from pluripotency has not been explored further. This screen allowed us to conclude that the platform we developed coupling CRISPR screening with ESC differentiation permits the discovery of known and novel genetic factors. The next step would then be to use this tool to find novel factors required for PGCLC differentiation.

a)



b)

Rank	Z-score	Gene
1	-2.798	Dnmt1
2	-2.16	Usp28
3	-2.098	Rnaseh2a
4	-2.069	Cdh1

Figure 8 – A CRISPR Screen identifies *Dnmt1*, *Usp28, Rnaseh2a* and *Cdh1* to be required for the exit from pluripotency.
a) Scatter plot representing the count of the four sgRNAs targeting the *Dnmt1* in the EpiLC population when compared to the ESC population.
b) Z-score rank of top depleted genes in the EpiLC population when compared to the ESC population.

3.8 - Upscaling of PGCLC differentiation to perform a genetic screen

In order to restrict the time of differentiation of ESCs to PGCLCs to the minimum so that DNA repair deficient cells with proliferative defects do not get outcompeted by other cells we decided to perform the PGCLC screen after 4 days of PGCLC differentiation. Also we observed that unlike after 4 days, aggregates after 6 days start to extrude extracelular matrix components which make these samples harder to process and therefore difficult to generate and FACS sort millions of PGCLCs. We were previously generating PGCLCs in Lipidure-Coat 96 well-plates according to the protocol described by Hayashi *et al.* (Hayashi et al. 2011) which yielded only 96 aggregates per plate corresponding to a few thousand PGCLCs. Therefore based on the method used by Ohta, H, *et al.* (Ohta et al. 2017) we adapted a new system for growing the PGCLC aggregates in 800 μm Aggrewell plates. These plates contain 800 μm diameter micro wells in which the aggregates can form, therefore permitting us to generate around 300 aggregates per well that corresponds to 7200 aggregates in a full plate and around 0.3 million PGCLCs after 4 days of differentiation. The aggregates we generated in Aggrewell plates achieved the same flow cytometry profile and percentage of PGCLCs as observed in Lipidure-Coat 96 well-plates (**Figure 9a**). Additionally, microscopic analysis did not reveal any morphological differences between the PGCLC aggregates grown in 96 well-plates compared to Aggrewell plates (**Figure 9b**). Altogether we can conclude that we have devised a method that permits to generate millions of PGCLCs so we can perform a robust genetic screen.

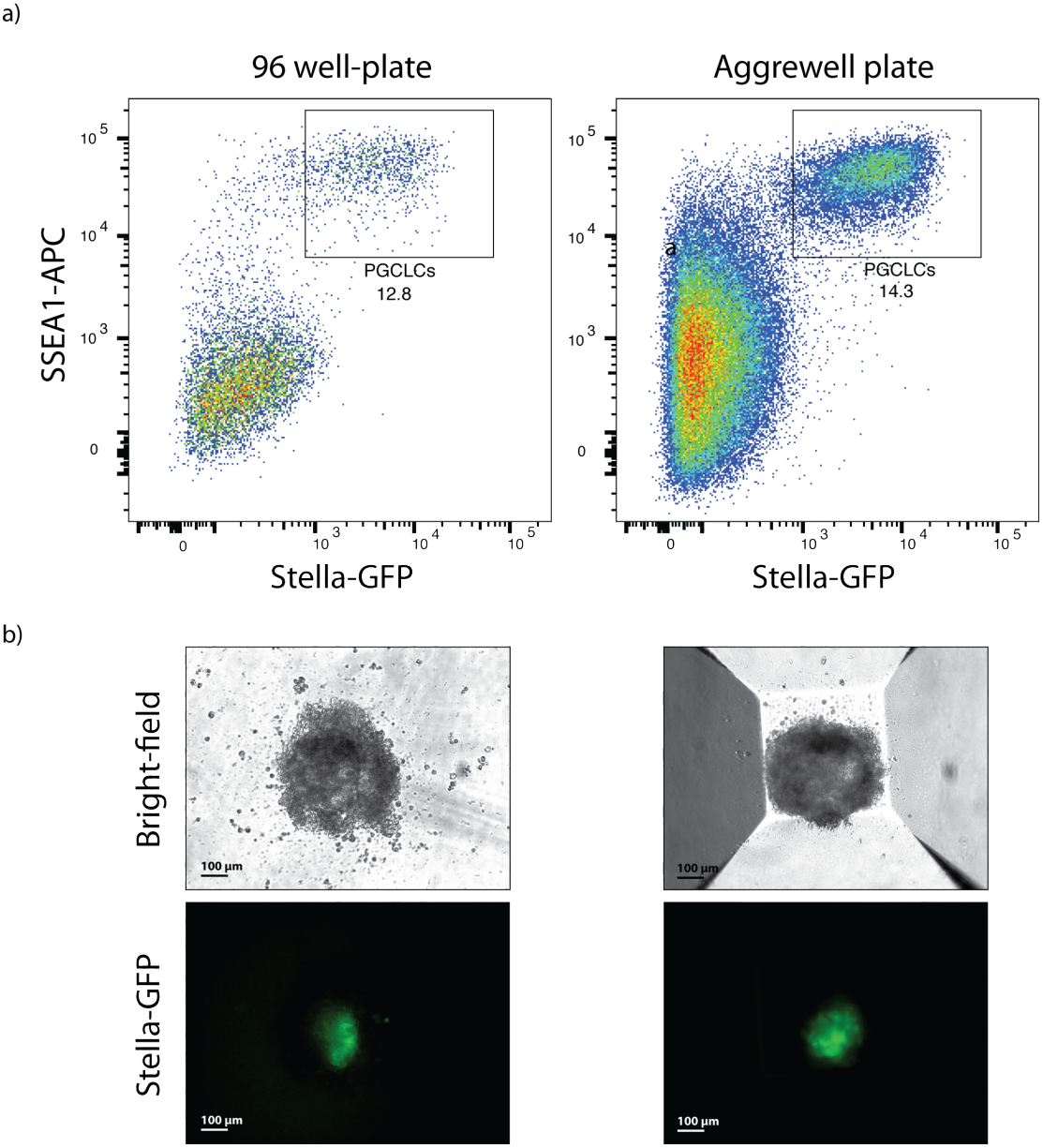


Figure 9 – Large scale production of PGCLCs using 800 µm Aggrewell plates.
a) Representative flow cytometric profile of PGCLCs induced in 96 well-plates and 800 µm Aggrewell plates from iCas9 8 ESCs after SSEA1 staining. The frequency of PGCLCs refers to the frequency of GFP and SSEA1 double positive population.
b) Bright-field and epifluorescence microcopy images of day 4 PGCLC induced from iCas9 clone 8 ESCs in both 96 well-plates and 800 µm Aggrewell plates.

Having optimised the concentration and duration of doxycycline treatment, we asked if we could observe a defect in the PGCLC population when targeting GFP after the ESC KO generation and differentiation step. This would ultimately provide us with some evidence regarding the efficiency of this screening strategy. To achieve this, the iCas9 8 cell line was transduced at a low MOI with sgRNAs targeting GFP, selected these with puromycin and treated them in the presence or absence of 2.5 μg/ml of doxycycline for 21 days. These cells were then induced to differentiate into EpiLCs and into PGCLCs and subsequently analysed by flow cytometry. **Figure 7a** and **b** show the presence of a GFP negative population in the experiment done in the presence of doxycycline wheres this GFP negative population is not observed in the untreated control. The GFP quantification in the SSEA1 positive cell population showed a striking reduction in the GFP signal, which led us to conclude that this system will permit the discovery of new factors required for the development of PGCLCs. The reduction in percentage of the GFP-positive cells in this experiment (around 40%) which is a proxy for gene disruption, is not as striking as the reduction of up to 80% observed in **Figure 10a and b** for CD117 disruption. This observation may be explained by an inferior efficiency of the sgRNAs used in both strategies. However, it is more likely that since there may be multiple copies of GFP in the Stella BAC9 cell line, it makes it more challenging to fully disrupt GFP in comparison to CD117 which has only two copies in the genome.

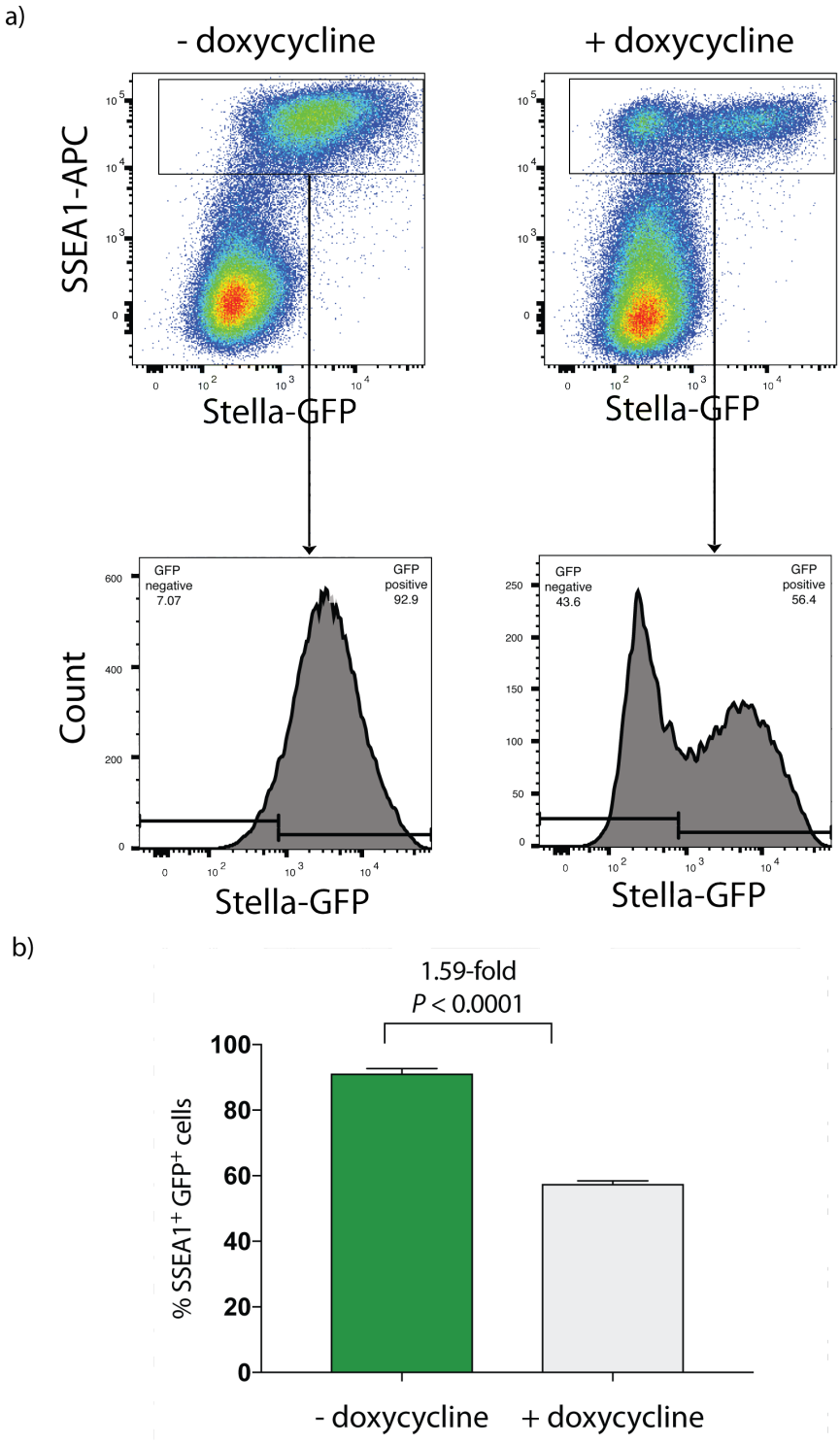


Figure 10 – Deletion of GFP in ESCs using optimised conditions for screening causes a reduction in the frequency of PGCLCs.
a) Representative flow cytometric profile of iCas9 8 day 4 PGCLCs transduced with a GFP sgRNA at a MOI inferior to 0.3 treated with 0 and 2.5 µg/ml of doxycycline for 21 days and after SSEA1 staining.
b) Quantification of the percentage of iCas9 8 day 4 PGCLCs transduced with a GFP sgRNA at a MOI inferior to 0.3 after treatment with 0 and 2.5 µg/ml doxycycline for 21 days and after SSEA1 staining. The data represent the mean of three independent experiments. The statistical analysis performed was an unpaired *t*-test.

3.9 - A CRISPR screen identifies novel repair factors in PGCLCs

Having developed all of the reagents and tools to perform the CRISPR screen and having shown that the developed strategy works for GFP, which is our reporter of PGCLCs, we performed the PGCLC screen. In order to do this, we infected the iCas9 cell line with the sgRNA library, which we then treated with doxycycline for 21 days. After this period, the ESCs were differentiated into EpiLCs. In order to generate 6.6 million PGCLCs, we had to plate 13 full Aggrewell plates since we generated around 0.02 million PGCLCs per well after plating 600.000 EpiLCs into one well of an Aggrewell plate. This therefore meant that we had to plate around 200 million EpiLCs into Aggrewell plates to then differentiate into PGCLCs and also collect 6.6 million EpiLCs for NGS. Finally, since from one 10 cm dish we could generate around 6 million of EpiLCs, we differentiated initially around 100 million ESCs to EpiLCs while also keeping 6.6 million for NGS. During the FACS purification loss of cells due to cell death caused the average representation of each guide in the PGCLC population to be slightly over 100-fold. Even though we did not reach the aimed representation of the sgRNAs in all populations of the screen, we performed subsequent NGS and compared the abundance of sgRNAs present in the EpiLCs with the sgRNAs present in the PGCLCs.

We ranked the genes corresponding to sgRNAs that were the most depleted in PGCLCs and observed that from the ones which did not cause embryonic lethality a considerable proportion (80%) were involved in fertility phenotypes described in the literature (**Figure 11 a and b**). No NTC sgRNAs were observed in this list.

Interestingly, we did not observe *Ercc1* to be one of the most depleted genes in the PGCLCs (rank = 368) (**Figure 11a and c**). This may be due to the fact that the reduction in PGCs observed in *Ercc1*–deficient embryos occurs later in PGC development (around E11.5) whereas the day 4 PGCLCs transcriptionally resemble E10.5 PGCs where a significant numeric reduction is not yet observed.

Also we observed some essential genes to be on the top of the list including: *Rfc1* (Replication Factor C Subunit 1), *Snprd1* (Small nuclear ribonucleoprotein D1 polypeptide) or *Smc5 (*Structural maintenance of chromosomes 5). The presence of sgRNAs targeting these genes may be explained by inefficient gene disruption which would cause the absence of depletion from the ESC population. Furthemore, the cells containing these sgRNAs can be either false-positives and therefore these genes would

have no effect on PGCLC formation or the consequence of having truncated but still functional versions of these genes that would confer cells an inabilty to generate PGCLCs.

However, we found several known factors to be required for PGC development to be highly depleted in the PGCLC population. This included *Pds5b* (Precocious dissociation of sisters 5b), which encodes a protein that interacts with regulatory component of cohesin and genetically or physically interacts with the cohesion factors *Scc1*, *Scc2* and *Scc3* (Sister chromatid cohesion), *Esco1* (Establishment of sister chromatid cohesion N-acetyltransferase 1), *Smc1* and *Smc3* (Structural maintenance of chromosomes 1 and 3) (Losada, Yokochi, and Hirano 2005; Panizza et al. 2000) to ensure proper sister chromatid cohesion during mitosis and meiosis (Panizza et al. 2000; Hartman et al. 2000; van Heemst and Heyting 2000). Furthermore, absence of *Pds5b* has been previously shown to be required for PGC development in mice (Zhang et al. 2007).

Also, we observed some new interesting factors. One of them is named *Spidr* (Scaffold protein involved in DNA repair) (**Figure 11e)**. SPIDR is known to interact with BLM and RAD51 and has been shown to cause increased rates of sister chromatid exchange and defects in HR (Wan et al. 2013). Furthermore, SPIDR has also been shown to interact with FIGNL1 and the authors claim that these factors act together in HR repair (Yuan and Chen 2013). Interestingly, patients in a consanguineous family with a mutation in SPIDR presented failure to reach puberty and exhibited very high levels of the hormones that stimulate gonad development (Smirin-Yosef et al. 2017). This is exactly the phenotype we would expect if this factor was to be required for the development of PGCs.

In the top of our list ranking number 8, we found *Tdg* which is a DNA glycosylase shown to play roles in maintaining epigenetic stability and DNA demethylation. However *Tdg* knockout mice die during embryonic development so the fertility phenotype cannot be assessed (Cortazar et al. 2011; Cortellino et al. 2011). Also, we found the *Msh2* (MutS homolog 2) (rank=2) and *Msh6* (MutS homolog 6) (rank=14) mismatch repair genes which together form the heterodimer MutSα. Even though both mouse models for these genes have been generated and were shown to be fertile (Reitmair et al. 1995; Edelmann et al. 1997), mice deficient in *Msh2* have been proposed to have germ cell defects in adults comparable to what is observed in *Ercc1*-deficient mice (Paul et al. 2007).

In addition to this, we found *Rev1* in the screen (rank=15) (**Figure 11a and f)**. *Rev1* is a deoxycytidyl transferase enzyme that also serves as a scaffold to recruit DNA polymerases involved in TLS (Pustovalova, Bezsonova, and Korzhnev 2012; Masuda et al. 2001). *Rev1*-deficient mice were previously generated and exhibited growth retardation and impaired class-switch recombination (Jansen et al. 2006). Interestingly, no germ cell or fertility defects were studied in detail in these mice. Overall we can conclude that we have optimised and developed a system which we used to perform a genetic screen to find novel DNA repair factors required for PGC development. We found known factors to be essential for PGC development such as *Pds5b* but more interestingly we found novel factors with potential roles in fertility such as *Spidr* or *Rev1*. The next logical step is to validate these hits which will inform us if these factors are indeed required for PGC development *in vitro* and *in vivo*.

a)

Rank	Z-score	Gene
1	-2.5364475	Rfc1
2	-2.5340903	Msh2
3	-2.314567	Dek
4	-2.2624423	Ube2c
5	-1.9900682	Cebpg
6	-1.9598061	Snrpd1
7	-1.9099297	Cul1
8	-1.8542189	Tdg
9	-1.8002815	Acly
10	-1.7629391	Nop16
11	-1.72293	Pa2g4
12	-1.7020209	Smc5
13	-1.6909486	**Pds5b**
14	-1.6560224	Msh6
15	-1.6397658	**Rev1**
16	-1.6258273	**Spidr**
...	...	...
368	-0.2928725	**Ercc1**

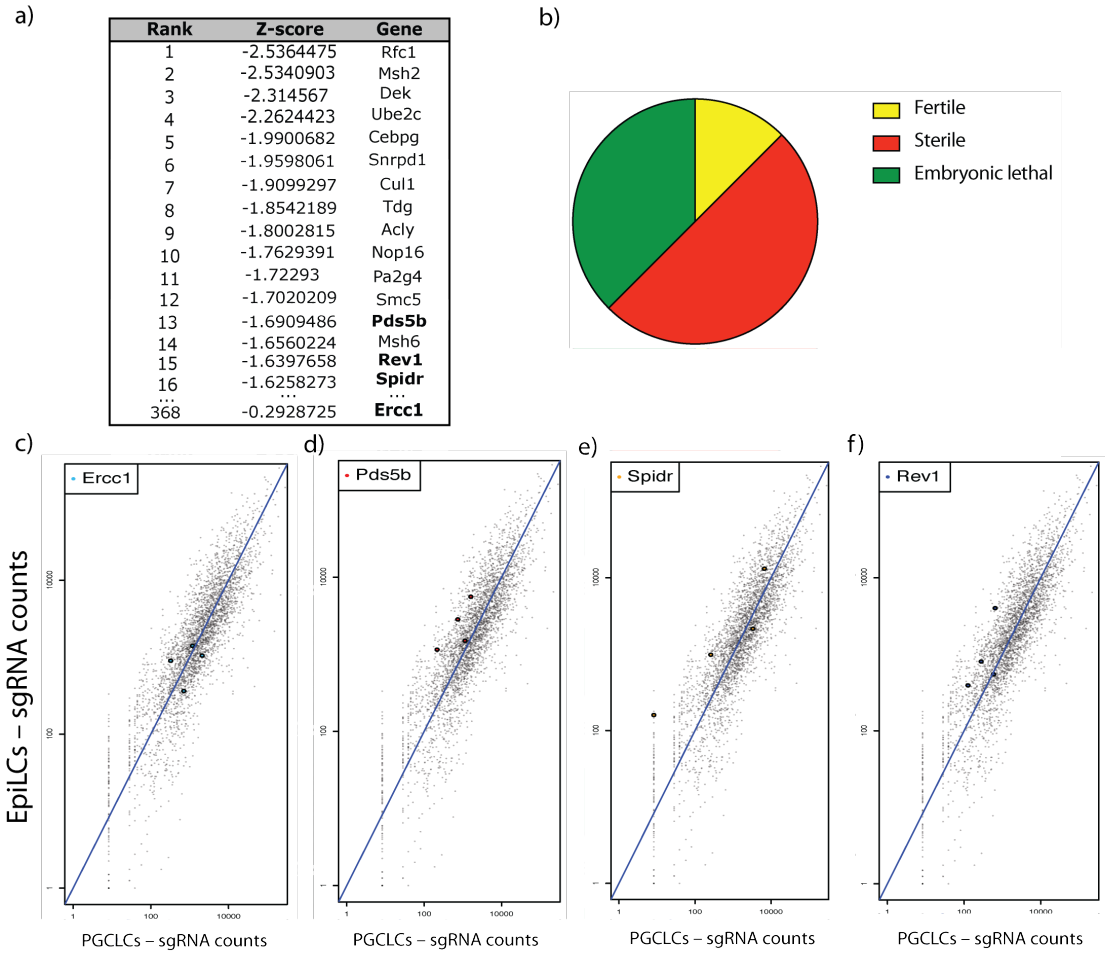


Figure 11 – A CRISPR Screen identifies *Rev1* and *Spidr* to be required for PGCLC development.
a) Z-score rank of top depleted genes in the EpiLC population when compared to PGCLCs.
b) Pie chart representing the proportion of genes from the table of top hits which when disrupted cause embryonic lethality in mice or that have been shown to be involved in sterility or fertility phenotypes.
c) Scatter plot representing the count of the four sgRNAs targeting *Ercc1* in the EpiLC population when compared to PGCLCs.
d) Scatter plot representing the count of the four sgRNAs targeting *Pds5b* in the EpiLC population when compared to PGCLCs.
e) Scatter plot representing the count of the four sgRNAs targeting *Spidr* in the EpiLC population when compared to PGCLCs.
f) Scatter plot representing the count of the four sgRNAs targeting *Rev1* in the EpiLC population when compared to PGCLCs.

Summary

The data presented here represent all of the steps of optimisation towards performing a CRISPR/Cas9 knockout screen using the PGCLC *in vitro* system. These steps included the cloning and assembly of a sgRNA library targeting genes involved in genome stability in the mouse genome, generating a doxycycline inducible Cas9 cell line and testing the cell line for CRISPR screening. First, we performed a screen in ESCs to test if the tools we generated would yield known factors required for ESC proliferation and to test if essential genes would be depleted as anticipated. As expected, we found sgRNAs targeting *Trp53* to be highly enriched in ESCs whereas the most depleted sgRNAs targeted essential genes. Having validated this step we upscaled the production of EpiLCs to generate enough cells for PGCLC differentiation. Having a tool to generate vast numbers of EpiLCs we performed a screen to test if differentiating ESCs to EpiLCs would yield known and novel factors required for ESC differentiation. We observed *Dnmt1* to be in the top of the list as observed in similar screens in the past. Finally, we upscaled the production of PGCLCs to perform a screen using this population. We performed a screen in PGCLCs, which permitted us to identify known factors required for PGC development including *Pds5b* but also putative new factors such as *Spidr* and *Rev1*.

Chapter 4 - Identification of *Rev1* as a critical factor required for early PGC development

4.1 - Disruption of *Rev1* in ESCs

One of the most interesting putative hits we obtained from the screen was *Rev1*. REV1 is a deoxycytidyl transferase enzyme that also acts as a scaffold to recruit DNA polymerases involved in TLS (Pustovalova, Bezsonova, and Korzhnev 2012; Masuda et al. 2001). *Rev1* is recruited to chromatin upon DNA damage via its BRCT domain (Guo et al. 2006) and is required to maintain cellular resistance to a number of DNA damaging agents including UV irradiation. However, its deoxycytidyl transferase activity is dispensable for maintaining resistance to DNA damage. REV1 interacts with other TLS DNA polymerases via its C-terminus, which implicates it as an adaptor for polymerase switching at the lesion site. This region is critical for cells to repair damage caused by DNA damaging agents. The potential discovery of the requirement for *Rev1* in PGCs is particularly interesting because it is known that *Rev7*-deficient mice and ESCs have reduced abilities to generate PGCs *in vivo* and *in vitro*, respectively (Pirouz, Pilarski, and Kessel 2013; Rahjouei et al. 2017; Watanabe et al. 2013). In addition to this, *Rev1*-deficient mice not only exhibit growth retardation and impaired class-switch recombination but have also been shown to be sterile. Since the origin of the sterility phenotype has not been explored, we decided to follow up on this putative hit, which would further validate the screen.

In order to perform the validation, we generated a new sgRNA that was not initially present in the sgRNA library used in the screen to therefore remove the consequence of off-target effects. Using this sgRNA we disrupted *Rev1* in our reporter ESC line to interrogate if absence of *Rev1* is required for PGC development *in vitro*. We chose to target the exon 10 since this exon encondes for the deoxycytidyl transferase domain, which is responsible for the cataltic activity of REV1 (**Figure 1a**). The targeted region was amplified by PCR, cloned into a TOPO Zero Blunt vector and

Sanger sequencing was performed to confirm the identity of the indels disrupting the genes (**Figure 1a**). Finally, to confirm biallelic disruption, we performed a colonogenic survival assay, which showed clear hypersensitivity to mitomycin C (**Figure 1b**) as previously described (Hicks et al. 2010).

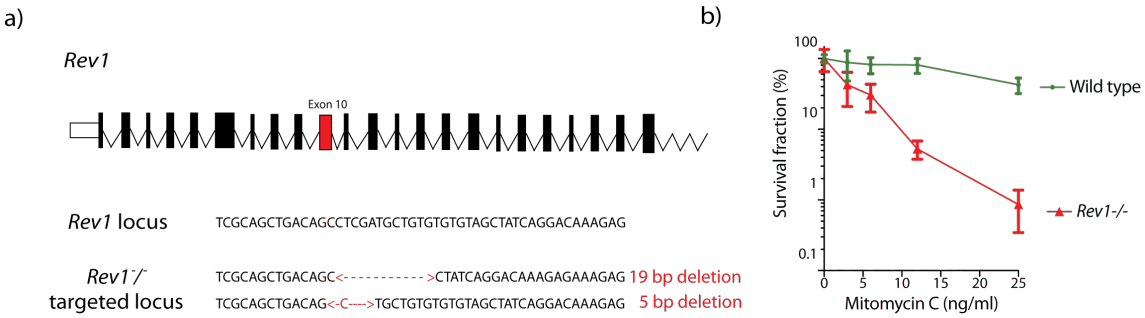


Figure 1 – Disruption of *Rev1* in ESCs.
a) Schema representing the *Rev1* locus with the targeted exon indicated in red and the sequence of disrupted alleles after CRISPR-mediated disruption. Coding exons are represented as black boxes.
b) Colonogenic survival assay with mitomycin C. ESCs were incubated with increasing doses of mitomycin C and the number of colonies observed were counted following 6 days. The data represent the mean of three independent experiments each carried out in duplicate made relative to the untreated of each cell line.

4.2 - *Rev1* is required for PGC development *in vitro*

In order to ascertain if *Rev1* is required for PGC development *in vitro* and validate our screen, we induced *Rev1*-deficient ESCs to differentiate to PGCLCs. Microscopic examination revealed a clear reduction in the number of *Stella-GFP* positive cells in the absence of *Rev1* (**Figure 2a**). Furthermore, the frequency of PGCLCs was determined by flow cytometry revealing that loss of *Rev1* leads to a 4-fold reduction in PGCLC frequency (**Figure 2b and c**). This set of experiments therefore validated one of the hits found in our screen and indicates the requirement for the TLS enzyme *Rev1* for the development of PGCs *in vitro*.

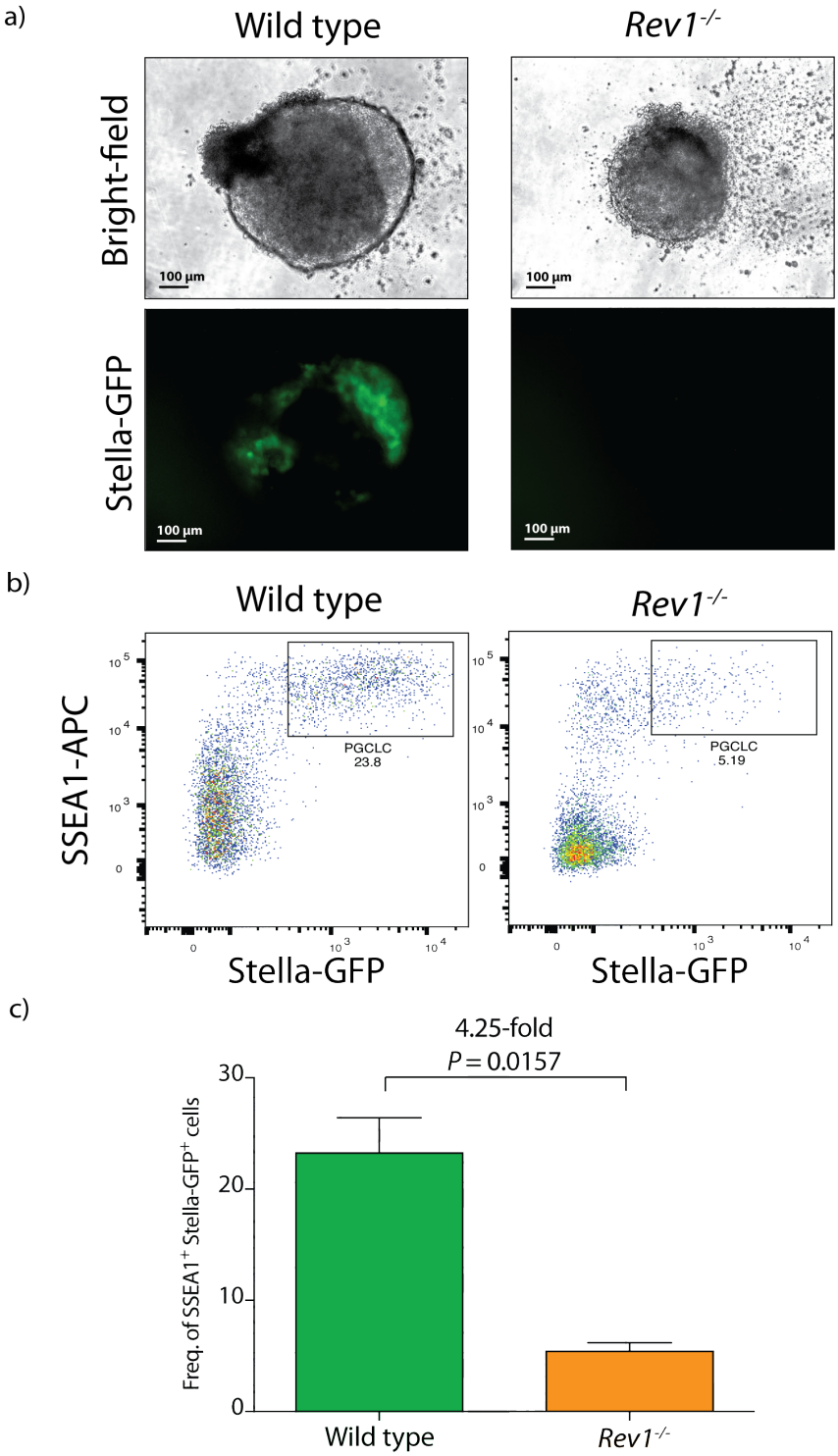


Figure 2 – Absence of *Rev1* impairs the development of PGCs *in vitro*.
a) Bright-field and epifluorescence microscopy images of day 4 Stella BAC9 Wild type and *Rev1*$^{-/-}$ PGCLCs.
b) Representative flow cytometry profiles of PGCLC induced from Stella BAC9 Wild type and *Rev1*$^{-/-}$ PGCLCs after SSEA1 stain. The percentage of PGCLC refers to the frequency of GFP and SSEA1 double positive population.
c) Quantification of the frequency of SSEA1 and GFP positive cells (PGCLCs) in the different cell lines shown. The data represent the mean of three independent experiments. The statistical analysis performed was an unpaired *t*-test.

4.3 - Rev1$^{-/-}$ mice display a numerical defect in PGCs which results in impaired fertility

Having observed that disruption of *Rev1* in ESCs caused a decrease in the number of PGCLCs *in vitro,* we tested whether this would also happen in mice. This served as a first validation of our hits *in vivo* and would provide further evidence that the system we setup is giving us new valuable insight into the biology of PGCs. One potential caveat of this system could be that the factors we found, namely DNA repair genes, would only have a reduction in PGCLC frequency since certain DNA repair genes impair cell proliferation. Also, using this system we only used one marker of PGCs – *Stella* – therefore making it possible that some of the factors we found directly affect the ability to express this marker rather than having an effect on PGC development. Finally, since we are using an *in vitro* system, many of the aspects such as the interaction with the niche cells in which PGCs are specified cannot be reproduced as when using a developing organism. This may also contribute for the presence of false-positive and false-negative putative hits in our screen. Therefore, the use of an organism as a model and a new reporter would bypass most limitations of using an *in vitro* system in order to validate putative hits from the screen. To achieve this goal, we first investigated the histology of the testis and ovaries of adult *Rev1*-deficient mice. In the *Rev1*-deficient testis we can observe a complete absence of germ cells in the seminiferous tubules with all tubules showing a Sertoli-cell only phenotype (**Figure 3a**). The absence of germ cells in the male gonads causes therefore these mice to become sterile as previously reported (Jansen et al. 2006). In the *Rev1*-deficient females we can observe absence of primary and maturating follicles therefore causing the absence of oocytes. This phenotype causes the sterility we observed by in the absence of *Rev1*. Then, in order to investigate if the absence of germ cells in adult mice was due to inability to develop PGCs, we performed timed matings between *Rev1* heterozygous mice carrying a PGC-reporter in which the expression of GFP is driven by a fragment of the Oct4 promoter (known as Gof18-GFP) (Szabo et al. 2002; Yeom et al. 1996). This would show that disruption of *Rev1* is not directly affecting the expression of *Stella* which would reduce the PGC numbers due to failure to activate the reporter rather than inability to correctly specify or develop PGCs. Also, since we performed the screen with day 4 PGCLCs, which transcriptionally resemble early PGCs, we chose to perform

the PGC analysis earlier in development, at E9.5, and not at E12.5 as done before for *Ercc1*-deficient embryos. This analysis would permit us therefore to uncover genes required for PGC specification and early development which should present a phenotype already at this developmental stage unlike *Ercc1*. In order to analyse E9.5 PGCs, we trypsinised and ressuspended the cells of the entire embryo which we analysed using flow cytometry (**Figure 3b and c**). Comparably to the *in vitro* system, we observed a 4.45-fold reduction in the ability to make PGCs in *Rev1*-deficient embryos while comparing to the Wild type littermates (**Figure 3c**). Therefore, these results demonstrated for the first time that *Rev1* is required during PGC development and the absence of this enzyme causes sterility in both sexes.

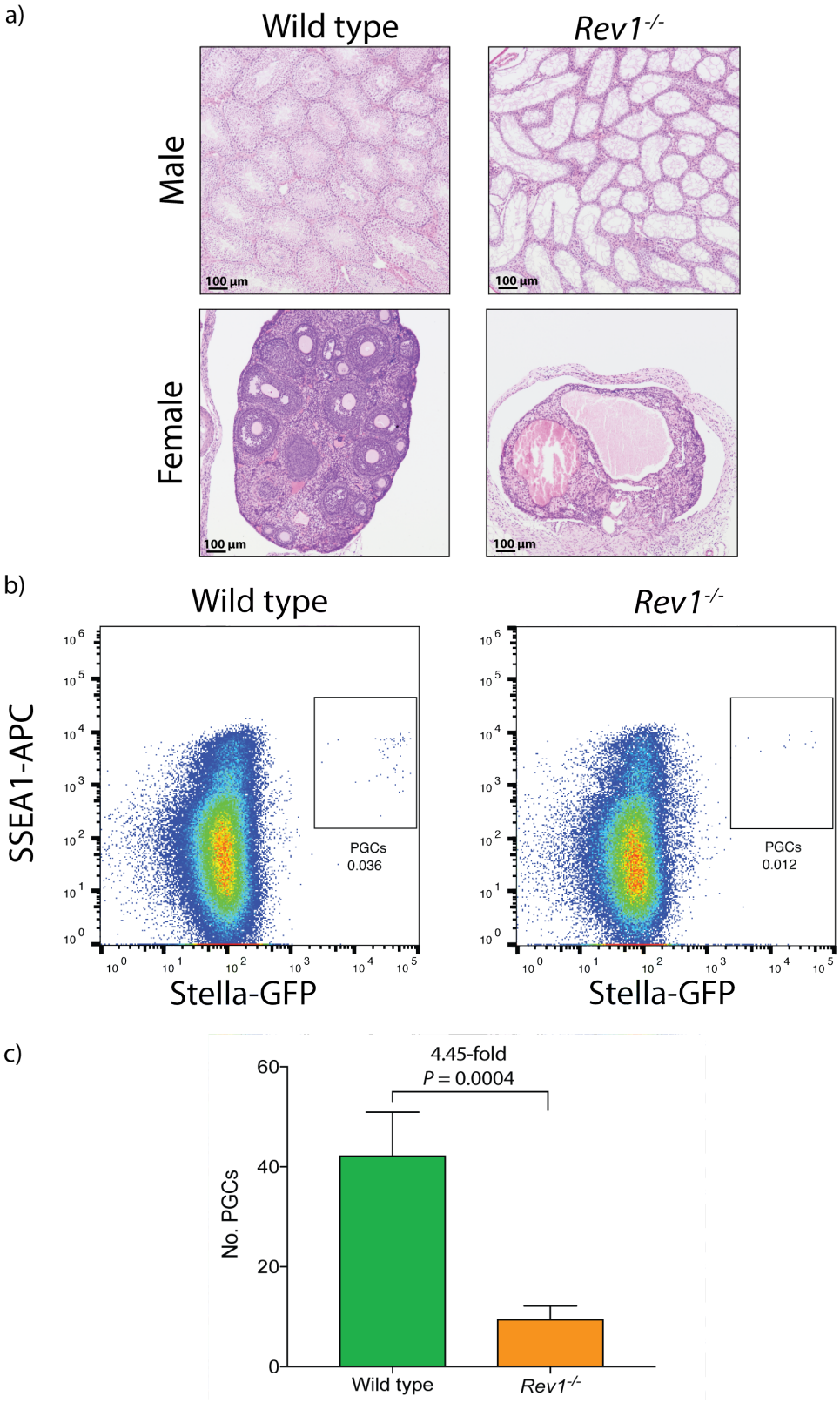


Figure 3 – *Rev1* deficient mice have an early PGC defect and are sterile.
a) Haematoxylin and eosin (H&E) stained sections of ovaries and testes of 8-week old Wild type and *Rev1*$^{-/-}$ mice.
b) Representative flow cytometric profile of PGCs obtained from E9.5 embryos after SSEA1 staining. The frequency of PGCs refers to the frequency of GFP and SSEA1 double positive population.
c) Quantification of the number of SSEA1 and GFP positive cells (PGCs) from Wild type and *Rev1*$^{-/-}$ embryos. The data represent the mean of three independent experiments.

4.4 - Generation of a deoxycytidyl transferase catalytically inactive *Rev1* cell line

In light of the data implicating the role of *Rev1* in the development of PGCs, the next step was to separate the functions of *Rev1* in order to understand which function of *Rev1* is required to maintain genome stability in PGCs. Interestingly, in yeast and bacteria the catalytic activity of *Rev1* is responsible for almost all of the point mutations (Cox and Horner 1982; Glassner et al. 1998). However, in vertebrates the importance of the catalytic activity of *Rev1* in promoting TLS seems to be less significant (Ross, Simpson, and Sale 2005). Therefore, in order to separate the catalytic activity of *Rev1* from its other functions *in vivo*, we aimed to generate a catalytically inactive form of *Rev1* in ESCs. Mice carrying a point mutation that generates a catalytically inactive *Rev1* exhibit a great decrease in G to C and C to G transversions in the immunoglobulin locus of B cells (Masuda et al. 2009). The *Rev1* catalytically inactive cell line ($Rev1^{D568A/E569A}$ from now on named $Rev1^{AA}$) was generated by substituting the catalytic aspartate and glutamate residues, responsible for the deoxycytidyl transferase activity, with two alanine residues. This substitutions have been shown to cause absence of enzymatic activity in terminally mismatched primer template assay using mouse Wild type and REV1AA purified proteins (Guo et al. 2003). In addition to this, the level of mRNA expression in both Wild type and $Rev1^{AA}$ spleen B-cells is equivalent (Masuda et al. 2009) however it has not been shown that the protein is equally stable as the Wild type. In order to generate this mutant cell line, we transfected ESCs with two plasmids: the sgRNA and Cas9-containing plasmid and the plasmid expressing tdTomato. We additionally co-transfected a single stranded oligonucleotide (ssODN) containing the mutation of interest in the *Rev1* exon 10, flanked by two homology arms. This sequence was designed to introduce a new *AciI* restriction site, which allows the screening for the correct integration of the ssODN (**Figure 4a and b**). We found a clone that after PCR amplification of the region and digestion with *AciI* revealed the correct integration of the ssODN and loss of the Wild type allele. This suggested that the ESCs were homozygous for the catalytically inactive form of *Rev1* (**Figure 4c**). To confirm that the ssODN was correctly integrated into the targeted locus we amplified the targeted region, cloned it into a TOPO Zero Blunt vector and confirmed the modification by Sanger sequencing (**Figure 4a and b**). Furthermore, we observed that unlike complete

absence of *Rev1*, the catalytic function of *Rev1* is not required to confer resistance to DNA crosslinking agents (**Figure 4c**) as previously described (Ross, Simpson, and Sale 2005) further suggesting that the REV1AA protein is still sufficiently stable to provide resistance to crosslinking agents.

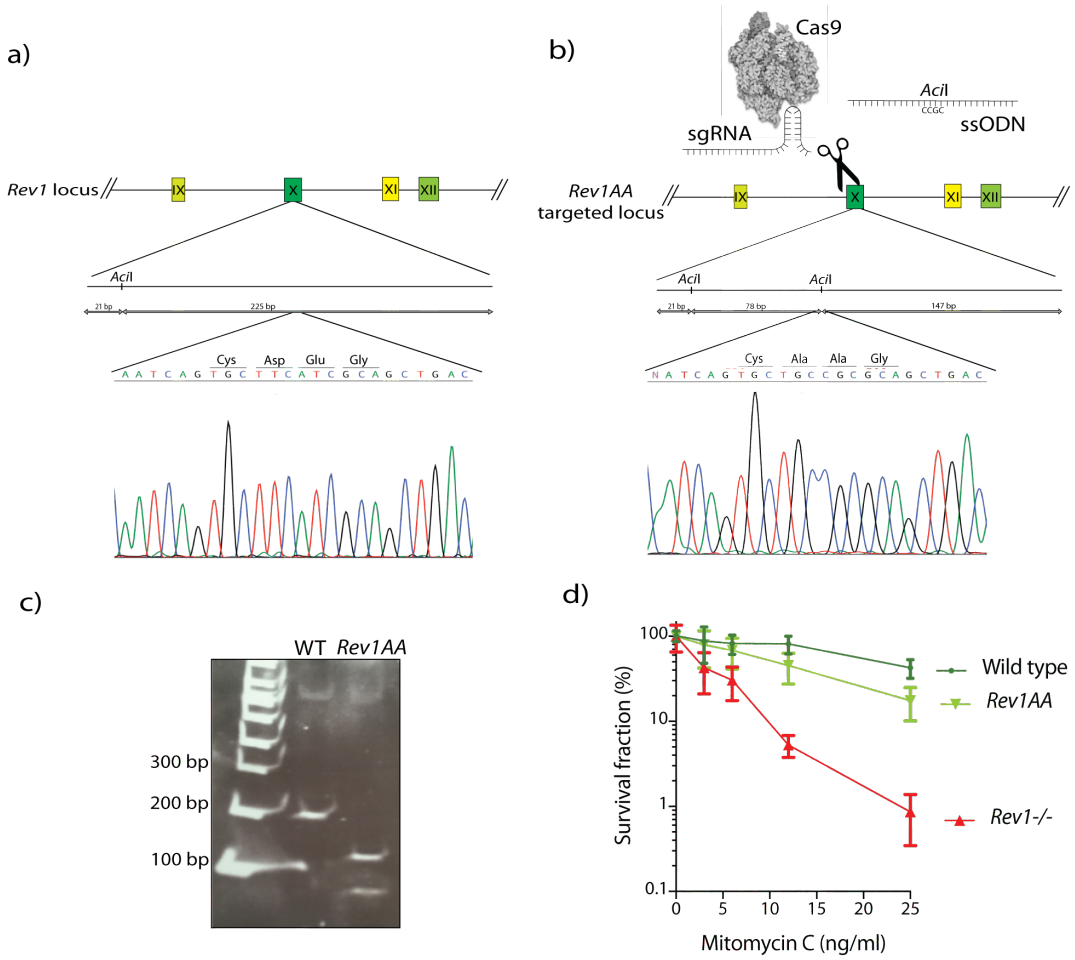


Figure 4 – Generation of a deoxycytidyl transferase catalitically inactive *Rev1AA* ESC line.
a) Schema representing the *Rev1* locus and the sequencing chromatogram covering the deoxycytidyl transferase domain coding region. Coding exons are represented as boxes.
b) Schema representing the *Rev1AA* targeted locus, the sequencing chromatogram spanning the deoxycytidyl transferase domain coding region after targeting with the Cas9 enzyme, the sgRNAs and the ssODN. The insertion of a new *Aci*I restriction site is also depicted. Coding exons are represented as boxes.
c) TBE urea polyacrylamide gel showing the PCR products of the *Rev1* targeted regions after *Aci*I digestion for 1 hour at 37°C in both Wild type and mutant ESC lines.
d) Colonogenic survival assay with mitomycin C. ESCs were incubated with increasing doses of mitomycin C and the number of colonies observed were counted following 6 days. The data represent the mean of three independent experiments each carried out in duplicate.

4.5 - The catalytic activity of *Rev1* is dispensable for PGCLC development

In order to clarify the role of the deoxycytidyl transferase domain of REV1 in germ cells, we used the same strategy as above, by looking at the capacity of *Rev1*$^{-/-}$ and *Rev1AA* ESCs to form PGCLCs (**Figure 5a and b**). The *Rev1AA* ESCs showed equal capacity to form PGCLCs as the parental Wild type ESCs (**Figure 5c**). This result allows us to conclude that the deoxycytidyl transferase activity of REV1 is not responsible for its requirement in PGCLC development.

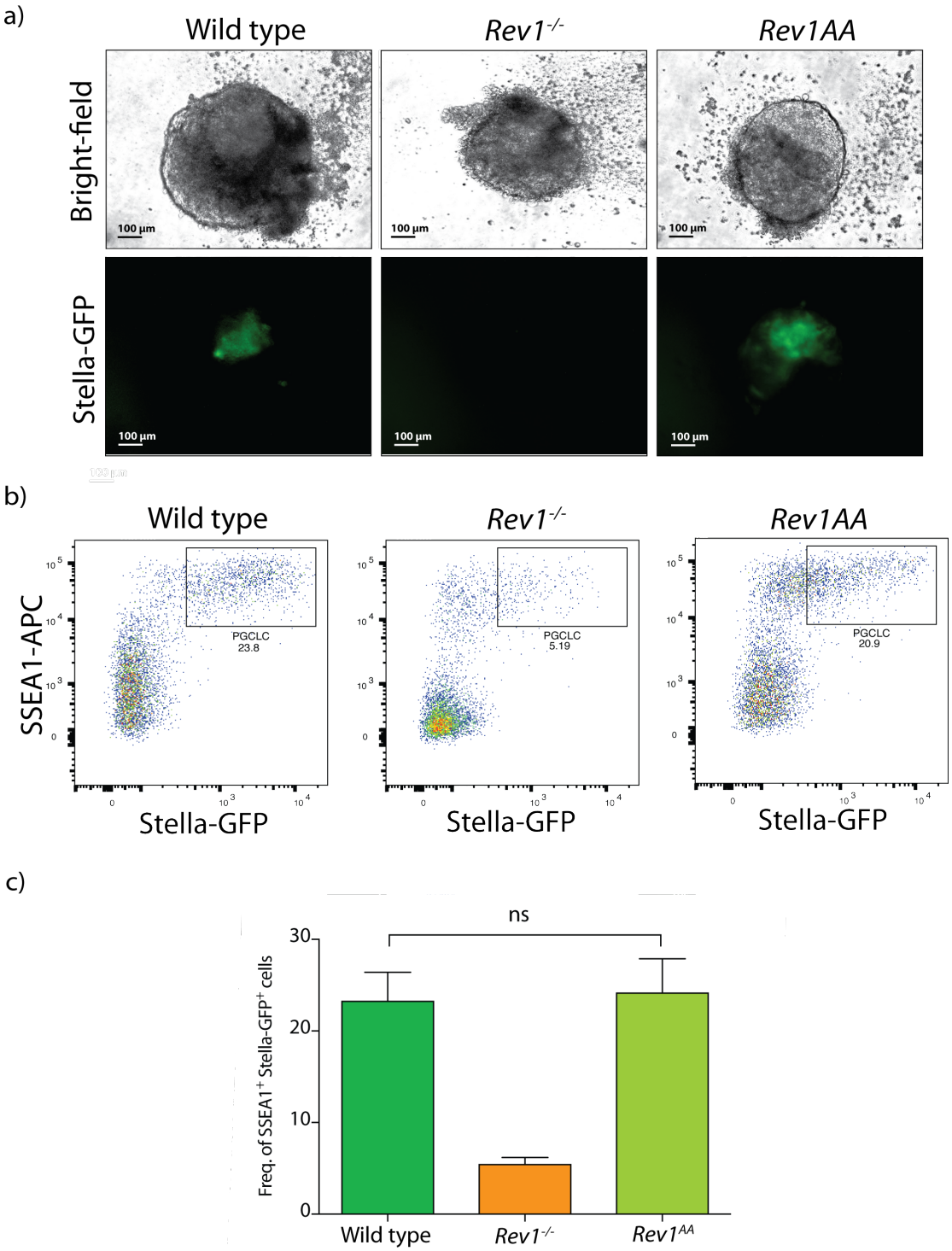
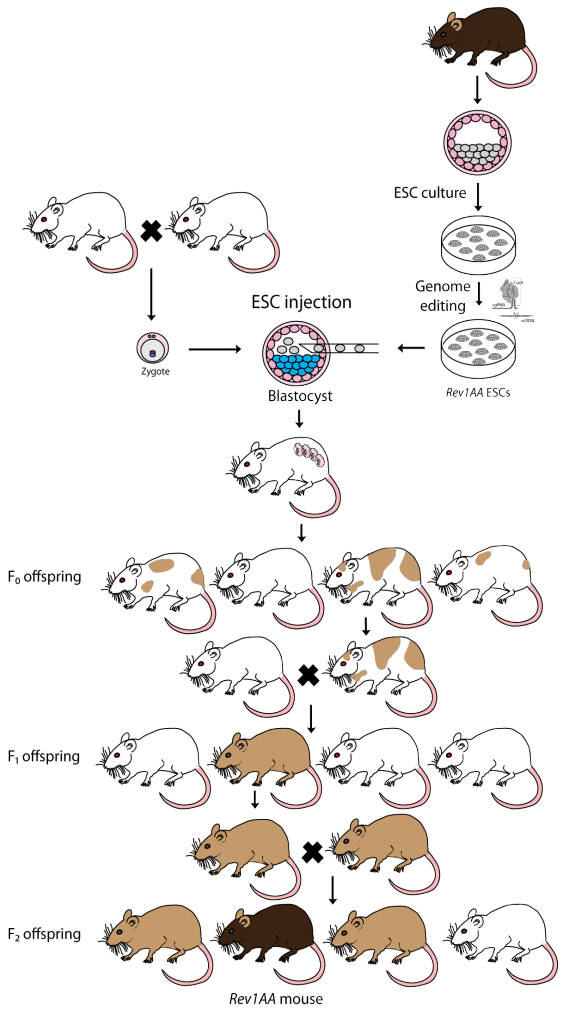


Figure 5 – REV1 catalytic activity is dispensable for PGCLC development *in vitro.*
a) Bright-field and epifluorescence microscopy images of day 4 Stella BAC9 Wild type, *Rev1*$^{-/-}$ and *Rev1AA* PGCLCs.
b) Representative flow cytometric profiles of PGCLCs induced from Stella BAC9 Wild type, *Rev1*$^{-/-}$ and *Rev1AA* PGCLCs after SSEA1 stain. The percentage of PGCLC refers to the frequency of GFP and SSEA1 double positive population.
c) Quantification of the frequency of SSEA1 and GFP positive cells (PGCLCs) in the different cell lines shown. The data represent the mean of three independent experiments. The statistical analysis performed was an unpaired *t*-test.

4.6 - Generation of *Rev1AA* mice

Having generated an ESC cell line which carries a point mutation inactivating the catalytic activity *Rev1* but still capable of generating PGCs *in vitro*, we decided to generate transgenic mice using this cell line so we could validate this phenotype *in vivo.* In order to achieve this, we aimed at generating transgenic mice by ESC injection into blastocysts. The *Rev1AA* ESCs were derived from a mixed background (B6CBA). Therefore when injected into blastocysts derived from B6-albino mice, which have white fur due to a mutation in the tyrosinase gene, the chimeric progeny of this mice may exhibit varying percentages of brown fur. This percentage permits to estimate the contribution of the ESC line to the different germ layers when injected into albino blastocysts. In order to generate *Rev1AA* mice, we then injected the *Rev1AA* ESCs into the ICM of E3.5 mouse embryos which were transplanted to pseudopregnant females (**Figure 6a**). After birth, chimeric mice were generated which permitted us to conclude that the *Rev1AA* ESCs contributed to the development of the chimeric mice (**Figure 6b**). These chimeras were backcrossed with B6-albino mice to allow germ line transmission of the *Rev1AA* allele into constitutive heterozygous mice. These mice were then interbred to obtain the homozygous *Rev1AA* mice.

a)



b)

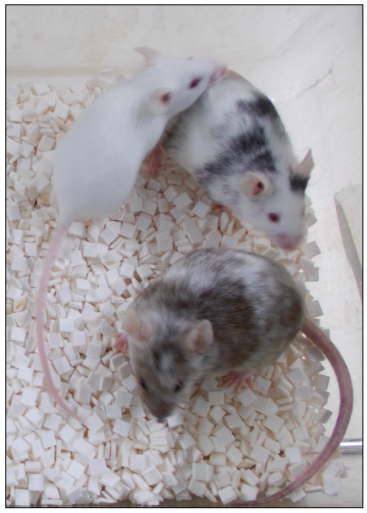


Figure 6 – Generation of *Rev1AA* mice.
a) Schema representing the strategy employed to generate *Rev1AA* mice. First, ESCs carrying the *Stella-GFP* reporter were derived from blastocysts of a B6CBA mixed background. Then, the point mutation in *Rev1* was introduced in the genome of these cells as previously described. Afterwards, ESCs were injected into blastocysts derived from B6-albino mice. These blastocysts containing *Rev1AA* ESCs were then implanted into a pseudo-pregnant female mouse. The progeny of this female mouse (F_0) yielded chimeric mice of different percentages. The highest percentage chimeras were then backcrossed with B6-albino mice to generate *Rev1AA* heterozygous mice. Provided the allele was transmitted via the germ line of the chimeric mice, the offspring of this cross (F_1) would yield constitutive heterozygous *Rev1AA* in a proportion of 1 in 4. The constitutive heterozygous *Rev1AA* mice were then crossed to generate the F_2 offspring in which a homozygous *Rev1AA* mouse is generated at a proportion of 1 in 4.
b) Representative image of chimeric mice with different percentages of chimerism (from 0 to 99%) which are attributed based on the colour of the fur.

4.7 - The catalytic activity of REV1 is dispensable for fertility in mice

Having generated the *Rev1AA* mouse allele, we asked if the deoxycytidyl transferase activity of *Rev1* is responsible for fertility and PGC development *in vivo*. In order to achieve this goal, we performed crosses between both female and male *Rev1AA* mice with Wild type mice and observed that the absence of the catalytic activity of REV1 does not affect fertility in adult mice (**Figure 7a**). Also, we investigated the histology of the testis and ovaries of 8-week old *Rev1AA* mice. H&E staining of *Rev1AA* testis revealed the presence of germ cells at all stages of development (**Figure 7b**). In the *Rev1AA* females we could observe by H&E staining of the ovaries the presence of all stages of maturation of follicles in a comparable manner as in the Wild type (**Figure 7b**). Finally, we assessed the reproductive capacity of both male and female *Rev1AA* mice. Then we performed timed matings between heterozygous $Rev1^{+/AA}$ mice and $Rev1^{+/-}$ to produce embryos with one copy of catalytically dead and Wild type embryos. We analysed the PGCs by flow cytometry at E9.5 (**Figure 7c and d**). Similarly to the *in vitro* system, we observed no reduction in the ability to make PGCs in *Rev1AA* mice compared to Wild type mice. These results show that the deoxycytidyl transferase activity of *Rev1* is not required for the development of PGCs in mice and therefore is dispensable for fertility. The catalytic activity of REV1 was shown to be dispensable for conferring resistance to DNA damaging agents in vertebrates (Ross, Simpson and Sale 2005). In addition, normal development of PGCs was shown to be dependent on different DNA repair pathways such as the FA repair pathway (Hill and Crossan 2019) and HR (Luo et al. 2015; Messiaen et al 2013). This suggests that an alternative function of REV1, which most likely confers resistance to DNA damage, may repair the damage encountered by PGCs. It has been shown that the C-terminus of REV1, which is required for a direct interaction with TLS enzymes and also with polymerase ζ via REV7, is essential for conferring cellular resistance to DNA damaging agents (Ross, Simpson and Sale 2005). Since Rev7 but not any of other TLS polymerases has been shown to be required for fertility and PGC development, one can hypothesize that Rev1 is required in PGCs to recruit DNA polymerase ζ via Rev7. In order to test this hypothesis, we would generate a mouse encoding a truncated version of Rev1 in the C-terminus or a point mutation which impairs the ability of REV1 to interact with the REV7. Then the number of PGCs in the gonads of these mouse embryos would be quantified to ascertain if the interaction of REV1 with DNA

polymerase ζ is indeed essential for the correct development of PGCs. Provided this is the case it would be of interest to explore whether REV7 is required in PGCs as a subunit of polymerase ζ therefore extending beyond the lesion encountered in PGCs or if there is a requirement for a different role of this multifunctional factor.

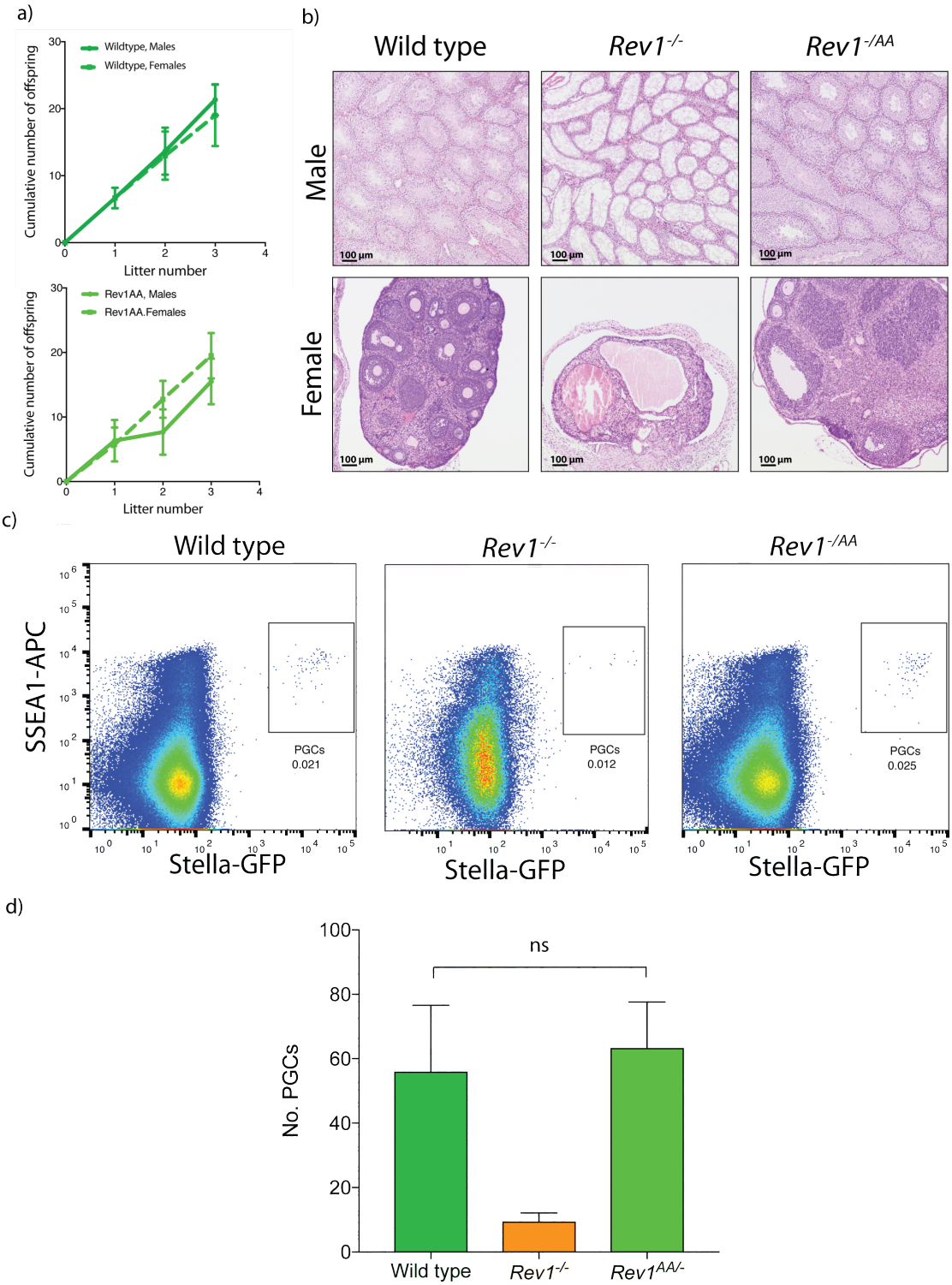


Figure 7 – The catalytic activity of REV1 is dispensable for PGC development and fertility.
a) Quantification of the accumulative number of offspring per liter when male or female Rev1AA or controls were mated with Wild type mice. The data represent the mean of three mice.
b) Haematoxylin and eosin (H&E) stained sections of ovaries and testes of 8-week old Wild type and *Rev1*$^{AA/AA}$ mice.
c) Representative flow cytometric profile of PGCs obtained from E9.5 embryos after SSEA1 staining. The frequency of PGCs refers to the frequency of GFP and SSEA1 double positive population.
d) Quantification of the number of SSEA1 and GFP positive cells (PGCs) from Wild type and *Rev1*$^{AA/-}$ embryos. The data represent the mean of three independent experiments. The statistical analysis performed was an unpaired *t*-test.

Summary

The data presented above validate and show that screening in PGCLCs is a useful system to explore new PGC biology. We validated a hit from our screen both *in vitro* and *in vivo*. This not only validates the system but also yields new insight into how the genome stability of PGCs is maintained. Using a different sgRNA as the ones used in the screen, we showed that disruption of *Rev1* in ESCs reduced the ability of cells to differentiate into PGCLCs. Then, using *Rev1*-deficient mice we found that absence of *Rev1* in adult mice caused loss of germ cells and infertility. In addition to this, we found that the sterility phenotype observed in *Rev1*-deficient mice is caused by a numerical reduction of PGCs early in development. Furthermore, we showed that the catalytic activity of REV1 is dispensable for PGC development both *in vitro* and *in vivo*. These result suggests that TLS may act in PGC development to promote mutagenesis in the mammalian germ line.

Chapter 5 - Discussion

5.1 - PGCLC system recapitulates the requirement for DNA repair *in vivo*

DNA repair has been studied for over 30 years since the initial discovery that the DNA molecule could be damaged and therefore would require some form of repair (Lindahl 1993). Most genetic studies have aimed to understand how DNA is repaired by treating DNA repair deficient cell lines with DNA damaging agents. This approach permits to study the role of a determined factor in an experimental setting which is far from resembling what happens in a physiological scenario. Our lab decided to study *Ercc1*-deficient mice and we observed that these mice exhibit a reduction in the number of PGCs during embryonic development. This provides an example of a requirement for DNA repair in a physiological system. Using an *in vitro* system, we were able to recapitulate the reduction in PGC numbers. The system we have employed therefore enables the study of DNA repair factors in a setting closer to physiology. This system has the advantages of avoiding the use of mice, permitting us to produce vast amounts of this very rare population of cells. The PGCLCs generated *in vitro* have the ability to start epigenetic reprogramming and have spermatogenic potential upon transplantation. Furthermore, this *in vitro* system permits the use of genetics to understand which DNA repair pathways are required for PGC development and permits to study genes required for development which would not be possible to study in PGCs using in a developing embryo. However, this system also has caveats, namely the extent to which the DNA demethylation process occurs *in vitro.* There is no consensus as to how far in development PGCLCs go with labs claiming this system recapitulates early PGC development (from E7.5-E9.5) whereas the DNA demethylation process is only finished by E13.5 in PGCs (Hargan-Calvopina et al. 2016). Also, the physiological context, cellular environment and dynamics observed in a developing embryo cannot be fully recapitulated using an *in vitro* system. Considering the limitations of this

system we were able to show that the genetic requirement for DNA repair in PGCs can be recapitulated using the PGCLC system. This provided us with a tool to study DNA repair in a system closer to physiology and that would permit the discovery of novel genetic factors required for PGC development.

5.2 - Identification of DNA repair factors required for the proliferation of mouse embryonic stem cells

We used an *in vitro* system to screen for novel DNA repair factors essential for PGC development. To achieve this, we compared the frequency of sgRNAs present in the EpiLCs and PGCLCs. However, this system first requires the generation of a pool of KOs in ESCs that subsequently differentiate to EpiLCs and then to PGCLCs. This permitted us to initially find factors that when depleted are required for proliferation in ESCs. First, we observed that the most depleted sgRNAs in ESCs targeted essential genes. This result allowed us to validate the screening strategy by showing that gene deletions occured in an efficient manner since genes which are essential for cell viability are depleted from the pool of ESCs. Furthermore, searching for the most enriched sgRNAs in ESCs we observed the four sgRNAs targeting *Trp53*. *Trp53* has been previously shown to be the most enriched gene identified in other CRISPR screens using ESCs (Li, Yu, et al. 2018; Hackett et al. 2018). Furthermore, it has been shown that in human pluripotent stem cells negative P53 mutations are acquired and expanded, further suggesting absence of P53 to confer a selective advantage in these cells (Merkle et al. 2017). The *Trp53* gene encodes for a transcription factor which in somatic cells responds to cellular stress signals by activating the transcription of genes that trigger cell cycle arrest or apoptosis which either promote repair or eliminate abnormal cells, respectively (Martinez et al. 1991; Yonish-Rouach et al. 1991). Interestingly, it has been suggested that in stem cells *Trp53* may have additional functions. It has been shown that in hematopoetic stem cells and neural stem cells p53 negatively regulates the proliferation and self-renewal of these cells and maintains their quiescent state (Liu et al. 2009; Meletis et al. 2006). Also, $Trp53^{-/-}$ mouse ESCs have been shown to have a proliferative advantage compared to $Trp53^{+/-}$ cells which is associated with decreased susceptibility to apoptosis (Sabapathy et al. 1997). Mouse ESCs have been shown to experience higher levels of DNA damage compared to MEFs, exemplified by the elevated levels of phospho-H2AX (histone H2A variant X) and chromatin bound RPA or RAD51 (Ahuja et al. 2016). This observation may be due to persistent replication stress caused by the high proliferative rate (Waisman et al. 2019). Nevertheless, mouse ESCs accumulate mutations at a slow rate when compared to MEFs, suggesting either a robust mechanism of repair of lesions and/or an efficient mechanism for clearing

damaged cells (West et al. 2009). Therefore, one can hypothesise that absence of *Trp53* in ESCs causes the survival of cells that otherwise would be cleared by apoptosis or would arrest in the cell cycle therefore ultimately increasing the total number of cells in culture. In order to test this we could initially generate *Trp53*-deficient mouse ESCs and record the growth advantage in comparison to Wild type cells. Having observed a growth advantage in the *Trp53*-deficient ESCs, we could then separate the functions of *Trp53* using complementation. Complementing the *Trp53*-deficient cells with Wild type and a mutant version of *Trp53* which is unable to promote cell cycle arrest but maintains apoptosis-inducing function (Toledo et al. 2006) we could assess if it is the ability to induce apoptosis or cell cycle arrest or both that rescue this phenotype.

Having observed a striking depletion of sgRNAs targeting essential genes and an enrichment for *Trp53* sgRNAs in the ESC population allowed us to conclude that the we have successfully generated a pool of KO ESCs to then differentiate into EpiLCs and PGCLCs. In addition to *Trp53*, we found *Chek2* and *Usp28* to confer a proliferative advantage to ESCs. *Chek2* is a DNA damage response cell cycle checkpoint regulator and a putative tumour suppressor. When active, *Chek2* is known to inhibit the *Cdc25c* phosphatase thereby preventing entry into mitosis (Chaturvedi et al. 1999) and also has been shown to stabilise the tumor suppressor protein *Trp53*, leading to cell cycle arrest in G1 (Dumaz et al. 2001). *Usp28* is a deubiquitinase involved in DNA damage induced apoptosis by specifically deubiquitinating proteins of the DNA damage pathway such as CLSPN (Claspin) (Zhang et al. 2006). USP28 has been shown to form a complex with 53BP1 (Tumor protein p53 binding protein 1) which stimulates *Trp53* DNA-binding activity (Cuella-Martin et al. 2016). Also, USP28 has been shown to directly deubiquitinate p53 *in vitro* and to stabilise p53 in cells (Fong et al. 2016). This therefore suggests that failure to activate *Trp53* via *Usp28* depletion causes the proliferative advantage observed in ESCs. Notably, sgRNAs targeting another kinase - *Atm* (Ataxia telangiectasia mutated) - which is upstream of *Chek2* in the phosphorylation cascade to activate TRP53, were also highly enriched in ESCs. These results further emphasize the requirement not only for *Trp53* but also of the downstream factors required for TRP53 activation to therefore mediate a robust mechanism of DNA damage repair and/or clearance of damaged cells in mouse ESCs.

The second most enriched set of sgRNAs in the ESC population corresponded to *Setd2*. *Setd2* is a histone methyltransferase that is specific for lysine-36 of histone H3, and methylation of this residue is associated with active chromatin (Sun et al.

2005). This protein also contains a novel transcriptional activation domain and has been found associated with hyperphosphorylated RNA pol II (Carvalho et al. 2013). *Setd2* has been shown to be required for endoderm differentiation (Zhang et al. 2014) but no phenotype has been observed regarding disruption of *Setd2* in mouse ESCs and proliferation. Interestingly, SETD2 is frequently mutated or deleted in several human tumours (Dalgliesh et al. 2010). In mouse models of colorectal cancer inactivation of *Setd2* promotes tumourigenesis by counteracting Wnt signaling, responsible for controlling proliferation and growth of cells, therefore facilitating self-renewal of intestinal stem/progenitor cells (Yuan et al. 2017). Therefore, one hypothesis for the selective advantage observed in *Setd2*-deficient cells could be that absence of *Setd2* disregulates Wnt signaling causes a proliferative advantage in mouse ESCs. To further support this hypothesis it has been shown that Wnt signaling is particulalry important in mouse ESC as the inhibitor of GSK3 used in the 2i media has been shown to activate the Wnt signaling pathway to promote ESC growth and viability (Ying et al. 2008). Our further steps will be to validate experimentally these results and explore what may be the basis of this effect.

Overall, we observed an enrichment for sgRNAs targeting genes involved in the activation of *Trp53* and *Trp53* itself. These results suggest that failure to activate *Trp53* in ESCs may allow cells that encounter DNA damage to progress in the cell cycle and to avoid clearance via apoptosis therefore gaining a proliferative advantage.

5.3 - Identification of DNA repair factors required for the exit of mouse embryonic stem cells from pluripotency

Then, we searched for factors required for the exit from pluripotency by comparing the frequency of sgRNAs in the ESCs with the sgRNAs in the EpiLC population. We found *Dnmt1* at the top of our gene list. *Dnmt1* is a maintenance DNA methyltransferase enzyme that transfers methyl groups to cytosine nucleotides of genomic DNA. DNMT1 is the major enzyme responsible for maintaining methylation patterns following DNA replication and shows a preference for hemi-methylated DNA (Hermann, Goyal, and Jeltsch 2004). It has been previously shown that absence of *Dnmt1* in ESCs causes apoptosis whenever ESCs are induced to generate embryoid bodies, therefore suggesting that *Dnmt1* is required for ESC differentiation (Panning and Jaenisch 1996). Furthermore, another screen identified *Dnmt1* to be required for the transition from ESC to EpiLC (Hackett et al. 2018). *Dnmt1*-deficient mouse ESCs show a global decrease in DNA methylation levels which does not affect proliferation but may have an effect on the initial ground state of the ESCs. Therefore, one potential explanation for the requirement of *Dnmt1* in ESC differentiation is that maintenance of DNA methylation by *Dnmt1* in ESCs may be crucial for propagating the changes in transcription that are necessary for differentiation. Methylation has been shown to be associated with transcriptionally silent chromatin by either impeding transcription factors to bind chromatin (Bell and Felsenfeld 2000) and recruitment of chromatin remodelling factors (Hendrich and Bird 1998) which would hence regulate transcription. Thus, in the absence of *Dnmt1* and in the presence of a hypomethylated genome, these changes cannot be performed and therefore the transcription of genes required for differentiation cannot be regulated. Also this could potentially explain why *Dnmt3a/b* did not emerge as top hit in this screen. *Dnmt3a/b*-deficiency in mouse ESCs results in only a gradual loss of methylation in genomic sequences that requires more than 70 passages to reduce 90% of global methylation (Chen et al. 2003). Furthermore, unlike *Dnmt1*-deficient mouse ESCs (Tucker et al. 1996), *Dnmt3a/b*-deficient cells are still able to give rise to teratomas suggesting that these cells can differentiate. In the future it would be interesting to explore if indeed *Dnmt1* is required to regulate the transcription of specific genes required for differentiation and to find which genes these may be. Having found *Dnmt1* as the most depleted gene in EpiLCs permitted us to

conclude that the screening strategy coupled with the differentiation assay is permitting us to find factors required for these transition states. In addition to *Dnmt1*, we found other factors, such as *Usp28*, *Rnaseh2a* and *Cdh1.* We previously identified *Usp28* have a selective advantage when disrupted in ESCs. Interestingly, this gene ranked second while comparing sgRNAs depleted in EpiLCs versus ESCs. This suggests that *Usp28* could be involved in the differentiation from ESCs to EpiLCs. However, since *Usp28*-deficient mice grow to adulthood and are able to live and reproduce (Richter et al. 2018) we should first validate this hit by differentiating *Usp28*-deficient ESCs to EpiLCs and assess if indeed we observe a reduction in EpiLC formation. If this is the case, it would be interesting to understand if the defect observed is due to activation of *Trp53* (which would be expected if *Trp53* would also came out as a hit) or if it has to do with other functions of this gene.

After *Usp28* in the gene rank list, we found *Rnaseh2a* to be depleted in the EpiLC population when comparing the sgRNAs with the ESC population. *Rnaseh2a* is a subunit of the heterotrimeric ribonuclease H enzyme (RNASEH2) (Jeong et al. 2004). RNASEH2 is the major source of ribonuclease H activity in mammalian cells and endonucleolytically cleaves ribonucleotides in genomic DNA (Hiller et al. 2012). RNASEH2 may also be important for the resolution of R-loops that arise during transcription (El Hage et al. 2010). In humans, mutations in RNASEH2 lead to Aicardi-Goutières syndrome which is a type I interferonopathy in which type I IFN production and clinical features resemble a congenital viral infection of the brain (Crow et al. 2015). Deletion of *Rnaseh2a* in mice causes embryonic lethality at E10.5 via a severe *Trp53*-mediated DNA damage response and with the first structural sign of embryonic defects occurring at the epiblast-to-gastrulation transition (Uehara et al. 2018). Since sgRNAs targeting the other subunits of the trimer were not initially present in the sgRNA library, we only observed the depletion of the subunit "a" of the trimer in EpiLCs. One explanation for the requirement of *Rnaseh2a* in EpiLCs may be that the DNA damage associated with absence of *Rnaseh2a* can be resolved in mouse ESCs but not in EpiLCs due to the differential use of DNA repair factors in these distinct cell types. Whereas ESCs would have robust DNA repair mechanisms to overcome such DNA damage it may be possible that EpiLCs induce an apoptotic response that would justify the reduction of sgRNAs targeting *Rnaseh2a* in this cell population. In addition to *Rnaseh2a*, we identified *Cdh1,* which encodes a cadherin. *Cdh1* was included in our sgRNA library since it has been shown to promote NER via positively regulating the

transcription of *Xpc* and DNA damage-binding protein 1 (*Ddb1*) (Qiang et al. 2016). CDH1 is a calcium-dependent cell-cell adhesion protein comprised of five extracellular cadherin repeats, a transmembrane region and a highly conserved cytoplasmic tail (Overduin et al. 1995). Loss of function mutations in this gene in human patients is correlated with cancer invasion and metastasis (Vleminckx et al. 1991). In addition to this, *Cdh1*-deficient mice fail to survive embryonic development showing severe abnormalities before implantation (Riethmacher, Brinkmann, and Birchmeier 1995) Interestingly, *Cdh1*-deficient mouse ESCs show impaired differentiation to embryoid bodies, which is in agreement with our screening results (Mohamet, Lea, and Ward 2010). However, the basis of this phenotype remains to be explored. In order to explore the requirement for *Cdh1* during ESC to EpiLC differentiation, one would have to generate *Cdh1*-deficient ESCs, differentiate these cells into EpiLC and assess the number of EpiLCs produced in comparison to the Wild type. If a decrease in EpiLCs is observed, we could then interrogate what is the nature of this decrease: is the absence of *Cdh1* causing cells to retain pluripotency, or to die during the differentiation protocol?

Using our screening strategy we have found potential novel factors required for the exit of pluripotency. Observing *Dnmt1* in the top of our list and knowing that absence of this factor impairs mouse ESC differentiation allows us to conlcude that by coupling the ESC differentiation system with CRISPR-screening we can find new factors required for the different cell fate transitions.

5.4 - A dropout screen identifies novel repair factors required for PGC development

Using the PGCLC differentiation system helped us to setup a CRISPR/Cas9 screening platform to find genome stability factors with novel functions in PGCs. In the past, a screen for genetic factors required in PGCs had been performed. This screen combined inhibitory RNAs with the use of ESCs carrying a *Stella-GFP* reporter which were differentiated into embryoid bodies that generated putative PGCs expressing *Stella-GFP* (West et al. 2009). Using this strategy a genetic requirement for *lin28* (lineage 28) in mouse PGCs was identified and validated for the first time. The main caveats of this screen are that only 30 genes were interrogated, which stronly limits the search for novel factors required for PGC development. Also the use of inhibitory RNAs does not fully deplete gene expression which may cause an increase in the false negative hits in the screen. In addition to this, the PGCs derived from embryoid bodies are generated via random differentiation and activation of the *Stella-GFP* reporter in contrast with the PGCLC system in which cytokines required for PGC development are used to induce PGC differentiation which more closely resembles the specification of PGCs *in vivo*. In addition to this screen, two other research labs employed the PGCLC differentiation system for screening for novel factors required for PGC development. One of the labs performed a targeted screen looking for histone modifiers in PGC fate using small interfering RNA (Mochizuki et al. 2018). This screen aimed at understanding how known factors such as PRDM14 and BLIMP1 regulate the PGC transcriptional programme via histone modifications, therefore permiting to understand in more detail the determination of epigenetic fate in PGCs. The caveats of this screen were the need of transducing each siRNA individually instead of using a pool. This therefore greatly diminished the number of genes that could be analysed. Furthermore, siRNA does not cause complete depletion of gene expression. In addition to this screen, a CRISPR/Cas9 screen has been previously employed to search for novel regulators required for PGC development (Hackett et al. 2018). This screen was performed using a genome-wide library of sgRNAs, however the authors did not succeed in employing a convenient system to produce the required number of PGCLCs to perform a valid screen. Therefore the authors were able to sort only 1.7 million PGCLCs resulting in each sgRNA to be present in a cell only 19 times on average (this is also named the

fold-representation of a library in a respective population). Since at low representation the library becomes skewed, causing erroneous interpretation of the dropout/enrichment of sgRNAs, the results of this screen become more difficult to interpret. We have upscaled the generation of PGCLCs by using the Aggrewell plate system that allows us to generate a vast amount of PGCLCs, coupled to the use of a smaller library of sgRNAs in order to more easily cover all the sgRNAs with a higher representation and therefore generating more reliable data. Using this approach we were able to find candidate genes such as *Pds5b*, which have been reported to have PGC defects when depleted from mice, but we also discovered a few novel factors. The factor we found with known PGC defects, named *Pds5b*, encodes for a protein that interacts with the conserved protein complex named cohesin. The cohesin complex holds together sister chromatids and facilitates accurate chromosome segregation during mitosis and meiosis. Additionally, mutations in cohesion proteins are associated with the developmental disorder Cornelia de Lange syndrome in humans. *Pds5b* knockout mice died shortly after birth (Zhang et al. 2007). These mice exhibited multiple congenital abnormalities, including heart defects, cleft palate, fusion of the ribs, short limbs, distal colon aganglionosis, abnormal migration and axonal projections of sympathetic neurons, and germ cell depletion, many of which are similar to abnormalities found in humans with Cornelia de Lange syndrome. Also, in vertebrates PDS5B, unlike PDS5A, contains a DNA-binding domain and localizes in the nucleolus (Zhang et al. 2007). Interestingly, the PGC defect is the only phenotype observed in *Pds5b*–deficient mice that is not observed in *Pds5a*-deficient mice (Zhang et al. 2009). This suggests that some function performed by *Pds5b* alone is essential for complete development of this population and is independent of the other phenotypes observed in *Pds5a*-deficient mice. Since cohesins have been shown to be involved in transcriptionally regulating gene expression, it would be interesting to test if in the absence of *Pds5b* in PGCs there are changes in the transcriptome that may ultimately lead to loss of this cell population.

In addition to *Pds5b*, we found a new factor with no published mouse phenotype. This factor is named *Spidr* and has been shown to play a role in HR. SPIDR has been shown to physically interact with BLM and RAD51 (Wan et al. 2013). Also, depletion of SPIDR from human cells causes an increase in sister chromatid exchange and defects in HR (Wan et al. 2013). Interestingly, two girls from a consanguineous family with mutations in the SPIDR gene have been shown to display failure to reach

puberty, primary ovarian insufficiency and very high levels of the hormones that stimulate gonad development (Smirin-Yosef et al. 2017). Such a phenotype would be expected from patients exhibiting failure to correctly develop PGCs during embryonic development. Furthermore, *Spidr*-deficient mice generated by the Jackson lab with phenotype descriptions available in the Mouse Genome Informatics online resource shows that both male and female adult mice are sterile.

Interestingly, other factors involved in HR such as *Mcm8* and *Mcm9* when depleted from mice also cause fertility problems related to PGC defects and it has been shown that *Mcm9* has a PGC numerical attrition as early as E11.5 (Luo and Schimenti 2015; Lutzmann et al. 2012). However, the requirement for *Mcm9* in PGCs has been shown to act on a different pathway as the Fanconi gene *Fancm* (Luo and Schimenti 2015). Therefore one can speculate that during PGC development there may be an additional requirement for HR factors, to deal with different sources of DNA damage and maintain the integrity of the genome. In the future it will be interesting to assess if *Spidr* is required for fertility in both sexes. Then it would be of interest to understand if this phenotype is caused by a numeric reduction of PGCs during embryonic development. Since HR is a less mutagenic form of DSB repair we can speculate that the germ line may devise this strategy to avoid mutations to be transmitted to the next generation.

5.5 - A dropout screen identifies *Rev1* as a novel factor required for PGC development

In addition to *Pds5b,* we found *Rev1* targeting sgRNAs to be depleted in the PGCLC population when compared to the same sgRNAs present in the EpiLCs. *Rev1* is a deoxycytidyl transferase enzyme that also acts as a scaffold to recruit DNA polymerases involved in TLS (Masuda et al. 2001; Pustovalova, Bezsonova, and Korzhnev 2012). *Rev1*-deficient mice have been generated previously and exhibited growth retardation and impaired class-switch recombination (Jansen et al. 2006). Interestingly, the germ cell or fertility defects observed in these mice have not been characterised until now. We have found a genetic requirement for *Rev1* in early PGC development. *Rev1*-deficient mice show a reduction in PGC numbers early in PGC development which culminates in loss of germ cells in both male and female adult mice.

Since *Rev1* and *Ercc1* both function in ICL repair one could hypothesise that these factors deal with the same DNA lesion during PGC development. However, the temporality and magnitude of the PGC numerical reduction observed in the *Rev1*-deficient embryos is distinct from the one observed in the absence of *Ercc1* (Hill, R.J. and Crossan, G.P. 2019). PGC numbers were reduced in *Ercc1*-deficient embryos at E11.5 whereas for *Rev1* the PGC defect emerged prior to E9.5. This is also observed in the histology of adult mice in which *Ercc1*-deficient mouse testis still have some tubules with germ cells whereas *Rev1*-deficient testis have tubules completely devoid of germ cells. Furthermore, unlike for *Rev1* targeting sgRNAs, we did not observe sgRNAs targeting *Ercc1* to be depleted in the PGCLC screen we performed. This may be because the PGC numerical reduction observed in *Ercc1*-deficient embryos occurs around E11.5 and the PGCLCs at day 4 resemble E10.5 PGCs, where a significant numerical reduction is not yet observed. These results therefore suggest that there may be more than one insult during germ cell development that requires distinct DNA repair pathways during different periods.

The discovery of the requirement for *Rev1* in PGCs is particularly interesting because it is known that *Rev7*-deficient mice and ESCs have reduced abilities to generate PGCs *in vivo* and *in vitro*, respectively (Pirouz, Pilarski, and Kessel 2013; Rahjouei et al. 2017; Watanabe et al. 2013). Therefore this suggests that the absence of *Rev7* impairs the development of PGCs due to its role in TLS, similarly to *Rev1*. To

corroborate this, we observed that the defect observed in PGCLCs in the absence of *Rev1* and *Rev7* is comparable. In the future, it will be interesting to test if the absence of the C-terminus region of *Rev1,* required for interacting with *Rev7,* is required for PGC development (Ross, Simpson, and Sale 2005). Also, it would be interesting to generate embryos deficient for both *Rev1* and *Rev7* to ascertain if these factors are required for the same function in PGC development.

Interestingly, from E8.5 to E9.5 there is a switch from the G2-arrest in the cell cycle to a proliferative state (Seki et al. 2007). Also, during this period the germ cells undergo an extensive epigenetic reprogramming process starting between E7.5 and E8.5. This epigenetic reprogramming may therefore pose a challenge to the genome stability for example via the release of alkylating or crosslinking agents in close proximity to DNA during histone demethylation. The DNA damage encountered during epigenetic reprogramming could therefore cause PGCs to be dependent on the TLS tolerance pathway. Interestingly, for *Ercc1* mutants the PGC defect emerges later at E11.5, which is when the second wave of DNA demethylation occurs. This may indicate that the challenge encountered early at E9.5 may only be observed later in the absence of *Ercc1* or there is a different challenge that occurs later in PGC development.

We have employed an *in vitro* system that permitted us to study the genetic requirement for DNA repair factors in PGCs. Using this system we performed a CRISPR/Cas9 screen which identified *Rev1* as a critical factor for PGCLC development. This selected hit from the screen was then validated both *in vitro* and then *in vivo* and we also showed that the catalytic activity of REV1 is dispensable for the development of PGCs. Interestingly, whilst aiming to find DNA repair factors that suppress mutagenesis in the germ line we found a factor belonging to the error-prone TLS pathway. In the future it will be interesting to explore if other TLS factos are also required during this temporal window in early PGC development. Furthermore, the observation that TLS is required early in PGC development suggests that the PGCs employ this mutagenic tolerance pathway to promote the evolution of the genome in mammals.

Chapter 6 - Materials and Methods

General techniques in molecular cloning

Molecular cloning was performed following standard protocols (Sambrook and Russel 2001). Ligation reactions were performed using high concentration T4 DNA ligases according to manufacturers instructions. Ligation reactions were transformed into chemically competent *E. coli* DH5α. DNA was prepared from bacteria using QIAprep Spin Miniprep kit (QIAgen) or QIAprep Maxiprep kit (Qiagen) following the manufacturers instructions.

Plasmids

The pLentiguide puro and px458 plasmids were used to express gene specific sgRNAs. For cloning sgRNA sequences into the pLentiguide Puro and px458 the restriction sites *Bsm*bI and *Bbs*I were used, respectively. In order to generate a doxycycline inducible Cas9 expressing vector we replaced the puromycin resistance cassette in the pCW Cas9 plasmid for the hygromycin resistance cassette in the pCW57-MCS1-2A-MCS2 (Hygro) vector using the restriction enzymes *Bam*HI and *Xba*I.

ESC culture, transfection and selection of clones

ESCs were grown in N2B27 medium complemented with PD0325901 (Axon Medchem) (MEKi), CHIR99021 (Axon Medchem) (GSK3i) and mLIF (MRC Cambridge Stem Cell Institute) – 2i + mLIF, at 37 °C in 10% CO_2. ESC were always grown in plates or dishes previously coated well with 16.6 μg/ml of fibronectin (Milipore) in PBS for 1 hour at 37 °C. Transfections were performed as follows: 0.3 x 10^6 ESC were plated the day before transfection in a well of a 6 well-plate and allowed to grow overnight at 37 °C in a 10% CO_2 incubator. Cells were transfected using Xfect™ Transfection Reagent (TaKaRa, Japan): 2 μg of DNA were mixed with the provided buffer and polymer for 10 minutes and then added to cells. After 4 hours of transfection, the media was replaced. One day after transfection, cells were transferred to a 10 cm dish previously coated fibronectin and containing 8 mL of media. The following day, when required, neomycin selection was performed by adding neomycin to a final concentration of 150 μg/ml. For puromycin, a final concentration of 0.5 μg/ml was used and for hygromycin 110 μg/ml. After 7-10 days colonies started to emerge which were handpicked and expanded into 96 well-plates.

DNA fragment analysis

Genomic DNA was extracted from cells in 96 well-plates using the DirectPCR Lysis Reagent (Viagen) with 1:100 proteinase K (Viagen): the cells were resuspended in the the lysis reagent for 45 minutes at 55 °C and subsequently for 45 minutes at 85 °C to inactivate the proteinase K. The genomic DNA was used for PCR amplification of the targeted locus with fluorescent primers tagged with FAM™ and HEX™ and an amplicon spanning 200-500 bp. The corresponding PCR products were diluted in Hi-Di™ Formamide (Invitrogen) and GeneScan™ 600 LIZ™ Dye Size Standard (Invitrogen) and analysed by capillary electrophoresis using a ABI 3130XL.

Gene	Targeted exon	Amplicon	FAM Fw primer (5' – 3')	HEX Rev primer (5' – 3')
Fan1	3	416	TAAAGACATGGGCTACCGGC	TCAGAGATCCAGCAGCCTCT
Mus81	9	300	TTCCTGCCTTCCCTGTCTCT	GTGTGGACATTGGCGAAACC
Snm1b	2	250	TGCTGAAGGAACCTGCTCTG	AGGTGAGCAAGGCACAAAGA

Table 2 – Primers used for screening of CRISPR-disrupted genome stability genes in mouse ESCs by fragment analysiss. This table shows sequence of the primers and respective amplicon size to screen for disruptions in exonic regions of *Fan1*, *Mus81* and *Snm1b.*

ESC induction into PGCLCs

In order to differentiate ESCs into PGCLCs we followed the protocol developed by Hayashi *et al.* (Hayashi et al. 2011). ESCs grown in 2i + mLIF grown in a well of a 6 well-plate were trypsinised with 0.4 ml of Tryple for 4 minutes at 37°C and washed with 1.6 ml of DMEM/F-12 (Invitrogen) supplemented with 0.15% BSA (Invitrogen). The cells were then counted and plated in a fibronectin coated 12 well-plate at 0.1 x 10^6 cells per well in 1 ml EpiLC medium (N2B27 supplemented with FGFb (MRC Cambridge Stem Cell Institute), mLIF and KnockOut Serum Replacement (Invitrogen) for 42 hours. The cells were fed after 16 hours. After 42 hours the cells were washed with 1 ml of Phosphate-buffered saline (PBS) and then trypsinised with 0.4 ml of TryPLE for 2 minutes at room temperature and washed with 1.6 ml of DMEM/F-12 (Invitrogen) supplemented with 0.15% BSA (Invitrogen). The EpiLCs were then spun down at 230 x g for 3 minutes and ressuspended in 1 ml of GK15 medium. For PGCLC differentiation in 96 well Lipidure-Coat plates EpiLC were plated at a density of 0.02

x 10^6 cells in PGCLC medium (GK15 containing BMP4 (PeproTech), SCF (R&D Systems), mLIF and mEGF (R&D Systems)). For PGCLC differentiation in Aggrewell 800 μm plates (StemCell Technologies) the plates were previously treated with 500 μl of Rinsing solution, according to manufacturer's instructions. Then the plates were spun down at 2000 x g for 5 min. The Rinsing solution was removed and the wells were washed with 2 ml of PBS. After washing, 1 ml of PGCLC media with 0.6 x 10^6 cells was added to the well, spun down at 100 x g for 3 minutes and allowed to grow at 37 °C in a 10% CO_2 incubator.

Ercc1 targeting vector assembly

Mouse genomic DNA was used as template to amplify the two homology arms for the *Ercc1* targeting vector by PCR using the LA Taq polymerase (TaKaRa, Japan). The Neo resistance cassette was amplified from the plasmid pL452. The amplification products were then purified on an agarose gel and assembled using a Gibson assembly reaction with PC1 digested with *Age*I and *Not*I. To confirm the correct assembly of the four DNA templates, Sanger sequencing was performed using the M13RP (5'-CAGGAAACAGCTATGAC-3') and TKRP (5'-TGCTTCCCTCTTGAAAACCACACTGCTCGAC-3') oligos.

sgRNA cloning

The px458 plasmid containing both Cas9 and GFP being expressed under a cytomegalovirus promoter and separated by a self-cleaving peptide (T2A) was digested with the restriction enzyme *Bbs*I. Then, the digestion product was run on an agarose gel and gel purified. Both the upper and lower oligo sequences corresponding to the sgRNA were phosphorylated with PNK and annealed by being incubated at 37°C for 1 hour followed by five minutes at 95°C and a temperature ramp from 95°C to 25°C at 0.1°C per second which permits the oligos to cool slowly allowing the annealing. The annealed sgRNAs were then ligated into the *Bbs*I digested px458. To confirm the correct assembly of the sgRNA into the px458 backbone, Sanger sequencing was performed using the U6 primer (5-GAGGGCCTATTTCCCATGATTCC–3').

Gene	sgRNA name	Oligo sequence (5' – 3')
Ercc1	Ercc1_12	CACCGATAGCATCATCGTAGCCCG
		AAACCGGGCTACGATGATGCTAT
	Ercc1_13	CACCGATAGCATCATCGTGAGCCCG
		AAACCGGGCTCACGATGATGCTAT
Fan1	Fan1_309	CACCGAACCACAGGGGTTAAGTCGG
		AAACCCGACTTAACCCCTGTGGTTC
	Fan1_310	CACCGTGCCTCCGACTTAACCCCTG
		AAACCAGGGGTTAAGTCGGAGGCAC
	Fan1_311	CACCGCTGGATTAAAATGAGTAAAC
		AAACGTTTACTCATTTTAATCCAGC
Mus81	Mus81_389	CACCGTCCAACGTGTAGCTTGCGTA
		AAACTACGCAAGCTACACGTTGGAC
	Mus81_392	CACCGGTGTAGCTTGCGTACGGTGT
		AAACACACCGTACGCAAGCTACACC
	Mus81_396	CACCGTGTAACTCTCGGAGCATTTC
		AAACGAAATGCTCCGAGAGTTACAC
Snm1b	Snm1b_703	CACCGTGGGAACTGTCGGATTAGC
		AAACTGCTAATCCGACAGTTCCCAC
	Snm1b_707	CACCGGGTTCTTCCTTCCCGACAAG
		AAACCTTGTCGGGAAGGAAGAACCC
Rev7	Rev7_367	CACCGACGTCAGCCACCACTGTGGT
		AAACACCACAGTGGTGGCTGACGTC
	Rev7_377	CACCGCCTGATTCTCTATGTGCGCG
		AAACCGCGCACATAGAGAATCAGGC
Rev1	Rev1_091	CACCGTAGCTACACACACAGCATCG
		AAACCGATGCTGTGTGTGTAGCTAC

Table 3 – sgRNA sequences used to generate CRISPR-disrupted DNA repair genes in mouse ESCs. Synthetic oligonucleotides were designed and chosen from the CRISPR Finder tool available on the Sanger website (https://www.sanger.ac.uk/htgt/wge/find_crisprs). The targeted genes, the sgRNAs and both the upper and lower oligo required for annealing each sgRNA are represented.

Colonogenic survival assay

ESCs were cultured in KOLIF medium and plated in duplicate in 6 well-plates at a concentration of 2.000, 200 and 20 cells per well. Increasing concentrations of mitomycin C were added to the medium and the cells were allowed to grow for 6 days. After 6 days, the cells were washed with 2 mL of PBS and fixed/stained with glutaraldehyde (6% v/v) and crystal violet (0.5% w/v.).

Sequencing of CRISPR deletions in ESCs

In order to determine the nature of the genetic perturbation induced in the locus of CRISPR/Cas9 targeted genes, genomic DNA was extracted from the putative clones using the Puregene protocol and the targeted genomic regions were amplified by PCR, cloned into a TOPO Zero Blunt vector (Life Technologies) and Sanger sequencing was performed.

Quantitative real-time PCR and gene expression analysis.

Total RNA was extracted from sorted cells using the Picopure kit (Life Technologies) and cDNA was synthesised using the Superscript III reverse transcriptase (Life Technologies) following the manufacturers instructions. Relative gene expression levels were normalised to *Gapdh*.

Western blotting analysis

Whole cell extracts of 2 x 10^6 ESCs were obtained using RIPA buffer and incubating the cells on ice for 15 minutes. The remaining cells were then sonicated for 10 cycles of 30 seconds using a Bioruptor. The extracts were then spun down (21.000 x g) at 4°C and the supernatant was collected. Protein concentration was determined using the DC Protein Assay (Biorad). Extracts were diluted in the same concentration for equal loading and run on a 4-12% Tris-Glycine Protein gel (Novex) for 1 hour at 200 V. Protein was transferred to a Nitrocellulose membranes of 0.2 µm pore size via wet electroblotting at 30V for 1 hour. Membranes were blocked with TBST containing 5% BSA and primarily incubated with rabbit monoclonal anti-ERCC1 antibody for 1 hour (1:100 dilution in TBST with 5% BSA (Invitrogen). Horse Radish Peroxidase (HRP) coupled anti rabbit IgG was used as a secondary antibody (1:2000 dilution in TBST). HRP signal detection was performed using ECL (GE Healthcare UK Ltd) and the membranes were exposed on Super RX films (Fujifilm).

Flow Cytometry

To obtain a single-cell suspension, the cells were collected to a 1.5 mL eppendorf and spun down at 230 x g for 10 minutes. The pellet was then incubated with ES Trypsin at 37°C for 5 minutes, trypsinised, incubated for 5 more minutes, added 2 µl of benzonase (Milipore), trypsined and incubated for 5 additional minutes. After trypsinisation, 5% foetal calf serum in PBS was added to inactivate the trypsin. The cells were then spun down at 230 x g for 10 minutes. Finally, the pellet was resuspended in 100 µl of conjugated antibody (1:100 dilution in PBS with 2.5% foetal calf serum) and allowed to stain for 10 min at room temperature. Subsequently, the cells were resuspended in 300 µl of PBS with 2,5% foetal calf serum and strained through 70 µm meshes. For PGCLC experiments, data were acquired using the Becton Dickinson LSRFortessa. For *in vivo* experiments data were acquired using the Sony Biotechnology Inc. Eclipse analyser. To sort mouse tissues and tissue culture cells the Sony Biotechnology Synergy High Speed Cell Sorter was used. Data analysis was performed using the FlowJo version v10.1r5.

Antibody	**Clone**	**Manufacturer**	**Species**	**Poly or monoclonal**	**Dilution**	**Fluorochrome**
anti-ERCC1	D-10	SantaCruz	mouse	monoclonal	1:100	None
anti-SSEA1	MC-480	BioLegend	mouse	monoclonal	1:100	Alexa Fluor 647
anti-Cas9	7A9 3A3	Cell Signaling	mouse	monoclonal	1:100	None
anti-CD117	2B8	BioLegend	rat	monoclonal	1:100	Alexa Fluor 647

Table 4 – List of antibodies. Details of the antibodies used for both Western blotting and Flow cytometry experiments.

Animal Husbandry

All animals were maintained in specific pathogen-free conditions. In individual experiments all mice were matched for age and gender. All animal experiments undertaken in this study were done so with the approval of the UK Home Office Licence 70/8325.

Mouse Genotyping

To genotype animals genomic DNA was obtained from the embryonic heads using PCR Tail Solution (Viagen Biotech). The DNA was then used for PCR using the following oligonucleotides: Ercc1F2 (5'-TGGCCTACAGTTACCCAGAACAA-3'), Ercc1R1 (5'-CTGGGGCAATTTTAGTGTCAGTG-3') and En2A (5'-GCTTCACTGAGTCTCTGGCATCTC-3') for *Ercc1* mice; and FL097 (5'-ATTGTGAGTCTCTAGCGTTTG-3'), FL098 5'-GCTGGAATTGAAATTCTAGG-3') and FL099 (5'-GCTTCCATTGCTCAGCGGTG-3') for *Rev1* mice. For genotyping *Rev1AA* embryos and mice a Taqman based assay was used.

Histology

Tissues were fixed in 10% neutral-buffered formalin for 24-36 hours and transferred to 70% ethanol. Fixed samples were embedded in paraffin and 4 μm sections cut, deparaffinised, rehydrated and stained with haematoxylin and eosin following standard methods. Images were captured using a Zeiss AxioPlan 2 microscope (Zeiss) and tissue architecture was scored blind.

Statistical analysis

The number of independent biological samples and technical repeats (n) are indicated in the figure legends. Unless otherwise stated, data are shown as mean ± standard error of the mean (s.e.m) and a student *t*-test was used to determine statistical significance. Analysis was perfomed using GraphPad Prism version 7.

Lentiviral production, titre determination and transduction

To produce lentiviruses, 1.5 x 10^6 HEK293T cells were plated in a p60 dish containing 5 ml of HEK293T media. The following day the cells were transfected with the lentiviral packaging and envelope plasmids (0.8 μg psPAX2 and 0.2 μg of pMD2.G, respectively) and 1 μg of plentiguide Puro plasmids containing the library of sgRNAs using polyethylenimine (PEI) in DMEM (Invitrogen). After 16 hours the media was replaced by 2i + mLIF. After 48 hours of transfection the media containing the viruses was filtered through a 0.45 μm filter and collected and stored at -80°C.

In order to determine the titre of the lentiviral particles obtained from the pooled library of sgRNAs in transducing units per ml (TU/ml) we performed a cologenic survival assay based on puromycin resistance. In order to achieve this, we first coated seven 10 cm dishes with 6 ml of 16.6 μg/ml of fibronectin in PBS for 1 hour at 37°C. Then, the lentiviral particles (stored at -80°C) were thawed at room temperature. During this period, the mouse ESCs previously grown in 10 cm dishes we were trypsinised with 5 ml of Tryple for 4 minutes at 37°C and washed with 20 ml of DMEM/F-12 (Invitrogen) supplemented with 0.15% BSA (Invitrogen). Then we performed 10-fold serial dilutions of the lentiviruses starting from 10^{-2} to 10^{-6} dilutions in 20 ml of 2i+mLIF media. The viable cells were then counted using a Beckman Coulter Vi-CELL™ Cell Viability Analyzer. 1.2 x 10^6 ESCs per condition were spun down at 230 x g for 10 minutes and resuspended with the respective lentivirus concentration in the presence of 8 μg/ml of polybrene (Milipore) in 8 ml of 2i+mLIF media. In addition to the 5 dilutions of lentiviruses 1.2 x 10^6 ESCs (from 10^{-2} to 10^{-6}) two other sets of ESCs were also resuspended in 2i+mLIF media containing no lentiviruses. There uninfected ESCs would allow us to first check that all ESCs would die in the presence of puromycin, whereas the other set would be used in the absence of both lentiviruses and

puromycin to check for plating efficiency. After mixing the ESCs with the lentiviruses and polybrene, fibronectin was removed from the 10 cm dishes and the ESCs were plated and allowed to grow in the 37°C incubator for 16 hours. Then, the media was replaced with 2i+mLIF media to remove remaining lentiviral particles and polybrene. The following day the media was replaced by adding 2i+mLIF supplemented with 0.5 µg/ml of puromycin (Invitrogen). After two days of puromycin treatment, the cells were replaced again with 2i+mLIF supplemented with 0.5 µg/ml of puromycin and after four days the media was finally replaced without supplementation of puromycin. One to two weeks after, colonies started to emerge and were counted using a light microscope. Finally, we obtained the transducing units per ml (TU/ml) by multiplying the number of colonies per ml of media with the dilution factor.

CRISPR/Cas9 Screening

45 x 10^6 iCas9 #8 ESCs were transduced with the lentiviral library at a MOI of 0.3 and 0.5 µg/ml of puromycin was added to the medium 2 days after to select for transductants. Selection was continued for 4 days post transduction. After selection of cells that stably expressed the sgRNAs, puromycin was removed from the media and the cells were grown for 21 days in 2.5 µg/ml of doxycycline to induce Cas9 expression and generate a pool of CRISPR knockouts. Subsequently, ESCs were induced to differentiate into EpiLCs in 10 cm dishes, fed after 16 hours and 42 hours after the EpiLCs were induced to differentiate into PGCLCs in Aggrewell 800 µm plates (StemCell Technologies). After 4 days of differentiation, the PGCLC aggregates were collected as follows: an entire plate was resuspended to a 50 ml falcon. Washed once with PBS by centrifuging at 230 x g for 5 minutes. Then the cell pellet was resuspended in 500 µl of ES Trypsin containing 3.3 µl of benzonase (Milipore) for 1 minute and 15 seconds at 37 °C. Subsequently, the cells were resuspended and incubated at 37°C for an additional 1 minute and 15 seconds. The cell suspension was finally resuspended in PBS containing 5% of Fetal Calf Serum, stained and sorted based on the *Stella-GFP* and SSEA1-APC signal using the Sony Biotechnology Synergy High Speed Cell Sorter. Sample cell pellets were frozen in every cell type and differentiation step. A library coverage of ≥ 2000 cells per sgRNA was maintained at every step. gDNA from cell pellets was isolated using the DNeasy Blood & Tissue kit (Qiagen) and genome-

integrated sgRNAs were PCR amplified using the Herculase II Fusion DNA Polymerase (Agilent). Multiplexing barcodes were added in a second round of PCR to distinguish between different cell type populations. The concentration of gel-purified PCR products was determined using a KAPA Library Quantication Kit (KAPABIOSYSTEMS). The PCR poducts were then used for sequencing on Illumina HiSeq4000 to determine sgRNA representation in each population sample.

Screen statistical analysis

First, in order to assign the sequencing reads to the different populations of the screen (plasmid pool, ESC, EpiLC and PGCLC) a demultiplexing step took the raw NGS data and assigned it to the different cell populations corresponding to the different barcodes used in the PCR amplification step. Then an algorithm is used to assign to every sequencing read the respective sgRNA and to count the number of each sgRNA targeting a specific gene in the different populations. Having the sgRNA count we used R to perform the analysis and graphs shown in Chapter 3 as described in Winter et al. 2016. The data for each population were normalized by quantile and the Z-scores for every gene were calculated. The Z-score represents the number of standard deviations from the mean a data point is. Only the value of three sgRNAs of a total of four was used to obtain the Z-score value of each gene, in order to avoid the effect of sgRNAs with low effect that were close to the mean. These data were represented as a list of genes with highest depletion in determined population and also in a graph showing the 4 sgRNAs targeting a specific gene and what is the distance between these sgRNAs and the average sgRNA in two populations.

General Reagents

DNA loading dye

50% glycerol
1x TBE
1% bromophenol blue (w/v)
1% xylene cyanol

RIPA buffer

20 mM Tris-HCl (pH 7.5)
150 mM NaCl
1 mM Na2EDTA
1 mM EGTA
1% NP-40
1% sodium deoxycholate
2.5 mM sodium pyrophosphate
1 mM b-glycerophosphate
1 mM Na3VO4
1 μg/ml leupeptin.

5x SDS-PAGE reducing sample buffer

300 mM Tris-HCl pH 6.8
10% SDS
25% glycerol
0.05% bromophenol blue
10% β-mercaptoethanol

Western blot transfer buffer

25 mM Tris
192 mM glycine
20% ethanol

2xTY

1.6% bactopeptone
1% yeast extract
0.5% NaCl

N2B27 medium

N2B27 medium is a basal medium used for both maintenance of ESCs and differentiation of EpiLCs. Preparation of this medium is based on Hayashi *et al.* (2013) instructions (Hayashi and Saitou 2013). Prepare each component, DMEM/F-12 + N2 and Neurobasal + B27, as follows:

DMEM/F-12 + N2

N2

3.6 ml DMEM/F-12
0.5 ml apo-transferrin (stock solution at 100 mg/ml) (Sigma-Aldrich)
0.33 ml 7.5% (wt/vol) BSA
16.5 µl progesterone (stock solution at 0.6 mg/ml) (Sigma-Aldrich)
50 µl putrescine (stock solution at at 160 mg/ml) (Sigma-Aldrich)
5 µl sodium selenite (stock solution at 7.5 nM) (Sigma-Aldrich)

495 ml of DMEM/F-12 (Invitrogen)
0.5 ml Insulin (stock solution at 25mg/ml) (Sigma-Aldrich)

Neurobasal + B27

480 ml Neurobasal (Invitrogen)
10 ml of B27 supplement minus vitamin A (Invitrogen)
5 ml penicillin-streptomycin
5 ml of l-Glutamine (stock solution at 200 mM) (Invitrogen)

For 1l of N2B27 medium:

mix 500 ml of DMEM/F-12 + N2 with 500 ml of Neurobasal + B27
add 1.8 ml of β-mercaptoethanol (stock solution at 50mM)

ES trypsin

Trypsin (Difco) 2.5g/L
EDTA 0.4g/L
NaCl 7g/L
KH2PO4 0.24g/L
KCl 0.37g/L
D-Glucose< 1g/L
Tris 3g/L
Phenol Red 1ml/L
Adjust pH to 7.6

Chapter 7 - References

Abe, K., C. Naruse, T. Kato, T. Nishiuchi, M. Saitou, and M. Asano. 2011. 'Loss of heterochromatin protein 1 gamma reduces the number of primordial germ cells via impaired cell cycle progression in mice', *Biol Reprod*, 85: 1013-24.

Agoulnik, A. I., B. Lu, Q. Zhu, C. Truong, M. T. Ty, N. Arango, K. K. Chada, and C. E. Bishop. 2002. 'A novel gene, Pog, is necessary for primordial germ cell proliferation in the mouse and underlies the germ cell deficient mutation, gcd', *Hum Mol Genet*, 11: 3047-53.

Ahuja, A. K., K. Jodkowska, F. Teloni, A. H. Bizard, R. Zellweger, R. Herrador, S. Ortega, I. D. Hickson, M. Altmeyer, J. Mendez, and M. Lopes. 2016. 'A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells', *Nat Commun*, 7: 10660.

Akhter, S., Y. C. Lam, S. Chang, and R. J. Legerski. 2010. 'The telomeric protein SNM1B/Apollo is required for normal cell proliferation and embryonic development', *Aging Cell*, 9: 1047-56.

Al-Minawi, A. Z., N. Saleh-Gohari, and T. Helleday. 2008. 'The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells', *Nucleic Acids Res*, 36: 1-9.

Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A. L. Borresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjord, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D. T. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lopez-Otin, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. Tutt, R. Valdes-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, Initiative Australian Pancreatic Cancer Genome, Icgc Breast Cancer Consortium, Icgc Mmml- Seq Consortium, Icgc PedBrain, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, and M. R. Stratton. 2013. 'Signatures of mutational processes in human cancer', *Nature*, 500: 415-21.

Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, P. J. Campbell, and M. R. Stratton. 2013. 'Deciphering signatures of mutational processes operative in human cancer', *Cell Rep*, 3: 246-59.

Alexandrov, L. B., and M. R. Stratton. 2014. 'Mutational signatures: the patterns of somatic mutations hidden in cancer genomes', *Curr Opin Genet Dev*, 24: 52-60.

Antoniou, A., P. D. Pharoah, S. Narod, H. A. Risch, J. E. Eyfjord, J. L. Hopper, N. Loman, H. Olsson, O. Johannsson, A. Borg, B. Pasini, P. Radice, S.

Manoukian, D. M. Eccles, N. Tang, E. Olah, H. Anton-Culver, E. Warner, J. Lubinski, J. Gronwald, B. Gorski, H. Tulinius, S. Thorlacius, H. Eerola, H. Nevanlinna, K. Syrjakoski, O. P. Kallioniemi, D. Thompson, C. Evans, J. Peto, F. Lalloo, D. G. Evans, and D. F. Easton. 2003. 'Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies', *Am J Hum Genet*, 72: 1117-30.

Auerbach, A. D. 1988. 'A test for Fanconi's anemia', *Blood*, 72: 366-7.

Auman, H. J., T. Nottoli, O. Lakiza, Q. Winger, S. Donaldson, and T. Williams. 2002. 'Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development', *Development*, 129: 2733-47.

Bae, J. B., S. S. Mukhopadhyay, L. Liu, N. Zhang, J. Tan, S. Akhter, X. Liu, X. Shen, L. Li, and R. J. Legerski. 2008. 'Snm1B/Apollo mediates replication fork collapse and S Phase checkpoint activation in response to DNA interstrand cross-links', *Oncogene*, 27: 5045-56.

Baker, D. J., G. Wuenschell, L. Xia, J. Termini, S. E. Bates, A. D. Riggs, and T. R. O'Connor. 2007. 'Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells', *J Biol Chem*, 282: 22592-604.

Bardwell, A. J., L. Bardwell, A. E. Tomkinson, and E. C. Friedberg. 1994. 'Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease', *Science*, 265: 2082-5.

Barnes, D. E., G. Stamp, I. Rosewell, A. Denzel, and T. Lindahl. 1998. 'Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice', *Curr Biol*, 8: 1395-8.

Battaglia, D. E., P. Goodwin, N. A. Klein, and M. R. Soules. 1996. 'Influence of maternal age on meiotic spindle assembly in oocytes from naturally cycling women', *Hum Reprod*, 11: 2217-22.

Bell, A. C., and G. Felsenfeld. 2000. 'Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene', *Nature*, 405: 482-5.

Belyi, V. A., P. Ak, E. Markert, H. Wang, W. Hu, A. Puzio-Kuter, and A. J. Levine. 2010. 'The origins and evolution of the p53 family of genes', *Cold Spring Harb Perspect Biol*, 2: a001198.

Bi, X., L. R. Barkley, D. M. Slater, S. Tateishi, M. Yamaizumi, H. Ohmori, and C. Vaziri. 2006. 'Rad18 regulates DNA polymerase kappa and is required for recovery from S-phase checkpoint-mediated arrest', *Mol Cell Biol*, 26: 3527-40.

Bluteau, D., J. Masliah-Planchon, C. Clairmont, A. Rousseau, R. Ceccaldi, C. Dubois d'Enghien, O. Bluteau, W. Cuccuini, S. Gachet, R. Peffault de Latour, T. Leblanc, G. Socie, A. Baruchel, D. Stoppa-Lyonnet, A. D. D'Andrea, and J. Soulier. 2016. 'Biallelic inactivation of REV7 is associated with Fanconi anemia', *J Clin Invest*, 126: 3580-4.

Boddy, M. N., P. H. Gaillard, W. H. McDonald, P. Shanahan, J. R. Yates, 3rd, and P. Russell. 2001. 'Mus81-Eme1 are essential components of a Holliday junction resolvase', *Cell*, 107: 537-48.

Boersma, V., N. Moatti, S. Segura-Bayona, M. H. Peuscher, J. van der Torre, B. A. Wevers, A. Orthwein, D. Durocher, and J. J. L. Jacobs. 2015. 'MAD2L2 controls DNA repair at telomeres and DNA breaks by inhibiting 5' end resection', *Nature*, 521: 537-40.

Bolcun-Filas, E., V. D. Rinaldi, M. E. White, and J. C. Schimenti. 2014. 'Reversal of female infertility by Chk2 ablation reveals the oocyte DNA damage checkpoint pathway', *Science*, 343: 533-6.
Borum, K. 1967. 'Oogenesis in the mouse. A study of the origin of the mature ova', *Exp Cell Res*, 45: 39-47.
Bostick, M., J. K. Kim, P. O. Esteve, A. Clark, S. Pradhan, and S. E. Jacobsen. 2007. 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells', *Science*, 317: 1760-4.
Brass, A. L., I. C. Huang, Y. Benita, S. P. John, M. N. Krishnan, E. M. Feeley, B. J. Ryan, J. L. Weyer, L. van der Weyden, E. Fikrig, D. J. Adams, R. J. Xavier, M. Farzan, and S. J. Elledge. 2009. 'The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus', *Cell*, 139: 1243-54.
Caldecott, K. W. 2008. 'Single-strand break repair and genetic disease', *Nat Rev Genet*, 9: 619-31.
Carvalho, S., A. C. Raposo, F. B. Martins, A. R. Grosso, S. C. Sridhara, J. Rino, M. Carmo-Fonseca, and S. F. de Almeida. 2013. 'Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription', *Nucleic Acids Res*, 41: 2881-93.
Cattell, E., B. Sengerova, and P. J. McHugh. 2010. 'The SNM1/Pso2 family of ICL repair nucleases: from yeast to man', *Environ Mol Mutagen*, 51: 635-45.
Cedar, H., and Y. Bergman. 2009. 'Linking DNA methylation and histone modification: patterns and paradigms', *Nat Rev Genet*, 10: 295-304.
Chang, H. H. Y., N. R. Pannunzio, N. Adachi, and M. R. Lieber. 2017. 'Non-homologous DNA end joining and alternative pathways to double-strand break repair', *Nat Rev Mol Cell Biol*, 18: 495-506.
Chaturvedi, P., W. K. Eng, Y. Zhu, M. R. Mattern, R. Mishra, M. R. Hurle, X. Zhang, R. S. Annan, Q. Lu, L. F. Faucette, G. F. Scott, X. Li, S. A. Carr, R. K. Johnson, J. D. Winkler, and B. B. Zhou. 1999. 'Mammalian Chk2 is a downstream effector of the ATM-dependent DNA damage checkpoint pathway', *Oncogene*, 18: 4047-54.
Chen, T., Y. Ueda, J. E. Dodge, Z. Wang, and E. Li. 2003. 'Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b', *Mol Cell Biol*, 23: 5594-605.
Chiquoine, A. D. 1954. 'The identification, origin, and migration of the primordial germ cells in the mouse embryo', *Anat Rec*, 118: 135-46.
Cleaver, J. E., V. Afzal, L. Feeney, M. McDowell, W. Sadinski, J. P. Volpe, D. B. Busch, D. M. Coleman, D. W. Ziffer, Y. Yu, H. Nagasawa, and J. B. Little. 1999. 'Increased ultraviolet sensitivity and chromosomal instability related to P53 function in the xeroderma pigmentosum variant', *Cancer Res*, 59: 1102-8.
Cockayne, E. A. 1936. 'Dwarfism with retinal atrophy and deafness', *Arch Dis Child*, 11: 1-8.
Cooper, D. L., R. S. Lahue, and P. Modrich. 1993. 'Methyl-directed mismatch repair is bidirectional', *J Biol Chem*, 268: 11823-9.
Cortazar, D., C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schuermann, A. L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird, and P. Schar. 2011. 'Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability', *Nature*, 470: 419-23.
Cortellino, S., J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F.

Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, and A. Bellacosa. 2011. 'Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair', *Cell*, 146: 67-79.

Cox, E. C., and D. L. Horner. 1982. 'Dominant mutators in Escherichia coli', *Genetics*, 100: 7-18.

Crow, Y. J., D. S. Chase, J. Lowenstein Schmidt, M. Szynkiewicz, G. M. Forte, H. L. Gornall, A. Oojageer, B. Anderson, A. Pizzino, G. Helman, M. S. Abdel-Hamid, G. M. Abdel-Salam, S. Ackroyd, A. Aeby, G. Agosta, C. Albin, S. Allon-Shalev, M. Arellano, G. Ariaudo, V. Aswani, R. Babul-Hirji, E. M. Baildam, N. Bahi-Buisson, K. M. Bailey, C. Barnerias, M. Barth, R. Battini, M. W. Beresford, G. Bernard, M. Bianchi, T. Billette de Villemeur, E. M. Blair, M. Bloom, A. B. Burlina, M. L. Carpanelli, D. R. Carvalho, M. Castro-Gago, A. Cavallini, C. Cereda, K. E. Chandler, D. A. Chitayat, A. E. Collins, C. Sierra Corcoles, N. J. Cordeiro, G. Crichiutti, L. Dabydeen, R. C. Dale, S. D'Arrigo, C. G. De Goede, C. De Laet, L. M. De Waele, I. Denzler, I. Desguerre, K. Devriendt, M. Di Rocco, M. C. Fahey, E. Fazzi, C. D. Ferrie, A. Figueiredo, B. Gener, C. Goizet, N. R. Gowrinathan, K. Gowrishankar, D. Hanrahan, B. Isidor, B. Kara, N. Khan, M. D. King, E. P. Kirk, R. Kumar, L. Lagae, P. Landrieu, H. Lauffer, V. Laugel, R. La Piana, M. J. Lim, J. P. Lin, T. Linnankivi, M. T. Mackay, D. R. Marom, C. Marques Lourenco, S. A. McKee, I. Moroni, J. E. Morton, M. L. Moutard, K. Murray, R. Nabbout, S. Nampoothiri, N. Nunez-Enamorado, P. J. Oades, I. Olivieri, J. R. Ostergaard, B. Perez-Duenas, J. S. Prendiville, V. Ramesh, M. Rasmussen, L. Regal, F. Ricci, M. Rio, D. Rodriguez, A. Roubertie, E. Salvatici, K. A. Segers, G. P. Sinha, D. Soler, R. Spiegel, T. I. Stodberg, R. Straussberg, K. J. Swoboda, M. Suri, U. Tacke, T. Y. Tan, J. te Water Naude, K. Wee Teik, M. M. Thomas, M. Till, D. Tonduti, E. M. Valente, R. N. Van Coster, M. S. van der Knaap, G. Vassallo, R. Vijzelaar, J. Vogt, G. B. Wallace, E. Wassmer, H. J. Webb, W. P. Whitehouse, R. N. Whitney, M. S. Zaki, S. M. Zuberi, J. H. Livingston, F. Rozenberg, P. Lebon, A. Vanderver, S. Orcesi, and G. I. Rice. 2015. 'Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1', *Am J Med Genet A*, 167A: 296-312.

Cuella-Martin, R., C. Oliveira, H. E. Lockstone, S. Snellenberg, N. Grolmusova, and J. R. Chapman. 2016. '53BP1 Integrates DNA Repair and p53-Dependent Cell Fate Decisions via Distinct Mechanisms', *Mol Cell*, 64: 51-64.

Dalgliesh, G. L., K. Furge, C. Greenman, L. Chen, G. Bignell, A. Butler, H. Davies, S. Edkins, C. Hardy, C. Latimer, J. Teague, J. Andrews, S. Barthorpe, D. Beare, G. Buck, P. J. Campbell, S. Forbes, M. Jia, D. Jones, H. Knott, C. Y. Kok, K. W. Lau, C. Leroy, M. L. Lin, D. J. McBride, M. Maddison, S. Maguire, K. McLay, A. Menzies, T. Mironenko, L. Mulderrig, L. Mudie, S. O'Meara, E. Pleasance, A. Rajasingham, R. Shepherd, R. Smith, L. Stebbings, P. Stephens, G. Tang, P. S. Tarpey, K. Turrell, K. J. Dykema, S. K. Khoo, D. Petillo, B. Wondergem, J. Anema, R. J. Kahnoski, B. T. Teh, M. R. Stratton, and P. A. Futreal. 2010. 'Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes', *Nature*, 463: 360-3.

de Graaf, B., A. Clore, and A. K. McCullough. 2009. 'Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA-protein crosslinks', *DNA Repair (Amst)*, 8: 1207-14.

De Muyt, A., L. Jessop, E. Kolar, A. Sourirajan, J. Chen, Y. Dayani, and M. Lichten. 2012. 'BLM helicase ortholog Sgs1 is a central regulator of meiotic recombination intermediate metabolism', *Mol Cell*, 46: 43-53.

de Vries, A., C. T. van Oostrom, F. M. Hofhuis, P. M. Dortant, R. J. Berg, F. R. de Gruijl, P. W. Wester, C. F. van Kreijl, P. J. Capel, H. van Steeg, and S. J. Verbeek. 1995. 'Increased susceptibility to ultraviolet-B and carcinogens of mice lacking the DNA excision repair gene XPA', *Nature*, 377: 169-73.

Demuth, I., M. Digweed, and P. Concannon. 2004. 'Human SNM1B is required for normal cellular response to both DNA interstrand crosslink-inducing agents and ionizing radiation', *Oncogene*, 23: 8611-8.

Dendouga, N., H. Gao, D. Moechars, M. Janicot, J. Vialard, and C. H. McGowan. 2005. 'Disruption of murine Mus81 increases genomic instability and DNA damage sensitivity but does not promote tumorigenesis', *Mol Cell Biol*, 25: 7569-79.

Di Leonardo, A., S. P. Linke, K. Clarkin, and G. M. Wahl. 1994. 'DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts', *Genes Dev*, 8: 2540-51.

Doench, J. G., N. Fusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, H. W. Virgin, J. Listgarten, and D. E. Root. 2016. 'Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9', *Nat Biotechnol*, 34: 184-91.

Drost, J. B., and W. R. Lee. 1995. 'Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human', *Environ Mol Mutagen*, 25 Suppl 26: 48-64.

Dudley, A. M., D. M. Janse, A. Tanay, R. Shamir, and G. M. Church. 2005. 'A global view of pleiotropy and phenotypically derived gene function in yeast', *Mol Syst Biol*, 1: 2005 0001.

Dumaz, N., D. M. Milne, L. J. Jardine, and D. W. Meek. 2001. 'Critical roles for the serine 20, but not the serine 15, phosphorylation site and for the polyproline domain in regulating p53 turnover', *Biochem J*, 359: 459-64.

Duxin, J. P., J. M. Dewar, H. Yardimci, and J. C. Walter. 2014. 'Repair of a DNA-protein crosslink by replication-coupled proteolysis', *Cell*, 159: 346-57.

Dycaico, M. J., G. S. Provost, P. L. Kretz, S. L. Ransom, J. C. Moores, and J. M. Short. 1994. 'The use of shuttle vectors for mutation analysis in transgenic mice and rats', *Mutat Res*, 307: 461-78.

Edelmann, W., K. Yang, A. Umar, J. Heyer, K. Lau, K. Fan, W. Liedtke, P. E. Cohen, M. F. Kane, J. R. Lipford, N. Yu, G. F. Crouse, J. W. Pollard, T. Kunkel, M. Lipkin, R. Kolodner, and R. Kucherlapati. 1997. 'Mutation in the mismatch repair gene Msh6 causes cancer susceptibility', *Cell*, 91: 467-77.

El Hage, A., S. L. French, A. L. Beyer, and D. Tollervey. 2010. 'Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis', *Genes Dev*, 24: 1546-58.

Evans, E., J. Fellows, A. Coffer, and R. D. Wood. 1997. 'Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein', *EMBO J*, 16: 625-38.

Fong, C. S., G. Mazo, T. Das, J. Goodman, M. Kim, B. P. O'Rourke, D. Izquierdo, and M. F. Tsou. 2016. '53BP1 and USP28 mediate p53-dependent cell cycle arrest in response to centrosome loss and prolonged mitosis', *Elife*, 5.

Fortini, P., B. Pascucci, E. Parlanti, R. W. Sobol, S. H. Wilson, and E. Dogliotti. 1998. 'Different DNA polymerases are involved in the short- and long-patch base excision repair in mammalian cells', *Biochemistry*, 37: 3575-80.

Fouquet, B., P. Pawlikowska, S. Caburet, C. Guigon, M. Makinen, L. Tanner, M. Hietala, K. Urbanska, L. Bellutti, B. Legois, B. Bessieres, A. Gougeon, A. Benachi, G. Livera, F. Rosselli, R. A. Veitia, and M. Misrahi. 2017. 'A homozygous FANCM mutation underlies a familial case of non-syndromic primary ovarian insufficiency', *Elife*, 6.

Fousteri, M., W. Vermeulen, A. A. van Zeeland, and L. H. Mullenders. 2006. 'Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo', *Mol Cell*, 23: 471-82.

Frank, K. M., N. E. Sharpless, Y. Gao, J. M. Sekiguchi, D. O. Ferguson, C. Zhu, J. P. Manis, J. Horner, R. A. DePinho, and F. W. Alt. 2000. 'DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway', *Mol Cell*, 5: 993-1002.

Fujiwara, T., N. R. Dunn, and B. L. Hogan. 2001. 'Bone morphogenetic protein 4 in the extraembryonic mesoderm is required for allantois development and the localization and survival of primordial germ cells in the mouse', *Proc Natl Acad Sci U S A*, 98: 13739-44.

Garg, P., and P. M. Burgers. 2005. 'Ubiquitinated proliferating cell nuclear antigen activates translesion DNA polymerases eta and REV1', *Proc Natl Acad Sci U S A*, 102: 18361-6.

German, J., S. Schonberg, S. Caskie, D. Warburton, C. Falk, and J. H. Ray. 1987. 'A test for Fanconi's anemia', *Blood*, 69: 1637-41.

Ginsburg, M., M. H. Snow, and A. McLaren. 1990. 'Primordial germ cells in the mouse embryo during gastrulation', *Development*, 110: 521-8.

Glassner, B. J., L. J. Rasmussen, M. T. Najarian, L. M. Posnick, and L. D. Samson. 1998. 'Generation of a strong mutator phenotype in yeast by imbalanced base excision repair', *Proc Natl Acad Sci U S A*, 95: 9997-10002.

Grabole, N., J. Tischler, J. A. Hackett, S. Kim, F. Tang, H. G. Leitch, E. Magnusdottir, and M. A. Surani. 2013. 'Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation', *EMBO Rep*, 14: 629-37.

Guibert, S., T. Forne, and M. Weber. 2012. 'Global profiling of DNA methylation erasure in mouse primordial germ cells', *Genome Res*, 22: 633-41.

Guo, C., P. L. Fischhaber, M. J. Luk-Paszyc, Y. Masuda, J. Zhou, K. Kamiya, C. Kisker, and E. C. Friedberg. 2003. 'Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis', *EMBO J*, 22: 6621-30.

Guo, C., E. Sonoda, T. S. Tang, J. L. Parker, A. B. Bielen, S. Takeda, H. D. Ulrich, and E. C. Friedberg. 2006. 'REV1 protein interacts with PCNA: significance of the REV1 BRCT domain in vitro and in vivo', *Mol Cell*, 23: 265-71.

Guo, F., X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T. P. Gu, B. Hu, C. P. Walsh, J. Li, F. Tang, and G. L. Xu. 2014. 'Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote', *Cell Stem Cell*, 15: 447-59.

Gurdon, J. B. 1962. 'The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles', *J Embryol Exp Morphol*, 10: 622-40.

Hackett, J. A., Y. Huang, U. Gunesdogan, K. A. Gretarsson, T. Kobayashi, and M. A. Surani. 2018. 'Tracing the transitions from pluripotency to germ cell fate with CRISPR screening', *Nat Commun*, 9: 4292.
Hackett, J. A., R. Sengupta, J. J. Zylicz, K. Murakami, C. Lee, T. A. Down, and M. A. Surani. 2013. 'Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine', *Science*, 339: 448-52.
Hajkova, P., K. Ancelin, T. Waldmann, N. Lacoste, U. C. Lange, F. Cesari, C. Lee, G. Almouzni, R. Schneider, and M. A. Surani. 2008. 'Chromatin dynamics during epigenetic reprogramming in the mouse germ line', *Nature*, 452: 877-81.
Hajkova, P., S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, and M. A. Surani. 2010. 'Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway', *Science*, 329: 78-82.
Haldane, J. B. 1947. 'The mutation rate of the gene for haemophilia, and its segregation ratios in males and females', *Ann Eugen*, 13: 262-71.
Hargan-Calvopina, J., S. Taylor, H. Cook, Z. Hu, S. A. Lee, M. R. Yen, Y. S. Chiang, P. Y. Chen, and A. T. Clark. 2016. 'Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation', *Dev Cell*, 39: 75-86.
Hartman, T., K. Stead, D. Koshland, and V. Guacci. 2000. 'Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in Saccharomyces cerevisiae', *J Cell Biol*, 151: 613-26.
Hassold, T., and P. Hunt. 2001. 'To err (meiotically) is human: the genesis of human aneuploidy', *Nat Rev Genet*, 2: 280-91.
Hayashi, K., H. Ohta, K. Kurimoto, S. Aramaki, and M. Saitou. 2011. 'Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells', *Cell*, 146: 519-32.
Hayashi, K., and M. Saitou. 2013. 'Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells', *Nat Protoc*, 8: 1513-24.
Hayashi, S., P. Lewis, L. Pevny, and A. P. McMahon. 2002. 'Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain', *Mech Dev*, 119 Suppl 1: S97-S101.
Hendrich, B., and A. Bird. 1998. 'Identification and characterization of a family of mammalian methyl-CpG binding proteins', *Mol Cell Biol*, 18: 6538-47.
Herman, R. K., and N. B. Dworkin. 1971. 'Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli', *J Bacteriol*, 106: 543-50.
Hermann, A., R. Goyal, and A. Jeltsch. 2004. 'The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites', *J Biol Chem*, 279: 48350-9.
Hicks, J. K., C. L. Chute, M. T. Paulsen, R. L. Ragland, N. G. Howlett, Q. Gueranger, T. W. Glover, and C. E. Canman. 2010. 'Differential roles for DNA polymerases eta, zeta, and REV1 in lesion bypass of intrastrand versus interstrand DNA cross-links', *Mol Cell Biol*, 30: 1217-30.
Hill, K. A., A. Halangoda, P. W. Heinmoeller, K. Gonzalez, C. Chitaphan, J. Longmate, W. A. Scaringe, J. C. Wang, and S. S. Sommer. 2005. 'Tissue-specific time courses of spontaneous mutation frequency and deviations in mutation pattern are observed in middle to late adulthood in Big Blue mice', *Environ Mol Mutagen*, 45: 442-54.
Hill, P. W. S., H. G. Leitch, C. E. Requena, Z. Sun, R. Amouroux, M. Roman-Trufero, M. Borkowska, J. Terragni, R. Vaisvila, S. Linnett, H. Bagci, G. Dharmalingham, V. Haberle, B. Lenhard, Y. Zheng, S. Pradhan, and P.

Hajkova. 2018. 'Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte', *Nature*, 555: 392-96.
Hiller, B., M. Achleitner, S. Glage, R. Naumann, R. Behrendt, and A. Roers. 2012. 'Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity', *J Exp Med*, 209: 1419-26.
Hilscher, B., W. Hilscher, B. Bulthoff-Ohnolz, U. Kramer, A. Birke, H. Pelzer, and G. Gauss. 1974. 'Kinetics of gametogenesis. I. Comparative histological and autoradiographic studies of oocytes and transitional prospermatogonia during oogenesis and prespermatogenesis', *Cell Tissue Res*, 154: 443-70.
Ho, C. K., G. Mazon, A. F. Lam, and L. S. Symington. 2010. 'Mus81 and Yen1 promote reciprocal exchange during mitotic recombination to maintain genome integrity in budding yeast', *Mol Cell*, 40: 988-1000.
Hodskinson, M. R., J. Silhan, G. P. Crossan, J. I. Garaycoechea, S. Mukherjee, C. M. Johnson, O. D. Scharer, and K. J. Patel. 2014. 'Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair', *Mol Cell*, 54: 472-84.
Hoege, C., B. Pfander, G. L. Moldovan, G. Pyrowolakis, and S. Jentsch. 2002. 'RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO', *Nature*, 419: 135-41.
Hoeijmakers, J. H. 2009. 'DNA damage, aging, and cancer', *N Engl J Med*, 361: 1475-85.
Holloway, J. K., J. Booth, W. Edelmann, C. H. McGowan, and P. E. Cohen. 2008. 'MUS81 generates a subset of MLH1-MLH3-independent crossovers in mammalian meiosis', *PLoS Genet*, 4: e1000186.
Hsia, K. T., M. R. Millar, S. King, J. Selfridge, N. J. Redhead, D. W. Melton, and P. T. Saunders. 2003. 'DNA repair gene Ercc1 is essential for normal spermatogenesis and oogenesis and for functional integrity of germ cell DNA in the mouse', *Development*, 130: 369-78.
Hu, J., J. H. Choi, S. Gaddameedhi, M. G. Kemp, J. T. Reardon, and A. Sancar. 2013. 'Nucleotide excision repair in human cells: fate of the excised oligonucleotide carrying DNA damage in vivo', *J Biol Chem*, 288: 20918-26.
Hu, X., L. Zhang, S. Q. Mao, Z. Li, J. Chen, R. R. Zhang, H. P. Wu, J. Gao, F. Guo, W. Liu, G. F. Xu, H. Q. Dai, Y. G. Shi, X. Li, B. Hu, F. Tang, D. Pei, and G. L. Xu. 2014. 'Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming', *Cell Stem Cell*, 14: 512-22.
Huang, X., C. V. Andreu-Vieyra, J. P. York, R. Hatcher, T. Lu, M. M. Matzuk, and P. Zhang. 2008. 'Inhibitory phosphorylation of separase is essential for genome stability and viability of murine embryonic germ cells', *PLoS Biol*, 6: e15.
Huang, Y., J. K. Kim, D. V. Do, C. Lee, C. A. Penfold, J. J. Zylicz, J. C. Marioni, J. A. Hackett, and M. A. Surani. 2017. 'Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition', *Elife*, 6.
Hultman, C. M., S. Sandin, S. Z. Levine, P. Lichtenstein, and A. Reichenberg. 2011. 'Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies', *Mol Psychiatry*, 16: 1203-12.
Hyldig, S. M., N. Croxall, D. A. Contreras, P. D. Thomsen, and R. Alberio. 2011. 'Epigenetic reprogramming in the porcine germ line', *BMC Dev Biol*, 11: 11.

Jansen, J. G., P. Langerak, A. Tsaalbi-Shtylik, P. van den Berk, H. Jacobs, and N. de Wind. 2006. 'Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice', *J Exp Med*, 203: 319-23.
Jeong, H. S., P. S. Backlund, H. C. Chen, A. A. Karavanov, and R. J. Crouch. 2004. 'RNase H2 of Saccharomyces cerevisiae is a complex of three proteins', *Nucleic Acids Res*, 32: 407-14.
Kafri, T., M. Ariel, M. Brandeis, R. Shemer, L. Urven, J. McCarrey, H. Cedar, and A. Razin. 1992. 'Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line', *Genes Dev*, 6: 705-14.
Kagiwada, S., K. Kurimoto, T. Hirota, M. Yamaji, and M. Saitou. 2013. 'Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice', *EMBO J*, 32: 340-53.
Kaina, B., A. Ziouta, K. Ochs, and T. Coquerelle. 1997. 'Chromosomal instability, reproductive cell death and apoptosis induced by O6-methylguanine in Mex-, Mex+ and methylation-tolerant mismatch repair compromised cells: facts and models', *Mutat Res*, 381: 227-41.
Kannouche, P. L., J. Wing, and A. R. Lehmann. 2004. 'Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage', *Mol Cell*, 14: 491-500.
Kapoor, M., and G. Lozano. 1998. 'Functional activation of p53 via phosphorylation following DNA damage by UV but not gamma radiation', *Proc Natl Acad Sci U S A*, 95: 2834-7.
Kato, T., and E. Nakano. 1981. 'Effects of the umuC36 mutation on ultraviolet-radiation-induced base-change and frameshift mutations in Escherichia coli', *Mutat Res*, 83: 307-19.
Kato, Y., K. G. Alavattam, H. S. Sin, A. R. Meetei, Q. Pang, P. R. Andreassen, and S. H. Namekawa. 2015. 'FANCB is essential in the male germline and regulates H3K9 methylation on the sex chromosomes during meiosis', *Hum Mol Genet*, 24: 5234-49.
Kelsey, G., and R. Feil. 2013. 'New insights into establishment and maintenance of DNA methylation imprints in mammals', *Philos Trans R Soc Lond B Biol Sci*, 368: 20110336.
Kim, J. M., Y. Kee, A. Gurtan, and A. D. D'Andrea. 2008. 'Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24', *Blood*, 111: 5215-22.
Kim, S., U. Gunesdogan, J. J. Zylicz, J. A. Hackett, D. Cougot, S. Bao, C. Lee, S. Dietmann, G. E. Allen, R. Sengupta, and M. A. Surani. 2014. 'PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos', *Mol Cell*, 56: 564-79.
Kirkwood, T. B. 1977. 'Evolution of ageing', *Nature*, 270: 301-4.
Knipscheer, P., M. Raschle, A. Smogorzewska, M. Enoiu, T. V. Ho, O. D. Scharer, S. J. Elledge, and J. C. Walter. 2009. 'The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair', *Science*, 326: 1698-701.
Kohler, S. W., G. S. Provost, A. Fieck, P. L. Kretz, W. O. Bullock, J. A. Sorge, D. L. Putman, and J. M. Short. 1991. 'Spectra of spontaneous and mutagen-induced mutations in the lacI gene in transgenic mice', *Proc Natl Acad Sci U S A*, 88: 7958-62.
Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, W. S. Wong, G.

Sigurdsson, G. B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D. F. Gudbjartsson, A. Helgason, O. T. Magnusson, U. Thorsteinsdottir, and K. Stefansson. 2012. 'Rate of de novo mutations and the importance of father's age to disease risk', *Nature*, 488: 471-5.

Krausz, C., A. Riera-Escamilla, C. Chianese, D. Moreno-Mendoza, E. Ars, O. Rajmil, R. Pujol, M. Bogliolo, I. Blanco, I. Rodriguez, I. Badell, E. Ruiz-Castane, and J. Surralles. 2019. 'From exome analysis in idiopathic azoospermia to the identification of a high-risk subgroup for occult Fanconi anemia', *Genet Med*, 21: 189-94.

Krokan, H. E., and M. Bjoras. 2013. 'Base excision repair', *Cold Spring Harb Perspect Biol*, 5: a012583.

Kroutil, L. C., K. Register, K. Bebenek, and T. A. Kunkel. 1996. 'Exonucleolytic proofreading during replication of repetitive DNA', *Biochemistry*, 35: 1046-53.

Kuchenbaecker, K. B., J. L. Hopper, D. R. Barnes, K. A. Phillips, T. M. Mooij, M. J. Roos-Blom, S. Jervis, F. E. van Leeuwen, R. L. Milne, N. Andrieu, D. E. Goldgar, M. B. Terry, M. A. Rookus, D. F. Easton, A. C. Antoniou, Brca, Brca Cohort Consortium, L. McGuffog, D. G. Evans, D. Barrowdale, D. Frost, J. Adlard, K. R. Ong, L. Izatt, M. Tischkowitz, R. Eeles, R. Davidson, S. Hodgson, S. Ellis, C. Nogues, C. Lasset, D. Stoppa-Lyonnet, J. P. Fricker, L. Faivre, P. Berthet, M. J. Hooning, L. E. van der Kolk, C. M. Kets, M. A. Adank, E. M. John, W. K. Chung, I. L. Andrulis, M. Southey, M. B. Daly, S. S. Buys, A. Osorio, C. Engel, K. Kast, R. K. Schmutzler, T. Caldes, A. Jakubowska, J. Simard, M. L. Friedlander, S. A. McLachlan, E. Machackova, L. Foretova, Y. Y. Tan, C. F. Singer, E. Olah, A. M. Gerdes, B. Arver, and H. Olsson. 2017. 'Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers', *JAMA*, 317: 2402-16.

Kudo, T., M. Kaneko, H. Iwasaki, A. Togayachi, S. Nishihara, K. Abe, and H. Narimatsu. 2004. 'Normal embryonic and germ cell development in mice lacking alpha 1,3-fucosyltransferase IX (Fut9) which show disappearance of stage-specific embryonic antigen 1', *Mol Cell Biol*, 24: 4221-8.

Kurimoto, K., Y. Yabuta, Y. Ohinata, M. Shigeta, K. Yamanaka, and M. Saitou. 2008. 'Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice', *Genes Dev*, 22: 1617-35.

Ladstatter, S., and K. Tachibana-Konwalski. 2016. 'A Surveillance Mechanism Ensures Repair of DNA Lesions during Zygotic Reprogramming', *Cell*, 167: 1774-87 e13.

Lahue, R. S., K. G. Au, and P. Modrich. 1989. 'DNA mismatch correction in a defined system', *Science*, 245: 160-4.

Lange, U. C., D. J. Adams, C. Lee, S. Barton, R. Schneider, A. Bradley, and M. A. Surani. 2008. 'Normal germ line establishment in mice carrying a deletion of the Ifitm/Fragilis gene family cluster', *Mol Cell Biol*, 28: 4688-96.

Larsen, N. B., A. O. Gao, J. L. Sparks, I. Gallina, R. A. Wu, M. Mann, M. Raschle, J. C. Walter, and J. P. Duxin. 2019. 'Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts', *Mol Cell*, 73: 574-88 e7.

Lawson, K. A., N. R. Dunn, B. A. Roelen, L. M. Zeinstra, A. M. Davis, C. V. Wright, J. P. Korving, and B. L. Hogan. 1999. 'Bmp4 is required for the generation of primordial germ cells in the mouse embryo', *Genes Dev*, 13: 424-36.

Lawson, K. A., and W. J. Hage. 1994. 'Clonal analysis of the origin of primordial germ cells in the mouse', *Ciba Found Symp*, 182: 68-84; discussion 84-91.
Lee, J., K. Inoue, R. Ono, N. Ogonuki, T. Kohda, T. Kaneko-Ishino, A. Ogura, and F. Ishino. 2002. 'Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells', *Development*, 129: 1807-17.
Lee, J. M., and A. Bernstein. 1993. 'p53 mutations increase resistance to ionizing radiation', *Proc Natl Acad Sci U S A*, 90: 5742-6.
Lemontt, J. F. 1971. 'Mutants of yeast defective in mutation induced by ultraviolet light', *Genetics*, 68: 21-33.
Lenain, C., S. Bauwens, S. Amiard, M. Brunori, M. J. Giraud-Panis, and E. Gilson. 2006. 'The Apollo 5' exonuclease functions together with TRF2 to protect telomeres from DNA repair', *Curr Biol*, 16: 1303-10.
Li, M., J. S. L. Yu, K. Tilgner, S. H. Ong, H. Koike-Yusa, and K. Yusa. 2018. 'Genome-wide CRISPR-KO Screen Uncovers mTORC1-Mediated Gsk3 Regulation in Naive Pluripotency Maintenance and Dissolution', *Cell Rep*, 24: 489-502.
Li, Y., M. A. Trush, and J. D. Yager. 1994. 'DNA damage caused by reactive oxygen species originating from a copper-dependent oxidation of the 2-hydroxy catechol of estradiol', *Carcinogenesis*, 15: 1421-7.
Li, Y., Z. Zhang, J. Chen, W. Liu, W. Lai, B. Liu, X. Li, L. Liu, S. Xu, Q. Dong, M. Wang, X. Duan, J. Tan, Y. Zheng, P. Zhang, G. Fan, J. Wong, G. L. Xu, Z. Wang, H. Wang, S. Gao, and B. Zhu. 2018. 'Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1', *Nature*, 564: 136-40.
Li, Z., T. P. Gu, A. R. Weber, J. Z. Shen, B. Z. Li, Z. G. Xie, R. Yin, F. Guo, X. Liu, F. Tang, H. Wang, P. Schar, and G. L. Xu. 2015. 'Gadd45a promotes DNA demethylation through TDG', *Nucleic Acids Res*, 43: 3986-97.
Limoli, C. L., E. Giedzinski, W. M. Bonner, and J. E. Cleaver. 2002. 'UV-induced replication arrest in the xeroderma pigmentosum variant leads to DNA double-strand breaks, gamma -H2AX formation, and Mre11 relocalization', *Proc Natl Acad Sci U S A*, 99: 233-8.
Limoli, C. L., R. Laposa, and J. E. Cleaver. 2002. 'DNA replication arrest in XP variant cells after UV exposure is diverted into an Mre11-dependent recombination pathway by the kinase inhibitor wortmannin', *Mutat Res*, 510: 121-9.
Lin, T., C. Chao, S. Saito, S. J. Mazur, M. E. Murphy, E. Appella, and Y. Xu. 2005. 'p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression', *Nat Cell Biol*, 7: 165-71.
Lindahl, T. 1976. 'New class of enzymes acting on damaged DNA', *Nature*, 259: 64-6.
———. 1993. 'Instability and decay of the primary structure of DNA', *Nature*, 362: 709-15.
Lindahl, T., and D. E. Barnes. 2000. 'Repair of endogenous DNA damage', *Cold Spring Harb Symp Quant Biol*, 65: 127-33.
Lindahl, T., and B. Nyberg. 1974. 'Heat-induced deamination of cytosine residues in deoxyribonucleic acid', *Biochemistry*, 13: 3405-10.
Lips, J., and B. Kaina. 2001. 'DNA double-strand breaks trigger apoptosis in p53-deficient fibroblasts', *Carcinogenesis*, 22: 579-85.
Listovsky, T., and J. E. Sale. 2013. 'Sequestration of CDH1 by MAD2L2 prevents premature APC/C activation prior to anaphase onset', *J Cell Biol*, 203: 87-100.

Liu, T., G. Ghosal, J. Yuan, J. Chen, and J. Huang. 2010. 'FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair', *Science*, 329: 693-6.
Liu, Y., S. E. Elf, Y. Miyata, G. Sashida, Y. Liu, G. Huang, S. Di Giandomenico, J. M. Lee, A. Deblasio, S. Menendez, J. Antipin, B. Reva, A. Koff, and S. D. Nimer. 2009. 'p53 regulates hematopoietic stem cell quiescence', *Cell Stem Cell*, 4: 37-48.
Losada, A., T. Yokochi, and T. Hirano. 2005. 'Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and Xenopus egg extracts', *J Cell Sci*, 118: 2133-41.
Luo, Y., S. A. Hartford, R. Zeng, T. L. Southard, N. Shima, and J. C. Schimenti. 2014. 'Hypersensitivity of primordial germ cells to compromised replication-associated DNA repair involves ATM-p53-p21 signaling', *PLoS Genet*, 10: e1004471.
Luo, Y., and J. C. Schimenti. 2015. 'MCM9 deficiency delays primordial germ cell proliferation independent of the ATM pathway', *Genesis*, 53: 678-84.
Lutzmann, M., C. Grey, S. Traver, O. Ganier, A. Maya-Mendoza, N. Ranisavljevic, F. Bernex, A. Nishiyama, N. Montel, E. Gavois, L. Forichon, B. de Massy, and M. Mechali. 2012. 'MCM8- and MCM9-deficient mice reveal gametogenesis defects and genome instability due to impaired homologous recombination', *Mol Cell*, 47: 523-34.
Lyon, M. F. 1961. 'Gene action in the X-chromosome of the mouse (Mus musculus L.)', *Nature*, 190: 372-3.
MacGregor, G. R., B. P. Zambrowicz, and P. Soriano. 1995. 'Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells', *Development*, 121: 1487-96.
Magnusdottir, E., S. Dietmann, K. Murakami, U. Gunesdogan, F. Tang, S. Bao, E. Diamanti, K. Lao, B. Gottgens, and M. Azim Surani. 2013. 'A tripartite transcription factor network regulates primordial germ cell specification in mice', *Nat Cell Biol*, 15: 905-15.
Martinez, J., I. Georgoff, J. Martinez, and A. J. Levine. 1991. 'Cellular localization and cell cycle regulation by a temperature-sensitive p53 protein', *Genes Dev*, 5: 151-9.
Masuda, K., R. Ouchida, Y. Li, X. Gao, H. Mori, and J. Y. Wang. 2009. 'A critical role for REV1 in regulating the induction of C:G transitions and A:T mutations during Ig gene hypermutation', *J Immunol*, 183: 1846-50.
Masuda, Y., M. Takahashi, N. Tsunekuni, T. Minami, M. Sumii, K. Miyagawa, and K. Kamiya. 2001. 'Deoxycytidyl transferase activity of the human REV1 protein is closely associated with the conserved polymerase domain', *J Biol Chem*, 276: 15051-8.
Matos, J., M. G. Blanco, S. Maslen, J. M. Skehel, and S. C. West. 2011. 'Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis', *Cell*, 147: 158-72.
Maya-Mendoza, A., P. Moudry, J. M. Merchut-Maya, M. Lee, R. Strauss, and J. Bartek. 2018. 'High speed of fork progression induces DNA replication stress and genomic instability', *Nature*, 559: 279-84.
Mayer, W., A. Niveleau, J. Walter, R. Fundele, and T. Haaf. 2000. 'Demethylation of the zygotic paternal genome', *Nature*, 403: 501-2.
McPherson, J. P., B. Lemmers, R. Chahwan, A. Pamidi, E. Migon, E. Matysiak-Zablocki, M. E. Moynahan, J. Essers, K. Hanada, A. Poonepalli, O. Sanchez-

Sweatman, R. Khokha, R. Kanaar, M. Jasin, M. P. Hande, and R. Hakem. 2004. 'Involvement of mammalian Mus81 in genome integrity and tumor suppression', *Science*, 304: 1822-6.
McWhir, J., J. Selfridge, D. J. Harrison, S. Squires, and D. W. Melton. 1993. 'Mice with DNA repair gene (ERCC-1) deficiency have elevated levels of p53, liver nuclear abnormalities and die before weaning', *Nat Genet*, 5: 217-24.
Meletis, K., V. Wirta, S. M. Hede, M. Nister, J. Lundeberg, and J. Frisen. 2006. 'p53 suppresses the self-renewal of adult neural stem cells', *Development*, 133: 363-9.
Melton, D. W., A. M. Ketchen, F. Nunez, S. Bonatti-Abbondandolo, A. Abbondandolo, S. Squires, and R. T. Johnson. 1998. 'Cells from ERCC1-deficient mice show increased genome instability and a reduced frequency of S-phase-dependent illegitimate chromosome exchange but a normal frequency of homologous recombination', *J Cell Sci*, 111 (Pt 3): 395-404.
Merkle, F. T., S. Ghosh, N. Kamitaki, J. Mitchell, Y. Avior, C. Mello, S. Kashin, S. Mekhoubad, D. Ilic, M. Charlton, G. Saphier, R. E. Handsaker, G. Genovese, S. Bar, N. Benvenisty, S. A. McCarroll, and K. Eggan. 2017. 'Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations', *Nature*, 545: 229-33.
Messiaen, S., A. Le Bras, C. Duquenne, V. Barroca, D. Moison, N. Dechamps, M. Doussau, A. L. Bauchet, M. J. Guerquin, G. Livera, J. Essers, R. Kanaar, R. Habert, and J. Bernardino-Sgherri. 2013. 'Rad54 is required for the normal development of male and female germ cells and contributes to the maintainance of their genome integrity after genotoxic stress', *Cell Death Dis*, 4: e774.
Michaelson, J. J., Y. Shi, M. Gujral, H. Zheng, D. Malhotra, X. Jin, M. Jian, G. Liu, D. Greer, A. Bhandari, W. Wu, R. Corominas, A. Peoples, A. Koren, A. Gore, S. Kang, G. N. Lin, J. Estabillo, T. Gadomski, B. Singh, K. Zhang, N. Akshoomoff, C. Corsello, S. McCarroll, L. M. Iakoucheva, Y. Li, J. Wang, and J. Sebat. 2012. 'Whole-genome sequencing in autism identifies hot spots for de novo germline mutation', *Cell*, 151: 1431-42.
Miyoshi, N., J. M. Stel, K. Shioda, N. Qu, J. Odajima, S. Mitsunaga, X. Zhang, M. Nagano, K. Hochedlinger, K. J. Isselbacher, and T. Shioda. 2016. 'Erasure of DNA methylation, genomic imprints, and epimutations in a primordial germ-cell model derived from mouse pluripotent stem cells', *Proc Natl Acad Sci U S A*, 113: 9545-50.
Mochizuki, K., Y. Hayashi, T. Sekinaka, K. Otsuka, Y. Ito-Matsuoka, H. Kobayashi, S. Oki, A. Takehara, T. Kono, N. Osumi, and Y. Matsui. 2018. 'Repression of Somatic Genes by Selective Recruitment of HDAC3 by BLIMP1 Is Essential for Mouse Primordial Germ Cell Fate Determination', *Cell Rep*, 24: 2682-93 e6.
Mohamet, L., M. L. Lea, and C. M. Ward. 2010. 'Abrogation of E-cadherin-mediated cellular aggregation allows proliferation of pluripotent mouse embryonic stem cells in shake flask bioreactors', *PLoS One*, 5: e12921.
Monk, M., M. Boubelik, and S. Lehnert. 1987. 'Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development', *Development*, 99: 371-82.
Monk, M., and A. McLaren. 1981. 'X-chromosome activity in foetal germ cells of the mouse', *J Embryol Exp Morphol*, 63: 75-84.

Morales-Ruiz, T., A. P. Ortega-Galisteo, M. I. Ponferrada-Marin, M. I. Martinez-Macias, R. R. Ariza, and T. Roldan-Arjona. 2006. 'DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases', *Proc Natl Acad Sci U S A*, 103: 6853-8.

Motro, B., D. van der Kooy, J. Rossant, A. Reith, and A. Bernstein. 1991. 'Contiguous patterns of c-kit and steel expression: analysis of mutations at the W and Sl loci', *Development*, 113: 1207-21.

Mu, D., M. Tursun, D. R. Duckett, J. T. Drummond, P. Modrich, and A. Sancar. 1997. 'Recognition and repair of compound DNA lesions (base damage and mismatch) by human mismatch repair and excision repair systems', *Mol Cell Biol*, 17: 760-9.

Murakumo, Y., Y. Ogura, H. Ishii, S. Numata, M. Ichihara, C. M. Croce, R. Fishel, and M. Takahashi. 2001. 'Interactions in the error-prone postreplication repair proteins hREV1, hREV3, and hREV7', *J Biol Chem*, 276: 35644-51.

Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. 'Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme', *Cell*, 102: 553-63.

Nakamura, T., Y. Arai, H. Umehara, M. Masuhara, T. Kimura, H. Taniguchi, T. Sekimoto, M. Ikawa, Y. Yoneda, M. Okabe, S. Tanaka, K. Shiota, and T. Nakano. 2007. 'PGC7/Stella protects against DNA demethylation in early embryogenesis', *Nat Cell Biol*, 9: 64-71.

Nakane, H., S. Takeuchi, S. Yuba, M. Saijo, Y. Nakatsu, H. Murai, Y. Nakatsuru, T. Ishikawa, S. Hirota, Y. Kitamura, and et al. 1995. 'High incidence of ultraviolet-B-or chemical-carcinogen-induced skin tumours in mice lacking the xeroderma pigmentosum group A gene', *Nature*, 377: 165-8.

Nakano, T., S. Morishita, A. Katafuchi, M. Matsubara, Y. Horikawa, H. Terato, A. M. Salem, S. Izumi, S. P. Pack, K. Makino, and H. Ide. 2007. 'Nucleotide excision repair and homologous recombination systems commit differentially to the repair of DNA-protein crosslinks', *Mol Cell*, 28: 147-58.

Narod, S. A. 2001. 'Hormonal prevention of hereditary breast cancer', *Ann N Y Acad Sci*, 952: 36-43.

Neale, M. J., and S. Keeney. 2006. 'Clarifying the mechanics of DNA strand exchange in meiotic recombination', *Nature*, 442: 153-8.

Nelson, J. R., C. W. Lawrence, and D. C. Hinkle. 1996a. 'Deoxycytidyl transferase activity of yeast REV1 protein', *Nature*, 382: 729-31.

———. 1996b. 'Thymine-thymine dimer bypass by yeast DNA polymerase zeta', *Science*, 272: 1646-9.

Nik-Zainal, S., P. Van Loo, D. C. Wedge, L. B. Alexandrov, C. D. Greenman, K. W. Lau, K. Raine, D. Jones, J. Marshall, M. Ramakrishna, A. Shlien, S. L. Cooke, J. Hinton, A. Menzies, L. A. Stebbings, C. Leroy, M. Jia, R. Rance, L. J. Mudie, S. J. Gamble, P. J. Stephens, S. McLaren, P. S. Tarpey, E. Papaemmanuil, H. R. Davies, I. Varela, D. J. McBride, G. R. Bignell, K. Leung, A. P. Butler, J. W. Teague, S. Martin, G. Jonsson, O. Mariani, S. Boyault, P. Miron, A. Fatima, A. Langerod, S. A. Aparicio, A. Tutt, A. M. Sieuwerts, A. Borg, G. Thomas, A. V. Salomon, A. L. Richardson, A. L. Borresen-Dale, P. A. Futreal, M. R. Stratton, P. J. Campbell, and Consortium Breast Cancer Working Group of the International Cancer Genome. 2012. 'The life history of 21 breast cancers', *Cell*, 149: 994-1007.

Nishino, T., K. Komori, Y. Ishino, and K. Morikawa. 2003. 'X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes', *Structure*, 11: 445-57.
Nolan, E., G. J. Lindeman, and J. E. Visvader. 2017. 'Out-RANKing BRCA1 in Mutation Carriers', *Cancer Res*, 77: 595-600.
Ohashi, E., Y. Murakumo, N. Kanjo, J. Akagi, C. Masutani, F. Hanaoka, and H. Ohmori. 2004. 'Interaction of hREV1 with three human Y-family DNA polymerases', *Genes Cells*, 9: 523-31.
Ohinata, Y., H. Ohta, M. Shigeta, K. Yamanaka, T. Wakayama, and M. Saitou. 2009. 'A signaling principle for the specification of the germ cell lineage in mice', *Cell*, 137: 571-84.
Ohinata, Y., B. Payer, D. O'Carroll, K. Ancelin, Y. Ono, M. Sano, S. C. Barton, T. Obukhanych, M. Nussenzweig, A. Tarakhovsky, M. Saitou, and M. A. Surani. 2005. 'Blimp1 is a critical determinant of the germ cell lineage in mice', *Nature*, 436: 207-13.
Ohta, H., K. Kurimoto, I. Okamoto, T. Nakamura, Y. Yabuta, H. Miyauchi, T. Yamamoto, Y. Okuno, M. Hagiwara, K. Shirane, H. Sasaki, and M. Saitou. 2017. 'In vitro expansion of mouse primordial germ cell-like cells recapitulates an epigenetic blank slate', *EMBO J*, 36: 1888-907.
Okada, T., E. Sonoda, M. Yoshimura, Y. Kawano, H. Saya, M. Kohzaki, and S. Takeda. 2005. 'Multiple roles of vertebrate REV genes in DNA repair and recombination', *Mol Cell Biol*, 25: 6103-11.
Okamoto, I., A. P. Otte, C. D. Allis, D. Reinberg, and E. Heard. 2004. 'Epigenetic dynamics of imprinted X inactivation during early mouse development', *Science*, 303: 644-9.
Overduin, M., T. S. Harvey, S. Bagby, K. I. Tong, P. Yau, M. Takeichi, and M. Ikura. 1995. 'Solution structure of the epithelial cadherin domain responsible for selective cell adhesion', *Science*, 267: 386-9.
Panizza, S., T. Tanaka, A. Hochwagen, F. Eisenhaber, and K. Nasmyth. 2000. 'Pds5 cooperates with cohesin in maintaining sister chromatid cohesion', *Curr Biol*, 10: 1557-64.
Panning, B., and R. Jaenisch. 1996. 'DNA hypomethylation can activate Xist expression and silence X-linked genes', *Genes Dev*, 10: 1991-2002.
Paul, C., J. E. Povey, N. J. Lawrence, J. Selfridge, D. W. Melton, and P. T. Saunders. 2007. 'Deletion of genes implicated in protecting the integrity of male germ cells has differential effects on the incidence of DNA breaks and germ cell loss', *PLoS One*, 2: e989.
Payer, B., S. M. Chuva de Sousa Lopes, S. C. Barton, C. Lee, M. Saitou, and M. A. Surani. 2006. 'Generation of stella-GFP transgenic mice: a novel tool to study germ cell development', *Genesis*, 44: 75-83.
Payer, B., M. Saitou, S. C. Barton, R. Thresher, J. P. Dixon, D. Zahn, W. H. Colledge, M. B. Carlton, T. Nakano, and M. A. Surani. 2003. 'Stella is a maternal effect gene required for normal early development in mice', *Curr Biol*, 13: 2110-7.
Phillips, D. H., A. Hewer, C. N. Martin, R. C. Garner, and M. M. King. 1988. 'Correlation of DNA adduct levels in human lung with cigarette smoking', *Nature*, 336: 790-2.
Pirouz, M., S. Pilarski, and M. Kessel. 2013. 'A critical function of Mad2l2 in primordial germ cell development of mice', *PLoS Genet*, 9: e1003712.
Poll, E. H., F. Arwert, H. Joenje, and A. H. Wanamarta. 1985. 'Differential sensitivity of Fanconi anaemia lymphocytes to the clastogenic action of cis-

diamminedichloroplatinum (II) and trans-diamminedichloroplatinum (II)', *Hum Genet*, 71: 206-10.
Pommier, Y., C. Redon, V. A. Rao, J. A. Seiler, O. Sordet, H. Takemura, S. Antony, L. Meng, Z. Liao, G. Kohlhagen, H. Zhang, and K. W. Kohn. 2003. 'Repair of and checkpoint response to topoisomerase I-mediated DNA damage', *Mutat Res*, 532: 173-203.
Popp, C., W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, and W. Reik. 2010. 'Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency', *Nature*, 463: 1101-5.
Pouliot, J. J., K. C. Yao, C. A. Robertson, and H. A. Nash. 1999. 'Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes', *Science*, 286: 552-5.
Prakash, S., R. E. Johnson, and L. Prakash. 2005. 'Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function', *Annu Rev Biochem*, 74: 317-53.
Prolla, T. A., S. M. Baker, A. C. Harris, J. L. Tsao, X. Yao, C. E. Bronner, B. Zheng, M. Gordon, J. Reneker, N. Arnheim, D. Shibata, A. Bradley, and R. M. Liskay. 1998. 'Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair', *Nat Genet*, 18: 276-9.
Pustovalova, Y., I. Bezsonova, and D. M. Korzhnev. 2012. 'The C-terminal domain of human Rev1 contains independent binding sites for DNA polymerase eta and Rev7 subunit of polymerase zeta', *FEBS Lett*, 586: 3051-6.
Putnam, C. D., T. K. Hayes, and R. D. Kolodner. 2009. 'Specific pathways prevent duplication-mediated genome rearrangements', *Nature*, 460: 984-9.
Qiang, L., P. Shah, M. H. Barcellos-Hoff, and Y. Y. He. 2016. 'TGF-beta signaling links E-cadherin loss to suppression of nucleotide excision repair', *Oncogene*, 35: 3293-302.
Rahbari, R., A. Wuster, S. J. Lindsay, R. J. Hardwick, L. B. Alexandrov, S. A. Turki, A. Dominiczak, A. Morris, D. Porteous, B. Smith, M. R. Stratton, Uk K. Consortium, and M. E. Hurles. 2016. 'Timing, rates and spectra of human germline mutation', *Nat Genet*, 48: 126-33.
Rahjouei, A., M. Pirouz, M. Di Virgilio, D. Kamin, and M. Kessel. 2017. 'MAD2L2 Promotes Open Chromatin in Embryonic Stem Cells and Derepresses the Dppa3 Locus', *Stem Cell Reports*, 8: 813-21.
Ravanat, J. L., T. Douki, and J. Cadet. 2001. 'Direct and indirect effects of UV radiation on DNA and its components', *J Photochem Photobiol B*, 63: 88-102.
Reitmair, A. H., R. Schmits, A. Ewel, B. Bapat, M. Redston, A. Mitri, P. Waterhouse, H. W. Mittrucker, A. Wakeham, B. Liu, and et al. 1995. 'MSH2 deficient mice are viable and susceptible to lymphoid tumours', *Nat Genet*, 11: 64-70.
Richter, K., T. Paakkola, D. Mennerich, K. Kubaichuk, A. Konzack, H. Ali-Kippari, N. Kozlova, P. Koivunen, K. M. Haapasaari, A. Jukkola-Vuorinen, H. R. Teppo, E. Y. Dimova, R. Bloigu, Z. Szabo, R. Kerkela, and T. Kietzmann. 2018. 'USP28 Deficiency Promotes Breast and Liver Carcinogenesis as well as Tumor Angiogenesis in a HIF-independent Manner', *Mol Cancer Res*, 16: 1000-12.
Riethmacher, D., V. Brinkmann, and C. Birchmeier. 1995. 'A targeted mutation in the mouse E-cadherin gene results in defective preimplantation development', *Proc Natl Acad Sci U S A*, 92: 855-9.

Ross, A. L., L. J. Simpson, and J. E. Sale. 2005. 'Vertebrate DNA damage tolerance requires the C-terminus but not BRCT or transferase domains of REV1', *Nucleic Acids Res*, 33: 1280-9.
Rougier, N., D. Bourc'his, D. M. Gomes, A. Niveleau, M. Plachot, A. Paldi, and E. Viegas-Pequignot. 1998. 'Chromosome methylation patterns during mammalian preimplantation development', *Genes Dev*, 12: 2108-13.
Sabapathy, K., M. Klemm, R. Jaenisch, and E. F. Wagner. 1997. 'Regulation of ES cell differentiation by functional and conformational modulation of p53', *EMBO J*, 16: 6217-29.
Saitou, M., S. C. Barton, and M. A. Surani. 2002. 'A molecular programme for the specification of germ cell fate in mice', *Nature*, 418: 293-300.
Saitou, M., S. Kagiwada, and K. Kurimoto. 2012. 'Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells', *Development*, 139: 15-31.
Sambrook, J. And Russell, D. 2001. Molecular Cloning: A Laboratory Manual, 3rd ed., Vols 1,2 and 3.
Sanjana, N. E., O. Shalem, and F. Zhang. 2014. 'Improved vectors and genome-wide libraries for CRISPR screening', *Nat Methods*, 11: 783-84.
Sargent, R. G., J. L. Meservy, B. D. Perkins, A. E. Kilburn, Z. Intody, G. M. Adair, R. S. Nairn, and J. H. Wilson. 2000. 'Role of the nucleotide excision repair gene ERCC1 in formation of recombination-dependent rearrangements in mammalian cells', *Nucleic Acids Res*, 28: 3771-8.
Sarkar, S., A. A. Davies, H. D. Ulrich, and P. J. McHugh. 2006. 'DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta', *EMBO J*, 25: 1285-94.
Sasaki, M. S. 1975. 'Is Fanconi's anaemia defective in a process essential to the repair of DNA cross links?', *Nature*, 257: 501-3.
Sasaki, M. S., and A. Tonomura. 1973. 'A high susceptibility of Fanconi's anemia to chromosome breakage by DNA cross-linking agents', *Cancer Res*, 33: 1829-36.
Sato, M., T. Kimura, K. Kurokawa, Y. Fujita, K. Abe, M. Masuhara, T. Yasunaga, A. Ryo, M. Yamamoto, and T. Nakano. 2002. 'Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells', *Mech Dev*, 113: 91-4.
Schaaper, R. M. 1988. 'Mechanisms of mutagenesis in the Escherichia coli mutator mutD5: role of DNA mismatch repair', *Proc Natl Acad Sci U S A*, 85: 8126-30.
Seisenberger, S., S. Andrews, F. Krueger, J. Arand, J. Walter, F. Santos, C. Popp, B. Thienpont, W. Dean, and W. Reik. 2012. 'The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells', *Mol Cell*, 48: 849-62.
Seki, Y., M. Yamaji, Y. Yabuta, M. Sano, M. Shigeta, Y. Matsui, Y. Saga, M. Tachibana, Y. Shinkai, and M. Saitou. 2007. 'Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice', *Development*, 134: 2627-38.
Selfridge, J., K. T. Hsia, N. J. Redhead, and D. W. Melton. 2001. 'Correction of liver dysfunction in DNA repair-deficient mice with an ERCC1 transgene', *Nucleic Acids Res*, 29: 4541-50.

Seligman, J., and D. C. Page. 1998. 'The Dazh gene is expressed in male and female embryonic gonads before germ cell sex differentiation', *Biochem Biophys Res Commun*, 245: 878-82.
Shalem, O., N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. Mikkelson, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, and F. Zhang. 2014. 'Genome-scale CRISPR-Cas9 knockout screening in human cells', *Science*, 343: 84-87.
Sharif, J., M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, and H. Koseki. 2007. 'The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA', *Nature*, 450: 908-12.
Shcherbakova, P. V., Y. I. Pavlov, O. Chilkova, I. B. Rogozin, E. Johansson, and T. A. Kunkel. 2003. 'Unique error signature of the four-subunit yeast DNA polymerase epsilon', *J Biol Chem*, 278: 43770-80.
Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero, and Y. Shi. 2004. 'Histone demethylation mediated by the nuclear amine oxidase homolog LSD1', *Cell*, 119: 941-53.
Shivji, M. K., V. N. Podust, U. Hubscher, and R. D. Wood. 1995. 'Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA', *Biochemistry*, 34: 5011-7.
Sijbers, A. M., W. L. de Laat, R. R. Ariza, M. Biggerstaff, Y. F. Wei, J. G. Moggs, K. C. Carter, B. K. Shell, E. Evans, M. C. de Jong, S. Rademakers, J. de Rooij, N. G. Jaspers, J. H. Hoeijmakers, and R. D. Wood. 1996. 'Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease', *Cell*, 86: 811-22.
Sim, Y. J., M. S. Kim, A. Nayfeh, Y. J. Yun, S. J. Kim, K. T. Park, C. H. Kim, and K. S. Kim. 2017. '2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms', *Stem Cell Reports*, 8: 1312-28.
Simpson, L. J., and J. E. Sale. 2003. 'Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line', *EMBO J*, 22: 1654-64.
Simsek, D., and M. Jasin. 2010. 'Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation', *Nat Struct Mol Biol*, 17: 410-6.
Sipos, A., F. Rasmussen, G. Harrison, P. Tynelius, G. Lewis, D. A. Leon, and D. Gunnell. 2004. 'Paternal age and schizophrenia: a population based cohort study', *BMJ*, 329: 1070.
Smirin-Yosef, P., N. Zuckerman-Levin, S. Tzur, Y. Granot, L. Cohen, J. Sachsenweger, G. Borck, I. Lagovsky, M. Salmon-Divon, L. Wiesmuller, and L. Basel-Vanagaite. 2017. 'A Biallelic Mutation in the Homologous Recombination Repair Gene SPIDR Is Associated With Human Gonadal Dysgenesis', *J Clin Endocrinol Metab*, 102: 681-88.
Smith, A. G., J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, and D. Rogers. 1988. 'Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides', *Nature*, 336: 688-90.
Smogorzewska, A., R. Desetty, T. T. Saito, M. Schlabach, F. P. Lach, M. E. Sowa, A. B. Clark, T. A. Kunkel, J. W. Harper, M. P. Colaiacovo, and S. J. Elledge. 2010. 'A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair', *Mol Cell*, 39: 36-47.
Sobol, R. W., D. E. Watson, J. Nakamura, F. M. Yakes, E. Hou, J. K. Horton, J. Ladapo, B. Van Houten, J. A. Swenberg, K. R. Tindall, L. D. Samson, and S.

H. Wilson. 2002. 'Mutations associated with base excision repair deficiency and methylation-induced genotoxic stress', *Proc Natl Acad Sci U S A*, 99: 6860-5.

Stelter, P., and H. D. Ulrich. 2003. 'Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation', *Nature*, 425: 188-91.

Stephens, P. J., C. D. Greenman, B. Fu, F. Yang, G. R. Bignell, L. J. Mudie, E. D. Pleasance, K. W. Lau, D. Beare, L. A. Stebbings, S. McLaren, M. L. Lin, D. J. McBride, I. Varela, S. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, M. A. Quail, J. Burton, H. Swerdlow, N. P. Carter, L. A. Morsberger, C. Iacobuzio-Donahue, G. A. Follows, A. R. Green, A. M. Flanagan, M. R. Stratton, P. A. Futreal, and P. J. Campbell. 2011. 'Massive genomic rearrangement acquired in a single catastrophic event during cancer development', *Cell*, 144: 27-40.

Stingele, J., R. Bellelli, and S. J. Boulton. 2017. 'Mechanisms of DNA-protein crosslink repair', *Nat Rev Mol Cell Biol*, 18: 563-73.

Stone, M. P., Y. J. Cho, H. Huang, H. Y. Kim, I. D. Kozekov, A. Kozekova, H. Wang, I. G. Minko, R. S. Lloyd, T. M. Harris, and C. J. Rizzo. 2008. 'Interstrand DNA cross-links induced by alpha,beta-unsaturated aldehydes derived from lipid peroxidation and environmental sources', *Acc Chem Res*, 41: 793-804.

Sugasawa, K., J. M. Ng, C. Masutani, S. Iwai, P. J. van der Spek, A. P. Eker, F. Hanaoka, D. Bootsma, and J. H. Hoeijmakers. 1998. 'Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair', *Mol Cell*, 2: 223-32.

Sugasawa, K., T. Okamoto, Y. Shimizu, C. Masutani, S. Iwai, and F. Hanaoka. 2001. 'A multistep damage recognition mechanism for global genomic nucleotide excision repair', *Genes Dev*, 15: 507-21.

Suh, E. K., A. Yang, A. Kettenbach, C. Bamberger, A. H. Michaelis, Z. Zhu, J. A. Elvin, R. T. Bronson, C. P. Crum, and F. McKeon. 2006. 'p63 protects the female germ line during meiotic arrest', *Nature*, 444: 624-8.

Sun, X. J., J. Wei, X. Y. Wu, M. Hu, L. Wang, H. H. Wang, Q. H. Zhang, S. J. Chen, Q. H. Huang, and Z. Chen. 2005. 'Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase', *J Biol Chem*, 280: 35261-71.

Surani, M. A., S. C. Barton, and M. L. Norris. 1984. 'Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis', *Nature*, 308: 548-50.

Szabo, P. E., K. Hubner, H. Scholer, and J. R. Mann. 2002. 'Allele-specific expression of imprinted genes in mouse migratory primordial germ cells', *Mech Dev*, 115: 157-60.

Tada, M., Y. Takahama, K. Abe, N. Nakatsuji, and T. Tada. 2001. 'Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells', *Curr Biol*, 11: 1553-8.

Takahashi, K., and S. Yamanaka. 2006. 'Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors', *Cell*, 126: 663-76.

Tam, P. P., and M. H. Snow. 1981. 'Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos', *J Embryol Exp Morphol*, 64: 133-47.

Tanaka, S. S., and Y. Matsui. 2002. 'Developmentally regulated expression of mil-1 and mil-2, mouse interferon-induced transmembrane protein like genes, during formation and differentiation of primordial germ cells', *Gene Expr Patterns*, 2: 297-303.

Thongthip, S., M. Bellani, S. Q. Gregg, S. Sridhar, B. A. Conti, Y. Chen, M. M. Seidman, and A. Smogorzewska. 2016. 'Fan1 deficiency results in DNA interstrand cross-link repair defects, enhanced tissue karyomegaly, and organ dysfunction', *Genes Dev*, 30: 645-59.

Tissier, A., P. Kannouche, M. P. Reck, A. R. Lehmann, R. P. Fuchs, and A. Cordonnier. 2004. 'Co-localization in replication foci and interaction of human Y-family members, DNA polymerase pol eta and REVl protein', *DNA Repair (Amst)*, 3: 1503-14.

Toledo, F., K. A. Krummel, C. J. Lee, C. W. Liu, L. W. Rodewald, M. Tang, and G. M. Wahl. 2006. 'A mouse p53 mutant lacking the proline-rich domain rescues Mdm4 deficiency and provides insight into the Mdm2-Mdm4-p53 regulatory network', *Cancer Cell*, 9: 273-85.

Toyooka, Y., N. Tsunekawa, Y. Takahashi, Y. Matsui, M. Satoh, and T. Noce. 2000. 'Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development', *Mech Dev*, 93: 139-49.

Tran, H. T., J. D. Keen, M. Kricker, M. A. Resnick, and D. A. Gordenin. 1997. 'Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants', *Mol Cell Biol*, 17: 2859-65.

Tsodikov, O. V., J. H. Enzlin, O. D. Scharer, and T. Ellenberger. 2005. 'Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1', *Proc Natl Acad Sci U S A*, 102: 11236-41.

Tsui, V., and W. Crismani. 2019. 'The Fanconi Anemia Pathway and Fertility', *Trends Genet*, 35: 199-214.

Tucker, K. L., D. Talbot, M. A. Lee, H. Leonhardt, and R. Jaenisch. 1996. 'Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene', *Proc Natl Acad Sci U S A*, 93: 12920-5.

Uehara, R., S. M. Cerritelli, N. Hasin, K. Sakhuja, M. London, J. Iranzo, H. Chon, A. Grinberg, and R. J. Crouch. 2018. 'Two RNase H2 Mutants with Differential rNMP Processing Activity Reveal a Threshold of Ribonucleotide Tolerance for Embryonic Development', *Cell Rep*, 25: 1135-45 e5.

Urist, M., T. Tanaka, M. V. Poyurovsky, and C. Prives. 2004. 'p73 induction after DNA damage is regulated by checkpoint kinases Chk1 and Chk2', *Genes Dev*, 18: 3041-54.

van Heemst, D., and C. Heyting. 2000. 'Sister chromatid cohesion and recombination in meiosis', *Chromosoma*, 109: 10-26.

van Overbeek, M., and T. de Lange. 2006. 'Apollo, an Artemis-related nuclease, interacts with TRF2 and protects human telomeres in S phase', *Curr Biol*, 16: 1295-302.

Vleminckx, K., L. Vakaet, Jr., M. Mareel, W. Fiers, and F. van Roy. 1991. 'Genetic manipulation of E-cadherin expression by epithelial tumor cells reveals an invasion suppressor role', *Cell*, 66: 107-19.

Waisman, A., F. Sevlever, M. Elias Costa, M. S. Cosentino, S. G. Miriuka, A. C. Ventura, and A. S. Guberman. 2019. 'Cell cycle dynamics of mouse embryonic stem cells in the ground state and during transition to formative pluripotency', *Sci Rep*, 9: 8051.

Wan, L., J. Han, T. Liu, S. Dong, F. Xie, H. Chen, and J. Huang. 2013. 'Scaffolding protein SPIDR/KIAA0146 connects the Bloom syndrome helicase with homologous recombination repair', *Proc Natl Acad Sci U S A*, 110: 10646-51.
Wang, T., J. J. Wei, D. M. Sabatini, and E. S. Lander. 2014. 'Genetic screens in human cells using the CRISPR-Cas9 system', *Science*, 343: 80-4.
Watanabe, K., S. Tateishi, M. Kawasuji, T. Tsurimoto, H. Inoue, and M. Yamaizumi. 2004. 'Rad18 guides poleta to replication stalling sites through physical interaction and PCNA monoubiquitination', *EMBO J*, 23: 3886-96.
Watanabe, N., S. Mii, N. Asai, M. Asai, K. Niimi, K. Ushida, T. Kato, A. Enomoto, H. Ishii, M. Takahashi, and Y. Murakumo. 2013. 'The REV7 subunit of DNA polymerase zeta is essential for primordial germ cell maintenance in the mouse', *J Biol Chem*, 288: 10459-71.
Weber, S., D. Eckert, D. Nettersheim, A. J. Gillis, S. Schafer, P. Kuckenberg, J. Ehlermann, U. Werling, K. Biermann, L. H. Looijenga, and H. Schorle. 2010. 'Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance', *Biol Reprod*, 82: 214-23.
Wechsler, T., S. Newman, and S. C. West. 2011. 'Aberrant chromosome morphology in human cells defective for Holliday junction resolution', *Nature*, 471: 642-6.
Weeda, G., I. Donker, J. de Wit, H. Morreau, R. Janssens, C. J. Vissers, A. Nigg, H. van Steeg, D. Bootsma, and J. H. Hoeijmakers. 1997. 'Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence', *Curr Biol*, 7: 427-39.
West, J. A., S. R. Viswanathan, A. Yabuuchi, K. Cunniff, A. Takeuchi, I. H. Park, J. E. Sero, H. Zhu, A. Perez-Atayde, A. L. Frazier, M. A. Surani, and G. Q. Daley. 2009. 'A role for Lin28 in primordial germ-cell development and germ-cell malignancy', *Nature*, 460: 909-13.
Western, P. S., D. C. Miles, J. A. van den Bergen, M. Burton, and A. H. Sinclair. 2008. 'Dynamic regulation of mitotic arrest in fetal male germ cells', *Stem Cells*, 26: 339-47.
Westerveld, A., J. H. Hoeijmakers, M. van Duin, J. de Wit, H. Odijk, A. Pastink, R. D. Wood, and D. Bootsma. 1984. 'Molecular cloning of a human DNA repair gene', *Nature*, 310: 425-9.
Whittle, C. A., and C. G. Extavour. 2017. 'Causes and evolutionary consequences of primordial germ-cell specification mode in metazoans', *Proc Natl Acad Sci U S A*, 114: 5784-91.
Winter, J., Breinig, M., Heigwer, F., Brügemann, D., Leible, S., Pelz, O., Zhan, T., Boutros, M., 2016 'caRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens', Bioinformatics, 32(4): 632-4.
Wong, J. C., N. Alon, C. McKerlie, J. R. Huang, M. S. Meyn, and M. Buchwald. 2003. 'Targeted disruption of exons 1 to 6 of the Fanconi Anemia group A gene leads to growth retardation, strain-specific microphthalmia, meiotic defects and primordial germ cell hypoplasia', *Hum Mol Genet*, 12: 2063-76.
Wu, L., and I. D. Hickson. 2003. 'The Bloom's syndrome helicase suppresses crossing over during homologous recombination', *Nature*, 426: 870-4.
Xu, G., J. R. Chapman, I. Brandsma, J. Yuan, M. Mistrik, P. Bouwman, J. Bartkova, E. Gogola, D. Warmerdam, M. Barazas, J. E. Jaspers, K. Watanabe, M. Pieterse, A. Kersbergen, W. Sol, P. H. N. Celie, P. C. Schouten, B. van den Broek, A. Salman, M. Nieuwland, I. de Rink, J. de Ronde, K. Jalink, S. J. Boulton, J. Chen, D. C. van Gent, J. Bartek, J. Jonkers, P. Borst, and S.

Rottenberg. 2015. 'REV7 counteracts DNA double-strand break resection and affects PARP inhibition', *Nature*, 521: 541-44.
Xue, Y., Y. Li, R. Guo, C. Ling, and W. Wang. 2008. 'FANCM of the Fanconi anemia core complex is required for both monoubiquitination and DNA repair', *Hum Mol Genet*, 17: 1641-52.
Yabuta, Y., K. Kurimoto, Y. Ohinata, Y. Seki, and M. Saitou. 2006. 'Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling', *Biol Reprod*, 75: 705-16.
Yamaguchi, S., K. Hong, R. Liu, A. Inoue, L. Shen, K. Zhang, and Y. Zhang. 2013. 'Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming', *Cell Res*, 23: 329-39.
Yamaguchi, S., H. Kimura, M. Tada, N. Nakatsuji, and T. Tada. 2005. 'Nanog expression in mouse germ cell development', *Gene Expr Patterns*, 5: 639-46.
Yamaji, M., Y. Seki, K. Kurimoto, Y. Yabuta, M. Yuasa, M. Shigeta, K. Yamanaka, Y. Ohinata, and M. Saitou. 2008. 'Critical function of Prdm14 for the establishment of the germ cell lineage in mice', *Nat Genet*, 40: 1016-22.
Yang, S. W., A. B. Burgin, Jr., B. N. Huizenga, C. A. Robertson, K. C. Yao, and H. A. Nash. 1996. 'A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases', *Proc Natl Acad Sci U S A*, 93: 11534-9.
Yang, W., and Y. Gao. 2018. 'Translesion and Repair DNA Polymerases: Diverse Structure and Mechanism', *Annu Rev Biochem*, 87: 239-61.
Yeom, Y. I., G. Fuhrmann, C. E. Ovitt, A. Brehm, K. Ohbo, M. Gross, K. Hubner, and H. R. Scholer. 1996. 'Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells', *Development*, 122: 881-94.
Ying, Q. L., J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen, and A. Smith. 2008. 'The ground state of embryonic stem cell self-renewal', *Nature*, 453: 519-23.
Ying, Y., X. M. Liu, A. Marble, K. A. Lawson, and G. Q. Zhao. 2000. 'Requirement of Bmp8b for the generation of primordial germ cells in the mouse', *Mol Endocrinol*, 14: 1053-63.
Ying, Y., X. Qi, and G. Q. Zhao. 2001. 'Induction of primordial germ cells from murine epiblasts by synergistic action of BMP4 and BMP8B signaling pathways', *Proc Natl Acad Sci U S A*, 98: 7858-62.
Ying, Y., and G. Q. Zhao. 2001. 'Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse', *Dev Biol*, 232: 484-92.
Yonish-Rouach, E., D. Resnitzky, J. Lotem, L. Sachs, A. Kimchi, and M. Oren. 1991. 'Wild-type p53 induces apoptosis of myeloid leukaemic cells that is inhibited by interleukin-6', *Nature*, 352: 345-7.
Yoshikiyo, K., K. Kratz, K. Hirota, K. Nishihara, M. Takata, H. Kurumizaka, S. Horimoto, S. Takeda, and J. Jiricny. 2010. 'KIAA1018/FAN1 nuclease protects cells against genomic instability induced by interstrand cross-linking agents', *Proc Natl Acad Sci U S A*, 107: 21553-7.
Yuan, H., N. Li, D. Fu, J. Ren, J. Hui, J. Peng, Y. Liu, T. Qiu, M. Jiang, Q. Pan, Y. Han, X. Wang, Q. Li, and J. Qin. 2017. 'Histone methyltransferase SETD2 modulates alternative splicing to inhibit intestinal tumorigenesis', *J Clin Invest*, 127: 3375-91.

Yuan, J., and J. Chen. 2013. 'FIGNL1-containing protein complex is required for efficient homologous recombination repair', *Proc Natl Acad Sci U S A*, 110: 10640-5.

Zakharyevich, K., S. Tang, Y. Ma, and N. Hunter. 2012. 'Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase', *Cell*, 149: 334-47.

Zhang, B., J. Chang, M. Fu, J. Huang, R. Kashyap, E. Salavaggione, S. Jain, S. Kulkarni, M. A. Deardorff, M. L. Uzielli, D. Dorsett, D. C. Beebe, P. Y. Jay, R. O. Heuckeroth, I. Krantz, and J. Milbrandt. 2009. 'Dosage effects of cohesin regulatory factor PDS5 on mammalian development: implications for cohesinopathies', *PLoS One*, 4: e5232.

Zhang, B., S. Jain, H. Song, M. Fu, R. O. Heuckeroth, J. M. Erlich, P. Y. Jay, and J. Milbrandt. 2007. 'Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome', *Development*, 134: 3191-201.

Zhang, D., K. Zaugg, T. W. Mak, and S. J. Elledge. 2006. 'A role for the deubiquitinating enzyme USP28 in control of the DNA-damage response', *Cell*, 126: 529-42.

Zhang, J., J. M. Dewar, M. Budzowska, A. Motnenko, M. A. Cohn, and J. C. Walter. 2015. 'DNA interstrand cross-link repair requires replication-fork convergence', *Nat Struct Mol Biol*, 22: 242-7.

Zhang, Y., S. Xie, Y. Zhou, Y. Xie, P. Liu, M. Sun, H. Xiao, Y. Jin, X. Sun, Z. Chen, Q. Huang, and S. Chen. 2014. 'H3K36 histone methyltransferase Setd2 is required for murine embryonic stem cell differentiation toward endoderm', *Cell Rep*, 8: 1989-2002.

Zhou, W., E. A. Otto, A. Cluckey, R. Airik, T. W. Hurd, M. Chaki, K. Diaz, F. P. Lach, G. R. Bennett, H. Y. Gee, A. K. Ghosh, S. Natarajan, S. Thongthip, U. Veturi, S. J. Allen, S. Janssen, G. Ramaswami, J. Dixon, F. Burkhalter, M. Spoendlin, H. Moch, M. J. Mihatsch, J. Verine, R. Reade, H. Soliman, M. Godin, D. Kiss, G. Monga, G. Mazzucco, K. Amann, F. Artunc, R. C. Newland, T. Wiech, S. Zschiedrich, T. B. Huber, A. Friedl, G. G. Slaats, J. A. Joles, R. Goldschmeding, J. Washburn, R. H. Giles, S. Levy, A. Smogorzewska, and F. Hildebrandt. 2012. 'FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair', *Nat Genet*, 44: 910-5.