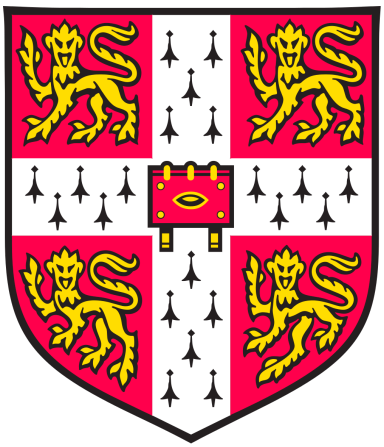
Loss of *rad51* in zebrafish (*Danio rerio*): a novel Fanconi anaemia model



Jan Gregor Botthof

Darwin College

University of Cambridge

May 2017

This dissertation is submitted for the degree of

Doctor of Philosophy

Loss of *rad51* in zebrafish (*Danio rerio*): a novel Fanconi anaemia model

Jan Botthof

Abstract

RAD51 is an indispensable homologous recombination protein, necessary for strand invasion and crossing over. It has recently been designated as a Fanconi anaemia (FA) gene, following the discovery of two patients carrying dominant negative mutations. FA is a hereditary DNA repair disorder characterised by various congenital abnormalities, progressive bone marrow failure and cancer predisposition. The cellular and molecular pathology of FA is poorly understood, resulting in a severe lack of effective treatment options. In this thesis, I describe the first viable vertebrate model of RAD51 loss. Phenotypic characterisation of zebrafish rad51 loss-of-function mutants showed that they develop key features of FA, including hypocellular kidney marrow, sensitivity to crosslinking agents and decreased size. Taking advantage of the unique properties of the zebrafish model, I show that some of these symptoms stem from both decreased proliferation, as well as increased apoptosis of embryonic haematopoietic stem and progenitor cells. Co-mutation of p53 was able to rescue the haematopoietic defects seen in the single mutants, but led to tumour development, underscoring the role of rad51 as a tumour suppressor. I further demonstrate that prolonged inflammatory stress can exacerbate the haematological impairment, leading to an additional decrease in kidney marrow cell numbers. In contrast, prolonged aldehyde-derived stress did not induce symptoms in the mutant fish. These findings strengthen the assignment of RAD51 as a Fanconi gene and provide more evidence for the notion that aberrant p53 signalling during embryogenesis leads to the haematological defects seen later in life in FA. It also strengthens the evidence for the involvement of haematopoietic stress, such as inflammation, in the development of bone marrow failure. Further research on this novel zebrafish FA model will lead to a deeper understanding of the molecular basis of bone marrow failure in FA and the cellular role of RAD51.

Declaration

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

It is not substantially the same as any 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. It does not exceed the prescribed word limit of 60000 words as prescribed by the Degree Committee for Clinical Medicine and Clinical Veterinary Medicine.

Some of the work presented in this thesis was kindly contributed by Ewa Bielczyk-Maczyńska (E.B.-M.). This data can be found in Figures 3.1C, 3.2, 3.3 and 4.1A. She was also responsible for generating the *rad51l1* mutant line.

Lauren Ferreira (L.F.) carried out the RNA-Seq workflow and Emmanouil Athanasiadis (E.A.) carried out the differential gene expression analysis.

Substantial parts of the work presented in this thesis have been published as "Loss of the homologous recombination gene *rad51* leads to Fanconi anemia-like symptoms in zebrafish" in Proceedings of the National Academy of Sciences of the United States of America. Published online before print on May 16th, 2017, doi:10.1073/pnas.1620631114.

Acknowledgements

Most of all, I would like to thank my primary supervisor, Ana Cvejić, for her support in developing the analytical, scientific and writing skills necessary to carry out the research presented here. Without her mentoring, I would not have matured into the scientist I am today. Moreover, I am grateful to my second supervisor Alan Warren, who provided several key ideas and insights that led to this work.

I would also like to thank all members of the Cvejić group for their encouragement and help. In particular, I am indebted to Ewa, who originally started the *rad51* project and provided a considerable amount of data. Moreover, I am thankful to Charlotte, who helped me in the initial stages of taking on this project. I would also like to acknowledge Manos' help with bioinformatics, as well as Lauren's and Shwethaa's experimental assistance. I am also grateful to Paulina, who proofread this manuscript and provided helpful feedback. I truly enjoyed working with all of you.

An additional thank you goes to the staff of the Sanger Institute fish facility, who have cared for my fish, and the staff of the Sanger Institute flow cytometry facility (Bee, Chris, Jenny and Sam), who have been vital in carrying out my FACS experiments. I would also like to thank Fengtang Yang for his help with chromosome spreads and Yvette Hooks for her help with histology.

Furthermore, I would like to acknowledge the generosity of the European Haematology association for funding my PhD studentship.

On a more personal level, I would like to thank my family, whose support and encouragement were vital in completing my academic studies. I thank my close friends (Beatrice, Carmen, Clemens, Daniel, Michael and Robert) for keeping me sane throughout both of my degrees. I would also like to thank all the artists and content producers who have helped make the countless hours of microscopy more bearable.

Finally, I would like to acknowledge all the people who have been unfairly omitted from this list or whose research has been overlooked in my text

Blut ist ein ganz besondrer Saft.

[Blood is juice of a very special kind.]

—"Mephistopheles" in *Faust I* by Johann Wolfgang von Goethe (1808)

Table of contents

Figures

Tables

Abbreviations

AAA+	ATPases associated with diverse cellular activities
AD	autosomal dominant
ADP	adenosine diphosphate
AGM	aorta-gonad mesonephros
ALDH	aldehyde dehydrogenase
ALM	Anterior lateral plate mesoderm
ALT	alternative lengthening of telomeres
AML	acute myeloid leukaemia
ANOVA	analysis of variance
AP	alkaline phosphatase
AR	autosomal recessive
ASCE	additional strand conserved E
ATP	adenosine triphosphate
ATR	ataxia telangiectasia and RAD3-related kinase
AV	annexin V
BIR	break-induced repair
BM	bone marrow
BMF	bone marrow failure
BMT	bone marrow transplantation
bp	basepair
BSA	bovine serum albumin
CHT	caudal haematopoietic tissue
CI	confidence interval
CPT	camptothecin

Abbreviations

CRISPR	Clustered regularly interspaced short palindromic repeats
DAMP	danger-associated molecular patterns
DAPI	4',6-diamidino-2-phenylindole
DEB	diepoxybutane
dHJ	double Holliday junction
DiQ	1,5-isoquinolinediole
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNAP	DNA polymerase
dpf	days post fertilisation
dpi	days post injection
dpt	days post transplantation
DSB	double-stranded break
dsDNA	double-stranded DNA
DTT	dithiothreitol
dUTP	deoxyuridine triphosphate
EGFP	enhanced GFP
ENU	N-ethyl-N-nitrosourea
ERCC	External RNA Controls Consortium
ES	embryonic stem
EtOH	ethanol
FA	Fanconi anaemia
FACS	fluorescence-activated cell sorting
FBS	foetal bovine serum
FSC	forward scatter
G-CSF	granulocyte-colony stimulating factor

GFP	green fluorescent protein
GO	gene ontology
GSH	glutathione-SH
GVHD	graft versus host disease
H&E	haematoxylin and eosin
HIV	human immunodeficiency virus
HJ	Holliday junction
HLA	human leukocyte antigen
hpf	hours post fertilisation
hpi	hours post injection
HPV	human papilloma virus
HR	homologous recombination
HSC	haematopoietic stem cell
HSCT	haematopoietic stem cell transplantation
HSPC	haematopoietic stem and progenitor cells
I	injected
ICL	interstrand crosslink
ICM	intermediary cell mass
IL	interleukin
IP	intraperitoneal
IS	immunostaining
ISH	*in-situ* hybridisation
IVF-PGD	*in-vitro* fertilisation-preimplantation genetic diagnostics
KASP	competitive allele-specific PCR
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	knockout

Abbreviations

LOF	loss of function
LPM	lateral plate mesoderm
LPS	lipopolysaccharide
LT-HSC	long-term haematopoietic stem cell
MAPK	mitogen activated protein kinase
MBT	mid-blastula transition
MDS	myelodysplastic syndrome
MeOH	methanol
MMC	mitomycin C
MPNST	malignant peripheral nerve sheath tumour
N2-ethyl-dG	N2-ethyl-2-deoxyguanosine
NBT/BCIP	nitro-blue tetrazolium and 5-bromo-4-chloro-3’-indolyphosphate
NER	nucleotide excision repair
NHEJ	non-homologous end joining
NI	non-injected
NK	natural killer
PAMP	pathogen-associated molecular pattern
PARP	poly ADP ribose polymerase
PB	peripheral blood
PBI	peripheral blood island
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Database
PdG	1,N2-propano-2-deoxyguanosine
PFA	paraformaldehyde
pH2AX	phospho-histone H2AX

PI	propidium iodide
pI:pC	polyinosinic:polycytidylic acid
PLM	posterior lateral mesoderm
PTW	phosphate buffered saline with Tween
QC	quality control
RNA	ribonucleic acid
RNS	reactive nitrogen species
ROS	reactive oxygen species
RT	room temperature
SCC	squamous cell carcinoma
SCT	stem cell transplantation
SD	standard deviation
SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SDSA	synthesis-dependent strand annealing
SEM	standard error of the mean
siRNA	small interfering RNA
SSC	side scatter
ssDNA	single-stranded DNA
TBST	tris-buffered saline with Tween
TGF	transforming growth factor
TLR	Toll-like receptor
TLS	translesion synthesis
TNF	tumour necrosis factor
TPM	transcripts per million
TUNEL	terminal deoxynucleotidyl transferase (TdT) dUTP nick-end labelling
VDA	ventral dorsal aorta

Abbreviations

WB	Western blot
WBC	white blood cell
WKM	whole kidney marrow
wpt	weeks post transplantation
WT	wild type
XLR	X-linked recessive
ZMP	zebrafish mutation project

1 Introduction

1.1 The history of Fanconi anaemia

The haematological disorder Fanconi anaemia was first described by the Swiss physician Guido Fanconi in 1927. In his initial paper called "Familiäre infantile perniziösartige Anämie (perniziöses Blutbild und Konstitution)" ["Familial infantile pernicious anaemia (pernicious blood count and constitution)"][1], he described a family with five children. Among them, three boys suffered from the disease. Fanconi's meticulous description already contained all the main features of the disease and suggested a hereditary cause – an exceptionally impressive feat considering the limitations of the day.

The affected boys showed what would later be considered typical congenital features of Fanconi anaemia, including microcephaly, café-au-lait-spots, hypoplastic testes and delayed development. The haematological symptoms of the patients consisted of anaemia with macrocytic erythrocytes and leukocytopenia, stemming from bone marrow dysfunction. Heavy bleeding, especially under the skin, was already apparent before anaemia developed, indicating early thrombocytopenia. All three boys died of their disease after a series of infections (the lack of leukocytes presumably made them susceptible to pathogens), culminating in bone marrow failure[1].

A further case reported by Uehlinger in 1929[2] made Fanconi realise that the disease affects all blood lineages, i.e. that it leads to pancytopenia. Later research led him to the erroneous conclusion that the malady is caused by chromosomal translocation[3], as the phenotype appeared to him to be too variable to be caused by one gene and affected families were not consanguineous. Subsequent discoveries disproved him, showing that the disease is caused by recessively inherited mutations in several genes (except for *FANCB*, which is X-linked[4] and some very rare dominant negative mutations).

In the early 1970s, the first steps toward a molecular understanding of the disease were underway. As shown by groups in the US and Japan, all Fanconi patients were sensitive to crosslinking agents, which leads to the accumulation of chromosomal damage[5,6].

It took a further two decades of study until the first FA complementation group was linked to a gene, namely *FANCC* [7]. However, this alone was not enough to explain the aetiology of the disease, as an understanding of the functional role of the Fanconi genes was lacking. This issue was resolved in the early 2000s, when the fundamental role of the Fanconi pathway in DNA repair was unravelled[8,9].

However, even though patients show strikingly increased DNA damage and the Fanconi genes play a major role in DNA repair, there are still divergent explanations as to what the major cause of bone marrow failure is. An important theory has long been that Fanconi anaemia cells are more sensitive to certain inflammatory cytokines[10,11], but other groups favour DNA damage by small aldehydes as the cause[12–15]. More recently, there has also been new evidence that Fanconi genes are involved in autophagy[16,17], telomere defects[18] and various other pathways. Overall, there is still no complete clarity about how all of these features are causally related to the observed gene mutations and how they lead to the congenital defects and bone marrow failure seen in Fanconi anaemia.

1.2 Symptoms

Despite being called Fanconi "anaemia", the symptoms of the disease are wide ranging and affect many more tissues than blood. In general, the symptoms can be classified into three different categories: congenital (i.e. birth) defects, haematological features and cancer predisposition.

1.2.1 Congenital features

About 75% of FA patients suffer from some form of physical abnormality from birth[19]. There is a wide variety of symptoms, with a large range in severity. The most common features include short stature, café-au-lait-spots on the skin and skeletal deformities, often affecting the thumbs and other fingers (See Figure 1.1 for example images of an FA patient and Table 1.1 for a comprehensive list of all known congenital symptoms and incidence rates). Typically, FA is diagnosed during early childhood. However, some individuals may not display any obvious symptoms until adulthood, where FA is occasionally discovered in people hypersensitive to chemo- and radiotherapy, but otherwise healthy[20,21]. Other patients suffer from severe disabilities throughout their life.

Intriguingly, children with severe congenital features tend to suffer from bone marrow failure earlier in life[22]. This inconsistency in the severity of symptoms has been linked to the high genetic variability of the disorder[22], which will be discussed later in this text.

Figure 1.1: FA patients have various congenital abnormalities. The images portrayed here are of a three-year-old toddler suffering from FA, showing typical symptoms. Visible on the left is the short stature for his age, a proportionally small head and a dislocated hip. In the middle, the small eyes and triangular face are apparent. In the top right image, hypo- and hyperpigmented areas "café-au-lait-spots" are visible. Shown in the bottom right are typical skeletal deformities affecting the hands, "dangling" thumbs. Reproduced with permission by Elsevier from Shimamura and Alter 2010[19].

Table 1.1: Summary of congenital features observed in FA patients. Based largely on Shimamura and Alter 2010[19], with additions indicated in the text. SD = standard deviation.

Abnormality	Details	Frequency
Low birth weight		5%
Microsomia	>60% are more than 2SD shorter than peers[23]. Only 10% of patients are taller than average.	40%
Skin	Generalised hyperpigmentation, café au lait spots, hypopigmented areas. Up to 12% of patients may develop Sweet's syndrome/acute neutrophilic dermatosis (red plaques and nodules).	40%
Skeletal: upper limbs, unilateral or bilateral	Thumbs: Absent or hypoplastic, bifid, duplicated, rudimentary, attached by a thread, triphalangeal, long, low set Radii: Absent or hypoplastic (only with abnormal thumbs), absent or weak pulse Hands: Flat thenar eminence, absent first metacarpal, clinodactyly, polydactyly Ulnae: Dysplastic, short	Thumbs: 35% Radii: 7% Hands: 5% Ulnae: 1%
Skeletal: Lower limbs	Feet: Toe syndactyly, abnormal toes, club feet Legs: Congenital hip dislocation	5%
Skeletal: Neck	Sprengel deformity, Klippel-Fiel anomaly, short, low hairline, webbed	1%
Skeletal: Spine	Spina bifida, scoliosis, hemivertebrae, abnormal ribs, coccygeal aplasia	2%
Craniofacial: Head	Microcephaly	20%
Craniofacial: Face	Triangular, birdlike, dysmorphic, micrognathia, mid-face hypoplasia	2%
Eyes	Small, cataracts, astigmatism, strabismus, epicanthal folds, hypo- and hypertelorism, ptosis	20%
Renal	Kidneys: horseshoe, ectopic or pelvic, abnormal, hypoplastic or dysplastic, absent; hydronephrosis or hydroureter	20%
Gonads: Males	Hypospadias, micropenis; undescended testes, absent testes. Fertility is reduced in males due to hypo or azoospermia.	25-64%[24]
Gonads: Females	bicornuate uterus, malposition, small ovaries. Pregnancy in females is possible, regardless of SCT status[25,26]	2%
Developmental delay	Intellectual disability, developmental delay	10%
Ears	Hearing loss (usually conductive secondary to middle ear bony anomalies); abnormal shape (dysplastic, atretic, narrow ear canal [i.e., external auditory meatus], abnormal pinna)	10-50%[27,28]
Cardiopulmonary	Congenital heart defect: patent ductus arteriosus, atrial septal defect, ventricular septal defect, coarctation of the aorta, truncus arteriosus, situs inversus	6%
Gastrointestinal	Oesophageal, duodenal, jejunal atresia; imperforate anus; tracheoesophageal fistula; annular pancreas; malrotation of the gut	5%
Central nervous system	Small pituitary, pituitary stalk interruption syndrome, absent corpus callosum, cerebellar hypoplasia, hydrocephalus, dilated ventricles	3%

One very important, but not immediately visible, defect in FA is the sensitivity of patient cells to crosslinking agents such as diepoxybutane (DEB) or mitomycin C (MMC) (Figure 1.2), which is an absolute diagnostic criterion[19].

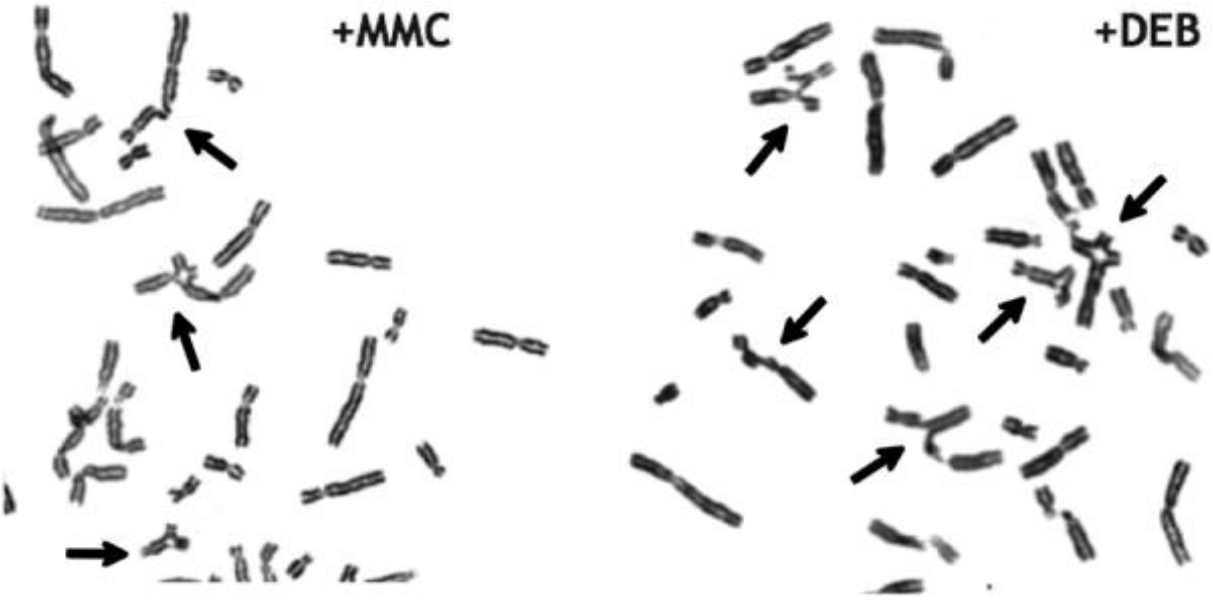


Figure 1.2: FA patients are sensitive to crosslinking agents. Shown are examples of FA patient chromosome spreads in response to MMC (left) and DEB (right). Visible characteristic are gaps, breaks and radial formations, which are pointed out with arrows. Reproduced with permission by Elsevier from Shimamura and Alter 2010[19].

1.2.2 Haematological features

Almost all FA patients develop haematological abnormalities at some point of their life. Progressive bone marrow failure (BMF, defined as the inability to make at least one blood lineage at levels required for survival) occurs in 90% of patients by age 40, with the median age of onset being about seven[29,30]. The range is very large however; babies may develop BMF[19,31], whereas some patients can be symptom free until adulthood[20,21,29], again highlighting the high variability of disease onset and severity of symptoms.

Early haematological signs of FA before complete BMF can be quite variable as well. More than 90% of FA patients develop macrocytosis early in life, which normally goes hand in hand with the aberrant expression of foetal haemoglobin[32]. Thrombocytopenia is also common and often occurs as the first haematological abnormality. It is frequently linked to macrocytosis as well[30]. Thrombocytopenia is often noticed first, due to the

spontaneous bleeding it can cause[32]. These uni-lineage cytopenias normally gradually develop into pancytopenia, followed by complete BMF[30]. On examination of the bone marrow, it usually appears hypocellular for the age of the patient, which is correlated with a decrease of the haematopoietic precursors. However, these cells typically appear morphologically normal, unless myelodysplastic syndrome (MDS) or acute myeloid leukaemia (AML) is developing[19].

1.2.3 FA and cancer

Patients who survive to adulthood without BMF, or have undergone stem cell transplantation, are at a high risk of developing both haematological, as well as solid malignancies. AML is particularly common, with a 500-fold increase in risk in FA patients compared to healthy controls[33–35]. The overall incidence of leukaemia and MDS reaches 9 and 7% respectively[36]. The cumulative incidence of leukaemia at age 40 is 33%[29] and the median age of onset is 11, a few years after the median onset of BMF[33]. Another study has estimated the mean age of development for leukaemias to be 14 and of MDS to be 17[37]. When looking at these numbers, it should be considered that most FA patients will develop BMF beforehand, i.e. they never progress to haematological malignancies solely because their original bone marrow is gone.

The risk of developing solid tumours is also increased markedly – the cumulative incidence of solid tumours reaches 28% at age 40[29]. This figure becomes even more dramatic if only patients alive at age 40 are considered – it goes up to 66%[29], as many patients die of other causes before developing solid tumours. The mean age of solid tumour development is 23[37]. Overall, BMF develops earliest, followed by leukaemias and solid cancers last[33]. The most common neoplasms in FA are squamous cell carcinomas (SCCs) of the head and neck, the anogenital region and the skin, followed by liver cancer and others[29]. Other cancers with significantly increased risk include cervical cancer (100-fold increase) and vulvar cancer (more than 1000-fold increase)[33,36,38].

As SCCs are the most common type of solid malignancy, a lot of research has been focussing on the causative factors in SCC development. Human papilloma virus (HPV) is a common cause of SCCs[39]. Some studies have suggested that FA patients are at an increased risk of HPV-linked SCCs[29], whereas others have found no such link[40].

However, HPV vaccination is still recommended for all FA affected children[32], especially as HPV infections are very common in FA patients[41].

Even though several of the FA genes are also breast cancer susceptibility genes (*FANCS/BRCA1*, *FANCD1/BRCA2*, *FANCN/PALB2, FANCJ/BRIP1* and *FANCO/RAD51C*), breast and ovarian cancer are rather rare in FA patients. This might be due to the hypogonadism seen in many FA patients[42].

1.3 Incidence

For a long time, the incidence of FA has been estimated to be 1:360000 live births[43]. This original model used a carrier frequency of 1:300 and an autosomal recessive mode of inheritance to determine this number. Newer data suggests a much higher carrier frequency in the United States of 1:181, leading to a predicted incidence of about 1:100000 live births[44].

Curiously, FA is more prevalent in males than females (1.2:1), a fact that cannot be solely explained by the X-linked inheritance of *FANCB* mutations[19].

In some communities, carrier frequencies are significantly higher than in the general population. Ashkenazi Jews have a carrier frequency greater than 1:100, with some unique mutations to this ethnic group due to several founder mutations[44–46]. Some sub-Saharan African ethnic groups also carry FA genes at a similar frequency, leading to an incidence of about 1:40000 live births, due to a founder mutation in Bantu-speaking populations[47]. Afrikaners (white South-Africans predominantly descended from Dutch and German settlers) also have a very high carrier frequency of about 1:80, which is thought to stem from a founder effect in the originally very small population[48]. The highest incidence of FA has been reported for a population of Spanish gypsies (carrier frequency 1:64-1:70). Like the other groups, this is due to a founder mutation about 600 years ago[49].

1.4 Prognosis

The median age of diagnosis for FA is 6.5 years[19], but as mentioned previously, adults do occasionally get diagnosed when they do not suffer from obvious congenital defects or

BMF[20,21,29]. The median survival time was 24 years in a 2003 study of the International Fanconi anaemia registry[29]. A 2010 study using the National Cancer Institute cohort found a significantly higher median survival of 33 years[35], with some patients living into their fifties. Indeed, one patient was only diagnosed at the age of 55 following their death from chemotherapy[50] and another patient was diagnosed at age 55 after being considered as a bone marrow donor for an affected sibling[19].

It is possible that some of the improvement in survival is due to better diagnosis of less severe cases, which might have been missed previously (a lead time bias). It cannot be discounted however, that the longer survival is at least partially a result of refined treatment methods, which will be discussed below.

1.5 Treatment

As mentioned previously, FA is a phenotypically extremely diverse disorder, with a wide range of symptoms. Because of this, the treatment of FA can never be limited to the haematological symptoms, but must involve the expertise of many specialists. This may include hand surgeons for repair of congenital bone defects or head and neck cancer specialists for tumour surveillance[21].

Currently, the decision to undertake treatment is mainly made on the basis of laboratory measurements of the blood. The present standard of care suggests that treatment is initiated when haemoglobin (Hb) < 8 g/dl, platelets < 30000/µl and neutrophil count < 500/µl. Treatment may be started earlier if problems such as anaemia, bleeding or infection occur[51].

Treatment modalities can be grouped into supportive therapies (trying to improve blood counts without addressing the underlying stem cell defect) and potentially curative therapies (repair or replacement of the HSCs) (Figure 1.3).

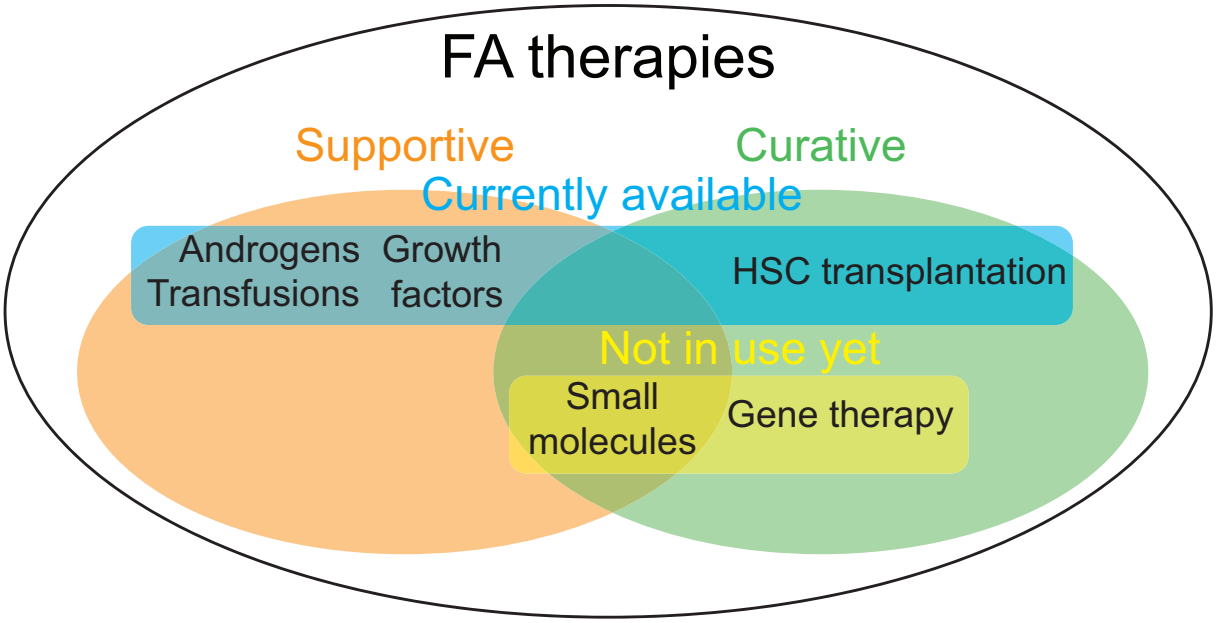


Figure 1.3: Summary of FA treatment approaches. Treatment modalities have been grouped into supportive (orange) and curative (green therapies). It is also indicated whether therapies are currently available (blue) or still being developed for clinical use (yellow).

1.5.1 Supportive therapies

1.5.1.1 Non-targeted therapies

The main supportive treatment modality are blood transfusions, which have been made possible by modern blood banking. This has helped greatly to deal with anaemia and bleeding caused by thrombocytopenia, but infections due to neutropenia are still a problem[52].

Another class of supportive drugs are haematopoietic growth factors, which stimulate the division of specific HSPC subsets[53]. These proteins have been used successfully to treat neutropenia, but granulocyte-colony stimulating factor (G-CSF) in particular has been associated with increased risk of leukaemia[54,55].

Androgens, such as oxymetholone, are used for the treatment of cytopenias in BMF in patients for which haematopoietic stem cell transplantation (HSCT) is not an option[56–59]. Similar to growth factors, they stimulate the production of mature blood cells[53]. This therapy has been linked to increased liver cancer and other side effects, such as virilisation and hirsutism[56]. In some patients, low-dose androgen therapy might

completely eliminate the need for a transplant[32]. However, as it decreases the chances of successful subsequent transplantation[60], it is more appropriate for patients in which this is not an option.

1.5.1.2 Targeted therapies/small molecules

Targeted therapies specifically aim to correct the pathways thought to be causative in BMF. They straddle the line between curative and supportive therapies – while they do not fix the underlying cause, they may eventually prevent BMF until the natural death of the patient. Small molecules have been used successfully to alleviate FA symptoms in mouse models by targeting oxidative stress[61,62]. Other potential drug targets that have been suggested are p38 kinase inhibition to target cytokine hypersensitivity[63,64] and transforming growth factor (TGF)β inhibition[65], which is particularly promising as inhibitors are in clinical trials for other diseases[66,67]. However, none of these treatments have reached routine clinical use yet.

1.5.2 Potentially curative therapies

1.5.2.1 Bone marrow transplantation

The only currently available curative therapy is haematopoietic stem cell transplantation (HSCT)[21,68]. The first transplantations of human leukocyte antigen (HLA)-identical donor marrow were carried out in the late 1970s and early 1980s[69]. This was followed by the development of cord-blood transplantations for FA in 1989[70]. Changes in the conditioning regimens have markedly improved survival to 94% five years post-transplantation[71] in a study of alternative donor HSCT, which is a major improvement compared to the ~30% survival in earlier trials[68,72]. This low success rate mainly stemmed from graft failure and graft versus host disease (GVHD)[68,71,73].

Another recent Dutch study also nicely illustrates how gradual improvements in therapy over the last 40 years have elevated survival post HSCT to over 90% in children (mean age 9.7 years)[74]. Particularly the introduction of the drug fludarabine to the standard conditioning regimen before transplantation considerably increased survival. The removal of irradiation was also helpful, as it is badly tolerated by FA patients[75]. Thus,

the current Dutch regimen of fludarabine and cyclophosphamide is highly successful and comparable to the recent American treatment regimens.

Even mismatched donor transplantation can have a success rate approaching 90% in FA affected children (mean age 11.1 years) if a more intensive treatment regimen is followed[76]. Adverse effects, such as GVHD were more common in this study. However, finding a suitable donor if no sibling is available limits the use of HSCT severely, especially as siblings need to be unaffected by FA[77,78]. Therefore, using HLA-mismatched donors is a good alternative for these patients.

An interesting advance to work around this limitation in the treatment of FA is the generation of so-called "saviour siblings". This refers to children from carrier couples selected via *in-vitro* fertilisation and pre-implantation genetic diagnosis (IVF-PGD) to be non-carriers. HSCs can be harvested from the umbilical cord of this unaffected child to treat its sick sibling[79]. However, apart from the ethical concerns with this technique, there are severe technical limitations. Most problematic is the extremely low IVF success rate per cycle (< 5%), as only healthy (and HLA matched) embryos can be used. This is particularly difficult to solve, as maternal age tends to be high, which increases the risk of aneuploidy[80].

Despite being a potentially very effective treatment, bone marrow transplantation (BMT) is not wholly without risks. Apart from the immediate side effects due to the conditioning regimen, there is some evidence that BMT increases the risk of tumour development[33,81]. There is also a risk of developing chronic or acute GVHD, immunodeficiency and various organ dysfunctions as a result of treatment[82].

The best time to transplant patients is currently being debated in the field. It is however agreed that it should be done before the development of MDS or AML, as it becomes considerably more challenging once malignant changes have occurred[77], but BMT itself also carries considerable risk. This is reflected in the low overall five-year survival in older patients ranging from 33 to 80% in several studies[83–86]. Some people even advocate HSCT before the development of BMF. A recent paper presented a decision analysis model to help deciding whether a patient should be pre-emptively transplanted[87]. This

model recommends pre-emptive transplantation only in children below five years of age, as later treatment-related mortality increases. This problem also applies to treatments seeking to delay BMT, such as androgen therapy, as older transplant recipients have worse outcomes[32,74]. A particularly striking example is a recent study which showed a five-year survival of only 54% in adult recipients (mean age 24), in stark contrast to the much higher survival discussed in younger recipients earlier in the text[88].

1.5.2.2 Gene therapy

In the future, the treatment of FA might become less dependent on the availability of donors, as gene therapy approaches become feasible and the patient's own cells can be used instead. FA should be very suitable for a gene therapy approach, as it is easy to identify which gene is defective in the patient and there is only one mutated gene per patient. Furthermore, FA is a haematological disease. This means HSCs can be easily obtained from the patient and transplanted back, enabling an *ex-vivo* gene therapy approach as an alternative to targeting cells while they are in the body[89]. The rapid advances in genome editing techniques such as the CRISPR-Cas9 system will probably lead to fast adoption in the clinic. Indeed, CRISPR-Cas9 has already been used to correct FA genes *in-vitro*[90].

Other efforts have focussed on correcting the mutations by supplying FA cells with a working copy of the mutant gene[91–98]. An especially interesting recent publication integrated *FANCA* into the AAVS1 locus[98] and transduced this construct into patient-derived fibroblasts using a retroviral vector. The cells were then reprogrammed to induced pluripotent stem cells (iPSCs), followed by differentiation into HSCs. Even though these FA fibroblasts were harder to de-differentiate to iPSCs than "normal" cells, they showed comparable expression and methylation of pluripotency genes. Furthermore, the iPSCs were able to form teratomas in mice, a common test for pluripotency. Patient-derived reprogrammed HSCs could potentially be transplanted back into the patient. Interestingly, FA cells are normally refractory to reprogramming, indicating that the FA pathway is important in maintaining "stemness"[99,100]. Reprogramming fibroblasts would be a good option if a patient does not have enough HSCs left to correct, which has caused issues with other gene therapy approaches[101,102]. However, retroviral vectors have considerable problems with transducing HSCs. This

issue was very problematic in a trial of gene therapy for FA[91], leading to the treatment being relatively ineffective in alleviating patient symptoms. Furthermore, there are safety concerns about the vectors[91], making genome editing based approaches more attractive in comparison. It remains to be seen whether more recently generated, improved retroviral vectors show better efficacy in the clinic [92,93].

HSC numbers in each patient will be very important. As $CD34^+$ HSCs may be depleted very early, it is possible to have too few for gene correction[101]. This might lead to two different treatment approaches: if enough $CD34^+$ cells are available, it would be preferable to fix the FA gene deficiency in these cells directly. In patients where this is not the case, reprogramming with an intermediate iPSC step might be the only possible solution, although reprogramming clearly introduces additional time, cost and safety considerations.

1.5.3 Novel challenges and future perspective

The current treatment regimens for FA are insufficient, as they suffer from lack of long-term effectivity (growth factors), have severe side effects (androgens) or are only possible when a suitable donor is found (HSCT). Because of this, a better understanding of the molecular mechanisms of the disease are necessary to develop better therapeutics until gene therapy becomes a widespread, cost-effective treatment method.

Improvements in the treatment of BMF will lead to new therapeutic challenges. As the lifespan of FA patients increases, the management of long-term complications stemming from HSCT and cancers in non-haematological tissues increases[103]. Cancer treatments are particularly problematic, as chemo- or radiation therapy are often tolerated badly or lead to more malignancies due to the DNA damage sensitivity of the patients[75]. As these long-term complications become more common, new treatments will need to be developed.

1.6 Fanconi anaemia genes

There are currently 21 FA genes recognised in the Fanconi Anaemia Mutation Database[104] (Table 1.2). Most patients have mutations in a subset of FA genes, which make up the majority of FA mutations. Particularly *FANCA* and *FANCC* mutations are

common, whereas some other complementation groups only have one or two known patients (See Table 1.2 and Figure 1.4). The main function of the FA genes is described below in the description of interstrand crosslink (ICL) repair by the FA pathway gene products.

Table 1.2: Fanconi genes in humans. Shown are their alternative names, chromosomal location, inheritance mode and patient number. These statistics are mainly based on the LOVD database[105], adding in the newly discovered *FANCV* and *U* genes. AD = autosomal dominant, AR = autosomal recessive, XLR = X-linked recessive.

Gene	Alternate name	Chromosome /band	Inheritance	Patients (total number)	Patients (%)
FANCA		16/q24.3	AR	1752	63.4
FANCB		X/p22.2	XLR	40	2.0
FANCC		9/q22.3	AR	323	11.4
FANCD1	*BRCA2*	13/q12-q13	AR	73	2.4
FANCD2		3/p25.3	AR	85	2.7
FANCE		6/p22-p21	AR	48	2.1
FANCF		11/p15	AR	47	1.5
FANCG	*XRCC9*	9/p13	AR	262	8.7
FANCI		15/q26.1	AR	32	1.4
FANCJ	*BRIP1, BACH1*	17/q22.2	AR	57	1.9
FANCL		2/p16.1	AR	12	0.5
FANCM		14/q21.3	AR	8	0.3
FANCN	*PALB2*	16/p12.1	AR	18	0.5
FANCO	*RAD51C*	17/q25.1	AR	6	0.1
FANCP	*SLX4*	16/p13.3	AR	14	0.4
FANCQ	*ERCC4*	16/p13.3	AR	4	0.1
FANCR	*RAD51*	15/q15.1	AD	2	0.1
FANCS	*BRCA1*	17/q21.31	AR	1	0.1
FANCT	*UBE2T*	1/q32.1	AR	6	0.2
FANCU	*ERCC2*	7/q36.1	AR	1	0.1
FANCV	*REV7*	1/p36.22	AR	1	0.1

1.6.1 Types of FA genes

The Fanconi anaemia genes can be classified into two groups: *bona-fide* FA genes, where at least two patients developed BMF during the course of their disease and FA-like genes, which recapitulate the main features of FA, but where no patient with BMF has been observed[106]. Four of the genes can be classified as FA-like: *FANCO*/*RAD51C*,

FANCR/RAD51, *FANCS/BRCA1* and *FANCU/XRCC2*. However, as there are very few patients for each of these genotypes (as few as one for *FANCS* and *FANCU*), it cannot be concluded that mutations in these genes invariably will not cause BMF, as many FA patients of other complementation groups never progress to BMF, or do so very late in life.

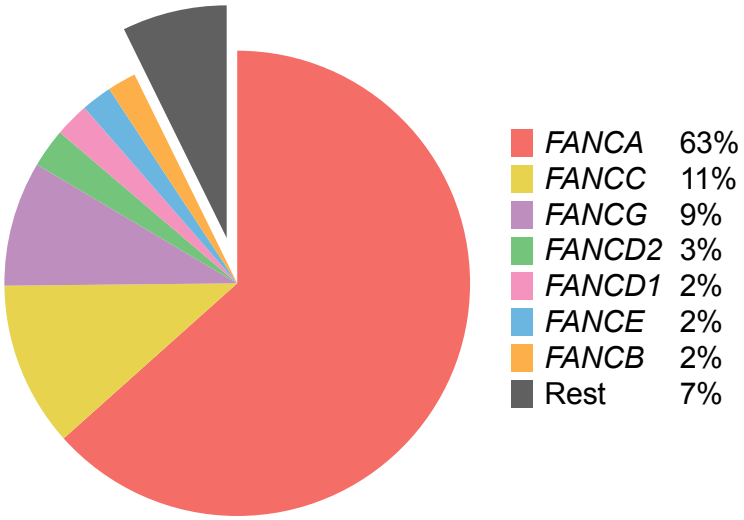


Figure 1.4: Mutations in a subset of Fanconi genes cause almost all cases of FA. This pie chart demonstrates the predominance of *FANCA*, *FANCC* and *FANCG* among FA patients. The data is based on Table 1.2. Shown individually in this chart are all complementation groups mutated in more than two percent of all patients.

1.6.2 Assignment of FA genes

Typically, a FA gene is assigned if the wild type version can correct the phenotype (i.e. features such as crosslinker sensitivity) in patient-derived cells. This process is not always straightforward, as illustrated by the case of *FANCM*. Even though it was initially described as a *bona-fide* FA gene[107] and sits at the core of the FA complex[108], it has recently been cast into doubt as to whether it actually qualifies as FA gene at all. This is because the patient in which it was described also carries homozygous *FANCA* mutations[109], complementation of which can rescue the phenotype. Furthermore, there are homozygous carriers for loss of function *FANCM* mutations who do not show any symptoms of FA (albeit increased risk for breast cancer)[110]. Because of this, it has been proposed to exclude it from the list of FA genes[106]. Nevertheless, *Fancm* mutant mice recapitulate many features of FA[111]. Together with its central role in the FA core complex,

this suggests that mutations in this gene at least have the potential to cause an FA phenotype.

A similar situation previously led to wrong assignment of the complementation group FANC-H, which turned out to be caused by mutations in *FANCA*[112]. This also explains the lack of a *FANCH* gene.

1.6.3 Genotype-phenotype variability

The large number of genes and the variations of different mutations within each gene means that the genetic causes of FA are very heterogeneous. This contributes to the variety of phenotypes seen. The influence of different gene mutations is particularly visible when the distribution of FA patients in the different mutation groups is considered. For example, *FANCD1*/*BRCA2* mutant patients make up around 2% of all FA patients. From the carrier frequency in the general population, it can be estimated that if all homozygous *FANCD1* mutant embryos survived to birth, the proportion of the *FANCD1* group would make up about 50% of all FA patients[22]. Other FA mutations are also found much less commonly than would be expected if they were compatible with life. There is only one known patient to survive with homozygous *FANCS*/*BRCA1* mutations[113], indicating that complete loss-of-function mutations are usually lethal and only very specific, hypomorphic mutations can be tolerated and still lead to the development of FA. Similarly, there are only two patients with *FANCR*/*RAD51* mutations[114,115], with very specific mutations, as other mutations presumably lead to an early embryonic death.

However, if a patient survives to term, there is considerable evidence that it is less important which gene is mutated, but rather what kind of mutation the patient carries – point mutations are usually less severe, whereas frameshifts or deletions tend to be very detrimental[22]. This is because point mutations are more likely to be hypomorphic, rather than complete loss-of-function mutations in comparison to frameshifts and deletions. An example of the variable phenotypes can be seen in patients carrying *FANCC* mutations. Patients with mutations in intron 4 (IVS4+4A>T) or point mutations in exon 14 (R458X and L554P) have a significantly worse prognosis, with a median onset of BMF at 2.7 and 2.1 years of age respectively. This is very early in comparison to other

FANCC mutations, which have a median onset of BMF around age seven, similar to other complementation groups[116].

The situation is further complicated by the fact that even in siblings with identical mutations, symptoms can be quite disparate, illustrating the importance of modifiers within the genetic background[22]. The effect of other alleles on the phenotype of FA patients is well demonstrated for the *ALDH2* gene, described in-depth in section 5.1.2.4. Another example for non-FA genes influencing disease phenotype are glutathione S-transferase (*GSTM1*) polymorphisms, which have been linked to considerably earlier (three vs. seven years) BMF in FANCC patients[117].

There are some FA mutations that lead to an especially increased risk of cancer development, particularly genes of the *BRCA* network[22]. Almost 100% of patients with *FANCD1*/*BRCA2* or *FANCN*/*PALB2* develop malignancies (mostly leukaemias) during their early childhood, before BMF[118–121]. Furthermore, *FANCS*/*BRCA1*, *FANCD1*/*BRCA2*, *FANCN*/*PALB2, FANCJ*/*BRIP1* and *FANCO*/*RAD51C* are breast cancer susceptibility genes when heterozygous[122–126]. Apart from these associations, there is little evidence that particular FA gene mutations lead to specific tumours in the different complementation groups.

1.7 Cellular pathology and bone marrow failure in Fanconi anaemia

Even though FA has been recognised as a disease for almost a century, there are still surprisingly many questions about why the majority of patients eventually suffer from BMF. Many hypotheses exist about the underlying cause, but most of them accept that the HSPCs are exhausted. The debate centres on what the main factor is in this stem cell depletion. FA has been likened to a premature ageing disorder, as many of the defects discussed below (e.g. cytokine overproduction and defective mitochondria) also commonly occur during normal ageing[127]. In particular, it shares many features with the premature ageing and BMF syndrome dyskeratosis congenita, including many of the congenital abnormalities and blood defects[127]. It has also been likened to a segmental progeroid syndrome of the blood[22]. The discussion in this section will be limited to the

haematopoietic defects contributing to eventual exhaustion of the HSPCs and will not cover congenital symptoms. Some of the main hypotheses proposed to explain the defects seen in FA are summarised in Figure 1.5.

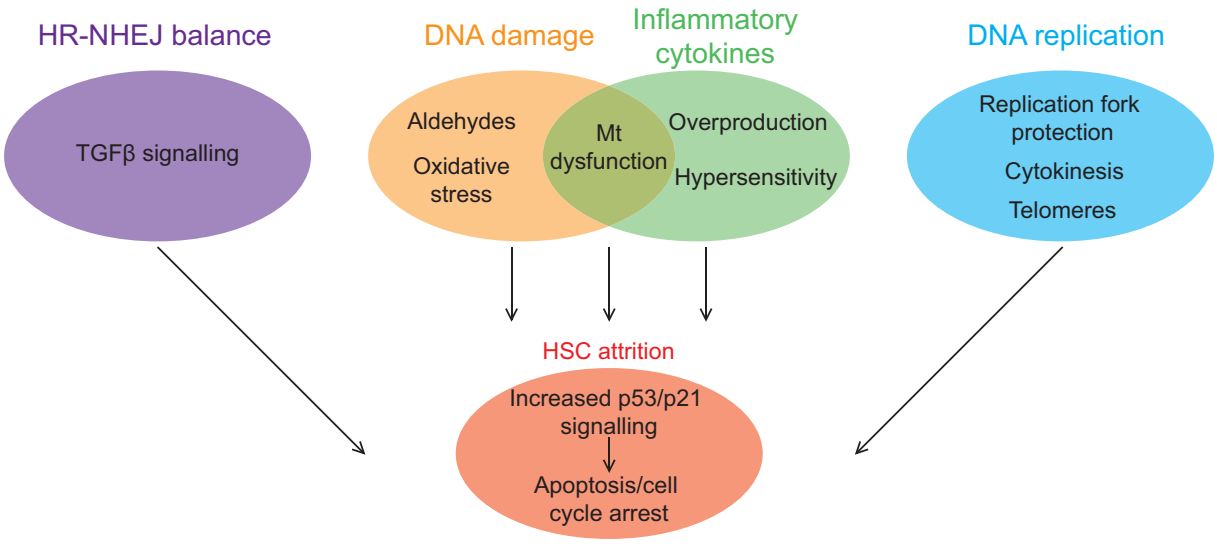


Figure 1.5: Summary of the different mechanisms that have been proposed as the trigger of BMF in FA. The different hypotheses are shown here as a Venn diagram. Mitochondrial (Mt) dysfunction has been linked to both DNA damage, as well as increased cytokine expression and therefore represents an overlap between those two options. All of these signals are thought to trigger apoptosis and cell cycle arrest, leading to HSC exhaustion.

1.7.1 Fanconi anaemia is caused by a HSPC defect

All blood lineages are affected in FA, which indicates that there is a general HSPC dysfunction, rather than a defect in differentiation of a uni/oligopotent progenitor[103]. Moreover, FA patients generally display low $CD34^+$ cell numbers in the bone marrow[91,101]. This is further corroborated by FA mouse models, many of which also show decreased HSPC numbers[15,62,128–131] and decreased engraftment ability upon transplantation[132,133]. Forcing HSPCs to enter the cell cycle by various means is able to cause BMF in mouse models[134]. Further evidence comes from cell culture models. FA patient-derived fibroblasts are refractory to reprogramming[99,100], suggesting that they cannot maintain pluripotency. Similarly, knocking down FA genes in embryonic stem (ES) cells leads to lineage defects[135], indicating FA genes are necessary to maintain stemness and multilineage potential.

1.7.2 HSPC defects start during embryonic/foetal development

It is normal for HSPC numbers to decrease during normal ageing[133], but elderly people do not usually suffer from BMF (albeit BMF is more common in elderly people than in children), in stark contrast to FA patients who almost invariably suffer from BMF during early childhood. This difference could be explained by three different possibilities. The first option is that patients start out with a decreased number of stem cells, which are depleted at the same rate as in healthy humans. The second option is that FA patients start with normal HSC numbers, but that the rate of depletion is increased, either to changes to the HSCs themselves or to their microenvironment. Finally, patients may suffer from a combination of these factors. The second option appears unlikely, as bone marrow cellularity is already reduced very shortly after birth[133]. In addition, foetal liver HSPCs derived from *Fancd2*$^{-/-}$ *and Fanc*$^{-/-}$ mice show decreased engraftment abilities when transplanted[136,137]. Taken together, these papers suggest that the haematopoietic defects seen in FA already start during embryonic haematopoiesis.

1.7.3 Signalling pathways mediating HSPC defects

Experiments looking at the foetal livers of Fanconi mice elucidated that the p53/p21 axis plays an important role in the lowered number of HSPCs. Levels of p21 are dramatically increased, suggesting an early proliferation defect leading to G0/G1 arrest[133]. Cultured *Fancg*$^{-/-}$ HSPCs also engraft significantly less well, a defect which can be rescued upon *p53* silencing[133]. However, knocking out *p21* in *Fancd2*$^{-/-}$ mice cannot rescue the haematopoietic phenotype (neither engraftment following transplantation nor proliferation defects were rescued), but in fact exacerbates it[138]. This argues against the role of p21, but it is possible that another p53 target mediates the defects.

Another connection between p53 and the FA pathway is that p53 actively downregulates the transcription of FA pathway genes[139]. This could lead to a harmful vicious circle. DNA damage induces p53, leading to the repression of FA genes. As the damage cannot be repaired due to the lack of repair proteins, the induction of p53 is increased further, exacerbating the problem rather than solving it. In combination, the upregulation of p53 in response to defective DNA repair[133] and the downregulation of the pathway in response to p53 suggests a positive feedback loop[139].

Even though the role of p53 is generally accepted by most of the field, there is evidence to the contrary. A study looking at foetal defects in *Fancd2*$^{-/-}$ mice found that while there are indeed proliferation defects, these are mediated via p38 instead of p53[137], in stark contrast to the results obtained by Ceccaldi *et al.*[133]. It is possible that the use of different FA models (*Fancg* and *Fancd2*) contributed to these differences, as well as different genetic backgrounds. Nevertheless, both models still predict decreased proliferation and increased apoptosis in the early embryonic HSCs within the foetal liver.

Nevertheless, it is clear that foetal FA HSPCs already show various impairments. As described above, FA has been likened to a premature ageing syndrome[127], so a comparison between the defects in aged and FA HSPCs is warranted. Indeed, a p53-dependent response to DNA damage and cellular stress has been implicated in HSC ageing in humans and mice[140–142]. These defects can cause genomic instability, BMF and haematopoietic malignancies in the elderly[140,141], akin to what is seen much earlier in FA patients. It may be that HSC number is not dependent on successful DNA repair, but their functionality is. These damaged HSCs are fine until they are forced to leave quiescence. This leads to their eventual exhaustion, as their self-renewal capacity is impaired. Such defects are especially problematic in times of stress where HSCs have to leave quiescence to replenish the blood[134,140]. In agreement with this, HSCs with hyperactive p53 are more similar to aged cells and have impaired functional properties. Moreover, the overall number of HSCs in p53 hyperactive mice is decreased and the remaining cells have to excessively proliferate to maintain working haematopoiesis[143].

DNA repair in general seems to be required to maintain the pool of HSPCs, as *Lig4*$^{-/-}$ (a gene in the non-homologous end joining (NHEJ) pathway) mice show decreased bone marrow cellularity, coupled with increased proliferation to maintain blood production[144]. It seems as if increased proliferation in response to reduced HSPC compartment is a general reaction in mice, which leads to a vicious circle. The already diminished HSPCs have to enter the cell cycle more often, thus slowly completely depleting the pool.

While it is clear that FA HSPCs are impaired in their functionality, it is less clear how the pathways leading to these defects are triggered. There are many hypotheses, the most common of which are discussed below. Broadly speaking, they can be grouped into

mechanisms where DNA damage is the primary cause of HSC exhaustion and those in which DNA damage is just coincident with other cellular defects, which are the ultimate trigger of cell death.

1.7.4 Aldehyde hypersensitivity

As discussed previously, the main (and best studied) function of the Fanconi pathway is the removal of ICLs from DNA. It is uncertain however what damaging agent is the predominant cause of ICLs in FA patients. The Patel group in Cambridge has accumulated considerable evidence in the past few years that the main damaging molecules are small aldehydes, which are generated during normal metabolism and are found in many dietary sources[145]. The evidence for aldehyde hypersensitivity of FA cells stems from cell culture experiments[12,14,146,147], mouse models[13–15,131] and epidemiological data[148,149]. The details of this hypothesis are outlined in detail in chapter 4, where I consider the effect of aldehyde-induced stress on my zebrafish FA model.

1.7.5 Oxidative stress

Even though there is a lot of genetic evidence for the role of aldehydes in causing DNA damage in FA, they are not the only damaging agent that can cause ICLs. There is considerable evidence that the FA pathway also serves to remove lesions induced by oxidation. The FA pathway assembles in response to oxidative stress[150] and some FA proteins are directly involved in mechanisms to reduce oxidative stress[151–153]. There is cell culture evidence that oxygen can cause chromosomal aberrations in FA cell lines[154] and that FA cells grow better when cultured in hypoxic conditions[155], which are disadvantageous to normal cells. Furthermore, FA cells gain more mutations when cultured under conditions of oxidative stress[156]. Additionally, double mutant *Fancc*$^{-/-}$, *Sod*$^{-/-}$ (superoxide dismutase) mice show hypocellular bone marrow and erythro/leukocytopenia, indicating that these mutations can cooperate to induce a stronger phenotype[157]. However, the number of phenotypic HSCs in these double mutants is normal, even though they show clear functional defects. Birth defects are also not observed, in contrast to the phenotypes seen in FA mice also carrying mutations in aldehyde metabolising enzymes. Oxidative stress overproduction in FA cells has been linked to mitochondrial defects[158] and lack of some FA proteins has been associated with

mitochondrial dysfunction[159]. This connects oxidative stress with another hypothesis on why FA HSCs get depleted: defective autophagy.

1.7.6 Autophagy defects

Autophagy is the process by which a cell recycles its unnecessary cellular constituents, both to gain energy in times of starvation and to remove damaged components. In particular, damaged mitochondria have to be removed from the cell. Mitochondria sustain a lot of damage, as they continually generate oxidative stress in the form of reactive oxygen species (ROS) as a side effect from the electron transport chain, which necessitates a turnover within about 10-25 days. Elimination of defective mitochondria is important, because otherwise mitochondrial DNA damage accumulates. This leads to a cell full of dysfunctional mitochondria, eventual bioenergetics failure and death of the cell. The importance of this process has led to the term “mitophagy” for the autophagic removal of mitochondria[160].

Defects in autophagy can lead to HSC attrition via the accumulation of damaged mitochondria and increased oxidative stress, which can eventually lead to bone marrow failure[161]. Indeed, the importance of autophagy in the maintenance of HSC functionality during ageing has been highlighted in a recent publication[162], which showed that defects in autophagy lead to impaired energy metabolism and are associated with loss of quiescence and stemness.

Evidence for mitochondrial dysfunction in FA comes from the fact that FA cells have a shift in metabolism towards glycolysis at the expense of oxidative phosphorylation[163]. Many Fancd2 interacting proteins are localised to mitochondria, suggesting a function in that compartment[164]. In addition to that, several FA genes (*FANCC*, *FANCF* and *FANCI*) were recently identified as potential autophagy regulators in a genome-wide small interfering RNA (siRNA) screen[165]. This discovery was initially surprising, as the canonical function of FA genes occurs solely in the nucleus. However, a combination of cell culture and *in-vivo* work showed that essentially all FA genes also have a cytosolic role in facilitating autophagy[16]. Another group validated these results and showed that defective autophagy in FA is due to defective mitochondrial fission[17]. Loss of mitophagy led to both increased oxidative stress via inflammosome activation, as well as ROS

generation, since defective mitochondria were not removed efficiently[16]. Furthermore, the involvement of mitochondria in cell death[166] could be an additional factor leading to HSC depletion in FA. These facts are especially intriguing, because this indicates that defective autophagy is both linked to increased oxidative stress, as well as increased inflammation; it could thus serve as the connecting point between the different hypotheses.

1.7.7 Abnormal inflammatory signalling

Inflammation is defined as a response to abnormal conditions, such as tissue damage or infection. Chronic inflammation in particular is linked to tissue dysfunction. One of the important features of inflammation is the local production of inflammatory cytokines by tissue-resident cells such as macrophages, which serve to recruit more immune cells to the site of inflammation and other things[167]. FA pathogenesis has been linked to inflammatory cytokines in multiple publications. There are two complementary aspects to this; FA bone marrow cells have shown both cytokine hypersensitivity[65,134,168,169], as well as increased production of inflammatory cytokines[11,170–174]. Both of these mechanisms might cooperate to induce BMF by increasing HSC cycling. The details of this are discussed in chapter 4, where I investigate the role of inflammatory stress in BMF.

1.7.8 Telomere defects in Fanconi anaemia

Telomere defects are common in diseases involving aplastic anaemia, with one third of all acquired aplastic anaemia patients showing shortened telomeres[18]. In agreement with this general trend, FA patients often display shortened telomeres, which tend to be about 50% as long as the average telomeres of healthy people of the same age. This puts them in the shorter, but still normal range of telomere length, in contrast to dyskeratosis congenital (DC) patients, who have much shorter telomeres yet[175]. However, while patients do show clear telomere defects, mouse models of FA generally fail to recapitulate this. Neither *Fancc*, *Fancg* nor *Fancj* mutant mice show any telomere shortening when unchallenged[176–178]. The fact that mouse telomeres tend to be considerably longer than human telomeres[177], as well as the shorter lifespan of mice may be contributing factors to this difference. The role of FA genes in alternative lengthening

of telomeres (ALT) was demonstrated using a gene knockdown approach. The researchers used siRNA in cultured cells relying on ALT to lengthen telomeres to show that at least *FANCA, FANCC* and *FANCD2* are necessary for ALT to occur. Another FA protein with an important role in telomere preservation is FANCP/SLX4, which participates in the negative regulation of telomere length, regulation of telomere recombination and the prevention of telomere defects[179–181]. Moreover, all the FA genes that are also homologous recombination (HR) genes are required for telomere replication and capping[182], so loss of HR proteins should affect telomeres even more than loss of core FA proteins. Cells lacking *RAD51* have shortened telomeres, as double stranded breaks remain unrepaired and hence telomeric DNA is lost. This is thought to be as replication forks have trouble progressing through telomeric DNA, leading to stalled replication forks[183,184]. The HR protein RAD51-BRCA1/2 axis is necessary for replication fork protection in conjunction with FANCD2, linking the FA pathway and HR in replication fork protection, as discussed in more depth below[185]. Consistent with this, HR-deficient cells also have more fragile telomeres[183,184]. It was previously thought that RAD51 also plays a role in the ALT mechanism of telomere lengthening, but a recent publication disputes this[186]. Overall, FA proteins play a role both in the ALT mechanism, as well as normal telomere lengthening.

1.7.9 Replication fork protection

The FA pathway and the RAD51-BRCA1/2 axis are also involved in normal replication fork protection. The first observations suggesting this possibility were undertaken when FA cells were treated with hydroxyurea, which does not cause ICLs, but does cause replication fork stalling, leading to the activation of the FA pathway[187]. Fittingly, FA pathway deficient cell lines[185], as well as cells derived from *Fancc*$^{-/-}$ mice[188] show increased instability at stalled replication forks. It has been shown that both the Fanconi core complex, including FANCD2-I are involved with fork stabilisation[185], as well as the HR proteins BRCA2 and RAD51[185,189]. As deficiency of the core complex genes can be compensated by overexpression of RAD51[185], it is likely that the core complex recruits BRCA1/2 and RAD51 as in ICL repair and the RAD51 protein complex then protects the fork from excess processing by nucleases, such as DNA2/WRN[114,189]. Interestingly, FANCD2 recruits the nuclease FAN1 via its ubiquitin modification[190,191], which was

previously thought to be necessary for ICL repair to stalled replication forks, where it is indispensable for fork stabilisation rather than crosslink removal[192,193]. This is corroborated by several earlier papers showing that FAN1 is not strictly necessary for ICL repair[194–196] and by the fact that FAN1 loss in humans and mice does not lead to Fanconi anaemia, but rather a kidney disease called karyomegalic interstitial nephritis[192,195]. It is currently unknown why the kidney is more sensitive to the loss of FAN1 than other tissues[195].

Other Fanconi proteins also play a role in the resolution of stalled replication forks; FANCJ is necessary for replication fork restarting[197]. This suggests that the whole pathway is involved with this process, as well as ICL repair.

1.7.10 Cytokinesis

Apart from protecting stalled replication forks, FA proteins also play other roles during cell division. Indeed, cytokinesis failure contributing to BMF has been observed in FA cells[198]. Several FA proteins have been shown to protect chromosomal fragile sites[187]. The FA proteins cooperate with bloom syndrome protein (BLM) to repair damaged chromosomes during mitosis and promote the efficient separation of the sister chromatids[199,200]. Furthermore, FA proteins can also associate with the mitotic apparatus, where they are necessary to prevent aneuploidy[201].

1.7.11 Aberrant TGFβ signalling

The FA pathway has extensive crosstalk with other DNA repair mechanisms[202]. Most importantly, it has been associated with suppression of the error-prone DNA repair mechanism NHEJ. Suppression of this pathway in FA-deficient cells can rescue some of the defects normally seen[203–205]. It has recently been shown that this switch to NHEJ is mediated via increased TGFβ signalling. In addition, compounds inhibiting TGFβ signalling can alleviate many of the defects seen in FA HSPCs, most importantly the aberrant p53 signalling. Interestingly, blocking TGFβ also rescued defects seen in FA cells upon inflammatory stress and aldehyde mediated DNA damage, suggesting that the TGFβ pathway is a shared mediator leading to stem cell depletion[65]. This indicates that TGFβ might be a master regulator in BMF and an attractive drug target[206]. Early

mouse studies on FA mice have already started, showing that two radiation mitigator drugs have the potential to modulate TGFβ induction[207].

1.7.12 Additional functions of the Fanconi anaemia pathway

The FA pathway is also involved in other cellular mechanisms, for which there is less evidence than the pathways described above. The FA complex has been implicated in the removal of aberrant DNA:RNA hybrids (R-loops)[208]. There is also limited evidence that FA cells are impaired in the epigenetic regulation of gene expression, similar to cancer-related changes[209].

1.7.13 Summary

FA proteins are involved in a multitude of cellular processes, many of which have been proposed to trigger the p53/p21 signalling that ultimately causes BMF (Figure 1.5). It is currently unclear which one of these factors (if any) is dominant in patients and eventually causes the depletion of HSCs in FA patients. However, multiple mechanisms common to several of the proposed causes of BMF have started to emerge, adding some pieces to the puzzle of why the bone marrow fails in FA patients. A better understanding of these changes will be necessary to develop novel treatments to lengthen the lifespan without BMF, potentially negating the need for costly and arduous BMT.

1.8 The Fanconi anaemia pathway

The DNA in each cell is constantly being damaged by various forms of insults, ranging from ultraviolet radiation to small oxygen radicals[210], leading to up to 10^5 lesions per cell per day[211]. One particularly problematic form of damage are ICLs, in which both strands of DNA are connected by covalent bonds, rather than the normal hydrogen bonds between the bases[212]. ICLs may lead to DNA breakages or chromosomal rearrangements if they are not repaired correctly[213]. A single ICL can potentially kill a cell if it cannot be repaired or bypassed. This is especially challenging, as ICLs are hard to repair due to the involvement of both strands of the DNA double helix[214]. This poses a huge challenge during DNA replication, because both strands need to be separated to act as a template[212]. As outlined above, one of the main hypotheses as to what causes ICLs in FA are small aldehydes, stemming from normal metabolic reactions. The Patel group in

Cambridge has done a lot of work on FA mouse models, and shown that stress caused by acetaldehyde and formaldehyde may be responsible for many of the Fanconi features[12–15]. Alcohol may be an especially big contributor, as it is metabolised to acetaldehyde[145]. FA patients carrying hypomorphic or null mutations in the more active mitochondrial aldehyde dehydrogenase (as is common in many Asian populations) tend to have stronger symptoms for this reason[148,149,215].

The main pathway to remove ICLs completely is via the FA pathway, coupled to replication during S-phase[216]. Other repair mechanisms that are available to the cell, such as nucleotide excision repair (NER), cannot fully remove ICLs[217], with the sole exception of the NEIL3 pathway. This glycosylase has recently been shown in *Xenopus* egg extracts to be able to remove at least some forms of ICLs (primarily less bulky ones), whereas the FA pathway in conjunction with HR is able to remove all types of ICLs[218]. Because of this new data on NEIL3, it has been proposed that other DNA glycosylases might also be able to remove certain types of ICLs[219]. Indeed, a recent publication highlighted another DNA glycosylase in bacteria that can also remove ICLs, showing the potential of this class of enzymes to participate in ICL repair[220]. As exciting as these new discoveries may be, this section will deal exclusively with the mechanisms of the canonical FA pathway.

There are four basic steps of repair using the FA pathway[221] (Figure 1.6). First, the lesion has to be recognised. This step is carried out by the FA core complex. Subsequently, the DNA is incised, which is done by genes related to nucleotide excision repair. The newly made gap in the DNA is now filled and the crosslink bypassed by translesion DNA polymerases. Finally, the strands are connected back together and the lesion removed by homologous recombination.

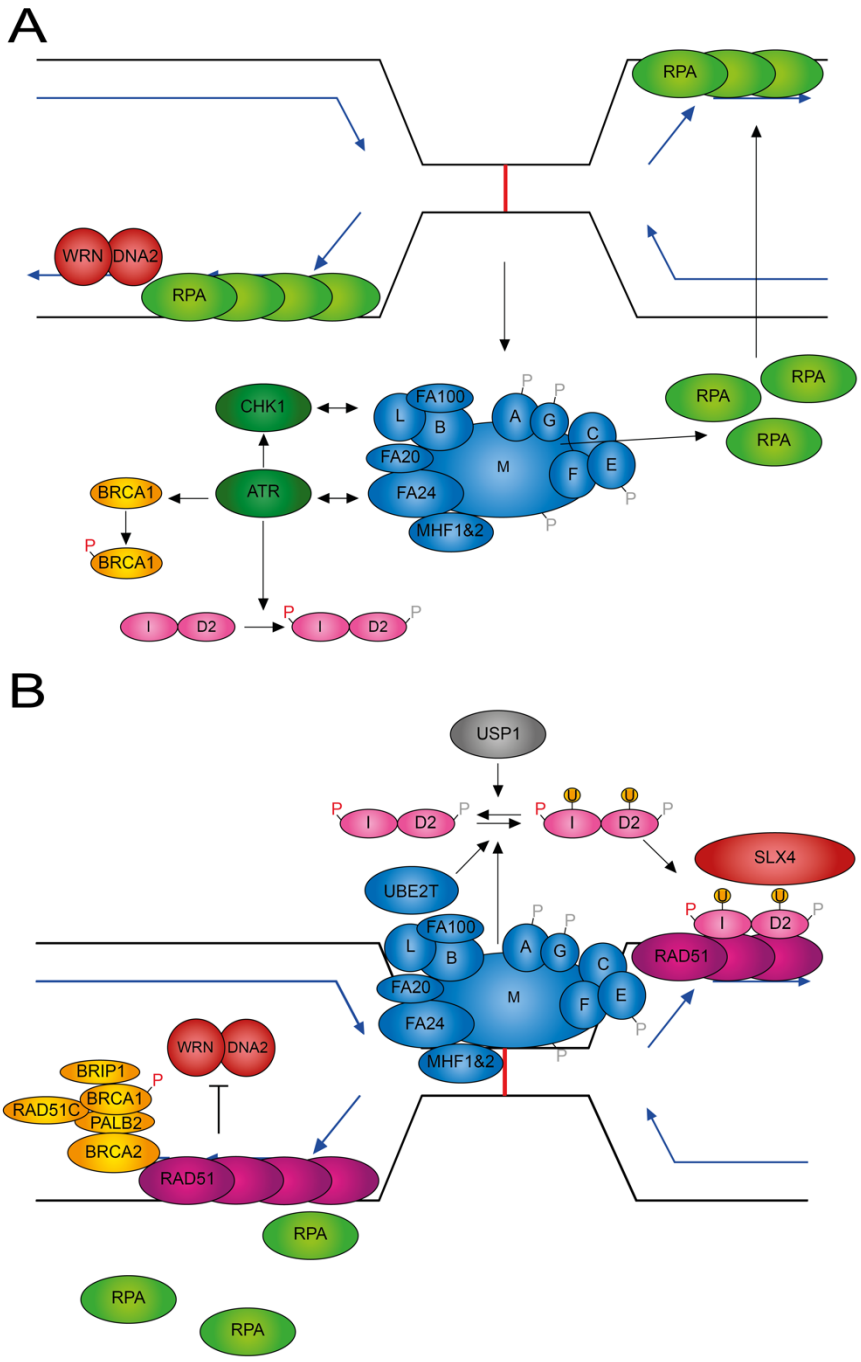
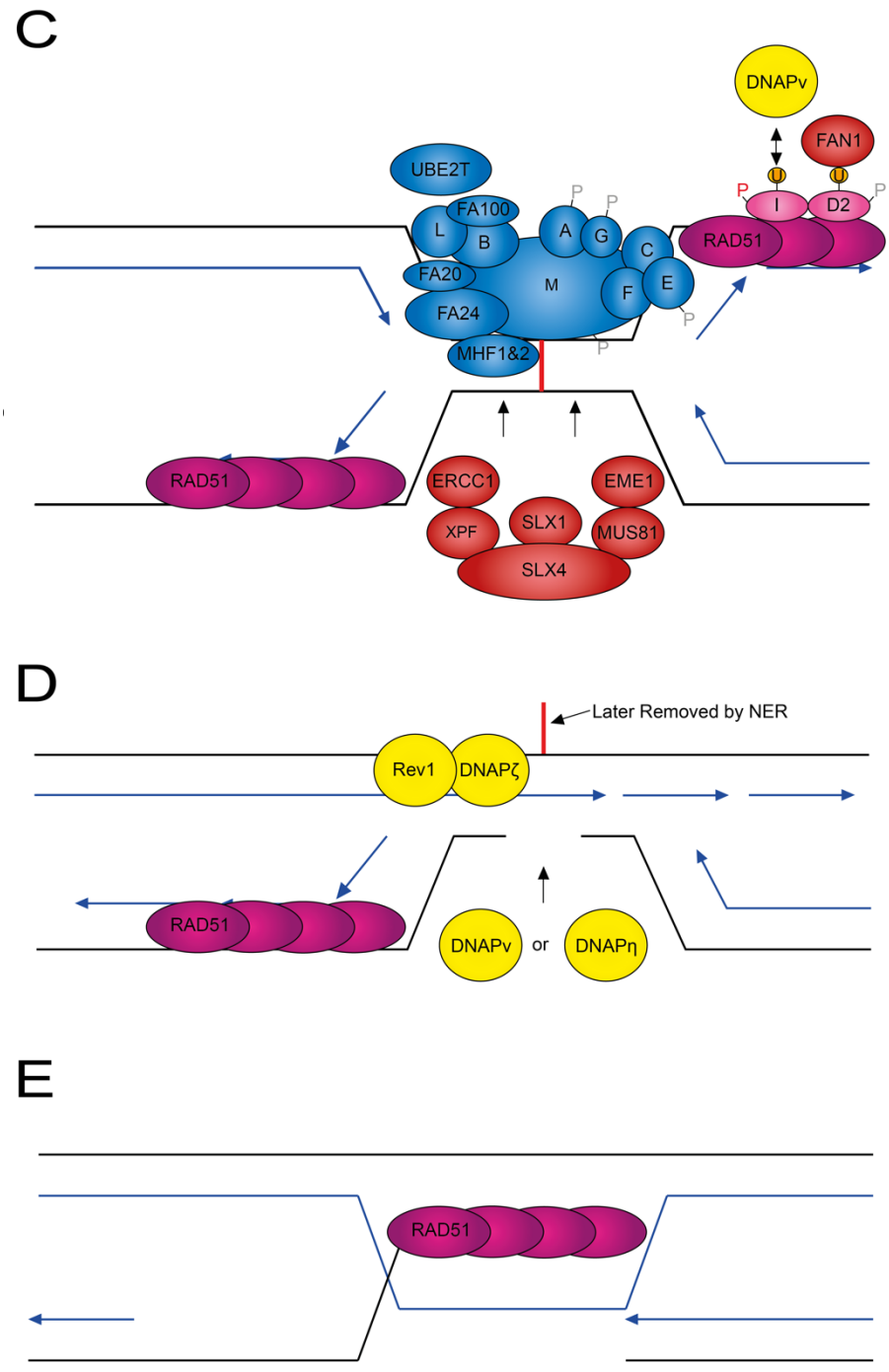


Figure 1.6: Overview of ICL repair by the FA pathway. Recognition of the lesion and activation of the FA pathway (A). FANCM recognises the lesion (red line connecting the DNA strands) and activates RPA filament formation and ATR activation. ATR in turn phosphorylates FANCD2-I, FANCS/BRCA1 and members of the FA core complex. Phosphorylation patterns are based on information from Wang *et al.* 2008[222]. Vital phosphates are shown in deep red, optional ones in grey. At the same time, nucleases such as WRN and DNA2, process the ends of the DNA for later HR. In the next step, the core complex monoubiquitylates FANCD2, in association with the E2 ubiquitin-conjugating enzyme FANCT/UBE2T (B). This process is reversible by the enzyme USP1. FANCD2-I binds to DNA and stabilises the RAD51 filament, whose formation was catalysed by a complex containing FANCS/BRCA1 and FANCD1/BRCA2.



This filament blocks excess processing of the DNA ends by DNA2. Once FANCD2-I has been ubiquitylated, it triggers nucleolytic incision by the SLX tri-nuclease complex (C), which cuts the DNA strand around the ICL. As soon as the DNA is incised, translesion synthesis mediated by several DNA polymerases can repair one strand across the lesion (D), leading to one intact DNA strand and another strand with a double-stranded break. In the final step of ICL repair, HR mediated among other things by RAD51 (E), repairs the second strand of the DNA.

1.8.1 Interstrand crosslink removal

1.8.1.1 Crosslink recognition

The FA pathway is triggered when two replication forks collide at an ICL during DNA replication in S-phase[223,224]. FANCM recognises the lesion, binds the crosslink and acts as an anchor for the rest of the FA core complex (Figure 1.6A). Because of this, it can be considered the "FA signalosome anchor"[108]. In addition to the FA proteins recruited subsequently, it is constitutively bound to its partners FAAP24/FA24 and MHF1-2[225–227]. Apart from functioning as the signalosome anchor, FANCM also triggers an ataxia telangiectasia and RAD3-related (ATR) kinase response and CHK1 dependent signalling[228–230]. Targets of ATR and CHK1 include FANCM[231] itself, as well as other Fanconi proteins, such as FANCI[232] and FANCE[233]. FANCM is also able to recruit RPA, which smooths out the DNA, to the lesion[234]. Once FANCM is bound, the rest of the core complex assembles, which is a mixture of bona-fide FA proteins and accessory proteins. There are 14 currently known members of the core complex (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM, FANCT, FAAP20/FA20, FAAP24/FA24, FAAP100/FA100, MHF1 and MHF2)[235]. The whole core complex acts as a scaffold for the E2 ubiquitin-conjugating enzyme FANCT/UBE2T and the E3 ubiquitin ligase FANCL, which monoubiquitylates the FANCD2-FANCI heterodimer on the FANCD2 subunit[52,236–239] (Figure 1.6B). FANCB and FAAP100 are bound to FANCL and modulate its activity[107,240,241]. FANCE's function during this process appears to be the direct binding of FANCD2 and holding it in place for the reaction to occur[8,242–244]. The role of many other FA genes in FANCD2 ubiquitylation is poorly defined, even that of the most commonly mutated gene in FA, FANCA[245]. FANCI may also be ubiquitylated, which might increase the efficiency of the signalling pathway, but is not essential[246]. It is unclear whether this happens at the same time as FANCD2 modification and whether the same proteins are responsible[245]. It has however been proposed that a dimeric FANCB-FAAP100-FANCL complex monoubiquitylates the FANCD2-I complex in a symmetric concurrent manner[245,247]. The ubiquitin on FANCD2 can be removed by the USP1-UAF1 heterodimer, adding extra regulation[248,249]. There are other modifications that happen to the FANCD2-I dimer apart from ubiquitin modification. Before this ubiquitylation event, FANCD2-I is also phosphorylated by ATR[246,250]. This

phosphorylation might be necessary for the localisation to chromatin[246]. Besides these modifications, sumoylation of the complex has also been reported. This appears to regulate removal of the complex from chromatin, to limit incisions to the site of damage and decrease off-target cuts[251]. As can be seen from the stringent regulation of this phase, FANCD2-I activation is a key step in controlling ICL repair. Next, the FANCD2-I heterodimer binds the DNA at the stalled replication fork and triggers the next phase of ICL repair: nucleolytic incision[216].

1.8.1.2 Nucleolytic incision

SLX4/FANCP is recruited to the chromatin, which directly binds the ubiquitin on FANCD2[252,253]. In turn, FANCP recruits several structure-specific nucleases. These include the ERCC4 (FANCQ/XPF)-ERCC1 complex[254,255], MUS81-EME1[256] and SLX1[256]. New data suggests that SLX4, SLX1, MUS81-EME1 and XPF-ERCC1 form the SMX tri-nuclease, which functions as complex responsible for nucleolytic incision during ICL repair[257]. In this complex, MUS81-EME1 seems to be the main nuclease, whereas SLX4 and XPF-ERCC1 tend to have more of a regulatory role in replication fork cleavage. SLX4's role seems to be to relax the substrate specificity of MUS81-EME1 and XPF-ERCC1 stimulates cleavage in an unknown way, independent of its nuclease activity[257]. Cleavage around the ICL by MUS81-EME1 leads to a double-stranded gap in the DNA (Figure 1.6C).

At this step, the nuclease FAN1 binds FANCD2 as well to protect the replication fork, but its role in ICL removal itself is disputed[192,193]. This situation is described in detail in section 1.7.9 on the role of FA proteins in replication fork protection.

1.8.1.3 Translesion synthesis

The next step in the repair process is to replicate the DNA around the lesion, a process termed "translesion synthesis"[235,258] (Figure 1.6D). There are three steps to this process: approach, insertion and extension. Approach refers to the DNA synthesis up to, but not including, the nucleotide opposite the lesion. The approach steps are catalysed by REV1 and DNAPζ (REV3-7)[259]. REV1 is a deoxycytidyl transferase[260,261] and plays an important role in regulating translesion synthesis (TLS) polymerase activity[262,263].

Insertion refers to the insertion of a nucleotide directly opposite to the lesion. The nuclease catalysing the insertion is still unknown, but DNAPη[264] and DNAPν, which binds RAD51 and FANCD2-ubiquitin, have been suggested[114,115]. This step is very error-prone and can reach up to 1% misincorporated nucleotides[265].

The last step of translesion synthesis, extension, is catalysed by DNAPζ[264,265]. This results in an intact DNA strand, which can be used as a template for homologous recombination (Figure 1.6E), as well as one strand containing a double-stranded break. The importance of this step has recently been underlined by the discovery of a Fanconi patient carrying biallelic *REV7* mutations, leading to the designation of *REV7* as *FANCV*[266].

1.9 Homologous recombination

Homologous recombination (HR) is the last phase of ICL repair, as well as a major pathway to repair double-stranded DNA (dsDNA) breaks outside of ICL repair[216]. The process can be divided into three fundamental phases: the presynaptic phase, the synaptic phase and the postsynaptic phase[267]. In the presynaptic phase, the DNA is prepared for RAD51 filament formation and the RAD51 monomers are assembled into a filament. In the synaptic phase, homologous DNA is searched and strand exchange catalysed by RAD51. In the postsynaptic phase, the gaps left by strand exchange are filled, leading to the completion of DNA repair. Note that some of the proteins involved differ when HR is used for other purposes, such as meiosis. This section will only discuss HR in the context of ICL repair. Rather than occurring after the processes mediated via the FA core complex, the early steps of HR have been shown to happen at the same time as some steps in the Fanconi pathway[114] (Figure 1.6).

1.9.1 Presynaptic phase

The first step of HR is the 5'-3' resection started by the MRE11-RAD50-NBS1 (MRN) complex in conjunction with CTIP and further continued by EXO1, BLM, WRN and DNA2[268] (Figure 1.6A). The multitude of proteins involved stems from the fact that there are two different pathways that can lead to resected DNA, with either EXO1 or DNA2 acting as the nuclease[269], however for the purposes of ICL repair, these are equivalent. The newly formed single-stranded DNA (ssDNA) is bound by RPA, which serves to

straighten out the DNA[270]. The resecting nucleases can resect the DNA too much, impairing DNA repair. Initially this was thought to be caused by MRE11, but now it is known to be mediated by DNA2[114]. While these nuclease complexes incise the DNA from both sides, RAD51 protects the DNA from excess processing from the nucleases, meaning it must already be coating the DNA beforehand[114,115,185,189] (Figure 1.6B). Proteins necessary for loading RAD51 onto the DNA include BRCA1 and 2 (FANCS and FANCD1), FANCN/PALB2 and FANCJ/BRIP1, which cooperate to load it into the ssDNA, replacing RPA in the process[271–274]. This underscores the importance of replication fork protection for the successful ICL repair, as lack of either of these proteins can cause FA. In the cytoplasm, RAD51 is usually found in oligomeric form, bound to BRCA2[275]. The following model has been proposed for BRCA2's role in loading RAD51 onto the DNA[276]: BRCA2 disrupts RAD51 oligomers by binding it via its BRC repeats. In this step, the TR2 region of BRCA2 is phosphorylated. Once a DSB has been detected, BRCA2 is dephosphorylated and RAD51 loaded onto the DNA. The interaction of the dephosphorylated TR2 region with the RAD51 filaments stabilises the newly formed filament structure[277–279]. Once HR has finished, BRCA2 is phosphorylated completely, helping with the disassembly of the RAD51 filament. The FA proteins FANCJ/BRIP1 and FANCN/PALB2 are important in this process by facilitating BRCA1 functionality and the BRCA1-BRCA2 interaction respectively. BRCA1 is itself necessary to recruit BRCA2 to the DNA[271–273,280]. BODL1 is also required to stabilise RAD51 at these forks and if lost, confers crosslinking sensitivity on the cell[281]. This process requires several accessory proteins, which positively and negatively regulate RAD51 function by mediating strand assembly or stabilising the RAD51 filament. These include RAD51B, FANCO/RAD51C, RAD51D, FANCU/XRCC2, XRCC3, RAD52, SWS1, SWSAP1, RAD54, RAD54B, SWI5, and SFR[282], as well as the ones mentioned above in the text. These interactions demonstrate the high amount or regulation required to carry out this process[282]. The importance of FA proteins in regulating this process is clear when the number involved in RAD51 filament formation and stabilisation is considered. Overall, resection results in single-stranded ends ready for the next phase, synapsis.

1.9.2 Synaptic phase

RAD51 molecules coat the ssDNA to form a nucleoprotein filament[283] that stretches the DNA out and promotes strand invasion into a homologous duplex strand, followed by strand exchange (synapsis)[284] (Figure 1.6E). Renkawitz *et al.* have proposed a unified model for how RAD51 searches for homologous DNA in the cell that they have termed "accelerated random search model"[267]. In this model, the RAD51 coated strand starts out by probing the surrounding DNA on the broken chromosome by undergoing very short strand invasions. If no homologous DNA is found, the DNA further away is probed. However, probing efficiency decreases as a function of distance very rapidly, as regions further away are encountered much less frequently by chance. This process might be accelerated by sliding of the RAD51 filament along the DNA, as the filament would not constantly have to dissociate and associate to probe large tracts of DNA. Sampling of multiple non-contiguous segments of DNA by one filament may also help with speeding up the process. Using such a process would ensure that homology search by RAD51 is efficient, as the DNA close to the break is searched first, but the mechanism can still cover a large swath of the genome. Once homologous DNA is found, strand exchange can occur. Strand exchange leads to a DNA formation called the "D-loop", in which the invading strand primes DNA synthesis[285]. As soon as RAD51 has fulfilled its function, it hydrolyses ATP to ADP, which leads to a conformational shift allowing dissociation from the DNA[286,287].

1.9.3 Postsynaptic phase

The postsynaptic phase, also termed the resolution phase, begins with DNA synthesis from the broken 3' end formed during the synaptic phase. The resulting structure can either be resolved via a double Holliday junction (dHJ) with crossover or non-crossover products, or the synthesis-dependent, strand annealing (SDSA) pathway, which prevents crossovers. Alternatively, a new replication fork can also be formed in a process called break-induced replication (BIR), which has the disadvantage that it can lead to loss of heterozygosity[285]. There are usually no crossovers in somatic cells, whereas they are common during meiosis[285]. In yeast cells, dHJ formation is blocked by Rad51, which biases the process to the SDSA pathway, where no crossovers occur[288]. This agrees with

experimental evidence showing that the SDSA pathway is preferred *in-vivo*[289]. However, dHJs can occasionally occur in somatic cells nonetheless[290].

1.9.3.1 Double Holliday junction resolution

The dHJ resolution model (Figure 1.7A), first proposed in this form by Szostak *et al.* in 1983[291], is the classical model of how HR is finished. After strand invasion, the invading strand primes DNA synthesis from its 3' end. D-loop extension allows the displaced complementary strand to anneal and its 3' end now also serves as a starting point for replication. The ends of the newly made DNA are now ligated, which leads to the formation of the dHJ. Ultimately, several nucleases can lead to the resolution of the dHJ, resulting in either crossover or non-crossover products. As both strands of DNA prime DNA synthesis here, the mechanism uses semi-conservative DNA replication[292].

1.9.3.2 Synthesis-dependent strand annealing

The basic process of SDSA (Figure 1.7B) was first described in a T4 Phage system by Formosa and Alberts in 1986[293]. As in dHJ resolution, it starts with extension of the invading strand at its 3' end. However, instead of also extending the displaced complementary strand at this step, the invading strand is further extended, leading to a translocating D-loop. The first strand being elongated is thereby displaced from its template strand by branch migration, as the "DNA bubble" moves along the DNA during extension. The second end is captured when the complementary ssDNA of the first strand is exposed by branch migration, which then anneals to the tail of the second strand as the D-loop collapses. The DNA on this strand can now be completely replicated and the breaks ligated. This process conserves the intact strand completely, thus newly made DNA is only on the repaired strand and the process is conservative[292]. This also prevents crossovers, unless second end capture happens by the second strand annealing to the D-loop, which can form another dHJ. However, this alternative way of completing SDSA and the resulting crossovers are extremely rare[289], so SDSA usually produces non-crossover products.

1.9.3.3 Break-induced replication

The BIR pathway (Figure 1.7C) of resolving HR intermediates can occur when one of the resected DSBs is homologous to the intact DNA. Here, the invading DNA strand serves

as starting point for a new complete replication fork, which has a leading and lagging strand as usually in the cell[294], whereas the other HR resolution methods only use the machinery for the leading strand. It is semi-conservative, like dHJ resolution[292] and requires the resolution of a single Holliday junction (HJ) to be completed[295].

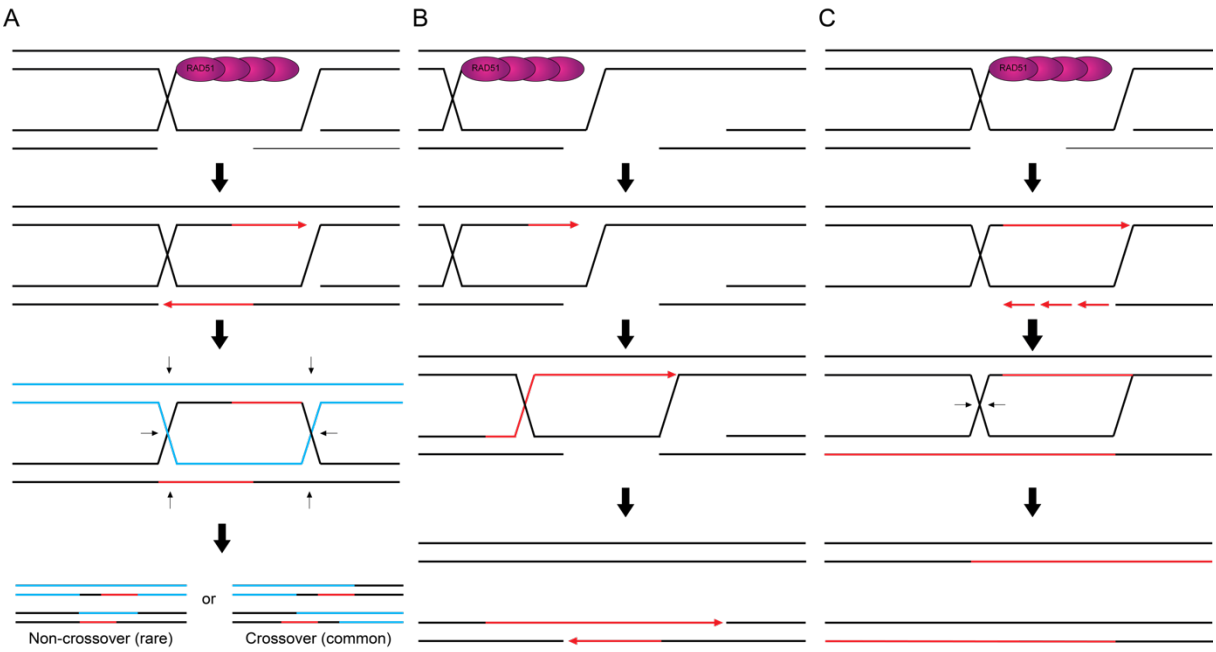


Figure 1.7: Ways of resolving HR. Newly made DNA is shown in red. (A) Classical dHJ resolution. A RAD51-coated DNA strand invades a homologous DNA duplex. A D-loop forms and is extended until the other broken strand can anneal, followed by the extension of this second strand. This extension leads to the formation of a dHJ, which can be resolved by cutting at several sites, indicated by different arrows. Depending on how the dHJ is cut (and the junction translocated), there can be crossover and non-crossover products. As indicated in the products, newly made DNA can be found on both new chromosomes, indicating the semi-conservative nature of this process. (B) SDSA pathway. As in dHJ resolution, RAD51-coated DNA invades a homologous duplex and forms a D-loop. However, in contrast to dHJ resolution, the D-loop is extended much longer, leading to a migrating DNA bubble. The newly made DNA can then hybridise with the broken end on the other DNA strand, acting as a template for repair of the other side of the break. Thus, newly made DNA is only found on one chromosome, making this way of repair conservative. (C) BIR pathway. Strand invasion triggers the generation of two completely new replication forks, with leading and lagging strand DNA synthesis. In the end, a single HJ is resolved to give rise to the repaired DNA strands. Both contain newly synthesised DNA (red), i.e. the process is semi-conservative.

1.10 RAD51

The RAD51 protein is one of the main components of the HR pathway. It plays a vital role in recombination, both for DNA repair, as well as in meiosis. It is in fact so important, that knocking out Rad51 in mice leads to early embryonic lethality[296,297]. This section will discuss the evolution, structure and mechanism of RAD51, with a focus on its role in the aetiology of FA.

1.10.1 Evolution

The human RAD51 protein was discovered as a homologue to the yeast recombinase Rad51, which in turn is a homologue of bacterial RecA[283,298,299]. The basic structure is highly conserved between all three proteins, particularly the "homologous core region"[298,300]. The yeast and human proteins are 83% identical[300]. Because of this similarity, most of the functional and structural studies of the enzymes have been carried out on the yeast and bacterial versions. When applicable, differences to the human protein will be highlighted, otherwise the proteins will be treated as interchangeable. These three proteins form part of a larger evolutionary family called the RecA/RAD51 like proteins. This family shares a common RecA/RAD51 core domain containing an ATP-binding Walker A and B motif[301]. The family is an evolutionarily ancient group, being found in eubacteria, eukaryotes and archaea[301,302]. Apart from RAD51 itself, it contains many of the RAD51 accessory proteins that play a role in the FA pathway and HR, including RAD51B, C and D, as well as XRCC2 and 3, all of which have more than 25% identity on the protein level[303]. The RecA/RAD51 group is part of the larger, ATPases associated with diverse cellular activities (AAA+) and additional strand conserved E (ASCE) family, which shares a common catalytic mechanism[304].

1.10.2 Structure-function relationship

RAD51's structure is highly similar to the bacterial RecA structure, yet there are some key differences in its domain architecture. In contrast to RecA, RAD51 starts with an N-terminal domain containing a helix-hairpin-helix domain, which mediates DNA binding. This is followed by the conserved RecA/RAD51 domain, which makes up the bulk of the protein. RAD51 ends with this domain, whereas RecA has an additional C-terminal domain, which facilitates DNA binding like the N-terminal RAD51 domain[301].

Crystallographic analysis showed that the catalytic domain of RAD51 and RecA are topologically identical and can be superimposed with negligible deviation between the structures[305]. The RecA domain is made up of an eight-stranded β-sheet flanked by α-helices. ATP binds near the C-terminus of a β-sheet, with its phosphates close to the N-terminus of an α-helix[306]. When bound to DNA, the RAD51 filament encases it in a helical filament with 103 Å pitch, a rise of 16.1 Å and a twist of 56.2°, consisting of 6.4 protomers per turn[307] (Figure 1.8).

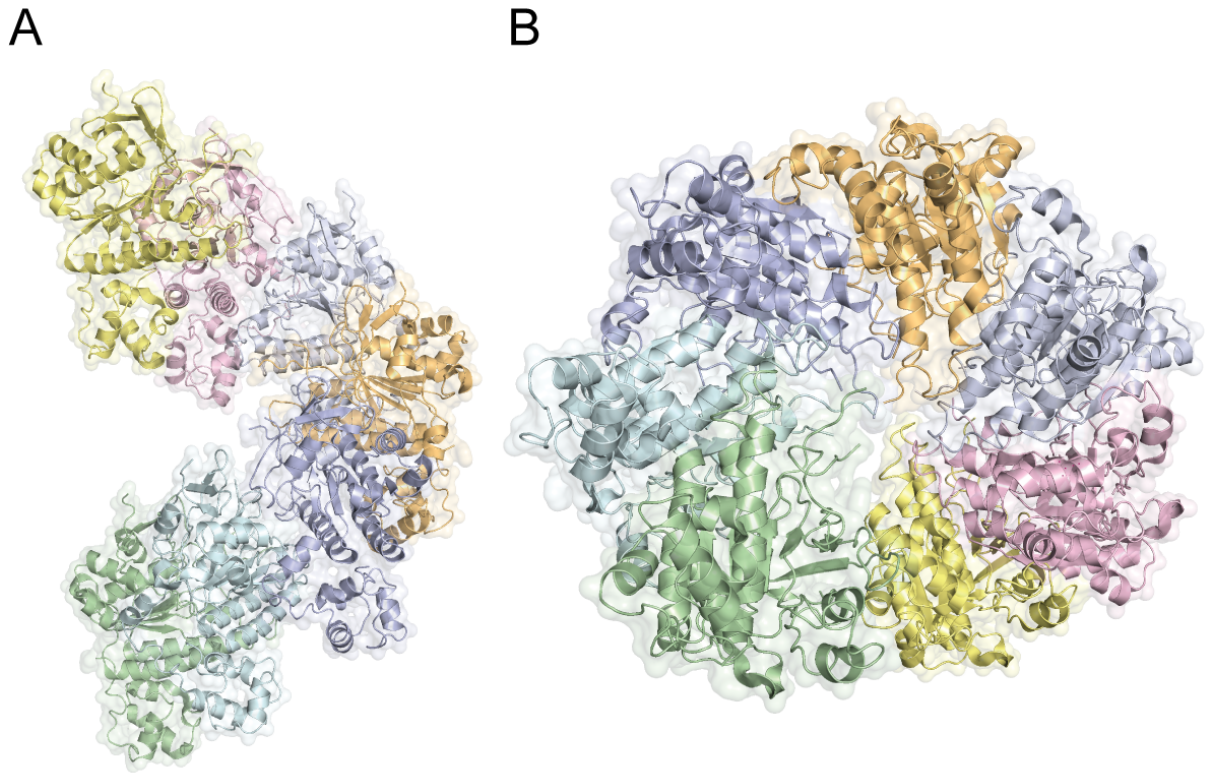


Figure 1.8: RAD51 crystal structures. RAD51 filament wound around ssDNA (not shown), forming a protomer seen from the side (A) and from the front (B) The individual RAD51 protomers are coloured differently. The helical shape of the filament is clearly apparent. DNA would normally fit into the channel in the middle, as visible in (B). Based on the crystal structures by Short *et al.* 2016[307]. PDB: 5JCZ.

These filaments are necessary to facilitate strand exchange[284], akin to bacterial RecA. Just like RecA interacts closely with the bacterial ssDNA binding protein SSB, Rad51 works together with RPA, which smooths out secondary structure in ssDNA and significantly stimulates Rad51 filament formation[308]. RAD51 has two DNA binding sites, one of which is necessary to bind the invading DNA strand, the other necessary to weakly bind DNA being probed for homology. Once homologous DNA is found, the

strand to be displaced is bound more tightly, allowing formation of the strand exchange complex[309]. Upon completion of strand exchange, RAD51 hydrolyses the bound ATP to ADP and the resulting conformational change leads to dissociation from the DNA[286,287]. Even though this mechanism is highly conserved, there are several small differences in the function of RecA and Rad51. RecA preferentially binds ssDNA or partially ssDNA[310], whereas Rad51 can bind under various conditions[284,298,311], including dsDNA. Furthermore, strand transfer occurs in opposite directions; RecA transfers 5'-3', whereas Rad51 transfers 3'-5'.

1.10.3 Function in meiosis and lymphopoiesis

Apart from its roles in somatic recombination, Rad51 is also seen on chromosomal DNA during meiotic prophase, where it facilitates cross overs between chromosomes[312–314]. However, here it does not act directly to catalyse strand exchange, but rather acts as an accessory factor to the highly similar protein DMC1, which catalyses meiotic strand exchange[315]. Because of this function, it is highly expressed in meiotic germ cells. Similarly, lymphoid tissues undergo considerable amounts of somatic recombination to generate diverse antibodies and T-cell receptors, leading to high expression of Rad51 there[300,316,317]. In somatic cells, Rad51 foci are normally only seen following DNA damage[312].

1.10.4 RAD51 in Fanconi anaemia

Recently, there have been two case reports of patients suffering from Fanconi anaemia-like symptoms linked to mutations in the RAD51[114,115]. One patient was a girl suffering from radial dysplasia, absent right thumb, pelvic left kidney, microcephaly and micropthalmia, café-au-lait-spots and increased DNA damage in response to crosslinking agents. This patient had above average intelligence[114]. The other patient displayed growth retardation, microcephaly, hydrocephalus, thumb abnormalities, imperforate anus and an abnormal left testicle. This patient displayed mild mental retardation, with an IQ of about 70[115]. Importantly, neither patient has yet developed BMF by age 13 and 23 respectively[114,115]. The discovery of patients carrying mutations in *RAD51* was surprising, as loss-of-function mutations of *Rad51* in mice invariably leads to early embryonic lethality[296,297]. This discrepancy was explained by the fact that both

patients only carry one mutant copy of the gene, which acts in a dominant negative fashion. HR is still functional in both patients, allowing survival to birth[114,115]. One mutation leads to the change of threonine 131 to proline (T131P) in the Walker A motif, leading to increased speed of ATP hydrolysis. The mutation causes the formation of unstable filaments and impairs the functionality of RAD51 in protecting the broken-down replication fork from excess processing by DNA2-WRN, which makes the appropriate repair of ICLs impossible. Interestingly, this patient upregulates the wild type form of the protein, seemingly to compensate for the mutant allele[114]. The other mutation leads to a change in alanine 293 to threonine (A293T), which leads to decreased, rather than increased ATPase function, but impairs DNA binding and also leads to unstable filament formation[115], presumably also impairing replication fork protection and thus ICL repair. Intriguingly, mapping of the mutations to a model of RAD51 in the filament revealed that both mutations map to the interface between the individual RAD51 protomers[307] (Figure 1.9). Phenylalanine 129 on one protomer interacts with histidine 294 from the adjacent protomer to stabilise the interaction between them. As shown in Figure 1.9, the disease-causing mutations map very closely to these key amino acids. In the case of the T131P mutation, a conformational distortion of the α-carbon backbone might lead to the F129 side chain, being oriented wrongly, obstructing the interaction with H294. The A293T mutation on the other hand might interfere with the orientation of histidine 294 itself, similarly destabilising the interface. It is intriguing that both mutations in very different parts of the primary structure map to the same location in the protein complex, leading to such similar defects. It is unclear what role the changed ATPase efficiencies play in this model, as the defects in filament formation seem to be mediated by abnormal protomer-protomer interactions rather than aberrant ATP hydrolysis[307].

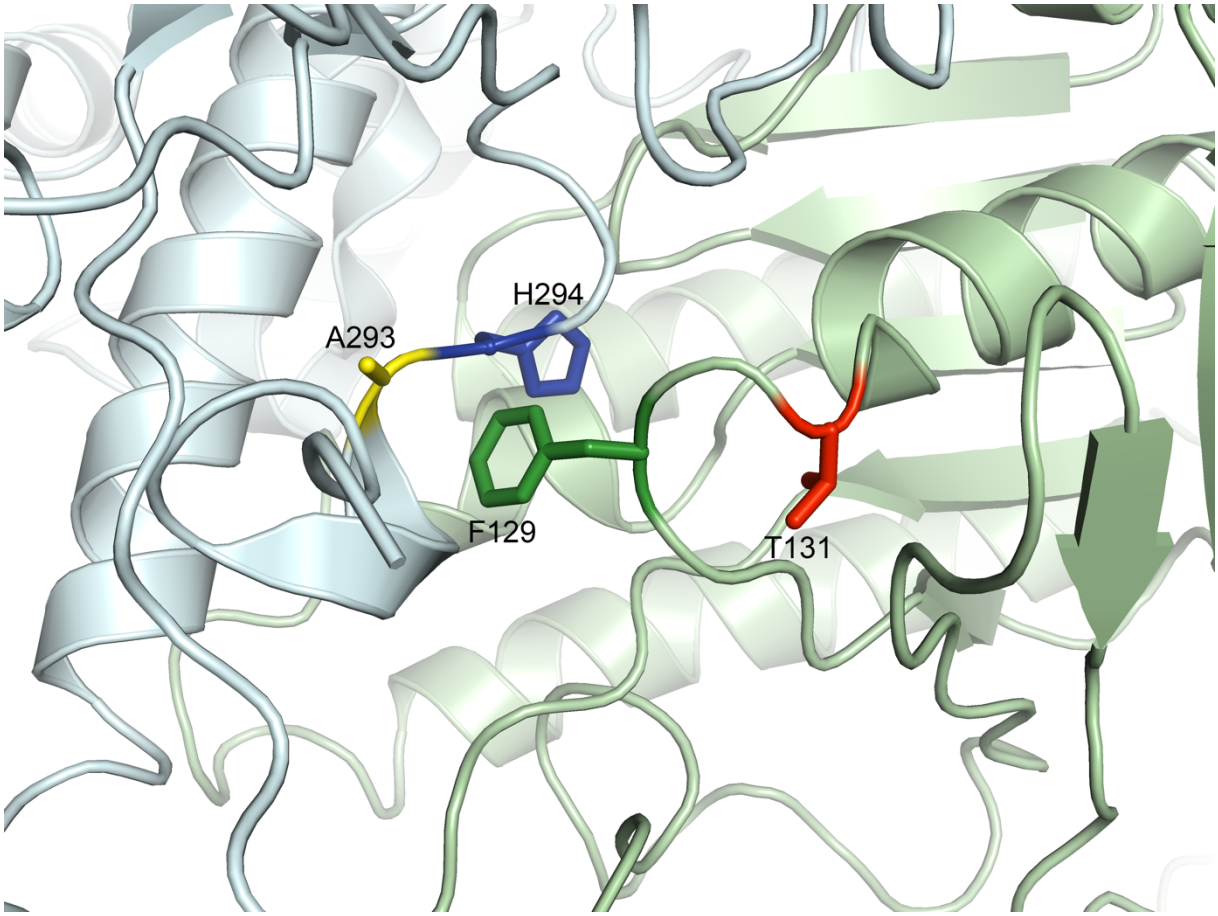


Figure 1.9: Location of the FA associated mutations on the RAD51 filament. The two different protomers are shown as cartoons in light blue and light green respectively. The side chains of FA-relevant amino acids are shown as stick models. F129 is in green, T131 in red, A293 in yellow and H294 in blue. The interaction between F129 and H294 can be seen in the middle between the protomers. It is apparent that the disease-causing T131P and A293T mutations are in close proximity on the filament, even though they are far apart on a single protomer and the primary amino acid sequence. Based on the crystal structures by Short *et al.* 2016[307]. PDB: 5JCZ.

1.11 Zebrafish as a model organism in haematology

Zebrafish (*Danio rerio*) are teleost fish in the cyprinid family (the carp or minnow family) of the class Actinopterygii (ray-finned fish), which are a subgroup of the Osteichthyes (bony fish)[318]. Their original range lies in the South-eastern Himalaya region[319]. They live in various different habitats, but most of them are associated with aquatic vegetation, such as rice fields[320]. The average natural life span of the zebrafish is unknown, but can exceed five years in rare cases in laboratories[320]. Teleosts, the group to which zebrafish belong, underwent a genome duplication event. This complex genome might be an explanation why teleosts are the most species-rich vertebrates[321,322].

1.11.1 Historical developments

The first known use of teleost fish for scientific research goes back to 19th century Austria. Joseph Oellacher, a physician, discovered the intra-embryonic generation of blood in trout embryos, whereas in other vertebrates haematopoiesis initially starts in the yolk[323]. Zebrafish were first mentioned in the literature as a scientific model in 1963[324]. Again, the paper was concerned with the early stages of haematopoiesis. It seems fitting that both the first paper using teleosts, as well as the first paper using zebrafish investigated blood formation, considering how popular the zebrafish model has proven to be with haematologists. Later, George Streisinger at the University of Oregon was the first to popularise the use of zebrafish as a model organism. Their high fecundity, small size and moderate space requirements[318], as well as their transparent embryos[325] made them ideal for use in research, so many laboratories started to set up their own colonies.

More recently, there have been many advances enabling even more sophisticated work on zebrafish. Their genome has been sequenced, which revealed that at least 70% of the human genome have at least one orthologue among the ~26000 genes in the zebrafish genome, including most known disease-causing genes[326]. The zebrafish mutation project (ZMP) at the Sanger Institute[327] is one of several resources to obtain various mutant lines. The CRISPR-Cas9 system has been adapted successfully for zebrafish[328,329], including a tissue-specific variant[330]. This technique has for example been used to supplement the Sanger ZMP pipeline[331]. Overall, the availability of these techniques and

the intrinsic advantages of the fish model make the zebrafish very attractive organisms for research purposes.

1.11.2 Zebrafish haematopoiesis

Haematopoiesis is highly conserved between fish and mammals[325,332,333]. As in mammals, there are two waves of haematopoiesis, called "primitive" and "definitive" wave[332,333].

The primitive wave (Figure 1.10A) starts from 12 hours post fertilisation (hpf) in the intermediary cell mass (ICM). It is a transient wave making erythrocytes[334] and leukocytes, which consist of mostly macrophages, but also some neutrophils[335–337]. From 12-16 hpf, erythroid and myeloid cells are made exclusively in the anterior lateral mesoderm and posterior lateral mesoderm respectively. Between 16 and 48 hpf, there is additional generation of primitive neutrophils and thrombocytes in the peripheral blood islands (PBI)[333]. The primitive wave is only relevant for very early embryogenesis, so the focus here will lie on the definitive wave.

The definitive wave (Figure 1.10B) starts around 33 hpf, when long-term haematopoietic stem cells (LT-HSCs) are formed from endothelial cells of the ventral wall of the dorsal aorta (VDA) in the anterior gonad mesonephros (AGM). Stem cells bud off from the VDA until approximately 54 hpf, with peak activity at 48 hpf [338–341]. This is dependent on the blood flow from the earlier primitive wave[342]. These newly made HSCs subsequently move to the caudal haematopoietic tissue (CHT), which serves as an intermediate place of haematopoiesis in which the HSCs expand greatly, akin to the foetal liver[343]. The CHT contributes to haematopoiesis roughly between 2 and 6 dpf [332]. At 3 dpf, the thymus is colonized by lymphoblasts stemming from the AGM and CHT[344]. The function of this organ is to serve as the production place of mature T-lymphocytes in the embryo and adult. Cells expressing the recombination-activating genes *rag1* and *rag2,* necessary for immunoglobulin and T-cell receptor production, can be seen there as early as 3.5 dpf [345]. From 4 dpf onwards, the pronephros, which later forms the fully mature mesonephros, is colonized by cells from the AGM and CHT[343,346]. The mesonephros is the fully mature form of the kidney in zebrafish, whereas in mammals it is completely replaced by the metanephros and atrophies[347].

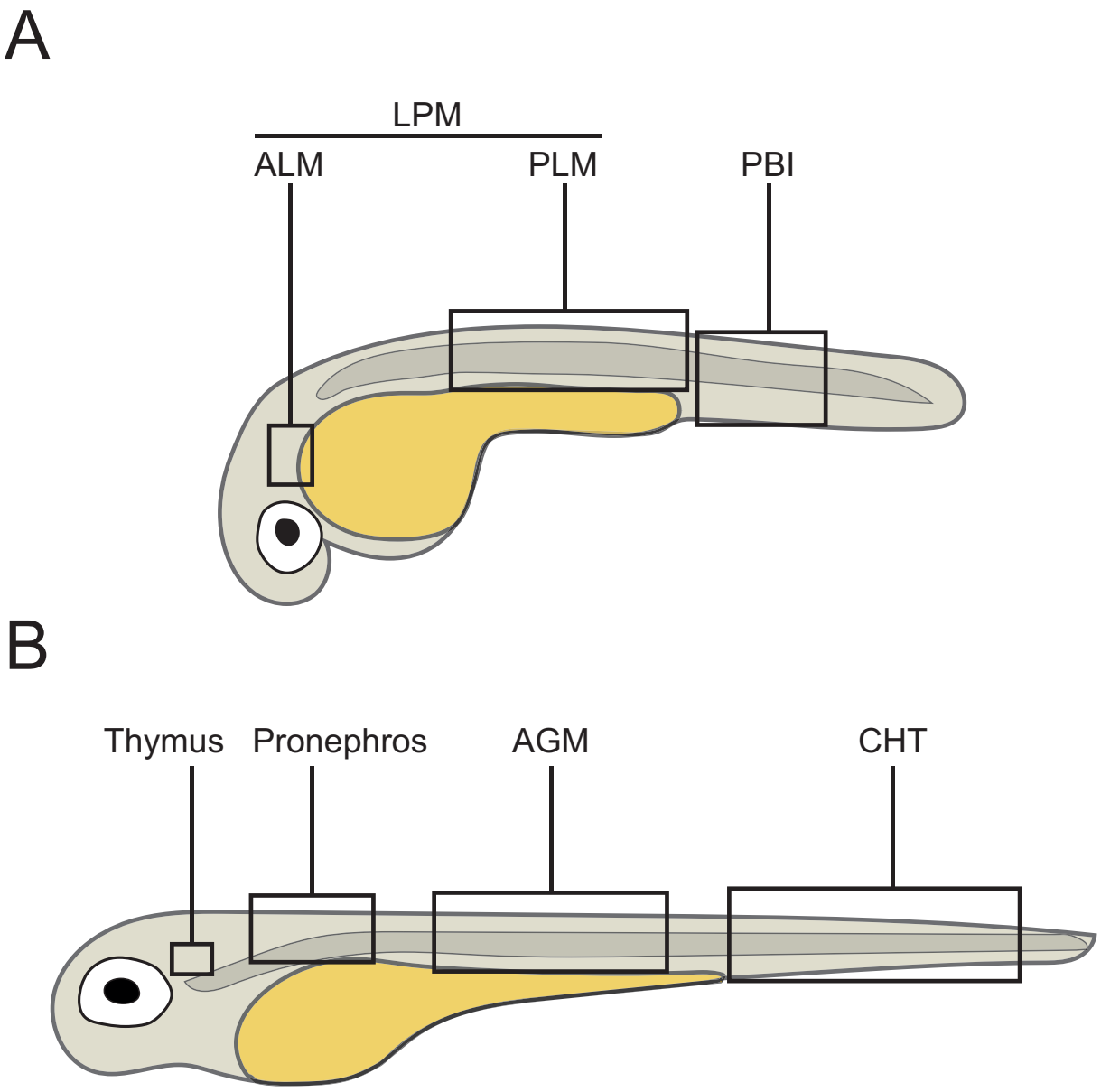


Figure 1.10: Locations of haematopoiesis in the zebrafish embryo. (A) Sites of the primitive wave in the zebrafish embryo. LPM = lateral plate mesoderm, ALM = anterior lateral mesoderm, PLM = posterior lateral mesoderm, PBI = peripheral blood island. (B) Sites of the definitive wave in the zebrafish embryo. AGM = aorta-gonad-mesenephros, CHT = caudal haematopoietic tissue.

In adult fish (Figure 1.11), haematopoiesis occurs in the kidney marrow, which is the equivalent of the mammalian bone marrow[333]. They have all major blood lineages found in humans, but some small differences do exist. Their erythrocytes are nucleated, but they undergo globin switching just like mammalian red blood cells[348]. Instead of platelets, they have thrombocytes, which are functionally identical but also nucleated. No megakaryocytes exist in zebrafish[349]. Granulocytes, such as neutrophils are also

found[350,351]. Other myeloid cells, such as monocytes/macrophages[336], dendritic cells[352] and mast cells[353] exist as well. The lymphocytic lineage is present too, with B-cells[354], T-cells[345], as well as natural killer (NK) cells[355–357].

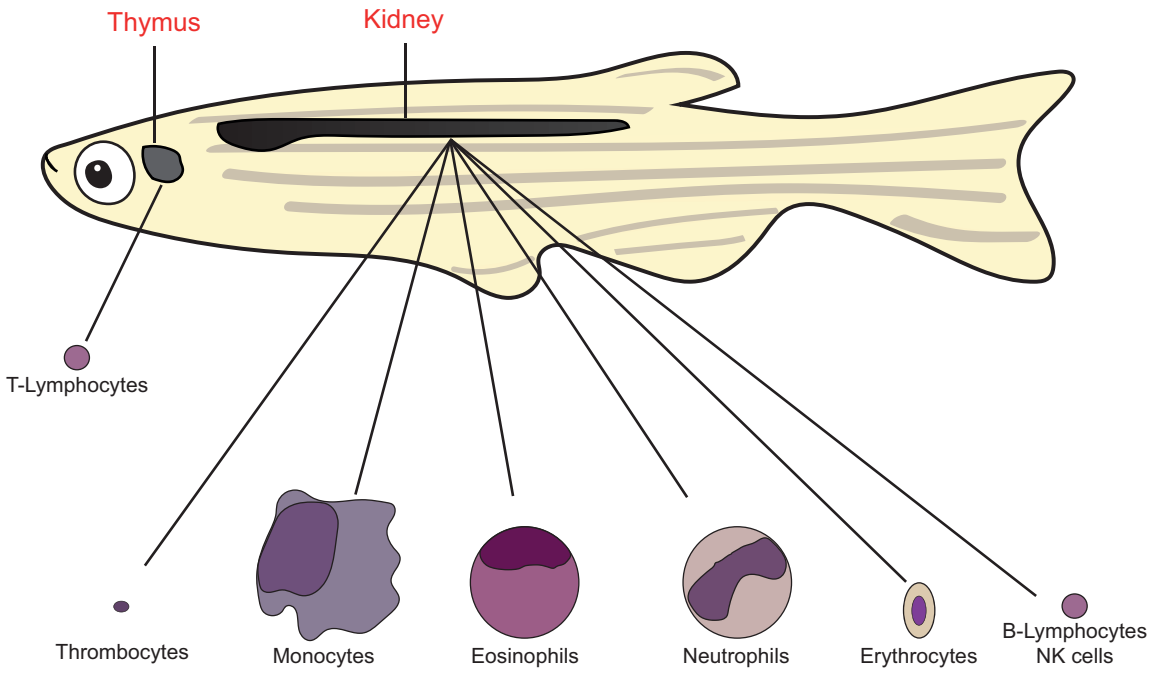


Figure 1.11: Sites of haematopoiesis in adult zebrafish. Zebrafish have all major blood lineages found in humans, which are mostly made in the kidney. T-lymphocytes are an exception, as they are formed in the thymus.

1.11.3 The Fanconi anaemia pathway in zebrafish

The Fanconi gene network is generally conserved in the zebrafish, with the exception of *BRCA1/FANCS*[358] (Table 1.3). There is evidence that the whole system must have evolved very early, at least before the divergence of ray fin and lobe fin fish approximately 450 million years ago[359,360]. FA genes seem to have reverted back to a single copy after the teleost genome duplication event. This is because none (bar one) of them is duplicated, which would be very unlikely to have happened by chance, considering the number of genes involved and that roughly 30% of the zebrafish genome is duplicated. This suggests the stochastics of the individual proteins are very important[359]. A recently discovered potential FA gene (*REV7/FANCV*) is in fact duplicated[266]. However, it has not been fully shown to be a *bona-fide* Fanconi gene and is not involved in the core Fanconi complex. This indicates that at least in the core complex the stochasticity of the proteins

is vital for the successful survival of an organism. This conservation of the pathway and the fact that they are vertebrates with similar organs to humans, make zebrafish good for modelling FA[359].

Table 1.3: Conservation of the FA gene network among several common model organisms. O indicates conservation, whereas X indicates the absence of this gene. A clear divide in conservation between lower organisms and vertebrates is seen. Zebrafish have all the FA genes that humans do, apart from *FANCS*/*BRCA1*. Data in this table is mostly based on Dong *et al.* 2015[358], with additional information from *FANCV* and *U* added via ENSEMBL searches.

Species	Fanconi Gene																			
	A	*B*	*C*	*D*	*E*	*F*	*G*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*	*U*	*V*
Homo sapiens (FANC...)																					
Mus musculus	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
Xenopus laevis	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
Danio rerio	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X	O	O	O
Drosophila melanogaster	X	X	X	X	O	X	X	X	O	X	O	O	X	O	X	O	O	X	O	X	O
Caenorhabditis elegans	X	X	X	O	O	X	X	X	O	O	X	X	X	O	O	O	O	X	O	X	X
Saccharomyces cerevisiae	X	X	X	X	X	X	X	X	X	O	X	O	X	O	O	O	O	X	X	X	X

Most of the Fanconi anaemia research on zebrafish so far has been carried out on embryos. Titus *et al.* showed that FA genes are mainly expressed in the developing nervous system, as well as the gonads[359]. Just like in humans[133], there is a connection between loss of the Fanconi pathway and excess *p53* activation. This was shown by knocking down *fancd2* in zebrafish embryos using morpholinos followed by irradiation. This led to a small eye and head phenotype, which could be rescued by *p53* knockdown[361]. Interestingly, it seems like Fanconi genes are very important in early development and are maternally provided because of that[359,361]. This is again consistent with the idea that the stem cell defect seen in FA is already apparent during embryonic development[133].

There are currently two zebrafish models in which Fanconi anaemia genes are lost; one involving loss of *fancl*[362] and one involving *brca2*[363,364] (this model was independently described by two different groups at almost the same time). Both models showed sex reversal with complete lack of female fish. This feature successfully rescued with *p53* co-

mutation[363,364]. A functional FA pathway seems to be necessary for the development of females and, explaining why the FA genes are highly expressed in the gonads of developing fish. Interestingly, *fancl* mutant males are fertile[362], indicating that HR, but not the FA pathway is necessary for successful male meiosis.

Interestingly, none of the mutants have been analysed in-depth for features of Fanconi anaemia. While *brca2* mutant fish display increased sensitivity to crosslinking agents[363,364], as do FA patients, there are no reports of any haematological abnormalities. This suggests the blood phenotype (if there is any) is subtle. This reflects the situation seen in mouse models of the disease, described in section 3.1.2. The lack of adult zebrafish without a characterised blood phenotype compelled me to study haematopoiesis in my *rad51* mutants more closely, leading to the first description of a zebrafish FA model using adult fish.

1.12 Aims

As discussed above, there are still significant gaps in our understanding of the causes of FA, as well as in our understanding of the *in-vivo* role of Rad51. In this thesis, I try to address these issues by pursuing the following aims:

1. Characterising the haematological and non-haematological phenotype of *rad51* mutant zebrafish

2. Using the zebrafish FA model to better understand the cellular and molecular basis of FA

3. Elucidating the effect of increased stress haematopoiesis on the haematological phenotype of the *rad51* mutant fish to gain a better understanding of the role of stress in the aetiology of FA

In the next chapter, the methods used to answer these questions are outlined. Chapters 3-5 describe and discuss the experiments conducted during this study. In the sixth and final chapter, these results are placed into the overall context of FA research.

2 Materials and methods

2.1 Zebrafish care and strains

Fish lines were maintained in the Sanger Institute zebrafish facility according to EU, Home Office and Animals Scientific Procedures Act 1986 regulations. Wild type fish were of the Tübingen long fin strain. The *rad51*^{sa23805} line was obtained from the Wellcome Trust Sanger Institute Zebrafish mutation project[327]. The *rad51b* mutant line was generated using the CRISPR-Cas9 technique as described below. The *tp53*^{zdf1} line[365] was generously provided by Sebastian Gerety. The *rag2*^{E450fs} line[366] was kindly provided by David Langenau. Apart from the mutant lines, I also used *Tg(itga2b:EGFP)*[367] (*itga2b* is an orthologue of human *CD41*), *Tg(gata1a:EGFP)*[368] and *Tg(sdf1a:dsRed)*[369] fish (*sdf1a* is also called *cxcl12a*). Lines were crossed as indicated in the main text. Embryos were incubated at 28.5°C in egg water (0.18 g/l of synthetic sea salt in sterile H_2O). When transparent embryos were required for assays, they were placed into egg water containing 0.0045% 1-phenyl-2-thiourea (Sigma-Aldrich) at about 7-8 hpf to inhibit melanocyte formation as described previously[370]. For breeding, adult fish were kept in small separate tanks with a mesh insert and removable barriers to control the exact time of the mating.

2.2 Genomic DNA extraction and genotyping

Adult fish were anaesthetised in 0.02% 3-amino-benzoic acid ethyl ester (tricaine) (Sigma-Aldrich) and the tip of the tail removed and sorted into 96-well plates. For embryos, either the head or the tail was kept for genotyping and also sorted into 96-well plates. DNA was extracted using the Hot Shot method. Samples were treated with alkaline lysis buffer (25 mM NaOH and 0.2 mM EDTA) at 95°C for 30 minutes to lyse cells and extract the DNA. This solution was neutralized by adding an equal volume of neutralization buffer (40 mM tris-HCl). I then used KASP (a commercial variant of competitive allele-specific PCR) genotyping assays (KBioscience) according to the manufacturer's protocol. The primers used for genotyping can be found in Table 2.1. Primers were acquired from Sigma-Aldrich or LGC genomics. Individual primers were mixed as follows: 24 µl primer 1 and 2 each, 60 µl common primer and 92 µl H_2O.

Analysis was carried out on a PHERAstar Plus (BMGlabtech) using KlusterCaller software (KBioscience).

Table 2.1: KASP Primers used for genotyping. Primers for *rad51l1* were ordered commercially from LGC genomics, hence the sequence is not known. All primers are listed in 5'-3' direction.

Allele	Primer 1	Primer 2	Common
***tp53*^{zdf1}**	GAAGGTGACCAAGTTCATGCTGATGGGCCTGCGGTTCA	GAAGGTCGGAGTCAACGGATTGGATGGGCCTGCGGTTCT	ACAACTGTGCTACTAAACTACATGTGCA
***rad51*^{sa23805}**	GAAGGTGACCAAGTTCATGCTCCTTCAACACTGACCATCAAACAC	GAAGGTCGGAGTCAACGGATTGCCTTCAACACTGACCATCAAACAT	GGTCATCATAGCGGAGGCCTGAT
rad51l1	Proprietary from LGC genomics		

2.3 Western blotting

Tissues for western blotting were dissected from fish and lysed in lysis buffer (50 mM tris pH 7.4, 100 mM NaCl, 5 mM $MgCl_2$, 0.5% NP40, 1 mM dithiothreitol (DTT)) with one tablet of proteinase inhibitor (Roche Diagnostics) per 10 ml of buffer for 30 minutes. Samples were then centrifuged at 14000 rpm for 10 minutes and the supernatant collected. The protein concentration was quantified using the Bradford assay. Each sample was diluted to the same concentration and stored at -80°C. Lithium dodecyl sulphate loading buffer to 1X final concentration and 25 µl/ml β-mercaptoethanol were added to the samples, which were then run on 12% NuPAGE polyacrylamide gels (Thermo Fisher Scientific) at 90 V for 20 minutes and then at 120 V until the marker reached the bottom of the gel. Proteins were transferred onto methanol-activated polyvinylidene fluoride membranes at 300 mA for 1.5 hours and blocked overnight in 2% bovine serum albumin (BSA)/tris-buffered saline with Tween (TBST) (50 mM tris pH 7.4, 0.05% Tween 20 (Sigma-Aldrich)). The membrane was then probed with the primary antibody (1:1000) for 1 hour, followed by horseradish peroxidase (HRP)-conjugated anti-mouse or goat secondary antibody for 1 hour. The membrane was washed extensively in TBST after each probing step. The signal was revealed using SuperSignal West Femto Maximum Sensitivity substrate (Thermo Scientific) and imaged using an GE Amersham ImageQuant LAS 4000. Used membranes were stripped

for 45 minutes at 50°C using pre-warmed stripping buffer (12.5 ml tris-HCl pH 6.8 0.5 M, 20 ml 10% SDS, 0.8 ml β-mercaptoethanol, 67.5 ml H_2O), followed by 2 hours of rinsing under running tap water and a ten-minute wash in TBST. Details on the antibodies used can be found in Table 2.2.

Table 2.2: List of antibodies. IS = immunostaining, WB = western blotting.

Target	Species	Conjugate	Company	Catalogue No.	Dilution used
Rad51	rabbit	none	Abcam	ab137323	1:1000 (WB), 1:200 (IS)
Rad51	rabbit	none	AnaSpec	55838	1:200
BrdU	rabbit	fluorescein	Roche	11202693001	1:100
pH2AX	rabbit	none	-	-	1:1000
Beta-actin	mouse	none	Sigma-Aldrich	012M4821	1:1000
Mouse IgG	rabbit	HRP	Abcam	ab97046	1:50000
Rabbit IgG	goat	HRP	Molecular Probes	G21234	1:50000

2.4 Immunostaining

Embryos were fixed at 2 dpf in 4% paraformaldehyde (PFA) in phosphate buffered saline (PBS) for 2 hours at room temperature, followed by extensive PBS + Tween 20 (PTW) washes. Samples were digested in 0.8 µl/ml proteinase K for 30 minutes at RT, followed by refixation in 4% PFA at room temperature (RT) and PTW washes. This step was omitted for freshly fertilised embryos. The samples were blocked in blocking solution (10% foetal bovine serum (FBS), 1% dimethylsulphoxide (DMSO), 0.1% Triton X-100 in PBS) for 1 hour at RT followed by overnight incubation with rabbit primary antibody in blocking solution (1:200) at 4°C. The embryos were then washed extensively in blocking solution and incubated with AF488-conjugated secondary antibody for 3 hours at RT. The stained embryos were washed in PBS and optionally incubated for 2 days at RT in 1 ng/ml Hoechst dye 33342. Details on the antibodies can be found in Table 2.2.

2.5 Microscopy

Pictures of flat mounted tails and live embryos were taken using a Zeiss Axio Zoom.V16 using Zen Imaging software. Pictures of stained slides were taken using a Leica DM 4000B using 10X and 20X non-immersion objectives and a 63X oil immersion objective as appropriate. An Olympus DP72 camera was used for image acquisition in conjunction with Olympus cellSens software. Chromosome spreads were imaged using a Leica DM 5000B microscope with a 100X oil immersion objective and SmartCapture software (Digital Scientific UK). Confocal images were acquired on a Leica SP-5 confocal microscope using a 40X water immersion lens and Leica LAS-AF software. Measurements on pictures, as well as maximum projections for fluorescence images were made using ImageJ software.

2.6 Embryo irradiation

Embryos were irradiated at 24 hpf in a Gammacell 1000 Elite Blood Irradiator (MDS Nordiron) at 750 cGy. Depending on the batch were then either collected and fixed after 3 hours for immunostaining, or kept until 2 dpf. At 2 dpf, comparison pictures were taken and embryos scored for phenotypic abnormalities.

2.7 Chromosome spreads

Embryos were treated with 5 µg/ml mitomycin C (MMC), 1 µg/ml diepoxybutane (DEB), 10 µM 1,5-isoquinolinediol (DiQ) or 10 nM camptothecin (CPT) (Sigma-Aldrich) in egg water between 4 and 24 hpf, followed by treatment with 4 mg/ml colchicine (Sigma-Aldrich) in egg water for 90 minutes. The embryos were then transferred to 1.1% sodium citrate and the yolk dissected away for 10 minutes. Dissected embryos were kept in sodium citrate on ice for another 10 minutes, followed by overnight fixation in 3:1 methanol (MeOH):acetic acid. The heads of the embryos were then separated and kept for genotyping. The individual tails were dissociated in 50% acetic acid by vigorously dissociating the tissue with dissecting forceps. The dissociated cells were then pipetted onto pre-warmed slides kept in a water bath at 50°C. After several seconds, a few drops of MeOH were added to stop the reaction. Following that, the slides were transferred into a drying chamber at 50°C. After complete drying of the samples, coverslips were

mounted onto slides using VECTASHIELD antifade mounting medium with 4',6-diamidino-2-phenylindole (DAPI, Vector Laboratories). Coverslips were fixed to the slide using nail varnish.

2.8 NHEJ inhibition

Non-homologous end joining was inhibited using the DNA ligase IV inhibitor SCR-7 (Sigma-Aldrich). SCR-7 was added to egg water at a range of concentrations (10, 25 and 75 μM). Embryos were incubated in these solutions between 4 and 24 hpf, at which point they were photographed and scored for abnormalities, followed by genotyping.

2.9 CRISPR-Cas9

Mutations in *rad51l1* were induced at exon two, leading to a seven basepair (bp) deletion (478-484delTGGGTCC in the cDNA). The targeted DNA sequence was 5'-GGATGTCCTGTCGGTCACCCAGG-3'. The sequence of the ssDNA oligonucleotides was 5'-TAGGATGTCCTGTCGGTCACCC-3' and 5'-AAACGGGTGACCGACAGGACAT-3' (Sigma-Aldrich). These were annealed and ligated with pDR274 vector (Addgene) linearised with BsaI (New England Biolabs) to make a guide RNA (gRNA) expression vector. gRNA was prepared with MAXIscript T7 kit (Life Technologies) using the DraI-linearised gRNA expression vector as a template, while Cas9 mRNA was synthesised using the mMESSAGE mMACHINE T7 kit (Ambion) and an pMLM3613 expression vector (Addgene) linearised with PmeI (New England Biolabs). Zebrafish embryos were injected at the 1-cell stage with 12.5 pg of gRNA and 160 pg of Cas9 mRNA. The presence of mutations was screened using Illumina miSeq.

2.10 Fish photography

Pictures of anaesthetised adult fish were taken using a Canon Eos 750D with a Canon Eos EF-S 60mm macro lens. The size was measured using ImageJ.

2.11 Semen collection

Adult male fish were anaesthetised using tricaine. Unconscious fish were dried off and placed on their backs in pre-made polyurethane foam moulds, allowing access to the

ventral side of the fish. Remaining water was dried off using a cotton bud. A 10 µl capillary was placed near the urogenital pore and the fish stroked gently from the anterior to the posterior end to stimulate the release of semen. Pictures of the capillaries were taken and used to estimate the volume of the emission. The cloudiness of the sample was used to determine the presence of spermatozoa.

2.12 Histology

Tissues for histology were fixed in 4% PFA overnight and if necessary, treated with 10% EDTA for decalcification over several days. Samples were processed by a Tissue-Tek VIP6 tissue processor (Sakura) and embedded in paraffin. Sections were obtained at 5 µm thickness using a Leica RM2235 rotary microtome. Slides were stained using standard Harris haematoxylin and eosin (H&E) staining (Sigma-Aldrich) and coverslipped on a Leica ST5010-CV5030 workstation.

2.13 RNA extraction and RT-qPCR

RNA was extracted using RNeasy columns (Qiagen) In brief, kidney samples were lysed in 350 µl RLT buffer and vortexed vigorously. An equal volume of 70% ethanol (EtOH) was added to each tube and the mixed sample transferred to silica columns. The samples were centrifuged at maximum speed for 15 seconds to bind to the column. This was followed by one wash with 700 µl RW1 buffer and two washes with 500 µl RPE buffer, interspersed by 15 second centrifugations at maximum speed. After the last wash, the columns were centrifuged for 2 minutes at maximum speed. Following that, the columns were transferred into fresh collection tubes and centrifuged at the same speed again for 1 minute. Finally, the tubes were placed into fresh RNAse free Eppendorf tubes and the RNA eluted by adding 30 µl of H_2O, followed by centrifugation at 8000 g for 1 minute. Eluted RNA was stored at -80°C until use. cDNA was generated using either the SuperScript III or the SuperScript VILO kit (Thermo Fisher). For SuperScript III, 2 µl of cDNA, 2µl of Enzyme 10 µl of buffer and 6 µl of H_2O was added per tube. The tube was then incubated at 25°C for 10 minutes, 50°C for 30 minutes and 85°C for 5 minutes. RNA was removed by adding 1 µl RNAse H and incubating at 37°C for 20 minutes. For SuperScript VILO, each tube contained 2 µl of RNA, 5 µl of VILO reaction mix and 13 µl of H_2O. Tubes were incubated at 25°C for 10 minutes, 45°C for 30 minutes and 85°C for

5 minutes. Following cDNA synthesis, samples were diluted threefold using H_2O. qPCR was carried out using 5 µl SYBR Green master mix (Thermo Fisher), 0.15 µl of 10 µM of forward and reverse primer each, 2.7 µl H_2O and 2 µl cDNA per sample in a QuantStudio 3 RT-PCR system (Thermo Fisher). Primers for qPCR analysis can be found in Table 2.3. Samples were run in duplicates and analysed using the ΔΔCt method[371]. The mean Ct value of the housekeeping genes *18s, actb1* and *eif1a* was used for normalisation. Statistical tests were carried out on the ΔCt values, as recommended in Yuan *et al.*[372]. $2^{-\Delta\Delta Ct}$ values were graphed with the geometric mean ±95% confidence intervals to estimate the fold change.

Table 2.3: Primers used for qPCR analysis. All primers are listed in 5'-3' direction.

Target	Forward	Reverse	Reference
18S	TCGCTAGTTGGCATCGTTTATG	CGGAGGTTCGAAGACGATCA	
actb1	CGAGCAGGAGATGGGAACC	CAACGGAAACGCTCATTGC	
brca2	AGCCAGACTTGTCCACAGAC	AGGCACACAGATAGCTCCTC	
csf1r	ATGACCATACCCAACTTTCC	AGTTTGTTGGTCTGGATGTG	373
eif1a	TGGCAAGGTCACAAAGTCTG	TCCGATGGGTTTTAATCAGC	
gadd45ab	AACATGAGACGTCTGGCAGA	AGATTGATGACTGGCACCCA	
il1b	GCTGGAGATCCAAACGGATA	ATACGCGGTGCTGATAAACC	374
il8	GTCGCTGCATTGAAACAGAA	AGGGGTCCAGACAGATCTCC	374
lig4	CTGGCGCCGTTTAACTTTCT	GTGCAGAGGTGAATGAACGG	
marco	ACGACAGCTTCGATAATTTG	AAAATACTGCTCTCGGTTCC	373
p21	GTGTCAGGAAAAGCAGCAGA	GACGCTTCTTGGCTTGGTAG	
p53	CGAGCCACTGCCATCTATAAG	TGCCCTCCACTCTTATCAAATG	
prkdc	TGAGTCTGACCGGGATGATC	GCTGCATTCTCTGGCTGTAC	
rad52	AGGATATGGGGTCAGCGAAG	GGCTCCAAATCGCTACGTTT	
rb1	GTATCTCTCTCCTGTCCGGC	CTGGATGTGAGGTCAGCAGA	
xrcc4	GGAGGAAGTACGGAAGATGAAC	CTGCATTCTGGCATCTCTTCT	

2.14 Single-cell suspension and cell counts

Single-cell suspensions were made as follows. Dissected zebrafish kidneys were placed in 5% FBS in PBS on ice and sequentially passed through a 40 µm cell strainer (Corning) using a plunger, followed by a second filtration using a 20 µm cell strainer (Partec). Cells were counted using Neubauer improved haemocytometers.

2.15 Kidney flow cytometry

Single-cell suspensions were made as described above. Flow cytometry/fluorescence activated cell sorting (FACS) was carried out on a BD LSRFortessa, a MoFlo XDP (Beckman Coulter) or a BD Influx (BD Biosciences). Debris was excluded using the forward (FSC) and side scatter (SSC) parameters (Figure 2.1A). Single cells were selected using FSC-A and FSC-W (Figure 2.1B). Dead cells were excluded using (Sigma-Aldrich) or propidium iodide (PI) at 1:1000 dilution (Figure 2.1C and D).

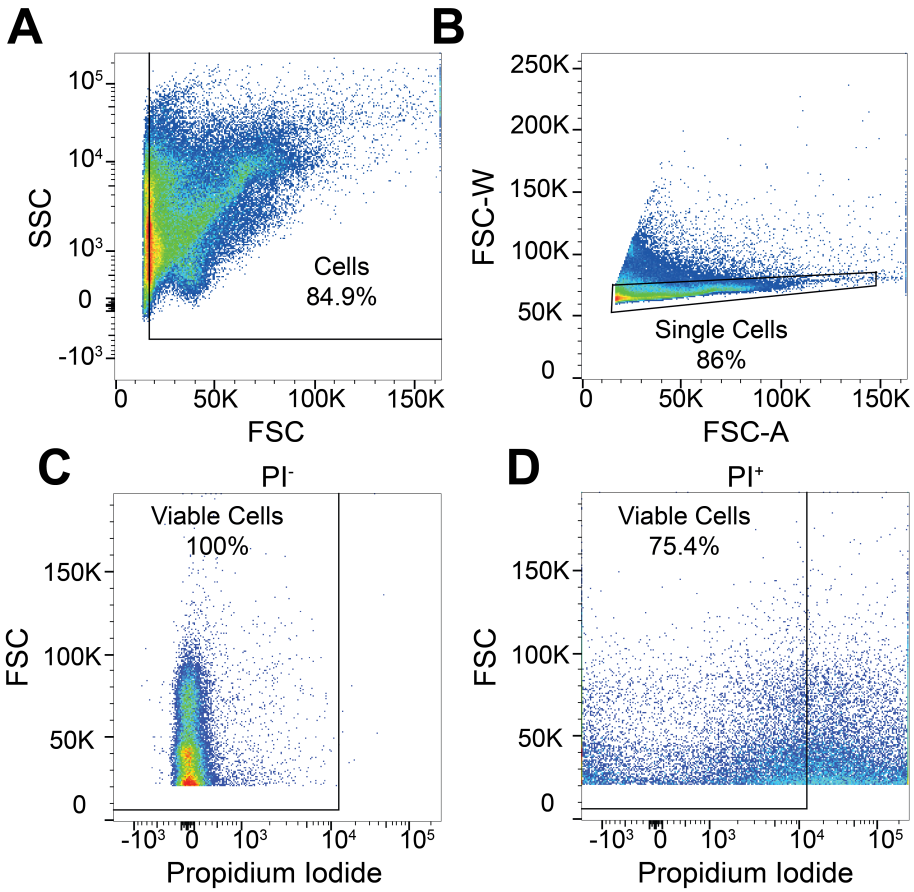


Figure 2.1: General FACS gating scheme. (A) Separation of cells from smaller debris using the FSC and SSC parameters. (B) Selection of single cells using the FSC-A and FSC-W parameters. (C and D) Selection of viable cells using PI on a PI^- control (C) and a PI^+ sample (D). Selection of viable cells using DAPI would work analogously, just using a different channel for detection.

Erythrocytes and committed erythrocytic progenitors were selected using the *Tg(gata1a:EGFP)* transgenic line[368] (Figure 2.2).

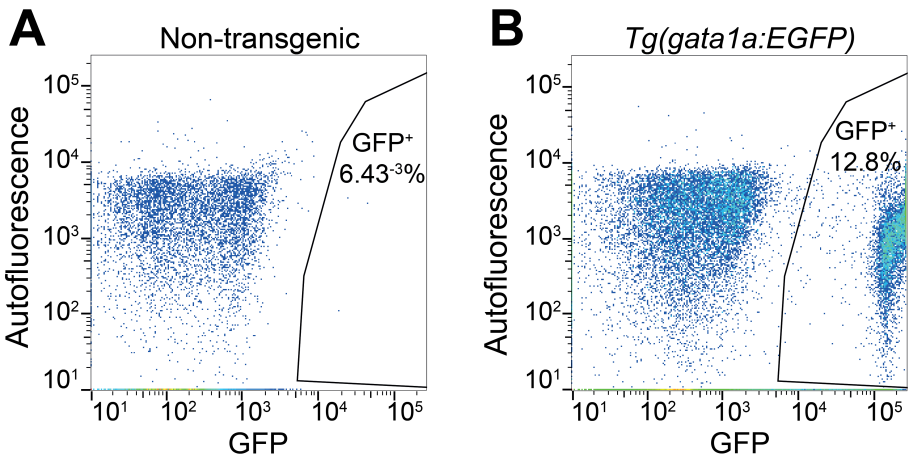


Figure 2.2: Selection of the erythrocytic lineage. (A) Gating for *Tg(gata1a:EGFP)*$^{+}$ cells on a non-transgenic control. (B) The same gate applied to a transgenic fish.

Cells of the thrombocytic lineage were selected by using the *Tg(itga2b:EGFP)* line[367] (Figure 2.3). In this line, thrombocytic progenitors are labelled in the GFP^{low} population and mature thrombocytes are labelled in the GFP^{high} population[375] (Figure 2.3C).

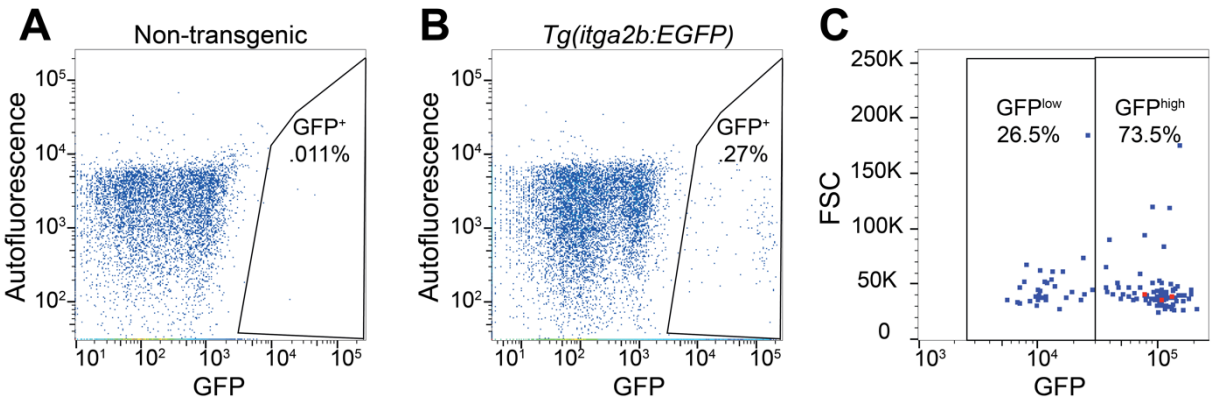


Figure 2.3: Selection of the thrombocytic lineage. (A) Gating for *Tg(itga2b:EGFP)*$^{+}$ cells on a non-transgenic fish. (B) The same gating applied to a transgenic fish. (C) GFP^{+} cells are split into GFP^{low} (thrombocytic progenitors) and GFP^{high} (mature thrombocytes) cells by selecting for two equally-sized gates at the 50% fluorescence mark.

Other blood lineages were selected on the basis of their FSC and SSC parameters, as first described by Traver *et al.*[376]. I used this gating considering new RNA-Seq information[377] that allowed me to redefine the identity of some of the populations (Figure 2.4).

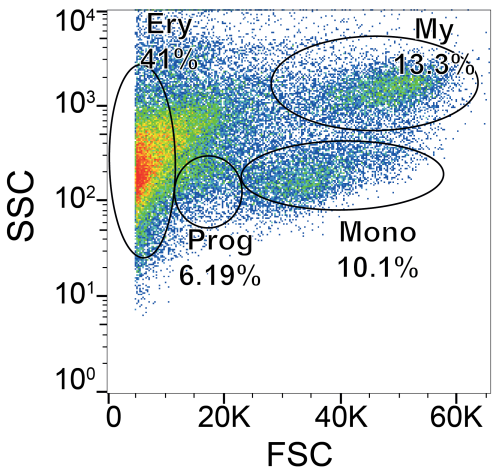


Figure 2.4: Identification of different blood lineages using the forward and side scatter parameters. Ery = Erythrocytes, Prog = Progenitors, My = Myeloid cells, Mono = Monocytes.

2.16 AnnexinV-propidium iodide assay

I used the Alexa Fluor 488 Annexin V/Dead Cell Apoptosis Kit (Thermo Fisher) per the manufacturer's instructions. In brief, kidney single cell suspensions were made as described above. Cells were spun down for 3 minutes at 3000 rpm and resuspended in 100 µl annexin buffer. One sample was split up to have controls without any stain, a control without PI and a control without antibody. All other samples received 5 µl annexin V antibody conjugated to AF488 (Roche), as well as 1 µl of 100 µg/ml PI. After 15 minutes of staining in the dark at RT, 400 µl of annexin binding buffer were added to each sample to stop the reaction. The samples were then filtered using 20 µm filters and analysed using FACS. Gating was set using the controls (Figure 2.5).

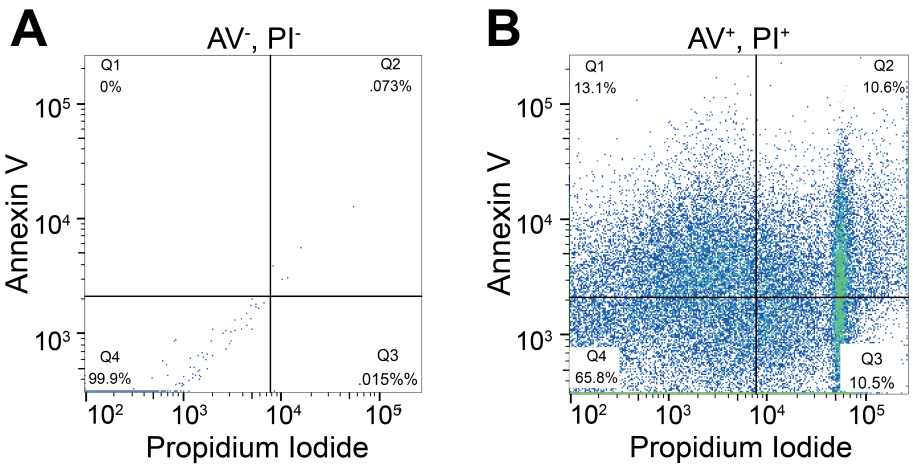


Figure 2.5: Gating for apoptotic cells. (A) Gating set on an unstained control, (B) gating on an AV-GFP and PI-stained control. Viable cells are in the bottom left (Q4), apoptotic cells in the top left (Q1) and dead cells are in the two gates on the right (Q2 and 3).

2.17 BrdU incorporation assays

2.17.1 Adult fish

Fish were subjected to intraperitoneal (IP) injection with 10 µl 10 mg/ml 5-bromo-2'-deoxyuridine (BrdU, Sigma-Aldrich) and culled after the time specified in the text. The kidney and blood were extracted and single-cell suspensions made as described earlier. These suspensions were spun down at 1500 rpm and fixed in 70% ethanol (EtOH) overnight. The samples were then treated with 2M HCl for 1 hour, followed by washes in tris (pH 9.5) and PBS (pH 6.8). The washed cells were incubated with anti-BrdU fluorescein-conjugated antibody at 1:100 in blocking solution (10% FBS, 1% DMSO, 0.1% Triton X-100) for 2 hours, after which the cells were resuspended in PBS. Samples were filtered again using 20 µm filters (Partec) and analysed using FACS (Figure 2.6). Details on the antibody can be found in Table 2.2.

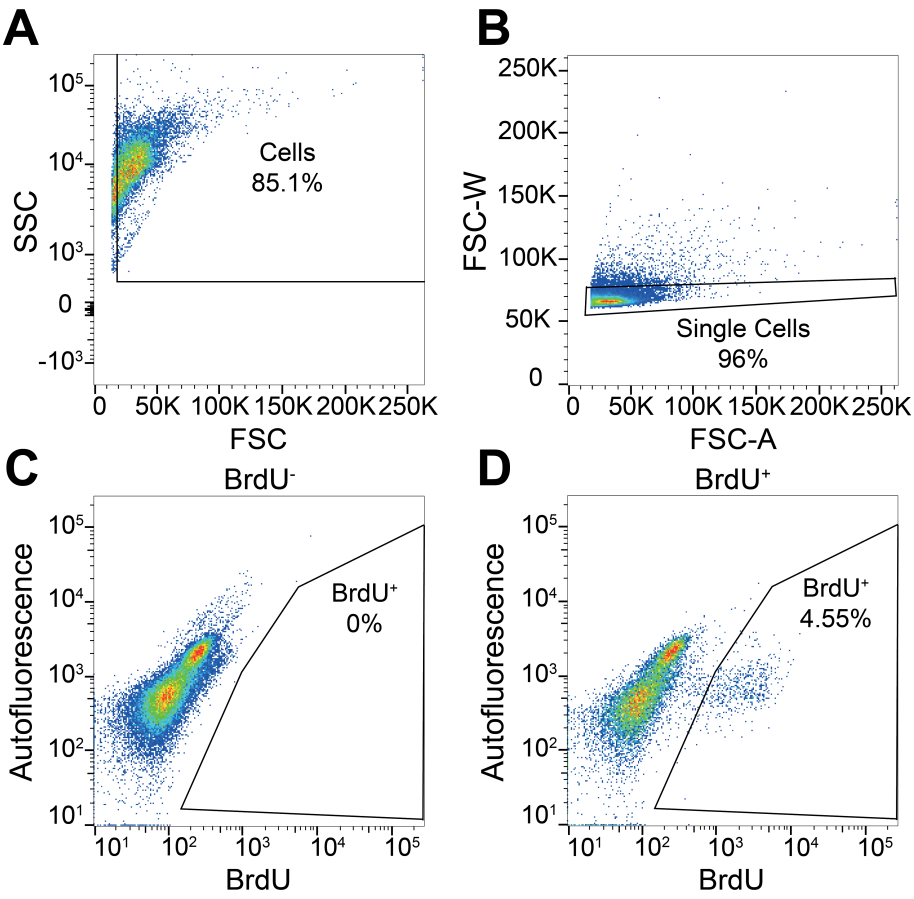


Figure 2.6: Gating for BrdU$^+$ cells on fixed cells. (A) Selection for cells, (B) singlets, (C) and BrdU$^+$ cells. (D) Same gating as in iii, but on a BrdU-injected sample.

2.17.2 Embryos

Embryos were dechorionated while chilled on ice for 15 minutes. This was followed by a 20-minute incubation in 10 mM BrdU dissolved in egg water on ice. They were placed at 28.5°C for 3 hours, followed by fixation in 4% PFA at 4°C overnight. They were then dehydrated in an increasing series of MeOH and stored at -20°C in 100% MeOH. After rehydration in a decreasing MeOH series, heads were cut off for genotyping. For 2 dpf embryos, the whole tails were then pooled according to genotype, whereas for 4 dpf embryos the CHT was cut out and pooled the same way. Samples were treated with 10 mM DTT in 1X Danieau's solution for 30 minutes at room temperature. They were then incubated in 1X liberase (Roche Diagnostics) in PBS for 3 hours at 37°C with regular vortexing. The reaction was stopped by replacing the solution with 5% FBS/PBS after spinning the tissue down at 3000 rpm for 3 minutes. Single cell suspensions were made

as described above, followed by fixation in 70% ethanol overnight. From here on, the staining process and analysis was identical to cells obtained from adults (see above).

2.18 Blood smears

Blood smears were carried out by severing the head of a culled fish on a microscope slide and smearing the blood along the length of the slide. After drying overnight, cells were fixed in -20°C cold methanol (MeOH) for 3 minutes at room temperature and stained in May-Grünwald solution (Sigma-Aldrich) diluted 1:1 in H_2O for 10 minutes at RT. This was followed by Giemsa solution diluted 5:1 in H_2O for 20 minutes at RT. Excess stain was removed by washing in H_2O.

Differential counts on blood smears were carried out using ImageJ. Ten pictures of each bloodsmear were taken and all cells from each picture counted blindly. Leukocyte numbers were normalised by dividing the number of leukocytes by the total number of cells.

2.19 Antisense probe synthesis

Template DNA for the probe was amplified by PCR and inserted into a plasmid. Plasmids were used to transform OneShot ready Cells (Life Technologies) according to the manufacturer's instruction. Transformed cells were then spread on plates carrying the appropriate antibiotic and incubated overnight at 37°C. Single colonies were picked and used to seed 25 ml Luria-Bertani broth containing the appropriate antibiotic, which was incubated at 37°C shaking overnight. The plasmid was extracted using a Qiagen MIDIPrep kit, according to the manufacturer's instructions. Depending on the orientation of the insert and possible restriction sites, 1 µg plasmid was digested using 300 U/µl of the appropriate restriction enzyme (New England Biolabs) using the corresponding New England Biolabs buffer at 37°C overnight. The product of this reaction was used to transcribe antisense probe using T3 or T7 enzyme at 37°C for 3 hours. The appropriate enzyme was selected according to the used vector and insert orientation (Table 2.4). The probe was then purified using an RNeasy kit (Qiagen), according to the manufacturers instruction. The resulting riboprobe was mixed 1:1 with hybridisation buffer (50% formamide (Sigma-Aldrich), 5X saline-sodium citrate, 150

µg/ml heparin, 5 mg/ml Torula RNA (Sigma-Aldrich) and 0.2% Tween 20) and stored at -20°C.

Table 2.4: Vectors used to generate ISH probes. Prom = promoter. Sizes are in basepairs. All primers are listed in 5'-3' direction.

Gene	Forward primer	Reverse primer	Vector name	Insert size	Antibiotic	Enzyme	Prom	Other info
cmyb	N/A	N/A	pCMV-SPORT 6.1	3271	Ampicillin	NruI	T7	IMAGE clone ID: 6790859
ea1-globin	ATGAGT CTCTCT GCCAA AGAC	TTATCT GTACTT CTCAG ACAT	pCR-Blunt II-TOPO	440	Kanamycin	SpeI	T7	
mpx	N/A	N/A	pBK-CMV	2830	Kanamycin	SalI	T7	IMAGE. clone ID: 6790891
rag1	TGTAAC CAACA CCTGA ACCCA	CCATCT TCTCAT CATAG CCTGT G	pCR4-TOPO	1200	Ampicillin Kanamycin	NotI	T3	

2.20 *In-situ* hybridisation (ISH)

Embryos were fixed in 4% PFA at 4°C overnight. They were then dehydrated in an increasing series of MeOH (25, 50, 75 and 100%) followed by storage at -20°C in 100% MeOH. After rehydration in a decreasing MeOH series, they were digested using 1.6 µl/ml proteinase K (Roche Diagnostics) in H_2O for one hour, followed by refixation in 4% PFA for 20 minutes. Samples were then incubated in hybridisation buffer for one hour at 56°C. This was exchanged for antisense probe diluted in hybridisation buffer (3:200), which was heated to 80°C for 15 minutes beforehand. The probe was left on overnight, followed by three 45-minute washes with preheated (56°C) washing solutions: first, 2X saline-sodium citrate/PTW, 0.1% Tween 20, 50% hybridisation buffer, second, 2X saline-sodium citrate/PTW, 0.2% Tween 20 and third, 0.2X saline-sodium citrate/PTW, 0.2% Tween 20. The samples were then blocked using blocking solution (5% FBS in PTW). Anti-digoxygenin-alkaline phosphatase (AP) conjugated antibody (Roche Diagnostics) was added at 1:2000 in blocking solution and incubated at 4°C overnight. Samples were washed several times using PTW, followed by a washing solution (0.1 M NaCl, 0.1 M tris-HCl at pH 9.5, 0.05 M $MgCl_2$ and 0.1% Tween 20) for

30 minutes, changing the solution once. One tablet of NBT/BCIP (nitro-blue tetrazolium and 5-bromo-4-chloro-3'-indolyphosphate) (Roche Diagnostics) was dissolved in 10 ml washing solution and added to the samples. Samples were incubated in the dark until sufficient staining was reached. Stained embryos were then washed extensively in PTW and subsequently stored in 80% glycerol. Embryos were sorted into high, medium and low staining categories before the head was removed from the embryo to use its tissue for genotyping. The tail was then flat mounted on a microscope slide for photography.

2.21 Terminal deoxynucleotidyl transferase (TdT) dUTP nick-end labelling (TUNEL)

Embryos for TUNEL assays were fixed in 4% PFA at 2 dpf, transferred to MeOH and rehydrated as for ISH described above. The embryos were then digested for 10 minutes with 0.8 µl/ml Proteinase K, followed by refixation for 20 minutes in 4% PFA. After washing the PFA off with extensive PTW washes, embryos were stained using the Roche Cell Death kit. A master mix of 10 µl enzyme solution and 90 µl label solution was prepared for each sample and added. They were incubated at 37°C for 30 min, followed by quick PTW washes to remove the label solution. All liquid was removed and the samples heated to 80°C for 10 minutes to inactivate endogenous AP. Embryos were quickly washed with PTW, followed by a 1 hour wash in blocking solution (5% FBS in PTW). After that, embryos were incubated in 200 µl blocking solution with AP-conjugated anti-fluorescein antibody (1:2000) overnight at 4°C. AP staining was revealed as for ISH, letting the embryos develop for 30 minutes in NBT/BCIP. Embryos were genotyped, followed by blinded scoring of 10 wild type and 10 mutant embryos per clutch by counting TUNEL foci in the CHT region of the tail. Details on the antibody can be found in Table 2.2.

2.22 Sudan black staining

Embryos were fixed in glutaraldehyde for 20 minutes at RT. Fixed embryos were incubated in Sudan black for 20 minutes at RT. Stained embryos were washed in 70% EtOH until only neutrophilic granules were stained, followed by storage in 80% glycerol.

2.23 Bulk RNA-Seq

Kidney cell suspensions were made as described above. I used two *rad51*$^{+/+}$ and *rad51*$^{-/-}$ fish each in a *Tg(itga2b:EGFP)* background. A BD Influx (BD Biosciences) flow cytometer was used to sort the cells according to fluorescence and their forward and side scatter properties. As described above, I sorted cells into GFP^{low} thrombocyte progenitors (Figure 2.3), erythrocytes, progenitors, myeloid cells/neutrophils and monocytes. I modified the gating slightly to take only the more mature erythrocytes, which tend to be higher in SSC value (Figure 2.7). After excluding debris, dead cells (using PI as a stain) and doublets, 50 cells of each gate were sorted per well, sorting 5 wells per cell type for each fish. Each well contained 2.3 µl of 0.2% Triton X-100 supplemented with 1 U/µl SUPERase In RNase inhibitor (Ambion).

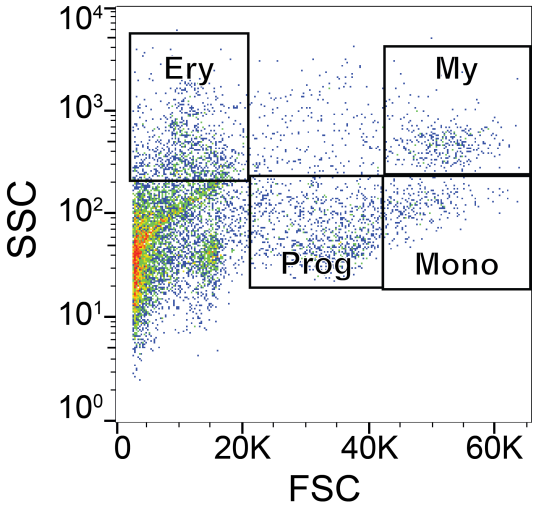


Figure 2.7: Gating strategy for the RNA-Seq experiment. This is a modification of the scheme presented in Figure XX, but with a more stringent gate for mature erythrocytes. Ery = Erythrocytes, Prog = Progenitors, My = Myeloid cells, Mono = Monocytes.

The whole transcriptome was amplified and libraries for sequencing were prepared using the Smart-seq2 protocol[378,379] using External RNA Controls Consortium (ERCC) spike in controls. In short, cDNA was made by reverse transcription using SMARTScribe enzyme (Clontech) (see cycling conditions in Table 2.5), an oligo-dT primer and a template-switching oligo. This step was followed by PCR amplification of the cDNA (see cycling conditions in Table 2.6) and a clean-up of the PCR product using Ampure XP beads (Beckman Coulter). Finally, libraries were generated using the Nextera XT DNA

sample preparation kit (Illumina) and a final clean-up using beads. The resulting pooled libraries were sequenced on the Illumina Hi-Seq2500 platform.

Table 2.5: Cycling conditions for cDNA synthesis. Steps 2-4 were repeated 24 times.

Step	Temperature (°C)	Time (seconds)
1	98	180
2	98	20
3	67	15
4	72	360
5	72	300
6	4	until required

Table 2.6: Cycling conditions for PCR amplification. Steps 2-5 were repeated 12 times.

Step	Temperature (°C)	Time (seconds)
1	72	180
2	95	30
3	95	10
4	55	30
5	72	60
6	72	300
7	10	until required

Reads were aligned to the zebrafish reference genome (Ensemble BioMart version 83) combined with the ERCC spike-ins sequences. Alignment and quantification was performed using Sailfish[380] version 0.9.0 with the default parameters using paired-end mode (parameter –l IU). Transcript Per Million (TPM) values reported by Sailfish were used for the quality control (QC) of the samples. All wells passed the QC since were found to contain more than 1,000 expressed genes (TPM>1) and less than 60% of ERCC or Mitochondrial content. In order to perform the differential expression analysis between wild type and mutant samples for each of the five different gating strategies, estimated counts for each well were used as input to the DEseq2 R package (version 1.14.1)[381]. For each of the five different gating strategies, statistically significant genes that scored $P < .001$ were used to perform Gene Ontology (GO) enrichment analysis. I searched the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database[382] using EnrichR[383] with the human orthologues of the zebrafish genes.

2.24 Stem cell transplantation

2.24.1 Genotyping of *rag2*^{E450fs} recipients

Recipient fish for transplantation were from the *rag2*^{E450fs} line generated by the Langenau laboratory[366]. Fish were genotyped using PCR followed by restriction enzyme digestion. For PCR, the KOD Hot-Start kit from Novagen was used at standard cycling conditions. DNA was amplified using the forward primer 5′-ACTGCTCTAGTTGCAATTCCT-3' and the reverse primer 5′-AGCTGGGGTCATCTTCAGT-3' to produce a 585-bp fragment. This was digested using the restriction enzyme XcmI at 37°C overnight followed by gel electrophoresis. WT fish produce one band at 585 bp. Heterozygotes produce three bands at 585, 372 and 212 bp. Homozygous mutants produce two bands at 372 and 212 bp.

2.24.2 Transplantation

Rag2$^{-/-}$ fish were normally irradiated sub-lethally with a 10 Gy dose from an IBL 437C irradiator using a caesium 137 source. Two days after irradiation, the fish were injected intraperitoneally with 1000 GFP^{low} cells in 10 µl 5% FBS/PBS. These were obtained from the kidneys of donor adults of the appropriate genotype carrying the *Tg(itga2b:EGFP)* transgene. Optionally, 500 stromal cells from *Tg(sdf1a:dsRed)* were added as well. Donor cells were sorted using FACS as described above. The fish were analysed for engraftment using FACS 4 or 16 weeks post transplantation. Any fluorescent cells in the kidney or peripheral blood of recipient fish were scored as engraftment.

2.25 Treatments to trigger stress haematopoiesis

2.25.1 Dose estimations

To estimate the tolerated maximum dose of the substances used, I injected different concentrations of acetaldehyde (Sigma-Aldrich) or polyinosinic:polydytidylic acid (pI:pC) (Sigma-Aldrich), followed by optional qPCR analysis of appropriate marker genes in different tissues at 6 hpi to show efficacy. The respective concentration and markers are indicated in the text.

2.25.2 Long-term pI:pC injections

Fish were injected with 10 mg/ml pI:pC once a week, totalling four injections. Fish were culled three days after the last injection. Blood smears from the peripheral blood were obtained and the kidney was removed for later analysis. Kidneys were processed to single cell suspensions for FACS, as described above. Cells were analysed using an Influx cytometer (BD Biosciences). Remaining cell suspensions after sorting were spun down and kept at -80°C for later gene expression analysis with qPCR.

2.25.3 Long-term pI:pC injections with BrdU

As in the other experiments, 10 mg/ml pI:pC and 10 mg/ml BrdU were used. BrdU was dissolved either in PBS or pI:pC solution. Depending on the experimental group, fish were injected with PBS only, BrdU only, pI:pC only or pI:pC + BrdU. The injection containing BrdU was always the last a fish received, and BrdU-injected fish were culled 1 day after the injection. Tissues were processed according to the normal BrdU staining protocol described above.

2.25.4 Long-term acetaldehyde injections

These injections followed the same scheme as the pI:pC injections, but substituted pI:pC for 1% acetaldehyde (Sigma-Aldrich) in PBS. Fish were culled after four injections, followed by cell counts, FACS and qPCR analysis on the kidney tissue.

2.25.5 Ethanol treatment

Embryos were grown in egg water containing 1.5% ethanol between 4 and 24 hpf. After this period, they were allowed to recover in normal egg water until 4 dpf. At this point, images were taken for scoring of the phenotype.

2.26 Statistical analysis

The data was analysed using Microsoft Excel, InVivoStat 5.3 and GraphPad Prism 6 software. Statistical methods were used as required and are indicated in the text. If necessary, results were corrected for multiple testing using appropriate methods as indicated in the text. Unless otherwise indicated, a (corrected) P-value $< .05$ was considered statistically significant.

2.27 Data visualisation

Flow cytometry data was analysed using FlowJo (Treestar). Graphs were generated in GraphPad Prism 6. Molecular images were created in PyMol using data deposited in the protein data bank (PDB, http://www.rcsb.org/pdb/home/home.do). Images were adjusted using Adobe Photoshop CC. Figures were assembled using Adobe Illustrator CC.

3 Characterisation of *rad51* mutant zebrafish

3.1 Introduction

3.1.1 Models of *rad51* deficiency

Initially, the study of *RAD51* in eukaryotes was limited to yeast. Mutant yeast strains were generated in both budding (*Saccharomyces cerevisiae*) and fission yeast (*Schizosaccharomyces pombe*), the classical workhorses of molecular biology. In both cases, this led to proliferation defects, radiation hypersensitivity and mitotic, as well as meiotic impairment. However, despite these severe impairments, the *RAD51* mutant strains were nevertheless completely viable[299,384,385].

After these early insights using yeast, two groups independently generated murine *Rad51* mutation models to study the role of the gene in mammals, publishing their papers only a few months apart. The first paper described the replacement of the fifth exon in *Rad51* with a neomycin cassette using a gene targeting approach, leading to a disrupted RecA domain[297]. They generated mice carrying the mutant allele and incrossed them to breed a total of 148 pups. However, none of these was a homozygous mutant for the *Rad51* allele. Further work using *in-vitro* fertilisation allowed them to show that *Rad51*$^{-/-}$ mice die early during embryonic development.

The second paper to be published[296] also used a gene targeting approach, this time deleting nucleotides 413-530 of the *Rad51* gene. They also did not find any homozygous mutant pups among 75 offspring. Their analysis showed that *Rad51* mutants develop normally until embryonic day E6, but are much smaller at E7.5, at which point the mother's body starts to resorb them. This decrease in size was caused by a proliferation defect starting at E5.5, which was accompanied by increased apoptosis starting from E7.5, as determined by BrdU incorporation and TUNEL assays respectively. These proliferation defects also extended to murine ES cells carrying the mutation, together with increased radiosensitivity. Furthermore, embryo-derived *Rad51* mutant cells rarely

entered the cell cycle and when they did, they often displayed an abnormal number of chromosomes. Most importantly, they showed that co-mutation with *Tp53* was able to rescue the mutation partially, as double mutants were able to survive until E9.5. The unexpected lethality of the mutation suggests that vertebrate cells are much more reliant on functional HR than lower organisms such as yeast, possibly due to the much larger genome, explaining the early embryonic death of the mutant mice.

Since then, neither conditional, nor hypomorphic, or mosaic *rad51* knockout mice have been generated to work around the embryonic lethality seen in full knockouts. While no published attempts of creating *rad51* knockouts in other vertebrate species exist, there is some evidence from *in-vitro* experiments suggesting that the cell death phenotype observed in mice is also present in other species. A study on the chicken cell line DT40 showed that loss of Rad51 leads to massive proliferation defects coupled to chromosomal damage, culminating in cell death[386].

In summary, these results show that lack of functional Rad51 is tolerable in lower eukaryotes, but leads to excess cell death and proliferation defects in at least some vertebrates, precluding the development of useful model systems to study the role of Rad51 in vertebrate cells.

3.1.2 FA model systems

In contrast to the situation for RAD51 models, there are multiple vertebrate *in-vitro* and *in-vivo* model systems for the study of FA. Nevertheless, their number is quite limited and their appropriateness often disputed. This lack of suitable genetic systems is a severe drawback to the study of the FA pathway in mammals[235].

Several patient-derived FA cell lines for *in-vitro* work exist, but their number is small. Indeed, the FA cell line repository at Oregon Health & Science University (http://www.ohsu.edu/research/fanconi-anemia/celllines.cfm/) only contains six human and eight mouse FA fibroblast lines (two of these are control lines). Furthermore, uncertainty about their physiological relevance and lack of specificity pose further problems to the validity of conclusions derived from these systems[235].

On the *in-vivo* front, almost all work has been on murine FA mutants. A rare, non-murine system is the *Xenopus* system established by the Walter group in Harvard. They used it to elucidate many of the molecular details during ICL repair[189,216,218]. The *Xenopus* system is mostly used for experiments on egg extract, making it mainly useful for *in-vitro* approaches. However, few other groups are focussing on this model with regards to FA.

There has also been limited *in-vivo* work in the FA field on zebrafish embryos[359,361], but this model system has rarely been utilised outside of some initial studies. As both the *Xenopus*, as well as the zebrafish model are rarely used, most of the work remains limited to murine systems.

Mice carrying mutations in *Fanca-p*[387], as well as *Fancr*[296,297] and *Fancs*[113] are currently available. Even though FA mutants for almost all FA genes exist (multiple different mutation sites are available for some of the alleles), there are some significant drawbacks to the use of murine FA models, the main issue being the low overlap between the murine and human disease phenotype[235,358,387,388], as well as a large variability of the phenotype depending on the genetic background of the mice[387].

While mice carrying FA mutations do not always develop congenital abnormalities, many of them do. Most commonly seen are small size, micropthalmia, reduced fertility and sub-Mendelian birth ratios[358,387]. These features do not cover all the congenital features described in patients (see section 1.2.1 for an in-depth discussion), but the symptoms that are present resemble their human equivalent reasonably well. Genes with such phenotypes include *Fanca*[389], *Fancc*[390], *Fancd2*[391], *Fancl*[392], *Fancm*[111] and *Fancp*[253]. However, not all mutations cause such defects; for example, homozygous *Fancg* mice show none of these features[393,394]. Another important feature of FA, cancer predisposition, is also recapitulated in many of the murine models. *Fanca*[389], *Fancd2*[391], *Fancf*[395] and *Fancs*[113] mice have been reported to form a variety of malignancies, ranging from leukaemias to epithelial tumours. As is described above for patients, severe mutations in many of the FA genes are usually lethal. Complete embryonic lethality is seen for *Fancl* (in a pure 129/Sv background)[392], *Fancn*[396,397], *Fanco*[398,399], *Fancr*[296,297], as well as null *Fancd1* and *Fancs* mutants[400]. The most characteristic feature of FA in

patients are the progressively worsening blood abnormalities leading to eventual complete failure of the bone marrow. However, almost all murine FA models show little to no haematopoietic features at all, unless they are challenged to undergo stress haematopoiesis by treatment with inflammation inducing or DNA damaging agents. Such treatments can even lead to complete BMF[10,65,130,134,387,388]. This disparity is presumably due to physiological differences between mice and humans (e.g. differences in life span, speed of metabolism), as well as the highly controlled environment and diet that laboratory mice experience, which may expose them to fewer DNA damaging agents. The only exceptions to the lack of blood defects unless challenged are hypomorphic *Fancd1* mutants, which display HSC proliferation defects[401], *Fancd2* mutants, which have a decreased number of HSCs[128] and *Fancp* mutants, which show a reduction in white blood cells and platelets[253]. The only murine FA model showing spontaneous BMF is a conditional *Fancs*/*Brca1* knockout model[113]. These conditional knockouts avoid the complete embryonic lethality of conventional *Fancs* mutations[400]. *Brca1* mutant mice develop macrocytic anaemia as early as one month after birth and have decreased WBC counts. This then progresses to spontaneous BMF in ~30% of the mice. Cancers (leukaemias and lymphomas) are also common, occurring in ~45% of the animals. These haematopoietic defects were linked to reduced HSPC number and function[113]. Due to the problems of recapitulating FA features in model systems, there is a significant need for the development of better model systems both for *in-vitro* and *in-vivo* research. I attempted to fill this gap with the *rad51* zebrafish model.

3.2 Results

3.2.1 Embryonic non-haematopoietic phenotypes

3.2.1.1 The *rad51*^{sa23805} allele leads to complete loss of functional Rad51 protein

To study the effects of Rad51 deficiency, I obtained zebrafish carrying the *rad51*^{sa23805} allele from the Wellcome Trust Sanger Institute Zebrafish Mutation Project (ZMP). The founding mutation was generated by ENU mutagenesis, as described in previously[402]. The *rad51*^{sa23805} allele has a C>T mutation at codon 203 in exon 7, leading to a premature stop codon in the catalytically important AAA+ ATPase domain (Figure 3.1A). To confirm the absence of full length Rad51 protein in homozygous *rad51* mutants, I carried out Western blotting on testis tissue. This organ was chosen as Rad51 has been previously described to be highly expressed in germ cells of other species[316], which I could confirm in zebrafish based on data from an unrelated RNA-Seq experiment[377]. The Western blot conclusively showed that full-length Rad51 is lost in *rad51*$^{-/-}$ fish (Figure 3.1B). I was unable to detect any new smaller bands, which would have indicated a truncated form of the protein. To verify the specificity of the antibody and thus the validity of the Western blotting results, immunostaining was carried out. This experiment, carried out by Ewa Bielczyk-Maczyńska (E.B.-M), used an anti-Rad51 antibody to detect Rad51 foci in irradiated *rad51*$^{+/-}$ incross embryos, followed by genotyping (Figure 3.1C). This strategy was used, as the formation of Rad51 foci is necessary for the repair of double-stranded breaks arising from irradiation-induced DNA damage. As expected, Rad51 foci appeared after irradiation in *rad51*$^{+/+}$ embryos, demonstrating the functionality of the antibody. In contrast, *rad51*$^{-/-}$ embryos did not develop any Rad51 foci, indicating that the protein is lost (Figure 3.1Ci and ii). The results of this experiment thus confirmed the Western blots and validated that *rad51*$^{-/-}$ fish completely lack functional Rad51.

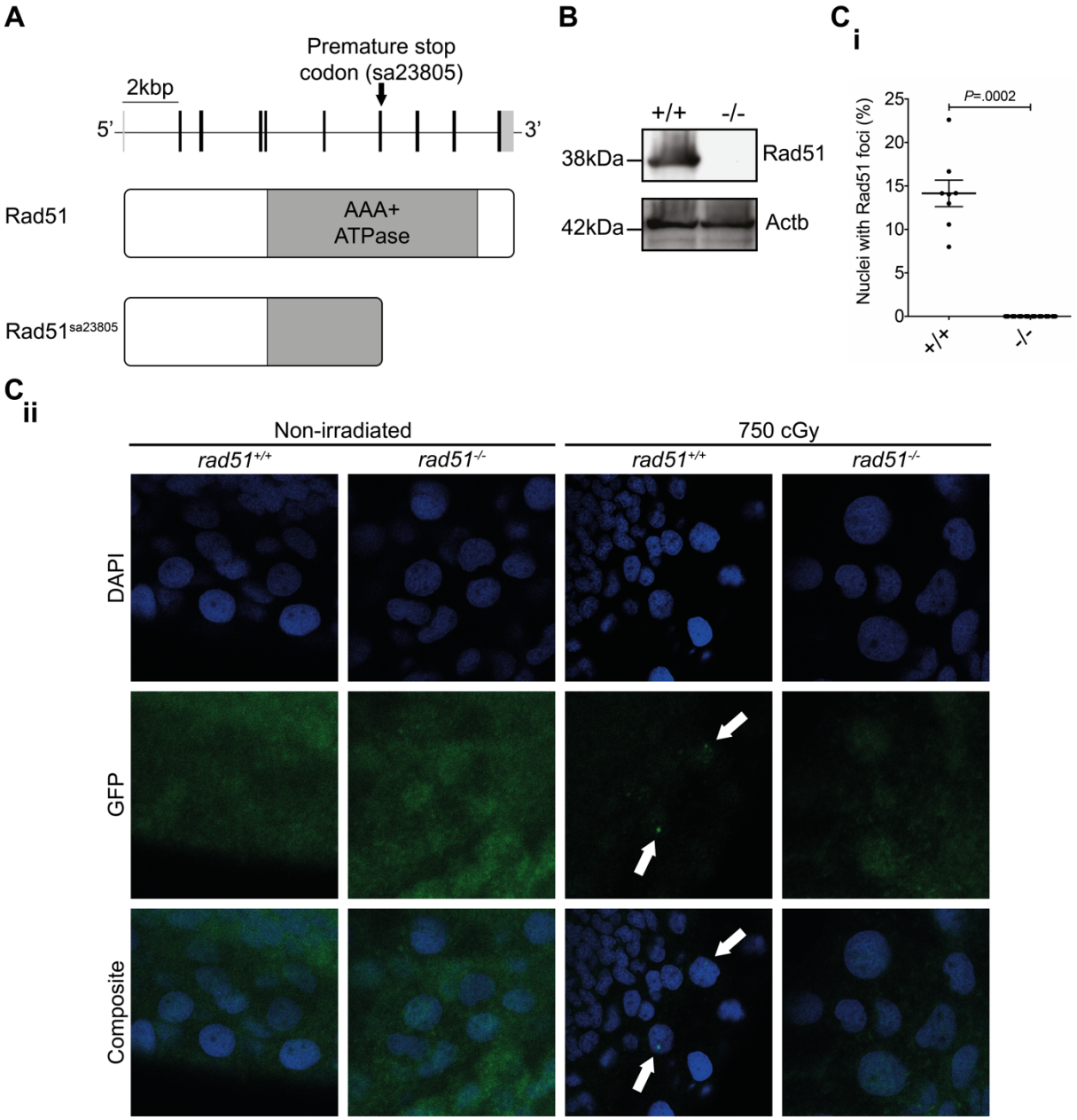


Figure 3.1: The *rad51*^{sa23805} allele leads to complete loss of functional Rad51 protein. (A) The *sa23805* allele contains a premature stop codon in exon seven, which leads to a truncation of the full-length Rad51 protein in the middle of the catalytically important AAA+ ATPase domain. (B) Western blot of testis tissue showing that the Rad51 protein is completely lost. (C) Immunostaining against Rad51 on embryos before and after irradiation. Quantification of Rad51 foci following irradiation (i) showed that mutants lack any Rad51 protein Foci are indicated using arrows in example images (ii). Mann-Whitney test, $P = .0002$, $n_{+/+} = 8$, $n_{-/-} = 8$. Images were obtained using a 40X water immersion objective Data shown in C was obtained by E.B.-M.

3.2.1.2 Embryos lacking Rad51 develop micropthalmia

Two of the most common features of FA are decreased height and micropthalmia. To investigate this, the body and eye size of 5 dpf embryos were measured (Figure 3.2A). While *rad51*$^{-/-}$ embryos were the same overall size as their wild type (WT) siblings (Figure 3.2B) at 5 dpf, their eyes were about 7% smaller on average (Figure 3.2C), already resembling the micropthalmia commonly seen in FA patients.

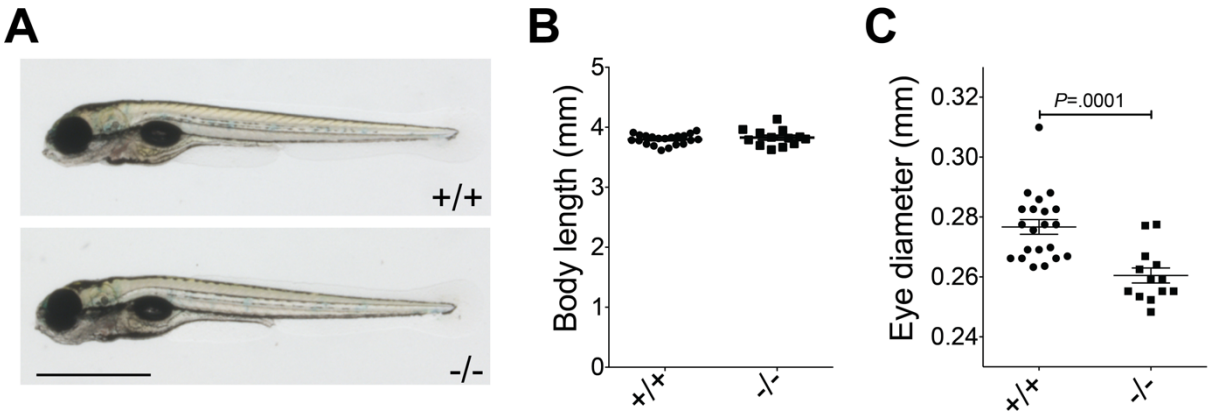


Figure 3.2: Embryos with *rad51* mutation are micropthalmic. (A) Representative 5 dpf embryos. Magnification = 16X. Measurement of total body length (B) showed no difference between WT and mutant embryos, but eye size (C) was significantly different between WT and mutant fish. Two-tailed Student's t-test, P = .0001, $n_{+/+}$ = 21, $n_{-/-}$ = 13. Bars represent mean ± SEM in both graphs. Data shown in this figure was obtained by E.B.-M.

3.2.1.3 The embryonic DNA damage response is impaired in *rad51* mutant embryos

After confirming the loss of Rad51, the next aim was to characterise the effects this would have on the DNA damage response during embryonic development. DNA damage remains unrepaired in *rad51* mutants, as immunostaining of a *rad51*$^{+/-}$ incross using a phospho-histone H2AX (pH2AX, a marker of double stranded breaks) antibody at 2 dpf showed (Figure 3.3A). Unrepaired DNA damage had a detrimental effect on the developing organism. Embryos lacking *rad51* developed small eyes and heads in response to gamma radiation at 2 dpf (Figure 3.3B), presumably because the rapidly dividing cells of the central nervous system are especially susceptible to DNA damage, leading to excess apoptosis. This strongly exacerbated the already decreased eye size in untreated embryos (Figure 3.2C) and also decreased the overall body size of the embryos, which

was previously unaffected (Figure 3.2B), making the irradiated phenotype easily recognisable by eye (Figure 3.3Bi).

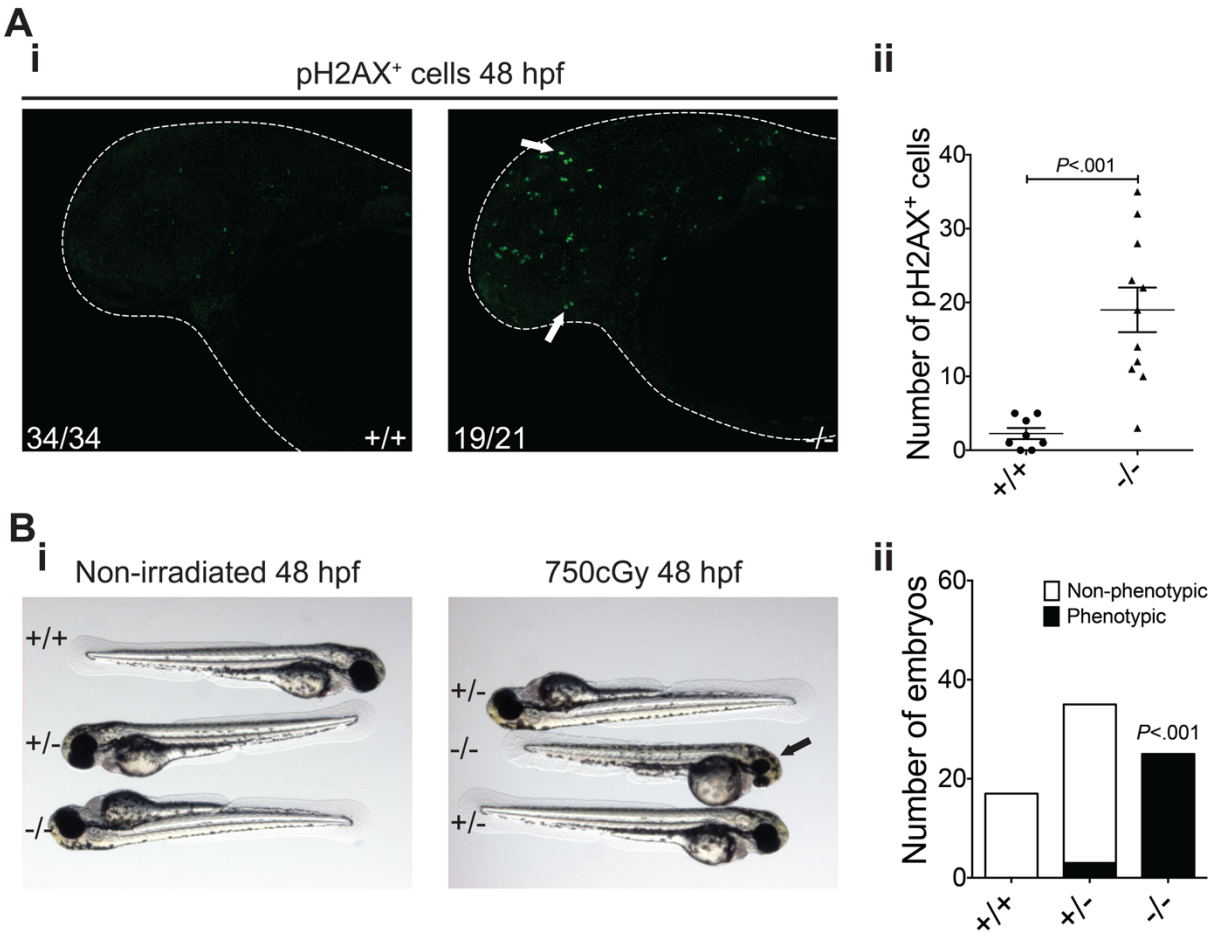


Figure 3.3: Loss of *rad51* leads to DNA damage sensitivity. (A) Immunostaining for pH2AX in wild type and mutant embryos with representative embryos (i) and quantification of foci (ii). White arrows indicate example foci. Images were obtained with a 40X water immersion objective. Two-tailed Student's t-test, $P < .0001$, $n_{+/+} = 8$, $n_{-/-} = 11$. Bars represent mean ± SEM. (B) Comparison of the response of 48 hpf wild type and mutant embryos to irradiation (i). The black arrow indicates the small head and eye phenotype, which is quantified in ii. Magnification = 16X. Two-tailed Fisher's exact test pooling wild types and heterozygotes as control group, $P < .001$, n = 67. Data shown in this figure was obtained by E.B.-M.

Next, I tested the response of the *rad51*$^{-/-}$ embryos to several DNA damaging drugs. After assessing the maximum tolerated dose for each drug on WT zebrafish embryos (defined as the highest tolerable concentration without any noticeable developmental abnormalities at 24 hpf), I treated the embryos with the drugs between 4 and 24 hpf, followed by chromosome spreads and genotyping of the individual embryos. I utilised two agents that are widely used in the diagnosis of FA, DEB and MMC, which induce

interstrand crosslinks, as well as the topoisomerase I inhibitor CPT and the PARP1 inhibitor DiQ. Chromosomes were blindly photographed and karyotyped by me, followed by genotyping to determine the mutation status of each spread. My results indicated that embryos lacking functional Rad51 cannot successfully repair ICLs, leading to characteristic chromosome breaks and radial structures seen exclusively in *rad51*$^{-/-}$ embryos (Figure 3.4A and B) upon DEB and MMC treatment, consistent with an FA-like phenotype and further demonstrating the DNA damage susceptibility of *rad51* mutants. MMC seemed to induce more premature chromatid separation events than DEB. CPT induced many DNA breaks, but relatively few radial structures (Figure 3.4C), showing the lack of functional HR. Treatment with DiQ however, induced little damage in both WT and mutant embryos (Figure 3.4D), indicating that PARP1 inhibition does not disproportionally affect *rad51* mutants.

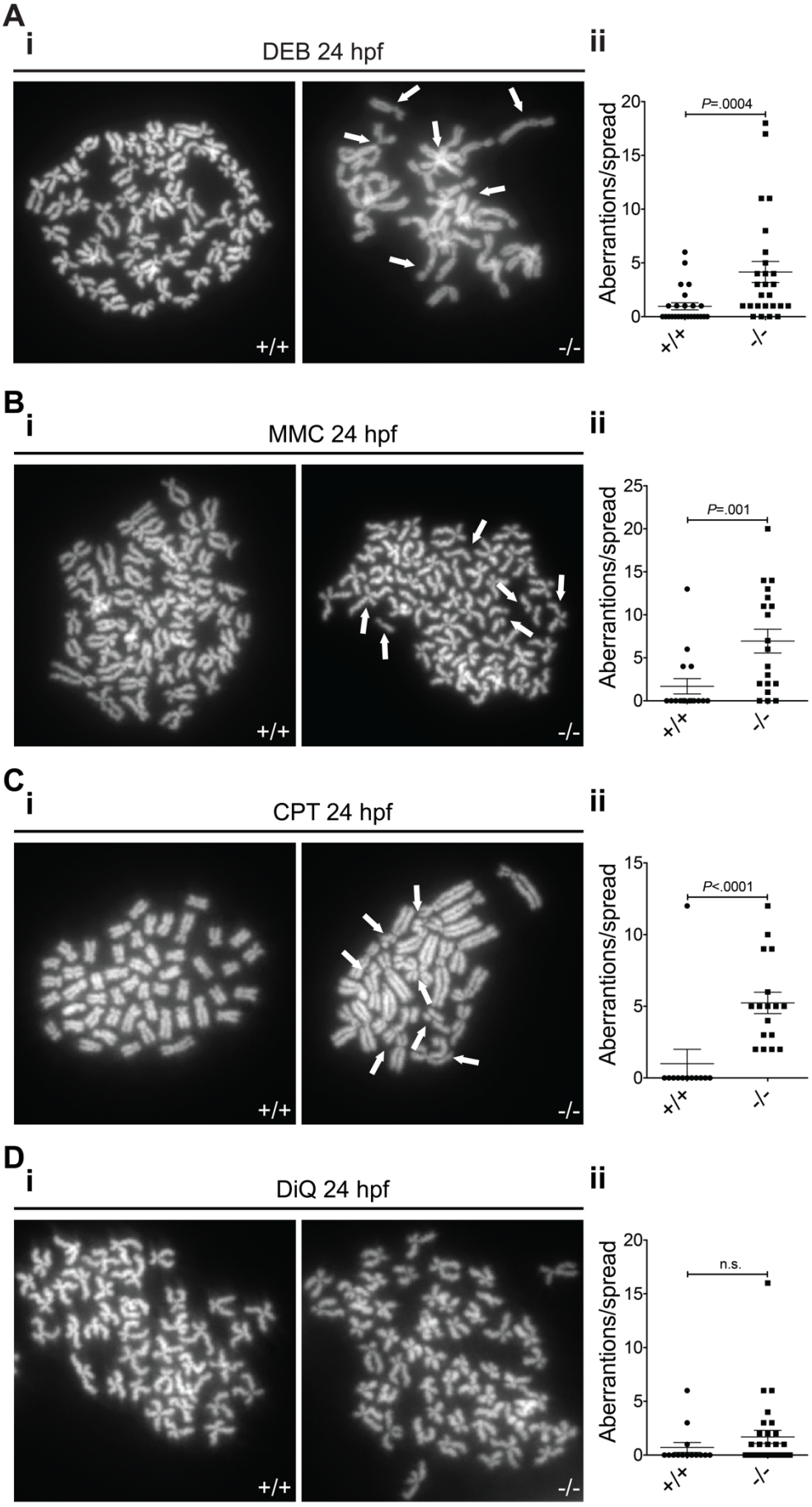

A
i
DEB 24 hpf
+/+
-/-
ii
Aberrantions/spread
P=.0004
+/+
-/-
B
i
MMC 24 hpf
+/+
-/-
ii
Aberrantions/spread
P=.001
+/+
-/-
C
i
CPT 24 hpf
+/+
-/-
ii
Aberrantions/spread
P<.0001
+/+
-/-
D
i
DiQ 24 hpf
+/+
-/-
ii
Aberrantions/spread
n.s.
+/+
-/-

Figure 3.4: *rad51* mutation leads to crosslinker sensitivity. (A) Chromosome spreads of 24 hpf wild type and mutant embryos treated with 1 µg/ml DEB for 20 hours. White arrows indicate characteristic damage (chromosome breaks and radial structures) in response to crosslinking agents (i). Quantification of the damage (ii). Mann-Whitney test, $P = .0004$, $n_{+/+} = 25$, $n_{-/-} = 26$. (B) Chromosome spreads of 24 hpf wild type and mutant embryos treated with 5 µg/ml MMC for 20 hours. White arrows indicate premature chromatid separation events (i). Quantification of the damage (ii). Mann-Whitney test, $P = .001$, $n_{+/+} = 16$, $n_{-/-} = 19$. (C) Chromosome spreads of 24 hpf wild type and mutant embryos treated with 1 nM CPT for 20 hours. White arrows indicate breaks and radial structures in response to topoisomerase inhibition (i). Quantification of the damage (ii). Mann-Whitney test, $P < .0001$, $n_{+/+} = 12$, $n_{-/-} = 17$. (D) Chromosome spreads of 24 hpf wild type and mutant embryos treated with 10 µM DiQ for 20 hours (i). Quantification of the damage (ii). Mann-Whitney test, $P = .13$, $n_{+/+} = 14$, $n_{-/-} = 29$. All images were obtained using a 100X oil immersion objective. Bars represent mean ± SEM in all graphs.

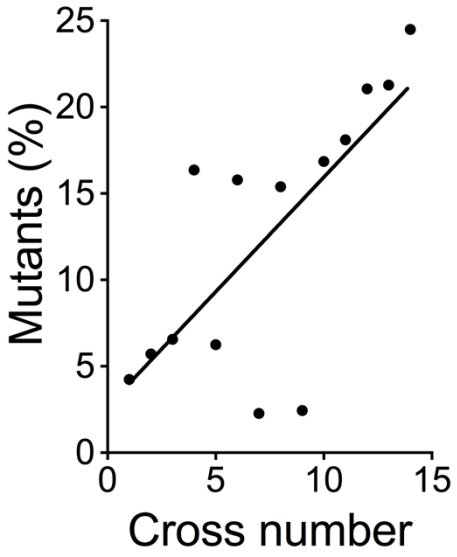


Figure 3.5: Outcrossing increases mutant survival. The percentage of mutants at 3 mpf in each cross (25% expected) is plotted against the cross number. A lower number tends to indicate less outcrossing, as older, more inbred lines get replaced by more outcrossed parent lines. The regression line shows a linear fit of the data. A significant regression equation was found. $F(1, 12) = 12.87$, $P = .004$, $R^2 = 0.52$.

I continued my investigation by growing *rad51* mutants to adulthood. In contrast to mice lacking *Rad51*, these fish survived to adulthood. In my first crosses, I observed sub-Mendelian survival of the *rad51*$^{-/-}$ fish. However, after several rounds of outcrossing, this effect went away (Figure 3.5). This means that *rad51* mutants do not inherently have a sub-Mendelian birth ratio, but might be more susceptible to the presence of other,

harmful mutations. Outcrossing was presumably able to prevent such deleterious mutations from accumulating and therefore brought survival back to expected levels.

Next, I investigated potential reasons for the survival of zebrafish *rad51* mutants to adulthood in contrast to the early death of mice lacking *Rad51*[296,297]. Early embryonic development is highly dependent on maternally-derived RNA and proteins[403]. Indeed, zebrafish development is reliant on maternally-derived factors until the mid-blastula transition (MBT) at approximately 4 hpf [404]. The first RNAs start being expressed at 2 hpf, reaching full activation of the zygotic genome around 3.7 hpf [405]. One possible reason for the survival of *rad51*$^{-/-}$ zebrafish embryos is a difference in the contribution of maternal RNA between mice and fish. I examined Rad51 expression using immunostaining with two different primary anti-Rad51 antibodies in freshly fertilised eggs stemming from a *rad51*$^{+/-}$ incross (Figure 3.6), as any protein present so early must stem from the mother. Embryos displayed a diffuse, but clearly recognisable signal corresponding to Rad51 in the cytoplasm. As embryos at this stage are not amenable to genotyping due to the low amount of DNA, I assumed a Mendelian ratio of mutants (one quarter) for the interpretation of the experiment. As all embryos stained with primary antibody (Figure 3.6B and C) showed considerably higher staining than the secondary antibody only controls (Figure 3.6A), I concluded that maternal contribution of Rad51 cannot be ruled out as a cause for the viability of *rad51* mutants to adulthood.

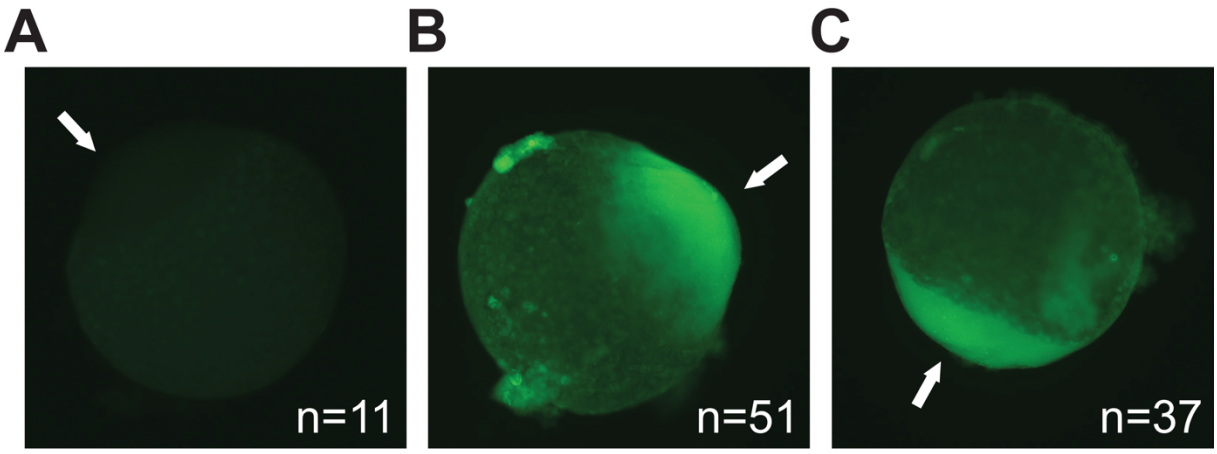


Figure 3.6: Rad51 is expressed in single-cell stage embryos. Representative Rad51 immunostained embryos derived from a *rad51*$^{+/-}$ incross. (A) Secondary only, n = 11, (B) Abcam primary, n = 51, (C) AnaSpec primary, n = 37. Magnification = 60X.

Finally, I considered the role of NHEJ in the survival of the mutant fish. To do that, I treated an incross of *rad51*$^{+/-}$ parents with a range of different concentrations of SCR-7, a DNA ligase IV inhibitor and scored the embryos at 24 hpf (Figure 3.7). At the lowest concentration, embryos were completely unaffected, whereas at the higher concentration all embryos died. At no concentration were the mutant embryos more sensitive than their WT siblings, indicating that NHEJ is dispensable for the survival of *rad51*$^{-/-}$ embryos.

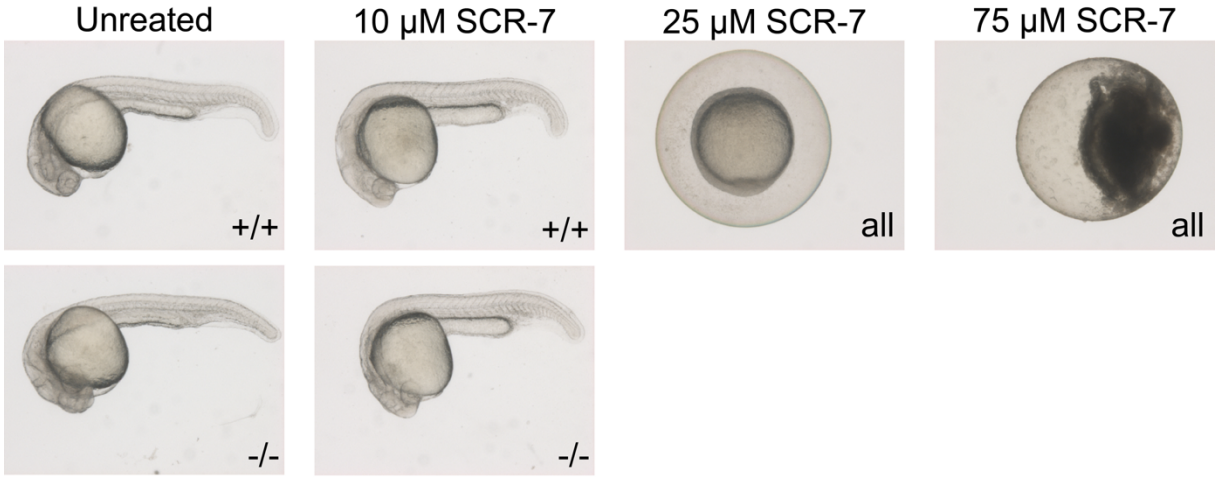


Figure 3.7: NHEJ inhibition does not lead to synthetic lethality in *rad51* mutants. Morphology of 24 hpf embryos treated with varying concentrations of SCR-7 for 20 hours. For each concentration, n = 40 treated embryos. Magnification = 16X.

3.2.1.4 Rad51b partially compensates for Rad51 loss

In addition to the *rad51* mutants, E.B.-M. generated a *rad51l1 (*the zebrafish orthologue of human *RAD51B)* mutant using CRISPR with a frameshift mutation in exon 2. As both mutants were viable and blood defects subtle in the case of *rad51* and not present at all in *rad51b* mutants, I decided to generate double *rad51, rad51b* mutants to study potential compound effects of the mutations. I carried out two independent matings of *rad51*$^{+/-}$, *rad51b*$^{+/-}$ fish, with a total of 204 fish surviving until 3 months post fertilisation (mpf) for genotyping. However, among these 204 fish there were no double mutants, even though 12.75 would have been expected due to Mendelian inheritance (Table 3.1). I repeated the same cross and collected embryos at 4 dpf. Again, zero out of 44 embryos were double mutants, even though 2.75 would have been expected. Finally, I repeated the cross and genotyped at 6 hpf (the earliest point the genotyping works reliably). Out of 47 embryos,

none were double mutants. This early embryonic synthetic lethality of co-mutation indicates that *rad51* and *rad51b* have at least partially redundant functionality, which may explain the viability of the single mutants.

Table 3.1: Co-mutation of *rad51* and *rad51b* is lethal. Punnet square of the genotyping results at 3 mpf stemming from two independent *rad51*$^{+/-}$, *rad51b*$^{+/-}$ incrosses. Actual numbers are in black, whereas expected numbers according to Mendel's rules are in blue. The lack of double mutants is highlighted in red. Overall n = 204.

rad51b \ *rad51*	+/+	+/-	-/-	
+/+	1/12.75	18/25.5	11/12.75	30/51
+/-	25/25.5	64/51	14/25.5	103/102
-/-	48/12.75	23/25.5	0/12.75	71/51
	74/51	105/102	25/51	

n = 204

3.2.2 Adult non-haematological phenotypes

After having shown that *rad51* mutants are viable to adulthood, presumably due to maternal effects, I proceeded with characterising the phenotype of the fish during adulthood.

3.2.2.1 *Rad51* mutants display congenital defects resembling FA

Just as for embryos, I measured the size of juvenile and adult *rad51* mutants (Figure 3.8). This revealed that in contrast to embryos, there was a size difference between WT and mutant fish starting in the juvenile period, which was maintained over the whole observed lifespan of the fish. Importantly, all adult fish appeared to be males, presumably due to sex determination defects during the juvenile phase. This was not due to female fish dying, as mutants were born at the expected ratio, indicating sex reversal. Zebrafish sex determination is explained in more detail in section 3.3.

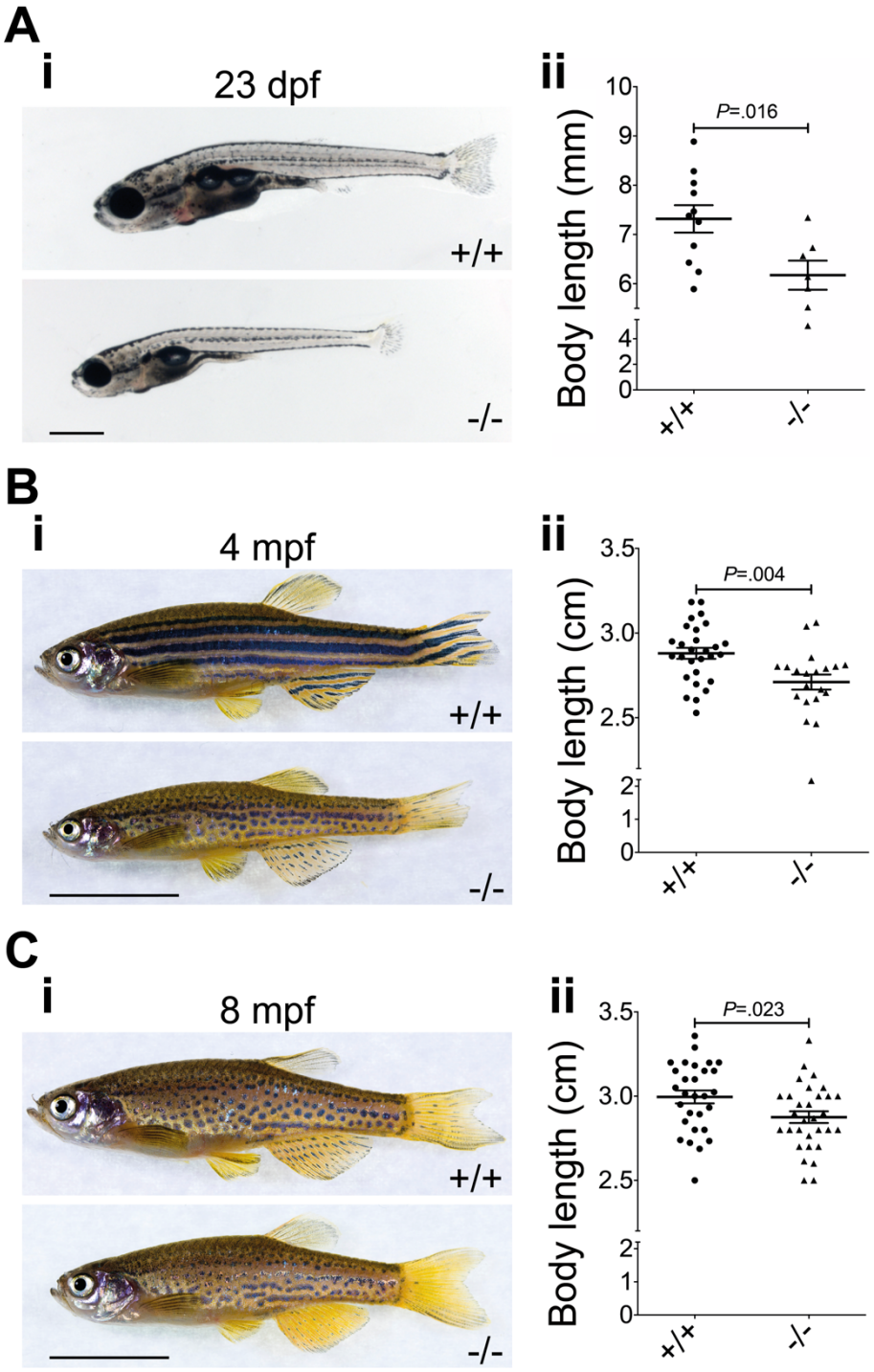


Figure 3.8: Adult *rad51* mutant fish are smaller than their wild type siblings. (A) Representative images of mutant and WT fish at 23 dpf (i). Magnification = 16X. Graph showing difference in size between the genotypes (ii). Two-tailed Student's t-test ($P = .0016$), $n_{+/+} = 11$, $n_{-/-} = 7$. Scale bar = 1 mm. (B) Comparison of mutant and wild type fish at 4 mpf (i). Graph showing difference in size between the genotypes (ii). Two-tailed Student's t-test ($P = .004$), $n_{+/+} = 28$, $n_{-/-} = 19$. Scale bar = 1 cm. (C) Representative images of mutant and WT fish at 8 mpf (i). Graph showing difference in size between the

genotypes (ii). Two-tailed Student's t-test (P = .023), $n_{+/+}$ = 29, $n_{-/-}$ = 31. Scale bar = 1 cm. Bars represent mean ± SEM in all graphs.

None of my attempts to breed these *rad51*$^{-/-}$ males resulted in any offspring, as all resulting eggs were unfertilised. Mating behaviour was not affected, as seen by a large number of unfertilised eggs whenever matings were attempted. I then tried to obtain sperm from the *rad51* mutants. As can be seen in Figure 3.9A, wild type fish had milky, thick sperm, whereas mutant fish only yielded thin, watery secretions, explaining their inability to father offspring. To investigate the cause of this lack of sperm, I obtained histological sections of 4 mpf *rad51* mutant testes, which were stained with standard H&E staining for visualisation (Figure 3.9B). As can be seen on these micrographs, fish lacking *rad51* cannot form mature spermatozoa, leading to many empty spaces in the testes.

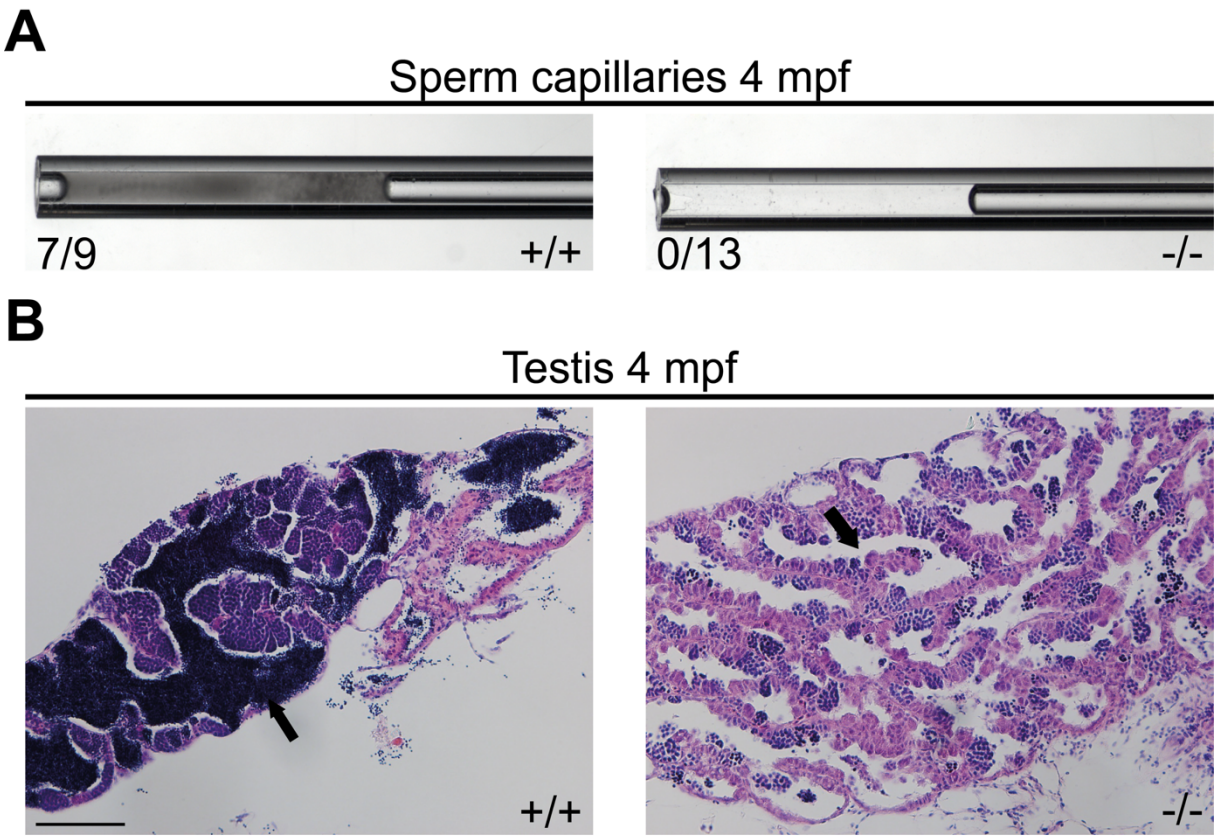


Figure 3.9: *rad51* mutant fish are all infertile males. (A) Capillaries used to obtain sperm from wild type and mutant fish. Numbers indicate the number of fish giving sperm among the fish tested. Two-tailed Fisher's exact test, P = .0002. $n_{+/+}$ = 9. $n_{-/-}$ = 13. (B) H&E stained histological sections of 4 mpf wild type and mutant kidneys using a 20X objective. The arrow points out the area where mature spermatozoa reside. Scale bar = 100 µm.

Other tissues appeared normal during dissection. For a more detailed look, I obtained histological sections of the intestine (Figure 3.10), which has highly proliferative cells that turn over very quickly. I reasoned that it might therefore be more affected by increased DNA damage sensitivity than many other tissues. However, no clear difference could be seen between WT and mutant intestine tissue, suggesting that this tissue has normal functionality.

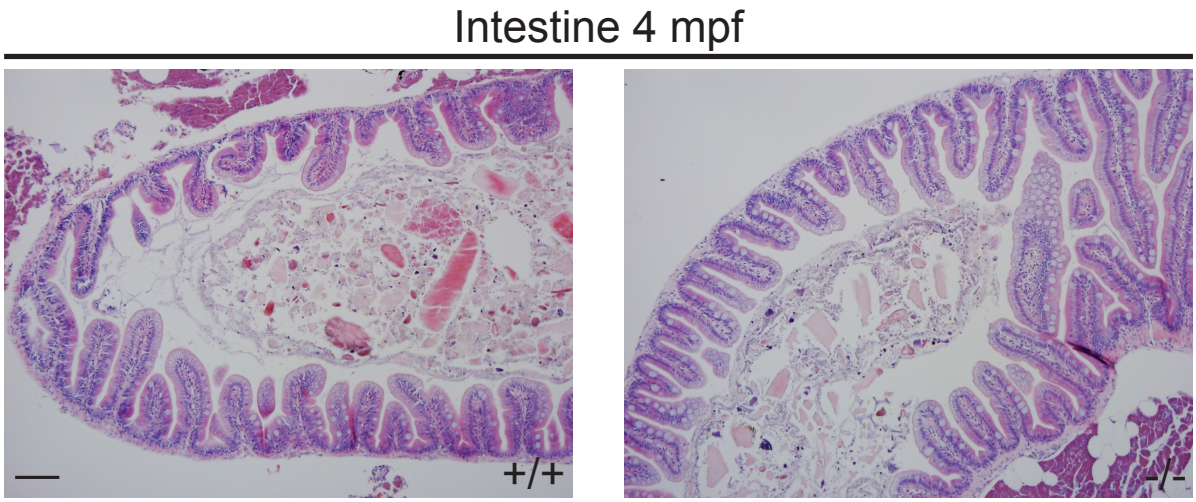


Figure 3.10: The intestine is unaffected in *rad51* mutant fish. Representative H&E stained histological sections of 4 mpf wild type and mutant intestines using a 10X objective. Scale bar = 100 µm.

3.2.2.2 Lack of rad51 leads to excess DNA damage in exposed tissues

To look at other genes linked to the FA pathway, I obtained tissue from the brain, eyes, fin and heart of *rad51*$^{+/+}$ and $^{-/-}$ fish, extracted the RNA and generated cDNA for qPCR. I considered a DNA damage marker (*gadd45ab),* NHEJ related genes (*lig4, prkdc, xrcc4*), genes encoding Rad51 binding proteins (*brca2* and *rad52*), cell cycle genes (*rb1*), as well as genes of the p53 pathway (*p21* and *p53*). Most genes were unchanged, but interestingly the skin showed a roughly tenfold, statistically significant upregulation of the *gadd45ab* gene (Figure 3.11), which is a marker for DNA damage. This shows that lack of Rad51 leads to increased DNA damage in adults as well, but this is only noticeable on the tissue level in organs exposed to a lot of DNA damage.

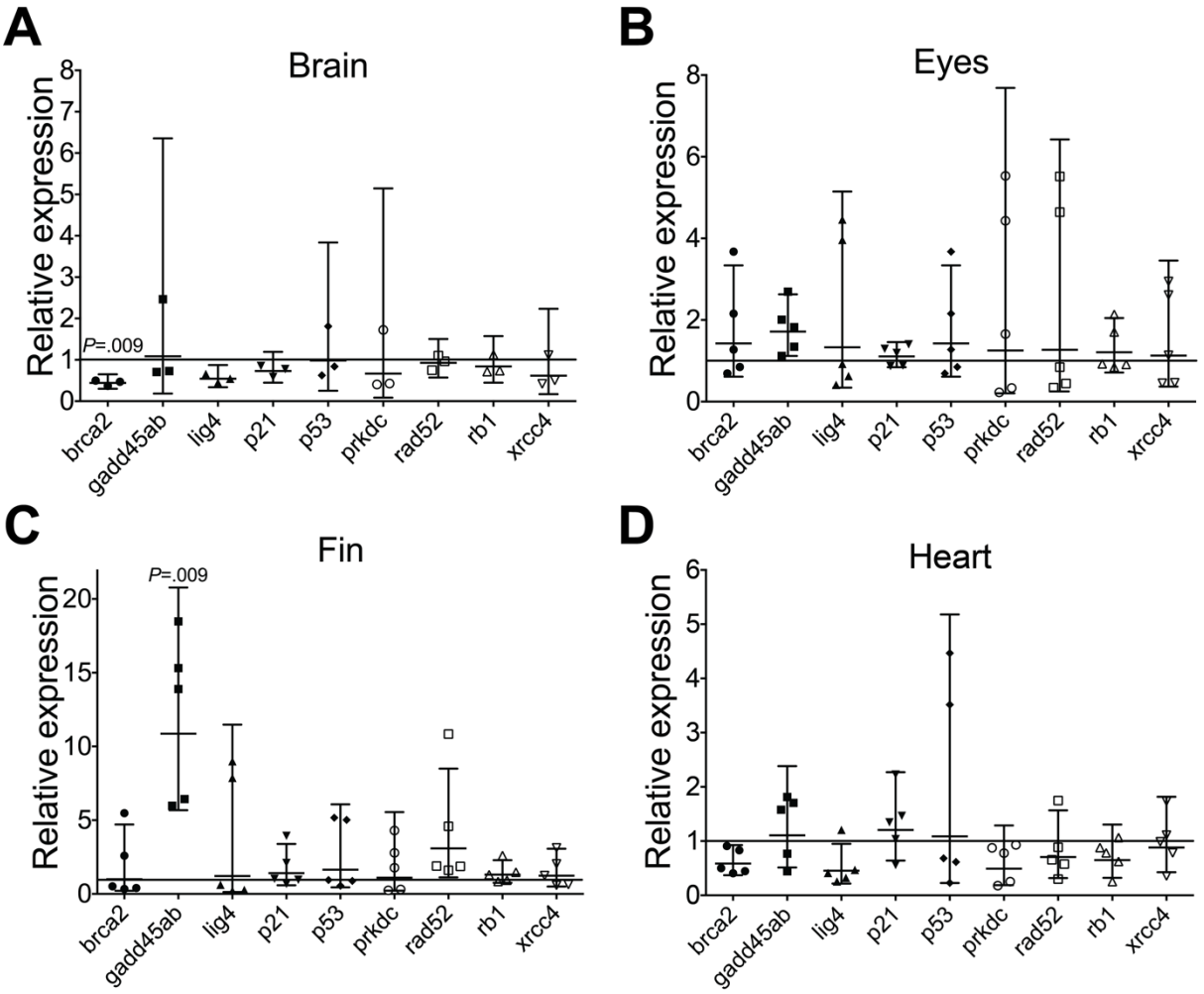


Figure 3.11: qPCR analysis of different tissues. (A) Brain. $n_{+/+} = 3$. $n_{-/-} = 3$. (B) Eye. $n_{+/+} = 4$. $n_{-/-} = 5$. (C) Fin. $n_{+/+} = 4$. $n_{-/-} = 5$. (D) Heart. $n_{+/+} = 4$. $n_{-/-} = 5$. Individual plots shown refer to the relative expression in mutants, the line at 1 represents the geometric mean of the WT fish. Bars represent the geometric mean ± 95% confidence interval to estimate fold changes. All *P*-values stem from two-tailed Student's t-tests with Bonferroni correction for multiple testing.

3.2.3 Haematological phenotypes

3.2.3.1 Adult *rad51* mutants display a hypocellular kidney marrow

The key features of FA are progressive marrow defects, leading to gradually worsening haematological impairments and eventual complete BMF. Because of this, I investigated the zebrafish kidney marrow (WKM) for defects. I observed the morphology of the WKM by obtaining histological sections of the kidney stained with H&E (Figure 3.12A). This allowed me to determine that the WKM was morphologically normal, without any obvious empty spaces or dysplastic cells. However, during the dissection of the kidney I

noticed that the *rad51*$^{-/-}$ kidneys appeared considerably smaller than the WT kidneys (Figure 3.12Bi). I quantified this size difference between 4 and 13 mpf by counting WKM cellularity with a haemocytometer (Figure 3.12Bii). This showed that mutants had roughly half as many cells in the kidney as WT fish and that this difference was maintained at the same level over the whole life of the fish. Interestingly, both WT and mutant WKM cellularity decreased significantly with age, a change which I did not expect.

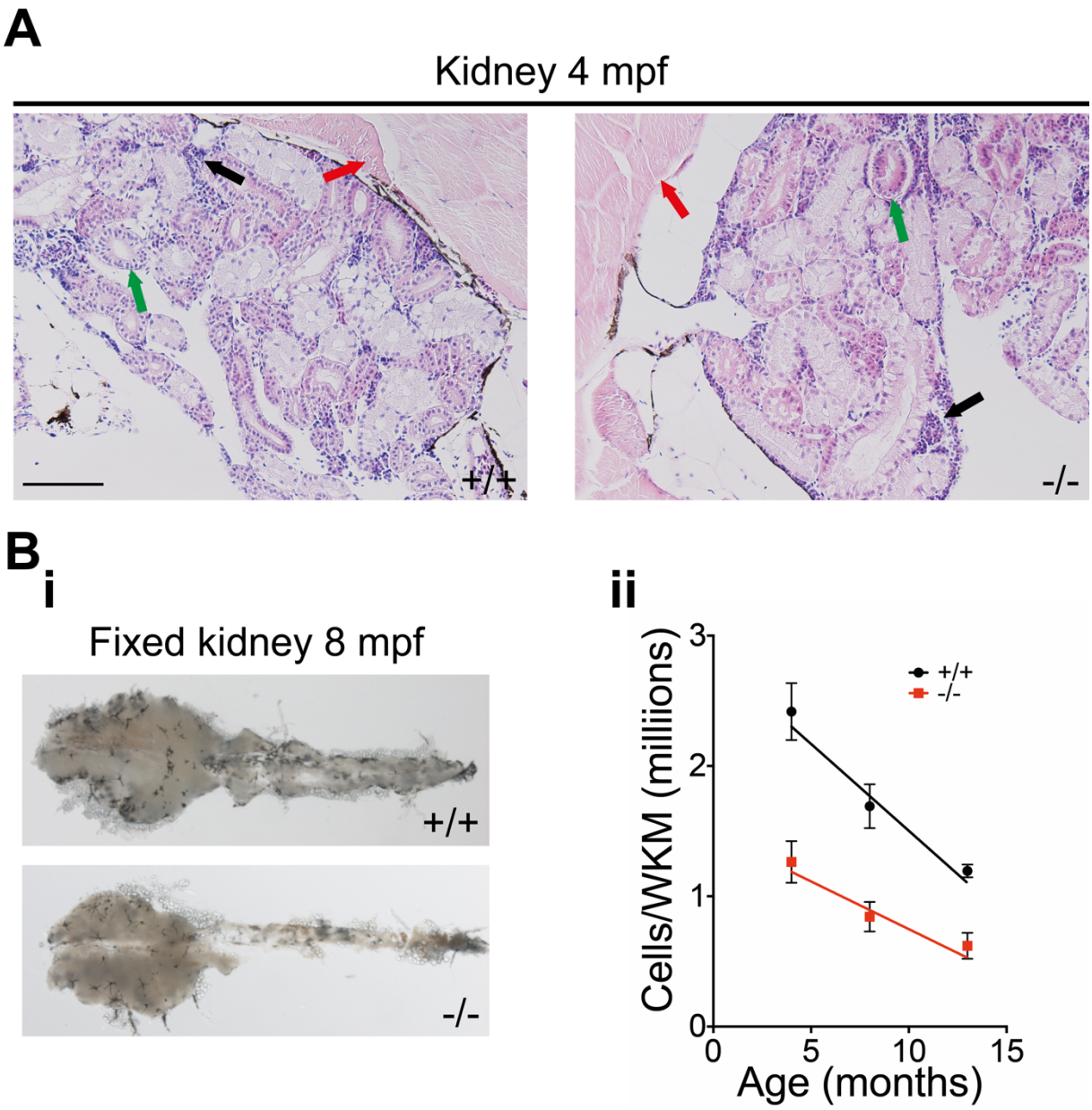


Figure 3.12: WKM cell number is halved in mutants. (A) Representative images of H&E stained histological sections of 4 mpf wild type and mutant kidneys using a 20X objective. Muscle (red arrow),

ducts/tubules (green arrow) and haematopoietic kidney marrow (black arrow) can be seen. Scale bar = 100 µm. (B) Fixed 8 mpf wild type and mutant kidneys (i). Magnification = 8.5X. Quantification of the number of total cells per freshly isolated kidney at different ages using a haemocytometer (ii). Two-way ANOVA was used and type III model fit. The test showed a significant influence of age ($F(1, 50) = 18.23$, $P < .0001$) and mutation status ($F(1,50) = 10.87$, $P = .0018$) on phenotype, but no interaction between these two factors ($F(1,50) = 1.54$, $P = .22$). 4 mpf $n_{+/+} = 6$, $n_{-/-} = 6$; 8 mpf; $n_{+/+} = 16$, $n_{-/-} = 16$; 13 mpf $n_{+/+} = 6$, $n_{-/-} = 4$. Bars represent mean ± SEM.

Next, I investigated the effect of age on the proportion of the different blood lineages present in the WKM (Figure 3.13). During ageing, the ratio of the different blood lineages in the WKM shifted towards the myeloid lineage (Figure 3.13B). However, this effect was stronger in mutants than in their WT siblings (Figure 3.13B).

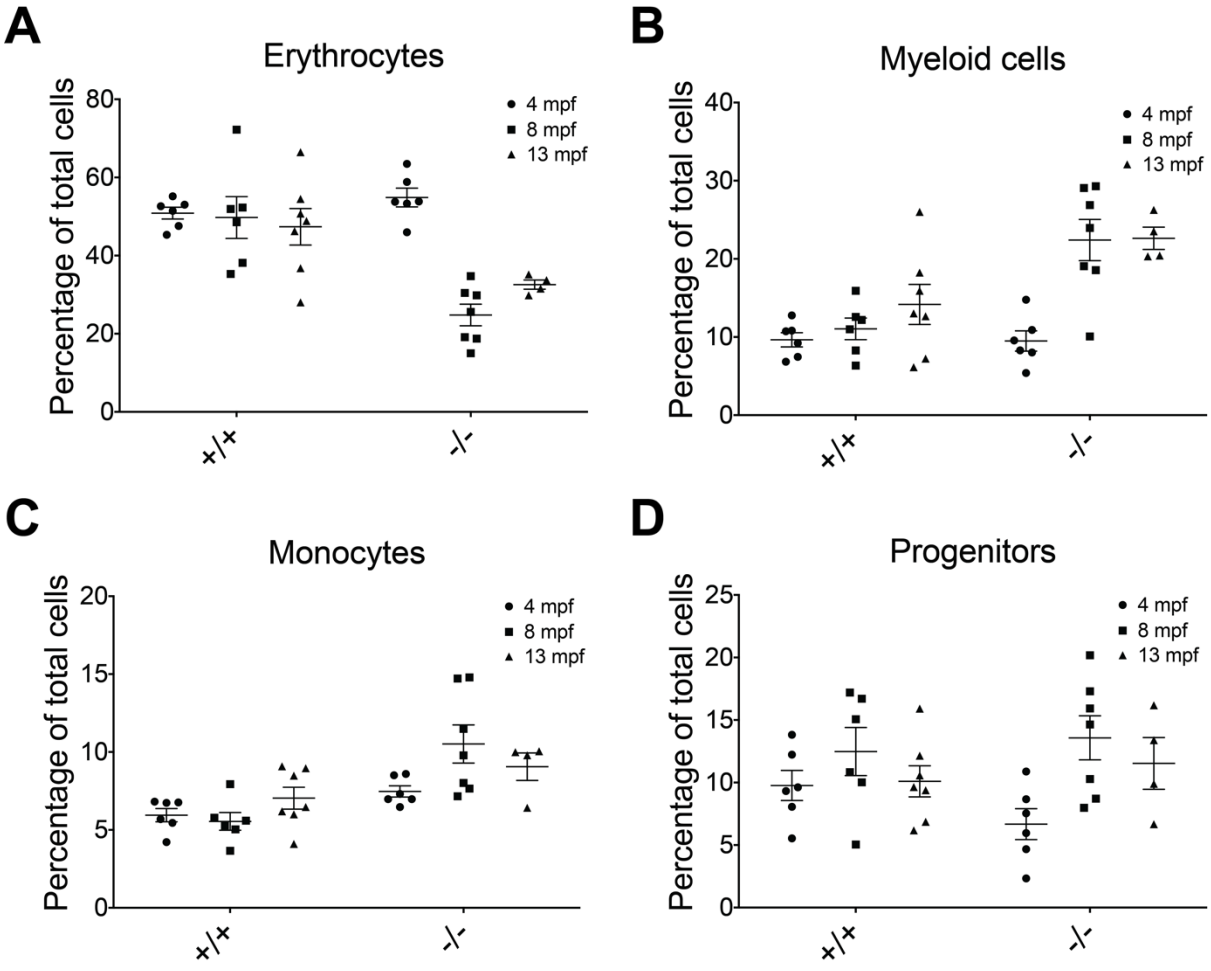


Figure 3.13: Ageing causes a shift towards myeloid cells in the WKM. (A) Percentage of erythrocytes in the WKM. Two-way ANOVA revealed a significant difference due to mutation status ($F(1, 30) = 15.7$, $P = .0004$) and of age ($F(2,30) = 10.69$, $P = .0003$), as well as an interaction between mutation status and age ($F(2,30) = 8.46$ $P < .0012$). (B) Percentage of myeloid cells in the WKM. Two-way ANOVA revealed a

significant difference due to mutation status ($F(1, 30) = 25.46$, $P = .0005$) and of age ($F(2, 30) = 10.52$, $P = .0003$), as well as an interaction between mutation status and age ($F(2, 30) = 4.524$ $P < .019$). (C) Percentage of monocytes in the WKM. Two-way ANOVA revealed a significant difference due to mutation status ($F(1, 30) = 18.76$, $P = .0002$), but not due to age ($F(2, 30) = 1.87$, $P = .17$) or an interaction between the two ($F(2, 30) = 2.87$, $P = .072$). (D) Percentage of progenitors in the WKM. Two-way ANOVA revealed a significant difference due to age ($F(2, 30) = 4.83$, $P = .015$), but not due to mutation status ($F(1, 30) = .02$, $P = .89$) or an interaction between the two ($F(2, 30) = 1.25$, $P = .3$). For all graphs: 4 mpf $n_{+/+} = 6$, $n_{-/-} = 6$; 8 mpf $n_{+/+} = 6$, $n_{-/-} = 7$; 13 mpf $n_{+/+} = 7$, $n_{-/-} = 4$. Bars represent mean ± SEM in all graphs.

3.2.3.2 *Rad51* mutants develop macrocytic erythrocytes with ageing

Subsequently I considered changes in the peripheral blood (PB). Neither WT nor mutant fish ever developed clear signs of anaemia (such as pallor) and neither did I find a higher death rate of mutants during the observed life span. To get a more quantifiable picture of the PB, I measured the number of erythrocytes in the PB at 4 mpf and found no significant difference (Figure 3.14A). To consider morphological defects as well, I obtained blood smears at different ages and quantified the nuclear and overall size of the erythrocytes (Figure 3.14B). This showed that cytoplasmic volume increased over time in the mutants (Figure 3.14Bi and ii). Changes were particularly visible when I calculated the ratio of the cytoplasmic to nuclear volume (Figure 3.14Biii). The ratio was essentially constant in both WT and mutant fish until 13 mpf, at which points the value increased considerably and significantly in the mutants only. This revealed that while the mutants did not develop acute anaemia until 13 mpf, they nevertheless developed macrocytic erythrocytes, an early sign of haematological impairments.

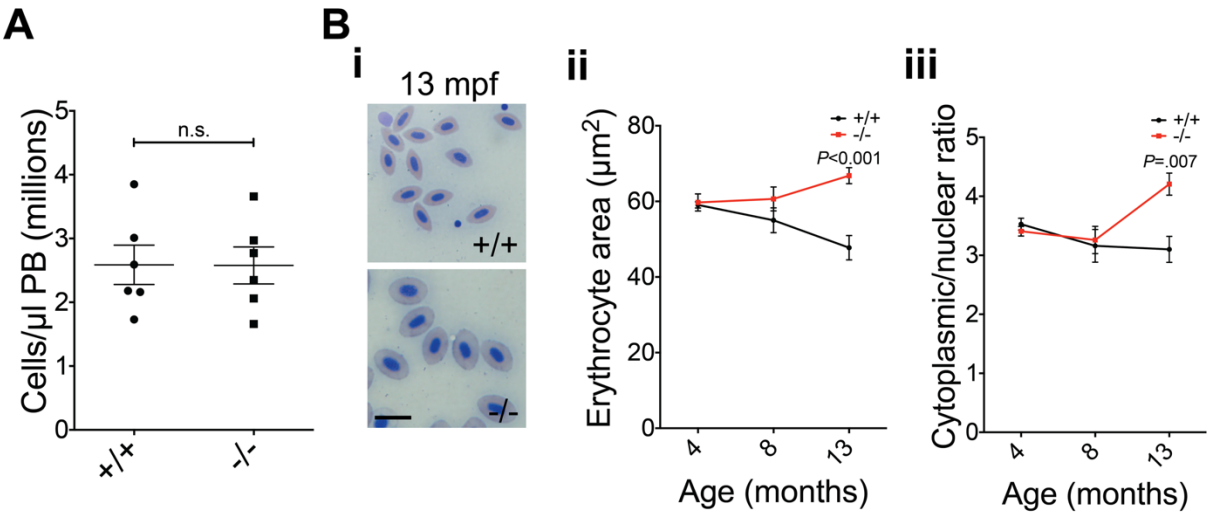


Figure 3.14: Mutants develop macrocytic erythrocytes. (A) Quantification of peripheral blood (PB) cells in wild type and mutant fish at 4 mpf. Two-tailed Student's t-test, $P = .98$, $n_{+/+} = 6$, $n_{-/-} = 6$. (B) Comparison of erythrocytes seen in blood smears of 13 mpf wild type (top) and mutant fish (bottom) (i). Scale bar = 10 µm. Quantification of the erythrocyte area (ii). There was a statistically significant interaction between age and mutation status ($F(1, 28) = 12.89$ $P = .0012$), no significant influence of age ($F(1, 28) = 180.76$, $P = .392$) and no significant influence of mutation status ($F(1, 28) = 2.88$, $P = .106$). Quantification of the cytoplasmic/nuclear ratio (iii). There was a statistically significant interaction between age and mutation status ($F(1, 28) = 449.7$ $P = 001.$), no significant influence of age ($F(1, 28) = 26.38$, $P = 39$) and no significant influence of mutation status ($F(1, 28) = 100.62$, $P = .1$). All data in B was analysed using two-way ANOVA and a type III model fit. *P*-values shown on the graphs stem from *post-hoc* Tukey multiple comparison tests. 4 mpf $n_{+/+} = 6$, $n_{-/-} = 6$; 8 mpf $n_{+/+} = 5$; $n_{-/-} = 5$, 13 mpf $n_{+/+} = 6$, $n_{-/-} = 4$. Bars represent mean ± SEM in all graphs.

3.2.3.3 Apoptosis is not increased in the WKM of adult *rad51*$^{-/-}$ fish

After establishing that *rad51* mutant fish develop kidney marrow hypocellularity, I aimed to gain a more mechanistic understanding of why this decrease happens in *rad51* mutants. The number of cells in the kidney is determined by the balance between the number of stem cells generated during development and the balance between apoptosis and proliferation during adulthood. I first assessed whether this difference was due to excess apoptosis in adult *rad51*$^{-/-}$ fish by carrying out Annexin V-Propidium Iodide staining on the dissected WKM followed by FACS (Figure 3.15). In this assay, cells about to undergo apoptosis are labelled because they express the protein marker annexin V, which is upregulated during apoptosis. This experiment revealed that there was no difference in the number of apoptotic cells between WT and mutant fish at 4 mpf (Figure

3.15B). Accordingly, apoptosis in the WKM of adult fish can be ruled out as a cause for the decreased WKM cellularity.

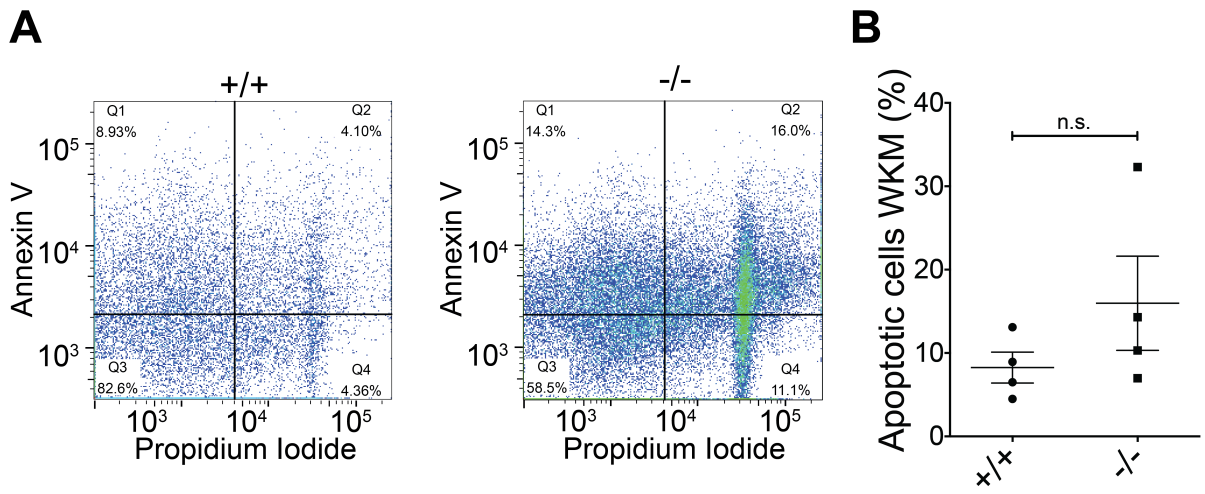


Figure 3.15: Apoptosis is not increased in the WKM of *rad51* mutants. (A) FACS profiles from an annexin V-PI assay to detect apoptotic cells. Representative results of WT (left) and mutant (right) WKM are shown. Apoptotic cells are found in Q1. (B) Quantification of apoptotic cells in the WKM. Two-tailed Student's t-test, $P = .24$, $n_{+/+} = 4$, $n_{-/-} = 4$. Bars represent mean ± SEM.

3.2.3.4 Establishing a BrdU pulse-chase system in adult zebrafish

After excluding apoptosis as a possible cause for the decreased WKM cellularity, I focussed on the proliferation rate of the HSPCs. To do this, I developed a novel BrdU injection assay using adult zebrafish, which allowed me to label newly generated cells. BrdU is a nucleoside analogue of thymidine. Upon injection, it is transported into all tissues by the blood, where it is incorporated into the DNA of dividing cells, replacing thymine. This allows the labelling of recently generated cells using BrdU-specific antibodies, which are conjugated to fluorophores or other labels. These can then be detected by flow cytometry or other methods. Before carrying out experiments on the mutant fish line, I aimed to establish the protocol on WT fish first. I started an initial trial experiment, in which I injected fish with 10 µl 10 mg/ml BrdU intraperitoneally and culled them four hours, six hours, one day, seven days, 14 days and 28 days post injection (dpi) (Figure 3.16A). After staining the WKM cells using anti-BrdU-fluorescein antibodies, I obtained FACS data to quantify the number of $BrdU^+$ cells. This showed that $BrdU^+$ cells started appearing very quickly in the WKM and PB, even after only 4 hours post injection (hpi). Peak labelling in the WKM was reached at 1 dpi, whereas it

was 7 dpi for the PB, although the error bars were very large due to the low number of fish used. At 28 dpi, there were only very few BrdU$^+$ cells present in the WKM. (Figure 3.16B). Because labelling in the WKM was highest at 1 dpi and almost undetectable at 28 dpi, I limited my analysis of the mutant line to this period.

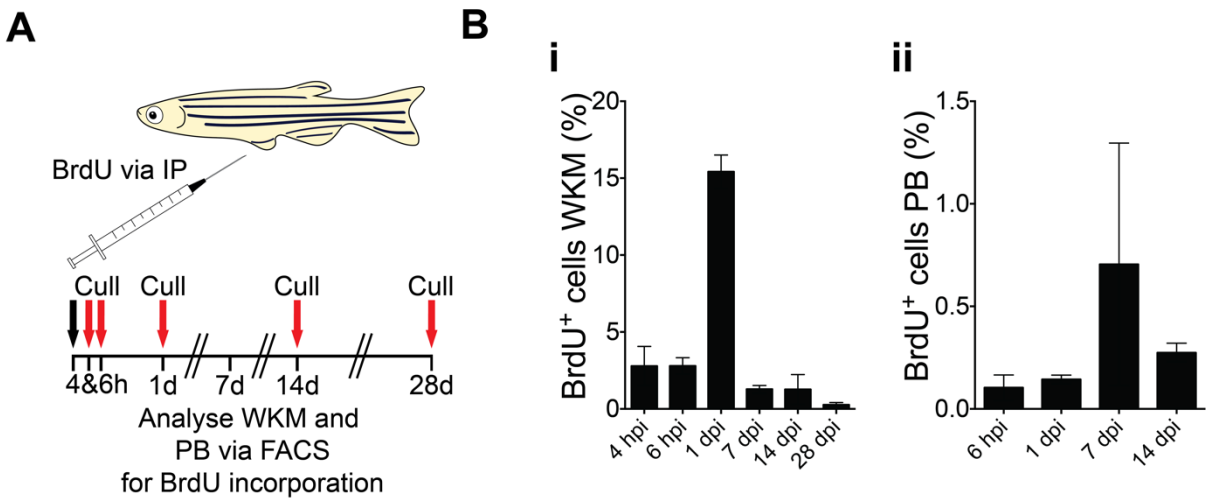


Figure 3.16: Establishing a BrdU pulse-chase system in zebrafish. (A) In a pilot experiment, I injected two WT fish for each time point (4 and 6 hours, 1, 7, 14 and 28 days post injection) with 10 mg/ml BrdU. (B) FACS analysis for BrdU incorporation. Percentage of BrdU-labelled cells in the WKM (i) and the PB (ii). n = 2 at each time point. Bars represent mean ± SEM in both graphs.

3.2.3.5 Proliferation is paradoxically increased in the WKM of *rad51*$^{-/-}$ fish

In my experimental design to compare WT and mutant fish, both groups were injected with BrdU and culled at 1, 7 or 14 dpi, followed by analysis for BrdU incorporation in the WKM and PB using FACS (Figure 3.17A). The results indicated that *rad51*$^{-/-}$ fish rapidly accumulated BrdU$^+$ cells in the WKM following injection, which were then diluted quickly to reach the same level as in the WT (Figure 3.17Bi), as the initial BrdU label was diluted in further divisions. In the PB, BrdU$^+$ erythrocytes appeared much quicker in the circulation in the mutants, but this level plateaued soon after (Figure 3.17Bii). Conversely, WT fish initially had fewer BrdU$^+$ cells in the PB, but then slowly reached the same level as the mutants (Figure 3.17Bii). Together, these results show that HSPCs in *rad51*$^{-/-}$ fish are proliferating significantly more than HSPCs in WT fish, even though overall WKM cellularity was reduced. Hence a decrease in proliferation cannot explain the lowered cell number in *rad51*$^{-/-}$ fish.

To accumulate more evidence for an increase in proliferation, I focussed on transgenic lines in which newly made blood cells are labelled (Figure 17C and D). In the *Tg(gata1a:EGFP)* line, newly made erythrocytes are labelled[368]. In agreement with my previous data, *rad51* mutants had about twofold more GFP$^+$ cells in the kidney at 4 mpf (Figure 3.17C), as shown by FACS analysis.

In the *Tg(itga2b/cd41:EGFP)* line[367], thrombocytic progenitors and mature thrombocytes are labelled in the GFP^{low} and GFP^{high} populations respectively[375]. My analysis showed that at 4 mpf, *rad51*$^{-/-}$ fish had about twice as many GFP^{low} thrombocytic progenitors than the WT fish, but the number of mature thrombocytes was not changed (Figure 3.17D). This shows that the progenitors of multiple lineages were dividing more quickly, i.e. the increase in proliferation was not driven by only one lineage.

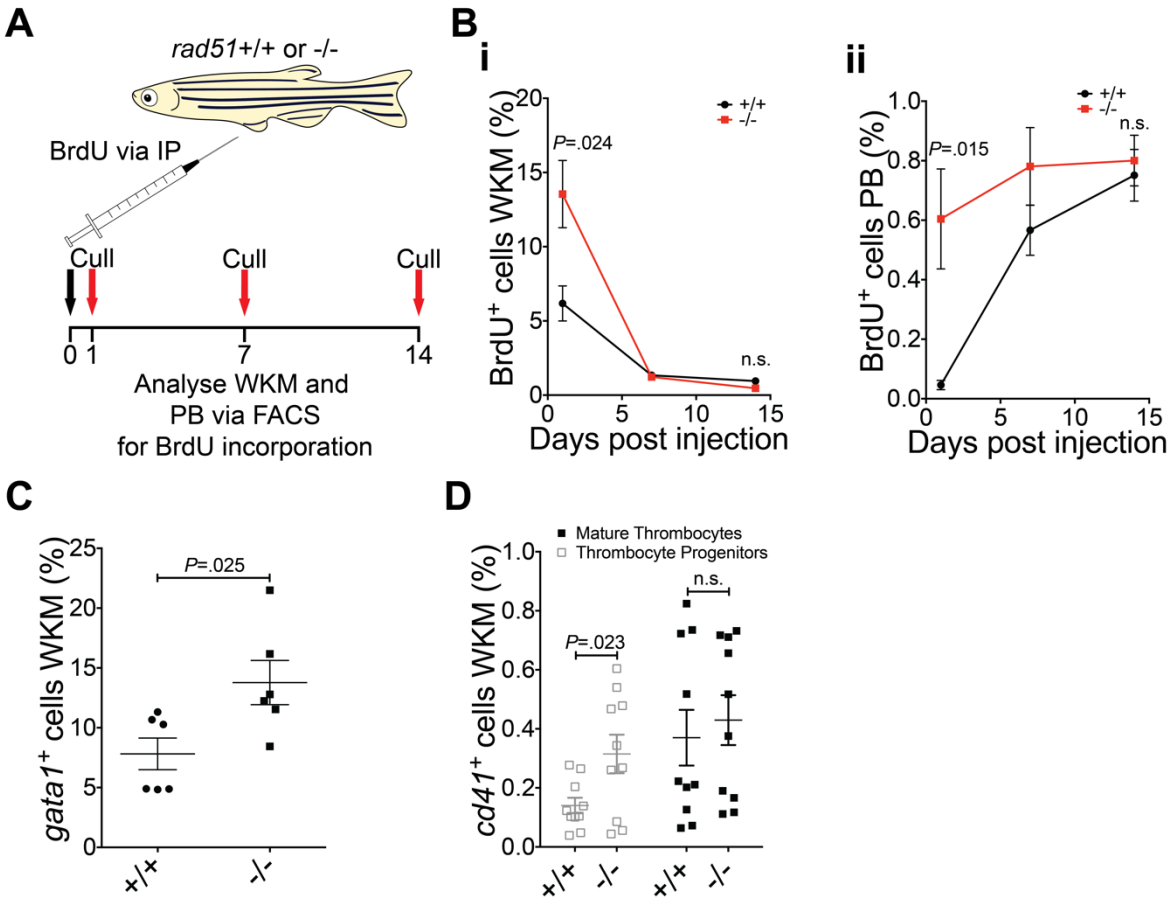


Figure 3.17: Proliferation is increased in the WKM of *rad51* mutants. (A) Schematic of the experimental design for the BrdU incorporation experiments. Fish were injected once with 10 mg/ml BrdU and culled after 1, 7 or 14 days to obtain the blood and kidney marrow for antibody staining and FACS analysis. (B) FACS results quantifying BrdU-labelled cells. (i) Percentage of BrdU$^+$ cells in the kidney. Two-

sided Student's t-test, $P = .024$ at 1 dpi and $P > .05$ at 14 dpi. 1 dpi $n_{+/+} = 5$, $n_{-/-} = 6$; 7 dpi $n_{+/+} = 6$, $n_{-/-} = 6$; 14 dpi $n_{+/+} = 5$, $n_{-/-} = 5$. (ii)Percentage of BrdU+ cells in the peripheral blood. Two-sided Student's t-test, $P = .0015$ at 1 dpi and $P = .64$ at 14 dpi. 1 dpi $n_{+/+} = 5$, $n_{-/-} = 6$; 7 dpi $n_{+/+} = 6$, $n_{-/-} = 5$; 14 dpi $n_{+/+} = 5$, $n_{-/-} = 5$. (C) Percentage of *gata1*:GFP^{+} cells in the kidney at 4 mpf. Two-sided Student's t-test, $P = .025$, $n_{+/+} = 6$, $n_{-/-} = 6$. (D) Percentage of GFP^{low} and GFP^{high} *cd41*:GFP^{+} cells in the kidney at 4 mpf, labelling thrombocytic progenitors and mature thrombocytes respectively. Two-tailed Student's t-test. Thrombocytic progenitors: $P = .023$, mature thrombocytes: $P =$ n.s., $n_{+/+} = 10$, $n_{-/-} = 10$. Bars represent mean ± SEM in all graphs.

As neither apoptosis, nor proliferation during adulthood could explain the hypocellular kidney marrow, I characterised the embryonic haematological development of the *rad51* mutants in detail.

3.2.3.6 *Rad51* mutants form fewer HSPCs during embryonic development

To assess embryonic HSPC levels, I carried out a *cmyb in-situ* hybridisation, which specifically labels HSPCs, on an incross from *rad51*$^{+/-}$ parents at 2 and 4 dpf (Figure 3.18A and B). After staining the embryos, they were blindly sorted into high, medium and low staining categories (20 embryos per category). The head and tail of each embryo were separated, followed by genotyping using the head tissue and photography of the tail, allowing me to connect ISH staining to genotype without biasing the sorting results. This revealed a small, but noticeable decrease in *rad51* mutants at 2 dpf (Figure 3.18A), which was exacerbated at 4 dpf (Figure 3.18B). To find out the cause of this decrease, I carried out a BrdU incorporation assay on the embryos, genotyped them and assessed proliferation levels (Figure 3.18C and D). This showed that there was a statistically significant decrease in proliferation at 2 dpf (roughly twofold) (Figure 3.18C), but not at 4 dpf (Figure 3.18D), although there was still a trend towards lower proliferation. This means that the main difference in proliferation happens very early between 2-4 dpf. I also assessed the levels of apoptosis at 2 dpf, using a TUNEL assay (Figure 3.18E and F) This time point was chosen as it corresponds to the time point where I saw the largest difference in proliferation. This showed that apoptosis was significantly increased (also roughly twofold) in *rad51*$^{-/-}$ embryos at 2 dpf (Figure 3.18F). Thus, the lower level of HSPC formation in *rad51* mutants stems from a combination of decreased proliferation and increased apoptosis during early embryonic development.

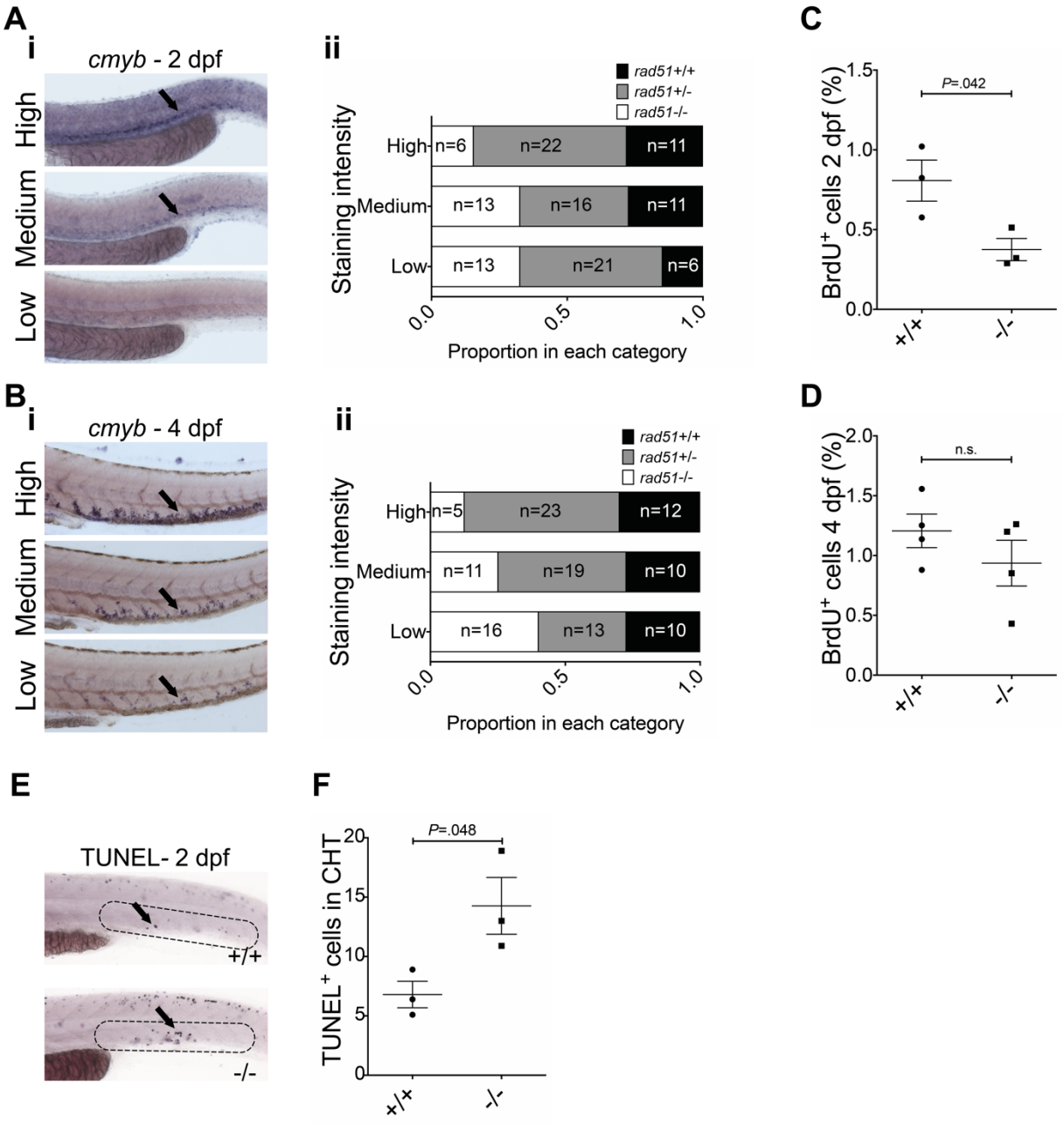


Figure 3.18: The *rad51*^{sa23805} HSPC defect starts during embryonic development. (A) Representative 2 dpf embryos stained using a *cmyb* ISH probe (i) and quantification of distribution of genotypes in the different staining categories (ii). The arrow shows HSPCs. n = 119 from two clutches. (B) Quantification of BrdU$^+$ cells in the tail at 2 dpf. Two-tailed Student's t-test, P = .042, $n_{+/+}$ = 3, $n_{-/-}$ = 3. (C) Representative 4 dpf embryos stained using a *cmyb* ISH probe (i) and quantification of distribution of genotypes in the different staining categories (ii). The arrow shows HSPCs. Representative images of the three different staining categories are shown in (i) and a quantification of the different genotypes in (ii). n = 119 from two clutches. n = 120 from two clutches. (D) Quantification of BrdU$^+$ cells in the CHT at 4 dpf. Two-tailed Student's t-test, P = .3, $n_{+/+}$ = 4, $n_{-/-}$ = 4. Bars represent mean ± SEM in B and D. (E) Representative images of TUNEL-stained 2 dpf embryos from a *rad51*$^{+/-}$ incross. Dotted lines indicate the area of the CHT that

was scored. Arrows indicate example TUNEL^{+} cells. (F) Quantification of three clutches of TUNEL-stained 2 dpf *rad51*$^{+/-}$ incrosses. Each clutch was scored blindly and consisted of 10 +/+ and 10 -/- embryos each. Bars represent mean ± SEM.

3.2.3.7 Most mature blood cells are unaffected in *rad51*$^{-/-}$ embryos

In addition to the HSPCs, I characterised the mature blood cells in the embryo using ISH, all at 5 dpf. As before, embryos were stained and then blindly sorted into different staining categories, followed by photography and genotyping. Erythrocytes were assessed using the *embryonic α1-globin* gene (*ea1*), lymphocytes using the recombination-activating gene (*rag1)* and myeloid cells using myeloperoxidase (*mpx*) expression (Figure 3.19). These ISH results indicated a drastic decrease in lymphocytes in mutant embryos, resulting in large numbers of mutants in the medium and low staining category (Figure 19A). In contrast, there was very little, if any differences between WT and mutant embryos (Figure 3.19B) for erythrocytes and a slight increase, if anything, for myeloid cells (Figure 3.19C). I verified the validity of the *mpx* ISH results by carrying Sudan black staining, which very specifically labels neutrophilic granules[406], on 5 dpf embryos (Figure 3.19D). Embryos were stained, genotyped and photographed. Ten embryos of each genotype were blindly scored; meaning the number of neutrophils was counted in each tail (Figure 3.19D). This did not indicate a statistically significant difference between WT and mutant fish either, consistent with ISH results. This confirmed the validity of the ISH approach for assessing the number of blood cells during embryonic definitive haematopoiesis. Taken together, these results mean that while HSPC numbers are affected during embryogenesis, most mature blood cells can be produced at normal numbers. Lymphocytes are an exception, since they require recombination to mature (see discussion in section 3.3).

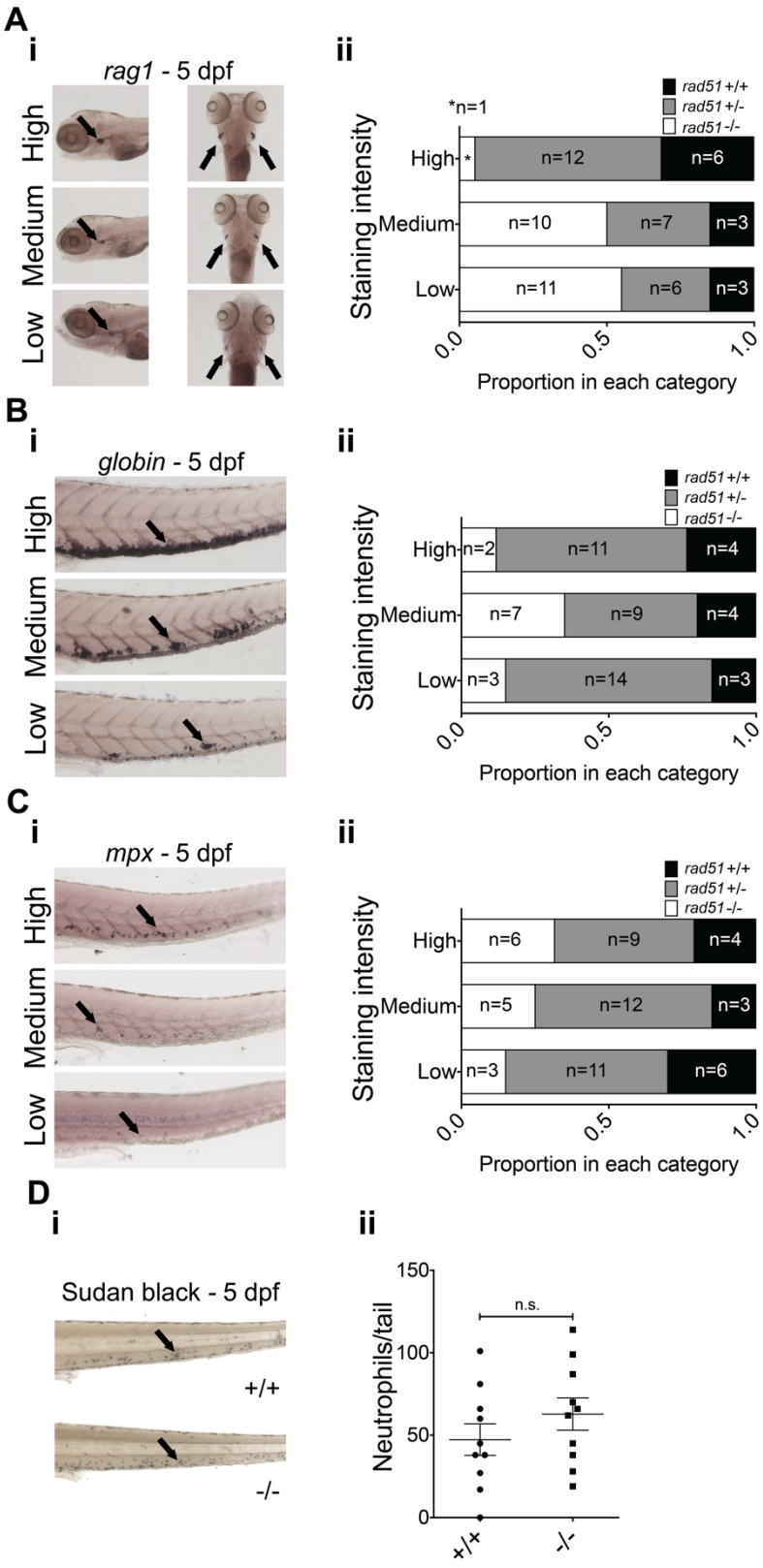


Figure 3.19: The number of lymphocytes is decreased in embryonic *rad51* mutants, whereas other lineages are unaffected. (A) Representative 5 dpf embryos stained using a *rag1* ISH probe (i) and quantification of distribution of genotypes in the different staining categories (ii). The arrow shows lymphocytes. n = 59 from one clutch. (B) Representative 5 dpf embryos stained using an *ea1 globin* ISH

probe (i) and quantification of distribution of genotypes in the different staining categories (ii). The arrow shows erythrocytes. n = 57 from one clutch. (C) Representative 5 dpf embryos stained using an *mpx* ISH probe (i) and quantification of distribution of genotypes in the different staining categories (ii). The arrow shows neutrophils. n = 60 from one clutch. (D) Representative 5 dpf embryos stained using Sudan black (i) and quantification of the number of neutrophils in WT and mutant embryos (ii). The arrow shows neutrophils. Two-tailed Student's t-test, $P = .27$, $n_{+/+} = 10$, $n_{-/-} = 10$. Bars represent mean ± SEM.

3.2.3.8 Haematopoietic stem cell transplantation experiments

Transplantation experiments are a widely-used method to determine the engraftment and self-renewal capacity of HSPCs[407]. This method is also commonly employed to assess whether HSPCs are functionally impaired[207,408,409]. Transplantation time-spans between 4 and 16 weeks are widely used to investigate the engraftment potential of progenitor cells and stem cells respectively. To determine these properties in the mutant fish, I aimed to establish a zebrafish transplantation protocol for the functional analysis of HSPCs in our laboratory. To this end, I employed the recently generated *rag2*^{E450fs} (homozygous carriers of this allele will be termed *rag2*$^{-/-}$ in this text) mutant line, which has a reduced number of functional B- and T-lymphocytes and therefore allows successful engraftment of non-immune matched cells even at sub-lethal doses of radiation[366].

As a first step to establish a transplantation protocol, I assessed the effect of varying doses of radiation (10, 20, 30 and 40 Gy) on the survival of five *rag2*$^{+/-}$ zebrafish (chosen as I did not have enough mutants) (Figure 3.20A). One fish of each group had to be culled due to side effects. In general, fish irradiated with lower doses (10 and 20 Gy) were livelier and appeared healthier. Three days post irradiation, I culled the fish and assessed cell numbers, as well as the proportion of different blood lineages in the kidney using haemocytometers and FACS (Figure 3.20B-F). From these studies, I ascertained that a dose of around 10 Gy was necessary to affect the number of progenitors (arguably the most important, as this would free niche space) and monocytes, whereas myeloid lineages were only affected from 20 Gy. Total cellularity was reduced at 10 and 20 Gy, but this effect only became statistically significant at 30 Gy. Due to concerns about side effects with higher doses of radiation and as progenitors were already affected at 10 Gy, I limited the radiation dose to 15 Gy for my next trial experiment.

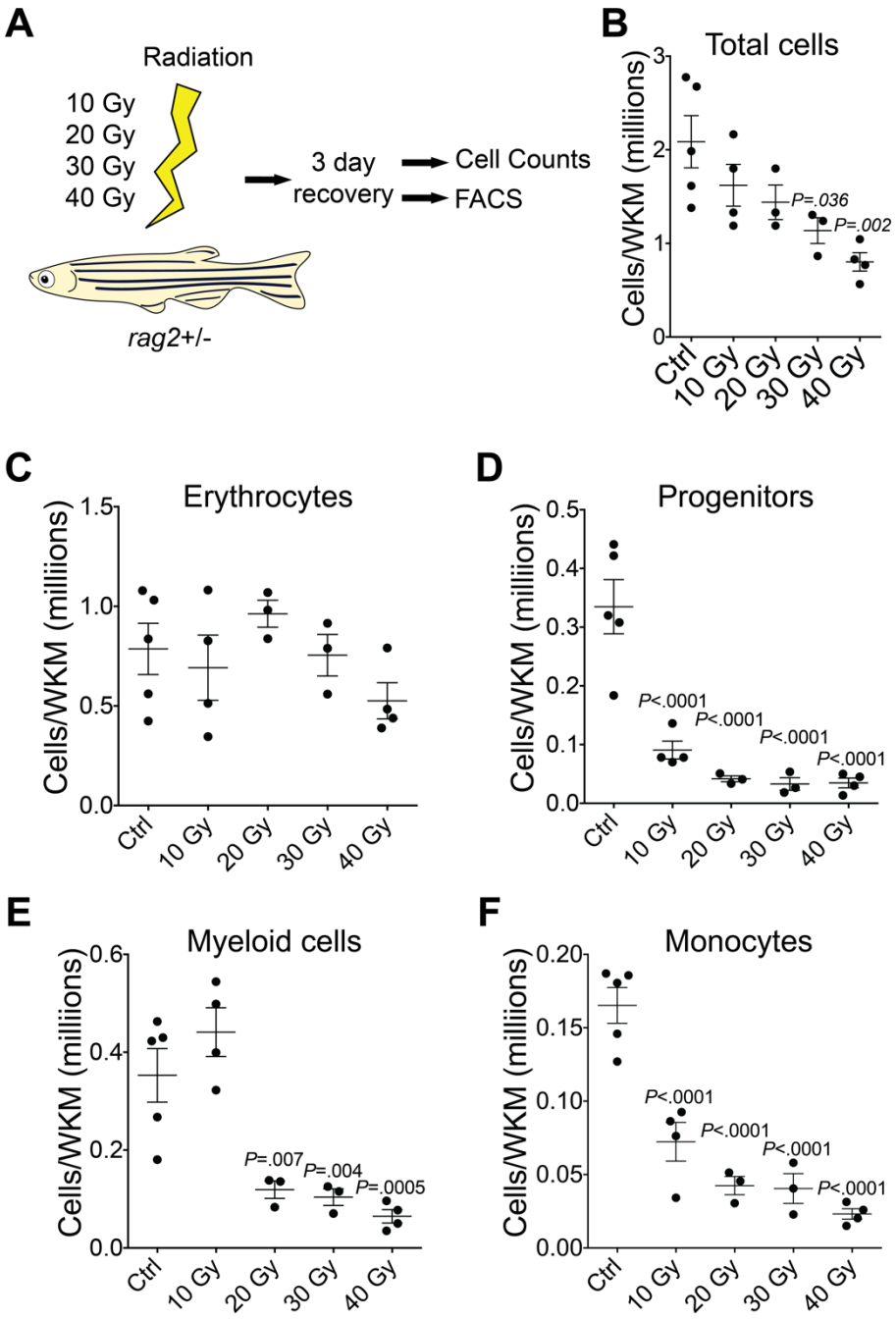


Figure 3.20: Establishing an irradiation dose for transplantation. (A) Schematic of the experiment. Fish heterozygous for *rag2*^{e450fs} were irradiated with different doses of gamma radiation. After a three-day recovery, the number of cells in the WKM was counted and the different blood lineages assessed using FACS. (B) Total number of cells in the WKM. $P = .007$. (C) Number of erythrocytes in the WKM. P = n.s. (D) Number of progenitors in the WKM. $P < .0001$. (E) Number of myeloid cells in the WKM. $P < .0001$. (F) Number of monocytes in the WKM. $P < .0001$. All data was analysed using one-way ANOVA and the resulting *P*-values are shown in the legend. *P*-values shown on the graphs stem from *post-hoc* Bonferroni multiple comparisons tests between the individual groups. For all graphs, $n_{Ctrl} = 5$, $n_{10\ Gy} = 4$, $n_{20\ Gy} = 3$, $n_{30\ Gy} = 3$, $n_{40\ Gy} = 4$. Bars represent mean ± SEM in all graphs. Ctrl = non-irradiated control fish.

Next, I wanted to introduce HSPC transplantation to the experiment and test the efficacy of the 15 Gy dose. I utilised the *Tg(itga2b:EGFP)* line, which has previously been shown to contain transplantable stem cells[339]. HSCs are enriched in the GFP^{low} population[375]. I planned to irradiate 9 mpf *rag2*$^{-/-}$ fish and transplant 1000 GFP^{low} cells intraperitoneally. In total, I irradiated ten fish at 15 Gy. One fish had to be culled three days post irradiation due to loss of balance and inability to swim. At four days post irradiation, another two fish had to be culled due to the same symptoms. The surviving fish were divided into two groups of four fish each. Both groups received 500 *Tg(sdf1a:dsRed)*-labelled stromal cells to improve engraftment of the HSCs. The groups differed in the number of GFP^{low} cells that were administered at the same time as the stromal cells. One group received 100 GFP^{low} cells each, whereas each fish in the other group received 1000.The stromal cells were omitted in later experiments due to issues with transgene silencing in the kidney marrow. At one day post transplantation (dpt), one of the low-dose recipients was found dead. At 4 dpt, a fish in the high dose group was found dead. At 7 dpt yet another fish of the low-dose group died, followed by another fish in the high dose group at 8 dpt. At 11 dpt another fish of the low-dose group had to be culled due to problems with maintaining balance while swimming. Finally, at 14 dpt another fish high dose group died. This led me to stop the experiment and cull the remaining two fish (one from the low and one from the high dose group) for analysis using FACS. Neither of the fish showed any engrafted cells. As I later learned via personal communication from David Langenau, whose laboratory created this line, *rag2* mutant fish tend to become very ill between 6-8 mpf, explaining the phenotype of the mutant fish, as they were 9 mpf at the start of the experiment. As a precaution, I only used 3-4 mpf fish in subsequent experiments and limited the radiation dose to 10 Gy to prevent any further deaths. I reasoned that ablating haematopoietic progenitors (as is possible with a 10 Gy dose (Figure 3.20D)) would be sufficient for transplantation in the absence of T-lymphocytes.

I then carried out another transplantation trial. I started by irradiating nine *rag2*$^{-/-}$ fish with 10 Gy (Figure 3.21A). One had to be culled shortly afterwards due to side effects from the radiation. The remaining eight fish were injected with 1000 GFP^{low} cells two days later to allow the fish some time for recovery. At 4 weeks post transplantation (wpt),

I assessed engraftment of the cells using FACS analysis of the PB and WKM. Fish were scored as engrafted if they showed clearly fluorescent cells in either the PB or WKM (Figure 3.21B) Of all injected fish, six out of eight retained fluorescent cells, corresponding to a 75% engraftment rate, although the number of engrafted cells varied greatly between the fish, ranging from two to several dozens of cells.

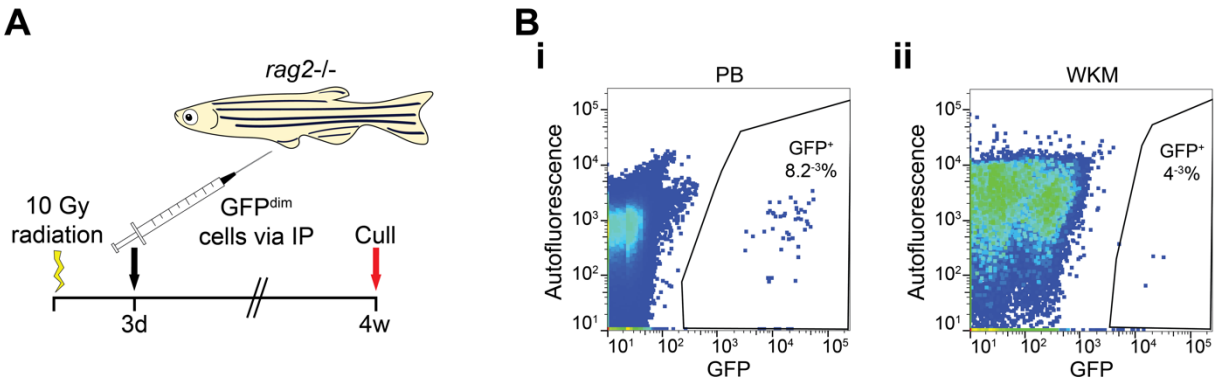


Figure 3.21: Initial trial transplantation trial. (A) Transplantation scheme. Fish mutant for *rag2* were irradiated with 10 Gy. After a three-day recovery, they were transplanted with 1000 GFP^{low} cells stemming from 4 mpf WT *Tg(itga2b:EGFP)* fish. Four weeks later, fish were culled and assessed for engraftment in the WKM and PB. (B) Typical FACS profile of the WKM (i) and PB (b) of a transplanted fish, showing the presence of GFP^{+} cells indicating engraftment. n = 8.

After completing these preliminary trials, I sought to use this technique to characterise the *rad51* mutant line. To do this, I irradiated ten mutant and WT fish each at 10 Gy, followed by transplantation of 1000 GFP^{low} cells per fish. After 16 weeks, I culled the recipients and assessed the WKM and PB for engraftment (Figure 3.22A). However, none of the fish retained any GFP^{+} cells (Figure 3.22B), suggesting that the transplanted cells did not contain any long-term engraftment potential, although the low number of fish means this can only be considered a very preliminary result. This lack of engraftment made it impossible for me to discern any differences between WT and mutant HSPCs in terms of their engraftment potential. Due to financial and time constraints, I did not pursue this avenue of research further.

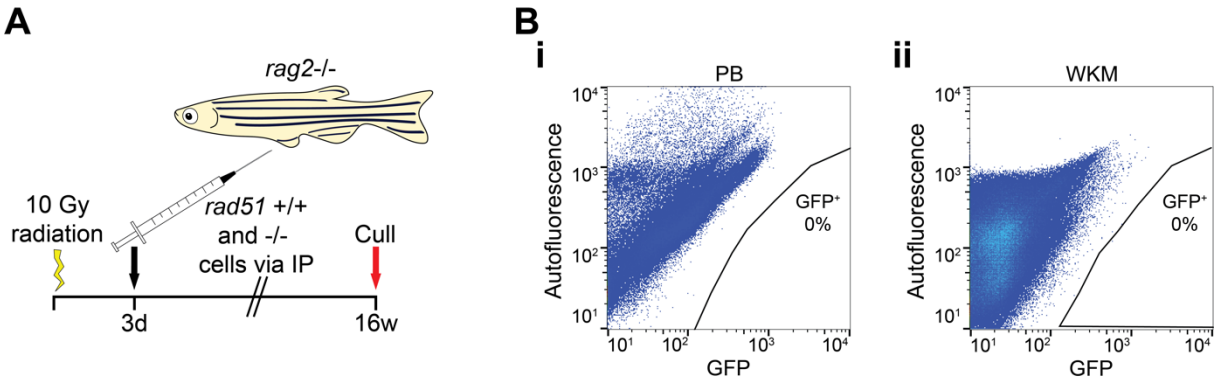


Figure 3.22: The GFP^{low} population of *Tg(itga2b:EGFP)* *WKM* lacks long-term engraftment potential. (A) Transplantation scheme. Fish mutant for *rag2* were irradiated with 10 Gy. After three days of recovery, they were transplanted with 1000 GFP^{low} cells stemming from 4 mpf *rad51* WT or mutant fish. 16 weeks later, fish were culled and assessed for engraftment in the WKM and PB. (B) Typical FACS profile of the WKM (i) and PB (ii) of a transplanted fish, showing the lack of GFP^{+} cells indicating engraftment. $n_{+/+} = 5$, $n_{-/-} = 5$.

3.2.3.9 RNA-Seq reveals a differential p53 response in different cell types

To complement my data, I carried out bulk RNA-Seq on the different cell types of the WKM. I used the same sorting approach as in one of our other publications[377] (see section 2.23 for a detailed description of the gating). By using that method, one can separate the major mature blood lineages, including HSPCs (progenitors). The basic principle of this technique is that the different blood cell types in the zebrafish kidney differ by their FSC and SSC parameters, allowing a rough determination of cell type. I was able to improve upon this method by verifying the identity of the different cell types using single-cell RNA-Seq, GO enrichment and cytospin analysis, as described previously[377], reassigning some of the gates. Lymphocytes and thrombocytes cannot be sorted based on the FSC and SSC parameters due to their low abundance in the WKM, so I focussed on erythrocytes, HSPCs, myeloid cells (mainly neutrophils) and monocytes. The erythroid and myeloid gate were almost exclusively made up by these cell types as shown by cytospins, but the monocyte and progenitor gates displayed considerably more heterogeneity. Additionally, I enriched for thrombocytic progenitors by using the *Tg(itga2b:EGFP)* line, sorting GFP^{low} cells. In total, I sorted four wells with 50 cells each per gate, for two WT and mutant fish. Following sorting, samples were prepared using the SmartSeq2 protocol[378,379] by L.F. and sequenced.

Sequencing data was processed and by E.A. Due to the size of the tables, the results are presented in the appendix. Following read-alignment and quality control, the reads of the different genotypes were pooled to compare WT and mutant gene expression, which was split up into up- and downregulated genes (Tables A1-10). I then used EnrichR[383] to search the KEGG pathway database[382] using differentially expressed genes with a non-adjusted $P < .001$ to discover pathways that are differentially expressed between WT and mutant cells. Up- and downregulated genes were searched separately. An adjusted *P*-value of .05 was used as cut-off for significance in the gene ontology (GO) term analysis.

Erythrocytes and *Tg(itga2b:EGFP)* GFP^{low} thrombocytic progenitors showed no significant enrichment among up- and downregulated genes (Tables A11-14).

For monocytes, an upregulation of ribosomal pathways (associated with proliferation) and p53 signalling were found (Table A15). For the downregulated genes, only non-specific terms related to secretion were found (Table A16).

For neutrophils, no GO terms were significant among the upregulated genes (Table A17). However, for the downregulated genes, oxidative phosphorylation, lysosomal genes and antigen processing/presentation were significant (Table A18).

Finally, I considered the gate containing mostly progenitors/HSPCs. Among the upregulated genes, both the p53 signalling pathway, as well as ribosomal genes were enriched (Table A19). For the downregulated genes on the other hand, a lot GO terms related to secretory pathways were found to be significantly enriched (Table A20).

Overall, this analysis highlighted alterations in pathways controlling proliferation and p53 signalling in the mutants, but many of the other changes were not easily relatable to the phenotype of the fish. The implications of these findings are discussed in more detail below.

3.3 Discussion

In the initial characterisation of the *rad51* mutant zebrafish line, I demonstrated that the mutation leads to a complete loss of protein and that while mutants are able to survive to adulthood, they are all infertile males. In terms of their FA-related phenotypes, I showed that mutant fish develop both haematological and non-haematological features resembling FA.

To characterise the response to DNA damage, *rad51* mutant embryos were exposed to gamma radiation, which led to the development of small eyes and heads, exacerbating the micropthalmia seen without the use of DNA damaging agents. A similar phenotype was seen in zebrafish embryos where *fancd2* was knocked down using morpholinos[361]. It is also reflective of the micropthalmia/microcephaly seen in some FA patients[19] and some FA mouse models[391,410], demonstrating the sensitivity of the rapidly dividing central nervous system to DNA damage. To counteract this, zebrafish embryos have a high expression of FA genes in the developing head and eyes, especially between the 15 somite and early long pec stages of development[359].

I also observed FA-typical chromosomal abnormalities in response to crosslinking agents commonly used in the diagnosis of FA (DEB and MMC). Similar results have been obtained for *fancd2* morphants[361], *brca2* mutants[363,364], as well as *fancl* mutants[362].

In addition, inhibition of topoisomerase I by CPT also led to increased DNA damage, consistent with defective HR repair.

In contrast, my results show that loss of *rad51* does not lead to higher sensitivity to PARP inhibitors. Previous research on mammalian cells showed that *BRCA1/2* mutant cells are more sensitive to PARP inhibition. One model proposes that this is due to the importance of PARP1 in reactivating stalled replication forks in HR-deficient cells[411]. That model is not compatible with my data, suggesting that PARP-BRCA synthetic lethality stems from functions unrelated to HR. *PARP1* is highly conserved between zebrafish and humans (~70%). A small molecule inhibitor such as DiQ should therefore be able to enter the zebrafish embryos and inhibit the enzyme the same way as in human cells. Further evidence for this is provided by the lethality of high doses of this drug in

WT and *rad51* mutant fish. In addition, the drug has previously been successfully used on adult zebrafish[412]. However, I was unable to obtain *brca2* mutant zebrafish to use as a positive control for DNA damage in response to PARP inhibition. This should be taken into account when interpreting my findings.

The survival of the *rad51* mutants to adulthood came as a surprise, since murine *Rad51* mutants die during early embryonic development[296,297]. This early embryonic death shows that Rad51 has vital functions during early mammalian development. However, murine and zebrafish development have considerable differences, the most obvious of which is the external development of zebrafish compared to the uterine development of mice. Another big difference is the timing at which zygotic genes dominate development compared to maternal genes. In mice, zygotic genes are switched on around one to two divisions, whereas zebrafish embryos undergo about ten rounds of divisions before zygotic transcripts are predominant[405]. This corresponds to the mid-blastula transition at about 3.7 hpf [404]. The immunostaining results showing that maternal Rad51 protein is present in all embryos regardless of mutation status mean that it is likely that maternal RNA rescue allows the survival of *rad51*$^{-/-}$ embryos. There is evidence that *fancd2* mRNA is also maternally provided[361], strengthening the hypothesis further. In agreement with this data, immature oocytes of adult zebrafish have been shown to have a high expression of FA genes[359]. This contrasts with human data, where FANCD2 protein was not found in oocytes[413]. If the same is the case for RAD51, this would explain the species differences in survival. A similar rescue has been proposed for *brca2* mutant zebrafish[363]. However, as it is impossible to generate fertile mutant females, I ultimately could not do an experiment that irrefutably proves maternal RNA/protein is what allows the survival.

I then proceeded with the characterisation of the adult mutants. My initial finding of sub-Mendelian birth ratios, which could be raised back to normal by outcrossing, indicates that my fish carried other mutations, which were recessive-lethal in combination with *rad51* mutations. I showed that surviving mutants have a decreased body size in comparison to their WT siblings, similar to the short stature seen in many FA patients[19] and some mouse models[253,389–391]. Together with the decreased eye size in

embryos and infertility, the *rad51* mutant line recapitulated key congenital features of FA.

Another intriguing finding of my research was that all mutants reared to adulthood were infertile males. I was able to conclude that this was due to sex reversal (i.e. mutants that should have developed as females became males), rather than female mutants dying, as the number of mutant fish was in accordance with Mendel's rules. This phenomenon has previously been described for *brca2*$^{-/-}$ fish[363,364]. Interestingly fish carrying homozygous mutations in the FA core complex gene *fancl* are also all males, but fertile[362]. Together, this evidence indicates that FA genes are necessary for the development of female zebrafish. HR genes are additionally necessary for successful meiosis in male zebrafish, besides being vital for the development of females. The requirement of DNA repair genes for sex determination is due to the complicated sex determination system in zebrafish, which consists of a mixture of environmental determination coupled to a polygenic sex determination system. During the larval period, zebrafish develop a juvenile ovary. If the fish is to become a male, the oocytes are eliminated by apoptosis[414]. Fish lacking FA genes have an increased rate of p53-driven apoptosis of the oocytes caused by unrepaired DNA damage, as demonstrated in the *fancl* model[362], leading to the development of all male mutants. The same is true for *brca2* mutant fish[363,364] and presumably for *rad51* as well, considering the role of Brca2 in loading Rad51 onto the DNA. It is possible that the common infertility in FA patients might stem from similar meiotic defects in part of the patient population.

I then utilised qPCR to investigate the expressions of genes linked to the DNA damage response. Intriguingly, *gadd45ab*, a marker for double-stranded breaks[415], was upregulated exclusively in fin tissue. This could stem from higher exposure of the skin to damaging agents due to its role as a barrier. It could also be that the DNA damage response pathways are slightly different in the various tissues, leading to failure of detection of DNA damage using *gadd45ab* as a marker gene.

In my characterisation of the haematological phenotype of the adult *rad51* mutants, I found that they exhibit decreased numbers of blood cells in the WKM and gradually develop macrocytic erythrocytes in the PB. I then showed that this decrease in WKM

cellularity was not caused by increased apoptosis or decreased proliferation of the adult cells. Paradoxically, mutant fish even displayed increased proliferation in the WKM. As this increased rate did not lead to increased cell numbers in the WKM, the newly made cells must be released into the PB. This suggests that the increased proliferation is a way to compensate for the WKM hypocellularity – a decreased number of haematopoietic progenitors proliferating at the normal rate would presumably not be enough to maintain homeostasis. In agreement with this, the fish were not anaemic. This fits with research done on *Fancd2*$^{-/-}$ mice, which showed increased cycling of HSPCs[62]. This rapid division might also be the cause of the macrocytic erythrocytes, as macrocytosis can be an early sign for the impairment of the haematopoietic system and is commonly seen before BMF in FA patients[416]. As I only observed my fish until 13 mpf, I do not know whether the fish ever develop anaemia in the PB or BMF at a later age. This would be an interesting area for further study. However, as most mouse FA models do not develop BMF during their natural life span[113,358,387], it is possible that my fish would also never progress to that stage. It is important to note that the two known patients carrying *RAD51* mutations, have not progressed to BMF so far[114,115] and that rare patients carrying mutations in other FA genes can live surprisingly long without ever experiencing blood-related pathologies[19,50]. Despite not showing overall anaemia, I saw an exacerbated shift to myeloid cells in the WKM, which is a common sign of ageing[417]. This fits with the view of FA as a segmental progeroid disorder, affecting only the blood[22,127]. Overall, the combination of key congenital and haematological symptoms in out model strengthens addition of *RAD51* to the Fanconi genes as *FANCR*.

A similar decrease in the number of blood-forming cells was seen in mice carrying homozygous mutations for the non-homologous end joining protein *Lig4*. These mice displayed decreased BM cell numbers and just like my fish, they had an approximately twofold increase in proliferation of the LT-HSCs[144]. This shows that both major pathways for the repair of double stranded breaks are important in maintaining HSCs, underlining the necessity of DNA repair in preserving the number of HSCs. It also suggests that hyperproliferation of HSPCs could be a common way to account for hypocellularity in the bone marrow.

Since the decrease in WKM cellularity could not be explained by defects seen in the adult fish, I turned to embryonic haematopoiesis to explain the changes seen later in life. My results showed a decrease in the number of HSPCs in embryos, caused by reduced proliferation with concurrently increased apoptosis. It was previously proposed by other groups that the number of HSCs in FA patients is already decreased at birth due to embryonic impairments[133,136,137]. However, my results are the first to demonstrate in an *in-vivo* model that the HSCs of FA patients are already decreased during embryonic development. In contrast to the work by Jean Soulier's group[133], who proposed decreased proliferation of HSCs as the main cause of cell number decreases in FA, I observed both decreased proliferation, as well as increased apoptosis.

When examining the mature blood cells during embryonic development, I found that erythroid and myeloid cells were unaffected, whereas *rag1*$^{+}$ T-lymphocytes in the thymus were drastically reduced. Lymphocytes undergo V(D)J recombination to generate a wide variety of receptors/antibodies during maturation. After V(D)J rearrangement, somatic diversification makes the receptors/antibodies even more diverse. One mechanism to do that is gene conversion. Data from rabbit cells suggests that Rad51 is a vital component of that process[418,419]. Other publications have implicated Rad51 in B-cell class switching, rather than somatic hypermutation[420]. Interestingly, my data suggests that these processes seem to be necessary for the generation of sufficient numbers of lymphocytes, rather than an additional step to generate more cellular diversity. It would be interesting to investigate whether cells that cannot undergo somatic diversification are dysfunctional, or whether there is another reason for the decrease in lymphocytes. The simultaneous decrease in HSPCs, but lack of changes in mature cells (bar the special case lymphocytes), indicates that there is no direct correlation between the number of stem and mature cells during embryonic development.

It is possible that this is also true during adulthood, indicating that there is no relationship between the number of HSCs and the number of mature blood cells in the PB. This is in agreement with studies showing that the number of functional HSCs declines with ageing[421,422], even though the number of phenotypic HSCs increases[423–428], because otherwise all elderly people should suffer from various cytopenias, which is not

the case. These observations also fit with a previous study that proposed that HSCs plays a minor role in the maintenance of normal (i.e. when not stressed) haematopoiesis[429].

Next, I wanted to determine whether the HSPC defects were cell-intrinsic, i.e. caused only by the cells themselves, or extrinsic, i.e. determined the environment of the HSPCs in their niche. To do this, I carried out transplantation experiments, using WT and mutant cells from *Tg(itga2b:EGFP)* carrying donors. However, my results showed no engraftment for WT fish at 16 wpt, making the comparison of WT and mutant engraftment impossible. I believe that this failure to engraft was due to a combination of the low amount of irradiation used in my trial and the low proportion of long-term repopulating cells in the GFP^{low} subpopulation. In contrast, Ma *et al.*[339], who used lethal irradiation, reported good engraftment at 16 wpt. When they used 1000 GFP^{low} cells, with carrier marrow (5-8 X 10^5 cells) they achieved 19% engraftment at 16 wpt (when dead fish are included in the analysis). Considering the low number of fish that were transplanted, it is conceivable that I just did not see any engraftment by chance. An interesting avenue for future research would be to repeat this experiment using a higher dose of radiation and more GFP^{low} cells. I have since refined the transplantation technique utilising the *Tg(runx1+23:mCherry)* transgenic line[338], routinely resulting in >80% engraftment at 16 wpt[377], indicating a higher proportion of HSCs in this transgenic line. A lack of time and financial resources prevented me from using this improved technique for the study of the *rad51* mutant fish.

Finally, I considered transcriptional changes in the different blood lineages using bulk RNA-Seq. I was able to investigate erythrocytes, myeloid cells, monocytes, HSPCs and GFP^{low} thrombocytic progenitors, but lymphocytes could not be considered due to their rarity in the kidney. My analysis revealed surprisingly few changes in the different blood lineages of mutant fish. The experimental design pooling only 50 cells per well might have caused this by increasing the variability within each biological replicate. This factor possibly contributed to the low number of differentially expressed genes. For future experiments, a higher number of cells per well would be preferable. However, even though the experimental design increased the ambiguity of my data, I was still able to detect differential responses to Rad51 loss in the different cell types. Of note is the

enrichment of antigen processing genes and oxidative phosphorylation among downregulated genes in neutrophils, which may suggest a less mature state of the cells. Most interestingly, both ribosomal genes, as well as the p53 signalling pathway were enriched among upregulated genes in HSPCs. The increase in ribosomal genes is interesting, as a previous publication of my research group identified high expression of ribosomal constituents to be associated with more immature cells[377], indicating that HSPCs in *rad51* mutant fish contain proportionally more immature cells than the WT WKM, indicating differentiation defects. In addition, the upregulation of p53 target genes indicates that the p53 pathway is being activated, fitting with increased haematopoietic stress. The role of p53-driven signalling in FA pathogenesis will be discussed in detail in the next chapter.

In summary, *rad51* mutant fish developed both congenital, as well as haematological symptoms resembling key features of FA. It can also be concluded that the assignment of *RAD51* as *FANCR* is appropriate. As *brca2* and *fancl* knockout fish have not been characterised in terms of haematological abnormalities[362–364], the *rad51* mutant line can therefore be considered the first characterised adult zebrafish FA model. After having established the appropriateness of my model, I went on to utilise it to understand the molecular pathology underlying FA.

4 BMF in FA: molecular signalling

4.1 Introduction

4.1.1 The p53 protein and pathway

The p53 protein is possibly the most famous protein in cancer biology, if not biology in general. Its moniker "p53" stems from the fact that it appears to have a mass of 53 kDa when run on an SDS-PAGE gel. In fact, its real molecular weight is only about 43.7 kDa (at least for the full-length form). This discrepancy is due to proline-rich regions in the protein, which impede migration through the gel[430]. This vital protein is encoded by the *TP53* gene in humans, which has orthologues in almost all vertebrate species[431]. Here, I will be using p53 and TP53 interchangeably.

The original discovery of p53 happened in the late 1970s during the study of cells transformed by Simian vacuolating virus 40 (SV40). In these cells, p53 was identified as the binding partner of the SV40 T antigen[432]. Subsequent experiments showed it to be highly expressed in cancer cells, whereas comparatively small amounts were found in non-transformed cells[433,434]. Due to this high expression in transformed cells and somewhat common dominant-negative mutations, initial reports erroneously considered it an oncogene rather than a tumour suppressor[430]. Following the cloning of the *TP53* gene in mice and humans for the first time in the early 1980s[435,436], evidence for the protein's role as a tumour suppressor started accumulating in the mid-1980s. At that point, it was shown that *P53* was commonly mutated in cancer cell lines and cancer patients, a hallmark of a tumour suppressor gene[437–441]. Further evidence came from overexpression experiments, which showed that *P53* overexpression can suppress transformation[442]. Finally, the discovery that germline *P53* loss-of-function mutations lead to the Li-Fraumeni cancer predisposition syndrome conclusively showed that p53 acts as a tumour suppressor rather than as an oncogene[443,444]. It is indeed one of the most commonly mutated genes in cancer, with a mutation rate ranging between 10 and 100% depending on the type of cancer[445]. Compared to solid tumours, *P53* mutations are relatively rare in haematological malignancies[446].

The p53 protein has various roles in the cell, which are reviewed in-depth elsewhere[447,448]. Its main function is as a transcription factor, controlling genes such as the proliferation suppressor *P21* and pro-apoptotic proteins such as *BAX*[449,450]. Many of its other targets are also involved in processes regulating cellular growth, metabolism and apoptosis[430]. It does not only promote transcription, but conversely can also repress genes[451].

Apart from its function in the nucleus as a transcription regulator, it also carries out many different non-transcriptional roles all over the cell. For example, p53 plays a vital role in the membrane permeabilisation of mitochondria during apoptosis[166].

Both transcriptional, as well as non-transcriptional roles of p53 are involved in the response to genome stress[452,453]. Because of this induction of p53 in response to damage to the DNA, it has been termed "guardian of the genome"[454].

The signalling pathways involving p53 are very diverse and there are few cellular mechanisms that do not involve it at least indirectly. Just to give some examples, p53 is involved in the antioxidant response[455], it acts as a negative regulator of inflammation[456] and mediates the toxic effects of radiation in haematopoietic tissues[456–458].

Due to the vital role of the p53 protein in the protection of genomic integrity, there is a complicated regulatory network controlling its activity. One of the most important regulatory player on which many signals converge is MDM2[430]. This protein acts as an E3 ubiquitin ligase for p53, contributing to the fast turnover of p53. MDM2 also functions in ubiquitin-independent modes to block the interaction of p53 with the rest of the transcriptional machinery[459]. It acts is conjunction with another important p53 regulator, MDM4, which can pose as a cofactor for MDM2[460]. A combination of these and other mechanisms allow the fine-tuning of p53 activity to the appropriate level.

4.1.2 The role of p53 in HSPCs

As in other tissues, p53 plays a large role in regulating the activity of haematopoietic cells. Depending on the cellular context, p53 may regulate the exit of HSPCs from the cell cycle, the apoptotic response or DNA repair[461–467]. The regulation of these factors by

p53 plays an important role in the ageing process of the haematopoietic system[140–142,144,468].

The p53 pathway seems to be particularly important in regulating quiescence and the cell cycle of HSCs[469]. Quiescence is, among other things, vital to maintain HSC functionality and proliferation of HSCs gradually leads to functional impairment during ageing[134,421,470,471]. As p53 suppresses cell cycle entry, it is not surprising that *Tp53* knockout mice have a two- to threefold larger HSC pool. This was linked to decreased cycling time and outcompetition of wild type cells *in-vitro*. However, these HSCs show a reduced repopulation ability, indicating functional impairment[463,472–474]. Similar results were obtained using pharmacological inhibition of p53 activity. These interventions lead to increased cell cycling and a decreased recovery time following BMT[475,476].

Another important function of p53 is to regulate the decision of whether a cell undergoes apoptosis. While apoptosis is usually beneficial to the organism (such as killing cells with DNA damage to prevent tumours), there are situations where it can be detrimental to the haematopoietic system. Aberrant p53 signalling may lead to genome instability and BMF[140,477]. For example, the toxicity of ionising radiation to the bone marrow is mediated via p53, whereas in other tissues such as the small intestine, p53 signalling post irradiation can serve as a survival signal[456–458,478]. This is why the bone marrow is extremely sensitive to ionising radiation. This sensitivity can also be used medically, as it allows the selective ablation of the bone marrow prior to HSC transplantation.

A further example of the sensitivity of the bone marrow to p53-driven apoptosis is the anaemia seen in the ribosomopathy Diamond-Blackfan anaemia, where p53-driven signalling in response to ribosomal stress leads to apoptosis of the erythroid progenitors. Similar mechanisms have been proposed for other ribosomopathies[479].

4.1.3 Zebrafish p53

The zebrafish genome contains an orthologue of the human *TP53* gene, with which it shares about 45% identity on the nucleotide level. Zebrafish *tp53* is particularly highly expressed during early development before 48 hpf, demonstrating the importance of

this gene during embryogenesis[480]. Despite this, zebrafish lacking *p53* survive to adulthood without any problems[365,481]. The role of *tp53* mutations in the adult zebrafish has also been studied. Homozygous fish carrying mutations in the DNA-binding domain (*tp53*^{M214K}) were studied in-depth for abnormalities. The only noticeable effect of *p53* mutation at the organismal level was the development of malignant peripheral nerve sheath tumours (MPNSTs) in slightly fewer than 30% of the adult fish, with the first tumours appearing about 8 mpf[365]. Mice carrying mutations in *Tp53* show similar features, albeit with different tumour types[482]. Apart from gross organismal defects, the fish showed many abnormalities in cellular signalling due to the loss of functional p53[365]. Overall, these similarities to the situation in mice and humans demonstrates the suitability of the zebrafish model to study p53 signalling.

4.1.4 The role of p53 in FA

There have been many studies looking at p53-driven signalling in FA cell lines, animal models and patient samples. Early studies were quite ambiguous, with many contradictory results. Some studies found that p53-mediated apoptosis following gamma irradiation is impaired in FA cells (i.e. cells were less likely to undergo apoptosis than normal[483,484]), whereas others found that p53-driven apoptosis is at least partly responsible for apoptosis in *Fancc* mutant mice. Similarly, some groups showed that p53 can induce the expression of FA genes such as *FANCC*[156], whereas others showed that p53 downregulates the FA pathway[139]. Apoptosis driven by p53 has also been implicated in germ cell defects in FA patients[485]. In addition, loss of p53 has been shown to accelerate tumour development in *Fancd2* and *Fancc* knockout mice[486,487], as well as *brca2* mutant zebrafish[363,364,488].

A more defined role for p53 in FA emerged in 2012, when Ceccaldi *et al.* proposed that an excessive p53/p21 response drives the bone marrow failure in FA patients[133]. First, they proved that the p53 protein is excessively induced in response to DNA damage in patient-derived FA cells. They then went on to show that this leads to p21-mediated cell cycle arrest. Engraftment defects upon transplantation of *Fancg*$^{-/-}$ cells were also rescued upon *p53* knockdown, with similar results obtained for *p53* knockdown on human FA cells. Because the defects seen in FA start early in life, they reasoned that an overactive

p53/p21 response in the foetal liver leads to a small initial HSC pool. Later in life, additional stress leads to further decreases in the stem cell pool, eventually exhausting it and causing BMF. This fits nicely with previous data from the group showing that FA patients with lower *p53* expression have less severe symptoms[489]. However, data from *p21, Fancd2* double knockout mice shows that these defects do not seem to be mediated via p21, as the phenotype of this line was exacerbated rather than alleviated as expected[138]. In addition, another paper has shown that HSC defects in the foetal liver of *Fancd2*$^{-/-}$ mice are independent of apoptosis of p53 activation, claiming that overly active MAPK (p38) is the main factor mediating blood abnormalities[137]. In conclusion, there is still no complete certainty about the role of p53-driven signalling in the pathogenesis of FA, even though there are strong indications that it is one of the main drivers of BMF.

In this part of this study, I investigated the role of p53 in the phenotype of the *rad51* mutant line. I show that haematological phenotypes were completely rescued upon *p53* co-mutation, adding further evidence that p53 plays a critical role in FA pathogenesis.

4.2 Results

4.2.1 Expression in embryos and adults

To look at the role of *p53* signalling in the abnormal phenotype of the *rad51* mutants, an *in-situ* hybridisation using a *tp53*-specific riboprobe was carried out at 2 dpf. Expression of *p53* in unchallenged *rad51* mutants was noticeably increased in the head region of the embryo (Figure 4.1A), underscoring the susceptibility of the developing nervous system.

Paradoxically, when I carried out qPCR analysis on the tissues of adult *rad51* mutants, *p53* expression was consistently decreased in the kidneys of the mutants (Figure 4.1B), suggesting that there is a functional need for the decrease in *p53* in most blood cells. Alternatively, there could be a selective advantage for cells expressing low levels of *p53*, either due to cells expressing high levels undergoing apoptosis, or because cells expressing low levels might proliferate at a higher rate.

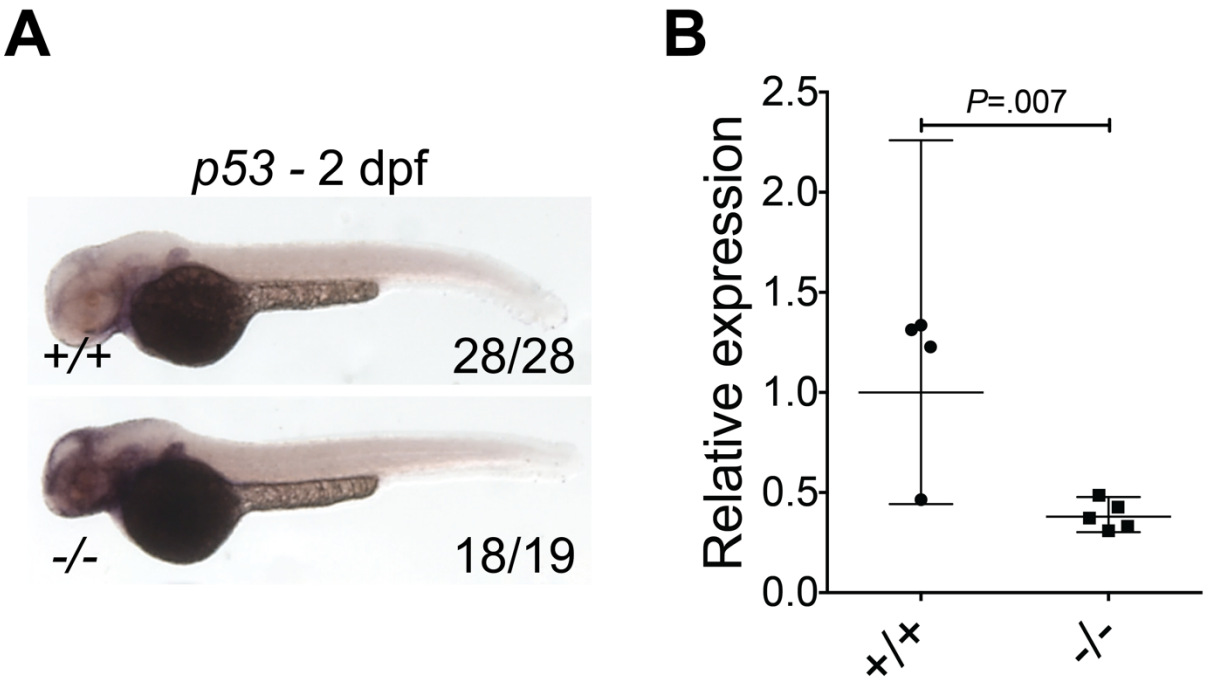


Figure 4.1: The expression of *p53* differs between embryos and adult fish. (A) ISH of 2 dpf embryos using a *p53* ISH probe. Upregulation of the transcript in the head region of the embryos is visible. $n_{+/+} = 28$, $n_{-/-} = 19$ (B) Expression of *tp53* in the adult WKM as assessed by qPCR analysis. The expression was roughly halved compared to the WT fish. Two-tailed Student's t-test, $P = .007$. $n_{+/+} = 4$, $n_{-/-} = 5$. Bars represent the geometric mean ± 95% CI to estimate fold changes. Data shown in A was obtained by E.B.-M

4.2.2 Characterisation of *rad51, p53* double mutants

I then decided to cross the *rad51* line with the *tp53*^{zdf1}/*tp53*^{M214K} line (*p53*$^{-/-}$), in which the *p53* gene is dysfunctional because of a M214K mutation[365], to generate double mutants for the study of the compound effects of combining these mutations. There were multiple compelling arguments to do this. For one, *p53* driven signalling has been implicated in the pathogenesis of FA[133], so I reasoned co-mutation might rescue some of the phenotypes I observed. Moreover, co-mutation of *brca2* and *p53* was able to successfully rescue the sex reversal seen in these fish and led to a high rate of cancer development[363,364,488]. Finally, the upregulation of *p53* in *rad51*$^{-/-}$ embryos (Figure 4.1A) led me to believe that p53 plays an important role in the phenotype of the mutant fish.

4.2.3 Co-mutation of *p53* partially rescues congenital phenotypes

In my initial characterisation of the offspring of *rad51*$^{+/-}$, *p53*$^{+/-}$ incrosses, I observed several female (four out of twelve fish) fish among the double mutants (Figure 4.2A). This was most likely caused by decreased apoptosis in the immature germ cells, allowing the development of oocytes and consequently ovaries. Zebrafish sex determination and the role of p53 in this process is described in more detail in sections 3.3 and 4.3 respectively. Fertility was not rescued, as I was unable to outcross both female and male double mutants to WT fish even after several attempted matings. The female gonads looked normal despite the infertility (Figure 4.2B), but in the males, the infertility matched with histological evidence showing lack of mature spermatozoa in the testes (Figure 4.2C).

Size measurements of my fish showed that both *rad51* single and *rad51, p53* double mutants were smaller than their wild type siblings, in contrast to *p53* single mutants, which were of normal size (Figure 4.2D). This indicates that the small size of *rad51* single mutants was unlinked to p53 signalling.

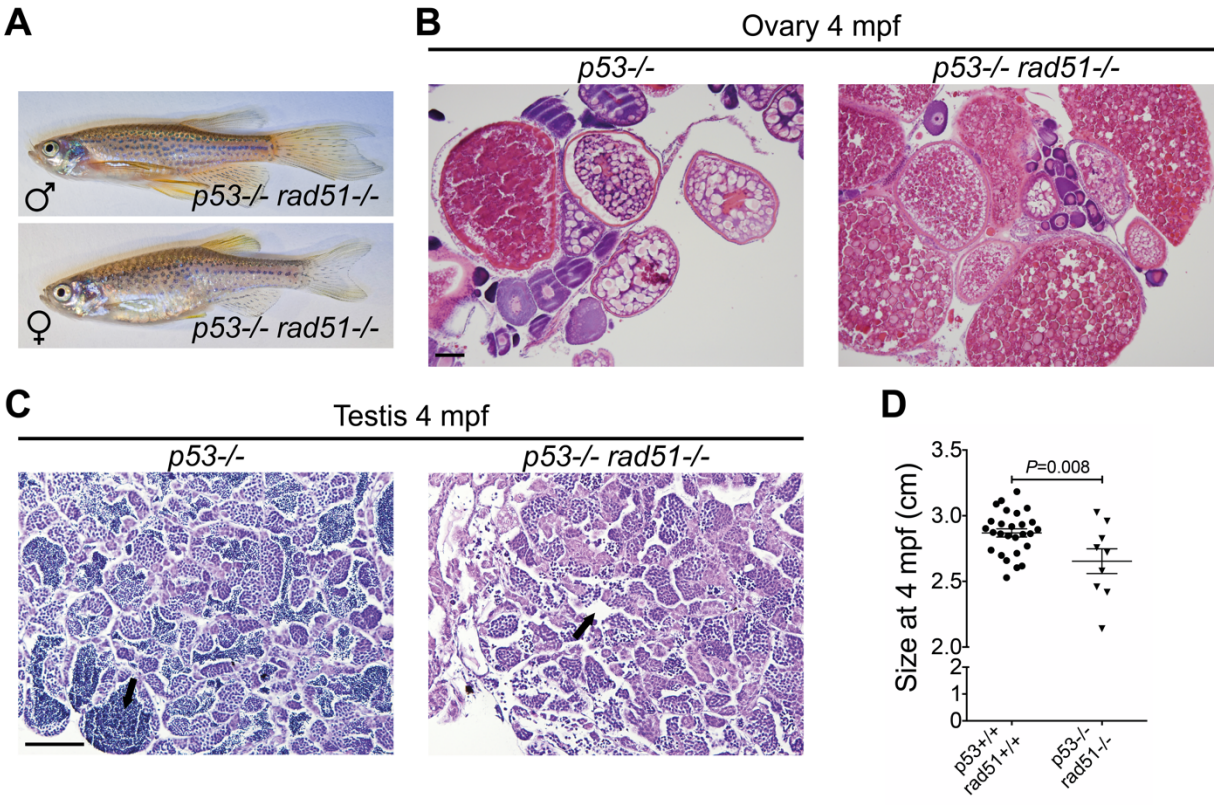


Figure 4.2: Co-mutation of *rad51* and *p53* can rescue some congenital phenotypes seen in the single mutants. (A) Male (top) and female (bottom) double mutants. (B) Histological sections of a *p53*$^{-/-}$ ovary (left) and a *p53, rad51*$^{-/-}$ ovary (right) at 4 mpf using a 10X objective. Oocytes of different maturation stages are apparent in both genotypes. Scale bar = 100 µm. (C) Histological sections of a *p53*$^{-/-}$ testis (left) and a *p53, rad51*$^{-/-}$ testis at 4 mpf using a 20X objective. (right). Arrows indicate mature spermatozoa (or lack thereof). Scale bar = 100 µm(D) Graph showing the size of WT and double mutant adults at 4 mpf, showing a small, but measurable difference. $n_{p53+/+\ rad51+/+} = 27$, $n_{p53-/-\ rad51-/-} = 9$. Two-tailed Student's t-test, $P = .008$. Bars represent mean ± SEM.

4.2.4 Double mutant fish develop tumours

Intriguingly, one third of the double mutants (three out of nine fish left to age) developed tumours. The earliest tumour appeared at 5 mpf and took the form of a tumour in the eye socket which led to the eye "bulging" out of its socket (Figure 4.3A). Two more tumours were detected upon histological examination at 8 mpf and these were located around the ovaries (Figure 4.3B). Further investigation of the histological slides revealed that the tumours resembled MPNSTs (common tumours in zebrafish in general), with very characteristic "spindle-like" cells (Figure 4.3A and B). Conversely, *p53*$^{-/-}$ fish have a higher latency of developing tumours (8 months) and only reach a

comparable tumour incidence at 16 mpf[365], indicating a compound effect of the *rad51* and *p53* mutations.

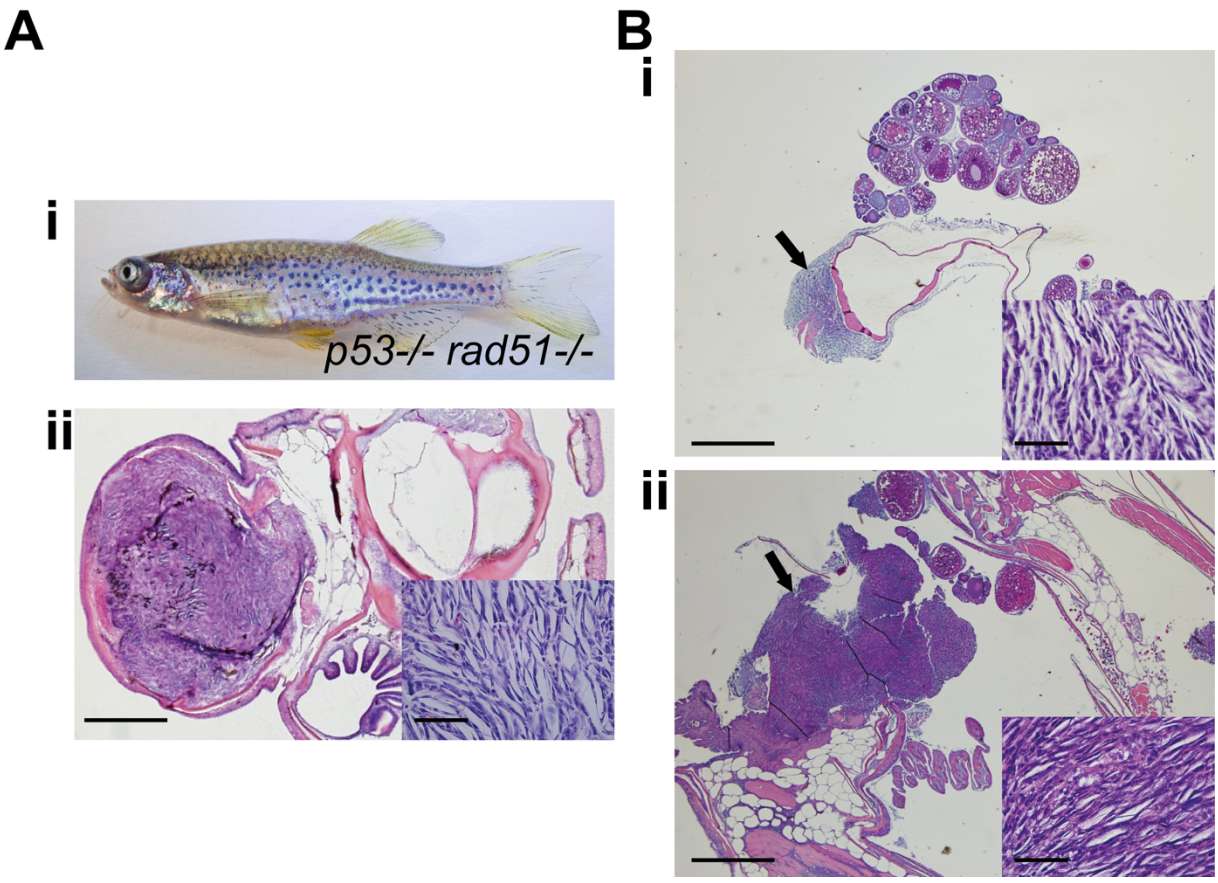


Figure 4.3: Double mutants are susceptible to tumourigenesis. (A) Tumour found directly behind the eye of a 5 mpf fish (i) and accompanying H&E stained histological sections of the tumour (ii) using 2.5 and 63X oil immersion objectives. In the insert characteristic spindle cells, consistent with a MPNST are readily apparent. (B) (i) and (ii) MPNSTs found upon histological examination two different 8 mpf female double mutants. Arrows indicate the tumour mass next to the normal ovary tissue. Overall, three out of nine aged double mutants developed tumours. Scale bar = 500 μm and 10 μm in the overall images and the inserts respectively.

4.2.5 Haematological defects in *rad51*$^{-/-}$ embryos are rescued when *p53* is co-mutated

I then went on to characterise the haematological phenotypes of the double mutants. As I saw a striking decrease in *cmyb*$^{+}$ cells in the CHT of 4 dpf *rad51* single mutants, I carried out an ISH on *p53*$^{+/-}$, *rad51*$^{+/-}$ incrosses, which was scored blindly as described above. Representative CHTs are shown in (Figure 4.4) and a quantification of staining intensity in (Table 4.1). It is apparent in these images that while *rad51*$^{-/-}$ embryos had considerably

decreased *cmyb* expression, this defect was absent in *p53*$^{-/-}$, *rad51*$^{-/-}$ embryos, which had similar staining to WT embryos.

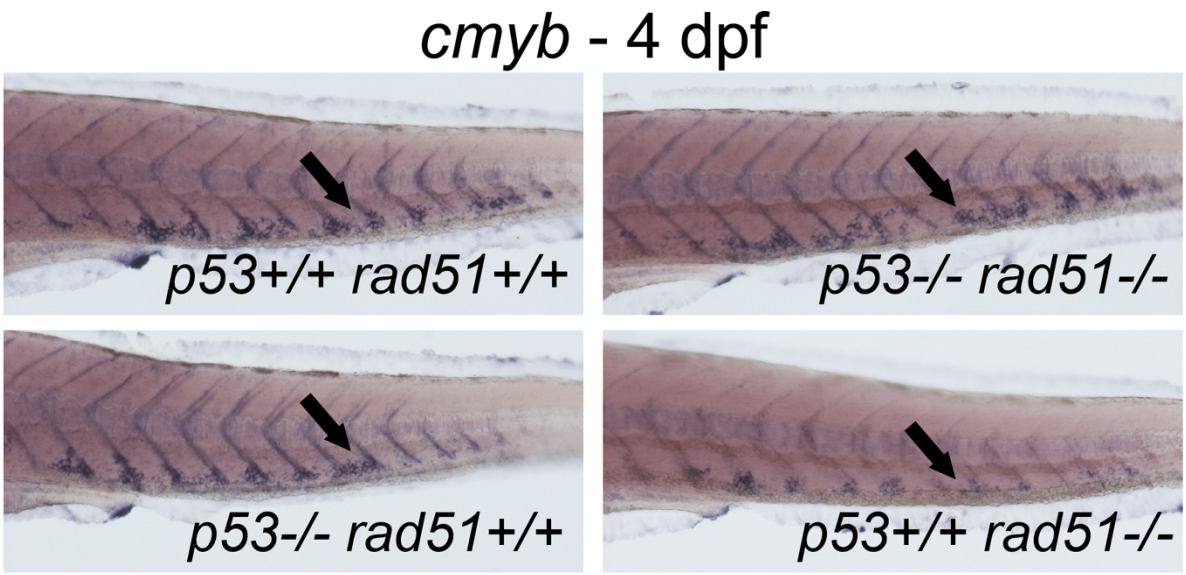


Figure 4.4: The embryonic HSPC defects in *rad51*^{sa23805} fish are rescued in a *p53* mutant background. Representative images of 4 dpf embryos resulting from incrosses of *p53*$^{+/-}$, *rad51*$^{+/-}$ parents stained using a *cmyb*-specific probe. The total number of embryos used (all genotypes) n = 237 from 4 clutches. For information about all genotypes please see Table 4.1.

Table 4.1: Quantification of the staining data presented in Figure 4.4. Three independent clutches with n = 60 were pooled to generate this table, leading to an overall n = 237 when accounting for unsuccessful genotyping. *p53*$^{-/-}$, *rad51*$^{+/+}$ fish are underrepresented in the high staining group (highlighted in orange), whereas *p53*$^{-/-}$, *rad51*$^{-/-}$ embryos (blue) stained similar to WTs.

		Staining intensity		
		High	Medium	Low
Genotype	*rad51*$^{+/+}$, *p53*$^{+/+}$	5	9	3
	rad51$^{+/+}$, *p53*$^{+/-}$	13	13	5
	rad51$^{+/+}$, *p53*$^{-/-}$	5	4	6
	rad51$^{+/-}$, *p53*$^{+/+}$	6	8	12
	rad51$^{+/-}$, *p53*$^{+/-}$	25	17	17
	rad51$^{+/-}$, *p53*$^{-/-}$	15	14	7
	rad51$^{-/-}$, *p53*$^{+/+}$	1	6	6
	rad51$^{-/-}$, *p53*$^{+/-}$	4	6	19
	rad51$^{-/-}$, *p53*$^{-/-}$	6	3	2
Unsuccessful genotyping		3		
Successful genotyping		237		

4.2.6 Co-mutation of *p53* also rescues adult WKM defects

Next, I investigated whether HSPC proliferation in the WKM of 4 mpf fish would also be rescued by *p53* co-mutation. To do this, I carried out the same BrdU incorporation assay described in section 3.2.3.5. As I did not have enough double mutants for a time course and BrdU incorporation was the highest at 1 dpi in the previous experiment, I focussed on this day for the analysis of BrdU incorporation. Strikingly, proliferation rates were reduced back to WT levels in the double mutants, rather than doubled as in the *rad51* single mutants (Figure 4.5A).

To see whether these changes would translate to a rescue of WKM cellularity, I carried out cell counts of the *rad51*, *p53* mutant fish. This experiment showed that kidney marrow cellularity reverted to WT levels upon co-mutation (Figure 4.5B). Interestingly, rather than having increased WKM cellularity, *p53* single mutants had the same cell number as WT fish (Figure 4.5B). The relative proportion of different blood cell lineages in the kidney was similarly unaffected (Figure 4.6).

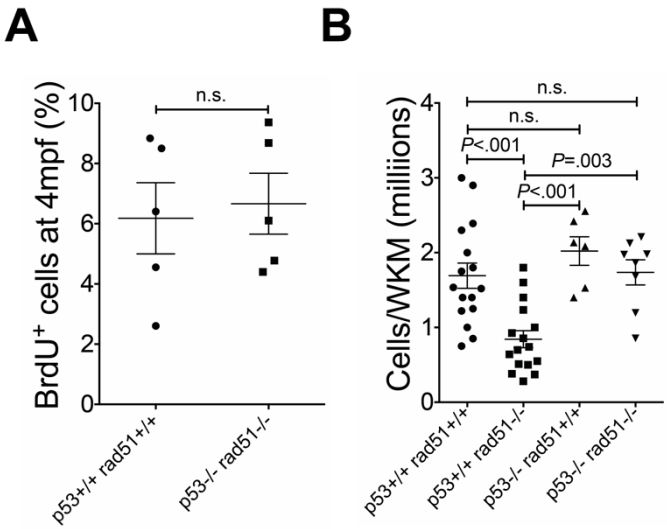


Figure 4.5: The adult WKM defects in *rad51*^{sa23805} fish are rescued in a *p53* mutant background. (A) Percentage of BrdU^{+} cells in 4 mpf kidneys at 1 dpi. Two-tailed Student's t-test, $P = .76$, $n_{p53+/+\ rad51+/+} = 5$, $n_{p53-/-\ rad51-/-} = 5$. (B) Number of total cells per kidney at 4 mpf quantified using a haemocytometer. Analysis using one-way ANOVA ($F(3, 43 = 10.45)$, $P < .0001$), individual p-values shown in the figure are from Tukey's, *post-hoc* test, $n_{p53+/+\ rad51+/+} = 16$, $n_{p53+/+\ rad51-/-} = 16$, $n_{p53-/-\ rad51+/+} = 6$, $n_{p53-/-\ rad51-/-} = 8$. Bars represent mean ± SEM in both graphs.

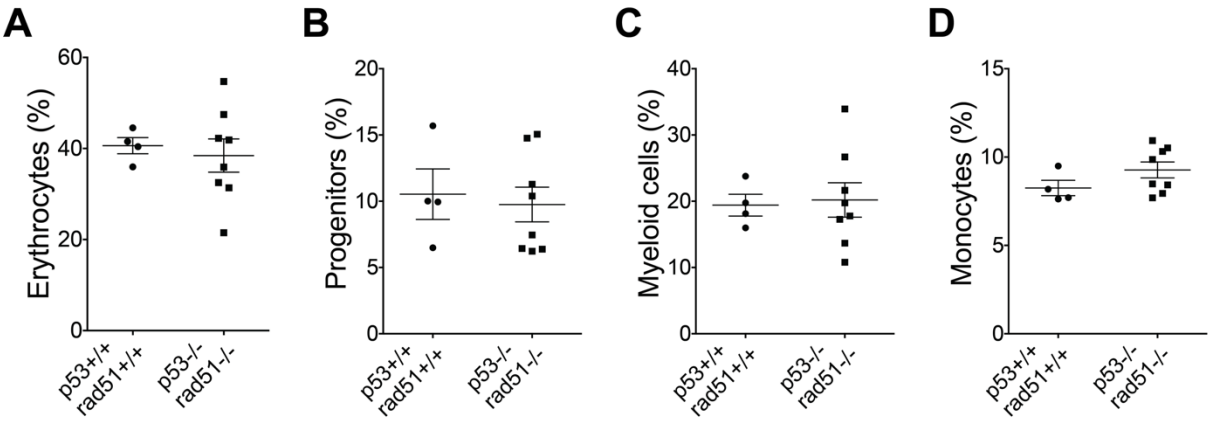


Figure 4.6: *p53* co-mutation reverts the blood lineage bias seen in single mutants. FACS results showing results WT and *p53*$^{-/-}$, *rad51*$^{-/-}$ fish at 8 mpf in terms of the proportions of different cell lineages present in the WKM. Shown are erythrocytes (A), progenitors (B), myeloid cells (C) and monocytes D. For all graphs, $n_{p53+/+\ rad51+/+}$ = 4, $n_{p53-/-\ rad51-/-}$ = 8. Bars represent mean ± SEM.

In summary, *p53* co-mutation completely rescued all haematological phenotypes investigated, suggesting that the deleterious effects of *rad51* mutation are mediated via a p53-driven mechanism. Interestingly, not all phenotypic features were rescued; neither size defects, nor infertility were fixed by co-mutating *p53*. This indicates that some FA features may be p53-driven, whereas others are not.

4.2.7 Loss of *p53* on its own can also affect haematopoiesis

I carried out a *cmyb* ISH in which *rad51*$^{+/-}$, *p53*$^{-/-}$ fish were incrossed (Figure 4.7A), to see whether *rad51* loss would still have an effect in a *p53* mutant background. The results of this experiment were inconclusive – there were slightly more *rad51* mutants in the low staining group, but the effect was not nearly as strong as in a *p53*$^{+/+}$ background, providing further evidence for the rescue of the HSPC defects. To clarify whether *p53* loss on its own can affect HSPC numbers, I carried out a *cmyb* ISH on a *p53*$^{+/-}$, *rad51*$^{+/+}$ incross (Figure 4.7B). This showed that *p53* mutation alone was able to increase the number of HSPCs in comparison to WT embryos, in contrast to the adult situation, where *p53* mutation on its own was unable to increase WKM cell numbers (Figure 4.5).

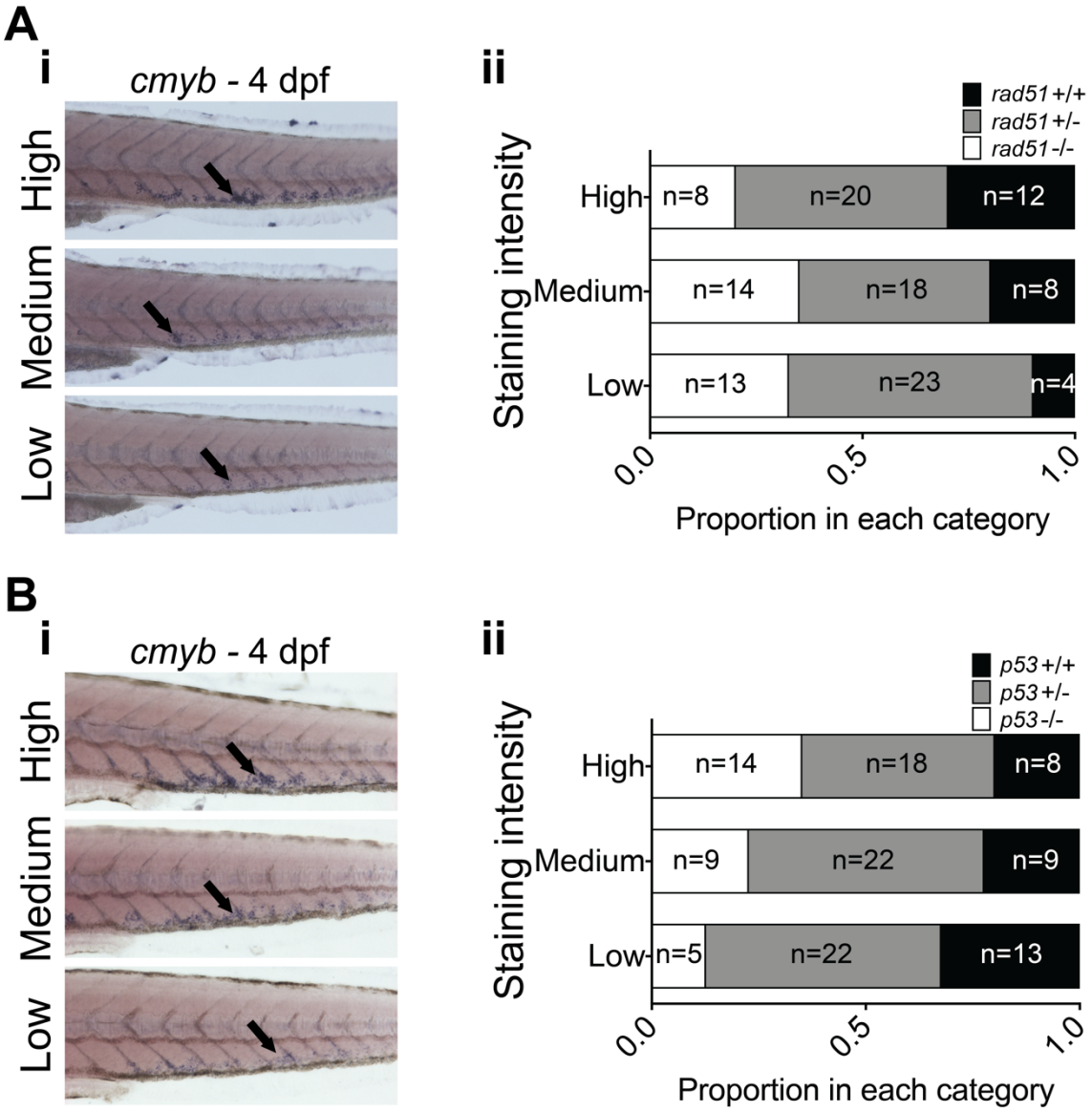


Figure 4.7: *p53* mutation can increase HSPC numbers independently from *rad51* mutation. (A) Representative images of 4 dpf embryos resulting from incrosses of $p53^{-/-}$, $rad51^{+/-}$ parents stained using a *cmyb*-specific probe (i) and quantification of the data (ii). The arrow shows HSPCs. n = 119 from two clutches. (B) Representative images of 4 dpf embryos resulting from incrosses of $p53^{+/-}$ parents stained using a *cmyb*-specific probe (i) and quantification of the data (ii). The arrow shows HSPCs. n = 120 from two clutches.

I also considered the mature blood lineages during the definitive wave of embryonic haematopoiesis. Neutrophils were increased upon *p53* mutation in 5 dpf embryos (Figure 4.8 A and B), but erythrocytes and lymphocytes were unaffected (Figure 4.8 C and D). These results imply that *p53* mutation on its own can affect the haematopoietic lineage output in zebrafish.

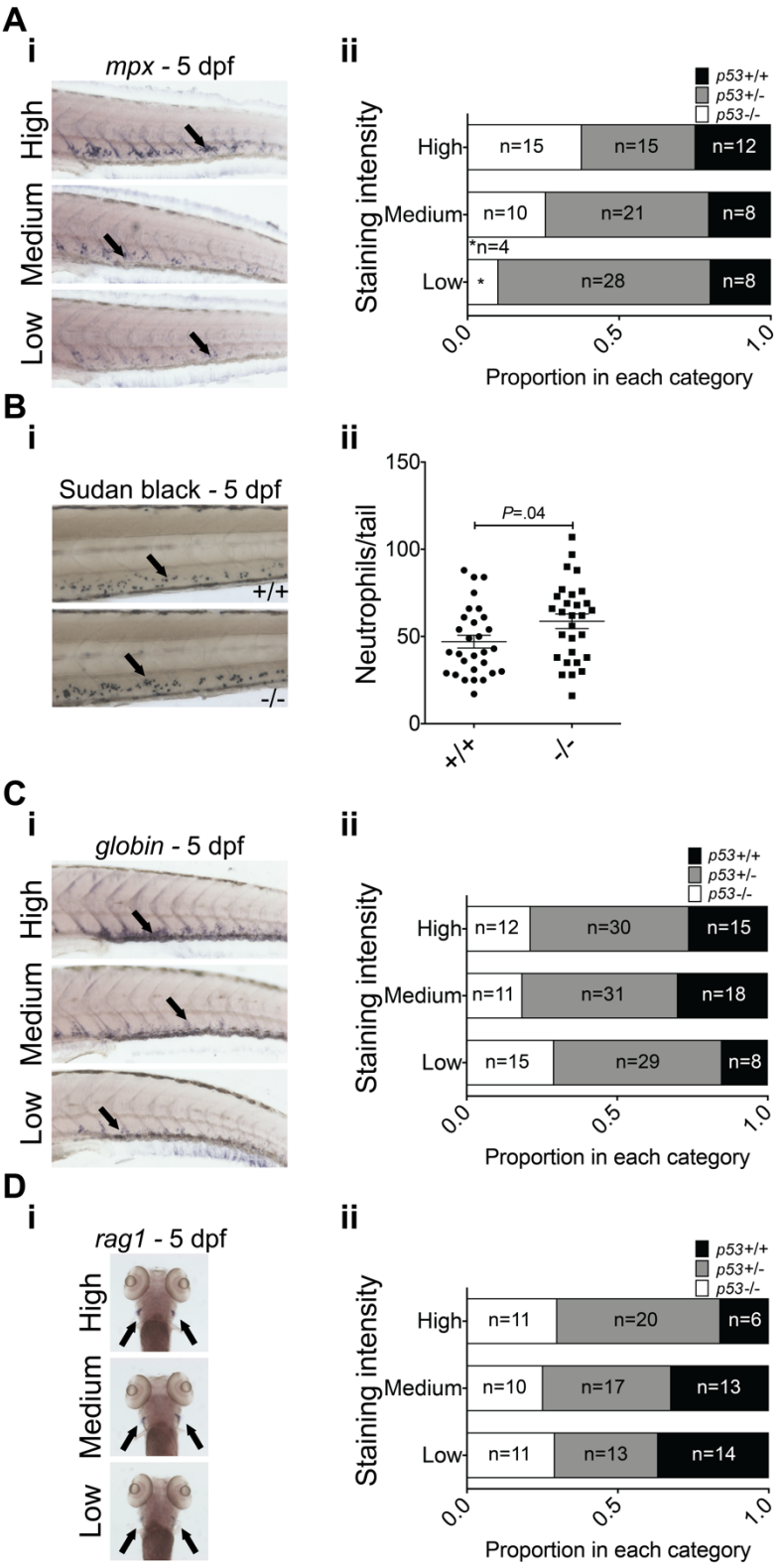


Figure 4.8: Neutrophils are increased in *p53* mutant embryos. (A) Representative embryos stemming from a $p53^{+/-}$ incross stained using an *mpx* ISH probe (i) and quantification of distribution of genotypes in the different staining categories (ii). The arrow shows neutrophils. n = 119 from two clutches. (B) Representative embryos stemming from a $p53^{+/-}$ incross stained using Sudan black (i) and quantification of the number of neutrophils in WT and mutant embryos (ii). The arrow shows neutrophils. Two-tailed

t-test, $P = .04$. $n_{+/+} = 29$, $n_{-/-} = 29$. Bars represent mean ± SEM. (C) Representative embryos stemming from a *p53*$^{+/-}$ incross stained using an *ea1 globin* ISH probe (i) and quantification of distribution of genotypes in the different staining categories (ii). The arrow shows erythrocytes. n = 169 from three clutches. (D) Representative embryos stemming from a *p53*$^{+/-}$ incross stained using a *rag1* ISH probe (i) and quantification of distribution of genotypes in the different staining categories (ii). The arrow shows lymphocytes. n = 115 from two clutches.

4.3 Discussion

In this part of the thesis, I showed that p53-driven signalling mediates many of the defects seen in *rad51* mutant zebrafish, including sex reversal and WKM hypocellularity and that co-mutation of *p53* can rescue many, but not all, of the observed abnormalities.

When considering the expression of *p53,* I found an increase in the developing nervous system. This fits with the small eye and head phenotype observed upon irradiation. It also agrees with the frequent micropthalmia/microcephaly, as well as occasional mental disability seen in FA patients in the clinic[19] and similar symptoms seen in mouse models[358,387]. This is most likely due to the rapid proliferation of cells in the developing nervous system[490]. Curiously, in adult WKM, *p53* expression was consistently decreased, which is very counterintuitive at first glance. A similar phenotype was observed in *Lig4* mutant mice, which develop BM hypocellularity and a hyperproliferative phenotype of the HSPCs to compensate for this[144]. I believe that the most likely explanation for this downregulation is the need to suppress further p53 signalling to maintain kidney homeostasis, i.e. to generate enough blood cells to stay alive. Interestingly, my RNA-Seq results presented in the previous chapter suggest that p53-driven signalling might still be upregulated in certain subsets cells enriched for HSPCs in the WKM. It may well be the case that decreased p53 signalling is necessary for the survival of the more mature, rapidly dividing progenitor cells, but this is speculative. It is unclear how this differential expression between the different blood lineages is regulated. Furthermore, the limitations of my RNA-Seq experiment should be taken into account when interpreting these findings.

I then went on to generate *rad51*, *p53* double mutants. I observed that the sex reversal and haematopoietic defects were rescued in these fish, whereas the small size and infertility of males were not rescued. In addition, double mutants developed tumours at a high rate.

The rescue of sex reversal, but not of infertility is equivalent to what has been observed in *brca2* and *brca2, p53* mutant fish[363,364] and is due to the premature apoptosis of oocytes, coupled to defective meiosis in immature sperm cells. Interestingly, *fancl*

mutant zebrafish also display sex reversal that can be rescued upon *p53* co-mutation, but the mutant males are fertile due to their intact meiosis[362]. This demonstrates the necessity of the FA proteins for zebrafish sex development and that they are dispensable for meiosis. In contrast, Rad51 is required for both sex development, as well as meiosis. Infertility is also a common symptom of FA[19] and it is intriguing to speculate whether some of these cases are due to meiotic defects in the germ cells, at least where FA genes are mutated that also function in HR. However, due to the differences in sex determination between humans and zebrafish, the applicability of these results to humans remains to be explored.

I also showed that *rad51, p53* mutant fish develop malignant peripheral nerve sheath tumours. This is the same kind of tumour commonly found in *p53* single mutants[365] and *brca2, p53* double mutants[363,364,488]. In contrast to single mutants, double mutants developed the tumours much earlier, starting at 5 mpf. The first tumours in single mutants have been reported to arise from 10 mpf onwards, rising to about 30% incidence at 16.5 mpf [365], whereas the *rad51*$^{-/-}$, *p53*$^{-/-}$ fish reached this rate at 8 mpf. This shows that loss of *rad51* can speed up tumour development and demonstrates its role as a tumour suppressor gene. Interestingly, two out of three of the tumours I observed were found in ovarian tissue, which agrees with findings on *brca2* knockout fish[363,364,488]. *BRCA1/2* are well known ovarian cancer predisposition genes[491,492]. Furthermore, the *RAD51* paralogues *RAD51B, C* and *D* have also been implicated as risk factors for ovarian cancer[122,493,494]. Some studies have found a correlation between *RAD51* polymorphisms and ovarian cancer[495], whereas a more robust meta-analysis found no correlation[496]. My data suggests that another look at hypomorphic *RAD51* alleles and ovarian cancer risk may be warranted. Overall it is clear however, that defects in the repair of doubles stranded breaks play an important role in the development of ovarian cancer.

In terms of haematological features, co-mutation of *p53* rescued HSC numbers in embryos. This translated to normal WKM cell numbers during adulthood, as well as normalised proliferation rates. These findings provide further evidence for the work presented by Ceccaldi *et al.* in 2012[133], who proposed a model in which excess p53 signalling during embryonic development leads to the later defects in FA. The normalisation in proliferation suggests that the increased cycling in *rad51* single mutants

is indeed to compensate for lower WKM cellularity, rather than due to increased unrepaired DNA damage. This is as *p53* co-mutation normalised WKM cellularity, but should have no influence on the amount of damage a cell receives. My results clearly show that embryonic abnormalities lead to decreased HSC counts in FA patients, rather than just depletion after birth caused by stressors such as DNA damage.

My conclusions disagree with the findings reported by Yoon *et al.*[137], who proposed that p53-driven pathways play little to no role in FA pathogenesis and that p38 MAPK inhibition can stimulate FA HSC engraftment after transplantation. While it is plausible that MAPK inhibition could alleviate engraftment defects (in particular since p38 is a known regulator of HSC metabolism and proliferation[497]), the data presented showing no involvement of p53 was unconvincing. The authors concluded the irrelevance of p53 on the basis that the two downstream target genes they investigated (*Puma* and *Noxa)* were unchanged. Furthermore, they claimed that HSC proliferation was not impaired based on very broadly labelled haematopoietic cells, potentially missing proliferation defects in actual stem cells. Overall, I find the evidence from genetic ablation of p53 to be more convincing.

Finally, I considered the role of p53 on its own in HSPC numbers, showing that *p53* mutation can increase the number of embryonic HSPCs and neutrophils, but does not affect WKM cell numbers during adulthood. This is in agreement with work on murine models, in which loss-of-function (LOF) mutations of *p53* led to two- to threefold increases in the number of HSCs, which were however functionally impaired[463,472,473]. Conversely, hyperactivation of p53 can lead to decreased HSC numbers in aged mice[143]. It is unclear why only embryonic HSPCs and neutrophils were affected, but no other lineages. Together, this underscores the vital role p53 plays in the regulation of HSC number and activity. It also highlights potential differences between embryonic and adult haematopoiesis.

In conclusion, my results confirm previous *in-vitro* work implicating p53 in FA pathogenesis in an *in-vivo* context, as well as highlighting its general role in haematopoiesis.

5 BMF in FA: initial triggers

5.1 Introduction

After investigating the signalling pathways that lead to apoptosis and senescence in FA cells, I focussed my attention on the initial causes of these signals. As alluded to in earlier parts of this thesis, there are many hypotheses as to why the bone marrow fails in FA. Essentially, every molecular abnormality seen in FA cells has been proposed as the cause for the progressive BM dysfunction that most FA patients eventually succumb to. In this study, I explore two of the major hypotheses, which are almost complete opposites. On one hand we have inflammation, which has been proposed to drive HSC proliferation and thereby cause their eventual exhaustion. On the other hand, there is aldehyde-induced DNA damage, which is thought to induce senescence and apoptosis of HSCs. I will discuss previous work on both hypotheses in turn, followed by an experimental investigation into both.

5.1.1 Inflammation in FA pathogenesis

The term inflammation describes the biological reaction in response to a disruption in tissue homeostasis[167]. Inflammation commonly includes tissue destruction and recruitment of blood constituents (plasma protein, fluid and leukocytes) into perturbed tissues. The entry of cells into the tissue is mediated via changes in the vasculature. These include vasodilation, increased permeability of the vascular system and increased blood flow. In combination, this leads to the classical symptoms of local inflammation, as already defined by Galen: calor, dolor, rubor, tumour (heat/fever, pain, reddening and swelling)[498].

The main causes of inflammation are infection, injury, trauma and exposure to foreign particles (e.g. fine dust in the lung leading to chronic inflammation)[498]. It is not completely clear why different stimuli such as infection and traumatic injury should give rise to very similar responses. There are two hypotheses as to why that might be the case. The first proposes that infections often follow traumatic injury, so it would be favourable

to respond to any injury with inflammation[499]. The other suggests that pathogens and wounding just trigger similar molecular signalling pathways[500].

There are several steps in the inflammatory response, which have been highly conserved during evolution[498]. First, pathogen-associated molecular patterns (PAMPs) are recognised by cells at the site of infection. Next, these cells release alarmins or danger-associated molecular patterns (DAMPs), which alert the innate immune system. Immune cells start producing inflammatory cytokines, such as interleukin (IL)1b, IL6 or tumour necrosis factor (TNF)α. Molecules like these lead to the recruitment of neutrophils, monocytes and other cells to the site of infection. At the site of infection, neutrophils create a cytotoxic environment by degranulating, which leads to the release of ROS and reactive nitrogen species (RNS), as well as proteinases. In this environment, both host and invading cells are killed or attenuated. Macrophages remove infected host cells and invaders. Normally, inflammation is resolved once the immediate cause has been removed. If this does not happen, unregulated inflammation can lead to unwanted side effects, including collateral damage to the tissue and pathological changes[498].

5.1.1.1 The effects of inflammation on HSCs

One of the major characteristics of inflammation is that more effector cells are needed at the site of tissue injury. This means that haematopoietic progenitors have to cycle more to provide a sufficient number of cells to fight off injection. It has recently been shown that HSCs are directly affected by inflammatory signalling[501]. Among others, interferons, IL1, TNF, Toll-like receptor (TLR) -ligands and G-CSF can stimulate proliferation and differentiation of HSCs[501–504], which can come at the expense of HSC self-renewal[505]. It has been shown that HSCs are negatively affected when they leave quiescence in the sense that they gradually lose their functionality every time they are activated[421]. This is why chronic inflammation may be deleterious to HSCs – the excess cycling can bring them out of quiescence[65,134,506] and lead to p53-dependent apoptosis even in WT cells[504]. Moreover, FA cells are also thought to be both more sensitive to inflammation, as well as to express more inflammatory cytokines. In addition to becoming less functional with each division, they are affected even more strongly, as

they cannot repair DNA damage that accumulates during DNA replication[65,134,506]. Both of these aspects are discussed in turn below.

5.1.1.2 Cytokine hypersensitivity

There is substantial evidence that FA cells react excessively to inflammatory stress mediated via pro-inflammatory cytokines. Previous research has especially focussed on the role of TNFα and interferon gamma (IFNγ). In cell culture experiments, TNFα can elicit cytotoxicity[168,507] and can induce apoptotic and necrotic cell death[10,508,509]. The FANCC protein has been implicated in regulating the apoptotic response to TNFα and Fas ligand[168]. Furthermore, it has been implicated in regulating protein kinase R activation, as well as interacting with HSP70 and STAT1, which plays an important role in IFN signalling[169,507,510].

Apart from this *in-vitro* evidence for cytokine hypersensitivity, there has also been considerable work done on animal models. Inflammatory stress induced by pI:pC can induce BMF in FA mice[65,134]. Similar results can be obtained by using the vaccine adjuvant alum, which triggers repeated emergency granulopoiesis leading to BMF in FA mice[511], as well as with TNFα[512] and IFNγ[513]. Increased TNFα also induces leukaemic clonal evolution in murine FA stem cells upon transplantation[514], demonstrating the role of inflammation in the development of malignancies.

5.1.1.3 Cytokine overproduction

Macrophages derived from the BM of FA mice have been shown to overproduce TNFα and IL1β[63,170]. FANCC has been implicated to regulate TNFα production in monocytes by suppressing TLR8 activity[171]. Moreover, FANCD2 can regulate the expression of TNFα by binding to its promoter[515]. In a mouse study of FA, it was shown that the LINE-1 retrotransposon becomes hyperactive in FA, which leads to the accumulation of excess nucleic acids in the cytoplasm. These are recognised by receptors that normally sense PAMPs, leading to the production of excess inflammatory cytokines[172]. Apart from this experimental data, there is also support from clinical research suggesting an overproduction of inflammatory cytokines in FA. Patients express considerably higher amounts of TNFα[173,174,516] and IL1β[517]. However, in contrast to these studies, Matsui *et al.* did not find overproduction of inflammatory cytokines in FA patients when their cells

were unchallenged. Challenge with lipopolysaccharide (LPS) led to an increase in cytokines, but this was also seen in healthy controls, suggesting that cytokine overproduction is just a general feature of BM dysfunction and not specific to FA[518]. In agreement with this, irradiation of mice induces high levels of TNFα, which mediates apoptosis in the bone marrow that can be rescued upon TNFα co-mutation[519].

It is currently disputed whether cytokine hypersensitivity and DNA repair defects are linked in FA[520], although there is limited evidence that cytokine overproduction and ICL sensitivity are unlinked at least in macrophages[521].

My results investigating the role of inflammation in the pathogenesis in FA suggest that increased cytokine sensitivity plays a contributing role in BMF, but intrinsic cytokine overproduction does not. These results are described in detail in section 5.2.1.

5.1.2 Aldehydes in FA pathogenesis

As discussed previously, the main (and best studied) function of the Fanconi pathway is the removal of ICLs from DNA. It is uncertain however what damaging agent is the predominant cause of ICLs in FA patients. The Patel group in Cambridge has accumulated considerable evidence in the past few years that the main damaging molecules are small aldehydes. This section will deal with the different types of aldehydes relevant to biology and how they might contribute to FA pathogenesis.

5.1.2.1 Sources of aldehydes

Aldehydes are ubiquitous molecules defined by their reactive aldehyde functional group. They can be reduced to alcohols or oxidised to organic acids. They are a very diverse group of molecules, found in nature and the chemical industry alike. While a lot of aldehydes are safe to ingest and are even found in food, many substances of this class are quite toxic due to their reactivity with macromolecules. It is impossible to avoid toxic aldehydes completely. Environmental sources, such as motor vehicle emissions, cigarette smoke, industrial chemicals and even rainwater contain large amounts of aldehydes, in particular formaldehyde and acetaldehyde. Our diet is also a major source of aldehydes. Acetaldehyde is found naturally in many fruits and vegetables. Cinnamaldehyde gives cinnamon its characteristic fragrance, while benzaldehyde gives

almonds their flavour. Many food components become metabolised to aldehydes as well. The clearest examples are of course alcohols such as methanol in fruits, which is converted to formaldehyde, and ethanol from alcoholic beverages, which is converted to acetaldehyde. Other metabolic reactions also lead to acetaldehyde products. Lipids may be peroxidised to aldehydes, carbohydrates oxidised to glyoxal or methylglyoxal and amines oxidised to various aldehydes. Myeloperoxidase found in neutrophils can also catalyse the formation of various aldehydes. Many drugs are already aldehydes or may be metabolised to aldehydes. Examples include the HIV drug Abacavir and some anticancer drugs that are metabolised to acrolein by P450[145].

5.1.2.2 Aldehyde scavengers

Due to the ubiquitous nature of aldehydes, a wide range of aldehyde scavenging enzymes has evolved. There are nine aldehyde oxidising enzymes, seven aldehyde reducing enzymes and three glutathione (GSH) dependent enzymes[145].

Most aldehyde oxidases can be found in the aldehyde dehydrogenase (ALDH) family. This includes the slow cytosolic ALDH1 and the much faster mitochondrial ALDH2, which is often dysfunctional in Asian ethnic groups. Among other things, these enzymes process acetaldehyde to acetate. Aldehyde reducing enzymes includes alcohol dehydrogenases. Furthermore, there are GSH-dependent enzymes like ADH5, which processes formaldehyde[145].

5.1.2.3 DNA damage induced by aldehydes

Many of these aldehydes are mutagenic and genotoxic. That is because they can form various forms of adducts on DNA. The simplest (and best studied) case is acetaldehyde. There are two main forms of adducts induced by acetaldehyde on the DNA. One option is N2-ethyl-2-deoxyguanosine (N2-ethyl-dG), formed from acetaldehyde and deoxyguanosine. However, this adduct is not particularly mutagenic, as it can be bypassed by translesion polymerases such as DNAPδ and η, with high fidelity. Another, more dangerous adduct is 1,N2-propano-2-deoxyguanosine (PdG), which is much more mutagenic and genotoxic. PdG can be a precursor for interstrand DNA-DNA or DNA-protein crosslinks, which are the substrates of the FA pathway. PdG itself can be

removed by nucleotide excision repair. Other aldehydes can form similar types of adducts in principle[522].

5.1.2.4 Evidence for aldehydes as main damaging molecules in FA pathogenesis

FA pathway deficient cells are hypersensitive to both acetaldehyde, as well as formaldehyde in culture[12,14,146,147]. Moreover, mutations in *fancd2* and *adh5* (the gene encoding for the alcohol dehydrogenase necessary for formaldehyde metabolism) are synthetic lethal in chicken DT40 cells[12].

Further evidence for the importance of aldehydes in the pathogenesis of FA stems from mouse models. Co-mutating *Fancd2* and *Aldh2* (acetaldehyde dehydrogenase) results in embryonic lethality in *Aldh2*$^{-/-}$ mothers, but embryos can be carried to term in Aldh2$^{+/-}$ mothers. However, the surviving *Fancd2*$^{-/-}$ *Aldh2*$^{-/-}$ all invariably develop acute leukaemia or BMF due to a 600-fold reduction in HSPC numbers[14,15]. Similar results were obtained in *Fanca*$^{-/-}$ *Aldh2*$^{-/-}$ mice[13]. Loss of *Adh5* in conjunction with FA mutations led to even more serious defects in double mutant mice, with invariable BMF due to a 950-fold reduction in HSPCs[131]. This indicates that several types of aldehydes can induce similar damage.

Lastly, there is also epidemiological evidence for the importance of aldehydes in FA. In humans, there are several different alleles of *ALDH2*. In Southeast Asia, a dominant negative, inactive isoform of is common. Carriers of this allele show the so-called "Asian flushing" syndrome in response to alcohol, as the alcohol can be processed to acetaldehyde quickly, but the further metabolism is slow, leading to accumulation of the toxic acetaldehyde[523]. FA patients with this form of *ALDH2* show a more severe phenotype, i.e. they have earlier BMF and stronger congenital features[148,149]. This again underlines the importance of the pathway during early development.

My own results investigating the role of aldehyde-induced DNA damage *in-vivo* do not suggest a major role in the development of BMF. These results are described in detail in section 5.2.2 and the difference in outcome when compared to the mouse models of aldehyde-induced damage is examined in the discussion of this chapter (section 5.3).

5.2 Results

5.2.1 Modelling the influence of prolonged inflammation in FA

5.2.1.1 Establishing a zebrafish model of prolonged inflammation

To investigate the role of inflammation in BMF, I developed a model for prolonged inflammation in zebrafish. To do this, I first selected an appropriate compound to induce inflammation. In the literature, various molecules are used, but pI:pC appealed most to me, as it is very stable, non-toxic and has previously been used successfully to induce inflammatory stimuli in fish and other species[524–527]. I first began by assessing the maximum tolerated single dose of pI:pC when intraperitoneally injected into WT fish. I determined that even the highest soluble concentration of pI:pC in PBS (10 mg/ml) is well tolerated by the fish when 10 µl are injected. In parallel, I began assessing the efficacy of the injections in terms of eliciting a robust inflammatory response (Figure 5.1A). I began by looking at the expression of several genes linked to inflammation (*il1b, il8, tnfa, p53*) in the kidney at 6 hpi (Figure 5.1B). These results revealed a robust upregulation of the interleukins, as well as an upregulation of *p53*. The expression of TNFα was unchanged. Because of this upregulation of inflammatory genes, I concluded that bolus pI:pC injections are efficacious at triggering inflammation in zebrafish.

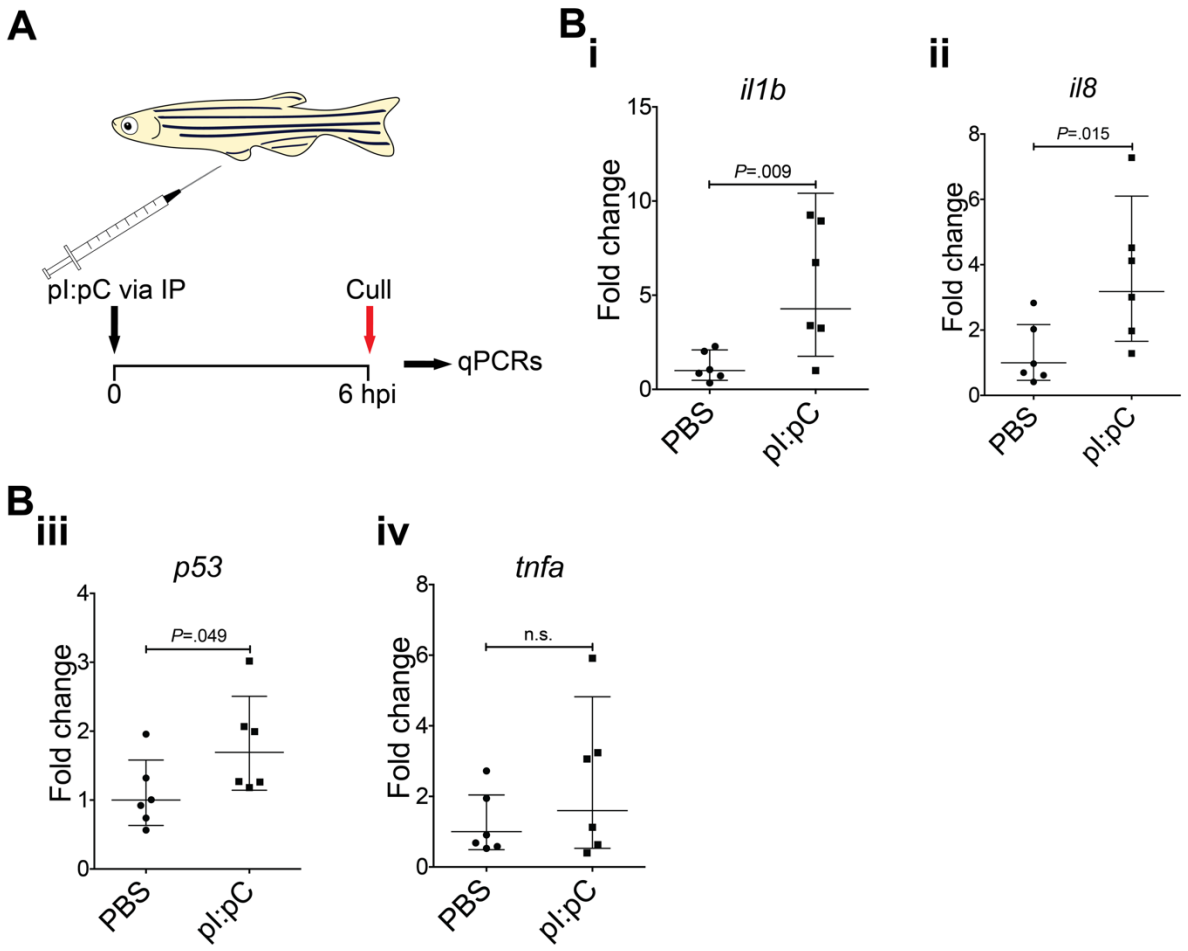


Figure 5.1: A bolus injection of pI:pC can induce the expression of marker genes for inflammation. (A) Injection scheme for the experiments. Fish were injected with 10 µl 10 mg/ml pI:pC and culled at 6 hpi to extract tissue for qPCR analysis. (B) Expression of inflammatory markers in the WKM at 6 hpi. (i) *il1b*. *P* = .009. (ii) *il8*. *P* = .015. (iii) *p53*. *P* = .049. (iv) *tnfa*. *P* = .38. PBS; n = 6, pI:pC; n = 6. *P*-values stem from two-sided Student's t-test and bars represent the geometric mean ± 95% CI in all graphs.

After establishing that single injections can robustly induce acute inflammation, I aimed to expand this model to prolonged inflammation. My initial test was scheduled to span four weeks, with weekly injections totalling to five rounds of injections, with culling, followed by qPCR and FACS analysis three days after the last injection. While the initial injections were tolerated well, after four rounds of injections, some fish started showing signs of bleeding and inflammation at the site of injection, with two out of five even dying due to the effects of this stress (Figure 5.2A). Because of this, the experiment had to be stopped early and fish were culled five days after the last injection. Again, RNA from the WKM and intestine was used for qPCR analysis, showing increases in several inflammatory markers (Figure 5.2B). However, the results were not statistically

significant, presumably due to the long time between the last injection and obtaining the tissue in combination with a decreased number of fish due to the unexpected deaths. Nevertheless, together with the very strong side effects seen in the injected fish, these results were encouraging enough to begin applying my model to the *rad51* mutant line.

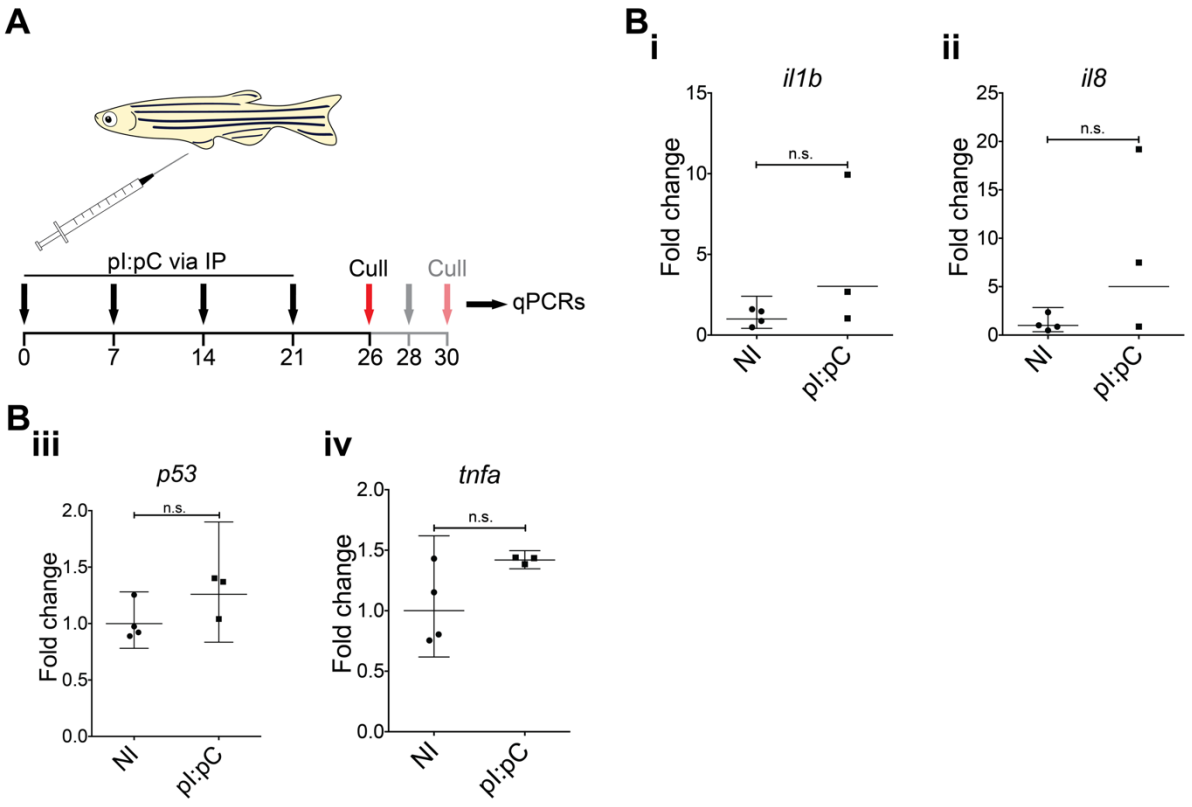


Figure 5.2: Long-term pI:pC injection trial. (A) Outline of the injection schedule. Fish were injected with 10 µl 10 mg/ml pI:pC once a week, totalling four injections, followed by culling five days after the fourth injection. In lighter colours, the initially planned experiment is outlined, continuing with a fifth injection and culling at three days after the last injection. (B) qPCR analysis of inflammatory genes, again showing increases in two out of three fish (i) *il1b*. $P = .072$. (ii) *il8*, $P = .11$. (iii) *p53*, $P = .133$. (iv) *tnfa*, $P = .11$. All were tested using a two-sided Student's t-test, $n_{NI} = 4$, $n_I = 3$. Bars represent the geometric mean ± 95% CI in all graphs, except for Bi and ii, where the 95% CI is too large to be displayed without skewing the graph.

5.2.1.2 Prolonged inflammation exacerbates the phenotype of *rad51* mutant fish

I modified my injection schedule to take into account the severe side effects seen during the first trial, leading to a schedule of four weekly injections followed by culling three days after the last injection (Figure 5.3A) (i.e. one injection fewer than planned before).

Overall, the experiment consisted of four groups: non-injected *rad51*$^{+/+}$ fish, non-injected *rad51*$^{-/-}$ fish, pI:pC injected *rad51*$^{+/+}$ fish and pI:pC injected rad51$^{-/-}$ fish. Following the injection period, all fish were culled. From the dead fish, I obtained blood smears for observing changes in the PB as well as, kidneys for cell counts, FACS and qPCR.

I confirmed the successful effect of the injections by assessing *il1b* and *il8* expression following the experimental schedule outlined in Figure 5.3A. Interestingly, both markers were decreased upon injection (Figure 5.3B), indicating an unexpected downregulation of cytokines involved in acute inflammation at this time point after injection This was in stark contrast to my previous experiments, where I used a bolus pI:pC injection on WT fish and assessed the expression of *il1b* and *il8* at 6 hpi (Figure 5.1). This showed a robust response in both WT and mutant fish (Figure 5.3B). Importantly, there was no difference in the expression of inflammatory genes between uninjected WT and mutant fish (Figure 5.3B).

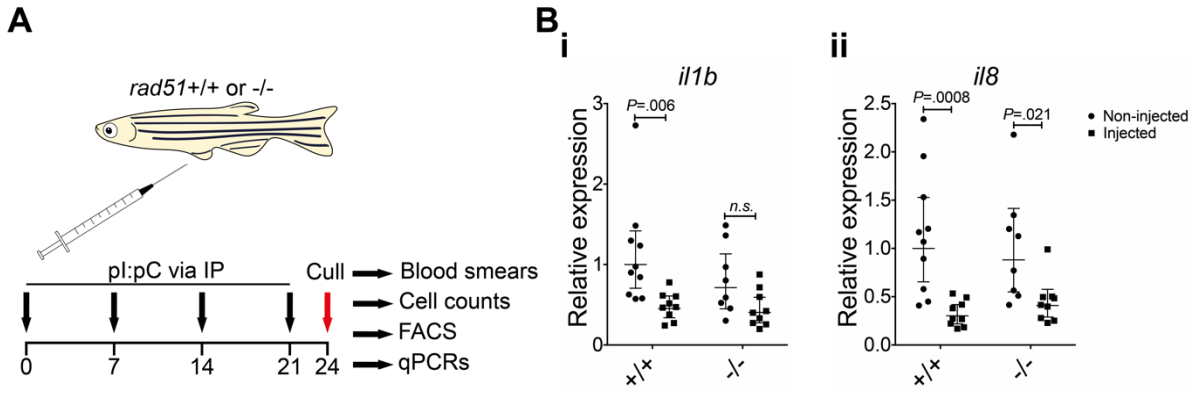


Figure 5.3: Repeated pI:pC injections elicit a different inflammatory response than single injections. (A) Schematic of the experimental design. Both wild type and *rad51*$^{-/-}$ fish were injected every seven days with 10 µl 10 mg/ml pI:pC, four injections in total. All fish were culled three days after the last injection. Control fish were not injected (B) Expression of acute inflammatory markers *il1b* (i) and *il8* (ii) after my injection schedule. (i) Two-way ANOVA showed a statistically significant effect of injection status ($F(1, 32) = 17.7$, $P = .0002$), but not of mutation status ($F(1, 32) = 1.97$, $P = .17$), or an interaction between the two ($F(1, 32) = .485$, $P = .49$). (ii) Two-way ANOVA showed a statistically significant effect of injection status ($F(1, 32) = 32.74$, $P < .0001$), but not of mutation status ($F(1, 32) = .26$, $P = .61$, or an interaction between the two ($F(1, 32) = 1.55$, $P = .22$). In both graphs, *P*-values shown stem from *post-hoc* Tukey multiple comparison tests. *rad51*$^{+/+}$, $n_{NI} = 10$, $n_I = 9$; *rad51*$^{-/-}$, $n_{NI} = 8$, $n_I = 9$. Bars represent the geometric mean ± 95% CI.

To assess changes in the PB, I obtained blood smears during the culling. Slides were then blindly scored for the different blood lineages (See example images in the panels labelled "i" in Figure 5.4 for examples of different blood cell types scored). This showed a statistically significant increase in monocytes upon injection in both WT and mutant fish (Figure 5.4A). There were no changes in the other blood lineages considered (Figure 5.4B-D).

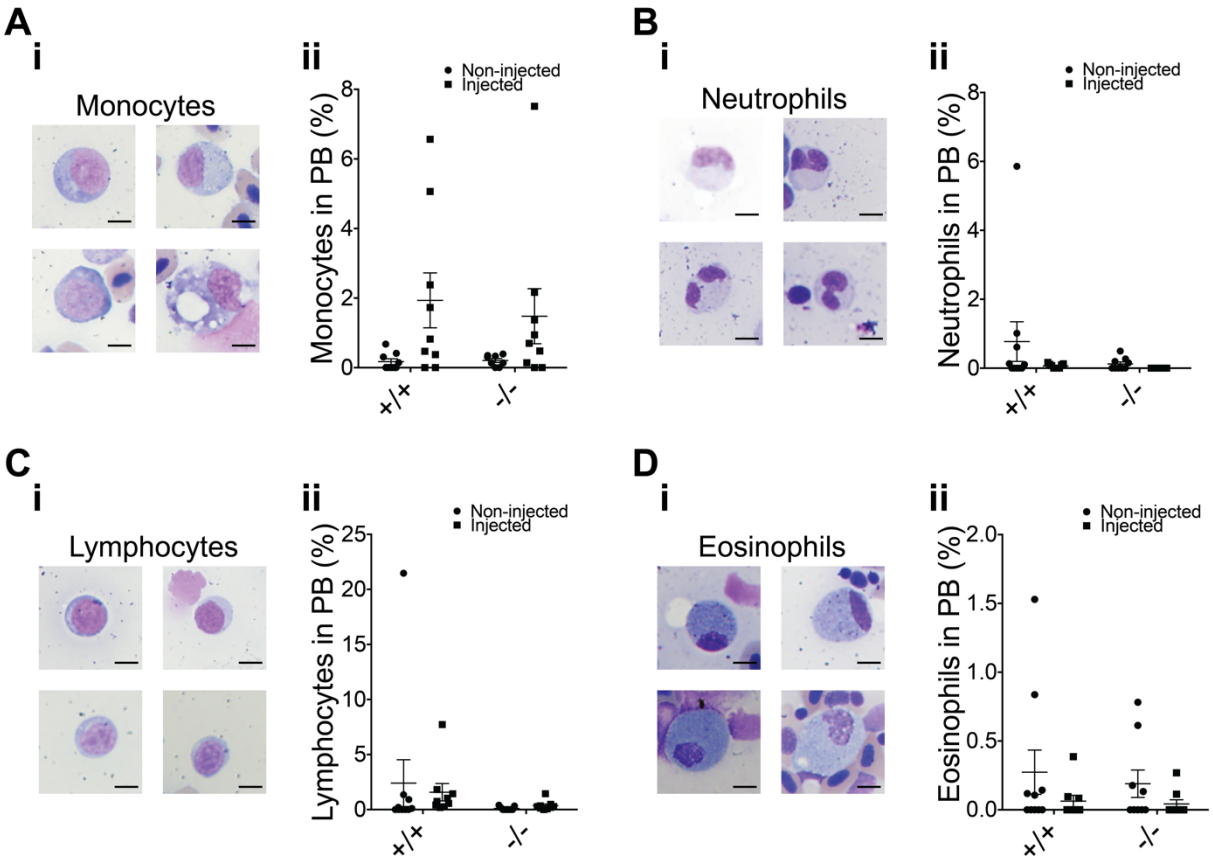


Figure 5.4: Prolonged inflammation leads to an increase of monocytes in the PB. (A) Changes in the monocytes. (i) Example pictures of cells scored as monocytes. Visible are large cells, with large nuclei. They occasionally contained large vacuoles in the cytoplasm. (ii) Quantification of monocytes in the PB. Two-way ANOVA revealed an influence of injection status ($F(1, 31) = 6.897$, $P = .133$), but no influence of mutation status ($F(1, 31) = .13$, $P = .72$) or an interaction between these factors ($F(1, 31) = .18$, $P = .67$). (B) Changes in the neutrophils. (i) Example pictures of cells scored as neutrophils. Visible are large cells with a normally multi-lobed nucleus. (ii) Quantification of neutrophils in the PB. Two-way ANOVA found no statistically significant difference between the groups in terms of injection status ($F(1, 31) = 2$, $P = 17.$), mutation status ($F(1, 31) = 1.4$, $P = .24$), or an interaction between the two ($F(1, 31) = .99$, $P = .33$). (C) Changes in the lymphocytes. (i) Example pictures of cells scored as lymphocytes. Visible are small cells with large nuclei. (ii) Quantification of lymphocytes in the PB. Two-way ANOVA found no statistically significant difference between the groups in terms of injection status ($F(1, 31) = 11$, $P = .74$), mutation

status (F (1, 31) = 2.16, *P* = 15.), or an interaction between the two (F (1, 31) = .27, *P* = .61). (D) Changes in the eosinophils. (i) Example pictures of cells scored as eosinophils. Visible are large cells with a comparatively small nucleus and a strongly stained, granular cytoplasm. (ii) Quantification of eosinophils in the PB. Two-way ANOVA found no statistically significant difference between the groups in terms of injection status (F (1, 31) = 3.66, *P* = .065), mutation status (F (1, 31) = .27, P = .61), or an interaction between the two (F (1, 31) = .11, P = .75). Bars represent mean ± SEM in all graphs. rad51+/+ n_{NI} = 9, n_I = 8; rad51-/- n_{NI} = 8, n_I = 9. Pictures were taken using a 63X oil immersion objective. Scale bars = 10 μm.

For the FACS analysis, I used standard gating that allows one to differentiate between the different blood lineages in the kidney[377] (Figure 2.4), with slight modifications from the original procedure to account for monocytes and HSPCs. This gating was similar to the one used for my RNA-Seq experiment discussed previously (see section 3.2.3.9 for details). The FACS results were then combined with absolute cell counts of the WKM, to obtain an estimate of the absolute number of each investigated blood lineage (Figure 5.5A).

Interestingly, WT fish expanded their monocyte population at the expense of the erythrocytes, whereas *rad51*$^{-/-}$ fish displayed a decrease in erythrocytes without being able to increase the monocytic population (Figure 5.5A). In agreement with this, WT fish increased the expression of the monocyte markers *marco* and *csfr1* in the WKM upon injection (Figure 5.5B).

Furthermore, overall kidney cellularity was decreased upon pI:pC injection in *rad51* mutants, but not in WT fish (Figure 5.5C). This shows that prolonged inflammatory stress can indeed exacerbate the kidney hypocellularity of the mutant fish even further, providing further evidence for the role of inflammation in BMF of FA patients.

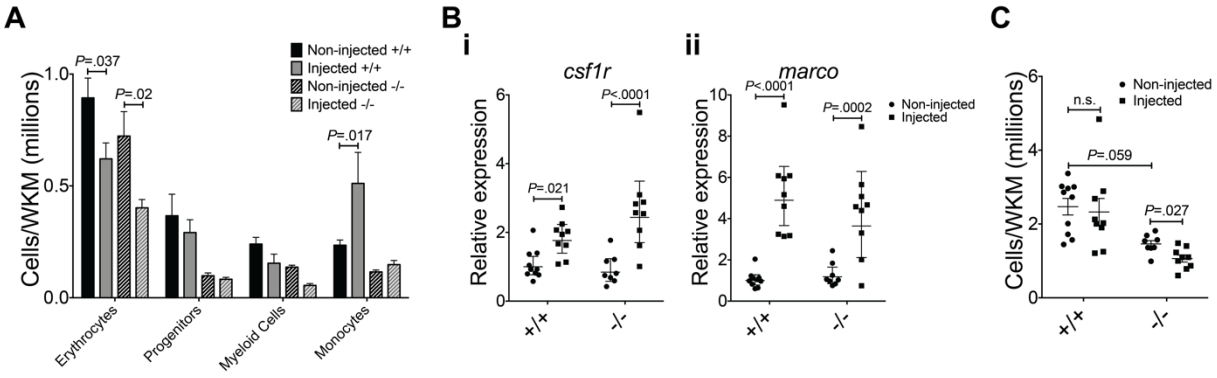


Figure 5.5: Prolonged inflammation leads to an exacerbation of the mutant phenotype. (A) Absolute number of cells belonging to different blood lineages in the kidney gained by combining FACS data with the cell counts shown in A. Statistical tests were carried out individually for each cell type, using two-way ANOVA. *P*-value shown on the graph stems from a *post-hoc* Šidak multiple comparison test, comparing non-injected to injected fish within each genotype. *rad51*$^{+/+}$, $n_{NI} = 10$, $n_I = 9$; *rad51*$^{-/-}$, $n_{NI} = 8$, $n_I = 9$. Erythrocytes: There was a significant effect of injection status ($F(1,32) = 13.66$, $P = .0008$), a significant effect of mutation status ($F(1,32) = 5.9$, $P = .021$), but no significant interaction between the two ($F(1,32) = .09$, $P = .77$). Progenitors: There was a significant effect of mutation status ($F(1,32) = 15.04$, $P = .0005$), but not of injection status ($F(1,32) = .54$, $P = .47$) or an interaction between the two ($F(1,32) = .24$, $P = .63$). Myeloid cells: There was a significant effect of injection status ($F(1,32) = 9.4$, $P = .004$), a significant effect of mutation status ($F(1,32) = 13.84$, $P = .0008$), but no significant interaction between the two ($F(1,32) = .007$, $P = .93$). Monocytes: There was a significant effect of injection status ($F(1,32) = 4.67$, $P = .038$), a significant effect of mutation status ($F(1,32) = 11.36$, $P = 002.$), but no significant interaction between the two ($F(1,32) = 2.89$, $P = .098$). Bars represent mean ±SEM. (B) Expression of the monocyte/macrophage markers *csf1r* (i) and *marco* (ii) in the WKM. (i) Expression of *csf1r*. Two-tailed ANOVA showed a statistically significant effect of injection status ($F(1, 32) = 36.74$, $P < .0001$), but not of mutation status ($F(1, 32) = .33$, $P = .57$) or an interaction between these factors ($F(1, 32) = 3.37$, $P = .076$). (ii) Expression of *marco*. Two-tailed ANOVA showed a statistically significant effect of injection status ($F(1, 32) = 71.84$, $P < .0001$), but not of mutation status ($F(1, 32) = 16$, $P = .69$) or an interaction between these factors ($F(1, 32) = 2.1$, $P = .16$). (C) The total number of cells in the kidney in injected and non-injected fish. Two-way ANOVA was carried out on the reciprocal of the data to fulfil the requirement of homoscedasticity as measured by Bartlett's test (before transformation: $P = .0002$, after transformation: $P = .095$). There was a statistically significant effect of mutation status ($F(1, 32) = 29.86$, $P < .0001$) and of injection status ($F(1, 32) = 6.778$, $P = .014$), but not due to an interaction between the two ($F(1, 32) = 2.95$, $P = .095$). *P*-value shown on the graph stems from a *post-hoc* Tukey multiple comparison test. For all groups, n is the same as in A. Bars represent mean ± SEM in all graphs.

As I previously saw that the WKM hypocellularity was caused by *p53*-driven defects, I investigated the expression of *p53* and its downstream partner *p21* in the WKM (Figure 5.6A). This showed that as before, *p53* was decreased in the *rad51*$^{-/-}$ fish before pI:pC injection. However, upon pI:pC injection, *p53* was massively upregulated exclusively in the mutants. A similar pattern was apparent for *p21,* albeit not to the same degree. Cell viability as measured by PI incorporation during FACS were decreased to the same degree upon pI:pC injection in both WT and mutant fish (Figure 5.6B). This indicated to me that *p53*-driven cell cycle arrest combined with a lot of monocytes leaving the WKM was the main driver of WKM hypocellularity in the *rad51* mutants upon inflammation.

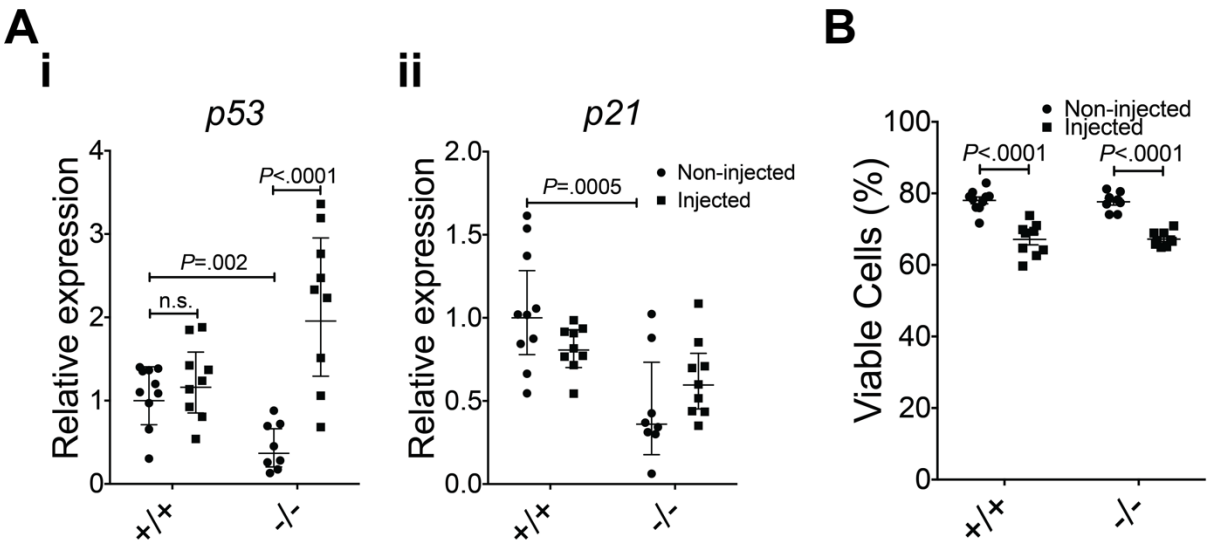


Figure 5.6: The p53 pathway responds differently to prolonged inflammation in WT and mutant fish. (A) Relative expression of genes linked to apoptosis and proliferation. (i) Expression of *p53*. Two-way ANOVA showed a significant influence of injection status (F (1, 32) = 18.16, *P* = .0002) and of an interaction between mutation status and injection status (F (1, 32) = 25.99, *P* < .0001), but not of mutation status on its own (F (1, 32) = 1.81, *P* = .19). (ii) Expression of *p21*.Two-way ANOVA showed a significant influence of mutation status (F (1, 32) = 8.153, *P* = .0003) and of an interaction between mutation status and injection status (F (1, 32) = 4.96, *P* = .033), but not of injection status on its own (F (1, 32) = .79, *P* = .38). *P*-value shown on the graphs stems from a *post-hoc* Tukey multiple comparison test. Bars represent geometric mean ± 95% CI. (B) Viable cells as determined by PI-staining. Two-way ANOVA revealed a significant influence of injection status (F (1, 32) = 100.1, *P* < .0001), but not of mutation status (F (1, 32) = .019, *P* = .88) or an interaction between the two (F (1, 32) = .041, *P* = .84). *P*-values on the graph stem from a *post-hoc* Tukey multiple comparison test. Bars represent mean ± SEM. Numbers are the same as in Figure 5.5.

To test this hypothesis, I conducted another long-term, pI:pC injection experiment, but this time I co-injected BrdU (Figure 5.7A). Due to constraints in fish numbers, only three groups were used: non-injected *rad51*$^{+/+}$ fish, pI:pC injected *rad51*$^{+/+}$ fish and pI:pC injected *rad51*$^{-/-}$ fish. During the injection round one day before culling, BrdU was co-administered as appropriate. Interestingly, pI:pC administration had no effect on proliferation rates in WT fish (Figure 5.7B). Furthermore, the results demonstrate that even one pI:pC injection was enough to decrease the normally increased proliferation in the kidney of *rad51*$^{-/-}$ fish (apart from one outlier demonstrating the normally increased proliferation), a change that was maintained after multiple injection rounds. This shows that the further reduction in WKM cell numbers was indeed driven by decreased proliferation caused by exacerbated p53/p21 signalling.

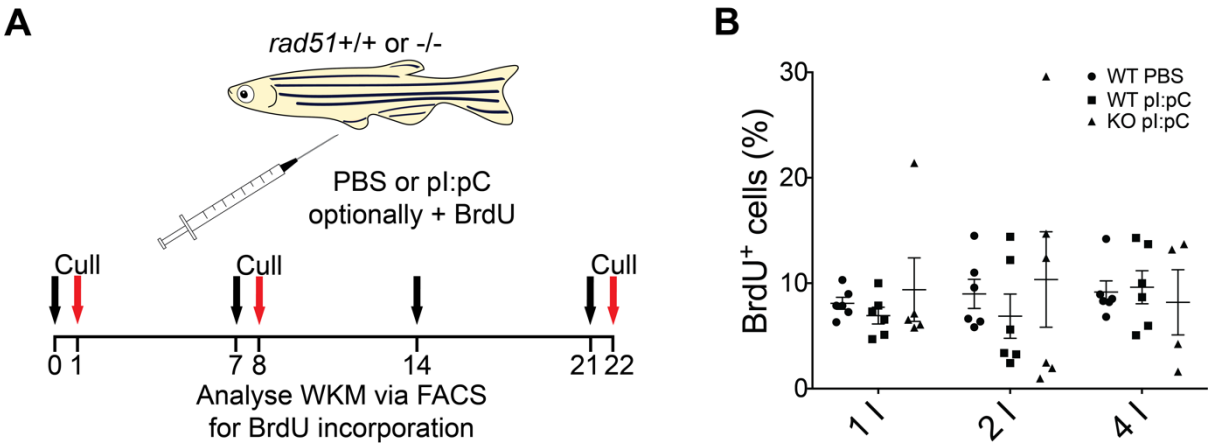


Figure 5.7: Prolonged inflammation can abrogate the excess proliferation in the mutant WKM. (A) Injection scheme for determining cell proliferation in pI:pC injected fish. The experiment was started with a cohort of 36 WT and 18 mutant fish. 12 WT were injected with PBS, 12 with pI:pC, 6 with PBS+BrdU and 6 with pI:pC+BrdU. 12 mutants were injected with pI:pC and 6 with pI:pC+BrdU. On the next day, the BrdU co-injected fish were culled and analysed. On day 7, 6 WT were injected with PBS, 6 with pI:pC, 6 with PBS+BrdU and 6 with pI:pC+BrdU. 6 mutants were injected with pI:pC and 6 with pI:pC+BrdU. On the next day, the BrdU co-injected fish were culled and analysed. On day 14, 6 WT fish were injected with PBS and 6 with pI:pC. 6 mutants were injected with pI:pC. On day 21, the same fish were injected again with the same substances and BrdU. All remaining fish were culled the next day and analysed. (B) Percentage of BrdU^{+} cells in the WKM of wild type (WT) and knockout (KO) fish in response to pI:pC. 1I = one injection, 2I = two injections, 4I = four injections. Bars represent mean ± SEM. Numbers are outlined in A.

5.2.2 Modelling the role of acetaldehyde-induced damage in FA

5.2.2.1 Acetaldehyde dose estimation

I first established that WT zebrafish can tolerate doses up to 10 µl 10% acetaldehyde intraperitoneally, albeit showing redness at the site of injection and abdominal muscle contractions up to five minutes after the injection, in combination with signs of stress. However, after that initial period, injected fish generally showed no further impairments.

5.2.2.2 Aldehyde-derived stress in *rad51* mutant zebrafish

In an initial experiment, I injected nine WT and eight mutant fish with 10 µl 10% acetaldehyde each. The WT fish generally tolerated the injections well. The mutants on the other hand showed drastic side effects right after injection, including extreme arching of the back, bleeding in the abdomen and clear signs of distress and pain. Two fish immediately succumbed to the side effects caused by the injection. Of the remaining six mutant fish, all but one died or had to be culled due to symptoms breaching the constraints of the Home Office licence this work was performed under, whereas only one WT had to be culled. This showed that *rad51* mutants are more susceptible to the acute toxicity of acetaldehyde than their WT siblings.

Because of the striking symptoms displayed by the injected mutants, I decided to drastically limit the dose of injected acetaldehyde for the next experiment. I repeated the same injection schedule as for prolonged inflammation, but substituted pI:pC with 10 µl 1% acetaldehyde (Figure 5.8A). Out of eight injected WT fish, none died and out of nine injected mutants, only one had to be culled due to adverse effects. After four rounds of injections, the fish were culled and the kidney dissected for WKM cell counts, FACS and qPCR analysis. The results were then compared to those obtained from non-injected control fish from the same clutch.

The number of cells in the different blood lineages was not affected in the mutants upon injection (Figure 5.8B) and neither was overall WKM cellularity (Figure 5.8C), whereas WT fish displayed increased numbers in all lineages (Figure 5.8B and C). Cell viability, however, was lowered in both WT and mutant fish upon acetaldehyde injection (Figure 5.8D). The levels of *p53* transcription were unaffected (Figure 5.8E). Overall, these data

suggest that prolonged acetaldehyde-mediated stress does not affect mutants more than their WT siblings, at least where haematological cells are concerned.

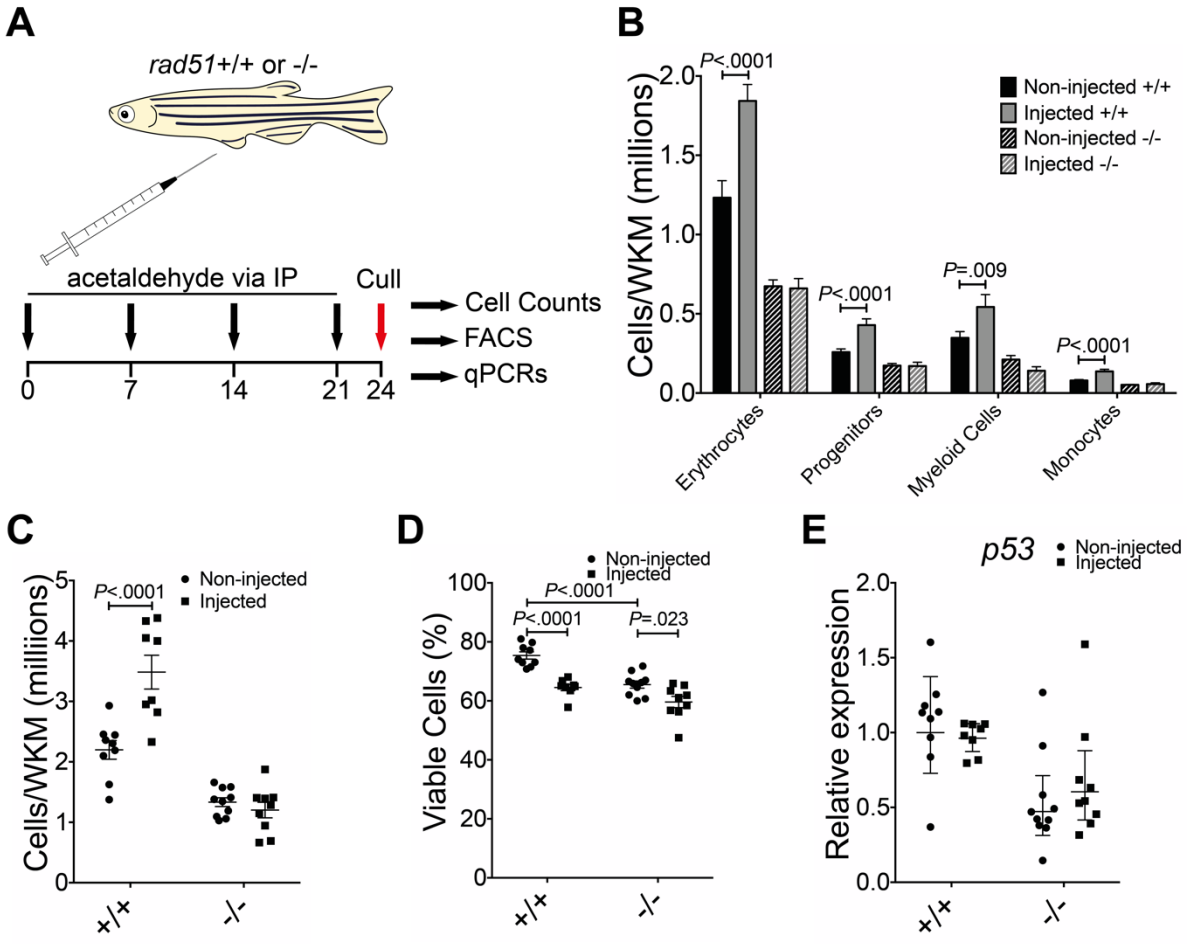


Figure 5.8: Mutation in *rad51* does not lead to acetaldehyde sensitivity. (A) Schematic of the experimental design. Both wild type and *rad51*$^{-/-}$ fish were injected every seven days with 10 µl 1% acetaldehyde, four injections in total. All fish were culled 3 days after the last injection. Control fish were not injected. (B) Absolute number of cells belonging to different blood lineages in the kidney gained by combining FACS data with the cell counts shown in A. Statistical tests were carried out individually for each cell type, using two-way ANOVA. Erythrocytes: There was a statistically significant influence of injection status ($F(1,32) = 13.68$), $P = .0008$), mutation status ($F(1,32) = 116.2$), $P < .0001$) and an interaction between the two ($F(1,32) = 14.91$), $P = .0005$). Progenitors: There was a statistically significant influence of injection status ($F(1,32) = 11.68$), $P = 0017.$), mutation status ($F(1,32) = 49.46$), $P < 0001.$) and an interaction between the two ($F(1,32) = 12.49$), $P = .0013$). Myeloid cells: There was a statistically significant effect of mutation status ($F(1,32) = 37.4$), $P < .0001$) and of an interaction between mutation and injection status ($F(1,32) = 9$), $P = .0052$), but not of injection status on its own ($F(1,32) = 1.98$), $P = .17$). Monocytes: There was a statistically significant influence of injection status ($F(1,32) = 16.14$), $P = .0003$), mutation status ($F(1,32) = 48.02$), $P < .0001$) and an interaction between the two ($F(1,32) = 11.68$),

P = .0017). *P*-value shown on the graph stems from a *post-hoc* Šidak multiple comparison test, comparing non-injected to injected fish within each genotype. For all groups, n is the same as in A. (C) The total number of cells in the kidney in injected and non-injected fish. Statistical testing was carried out using two-way ANOVA. There was a statistically significant effect of mutation status ($F(1, 32) = 91.2$, $P < .0001$) and of injection status ($F(1, 32) = 12.29$, $P = .0014$), as well as the interaction between the two ($F(1, 32) = 18.49$, $P = .0001$). *P*-value shown on the graph is from a *post-hoc* Tukey multiple comparison test. (D) Viable cells as determined by PI-staining. Two-way ANOVA revealed a significant influence of mutation status ($F(1, 32) = 27.47$, $P < .0001$), as well as injection status ($F(1, 32) = 35.28$, $P < .0001$), but not of an interaction between the two ($F(1, 32) = 3.11$, $P = .087$). *P*-values on the graph stem from a *post-hoc* Tukey multiple comparison test. Bars represent mean ± SEM in B, C and D. (E) Relative gene expression of *p53*. Two-way ANOVA revealed a significant influence of mutation status ($F(1, 32) = 16.67$, $P = .0003$), but not of injection status ($F(1, 32) = .49$, $P = .49$) or an interaction between these factors ($F(1, 32) = .92$, $P = .34$). Bars represent geometric mean ± 95% CI to estimate fold changes. For all graphs, *rad51*$^{+/+}$, $n_{NI} = 9$, $n_I = 8$; *rad51*$^{-/-}$, $n_{NI} = 10$, $n_I = 9$.

5.2.2.3 Alcohol-derived stress in *rad51* mutant embryos

After establishing that aldehyde-derived stress does not play a role in BMF of adult fish, I also considered the effect of similar stress on embryonic development. This was done as the main impairments of FA have been proposed to happen during early development[133] and the major works studying aldehyde-induced stress utilised genomic deletion of aldehyde dehydrogenase, leading to early embryonic death[13–15,131]. To estimate the effect of aldehyde-induced stress on embryonic *rad51* mutants, I treated incrosses of *rad51*$^{+/-}$ parents with varying concentrations of alcohol. EtOH rather than acetaldehyde itself was chosen, as it is metabolised to acetaldehyde and not as volatile as that chemical, making the treatment of embryos safer and concentration in solution more consistent. Embryos were transiently treated with EtOH and left to recover until 4 dpf. At this point, pictures were taken, the embryos genotyped and blindly scored. I found that EtOH treatment decreases the overall length of the embryos slightly (but statistically significantly). However, this effect was not influenced by genotype (Figure 5.9). This result indicates that while aldehyde mediated damage is harmful to the developing embryo, it does not affect *rad51* mutants more than their WT siblings.

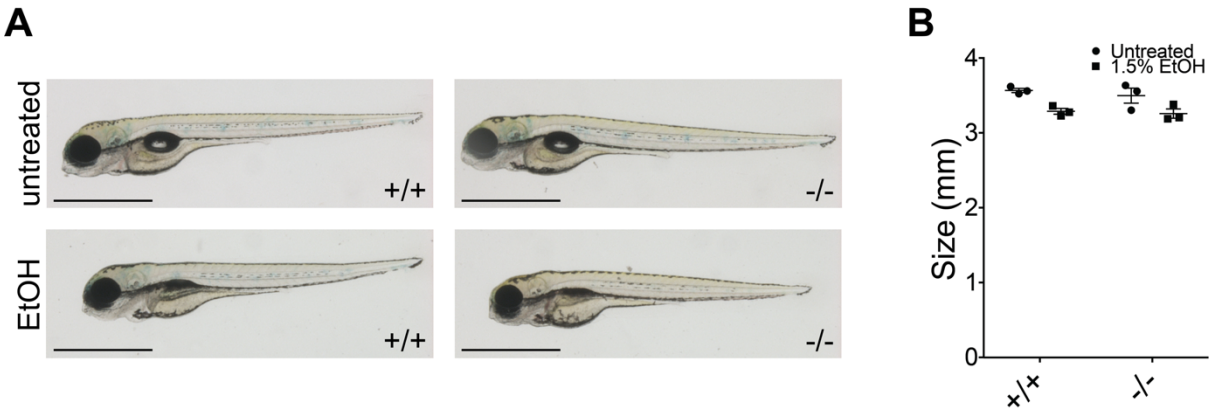


Figure 5.9: Ethanol does not affect *rad51* mutant embryos more strongly than their WT siblings. (A) Example images of untreated and EtOH treated WT and mutant embryos taken at 4 dpf. EtOH was added at 1.5% between 4 and 24 hpf. (B) Size measurements of the embryos. The data stems from three separate clutches, which were divided into a treatment and a non-treatment group. Each clutch consisted of 50 embryos in total. The points plotted represent the mean of that genotype in that condition. Two-tailed ANOVA showed a statistically significant effect of EtOH treatment ($F(1, 8) = 16.75$, $P = .004$), but not of mutation status ($F(1, 8) = .62$, $P = .46$) or an interaction between these factors ($F(1, 8) = .093$, $P = .77$). Bars represent mean ± SEM. Magnification = 16X. Scale bars = 1 mm.

5.3 Discussion

In this section of the thesis, I focussed on the effect of inflammation- and aldehyde-induced stress on *rad51* mutant fish, to explore the role of these factors in BMF in FA patients. I first established an injection scheme for prolonged inflammation, which I then applied to my mutant model. Subsequently, I adapted this injection schedule for acetaldehyde to study the role of aldehydes in FA pathogenesis.

My results showed that inflammation can further exacerbate the haematological phenotype of *rad51* mutants, driven by p53-mediated signalling. This fits with previous publications, which showed that repeated pI:pC injections can cause BMF in FA mouse models[65,134]. In line with the published literature, WT fish were able to respond to inflammation by increasing the production of monocytes/macrophages, which was stimulated by pI:pC and is associated with chronic inflammation[528]. When I challenged fish with a single pI:pC injection, I saw an upregulation of markers of acute inflammation (*il1b* and *il8*). In contrast, prolonged inflammation led to a decrease in these marker genes, coupled to an upregulation of monocyte specific genes (*csfr1* and *marco*), fitting with the increase in monocytes. These findings are consistent with chronic inflammation, in which the inflammatory response shifts from granulocytes to mononuclear cells such as monocytes[529–531]. Intriguingly, the mutant fish were unable to respond appropriately to inflammation and instead displayed a decline in WKM cellularity, resembling the early stages of BMF. This strengthens the case for the hypothesis that FA HSCs are more sensitive to inflammatory stress[10,65,134,168,507–509,511–513] and is further proof for the suitability of *rad51* mutant fish for the study of FA pathogenesis.

Importantly, neither did mutant fish express more inflammatory cytokines than their WT siblings when unstimulated, nor did they show a different response upon the induction of inflammation. This shows that differences in the expression of inflammatory cytokines are not the deciding factor in the blood abnormalities in FA, going against many previously published studies[63,170,171,173,174,515–517]. However, my data fits well with more recent clinical evidence that saw no difference in the amount of inflammatory cytokines produced by FA patient cells when unchallenged. Similar to

rad51 mutant fish, patient-derived cells increased inflammatory cytokine production upon challenge with LPS, but to the same degree as healthy cells[518]. Overall, my model reflects this new clinical data well and provides further evidence that it is cytokine hypersensitivity, rather than overexpression that is problematic in FA.

When I checked for other potential reasons for the decrease in WKM cellularity I saw an upregulation of *p53* in response to inflammation exclusively in the mutant fish, even though they downregulate *p53* expression when unchallenged. This data provides yet more evidence for the hypothesis that p53 is the major driver in HSC attrition in FA patients[133].

One apparent difference between my inflammation model and the published literature, was a lack of increased proliferation in the *rad51* mutants upon pI:pC treatment, whereas HSC cycling is increased in mouse models of inflammation in WT and FA mice[65,134,501,503,505,506,532]. The most important factor contributing to this difference is that I did not enrich for HSCs when looking at the proliferation rate, in contrast to the publications using murine models. It is likely that more mature blood progenitors and HSCs respond differently to inflammatory stress, especially since the proliferation rate is already increased in the mutant. Judging from my data, additional signals to proliferate trigger a p53 response leading to apoptosis and exhaustion of proliferative progenitors, which is reflected in the decreased WKM cell number. This overall decrease in cell numbers, as well as the inability to respond appropriately to chronic inflammation by making more monocytes suggests a HSC defect.

There are, however, a few limitations to my inflammation model. The effect of inflammation on the PB is hard to ascertain using this method, as blood smears were not quantitative enough to yield conclusive data on the changes in the PB. In addition, while I improved the original method of estimating blood lineages in the kidney[376] using single-cell RNA-Seq data[377], the method is still far from exact and can only be viewed as a rough approximation of blood lineages in the WKM. In contrast to previous studies[65,134], I was unable to elicit complete BMF during the duration of my experiment. Further pI:pC injections might have led to complete ablation of the WKM, but due to

animal welfare concerns and the limits of the Home Office licence, I was unable to investigate this any further.

My model might also have applications in the study of inflammatory signalling in general. Current inflammation models using adult fish are mostly based on infecting the fish with various microorganisms[533]. The pI:pC injection model offers the advantage of decreased risk to staff and other fish in the facility as well as being more well-defined in terms of the inducing agent.

I went on to explore another major hypothesis explaining BMF in FA, the accumulation of excess DNA damage stemming from small aldehydes. My results suggest that *rad51* mutants are not excessively sensitive to aldehyde-induced stress.

Initial experiments with very high doses of acetaldehyde showed rapid mortality in *rad51* mutants. However, this is unlikely to be blood-related due to the fast death of the fish. A more likely cause of death is massive necrosis around the site of injection, consistent with the bleeding and reddening at the injection site. This was possibly as the high dose of acetaldehyde overwhelmed the capacity of metabolising enzymes.

Next, I modelled prolonged aldehyde-derived stress in adult fish using repeated injections with a lowered dose of acetaldehyde. This experiment showed no difference in the response between WT and mutant fish, in contrast to previous research proposing aldehydes as the main culprit behind BMF in FA. Previous *in-vivo* studies looking at the role of aldehydes employed double knockout mice in FA genes and aldehyde scavenging enzymes[13–15,131] saw striking defects starting during early embryonic development, especially in *aldh2*$^{-/-}$ mothers. However, my experiment is not directly comparable to the ones carried out on mice. Importantly, my fish were capable of metabolising harmful aldehydes (they did not carry *aldh* mutations) and my experiment was done solely on adult fish, whereas the mice died *in utero* or shortly after birth.

To exclude the possibility of differential aldehyde sensitivity during embryonic development and adulthood, I treated *rad51* mutant embryos with EtOH. My results did not show a difference between wild types and mutants. From this outcome, it can be concluded that on the organismal level, FA patients are not more susceptible to

aldehydes. Due to the large number of enzymes that metabolise aldehydes[145], it is a reasonable assumption that most, if not all, aldehydes are removed before they can cause damage to the genome. If that is the case, the results obtained from *Aldh2* and *Adh5* knockout mice[13–15,131], while correct, are not necessarily relevant to FA patients, apart from rare cases involving patients carrying *ALDH2* polymorphisms[148,149]. These results are also compatible with *in-vitro* studies showing higher susceptibility of FA cells to aldehyde[12,14], but on an organismal level the aldehydes would be removed before causing relevant damage. In conclusion, although aldehyde-derived DNA damage cannot be removed efficiently in FA cells, the burden of aldehyde-induced damage *in-vivo* does not appear to contribute to BMF in a major way. More studies will be needed to determine how much DNA damage by aldehydes contributes to the FA phenotypes in "normal" patients.

The main molecular function of the FA pathway is to remove ICLs, as demonstrated by the striking effects of crosslinking agents on FA chromosomes. However, mice with FA mutations fail to develop BMF[10,130,387,388], unless crosslinking agents are artificially added, the protection against metabolic crosslinkers is removed[13–15,131] or the HSCs are forced to divide[65,134]. This poses the question of why division and excess DNA damage lead to the same outcome. Walter *et al.* recently proposed a model in which HSC exit from quiescence induces DNA damage[134] and causes the death of the cells. It has also been shown that quiescent HSCs accumulate DNA damage that is normally repaired upon re-entry of the cell cycle[471], probably because quiescent HSCs use more error-prone repair mechanisms, such as NHEJ, rendering them more prone to mutagenesis[465]. Other groups have shown that increased HSC cycling leads to excess oxidative stress stemming from the mitochondria[134] and that replication stress contributes to functional impairments in HSCs in general[408]. It is probable that a lot of this damage cannot be repaired in FA cells, leading to senescence or apoptosis. Together, this evidence suggests that excess division of FA HSCs in response to haematopoietic stress triggers BMF due to unrepairable DNA damage in their genome. This damage might be the direct result of proliferation, or be accumulated during quiescence.

Overall, this chapter confirmed the role of inflammatory stress in BMF. In contrast, my data does not indicate a role for small aldehydes in BMF aetiology in FA patients.

6 Discussion

6.1 Insights into the aetiology of FA

6.1.1 A model for BMF in FA

In chapter 3, I outlined the generation and characterisation of the first zebrafish FA model, utilising *rad51* knockout fish. I then exploited this new model system to investigate the cause of BMF in FA patients in chapters 4 and 5. I found that HSC numbers are already decreased early in life and that additional haematopoietic stress caused by inflammation can lead to further decreases in blood cell numbers. In combination with the data discussed above, I propose the following model for FA pathogenesis (Figure 6.1): During embryonic and foetal development, the HSCs in the growing organism accumulate DNA damage due to their inability to undergo efficient ICL repair[465]. Previous research has implicated small aldehydes in this process[12–14,131,146,147], but my own data does not support this notion. When HSCs leave quiescence during periods of high proliferation, as is normal during early development, the p53 signalling pathway is triggered due to the previously sustained DNA damage[133,534]. It is likely that proliferation itself contributes to the increased DNA damage as well[134,506]. This induces both apoptosis, as well as cell cycle arrest of the HSCs, leading to a lowered number of HSCs at birth. Additional haematopoietic stress, such as infections or treatment with growth factors[421,501–504], trigger further HSC division. Intrinsic overexpression of inflammatory cytokines does not seem to play a role in this process, but rather increased sensitivity of the FA HSCs to such stimuli. Eventually, all functional HSCs are exhausted, leading to BMF.

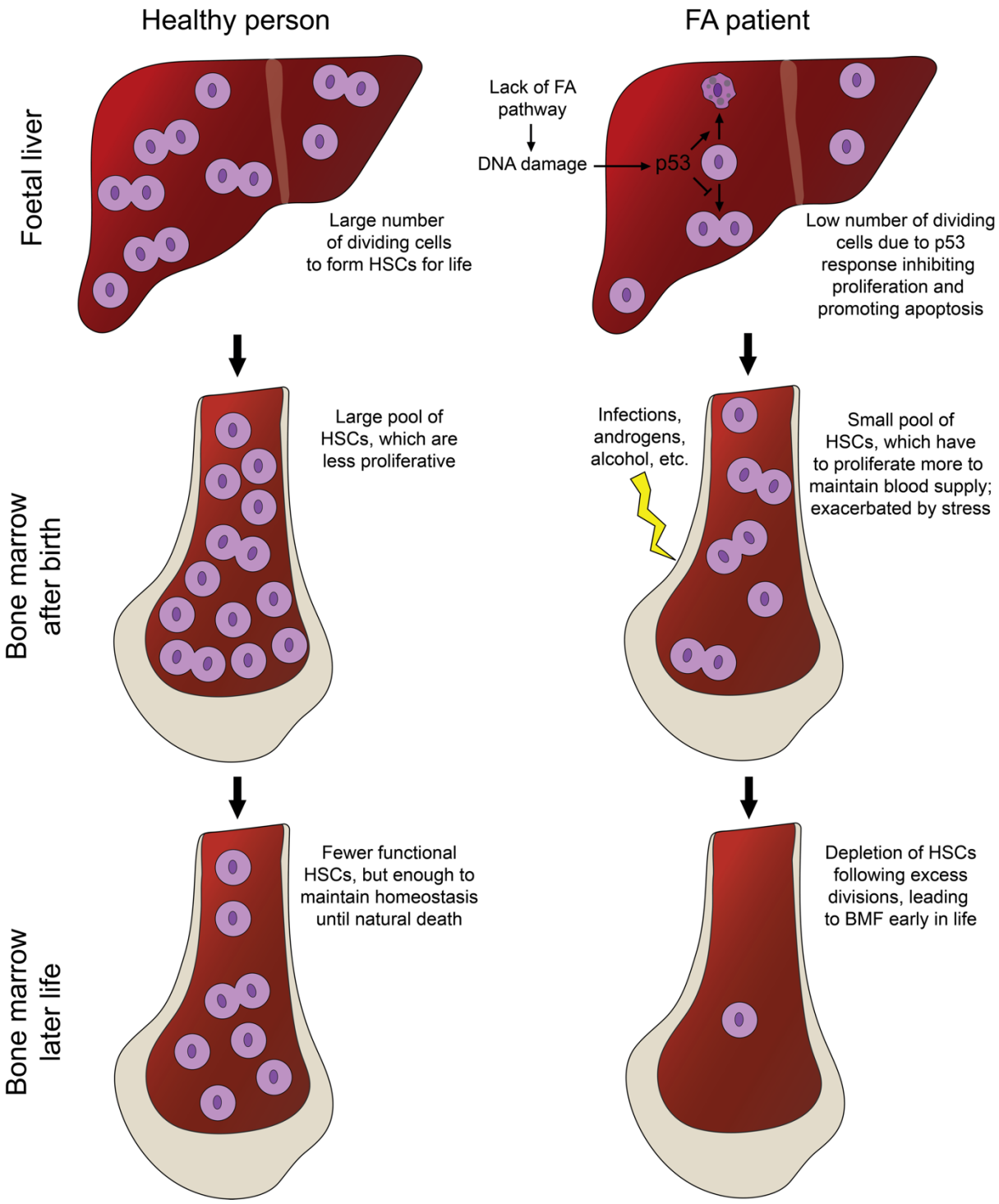


Figure 6.1: Proposed model of FA pathogenesis. In a healthy person (left), a large number of HSCs is formed during embryonic and foetal development in the liver. The number of functional HSCs gradually decreases during adult life, but is normally sufficient to supply enough blood cells to maintain homeostasis, except in a disease context. Conversely in FA patients (right), DNA damage during embryonic and foetal development triggers an excess p53 response. This has two potential consequences: HSCs stop cycling (which seems to be the main mechanism in FA) or they can undergo apoptosis. Because of this, people affected by FA start out with a lowered number of HSCs right after birth, which must divide excessively to maintain a normal number of mature blood cells. This leads to their premature exhaustion,

triggering BMF. Various factors driving HSC division, such as infections, may contribute to the early BMF seen in FA sufferers.

6.1.2 Implications for FA patients

This model has important implications for the development of novel therapeutics to treat FA. The early loss of HSCs during foetal development means that treatment aimed at conserving the patient's own haematopoietic system would have to be started very early to protect the HSC pool – ideally already during gestation. There is clearly a risk of developmental defects when giving drugs so early, but if treatment was started too late, the HSC pool may already be too small to be maintained regardless of treatment. The model also suggests that the p53 pathway might be an interesting target for therapeutics. Inhibitors for p53 are readily available and have been used with success in mouse models without increased tumour development[475,535], including an application to increase HSC numbers[475]. As the potential for side effects when inhibiting the whole p53 pathway, especially in patients already at high risk of malignancies, is considerable, downstream targets of p53 might be preferable.

Another important point is that all drugs that induce HSC cycling, such as growth factors and androgens, should be avoided if possible. Although data derived from animal models should not be used to change treatment regimens on its own, there is considerable clinical evidence highlighting the disadvantages of using such drugs. Importantly, some studies have cautioned against the use of androgens unless unavoidable, as they have been associated with increased liver cancer rates[56,58], lower BMT success following androgen treatment[60] and most severely earlier BMF[59]. There is also evidence for severe side effects caused by growth factors. Long-term administration of G-CSF has been associated with BM hypoplasia[53] and an increased risk of haematological malignancies[54,55]. Furthermore, G-CSF and thrombopoietin had similar efficacy in eliciting BMF compared to pI:pC-induced inflammation[134] in FA mice. This issue highlights the urgent need for novel FA therapeutics.

In addition to growth factors and androgens, other unnecessary activation of HSCs should be avoided in FA patients. However, as even things such as psychosocial stress[536]

or inflammation as it occurs during infections[65,134,505,512,513,534,537] can trigger HSC activation, this is an almost impossible task.

6.1.3 Limitations and open questions

There are, however, some limitations to the proposed model. It adequately describes why FA patients experience almost universal BMF, but several gaps in our understanding of the disease aetiology remain. Importantly, my model does not explain some of the congenital features of the disease, such as decreased stature, which were not rescued upon *p53* co-mutation in my zebrafish model. It also does not address the differences between FA and healthy cells in terms of autophagy, telomere maintenance and other cellular pathways described in earlier chapters. It remains to be elucidated to what extent, if at all, these abnormalities contribute to BMF in FA. They are likely implicated in some of the congenital features of FA that I found to be p53-independent. In general, the connection between BMF and congenital abnormalities has scarcely been explored. Increased data on these features will therefore be vital to develop a model that explains the whole disease phenotype rather than just the haematopoietic features.

It is also not certain what the main ICL-inducing agent is and to what extent ICLs are necessary to trigger BMF. Genetic knockout models implicate small aldehydes[12–14,131,146,147], but my own work discussed in chapter 5 casts doubt on these earlier findings. If the main cause of ICLs was better understood, it might be possible to develop drugs specifically lowering the ICL burden, thereby extending HSC survival without risking tumour development.

Moreover, it is unclear which downstream signalling from p53 is responsible for mediating HSC senescence and apoptosis. Some papers have implicated p21[133], but this is disputed by others[138]. My own data was unable to rule out p21-independent causes, but the decreased proliferation of HSCs suggests at least some involvement. Understanding the downstream signalling from p53 in FA patients will be vital for the creation of effective drugs to combat BMF, as these would likely be better drug targets than p53 itself.

Finally, it is unclear what other factors, such as the genetic background, diet and others, determine the timing of BMF. A better understanding of these factors would lead to improvements in the ability to predict the course of the disease and possibly enable lifestyle changes, leading to a better quality of life and longer survival.

Apart from the inherent limitations of the model, there is always the question of how well data derived from model organisms will ultimately translate to human patients. In general, FA mouse models replicate the features of the disease poorly and few show any blood abnormalities. Cellular phenotypes (i.e. abnormal signalling, etc.), conversely, are reasonably similar[358,387]. Just like the FA mice do not develop BMF, my fish do not develop kidney marrow failure when unchallenged. Nevertheless, they do show several haematological phenotypes, making them closer to the human disease than most mouse models. However, studies confirming my findings in patient-derived samples would be appropriate wherever practicable.

Nevertheless, there is strong evidence that data from zebrafish should be very applicable to humans. In general, haematopoiesis is remarkably conserved between zebrafish and humans, reaching up to 90% conservation of gene expression in HSPCs[377]. As HSCs are the cell type that is ultimately affected in FA, this indicates that cell-based results obtained in zebrafish should be highly analogous to the situation in humans. Moreover, the Fanconi genes are almost all conserved[358], highlighting the similarity in terms of ICL repair between fish and humans. Taking both of these findings together, there is a high degree of conservation in both the cellular, as well of the molecular aspects of FA, making the model appropriate.

6.1.4 Future research on my FA model

I believe my *rad51* knockout FA model can be a valuable asset to the FA field. They are particularly suitable for investigating the relation between embryonic and adult phenotypes, as I demonstrated in the preceding chapters.

In addition to such studies, zebrafish are also ideal organisms for high-throughput screening. Their rapid growth, small size and high fecundity makes them very cheap to raise in large numbers[538,539]. Moreover, zebrafish embryos develop externally and as they

are aquatic, drugs for screening can simply be added to the culture medium provided they are sufficiently soluble[540]. Due to their small size, individual embryos can be cultured in 96 well plates for up to a week of age, allowing automation of the screening process[540,541]. Pigmentation can be stopped during the first days of growth[370], facilitating the readout of fluorescent reporters[540,541] due to the optic clarity of the embryos. Apart from these purely technical considerations, zebrafish embryos are also more similar to humans than immortalised cell lines typically used for these kinds of screen[359]. Drugs need to pass the epithelium to enter the embryo, where early precursors of the liver and nephric system metabolise and excrete the drugs respectively. Other tissues, such as blood and nerve cells, exist as well and allow the study of tissue-specific effects[541]. In combination with available lines labelling HSPCs in the embryo (*Tg(cmyb:EGFP)*[542] and *Tg*(*runx1+23:mCherry*)[338]), the *rad51* mutant line could be straightforwardly used to screen for molecules that could alleviate the HSC defects during embryonic development. After identifying potential targets, it would be easy to rear drug-treated embryos to adulthood, to see whether embryonic changes directly translate to benefits in adult animals, a great advantage to cell-based high-throughput screens.

6.2 Looking beyond FA

Although the main focus of my work was on characterising the molecular basis of FA, the *rad51* mutation model also allows insight into a range of other molecular processes, including the role of Rad51, and zebrafish sex determination. This model also demonstrates some of the advantages of using zebrafish for research purposes in general. In this case, their viability without *rad51* was vital to carry out my study. Moreover, the ease of investigating embryonic development enabled me to directly visualise the developing haematopoietic system, which would have been very difficult, if not impossible in other organisms.

6.2.1 A novel system to study Rad51

As mouse mutants proved lethal when *Rad51* was lost[296,297], zebrafish are the only known vertebrates that can survive without *rad51*. The survival of these zebrafish, while puzzling, might yield insights into the differences in HR between vertebrates. In this document, I considered the functional overlap between *rad51* and its paralogues in the

survival of my fish, as well as the mechanism of PARP1 inhibition. Other potential avenues of investigation include the discovery of non-HR pathways involving Rad51 in vertebrate cells, as any pathway involving this protein should be dysfunctional. For example, RAD51 has been linked to replication fork protection in ICL repair, before it functions in HR[114]. It would be intriguing if the protein turns out to be involved in other pathways as well.

This study also highlighted the role of Rad51 as a tumour suppressor. The HR pathway is defective in many different types of cancer, including bowel, prostate, breast and ovarian cancer[543]. My model could be useful in further understanding the pathology of such malignancies, as well as for the discovery of novel drugs targeting these cancers.

6.2.2 Zebrafish sex determination and meiosis

My research on the *rad51* mutant line corroborated findings from previous studies, showing that successful ICL repair is necessary for the development of female fish[363,364]. Increased p53-driven apoptosis causes the death of immature oocytes, leading to virilisation of the fish[363,364]. In agreement with the other studies, I also saw a rescue of the sex reversal phenotype upon *p53* co-mutation. It is possible that other interventions preventing the apoptosis of germ cells would lead to similar results. Interestingly, meiosis only depends on HR, but not ICL repair[362]. Meiotic defects have been described for *Brca2*[544] and *Fanca*[410] knockout mice, underlining the applicability of these results to mammalian systems. The *rad51* mutant line is therefore a suitable model to study the role of HR and ICL repair in developing germ cells.

6.3 Conclusion

In this thesis, I outlined the generation of the first viable vertebrate *rad51* mutation model. During the characterisation of my model, I showed that these fish develop many symptoms resembling the haematological disease Fanconi anaemia. I went on to utilise this system to increase our understanding of the underlying molecular biology of this disease. My results implicate a hyperactive p53 response during embryonic development as the main cause of the illness, exacerbated by haematopoietic stress that induces HSC cycling as the cause of BMF. Based on this evidence, I was able to propose a refined

model of how mutations in the FA pathway induce BMF in patients. In doing so, I contributed to the literature by developing a useful new tool for studying the cellular and molecular pathology of FA. In addition to my findings on signalling pathways involved in FA pathogenesis, I was also able to provide further evidence for the role of inflammation in triggering BMF. Further research will be needed to make full use of this model and apply it to the discovery of new treatment options. Apart from the study of FA, there are several other possible uses of this fish line, including the examination of the mechanisms underpinning homologous recombination, as well as meiosis.

References

1. Fanconi, G. Familiäre infantile perniziosaartige Anämie (perniziöses Blutbild und Konstitution). *Jahrb. für Kinderheilkd. und Phys. Erziehung* **117,** 257–280 (1927).

2. Uehlinger, E. Konstitutionelle infantile (perniciosaartige) Anämie. *Klin.Wochenschr.* **8,** 1501–1503 (1929).

3. Fanconi, G. Hypothesis of Chromosomal Translocation as a Genetic Interpretation of Fanconi's Familial Constitutional Panmyelopathy. *Helv.Paediatr.Acta* **19,** 29–33 (1964).

4. Meetei, A. R. *et al.* X-linked inheritance of Fanconi anemia complementation group B. *Nat.Genet.* **36,** 1219–1224 (2004).

5. Auerbach, A. D. & Wolman, S. R. Susceptibility of Fanconi's anaemia fibroblasts to chromosome damage by carcinogens. *Nature* **261,** 494–496 (1976).

6. Sasaki, M. S. & Tonomura, A. A high susceptibility of Fanconi's anemia to chromosome breakage by DNA cross-linking agents. *Cancer Res.* **33,** 1829–1836 (1973).

7. Strathdee, C. A., Gavish, H., Shannon, W. R. & Buchwald, M. Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **356,** 763–767 (1992).

8. Pace, P. *et al.* FANCE: the link between Fanconi anaemia complex assembly and activity. *EMBO J.* **21,** 3414–3423 (2002).

9. Grompe, M. & D'Andrea, A. Fanconi anemia and DNA repair. *Hum.Mol.Genet.* **10,** 2253–2259 (2001).

10. Haneline, L. S. *et al.* Multiple inhibitory cytokines induce deregulated progenitor growth and apoptosis in hematopoietic cells from Fac-/- mice. *Blood* **91,** 4092–4098 (1998).

11. Dufour, C. *et al.* TNF-alpha and IFN-gamma are overexpressed in the bone marrow of Fanconi anemia patients and TNF-alpha suppresses erythropoiesis in vitro. *Blood* **102,** 2053–2059 (2003).

12. Rosado, I. V, Langevin, F., Crossan, G. P., Takata, M. & Patel, K. J. Formaldehyde catabolism is essential in cells deficient for the Fanconi anemia DNA-repair pathway. *Nat. Struct. Mol. Biol.* **18,** 1432–1434 (2011).

13. Oberbeck, N. *et al.* Maternal aldehyde elimination during pregnancy preserves the fetal genome. *Mol.Cell* **55,** 807–817 (2014).

References

14. Langevin, F., Crossan, G. P., Rosado, I. V, Arends, M. J. & Patel, K. J. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* **475,** 53–58 (2011).

15. Garaycoechea, J. I. *et al.* Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* **489,** 571–575 (2012).

16. Sumpter, R. J. *et al.* Fanconi Anemia Proteins Function in Mitophagy and Immunity. *Cell* **165,** 867–881 (2016).

17. Shyamsunder, P. *et al.* Impaired mitophagy in Fanconi anemia is dependent on mitochondrial fission. *Oncotarget* **7,** 58065–58074 (2016).

18. Calado, R. T. & Young, N. S. Telomere maintenance and human bone marrow failure. *Blood* **111,** 4446–4455 (2008).

19. Shimamura, A. & Alter, B. P. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood Rev.* **24,** 101–122 (2010).

20. Green, A. M. & Kupfer, G. M. Fanconi anemia. *Hematol.Oncol.Clin.North Am.* **23,** 193–214 (2009).

21. Alter, B. P. Diagnosis, genetics, and management of inherited bone marrow failure syndromes. *Hematol. Am.Soc.Hematol.Educ.Program.* 29–39 (2007).

22. Neveling, K., Endt, D., Hoehn, H. & Schindler, D. Genotype–phenotype correlations in Fanconi anemia. *Mutat. Res./Fundam. Mol. Mech. Mutag.* **668,** 73–91 (2009).

23. Rose, S. R. *et al.* Endocrine phenotype of children and adults with Fanconi anemia. *Pediatr. Blood Cancer* **59,** 690–696 (2012).

24. Giri, N., Batista, D. L., Alter, B. P. & Stratakis, C. A. Endocrine abnormalities in patients with Fanconi anemia. *J. Clin. Endocrinol. Metab.* **92,** 2624–2631 (2007).

25. Alter, B. P. *et al.* Fanconi's anaemia and pregnancy. *Br.J.Haematol.* **77,** 410–418 (1991).

26. Dalle, J. H. *et al.* Successful pregnancies after bone marrow transplantation for Fanconi anemia. *Bone Marrow Transpl.* **34,** 1099–1100 (2004).

27. Santos, F., Selesnick, S. H. & Glasgold, R. A. Otologic manifestations of Fanconi anemia. *Otol. Neurotol.* **23,** 873–875 (2002).

28. Jose Vale, M. *et al.* Audiologic abnormalities of Fanconi anaemia. *Acta Otolaryngol.* **128,** 992–996 (2008).

29. Kutler, D. I. *et al.* A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood* **101,** 1249–1256 (2003).

30. Butturini, A. *et al.* Hematologic abnormalities in Fanconi anemia: an International Fanconi Anemia Registry study. *Blood* **84,** 1650–1655 (1994).

31. Landmann, E., Bluetters-Sawatzki, R., Schindler, D. & Gortner, L. Fanconi anemia in a neonate with pancytopenia. *J.Pediatr.* **145,** 125–127 (2004).

32. *Fanconi Anemia: Guidelines for Diagnosis and Management*. (Fanconi Anemia Research Fund, Inc., 2014).

33. Rosenberg, P. S., Greene, M. H. & Alter, B. P. Cancer incidence in persons with Fanconi anemia. *Blood* **101,** 822–826 (2003).

34. Tamary, H. *et al.* Frequency and natural history of inherited bone marrow failure syndromes: the Israeli Inherited Bone Marrow Failure Registry. *Haematologica* **95,** 1300–1307 (2010).

35. Alter, B. P. *et al.* Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. *Br.J.Haematol.* **150,** 179–188 (2010).

36. Alter, B. P. Cancer in Fanconi anemia, 1927–2001. *Cancer* **97,** 425–440 (2003).

37. Alter, B. P. Fanconi's anemia and malignancies. *Am.J.Hematol.* **53,** 99–110 (1996).

38. Rosenberg, P. S., Alter, B. P. & Ebell, W. Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. *Haematologica* **93,** 511–517 (2008).

39. Zur Hausen, H. Papillomavirus infections—a major cause of human cancers. *Biochim. Biophys. Acta (BBA)-Reviews Cancer* **1288,** F55–F78 (1996).

40. van Zeeburg, H. J. *et al.* Clinical and molecular characteristics of squamous cell carcinomas from Fanconi anemia patients. *J.Natl.Cancer Inst.* **100,** 1649–1653 (2008).

41. Sauter, S. L. *et al.* Oral human papillomavirus is common in individuals with Fanconi anemia. *Cancer Epidemiol.Biomarkers Prev.* **24,** 864–872 (2015).

42. D'Andrea, A. D. Susceptibility pathways in Fanconi's anemia and breast cancer. *N.Engl.J.Med.* **362,** 1909–1919 (2010).

43. Swift, M. Fanconi's anaemia in the genetics of neoplasia. *Nature* **230,** 370–373 (1971).

44. Rosenberg, P. S., Tamary, H. & Alter, B. P. How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel. *Am. J. Med. Genet. A* **155,** 1877–1883 (2011).

45. Verlander, P. C. *et al.* Carrier frequency of the IVS4 + 4 A-->T mutation of the Fanconi anemia gene *FAC* in the Ashkenazi Jewish population. *Blood* **86,** 4034–4038 (1995).

46. Kutler, D. I. & Auerbach, A. D. Fanconi anemia in Ashkenazi Jews. *Fam. Cancer* **3,** 241–248 (2004).

47. Morgan, N. V *et al.* A common Fanconi anemia mutation in black populations of sub-Saharan Africa. *Blood* **105,** 3542–3544 (2005).

48. Rosendorff, J. *et al.* Fanconi anemia: another disease of unusually high prevalence in the Afrikaans population of South Africa. *Am.J.Med.Genet.* **27,** 793–797 (1987).

49. Callen, E. *et al.* A common founder mutation in FANCA underlies the world's highest prevalence of Fanconi anemia in Gypsy families from Spain. *Blood* **105,** 1946–1949 (2005).

50. Stevens, H., Chyn Chua, C., Wallis, M., Hew, S. & Grigg, A. Fanconi anemia in 55-year-old identical twins first presenting as fatal post-chemotherapy pancytopenia. *Am.J.Hematol.* **91,** 1273–1276 (2016).

51. *Standards of care for fanconi anaemia affected individuals and their families*. (Fanconi Hope Charitable Trust, 2009).

52. Longerich, S., Li, J., Xiong, Y., Sung, P. & Kupfer, G. M. Stress and DNA repair biology of the Fanconi anemia pathway. *Blood* **124,** 2812–2819 (2014).

53. Shukla, P., Ghosh, K. & Vundinti, B. R. Current and emerging therapeutic strategies for Fanconi anemia. *Hugo J.* **6,** 1 (2012).

54. MacMillan, M. L. & Wagner, J. E. Haematopoeitic cell transplantation for Fanconi anaemia–when and how? *Br.J.Haematol.* **149,** 14–21 (2010).

55. Myers, K. C. & Davies, S. M. Hematopoietic stem cell transplantation for bone marrow failure syndromes in children. *Biol. Blood Marrow Transplant.* **15,** 279–292 (2009).

56. Velazquez, I. & Alter, B. P. Androgens and liver tumors: Fanconi's anemia and non-Fanconi's conditions. *Am. J. Hematol.* **77,** 257–267 (2004).

57. Scheckenbach, K. *et al.* Treatment of the bone marrow failure in Fanconi anemia patients with danazol. *Blood Cells Mol. Dis.* **48,** 128–131 (2012).

58. Paustian, L. *et al.* Androgen therapy in Fanconi anemia: A retrospective analysis of 30 years in Germany. *Pediatr. Hematol. Oncol.* **33,** 5–12 (2016).

59. Zhang, Q.-S. *et al.* Oxymetholone therapy of fanconi anemia suppresses osteopontin transcription and induces hematopoietic stem cell cycling. *Stem cell reports* **4,** 90–102 (2015).

60. Dufour, C. & Svahn, J. Fanconi anaemia: new strategies. *Bone Marrow Transplant.* **41,** S90–S95 (2008).

61. Zhang, Q.-S. *et al.* Tempol protects against oxidative damage and delays epithelial tumor onset in Fanconi anemia mice. *Cancer Res.* **68,** 1601–1608 (2008).

62. Zhang, Q.-S. *et al. Fancd2-/-* mice have hematopoietic defects that can be partially corrected by resveratrol. *Blood* **116,** 5140–5148 (2010).

63. Anur, P. *et al.* p38 MAPK inhibition suppresses the TLR-hypersensitive phenotype in FANCC- and FANCA-deficient mononuclear phagocytes. *Blood* **119,** 1992–2002 (2012).

64. Svahn, J. *et al.* p38 Mitogen-activated protein kinase inhibition enhances in vitro erythropoiesis of Fanconi anemia, complementation group A-deficient bone marrow cells. *Exp. Hematol.* **43,** 295–299 (2015).

65. Zhang, H. *et al.* TGF-β Inhibition Rescues Hematopoietic Stem Cell Defects and Bone Marrow Failure in Fanconi Anemia. *Cell Stem Cell* **18,** 668–681 (2016).

66. Akhurst, R. J. & Hata, A. Targeting the TGFβ signalling pathway in disease. *Nat. Rev. Drug Discov.* **11,** 790–811 (2012).

67. Grafe, I. *et al.* Excessive transforming growth factor-β signaling is a common mechanism in osteogenesis imperfecta. *Nat. Med.* **20,** 670–675 (2014).

68. Gluckman, E. *et al.* Bone marrow transplantation for Fanconi anemia. *Blood* **86,** 2856–2862 (1995).

69. Gluckman, E. *et al.* Bone marrow transplantation in Fanconi anaemia. *Br. J. Haematol.* **45,** 557–64 (1980).

70. Gluckman, E. *et al.* Hematopoietic reconstitution in a patient with Fanconi's anemia by means of umbilical-cord blood from an HLA-identical sibling. *N.Engl.J.Med.* **321,** 1174–1178 (1989).

71. MacMillan, M. L. *et al.* Alternative donor hematopoietic cell transplantation for Fanconi anemia. *Blood* **125,** 3798–3804 (2015).

72. Guardiola, P. *et al.* Outcome of 69 allogeneic stem cell transplantations for Fanconi anemia using HLA-matched unrelated donors: a study on behalf of the European Group for Blood and Marrow Transplantation. *Blood* **95,** 422–429 (2000).

73. Guardiola, P. *et al.* Acute graft-versus-host disease in patients with Fanconi anemia or acquired aplastic anemia undergoing bone marrow transplantation from HLA-identical sibling donors: risk factors and influence on outcome. *Blood* **103,** 73–77 (2004).

74. Smetsers, S. E., Smiers, F. J., Bresters, D., Sonnevelt, M. C. & Bierings, M. B. Four decades of stem cell transplantation for Fanconi anaemia in the Netherlands. *Br. J. Haematol.* **174,** 952–961 (2016).

75. Alter, B. P. Radiosensitivity in Fanconi's anemia patients. *Radiother. Oncol.* **62,** 345–347 (2002).

76. Chao, M. M. *et al.* Outcomes of mismatched and unrelated donor hematopoietic stem cell transplantation in Fanconi anemia conditioned with chemotherapy only. *Ann.Hematol.* **94,** 1311–1318 (2015).

77. MacMillan, M. L., Hughes, M. R., Agarwal, S. & Daley, G. Q. Cellular Therapy for Fanconi Anemia: The Past, Present, and Future. *Biol. Blood Marrow Transplant.* **17,** S109–S114 (2011).

78. Stepensky, P. *et al.* Bone marrow transplantation for Fanconi anemia using fludarabine-based conditioning. *Biol. Blood Marrow Transplant.* **17,** 1282–1288 (2011).

79. Verlinsky, Y., Rechitsky, S., Schoolcraft, W., Strom, C. & Kuliev, A. Preimplantation diagnosis for Fanconi anemia combined with HLA matching. *JAMA* **285,** 3130–3133 (2001).

80. Trujillo, J. P. & Surralles, J. Savior siblings and Fanconi anemia: analysis of success rates from the family's perspective. *Genet. Med.* **17,** 935–938 (2015).

81. Socié, G. *et al.* Transplantation for Fanconi's anaemia: long-term follow-up of fifty patients transplanted from a sibling donor after low-dose cyclophosphamide and thoraco-abdominal irradiation for conditioning. *Br. J. Haematol.* **103,** 249–255 (1998).

82. Mohty, M. & Apperley, J. F. Long-term physiological side effects after allogeneic bone marrow transplantation. *Hematol. Am. Soc. Hematol. Educ. Progr.* **2010,** 229–236 (2010).

83. Ayas, M. *et al.* Allogeneic stem cell transplantation in Fanconi anemia patients presenting with myelodysplasia and/or clonal abnormality: update on the Saudi experience. *Bone Marrow Transplant.* **41,** 261–265 (2008).

84. Mitchell, R. *et al.* Haematopoietic cell transplantation for acute leukaemia and advanced myelodysplastic syndrome in Fanconi anaemia. *Br. J. Haematol.* **164,** 384–395 (2014).

85. Beier, R. *et al.* Minimal antileukaemic treatment followed by reduced-intensity conditioning in three consecutive children with Fanconi anaemia and AML. *Bone Marrow Transplant.* **50,** 463–464 (2015).

86. Ayas, M. *et al.* Allogeneic hematopoietic cell transplantation for fanconi anemia in patients with pretransplantation cytogenetic abnormalities, myelodysplastic syndrome, or acute leukemia. *J. Clin. Oncol.* **31,** 1669–1676 (2013).

87. Khan, N. E., Rosenberg, P. S. & Alter, B. P. Preemptive Bone Marrow Transplantation and Event-Free Survival in Fanconi Anemia. *Biol. Blood Marrow*

Transplant. **22,** 1888–1892 (2016).

88. Kamranzadeh Fumani, H. *et al.* Allogeneic Hematopoietic Stem Cell Transplantation for Adult Patients with Fanconi Anemia. *Mediterr. J. Hematol. Infect. Dis.* **8,** e2016054 (2016).

89. Naldini, L. Ex vivo gene transfer and correction for cell-based therapies. *Nat. Rev. Genet.* **12,** 301–315 (2011).

90. Osborn, M. J. *et al.* Fanconi anemia gene editing by the CRISPR/Cas9 system. *Hum.Gene Ther.* **26,** 114–126 (2014).

91. Liu, J. M. *et al.* Engraftment of hematopoietic progenitor cells transduced with the Fanconi anemia group C gene (FANCC). *Hum.Gene Ther.* **10,** 2337–2346 (1999).

92. Jacome, A. *et al.* Lentiviral-mediated genetic correction of hematopoietic and mesenchymal progenitor cells from Fanconi anemia patients. *Mol. Ther.* **17,** 1083–1092 (2009).

93. González-Murillo, A. *et al.* Development of lentiviral vectors with optimized transcriptional activity for the gene therapy of patients with Fanconi anemia. *Hum. Gene Ther.* **21,** 623–630 (2010).

94. Becker, P. S. *et al.* Preclinical correction of human Fanconi anemia complementation group A bone marrow cells using a safety-modified lentiviral vector. *Gene Ther.* **17,** 1244–1252 (2010).

95. Molina-Estevez, F. J. *et al.* Lentiviral-Mediated Gene Therapy in Fanconi Anemia-A Mice Reveals Long-Term Engraftment and Continuous Turnover of Corrected HSCs. *Curr. Gene Ther.* **15,** 550–562 (2015).

96. Frecha, C. *et al.* A novel lentiviral vector targets gene transfer into human hematopoietic stem cells in marrow from patients with bone marrow failure syndrome and in vivo in humanized mice. *Blood* **119,** 1139–1150 (2012).

97. Chakkaramakkil Verghese, S., Goloviznina, N. A. & Kurre, P. Phenotypic correction of Fanconi anemia cells in the murine bone marrow after carrier cell mediated delivery of lentiviral vector. *Stem Cell Res. Ther.* **7,** 170 (2016).

98. Rio, P. *et al.* Targeted gene therapy and cell reprogramming in Fanconi anemia. *EMBO Mol.Med.* **6,** 835–848 (2014).

99. Muller, L. U. *et al.* Overcoming reprogramming resistance of Fanconi anemia cells. *Blood* **119,** 5449–5457 (2012).

100. Raya, Á. *et al.* Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature* **460,** 53–59 (2009).

101. Kelly, P. F. *et al.* Stem cell collection and gene transfer in Fanconi anemia. *Mol.*

Ther. **15,** 211–219 (2007).

102. Croop, J. M. *et al.* Mobilization and collection of peripheral blood CD34+ cells from patients with Fanconi anemia. *Blood* **98,** 2917–2921 (2001).

103. Kee, Y. & D'Andrea, A. D. Molecular pathogenesis and clinical management of Fanconi anemia. *J. Clin. Invest.* **122,** 3799–806 (2012).

104. Auerbach, A. & Smogorzewska, A. Fanconi Anemia Mutation Database. Available at: http://www2.rockefeller.edu/fanconi/. (Accessed: 6th May 2017)

105. Fokkema, I. F. *et al.* LOVD v. 2.0: the next generation in gene variant databases. *Hum.Mutat.* **32,** 557–563 (2011).

106. Bogliolo, M. & Surrallés, J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Curr.Opin.Genet.Dev.* **33,** 32–40 (2015).

107. Meetei, A. R. *et al.* A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat.Genet.* **35,** 165–170 (2003).

108. Niedernhofer, L. J. *et al.* The Fanconi anemia signalosome anchor. *Mol. Cell* **25,** 487–90 (2007).

109. Singh, T. R. *et al.* Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. *Blood* **114,** 174–180 (2009).

110. Lim, E. T. *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* **10,** e1004494 (2014).

111. Bakker, S. T. *et al. Fancm*-deficient mice reveal unique features of Fanconi anemia complementation group M. *Hum.Mol.Genet.* **18,** 3484–3495 (2009).

112. Joenje, H. *et al.* Complementation analysis in Fanconi anemia: assignment of the reference FA-H patient to group A. *Am. J. Hum. Genet.* **67,** 759–762 (2000).

113. Vasanthakumar, A. *et al. Brca1* deficiency causes bone marrow failure and spontaneous hematologic malignancies in mice. *Blood* **127,** 310–313 (2016).

114. Wang, A. T. *et al.* A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. *Mol.Cell* **59,** 478–490 (2015).

115. Ameziane, N. *et al.* A novel Fanconi anaemia subtype associated with a dominant-negative mutation in *RAD51*. *Nat. Commun.* **6,** 8829 (2015).

116. Gillio, A. P., Verlander, P. C., Batish, S. D., Giampietro, P. F. & Auerbach, A. D. Phenotypic consequences of mutations in the Fanconi anemia *FAC* gene: an International Fanconi Anemia Registry study. *Blood* **90,** 105–110 (1997).

117. Davies, S. M. *et al.* GST genotype may modify clinical phenotype in patients with Fanconi anaemia. *Br. J. Haematol.* **131,** 118–122 (2005).

118. Wagner, J. E. *et al.* Germline mutations in BRCA2: shared genetic susceptibility to breast cancer, early onset leukemia, and Fanconi anemia. *Blood* **103,** 3226–3229 (2004).

119. Alter, B. P. The association between FANCD1/BRCA2 mutations and leukaemia. *Br.J.Haematol.* **133,** 446–448 (2006).

120. Alter, B. P., Rosenberg, P. S. & Brody, L. C. Clinical and molecular features associated with biallelic mutations in *FANCD1/BRCA2*. *J.Med.Genet.* **44,** 1–9 (2007).

121. Reid, S. *et al.* Biallelic mutations in *PALB2* cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat.Genet.* **39,** 162–164 (2007).

122. Meindl, A. *et al.* Germline mutations in breast and ovarian cancer pedigrees establish *RAD51C* as a human cancer susceptibility gene. *Nat. Genet.* **42,** 410–414 (2010).

123. Erkko, H. *et al.* A recurrent mutation in PALB2 in Finnish cancer families. *Nature* **446,** 316–319 (2007).

124. Seal, S. *et al.* Truncating mutations in the Fanconi anemia J gene *BRIP1* are low-penetrance breast cancer susceptibility alleles. *Nat.Genet.* **38,** 1239–1241 (2006).

125. Wooster, R. *et al.* Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* **265,** 2088–2090 (1994).

126. Hall, J. M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250,** 1684–1689 (1990).

127. Brosh, R. M., Bellani, M., Liu, Y. & Seidman, M. M. Fanconi Anemia: A DNA repair disorder characterized by accelerated decline of the hematopoietic stem cell compartment and other features of aging. *Ageing Res. Rev.* **33,** 67–75 (2017).

128. Parmar, K. *et al.* Hematopoietic stem cell defects in mice with deficiency of Fancd2 or Usp1. *Stem Cells* **28,** 1186–1195 (2010).

129. Du, W. *et al. Fancb* deficiency impairs hematopoietic stem cell function. *Sci.Rep.* **5,** 18127 (2015).

130. Carreau, M. *et al.* Bone marrow failure in the Fanconi anemia group C mouse model after DNA damage. *Blood* **91,** 2737–2744 (1998).

131. Pontel, L. B. *et al.* Endogenous formaldehyde is a hematopoietic stem cell genotoxin and metabolic carcinogen. *Mol.Cell* **60,** 177–188 (2015).

132. Haneline, L. S. *et al.* Loss of *FancC* function results in decreased hematopoietic

stem cell repopulating ability. *Blood* **94,** 1–8 (1999).

133. Ceccaldi, R. *et al.* Bone marrow failure in Fanconi anemia is triggered by an exacerbated p53/p21 DNA damage response that impairs hematopoietic stem and progenitor cells. *Cell Stem Cell* **11,** 36–49 (2012).

134. Walter, D. *et al.* Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature* **520,** 549–552 (2015).

135. Tulpule, A. *et al.* Knockdown of Fanconi anemia genes in human embryonic stem cells reveals early developmental defects in the hematopoietic lineage. *Blood* **115,** 3453–3462 (2010).

136. Kamimae-Lanning, A. N., Goloviznina, N. A. & Kurre, P. Fetal origins of hematopoietic failure in a murine model of Fanconi anemia. *Blood* **121,** 2008–2012 (2013).

137. Yoon, Y. me, Storm, K. J., Kamimae-Lanning, A. N., Goloviznina, N. A. & Kurre, P. Endogenous DNA Damage Leads to p53-Independent Deficits in Replicative Fitness in Fetal Murine *Fancd2*–/– Hematopoietic Stem and Progenitor Cells. *Stem Cell Reports* **7,** 840–853 (2016).

138. Zhang, Q.-S. *et al.* Fancd2 and p21 function independently in maintaining the size of hematopoietic stem and progenitor cell pool in mice. *Stem Cell Res.* **11,** 687–692 (2013).

139. Jaber, S., Toufektchan, E., Lejour, V., Bardot, B. & Toledo, F. p53 downregulates the Fanconi anaemia DNA repair pathway. *Nat. Commun.* **7,** 11091 (2016).

140. Niedernhofer, L. J. DNA repair is crucial for maintaining hematopoietic stem cell function. *DNA Repair (Amst).* **7,** 523–529 (2008).

141. Rossi, D. J. *et al.* Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. *Nature* **447,** 725–729 (2007).

142. Zhang, X., Li, J., Sejas, D. P. & Pang, Q. Hypoxia-reoxygenation induces premature senescence in FA bone marrow hematopoietic cells. *Blood* **106,** 75–85 (2005).

143. Dumble, M. *et al.* The impact of altered p53 dosage on hematopoietic stem cell dynamics during aging. *Blood* **109,** 1736–1742 (2007).

144. Nijnik, A. *et al.* DNA repair is limiting for haematopoietic stem cells during ageing. *Nature* **447,** 686–690 (2007).

145. O'Brien, P. J., Siraki, A. G. & Shangari, N. Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. *Crit. Rev. Toxicol.* **35,** 609–62 (2005).

146. Mechilli, M., Schinoppi, A., Kobos, K., Natarajan, A. T. & Palitti, F. DNA repair

deficiency and acetaldehyde-induced chromosomal alterations in CHO cells. *Mutagenesis* **23,** 51–56 (2008).

147. Ridpath, J. R. *et al.* Cells deficient in the FANC/BRCA pathway are hypersensitive to plasma levels of formaldehyde. *Cancer Res.* **67,** 11117–11122 (2007).

148. Hira, A. *et al.* Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. *Blood* **122,** 3206–3209 (2013).

149. Yabe, M. *et al.* The phenotype and clinical course of Japanese Fanconi Anaemia infants is influenced by patient, but not maternal *ALDH2* genotype. *Br. J. Haematol.* **175,** 457–461 (2016).

150. Park, S. J. *et al.* Oxidative stress/damage induces multimerization and interaction of Fanconi anemia proteins. *J.Biol.Chem.* **279,** 30053–30059 (2004).

151. Kruyt, F. A. *et al.* Abnormal microsomal detoxification implicated in Fanconi anemia group C by interaction of the FAC protein with NADPH cytochrome P450 reductase. *Blood* **92,** 3050–3056 (1998).

152. Futaki, M. *et al.* The FANCG Fanconi anemia protein interacts with CYP2E1: possible role in protection against oxidative DNA damage. *Carcinogenesis* **23,** 67–72 (2002).

153. Mukhopadhyay, S. S. *et al.* Defective mitochondrial peroxiredoxin-3 results in sensitivity to oxidative stress in Fanconi anemia. *J.Cell Biol.* **175,** 225–235 (2006).

154. Joenje, H., Arwert, F., Eriksson, A. W., de Koning, H. & Oostra, A. B. Oxygen-dependence of chromosomal aberrations in Fanconi's anaemia. *Nature* **290,** 142–3 (1981).

155. Schindler, D. & Hoehn, H. Fanconi Anemia Mutation Causes Cellular Susceptibility to Ambient Oxygen. *Am. J. Hum. Genet* **43,** 429–43 (1988).

156. Liebetrau, W., Budde, A., Savoia, A., Grummt, F. & Hoehn, H. p53 activates Fanconi anemia group C gene expression. *Hum. Mol. Genet.* **6,** 277–283 (1997).

157. Hadjur, S. *et al.* Defective hematopoiesis and hepatic steatosis in mice with combined deficiencies of the genes encoding Fancc and Cu/Zn superoxide dismutase. *Blood* **98,** 1003–1011 (2001).

158. Kumari, U., Jun, W. Y., Bay, B. H. & Lyakhovich, A. Evidence of mitochondrial dysfunction and impaired ROS detoxifying machinery in Fanconi anemia cells. *Oncogene* **33,** 165–172 (2014).

159. Pagano, G. *et al.* From clinical description, to in vitro and animal studies, and backward to patients: oxidative stress and mitochondrial dysfunction in Fanconi anemia. *Free Radic. Biol. Med.* **58,** 118–125 (2013).

160. Lemasters, J. J. Selective mitochondrial autophagy, or mitophagy, as a targeted

defense against oxidative stress, mitochondrial dysfunction, and aging. *Rejuvenation Res.* **8,** 3–5 (2005).

161. Mortensen, M. *et al.* The autophagy protein Atg7 is essential for hematopoietic stem cell maintenance. *J.Exp.Med.* **208,** 455–467 (2011).

162. Ho, T. T. *et al.* Autophagy maintains the metabolism and function of young and old stem cells. *Nature* **543,** 205–210 (2017).

163. Cappelli, E. *et al.* Defects in mitochondrial energetic function compels Fanconi Anaemia cells to glycolytic metabolism. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1863,** 1214–1221 (2017).

164. Zhang, T. *et al.* Fancd2 in vivo interaction network reveals a non-canonical role in mitochondrial function. *Sci. Rep.* **7,** 45626 (2017).

165. Orvedahl, A. *et al.* Image-based genome-wide siRNA screen identifies selective autophagy factors. *Nature* **480,** 113–117 (2011).

166. Vaseva, A. V & Moll, U. M. The mitochondrial p53 pathway. *Biochim. Biophys. Acta* **1787,** 414–420 (2009).

167. Medzhitov, R. Origin and physiological roles of inflammation. *Nature* **454,** 428–435 (2008).

168. Koh, P. S., Hughes, G. C., Faulkner, G. R., Keeble, W. W. & Bagby, G. C. The Fanconi anemia group C gene product modulates apoptotic responses to tumor necrosis factor-α and Fas ligand but does not suppress expression of receptors of the tumor necrosis factor receptor superfamily. *Exp.Hematol.* **27,** 1–8 (1999).

169. Pang, Q. *et al.* The Fanconi anemia protein FANCC binds to and facilitates the activation of STAT1 by gamma interferon and hematopoietic growth factors. *Mol.Cell.Biol.* **20,** 4724–4735 (2000).

170. Garbati, M. R. *et al.* FANCA and FANCC modulate TLR and p38 MAPK-dependent expression of IL-1β in macrophages. *Blood* **122,** 3197–3205 (2013).

171. Vanderwerf, S. M. *et al.* TLR8-dependent TNF-α overexpression in Fanconi anemia group C cells. *Blood* **114,** 5290–5298 (2009).

172. Brégnard, C. *et al.* Upregulated LINE-1 Activity in the Fanconi Anemia Cancer Susceptibility Syndrome Leads to Spontaneous Pro-inflammatory Cytokine Production. *EBioMedicine* **8,** 184–194 (2016).

173. Rosselli, F., Sanceau, J., Gluckman, E., Wietzerbin, J. & Moustacchi, E. Abnormal lymphokine production: a novel feature of the genetic disease Fanconi anemia. II. In vitro and in vivo spontaneous overproduction of tumor necrosis factor α. *Blood* **83,** 1216–1225 (1994).

174. Schultz, J. C. & Shahidi, N. T. Tumor necrosis factor-α overproduction in

Fanconi's anemia. *Am.J.Hematol.* **42,** 196–201 (1993).

175. Alter, B. P., Giri, N., Savage, S. A. & Rosenberg, P. S. Telomere length in inherited bone marrow failure syndromes. *Haematologica* **100,** 49–54 (2015).

176. Rhee, D. B. *et al.* FANCC suppresses short telomere-initiated telomere sister chromatid exchange. *Hum.Mol.Genet.* **19,** 879–887 (2010).

177. Franco, S. *et al.* Telomere dynamics in Fancg-deficient mouse and human cells. *Blood* **104,** 3927–3935 (2004).

178. Matsuzaki, K., Borel, V., Adelman, C. A., Schindler, D. & Boulton, S. J. FANCJ suppresses microsatellite instability and lymphomagenesis independent of the Fanconi anemia pathway. *Genes Dev.* **29,** 2532–2546 (2015).

179. Sarkar, J. *et al.* SLX4 contributes to telomere preservation and regulated processing of telomeric joint molecule intermediates. *Nucleic Acids Res.* **43,** 5912–5923 (2015).

180. Wan, B. *et al.* SLX4 assembles a telomere maintenance toolkit by bridging multiple endonucleases with telomeres. *Cell Rep.* **4,** 861–869 (2013).

181. Wilson, J. S. J. *et al.* Localization-dependent and-independent roles of SLX4 in regulating telomeres. *Cell Rep.* **4,** 853–860 (2013).

182. Tacconi, E. M. C. & Tarsounas, M. How homologous recombination maintains telomere integrity. *Chromosoma* **124,** 119–130 (2015).

183. Badie, S. *et al.* BRCA2 acts as a RAD51 loader to facilitate telomere replication and capping. *Nat. Struct. Mol. Biol.* **17,** 1461–1469 (2010).

184. Zimmer, J. *et al.* Targeting BRCA1 and BRCA2 deficiencies with G-quadruplex-interacting compounds. *Mol.Cell* **61,** 449–460 (2016).

185. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22,** 106–116 (2012).

186. Dilley, R. L. *et al.* Break-induced telomere synthesis underlies alternative telomere maintenance. *Nature* **539,** 54–58 (2016).

187. Howlett, N. G., Taniguchi, T., Durkin, S. G., D'Andrea, A. D. & Glover, T. W. The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Hum.Mol.Genet.* **14,** 693–701 (2005).

188. Luebben, S. W., Kawabata, T., Johnson, C. S., O'Sullivan, M. G. & Shima, N. A concomitant loss of dormant origins and FANCC exacerbates genome instability by impairing DNA replication fork progression. *Nucleic Acids Res.* **42,** 5605–5615 (2014).

189. Long, D. T., Raschle, M., Joukov, V. & Walter, J. C. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science* **333,** 84–87 (2011).

190. Liu, T., Ghosal, G., Yuan, J., Chen, J. & Huang, J. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science* **329,** 693–696 (2010).

191. Smogorzewska, A. *et al.* A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol.Cell* **39,** 36–47 (2010).

192. Lachaud, C. *et al.* Karyomegalic interstitial nephritis and DNA damage-induced polyploidy in Fan1 nuclease-defective knock-in mice. *Genes Dev.* **30,** 639–644 (2016).

193. Lachaud, C. *et al.* Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to prevent genome instability. *Science* **351,** 846–849 (2016).

194. Yoshikiyo, K. *et al.* KIAA1018/FAN1 nuclease protects cells against genomic instability induced by interstrand cross-linking agents. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21553–21557 (2010).

195. Zhou, W. *et al. FAN1* mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nat.Genet.* **44,** 910–915 (2012).

196. Duxin, J. P., Dewar, J. M., Yardimci, H. & Walter, J. C. Repair of a DNA-Protein Crosslink by Replication-Coupled Proteolysis. *Cell* **159,** 346–357 (2014).

197. Raghunandan, M., Chaudhury, I., Kelich, S. L., Hanenberg, H. & Sobeck, A. FANCD2, FANCJ and BRCA2 cooperate to promote replication fork recovery independently of the Fanconi Anemia core complex. *Cell Cycle* **14,** 342–353 (2015).

198. Vinciguerra, P., Godinho, S. A., Parmar, K., Pellman, D. & D'Andrea, A. D. Cytokinesis failure occurs in Fanconi anemia pathway-deficient murine and human bone marrow hematopoietic cells. *J.Clin.Invest.* **120,** 3834–3842 (2010).

199. Chan, K. L., Palmai-Pallag, T., Ying, S. & Hickson, I. D. Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nat. Cell Biol.* **11,** 753–760 (2009).

200. Naim, V. & Rosselli, F. The FANC pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities. *Nat. Cell Biol.* **11,** 761–768 (2009).

201. Nalepa, G. *et al.* Fanconi anemia signaling network regulates the spindle assembly checkpoint. *J.Clin.Invest.* **123,** 3839–3847 (2013).

202. Haynes, B., Saadat, N., Myung, B. & Shekhar, M. P. V. Crosstalk between translesion synthesis, Fanconi anemia network, and homologous recombination

repair pathways in interstrand DNA crosslink repair and development of chemoresistance. *Mutat. Res. Mutat. Res.* **763,** 258–266 (2015).

203. Adamo, A. *et al.* Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia. *Mol.Cell* **39,** 25–35 (2010).

204. Pace, P. *et al.* Ku70 corrupts DNA repair in the absence of the Fanconi anemia pathway. *Science* **329,** 219–223 (2010).

205. Howard, S. M., Yanez, D. A. & Stark, J. M. DNA damage response factors from diverse pathways, including DNA crosslink repair, mediate alternative end joining. *PLoS Genet.* **11,** e1004943 (2015).

206. Río, P. & Bueren, J. A. TGF-β: a master regulator of the bone marrow failure puzzle in Fanconi anemia. *Stem Cell Investig.* **3,** 75–75 (2016).

207. Epperly, M. W. *et al.* Induction of TGF-β by Irradiation or Chemotherapy in Fanconi Anemia (FA) Mouse Bone Marrow Is Modulated by Small Molecule Radiation Mitigators JP4-039 and MMS350. *In Vivo (Brooklyn).* **31,** 159–168 (2017).

208. García-Rubio, M. L. *et al.* The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genet.* **11,** e1005674 (2015).

209. Belo, H. *et al.* Epigenetic Alterations in Fanconi Anaemia: Role in Pathophysiology and Therapeutic Potential. *PLoS One* **10,** e0139740 (2015).

210. Hoeijmakers, J. H. J. DNA damage, aging, and cancer. *N. Engl. J. Med.* **361,** 1475–1485 (2009).

211. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715 (1993).

212. Noll, D. M., Mason, T. M. & Miller, P. S. Formation and repair of interstrand cross-links in DNA. *Chem.Rev.* **106,** 277–301 (2006).

213. Deans, A. J. & West, S. C. DNA interstrand crosslink repair and cancer. *Nat. Rev. Cancer* **11,** 467–480 (2011).

214. Schärer, O. D. DNA Interstrand Crosslinks: Natural and Drug-Induced DNA Adducts that Induce Unique Cellular Responses. *Chembiochem* **6,** 27–32 (2005).

215. Van Wassenhove, L. D., Mochly-Rosen, D. & Weinberg, K. I. Aldehyde Dehydrogenase 2 in Aplastic Anemia, Fanconi Anemia and Hematopoietic Stem Cells. *Mol.Genet.Metab.* **119,** 28–36 (2016).

216. Knipscheer, P. *et al.* The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* **326,** 1698–1701 (2009).

217. Mu, D. *et al.* DNA interstrand cross-links induce futile repair synthesis in mammalian cell extracts. *Mol.Cell.Biol.* **20,** 2446–2454 (2000).

218. Semlow, D. R., Zhang, J., Budzowska, M., Drohat, A. C. & Walter, J. C. Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell* **167,** 498–511.e14 (2016).

219. Yang, Z. *et al.* A role for the base excision repair enzyme NEIL3 in replication-dependent repair of interstrand DNA cross-links derived from psoralen and abasic sites. *DNA Repair (Amst).* **52,** 1–11 (2017).

220. Mullins, E. A., Warren, G. M., Bradley, N. P. & Eichman, B. F. Structure of a DNA glycosylase that unhooks interstrand cross-links. *Proc. Natl. Acad. Sci.* **114,** 4400–4405 (2017).

221. Kottemann, M. C. & Smogorzewska, A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature* **493,** 356–363 (2013).

222. Wang, W. A major switch for the Fanconi anemia DNA damage–response pathway. *Nat. Struct. Mol. Biol.* **15,** 1128–1130 (2008).

223. Zhang, J. & Walter, J. C. Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair* **19,** 135–142 (2014).

224. Zhang, J. *et al.* DNA interstrand cross-link repair requires replication-fork convergence. *Nat. Struct. Mol. Biol.* **22,** 242–247 (2015).

225. Coulthard, R. *et al.* Architecture and DNA recognition elements of the Fanconi anemia FANCM-FAAP24 complex. *Structure* **21,** 1648–1658 (2013).

226. Collis, S. J. *et al.* FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. *Mol.Cell* **32,** 313–324 (2008).

227. Ciccia, A. *et al.* Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Mol.Cell* **25,** 331–343 (2007).

228. Collins, N. B. *et al.* ATR-dependent phosphorylation of FANCA on serine 1449 after DNA damage is important for FA pathway function. *Blood* **113,** 2181–2190 (2009).

229. Luke-Glaser, S., Luke, B., Grossi, S. & Constantinou, A. FANCM regulates DNA chain elongation and is stabilized by S-phase checkpoint signalling. *EMBO J.* **29,** 795–805 (2010).

230. Schwab, R. A., Blackford, A. N. & Niedzwiedz, W. ATR activation and replication fork restart are defective in FANCM-deficient cells. *EMBO J.* **29,** 806–818 (2010).

231. Sobeck, A., Stone, S., Landais, I., de Graaf, B. & Hoatlin, M. E. The Fanconi anemia protein FANCM is controlled by FANCD2 and the ATR/ATM pathways. *J.Biol.Chem.* **284,** 25560–25568 (2009).

232. Shigechi, T. *et al.* ATR-ATRIP kinase complex triggers activation of the Fanconi

anemia DNA repair pathway. *Cancer Res.* **72,** 1149–1156 (2012).

233. Wang, X. *et al.* Chk1-mediated phosphorylation of FANCE is required for the Fanconi anemia/BRCA pathway. *Mol.Cell.Biol.* **27,** 3098–3108 (2007).

234. Huang, M. *et al.* The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response. *Mol. Cell* **39,** 259–268 (2010).

235. Ceccaldi, R., Sarangi, P. & D'Andrea, A. D. The Fanconi anaemia pathway: new players and new functions. *Nat. Rev. Mol. Cell Biol.* **17,** 337–349 (2016).

236. Meetei, A. R., Yan, Z. & Wang, W. FANCL replaces BRCA1 as the likely ubiquitin ligase responsible for FANCD2 monoubiquitination. *Cell Cycle* **3,** 174–176 (2004).

237. Hira, A. *et al.* Mutations in the gene encoding the E2 conjugating enzyme UBE2T cause Fanconi anemia. *Am. J. Hum. Genet.* **96,** 1001–1007 (2015).

238. Miles, J. A. *et al.* The Fanconi Anemia DNA repair pathway is regulated by an interaction between ubiquitin and the E2-like fold domain of FANCL. *J. Biol. Chem.* **290,** 20995–21006 (2015).

239. Virts, E. L. *et al.* AluY-mediated germline deletion, duplication and somatic stem cell reversion in *UBE2T* defines a new subtype of Fanconi anemia. *Hum.Mol.Genet.* **24,** 5093–5108 (2015).

240. Rajendra, E. *et al.* The genetic and biochemical basis of FANCD2 monoubiquitination. *Mol. Cell* **54,** 858–869 (2014).

241. Huang, Y. *et al.* Modularized functions of the Fanconi anemia core complex. *Cell Rep.* **7,** 1849–1857 (2014).

242. Nookala, R. K., Hussain, S. & Pellegrini, L. Insights into Fanconi Anaemia from the structure of human FANCE. *Nucleic Acids Res.* **35,** 1638–1648 (2007).

243. Gordon, S. M., Alon, N. & Buchwald, M. FANCC, FANCE, and FANCD2 Form a Ternary Complex Essential to the Integrity of the Fanconi Anemia DNA Damage Response Pathway. *J. Biol. Chem.* **280,** 36118–36125 (2005).

244. Polito, D. *et al.* The carboxyl terminus of FANCE recruits FANCD2 to the Fanconi Anemia (FA) E3 ligase complex to promote the FA DNA repair pathway. *J. Biol. Chem.* **289,** 7003–7010 (2014).

245. Swuec, P. & Costa, A. DNA replication and inter-strand crosslink repair: Symmetric activation of dimeric nanomachines? *Biophys. Chem.* (2016). doi:10.1016/j.bpc.2016.11.001

246. Ishiai, M. *et al.* FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **15,** 1138–1146 (2008).

247. van Twest, S. *et al.* Mechanism of Ubiquitination and Deubiquitination in the

Fanconi Anemia Pathway. *Mol. Cell* **65,** 247–259 (2017).

248. Nijman, S. M. B. *et al.* The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Mol.Cell* **17,** 331–339 (2005).

249. Cohn, M. A. *et al.* A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway. *Mol.Cell* **28,** 786–797 (2007).

250. Andreassen, P. R., D'Andrea, A. D. & Taniguchi, T. ATR couples FANCD2 monoubiquitination to the DNA-damage response. *Genes Dev.* **18,** 1958–1963 (2004).

251. Gibbs-Seymour, I. *et al.* Ubiquitin-SUMO circuitry controls activated fanconi anemia ID complex dosage in response to DNA damage. *Mol.Cell* **57,** 150–164 (2015).

252. Yamamoto, K. N. *et al.* Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 6492–6496 (2011).

253. Crossan, G. P. *et al.* Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. *Nat. Genet.* **43,** 147–152 (2011).

254. Hodskinson, M. R. G. *et al.* Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair. *Mol.Cell* **54,** 472–484 (2014).

255. Douwel, D. K. *et al.* XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol.Cell* **54,** 460–471 (2014).

256. Wyatt, H. D. M., Sarbajna, S., Matos, J. & West, S. C. Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells. *Mol.Cell* **52,** 234–247 (2013).

257. Wyatt, H. D. M., Laister, R. C., Martin, S. R., Arrowsmith, C. H. & West, S. C. The SMX DNA Repair Tri-nuclease. *Mol. Cell* **65,** 848–860.e11 (2017).

258. Kim, H. & D'Andrea, A. D. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes Dev.* **26,** 1393–1408 (2012).

259. Hicks, J. K. *et al.* Differential roles for DNA polymerases eta, zeta, and REV1 in lesion bypass of intrastrand versus interstrand DNA cross-links. *Mol.Cell.Biol.* **30,** 1217–1230 (2010).

260. Nelson, J. R., Lawrence, C. W. & Hinkle, D. C. Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* **382,** 729–731 (1996).

261. Haracska, L., Prakash, S. & Prakash, L. Yeast Rev1 protein is a G template-specific DNA polymerase. *J.Biol.Chem.* **277,** 15546–15551 (2002).

262. Waters, L. S. *et al.* Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol.Mol.Biol.Rev.* **73,** 134–154 (2009).

263. Sharma, S. *et al.* REV1 and polymerase ζ facilitate homologous recombination repair. *Nucleic Acids Res.* **40,** 682–691 (2012).

264. Mogi, S., Butcher, C. E. & Oh, D. H. DNA polymerase η reduces the γ-H2AX response to psoralen interstrand crosslinks in human cells. *Exp.Cell Res.* **314,** 887–895 (2008).

265. Budzowska, M., Graham, T. G., Sobeck, A., Waga, S. & Walter, J. C. Regulation of the Rev1-pol ζ complex during bypass of a DNA interstrand cross-link. *EMBO J.* **34,** 1971–1985 (2015).

266. Bluteau, D. *et al.* Biallelic inactivation of *REV7* is associated with Fanconi anemia. *J. Clin. Invest.* **127,** 1117–1117 (2017).

267. Renkawitz, J., Lademann, C. A. & Jentsch, S. Mechanisms and principles of homology search during recombination. *Nat. Rev. Mol. Cell Biol.* **15,** 369–383 (2014).

268. Mimitou, E. P. & Symington, L. S. DNA end resection—unraveling the tail. *DNA Repair* **10,** 344–348 (2011).

269. Nimonkar, A. V *et al.* BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev.* **25,** 350–362 (2011).

270. Chen, R. & Wold, M. S. Replication protein A: Single-stranded DNA's first responder. *Bioessays* **36,** 1156–1161 (2014).

271. Xia, B. *et al.* Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol.Cell* **22,** 719–729 (2006).

272. Sy, S. M., Huen, M. S. & Chen, J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 7155–7160 (2009).

273. Zhang, F., Fan, Q., Ren, K. & Andreassen, P. R. PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. *Mol.Cancer.Res.* **7,** 1110–1118 (2009).

274. Yang, H. *et al.* BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **297,** 1837–1848 (2002).

275. Reuter, M. *et al.* BRCA2 diffuses as oligomeric clusters with RAD51 and changes mobility after DNA damage in live cells. *J. Cell Biol.* **207,** 599–613 (2014).

276. Lord, C. J. & Ashworth, A. RAD51, BRCA2 and DNA repair: a partial resolution. *Nat. Struct. Mol. Biol.* **14,** 461–462 (2007).

277. Hucl, T. *et al.* A syngeneic variance library for functional annotation of human variation: application to BRCA2. *Cancer Res.* **68,** 5023–5030 (2008).

278. Ayoub, N. *et al.* The carboxyl terminus of Brca2 links the disassembly of Rad51 complexes to mitotic entry. *Curr. Biol.* **19,** 1075–1085 (2009).

279. Schlacher, K. *et al.* Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145,** 529–542 (2011).

280. Zhang, F. *et al.* PALB2 links BRCA1 and BRCA2 in the DNA-damage response. *Curr. Biol.* **19,** 524–529 (2009).

281. Higgs, M. R. *et al.* BOD1L is required to suppress deleterious resection of stressed replication forks. *Mol.Cell* **59,** 462–477 (2015).

282. Kowalczykowski, S. C. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harb. Perspect. Biol.* **7,** a016410 (2015).

283. Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science* **259,** 1896–1899 (1993).

284. Sung, P. Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* **265,** 1241–1243 (1994).

285. Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu.Rev.Genet.* **44,** 113–139 (2010).

286. Hilario, J., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. Direct imaging of human Rad51 nucleoprotein dynamics on individual DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 361–368 (2009).

287. van Mameren, J. *et al.* Counting RAD51 proteins disassembling from nucleoprotein filaments under tension. *Nature* **457,** 745–748 (2009).

288. Wu, Y., Kantake, N., Sugiyama, T. & Kowalczykowski, S. C. Rad51 protein controls Rad52-mediated DNA annealing. *J.Biol.Chem.* **283,** 14883–14892 (2008).

289. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* **63,** 349–404 (1999).

290. Bzymek, M., Thayer, N. H., Oh, S. D., Kleckner, N. & Hunter, N. Double Holliday junctions are intermediates of DNA break repair. *Nature* **464,** 937–941 (2010).

291. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33,** 25–35 (1983).

292. Maher, R. L., Branagan, A. M. & Morrical, S. W. Coordination of DNA replication and recombination activities in the maintenance of genome stability. *J. Cell. Biochem.* **112,** 2672–2682 (2011).

293. Formosa, T. & Alberts, B. M. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. *Cell* **47,** 793–806 (1986).

294. Lydeard, J. R. *et al.* Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. *Genes Dev.* **24,** 1133–1144 (2010).

295. McEachern, M. J. & Haber, J. E. Break-Induced Replication and Recombinational Telomere Elongation in Yeast. *Annu. Rev. Biochem.* **75,** 111–135 (2006).

296. Lim, D. S. & Hasty, P. A mutation in mouse *rad51* results in an early embryonic lethal that is suppressed by a mutation in *p53*. *Mol.Cell.Biol.* **16,** 7133–7143 (1996).

297. Tsuzuki, T. *et al.* Targeted disruption of the *Rad51* gene leads to lethality in embryonic mice. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 6236–6240 (1996).

298. Benson, F. E., Stasiak, A. & West, S. C. Purification and characterization of the human Rad51 protein, an analogue of E. coli RecA. *EMBO J.* **13,** 5764–5771 (1994).

299. Shinohara, A., Ogawa, H. & Ogawa, T. Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **69,** 457–470 (1992).

300. Shinohara, A. *et al.* Cloning of human, mouse and fission yeast recombination genes homologous to *RAD51* and *recA*. *Nat.Genet.* **4,** 239–243 (1993).

301. Lin, Z., Kong, H., Nei, M. & Ma, H. Origins and evolution of the *recA*/*RAD51* gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 10328–10333 (2006).

302. Chintapalli, S. V *et al.* Reevaluation of the evolutionary events within recA/RAD51 phylogeny. *BMC Genomics* **14,** 240 (2013).

303. Miller, K. A., Sawicka, D., Barsky, D. & Albala, J. S. Domain mapping of the Rad51 paralog protein complexes. *Nucleic Acids Res.* **32,** 169–178 (2004).

304. Erzberger, J. P. & Berger, J. M. Evolutionary relationships and structural mechanisms of AAA proteins. *Annu.Rev.Biophys.Biomol.Struct.* **35,** 93–114 (2006).

305. Pellegrini, L. *et al.* Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* **420,** 287–293 (2002).

306. Story, R. M. & Steitz, T. A. Structure of the recA protein–ADP complex. *Nature* **355,** 374–376 (1992).

307. Short, J. M. *et al.* High-resolution structure of the presynaptic RAD51 filament on single-stranded DNA by electron cryo-microscopy. *Nucleic Acids Res.* **44,** 9017–9030 (2016).

308. Baumann, P. & West, S. C. The human Rad51 protein: polarity of strand transfer

and stimulation by hRP-A. *EMBO J.* **16,** 5198–5206 (1997).

309. Mazin, A. V & Kowalczykowski, S. C. The specificity of the secondary DNA binding site of RecA protein defines its role in DNA strand exchange. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10673–10678 (1996).

310. Howard-Flanders, P., West, S. C. & Stasiak, A. Role of RecA protein spiral filaments in genetic recombination. *Nature* **309,** 215–9 (1984).

311. Baumann, P., Benson, F. E. & West, S. C. Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro. *Cell* **87,** 757–766 (1996).

312. Haaf, T., Golub, E. I., Reddy, G., Radding, C. M. & Ward, D. C. Nuclear foci of mammalian Rad51 recombination protein in somatic cells after DNA damage and its localization in synaptonemal complexes. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 2298–2302 (1995).

313. Plug, A. W., Xu, J., Reddy, G., Golub, E. I. & Ashley, T. Presynaptic association of Rad51 protein with selected sites in meiotic chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 5920–5924 (1996).

314. Barlow, A. L., Benson, F. E., West, S. C. & Hulten, M. A. Distribution of the Rad51 recombinase in human and mouse spermatocytes. *EMBO J.* **16,** 5207–5215 (1997).

315. Cloud, V., Chan, Y.-L., Grubb, J., Budke, B. & Bishop, D. K. Rad51 Is an Accessory Factor for Dmc1-Mediated Joint Molecule Formation During Meiosis. *Science* **337,** 1222–1225 (2012).

316. Bezzubova, O., Shinohara, A., Mueller, R. G., Ogawa, H. & Buerstedde, J. M. A chicken RAD51 homologue is expressed at high levels in lymphoid and reproductive organs. *Nucleic Acids Res.* **21,** 1577–1580 (1993).

317. Morita, T. *et al.* A mouse homolog of the *Escherichia coli recA* and *Saccharomyces cerevisiae RAD51* genes. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 6577–6580 (1993).

318. Nusslein-Volhard, C. & Dahm, R. *Zebrafish*. (Oxford University Press, 2002).

319. Talwar, P. K. & Jhingran, A. G. *Inland fishes of India and adjacent countries*. (CRC Press, 1991).

320. Lawrence, C. The husbandry of zebrafish (*Danio rerio*): A review. *Aquaculture* **269,** 1–20 (2007).

321. Meyer, A. & Schartl, M. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr.Opin.Cell Biol.* **11,** 699–704 (1999).

322. Amores, A. *et al.* Zebrafish hox clusters and vertebrate genome evolution.

Science **282,** 1711–1714 (1998).

323. Oellacher, J. Beiträge zur Entwicklungsgeschichte der Knochenfische nach Beobachtungen am Bachforelleneie. *Zeitschrift für wissenschaftliche Zool.* **23,** 1–115 (1872).

324. Colle-Candevelde, A. Blood Anlage in Teleostei. *Nature* **198,** 1223–1223 (1963).

325. Carroll, K. J. & North, T. E. Oceans of opportunity: Exploring vertebrate hematopoiesis in zebrafish. *Exp. Hematol.* **42,** 684–696 (2014).

326. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496,** 498–503 (2013).

327. Kettleborough, R. N. W. *et al.* A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* **496,** 494–497 (2013).

328. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat.Biotechnol.* **31,** 227–229 (2013).

329. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* **23,** 465–472 (2013).

330. Ablain, J., Durand, E. M., Yang, S., Zhou, Y. & Zon, L. I. A CRISPR/Cas9 Vector System for Tissue-Specific Gene Disruption in Zebrafish. *Dev. Cell* **32,** 756–764 (2015).

331. Brocal, I. *et al.* Efficient identification of CRISPR/Cas9-induced insertions/deletions by direct germline screening in zebrafish. *BMC Genomics* **17,** 1 (2016).

332. Paik, E. J. & Zon, L. I. Hematopoietic development in the zebrafish. *Int.J.Dev.Biol.* **54,** 1127–1137 (2010).

333. Ellett, F. & Lieschke, G. J. Zebrafish as a model for vertebrate hematopoiesis. *Curr. Opin. Pharmacol.* **10,** 563–570 (2010).

334. Detrich 3rd, H. W. *et al.* Intraembryonic hematopoietic cell migration during vertebrate development. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 10713–10717 (1995).

335. Lieschke, G. J. *et al.* Zebrafish SPI-1 (PU. 1) marks a site of myeloid development independent of primitive erythropoiesis: implications for axial patterning. *Dev.Biol.* **246,** 274–295 (2002).

336. Herbomel, P., Thisse, B. & Thisse, C. Ontogeny and behaviour of early macrophages in the zebrafish embryo. *Development* **126,** 3735–3745 (1999).

337. Le Guyader, D. *et al.* Origins and unconventional behavior of neutrophils in developing zebrafish. *Blood* **111,** 132–141 (2008).

338. Tamplin, O. J. *et al.* Hematopoietic Stem Cell Arrival Triggers Dynamic Remodeling of the Perivascular Niche. *Cell* **160,** 241–252 (2015).

339. Ma, D., Zhang, J., Lin, H. F., Italiano, J. & Handin, R. I. The identification and characterization of zebrafish hematopoietic stem cells. *Blood* **118,** 289–297 (2011).

340. Bertrand, J. Y., Kim, A. D., Teng, S. & Traver, D. CD41+ cmyb+ precursors colonize the zebrafish pronephros by a novel migration route to initiate adult hematopoiesis. *Development* **135,** 1853–1862 (2008).

341. Bertrand, J. Y. *et al.* Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464,** 108–111 (2010).

342. North, T. E. *et al.* Hematopoietic stem cell development is dependent on blood flow. *Cell* **137,** 736–748 (2009).

343. Murayama, E. *et al.* Tracing hematopoietic precursor migration to successive hematopoietic organs during zebrafish development. *Immunity* **25,** 963–975 (2006).

344. Hess, I. & Boehm, T. Intravital imaging of thymopoiesis reveals dynamic lympho-epithelial interactions. *Immunity* **36,** 298–309 (2012).

345. Willett, C. E., Zapata, A. G., Hopkins, N. & Steiner, L. A. Expression of Zebrafish *rag* Genes during early development identifies the thymus. *Dev.Biol.* **182,** 331–341 (1997).

346. Burns, C. E. *et al.* Isolation and characterization of runxa and *runxb*, zebrafish members of the runt family of transcriptional regulators. *Exp.Hematol.* **30,** 1381–1389 (2002).

347. Drummond, I. Making a zebrafish kidney: a tale of two tubes. *Trends Cell Biol.* **13,** 357–65 (2003).

348. Ganis, J. J. *et al.* Zebrafish *globin* switching occurs in two developmental stages and is controlled by the LCR. *Dev. Biol.* **366,** 185–194 (2012).

349. Jagadeeswaran, P., Sheehan, J. P., Craig, F. E. & Troyer, D. Identification and characterization of zebrafish thrombocytes. *Br.J.Haematol.* **107,** 731–738 (1999).

350. Lieschke, G. J., Oates, A. C., Crowhurst, M. O., Ward, A. C. & Layton, J. E. Morphologic and functional characterization of granulocytes and macrophages in embryonic and adult zebrafish. *Blood* **98,** 3087–3096 (2001).

351. Bennett, C. M. *et al.* Myelopoiesis in the zebrafish, Danio rerio. *Blood* **98,** 643–651 (2001).

352. Lugo-Villarino, G. *et al.* Identification of dendritic antigen-presenting cells in the zebrafish. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15850–15855 (2010).

353. Dobson, J. T. *et al.* Carboxypeptidase A5 identifies a novel mast cell lineage in the zebrafish providing new insight into mast cell fate determination. *Blood* **112,** 2969–2972 (2008).

354. Page, D. M. *et al.* An evolutionarily conserved program of B-cell development and activation in zebrafish. *Blood* **122,** e1–e11 (2013).

355. Yoder, J. A. *et al.* Immune-type receptor genes in zebrafish share genetic and functional properties with genes encoded by the mammalian leukocyte receptor cluster. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 6771–6776 (2001).

356. Yoder, J. A. *et al.* Resolution of the novel immune-type receptor gene cluster in zebrafish. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 15706–15711 (2004).

357. Muire, P. J., Hanson, L., Yoder, J. & Petrie-Hanson, L. Transcript analysis of natural killer (NK) cell specific genes in the liver, kidney and spleen tissues of *rag1*–/– mutant zebrafish in response to *in vivo* administration of TLR ligands. *J. Immunol.* **196,** 216.4-216.4 (2016).

358. Dong, H. *et al.* Update of the human and mouse Fanconi anemia genes. *Hum. Genomics* **9,** 32 (2015).

359. Titus, T. A. *et al.* The Fanconi anemia/BRCA gene network in zebrafish: embryonic expression and comparative genomics. *Mutat. Res./Fundam. Mol. Mech. Mutag.* **668,** 117–132 (2009).

360. Titus, T. A. *et al.* The Fanconi anemia gene network is conserved from zebrafish to human. *Gene* **371,** 211–223 (2006).

361. Liu, T. X. *et al.* Knockdown of zebrafish Fancd2 causes developmental abnormalities via p53-dependent apoptosis. *Dev. Cell* **5,** 903–914 (2003).

362. Rodríguez-Marí, A. *et al.* Sex reversal in zebrafish *fancl* mutants is caused by Tp53-mediated germ cell apoptosis. *PLoS Genet.* **6,** e1001034 (2010).

363. Shive, H. R. *et al. brca2* in Zebrafish Ovarian Development, Spermatogenesis, and Tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 19350–19355 (2010).

364. Rodríguez-Marí, A. *et al.* Roles of *brca2* (*fancd1*) in oocyte nuclear architecture, gametogenesis, gonad tumors, and genome stability in zebrafish. *PLoS Genet.* **7,** e1001357 (2011).

365. Berghmans, S. *et al. tp53* Mutant Zebrafish Develop Malignant Peripheral Nerve Sheath Tumors. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 407–412 (2005).

366. Tang, Q. *et al.* Optimized cell transplantation using adult *rag2* mutant zebrafish. *Nat. Methods* **11,** 821–824 (2014).

367. Lin, H. F. *et al.* Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. *Blood* **106,** 3803–3810 (2005).

368. Long, Q. *et al.* GATA-1 expression pattern can be recapitulated in living transgenic zebrafish using GFP reporter gene. *Development* **124,** 4105–4111 (1997).

369. Glass, T. J. *et al.* Stromal cell-derived factor-1 and hematopoietic cell homing in an adult zebrafish model of hematopoietic cell transplantation. *Blood* **118,** 766–774 (2011).

370. Karlsson, J., von Hofsten, J. & Olsson, P.-E. E. Generating transparent zebrafish: a refined method to improve detection of gene expression during embryonic development. *Mar. Biotechnol.* **3,** 522–527 (2001).

371. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^(-Delta Delta C(T)) Method. *Methods* **25,** 402–408 (2001).

372. Yuan, J., Reed, A., Chen, F. & Stewart, C. N. Statistical analysis of real-time PCR data. *BMC Bioinformatics* **7,** 85 (2006).

373. Wittamer, V., Bertrand, J. Y., Gutschow, P. W. & Traver, D. Characterization of the mononuclear phagocyte system in zebrafish. *Blood* **117,** 7126–7135 (2011).

374. Kyritsis, N. *et al.* Acute inflammation initiates the regenerative response in the adult zebrafish brain. *Science* **338,** 1353–1356 (2012).

375. Macaulay, I. C. *et al.* Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep.* **14,** 966–977 (2016).

376. Traver, D. *et al.* Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat.Immunol.* **4,** 1238–1246 (2003).

377. Athanasiadis, E. I. *et al.* Single-cell RNA-Sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *bioRxiv* (2017). doi:10.1101/117960

378. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9,** 171–181 (2014).

379. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10,** 1096–1098 (2013).

380. Patro, R., Mount, S. M. & Kingsford, C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat. Biotechnol.* **32,** (2014).

381. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

382. Ogata, H., Goto, S., Fujibuchi, W. & Kanehisa, M. Computation with the KEGG pathway database. *Biosystems.* **47,** 119–128 (1998).

383. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list

enrichment analysis tool. *BMC Bioinformatics* **14,** 128 (2013).

384. Muris, D. F. *et al.* Cloning of human and mouse genes homologous to *RAD52*, a yeast gene involved in DNA repair and recombination. *Mutat. Res.* **315,** 295–305 (1994).

385. Saeki, T., Machida, I. & Nakai, S. Genetic control of diploid recovery after gamma-irradiation in the yeast Saccharomyces cerevisiae. *Mutat. Res.* **73,** 251–265 (1980).

386. Sonoda, E. *et al.* Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *EMBO J.* **17,** 598–608 (1998).

387. Bakker, S. T., de Winter, J. P. & te Riele, H. Learning from a paradox: recent insights into Fanconi anaemia through studying mouse models. *Dis.Model.Mech.* **6,** 40–47 (2013).

388. Parmar, K., D'Andrea, A. & Niedernhofer, L. J. Mouse models of Fanconi anemia. *Mutat. Res.* **668,** 133–140 (2009).

389. Wong, J. C. Y. *et al.* 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,** 2063–2076 (2003).

390. Chen, M. *et al.* Inactivation of *Fac* in mice produces inducible chromosomal instability and reduced fertility reminiscent of Fanconi anaemia. *Nat. Genet.* **12,** 448–451 (1996).

391. Houghtaling, S. *et al.* Epithelial cancer in Fanconi anemia complementation group D2 (Fancd2) knockout mice. *Genes Dev.* **17,** 2021–2035 (2003).

392. Agoulnik, A. I. *et al.* 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,** 3047–3053 (2002).

393. Yang, Y. *et al.* Targeted disruption of the murine Fanconi anemia gene, *Fancg/Xrcc9*. *Blood* **98,** 3435–3440 (2001).

394. Koomen, M. *et al.* Reduced fertility and hypersensitivity to mitomycin C characterize *Fancg/Xrcc9* null mice. *Hum. Mol. Genet.* **11,** 273–281 (2002).

395. Bakker, S. T. *et al. Fancf*-deficient mice are prone to develop ovarian tumours. *J. Pathol.* **226,** 28–39 (2012).

396. Rantakari, P. *et al.* Inactivation of *Palb2* gene leads to mesoderm differentiation defect and early embryonic lethality in mice. *Hum. Mol. Genet.* **19,** 3021–3029 (2010).

397. Bouwman, P. *et al.* Loss of p53 partially rescues embryonic development of *Palb2*

knockout mice but does not foster haploinsufficiency of *Palb2* in tumour suppression. *J. Pathol.* **224,** 10–21 (2011).

398. Kuznetsov, S. *et al.* RAD51C deficiency in mice results in early prophase I arrest in males and sister chromatid separation at metaphase II in females. *J. Cell Biol.* **176,** 581–592 (2007).

399. Smeenk, G. *et al.* Rad51C is essential for embryonic development and haploinsufficiency causes increased DNA damage sensitivity and genomic instability. *Mutat. Res.* **689,** 50–58 (2010).

400. Evers, B. & Jonkers, J. Mouse models of BRCA1 and BRCA2 deficiency: past lessons, current understanding and future prospects. *Oncogene* **25,** 5885–5897 (2006).

401. Navarro, S. *et al.* Hematopoietic dysfunction in a mouse model for Fanconi anemia group D1. *Mol. Ther.* **14,** 525–535 (2006).

402. Kettleborough, R. N. W., Bruijn, E. de, Eeden, F. van, Cuppen, E. & Stemple, D. L. High-throughput target-selected gene inactivation in zebrafish. *Methods Cell Biol.* **104,** 121–127 (2011).

403. Abrams, E. W. & Mullins, M. C. Early zebrafish development: it's in the maternal genes. *Curr. Opin. Genet. Dev.* **19,** 396–403 (2009).

404. Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119,** 447–456 (1993).

405. Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* **30,** 581–613 (2014).

406. Sheehan, H. L. & Storey, G. W. An improved method of staining leucocyte granules with Sudan black B. *J. Pathol. Bacteriol.* **59,** 336 (1947).

407. Ema, H., Uchinomiya, K., Morita, Y., Suda, T. & Iwasa, Y. Repopulation dynamics of single haematopoietic stem cells in mouse transplantation experiments: Importance of stem cell composition in competitor cells. *J. Theor. Biol.* **394,** 57–67 (2016).

408. Flach, J. *et al.* Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. *Nature* **512,** 198–202 (2014).

409. Down, J. D. & Ploemacher, R. E. Transient and permanent engraftment potential of murine hematopoietic stem cell subsets: differential effects of host conditioning with gamma radiation and cytotoxic drugs. *Exp. Hematol.* **21,** 913–921 (1993).

410. Wong, J. C. *et al.* 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–

2076 (2003).

411. Helleday, T. The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings. *Mol. Oncol.* **5,** 387–393 (2011).

412. Dhliwayo, N., Sarras, M. P., Luczkowski, E., Mason, S. M. & Intine, R. V. Parp Inhibition Prevents Ten-Eleven Translocase Enzyme Activation and Hyperglycemia-Induced DNA Demethylation. *Diabetes* **63,** 3069–3076 (2014).

413. Hölzel, M. *et al.* FANCD2 protein is expressed in proliferating cells of human tissues that are cancer-prone in Fanconi anaemia. *J. Pathol.* **201,** 198–203 (2003).

414. Liew, W. C. & Orbán, L. Zebrafish sex: a complicated affair. *Brief. Funct. Genomics* **13,** 172–187 (2014).

415. Hollander, M. C. *et al.* Analysis of the mammalian *gadd45* gene and its response to DNA damage. *J. Biol. Chem.* **268,** 24385–93 (1993).

416. Tischkowitz, M. D. & Hodgson, S. V. Fanconi anaemia. *J.Med.Genet.* **40,** 1–10 (2003).

417. Kuranda, K. *et al.* Age-related changes in human hematopoietic stem/progenitor cells. *Aging Cell* **10,** 542–546 (2011).

418. Knight, K. L. & Barrington, R. A. Somatic diversification of IgH genes in rabbit. *Immunol. Rev.* **162,** 37–47 (1998).

419. Barrington, R. A., Fasullo, M. & Knight, K. L. A role for RAD51 in the generation of immunoglobulin gene diversity in rabbits. *J. Immunol.* **162,** 911–919 (1999).

420. Li, M. J. *et al.* Rad51 expression and localization in B cells carrying out class switch recombination. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10222–10227 (1996).

421. Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H. & Moore, K. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell* **167,** 1296–1309.e10 (2016).

422. Geiger, H., de Haan, G. & Florian, M. C. The ageing haematopoietic stem cell compartment. *Nat. Rev. Immunol.* **13,** 376–389 (2013).

423. Beerman, I. *et al.* Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 5465–5470 (2010).

424. Cho, R. H., Sieburg, H. B. & Muller-Sieburg, C. E. A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells. *Blood* **111,** 5553–5561 (2008).

425. Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M. & de Haan, G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *J.*

Exp. Med. **208,** 2691–2703 (2011).

426. Morrison, S. J., Wandycz, A. M., Akashi, K., Globerson, A. & Weissman, I. L. The aging of hematopoietic stem cells. *Nat. Med.* **2,** 1011–1016 (1996).

427. Pang, W. W. *et al.* Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 20012–20017 (2011).

428. Sudo, K., Ema, H., Morita, Y. & Nakauchi, H. Age-associated characteristics of murine hematopoietic stem cells. *J. Exp. Med.* **192,** 1273–1280 (2000).

429. Busch, K. *et al.* Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature* (2015).

430. Levine, A. J. & Oren, M. The first 30 years of p53: growing ever more complex. *Nat. Rev. Cancer* **9,** 749–758 (2009).

431. Belyi, V. A. *et al.* The origins and evolution of the p53 family of genes. *Cold Spring Harb. Perspect. Biol.* **2,** a001198 (2010).

432. Lane, D. P. & Crawford, L. V. T antigen is bound to a host protein in SV40-transformed cells. *Nature* **278,** 261–263 (1979).

433. DeLeo, A. B. *et al.* Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 2420–2424 (1979).

434. Rotter, V., Witte, O. N., Coffman, R. & Baltimore, D. Abelson murine leukemia virus-induced tumors elicit antibodies against a host cell protein, P50. *J. Virol.* **36,** 547–555 (1980).

435. Chumakov, P. M., Iotsova, V. S. & Georgiev, G. P. [Isolation of a plasmid clone containing the mRNA sequence for mouse nonviral T-antigen]. *Dokl. Akad. Nauk SSSR* **267,** 1272–1275 (1982).

436. Matlashewski, G. *et al.* Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene. *EMBO J.* **3,** 3257–3262 (1984).

437. Wolf, D. & Rotter, V. Inactivation of p53 gene expression by an insertion of Moloney murine leukemia virus-like DNA sequences. *Mol. Cell. Biol.* **4,** 1402–1410 (1984).

438. Ben David, Y., Prideaux, V. R., Chow, V., Benchimol, S. & Bernstein, A. Inactivation of the *p53* oncogene by internal deletion or retroviral integration in erythroleukemic cell lines induced by Friend leukemia virus. *Oncogene* **3,** 179–185 (1988).

439. Mowat, M., Cheng, A., Kimura, N., Bernstein, A. & Benchimol, S. Rearrangements of the cellular p53 gene in erythroleukaemic cells transformed

by Friend virus. *Nature* **314,** 633–636 (1985).

440. Wolf, D. & Rotter, V. Major deletions in the gene encoding the p53 tumor antigen cause lack of p53 expression in HL-60 cells. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 790–794 (1985).

441. Baker, S. J. *et al.* Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244,** 217–221 (1989).

442. Finlay, C. A., Hinds, P. W. & Levine, A. J. The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57,** 1083–1093 (1989).

443. Malkin, D. *et al.* Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250,** 1233–1238 (1990).

444. Srivastava, S., Zou, Z. Q., Pirollo, K., Blattner, W. & Chang, E. H. Germ-line transmission of a mutated *p53* gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature* **348,** 747–749 (1990).

445. Rivlin, N., Brosh, R., Oren, M. & Rotter, V. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes Cancer* **2,** 466–474 (2011).

446. Peller, S. & Rotter, V. TP53 in hematological cancer: low incidence of mutations with significant clinical relevance. *Hum. Mutat.* **21,** 277–284 (2003).

447. Joerger, A. C. & Fersht, A. R. The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. *Annu. Rev. Biochem.* **85,** 375–404 (2016).

448. Bieging, K. T., Mello, S. S. & Attardi, L. D. Unravelling mechanisms of p53-mediated tumour suppression. *Nat. Rev. Cancer* **14,** 359–370 (2014).

449. El-Deiry, W. *WAF1,* a potential mediator of p53 tumor suppression. *Cell* **75,** 817–825 (1993).

450. Miyashita, E. M., Yang, B., Lam, K. M. ., Crawford, D. H. & Thorley-Lawson, D. A. A novel form of Epstein-Barr virus latency in normal B cells in vivo. *Cell* **80,** 593–601 (1995).

451. Ginsberg, D., Mechta, F., Yaniv, M. & Oren, M. Wild-type p53 can down-modulate the activity of various promoters. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 9979–9983 (1991).

452. Maltzman, W. & Czyzyk, L. UV irradiation stimulates levels of p53 cellular tumor antigen in nontransformed mouse cells. *Mol. Cell. Biol.* **4,** 1689–1694 (1984).

453. Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* **51,** 6304–6311 (1991).

454. Lane, D. P. p53, guardian of the genome. *Nature* **358,** 15–16 (1992).

455. Sablina, A. A. *et al.* The antioxidant function of the p53 tumor suppressor. *Nat. Med.* **11,** 1306–1313 (2005).

456. Gudkov, A. V & Komarova, E. A. Pathologies associated with the p53 response. *Cold Spring Harb. Perspect. Biol.* **2,** a001180 (2010).

457. Cui, Y. F. *et al.* Apoptosis in bone marrow cells of mice with different *p53* genotypes after gamma-rays irradiation in vitro. *J. Environ. Pathol. Toxicol. Oncol.* **14,** 159–163 (1995).

458. Wang, L. *et al.* Gamma-ray-induced cell killing and chromosome abnormalities in the bone marrow of *p53*-deficient mice. *Radiat. Res.* **146,** 259–266 (1996).

459. Shi, D. & Gu, W. Dual Roles of MDM2 in the Regulation of p53: Ubiquitination Dependent and Ubiquitination Independent Mechanisms of MDM2 Repression of p53 Activity. *Genes Cancer* **3,** 240–8 (2012).

460. Eischen, C. M. & Lozano, G. The Mdm Network and Its Regulation of p53 Activities: A Rheostat of Cancer Risk. *Hum. Mutat.* **35,** 728–737 (2014).

461. Abbas, T. & Dutta, A. p21 in cancer: intricate networks and multiple activities. *Nat. Rev. Cancer* **9,** 400–414 (2009).

462. Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21cip1/waf1. *Science* **287,** 1804–1808 (2000).

463. Liu, Y. *et al.* p53 Regulates Hematopoietic Stem Cell Quiescence. *Cell Stem Cell* **4,** 37–48 (2009).

464. Milyavsky, M. *et al.* A distinctive DNA damage response in human hematopoietic stem cells reveals an apoptosis-independent role for p53 in self-renewal. *Cell Stem Cell* **7,** 186–197 (2010).

465. Mohrin, M. *et al.* Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell* **7,** 174–185 (2010).

466. Viale, A. *et al.* Cell-cycle restriction limits DNA damage and maintains self-renewal of leukaemia stem cells. *Nature* **457,** 51–56 (2009).

467. Vousden, K. H. & Prives, C. Blinded by the light: the growing complexity of p53. *Cell* **137,** 413–431 (2009).

468. Blanpain, C., Mohrin, M., Sotiropoulou, P. A. & Passegue, E. DNA-damage response in tissue-specific and cancer stem cells. *Cell Stem Cell* **8,** 16–29 (2011).

469. Pant, V., Quintas-Cardama, A. & Lozano, G. The p53 pathway in hematopoiesis: lessons from mouse models, implications for humans. *Blood* **120,** 5118–5127 (2012).

470. Qiu, J., Papatsenko, D., Niu, X., Schaniel, C. & Moore, K. Divisional History and Hematopoietic Stem Cell Function during Homeostasis. *Stem Cell Reports* **2,** 473–490 (2014).

471. Beerman, I., Seita, J., Inlay, M. A., Weissman, I. L. & Rossi, D. J. Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. *Cell Stem Cell* **15,** 37–50 (2014).

472. TeKippe, M., Harrison, D. E. & Chen, J. Expansion of hematopoietic stem cell phenotype and activity in Trp53-null mice. *Exp. Hematol.* **31,** 521–527 (2003).

473. Akala, O. O. *et al.* Long-term haematopoietic reconstitution by *Trp53-/- p16Ink4a-/-p19Arf-/-* multipotent progenitors. *Nature* **453,** 228–232 (2008).

474. Chen, J. *et al.* Enrichment of hematopoietic stem cells with SLAM and LSK markers for the detection of hematopoietic stem cell function in normal and Trp53 null mice. *Exp. Hematol.* **36,** 1236–1243 (2008).

475. Leonova, K. I. *et al.* A small molecule inhibitor of p53 stimulates amplification of hematopoietic stem cells but does not promote tumor development in mice. *Cell Cycle* **9,** 1434–1443 (2010).

476. Wlodarski, P. *et al.* Role of p53 in hematopoietic recovery after cytotoxic treatment. *Blood* **91,** 2998–3006 (1998).

477. Rossi, D. J., Jamieson, C. H. M. & Weissman, I. L. Stems cells and the pathways to aging and cancer. *Cell* **132,** 681–696 (2008).

478. Komarova, E. A. *et al.* Dual effect of p53 on radiation sensitivity in vivo: p53 promotes hematopoietic injury, but protects from gastro-intestinal syndrome in mice. *Oncogene* **23,** 3265–3271 (2004).

479. Narla, A. & Ebert, B. L. Ribosomopathies: human disorders of ribosome dysfunction. *Blood* **115,** 3196–3205 (2010).

480. Cheng, R. *et al.* Zebrafish (*Danio rerio*) p53 tumor suppressor gene: cDNA sequence and expression during embryogenesis. *Mol. Mar. Biol. Biotechnol.* **6,** 88–97 (1997).

481. Langheinrich, U., Hennen, E., Stott, G. & Vacun, G. Zebrafish as a model organism for the identification and characterization of drugs and genes affecting p53 signaling. *Curr. Biol.* **12,** 2023–2028 (2002).

482. Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356,** 215–221 (1992).

483. Rosselli, F. *et al.* p53-dependent pathway of radio-induced apoptosis is altered in Fanconi anemia. *Oncogene* **10,** 9–17 (1995).

484. Kruyt, F. A., Dijkmans, L. M., van den Berg, T. K. & Joenje, H. Fanconi anemia

genes act to suppress a cross-linker-inducible p53-independent apoptosis pathway in lymphoblastoid cell lines. *Blood* **87,** 938–48 (1996).

485. Luo, Y. *et al.* Hypersensitivity of primordial germ cells to compromised replication-associated DNA repair involves ATM-p53-p21 signaling. *PLoS Genet.* **10,** e1004471 (2014).

486. Freie, B. *et al.* Fanconi anemia type C and p53 cooperate in apoptosis and tumorigenesis. *Blood* **102,** 4146–4152 (2003).

487. Houghtaling, S. *et al.* Heterozygosity for p53 (*Trp53*+/-) accelerates epithelial tumor formation in fanconi anemia complementation group D2 (*Fancd2*) knockout mice. *Cancer Res.* **65,** 85–91 (2005).

488. Shive, H. R., West, R. R., Embree, L. J., Golden, C. D. & Hickstein, D. D. *brca2* and *tp53* Collaborate in Tumorigenesis in Zebrafish. *PLoS One* **9,** (2014).

489. Ceccaldi, R. *et al.* Spontaneous abrogation of the G(2)DNA damage checkpoint has clinical benefits but promotes leukemogenesis in Fanconi anemia patients. *J.Clin.Invest.* **121,** 184–194 (2011).

490. Rice, D. & Barone Jr, S. Critical periods of vulnerability for the developing nervous system: evidence from humans and animal models. *Environ. Heal. Perspect.* **108 Suppl,** 511–533 (2000).

491. Lakhani, S. R. *et al.* Pathology of ovarian cancers in BRCA1 and *BRCA2* carriers. *Clin. Cancer Res.* **10,** 2473–2481 (2004).

492. Sowter, H. M. & Ashworth, A. *BRCA1* and *BRCA2* as ovarian cancer susceptibility genes. *Carcinogenesis* **26,** 1651–1656 (2005).

493. Song, H. *et al.* Contribution of Germline Mutations in the *RAD51B, RAD51C*, and *RAD51D* Genes to Ovarian Cancer in the Population. *J. Clin. Oncol.* **33,** 2901–2907 (2015).

494. Loveday, C. *et al.* Germline mutations in *RAD51D* confer susceptibility to ovarian cancer. *Nat. Genet.* **43,** 879–882 (2011).

495. Romanowicz-Makowska, H., Smolarz, B., Połać, I. & Sporny, S. Single nucleotide polymorphisms of *RAD51 G135C, XRCC2* Arg188His and *XRCC3* Thr241Met homologous recombination repair genes and the risk of sporadic endometrial cancer in Polish women. *J.Obstet.Gynaecol.Res.* **38,** 918–924 (2012).

496. Hu, X. & Sun, S. RAD51 Gene 135G/C polymorphism and ovarian cancer risk: a meta-analysis. *Int. J. Clin. Exp. Med.* **8,** 22365–22370 (2015).

497. Karigane, D. *et al.* p38α Activates Purine Metabolism to Initiate Hematopoietic Stem/Progenitor Cell Cycling in Response to Stress. *Cell Stem Cell* **19,** 192–204 (2016).

498. Ashley, N. T., Weil, Z. M. & Nelson, R. J. Inflammation: Mechanisms, Costs, and Natural Variation. *Annu. Rev. Ecol. Evol. Syst* **43,** 385–406 (2012).

499. Nathan, C. Points of control in inflammation. *Nature* **420,** 846–852 (2002).

500. Bianchi, M. E. DAMPs, PAMPs and alarmins: all we need to know about danger. *J. Leukoc. Biol.* **81,** 1–5 (2007).

501. Schuettpelz, L. G. & Link, D. C. Regulation of Hematopoietic Stem Cell Activity by Inflammation. *Front. Immunol.* **4,** 204 (2013).

502. Essers, M. A. G. *et al.* IFNα activates dormant haematopoietic stem cells in vivo. *Nature* **458,** 904–908 (2009).

503. Sato, T. *et al.* Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon–dependent exhaustion. *Nat. Med.* **15,** 696–700 (2009).

504. Pietras, E. M., Warr, M. R. & Passegue, E. Cell cycle regulation in hematopoietic stem cells. *J.Cell Biol.* **195,** 709–720 (2011).

505. Pietras, E. M. *et al.* Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal. *Nat. Cell Biol.* **18,** 607–618 (2016).

506. Kaschutnig, P. *et al.* The Fanconi anemia pathway is required for efficient repair of stress-induced DNA damage in haematopoietic stem cells. *Cell Cycle* **14,** 2734–2742 (2015).

507. Pang, Q. FANCC interacts with Hsp70 to protect hematopoietic cells from IFN-gamma/TNF-alpha-mediated cytotoxicity. *EMBO J.* **20,** 4478–4489 (2001).

508. Rathbun, R. K. *et al.* Inactivation of the Fanconi anemia group C gene augments interferon-gamma-induced apoptotic responses in hematopoietic cells. *Blood* **90,** 974–985 (1997).

509. Wang, J. *et al.* Overexpression of the fanconi anemia group C gene (*FAC*) protects hematopoietic progenitors from death induced by Fas-mediated apoptosis. *Cancer Res.* **58,** 3538–3541 (1998).

510. Pang, Q. *et al.* Role of double-stranded RNA-dependent protein kinase in mediating hypersensitivity of Fanconi anemia complementation group C cells to interferon gamma, tumor necrosis factor-alpha, and double-stranded RNA. *Blood* **97,** 1644–1652 (2001).

511. Hu, L., Huang, W., Hjort, E. & Eklund, E. A. Increased Fanconi C expression contributes to the emergency granulopoiesis response. *J.Clin.Invest.* **123,** 3952–3966 (2013).

512. Si, Y. *et al.* Continuous in vivo infusion of interferon-gamma (IFN-gamma)

enhances engraftment of syngeneic wild-type cells in *Fanca*-/- and *Fancg*-/- mice. *Blood* **108,** 4283–4287 (2006).

513. Li, X. *et al.* Continuous in vivo infusion of interferon-gamma (IFN-gamma) preferentially reduces myeloid progenitor numbers and enhances engraftment of syngeneic wild-type cells in *Fancc*-/- mice. *Blood* **104,** 1204–1209 (2004).

514. Li, J. *et al.* TNF-alpha induces leukemic clonal evolution ex vivo in Fanconi anemia group C murine stem cells. *J.Clin.Invest.* **117,** 3283–3295 (2007).

515. Matsushita, N. *et al.* Direct inhibition of TNF-α promoter activity by Fanconi anemia protein FANCD2. *PLoS One* **6,** e23324 (2011).

516. Dufour, C. *et al.* TNF-α and IFN-γ are overexpressed in the bone marrow of Fanconi anemia patients and TNF-α suppresses erythropoiesis in vitro. *Blood* **102,** 2053–2059 (2003).

517. Ibanez, A. *et al.* Elevated levels of IL-1beta in Fanconi anaemia group A patients due to a constitutively active phosphoinositide 3-kinase-Akt pathway are capable of promoting tumour cell proliferation. *Biochem.J.* **422,** 161–170 (2009).

518. Matsui, K., Giri, N., Alter, B. P. & Pinto, L. A. Cytokine production by bone marrow mononuclear cells in inherited bone marrow failure syndromes. *Br.J.Haematol.* **163,** 81–92 (2013).

519. Cachaço, A. S. *et al.* TNF-α regulates the effects of irradiation in the mouse bone marrow microenvironment. *PLoS One* **5,** e8980 (2010).

520. Garaycoechea, J. I. & Patel, K. J. Why does the bone marrow fail in Fanconi anemia? *Blood* **123,** 26–34 (2014).

521. Garbati, M. R. *et al.* Cytokine overproduction and crosslinker hypersensitivity are unlinked in Fanconi anemia macrophages. *J.Leukoc.Biol.* **99,** 455–465 (2016).

522. Brooks, P. J. & Theruvathu, J. A. DNA adducts from acetaldehyde: implications for alcohol-related carcinogenesis. *Alcohol* **35,** 187–193 (2005).

523. Larson, H. N., Weiner, H. & Hurley, T. D. Disruption of the coenzyme binding site and dimer interface revealed in the crystal structure of mitochondrial aldehyde dehydrogenase 'Asian' variant. *J.Biol.Chem.* **280,** 30550–30556 (2005).

524. Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF-κB by Toll-like receptor 3. *Nature* **413,** 732–738 (2001).

525. Magee, W. E. & Griffith, M. J. The liver as a site for interferon production in response to poly I: poly C. *Life Sci.* **11,** 1081–1086 (1972).

526. Xiong, R., Nie, L., Xiang, L. X. & Shao, J. Z. Characterization of a PIAS4 homologue from zebrafish: insights into its conserved negative regulatory

mechanism in the TRIF, MAVS, and IFN signaling pathways during vertebrate evolution. *J.Immunol.* **188,** 2653–2668 (2012).

527. Fortier, M. E. *et al.* The viral mimic, polyinosinic:polycytidylic acid, induces fever in rats via an interleukin-1-dependent mechanism. *Am.J.Physiol.Regul.Integr.Comp.Physiol.* **287,** R759-66 (2004).

528. Manetti, R. *et al.* Polyinosinic acid:polycytidylic acid promotes T helper type 1-specific immune responses by stimulating macrophage production of interferon-α and interleukin-12. *Eur.J.Immunol.* **25,** 2656–2660 (1995).

529. Melnicoff, M. J., Horan, P. K. & Morahan, P. S. Kinetics of changes in peritoneal cell populations following acute inflammation. *Cell. Immunol.* **118,** 178–191 (1989).

530. Ryan, G. B. & Majno, G. Acute inflammation. A review. *Am. J. Pathol.* **86,** 183–276 (1977).

531. Kaplanski, G. IL-6: a regulator of the transition from neutrophil to monocyte recruitment during inflammation. *Trends Immunol.* **24,** 25–29 (2003).

532. Essers, M. A. G. *et al.* IFNα activates dormant haematopoietic stem cells in vivo. *Nature* **458,** 904–908 (2009).

533. Novoa, B. & Figueras, A. in *Current Topics in Innate Immunity II* (eds. Lambris, J. D. & Hajishengallis, G.) 253–275 (Springer New York, 2012).

534. Pietras, E. M. *et al.* Re-entry into quiescence protects hematopoietic stem cells from the killing effect of chronic exposure to type I interferons. *J. Exp. Med.* **211,** 245–262 (2014).

535. Komarov, P. G. *et al.* A chemical inhibitor of p53 that protects mice from the side effects of cancer therapy. *Science* **285,** 1733–1737 (1999).

536. Heidt, T. *et al.* Chronic variable stress activates hematopoietic stem cells. *Nat. Med.* **20,** 754–758 (2014).

537. Flach, J. *et al.* Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. *Nature* **512,** 198–202 (2014).

538. Stern, H. M. & Zon, L. I. Cancer genetics and drug discovery in the zebrafish. *Nat. Rev. Cancer* **3,** 533–539 (2003).

539. Zon, L. I. & Peterson, R. T. In vivo drug discovery in the zebrafish. *Nat. Rev. Drug Discov.* **4,** 35–44 (2005).

540. Parng, C., Seng, W. L., Semino, C. & McGrath, P. Zebrafish: a preclinical model for drug screening. *Assay Drug Dev. Technol.* **1,** 41–48 (2002).

541. Rubinstein, A. L. Zebrafish: from disease modeling to drug discovery. *Curr. Opin. Drug Discov. Devel.* **6,** 218–23 (2003).

542. North, T. E. *et al.* Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature* **447,** 1007–1011 (2007).

543. Cerbinskaite, A., Mukhopadhyay, A., Plummer, E. R., Curtin, N. J. & Edmondson, R. J. Defective homologous recombination in human cancers. *Cancer Treat. Rev.* **38,** 89–100 (2012).

544. Sharan, S. K. *et al.* BRCA2 deficiency in mice leads to meiotic impairment and infertility. *Development* **131,** 131–142 (2004).

Appendix

Table A1: Differentially upregulated genes in the mutant *Tg(itga2b:EGFP)* GFP^{low} population. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000088641	*grn2*	*GRN*	-3.21064	0.77786	0.00004
ENSDARG00000105300	*BX927081.1*		-4.99657	1.28691	0.00010
ENSDARG00000002670	*ATG14*	*ATG14*	-5.38033	1.45565	0.00022
ENSDARG00000020730	*smpd4*	*SMPD4*	-5.10361	1.42013	0.00033
ENSDARG00000014731	*cacybp*	*CACYBP*	-4.44330	1.25307	0.00039
ENSDARG00000003495	*madd*	*MADD*	-4.41023	1.24480	0.00040
ENSDARG00000096127	*si:dkey-149m13.1*		-3.12261	0.88160	0.00040
ENSDARG00000055722	*bco2a*	*BCO2*	-2.65119	0.75876	0.00048
ENSDARG00000086712	*si:dkeyp-97b10.3*		-4.90511	1.42738	0.00059
ENSDARG00000075803	*slc41a2a*	*SLC41A2*	-4.07510	1.20066	0.00069
ENSDARG00000008976	*mdn1*	*MDN1*	-2.10434	0.63501	0.00092
ENSDARG00000087536	*znf407*	*ZNF407*	-4.78222	1.45507	0.00101
ENSDARG00000030830	*cmtr1*	*CMTR1*	-4.74105	1.49232	0.00149
ENSDARG00000059130	*gata1b*	*GATA1*	-2.76686	0.87466	0.00156
ENSDARG00000102436	*WIZ*	*WIZ*	-3.09365	1.00707	0.00213
ENSDARG00000075140	*CR318653.1*		-3.85393	1.25457	0.00213
ENSDARG00000104255	*arhgap27*	*ARHGAP27*	-3.74206	1.23763	0.00250
ENSDARG00000071196	*sdprb*	*SDPR*	-3.96298	1.31854	0.00265
ENSDARG00000016481	*ptpn2a*	*PTPN2*	-3.84578	1.29210	0.00292
ENSDARG00000019529	*parp1*	*PARP1*	-2.59181	0.87399	0.00302
ENSDARG00000103768	*si:dkeyp-28d2.4*		-3.86105	1.31415	0.00330
ENSDARG00000080009	*bahcc1*	*BAHCC1*	-2.01429	0.68852	0.00344
ENSDARG00000099521	*CU466240.1*		-4.42887	1.51911	0.00355
ENSDARG00000099359	*BX649497.2*		-3.48374	1.20492	0.00384
ENSDARG00000104953	*stt3a*	*STT3A*	-3.96841	1.37532	0.00391
ENSDARG00000094110	*si:ch211-284e20.4*		-3.33106	1.15556	0.00394
ENSDARG00000008904	*smarca2*	*SMARCA2*	-2.15960	0.75011	0.00399
ENSDARG00000063921	*mt-nd5*	*MT-ND5*	-1.53059	0.53209	0.00402
ENSDARG00000045297	*phb2a*	*PHB2*	-3.51376	1.23111	0.00432
ENSDARG00000042458	*rfc4*	*RFC4*	-3.73948	1.31289	0.00440
ENSDARG00000058587	*ccdc79*	*CCDC79*	-2.68611	0.94630	0.00453
ENSDARG00000095268	*si:dkey-261h17.1*		-3.20116	1.12863	0.00456
ENSDARG00000057787	*itgae.2*	*ITGAE*	-2.72101	0.97065	0.00506
ENSDARG00000080001	*si:dkey-208m12.2*		-3.38500	1.21010	0.00515
ENSDARG00000011770	*dhrs12*	*DHRS12*	-3.86711	1.38465	0.00522

ENSDARG00000103787	*chd1*	*CHD1*	-2.97632	1.07521	0.00564
ENSDARG00000075504			-3.66586	1.32956	0.00583
ENSDARG00000021664	*fzd3a*	*FZD3*	-2.97876	1.08283	0.00594
ENSDARG00000007244	*acp2*	*ACP2*	-3.18164	1.15667	0.00595
ENSDARG00000010477	*p2rx3a*	*P2RX3*	-3.50343	1.27862	0.00614
ENSDARG00000004184	*nf1b*	*NF1*	-2.49199	0.91082	0.00622
ENSDARG00000003449	*pde10a*	*PDE10A*	-3.80209	1.39128	0.00628
ENSDARG00000016867	*rnf128a*	*RNF148*	-3.06055	1.12252	0.00640
ENSDARG00000076586	*csf2rb*	*CSF2RB*	-2.46506	0.90905	0.00669
ENSDARG00000020606	*nfe2*	*NFE2*	-3.41368	1.26319	0.00688
ENSDARG00000012035	*zgc:100832*	*FBXO44*	-2.47787	0.91827	0.00697
ENSDARG00000016527	*helz2*	*HELZ2*	-3.33672	1.23892	0.00708
ENSDARG00000075980	*tmem125b*	*TMEM125*	-3.32091	1.24129	0.00746
ENSDARG00000008333	*znfl2a*		-2.69356	1.00892	0.00759
ENSDARG00000078624	*arhgef9b*	*ARHGEF9*	-2.35210	0.88205	0.00766
ENSDARG00000056678	*trim47*	*TRIM69*	-2.99633	1.13063	0.00805
ENSDARG00000097533	*si:dkey-29l4.4*		-2.72300	1.03277	0.00837
ENSDARG00000052190	*fdx1l*	*CTD-2369P2.10*	-2.67164	1.01716	0.00863
ENSDARG00000002369	*UBC*	*UBC*	-1.65672	0.63167	0.00872
ENSDARG00000062139	*eif2ak3*	*EIF2AK3*	-3.04968	1.16479	0.00884
ENSDARG00000006526	*fn1b*		-3.67090	1.40806	0.00913
ENSDARG00000103472	*nedd8l*		-2.87853	1.11075	0.00956

Table A2: Differentially downregulated genes in the mutant *Tg(itga2b:EGFP)* GFP^{low} population. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000000503	*stx1b*	*STX1B*	6.11001	1.26110	0.00000
ENSDARG00000068478	*gpx4a*	*GPX4*	5.85288	1.26965	0.00000
ENSDARG00000054578	*arl6ip1*	*ARL6IP1*	3.17018	0.69523	0.00001
ENSDARG00000053517	*EML5*	*EML5*	5.36563	1.19444	0.00001
ENSDARG00000103203	*vps28*	*VPS28*	5.40181	1.26452	0.00002
ENSDARG00000058354	*selt1a*	*SELT*	4.95022	1.19167	0.00003
ENSDARG00000099320	*si:dkey-284i7.2*		4.84260	1.17326	0.00004
ENSDARG00000031649	*sst3*		4.71697	1.15214	0.00004
ENSDARG00000018460	*mbnl2*	*MBNL2*	3.90972	0.96295	0.00005
ENSDARG00000098108	*dusp2*	*DUSP2*	5.00010	1.23284	0.00005
ENSDARG00000036848	*slc43a2a*	*SLC43A2*	4.18884	1.04068	0.00006
ENSDARG00000035557	*gabarapa*	*GABARAP*	3.14521	0.80101	0.00009
ENSDARG00000100743	*si:dkey-190j3.4*		4.54120	1.18217	0.00012
ENSDARG00000079374	*tjp1b*	*TJP1*	5.40336	1.41540	0.00013

ENSDARG00000007923	*ptpn1*	*PTPN1*	3.04912	0.80035	0.00014
ENSDARG00000008370	*csnk1da*		3.75439	0.98635	0.00014
ENSDARG00000069404	*pln*		4.26464	1.13026	0.00016
ENSDARG00000045909	*dynlt1*	*DYNLT1*	3.35902	0.89828	0.00018
ENSDARG00000005560	*ywhah*	*YWHAH*	3.13206	0.84600	0.00021
ENSDARG00000082142	*CR753862.1*		4.83569	1.32032	0.00025
ENSDARG00000058606	*sik1*	*CH507-42P11.8*	4.50009	1.23093	0.00026
ENSDARG00000103720	*ZFP36*	*ZFP36*	2.25793	0.61872	0.00026
ENSDARG00000100825	*calm3a*		2.79360	0.76901	0.00028
ENSDARG00000014320	*gucy2c*	*GUCY2C*	3.57602	0.98591	0.00029
ENSDARG00000104039	*errfi1*	*ERRFI1*	3.90807	1.07909	0.00029
ENSDARG00000004060	*bhlhe40*	*BHLHE40*	3.98608	1.12788	0.00041
ENSDARG00000053656	*lypla2*	*LYPLA2*	3.07593	0.87254	0.00042
ENSDARG00000098994	*CR352263.1*		4.03432	1.14535	0.00043
ENSDARG00000098380	*BX957331.1*		4.66642	1.32830	0.00044
ENSDARG00000002609	*rnf145a*	*RNF145*	4.55080	1.30657	0.00050
ENSDARG00000035253	*npr3*	*NPR3*	5.10625	1.46837	0.00051
ENSDARG00000074527	*chst15*	*CHST15*	4.94130	1.42816	0.00054
ENSDARG00000104077	*fcer1gl*	*FCER1G*	1.83374	0.53051	0.00055
ENSDARG00000036700			3.67168	1.06811	0.00059
ENSDARG00000076847	*tnrc6c1*	*TNRC6C*	2.26347	0.66105	0.00062
ENSDARG00000057456	*ppp3ccb*	*PPP3CC*	3.70745	1.08355	0.00062
ENSDARG00000020771	*tnr*	*TNR*	4.20224	1.25383	0.00080
ENSDARG00000015757	*tmem50a*	*TMEM50A*	3.72306	1.11352	0.00083
ENSDARG00000076667	*ccng1*	*CCNG1*	1.78161	0.53442	0.00086
ENSDARG00000103852	*si:ch211-215e19.7*		3.78015	1.13423	0.00086
ENSDARG00000018124	*psmd3*	*PSMD3*	2.36373	0.71044	0.00088
ENSDARG00000093303		*IFITM10*	4.39245	1.32400	0.00091
ENSDARG00000102998	*taf6*	*TAF6*	4.35802	1.31462	0.00092
ENSDARG00000003938	*rpa1*	*RPA1*	2.39537	0.72368	0.00093
ENSDARG00000104701	*map7d1b*	*MAP7D1*	3.85718	1.17146	0.00099
ENSDARG00000042796	*yy1a*	*YY1*	2.51769	0.76878	0.00106
ENSDARG00000079616	*cramp1l*	*CRAMP1*	4.28001	1.31534	0.00114
ENSDARG00000015059	*daam1a*	*DAAM1*	4.20451	1.29415	0.00116
ENSDARG00000053475	*ngb*	*NGB*	3.98501	1.23484	0.00125
ENSDARG00000076853	*AREG*	*AREG*	2.81184	0.87314	0.00128
ENSDARG00000008548	*arhgap12a*		4.13406	1.28374	0.00128
ENSDARG00000013505	*ube2kb*	*UBE2K*	3.32295	1.03232	0.00129
ENSDARG00000091917	*si:ch73-386o14.1*		4.39933	1.38770	0.00152
ENSDARG00000053753	*mff*	*MFF*	3.12331	0.98770	0.00157
ENSDARG00000071107	*wnt7bb*	*WNT7B*	4.06430	1.28657	0.00158
ENSDARG00000026865	*fam107b*	*FAM107B*	2.55375	0.80962	0.00161
ENSDARG00000018272	*wdr33*	*WDR33*	3.48630	1.10922	0.00167

ENSDARG00000062987	*tyw1*	*TYW1*	3.85641	1.23650	0.00182
ENSDARG00000104789		*KIAA0141*	3.97793	1.28532	0.00197
ENSDARG00000030440	*rsrp1*		2.90185	0.93956	0.00201
ENSDARG00000031952	*mb*	*MB*	3.87995	1.25770	0.00204
ENSDARG00000041078	*chka*	*CHKA*	3.56991	1.15888	0.00207
ENSDARG00000033950	*lamb2l*		2.73198	0.88842	0.00210
ENSDARG00000078458	*ppp1r37*	*PPP1R37*	2.91519	0.95880	0.00236
ENSDARG00000092809	*hoxc9a*	*HOXC9*	4.27298	1.40964	0.00244
ENSDARG00000037640	*aurkb*	*AURKB*	2.59699	0.85745	0.00246
ENSDARG00000010301	*b4galt6*	*B4GALT6*	2.45189	0.81276	0.00255
ENSDARG00000052747	*gpatch3*	*GPATCH3*	2.94464	0.98125	0.00269
ENSDARG00000035521	*sfrp1a*	*SFRP1*	3.97464	1.32900	0.00278
ENSDARG00000087688	*KLHL29*		4.00937	1.34313	0.00283
ENSDARG00000069752	*ckba*	*CKB*	4.07256	1.36555	0.00286
ENSDARG00000005627	*CABZ01068209.1*	*GBP1*	3.01889	1.01403	0.00291
ENSDARG00000003270	*dhps*	*DHPS*	3.48653	1.17951	0.00312
ENSDARG00000075853	*sh3kbp1*	*SH3KBP1*	2.56670	0.87271	0.00327
ENSDARG00000018698	*carm1*	*CARM1*	2.69504	0.91877	0.00335
ENSDARG00000057853	*atp6v0ca*	*ATP6V0C*	2.00666	0.68452	0.00337
ENSDARG00000068434	*h3f3b.1*		1.99269	0.68079	0.00342
ENSDARG00000041750	*ccdc92*	*CCDC92*	2.96596	1.01664	0.00353
ENSDARG00000017744	*smc2*	*SMC2*	3.27436	1.13247	0.00384
ENSDARG00000069696	*LRIF1*	*LRIF1*	4.11461	1.42420	0.00386
ENSDARG00000032614	*msi2b*	*MSI2*	3.02429	1.05354	0.00410
ENSDARG00000041068	*got2a*	*GOT2*	3.07109	1.07143	0.00415
ENSDARG00000038141	*atf4b*	*ATF4*	2.64803	0.93231	0.00451
ENSDARG00000039757	*mcfd2*	*MCFD2*	2.67942	0.94462	0.00456
ENSDARG00000061187	*cbx5*	*CBX5*	2.23427	0.78833	0.00459
ENSDARG00000021149	*cbr1l*		3.59934	1.27009	0.00460
ENSDARG00000042368	*kif26aa*	*KIF26A*	3.68292	1.31071	0.00496
ENSDARG00000054864	*aplp2*	*SPINT3*	2.38178	0.84902	0.00503
ENSDARG00000078179	*fndc3ba*	*FNDC3B*	2.52457	0.90295	0.00518
ENSDARG00000053257	*zgc:153733*		3.87178	1.38542	0.00520
ENSDARG00000026582	*iscub*	*ISCU*	3.17771	1.14118	0.00536
ENSDARG00000053358	*basp1*		3.64173	1.30799	0.00537
ENSDARG00000031200	*ppp2r5cb*	*PPP2R5C*	2.33917	0.84154	0.00544
ENSDARG00000027109	*zfr*	*ZFR*	2.07976	0.75135	0.00564
ENSDARG00000043795	*arhgdia*	*ARHGDIA*	1.20001	0.43410	0.00570
ENSDARG00000100813	*nudt12*	*NUDT12*	4.12354	1.49826	0.00592
ENSDARG00000054063	*arpc4*	*ARPC4*	1.74031	0.63423	0.00607
ENSDARG00000037760	*uncx4.1*	*UNCX*	3.46383	1.26492	0.00617
ENSDARG00000092752	*si:ch211-233a1.4*		2.96973	1.08822	0.00635
ENSDARG00000100752	*RAB21*	*RAB21*	3.05517	1.12479	0.00660

ENSDARG00000057577	*mbtps2*	*MBTPS2*	3.29768	1.21522	0.00665
ENSDARG00000074656	*ctssb.1*	*CTSS*	2.61687	0.96532	0.00671
ENSDARG00000100513	*rps27.2*	*RPS27L*	2.22941	0.82704	0.00703
ENSDARG00000052728	*sltm*	*SLTM*	1.59747	0.59803	0.00756
ENSDARG00000012513	*sdcbp2*	*SDCBP2*	2.15416	0.81539	0.00824
ENSDARG00000045417	*fam49bb*	*FAM49B*	2.25487	0.85663	0.00848
ENSDARG00000078014	*pacsin2*	*PACSIN2*	1.99375	0.75921	0.00864
ENSDARG00000100823	*selk*	*SELK*	2.13327	0.81368	0.00875
ENSDARG00000053668	*stag2b*	*STAG2*	1.99237	0.76129	0.00887
ENSDARG00000101755	*prkar1ab*		2.74281	1.04886	0.00892
ENSDARG00000000857	*mapk14a*	*MAPK14*	2.78500	1.06519	0.00893
ENSDARG00000056653	*fhl1b*	*FHL1*	3.16148	1.21008	0.00898
ENSDARG00000032619	*tob1a*	*TOB1*	2.20784	0.84775	0.00921
ENSDARG00000052000	*cav2*	*CAV2*	3.07262	1.18515	0.00953
ENSDARG00000053291	*pnrc2*	*PNRC2*	1.28908	0.49792	0.00963
ENSDARG00000040812	*ncf4*	*NCF4*	2.20765	0.85344	0.00969
ENSDARG00000069763	*etv5a*	*ETV5*	2.31028	0.89377	0.00974
ENSDARG00000056027	*hoxb8a*		2.82312	1.09475	0.00992

Table A3: Differentially upregulated genes in mutant erythrocytes. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000003383	*asb8*	*ASB8*	-6.83710	1.90899	0.00034
ENSDARG00000069464	*cox7a1*	*COX7A1*	-5.66943	1.75500	0.00124
ENSDARG00000036305	*phf23b*	*PHF23*	-6.32857	1.98530	0.00143
ENSDARG00000067717	*tomm7*	*TOMM7*	-5.29316	1.68174	0.00165
ENSDARG00000101216	*smarca2*	*SMARCA2*	-4.39017	1.45736	0.00259
ENSDARG00000008904	*meaf6*	*MEAF6*	-6.30431	2.09555	0.00263
ENSDARG00000015757	*tmem50a*	*TMEM50A*	-5.34355	1.79126	0.00285
ENSDARG00000069846	*zgc:162944*		-4.70562	1.58966	0.00308
ENSDARG00000062058	*slc12a7b*	*SLC12A7*	-5.68102	1.93342	0.00330
ENSDARG00000038576	*ube2d1b*	*UBE2D1*	-5.32100	1.82133	0.00348
ENSDARG00000098066	*si:dkey-92c21.2*		-6.05443	2.10732	0.00407
ENSDARG00000017366	*prdm4*	*PRDM4*	-5.12304	1.78681	0.00414
ENSDARG00000098924	*suz12b*	*SUZ12*	-6.07905	2.18416	0.00538
ENSDARG00000041878	*rab11ba*	*RAB11B*	-3.93368	1.42136	0.00565
ENSDARG00000101275	*si:dkey-117a16.1*		-5.88954	2.15396	0.00625
ENSDARG00000102012	*ppp1cbl*		-4.55539	1.67602	0.00657
ENSDARG00000058358	*krt8*	*KRT79*	-4.29423	1.58676	0.00680
ERCC-00071	*ERCC-00071*	*NA*	-4.59143	1.70669	0.00714
ENSDARG00000061985	*gabarapa*	*GABARAP*	-4.51065	1.70307	0.00808
ENSDARG00000063612	*psma1*	*RP11-140L24.4*	-3.86253	1.45897	0.00811

ENSDARG00000019420	*rbm47*	*RBM47*	-3.61850	1.37779	0.00863
ENSDARG00000035557	*antxr1c*	*ANTXRL*	-4.70837	1.79382	0.00867
ENSDARG00000102555	*etnk1*	*ETNK1*	-3.57391	1.36175	0.00868
ENSDARG00000101560	*si:dkey-33i11.9*		-4.62968	1.76788	0.00882
ENSDARG00000063636	*galnt11*	*GALNT11*	-4.56470	1.75975	0.00949
ENSDARG00000035751	*ipo7*	*IPO7*	-3.93842	1.52619	0.00986

Table A4: Differentially downregulated genes in mutant erythrocytes. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000056491	*ikzf5*	*IKZF5*	8.70139	1.95499	0.00001
ENSDARG00000006240	*slc27a1a*	*SLC27A1*	8.55867	2.01008	0.00002
ENSDARG00000014592	*DENND1A*	*DENND1A*	8.13607	1.95528	0.00003
ENSDARG00000024295	*slc11a2*	*SLC11A2*	6.79441	1.63499	0.00003
ENSDARG00000045399	*cct5*	*CCT5*	7.55616	1.83902	0.00004
ENSDARG00000029510	*timm17a*	*TIMM17A*	7.52138	1.91090	0.00008
ENSDARG00000034616	*mlf2*	*MLF2*	7.83002	2.01881	0.00011
ENSDARG00000041586	*dhx40*	*DHX40*	7.81632	2.03482	0.00012
ENSDARG00000069031	*plac8l1*	*PLAC8L1*	7.93215	2.07402	0.00013
ENSDARG00000012642	*ZNF208*		7.72797	2.04018	0.00015
ENSDARG00000034768	*phf3*	*PHF3*	7.55503	2.00503	0.00016
ENSDARG00000053646	*adrb3a*	*ADRB3*	7.82148	2.08331	0.00017
ENSDARG00000033489	*ube2j1*	*UBE2J1*	6.63239	1.81152	0.00025
ENSDARG00000075014	*sqstm1*	*SQSTM1*	6.03094	1.64747	0.00025
ENSDARG00000014582	*exoc3*	*EXOC3*	7.64940	2.09107	0.00025
ENSDARG00000099999	*tcf3b*	*TCF3*	6.63893	1.82170	0.00027
ENSDARG00000013931	*eif3m*	*EIF3M*	5.06388	1.41376	0.00034
ENSDARG00000032469	*ampd3b*	*AMPD3*	6.94325	1.94920	0.00037
ENSDARG00000086416	*med11*	*MED11*	7.44857	2.09352	0.00037
ENSDARG00000025338	*hagh*	*HAGH*	6.33132	1.79263	0.00041
ENSDARG00000071037	*pex13*	*PEX13*	7.02328	2.00980	0.00047
ENSDARG00000076483	*zgc:198241*		6.76603	1.93949	0.00049
ENSDARG00000016011	*gpcpd1*	*GPCPD1*	7.26676	2.09360	0.00052
ENSDARG00000001220	*mycbp2*	*MYCBP2*	4.90150	1.42115	0.00056
ENSDARG00000104789		*KIAA0141*	7.21676	2.09622	0.00058
ENSDARG00000104842	*CABZ01085069.1*		6.87110	2.00501	0.00061
ENSDARG00000013946	*ivns1abpb*	*IVNS1ABP*	6.73798	1.97013	0.00063
ENSDARG00000044619	*birc2*	*BIRC3*	5.30233	1.55471	0.00065
ENSDARG00000071570	*CCDC71*	*CCDC71*	7.33678	2.15912	0.00068
ENSDARG00000004161	*ik*	*IK*	5.60836	1.65116	0.00068
ENSDARG00000057353	*EHBP1L1*		5.51913	1.65046	0.00083
ENSDARG00000063229	*xpo1b*	*XPO1*	5.34107	1.60831	0.00090

ENSDARG00000020101	*psmc2*	*PSMC2*	7.09585	2.15684	0.00100
ENSDARG00000012818	*csnk2a2a*	*CSNK2A2*	6.74061	2.07567	0.00116
ENSDARG00000038154	*isca2*	*ISCA2*	6.15765	1.89681	0.00117
ENSDARG00000098391	*psmg3*	*PSMG3*	6.56397	2.02779	0.00121
ENSDARG00000013804	*capns1b*	*CAPNS2*	5.68198	1.75815	0.00123
ENSDARG00000079238	*trim59*	*TRIM59*	6.79111	2.10157	0.00123
ENSDARG00000019188	*ube2l3a*	*UBE2L3*	6.97128	2.15738	0.00123
ENSDARG00000052553	*lig3*	*LIG3*	6.97303	2.15859	0.00124
ENSDARG00000104726	*si:ch211-162e15.3*	*NCBP2-AS2*	6.71034	2.09207	0.00134
ENSDARG00000077405			5.41063	1.69949	0.00145
ENSDARG00000025858	*GOLM1*	*GOLM1*	6.65182	2.09286	0.00148
ENSDARG00000090600	*si:ch211-213a13.1*		5.34589	1.69011	0.00156
ENSDARG00000070426	*chac1*	*CHAC1*	6.15706	1.95205	0.00161
ENSDARG00000018742	*psme4b*	*PSME4*	6.30296	2.00626	0.00168
ENSDARG00000018190	*asna1*	*ASNA1*	5.60296	1.79274	0.00178
ENSDARG00000011934	*gyg1a*	*GYG1*	5.43643	1.74558	0.00184
ENSDARG00000012340	*ptpn11b*		6.05192	1.94516	0.00186
ENSDARG00000103343	*TRIM14*		6.48735	2.09509	0.00196
ENSDARG00000103203	*vps28*	*VPS28*	6.19986	2.00621	0.00200
ENSDARG00000086618	*psma3*	*PSMA3*	6.05933	1.96639	0.00206
ENSDARG00000044373	*atg4c*	*ATG4C*	6.40916	2.08831	0.00215
ENSDARG00000103200	*gpr107*	*GPR107*	6.61075	2.15886	0.00220
ENSDARG00000068698	*psenen*	*AC002398.9*	6.59402	2.15501	0.00221
ENSDARG00000023217	*crema*		6.33146	2.07327	0.00226
ENSDARG00000071013	*arl6ip6*	*ARL6IP6*	6.35279	2.08772	0.00234
ENSDARG00000012577	*waca*	*WAC*	5.43223	1.80088	0.00256
ENSDARG00000093058	*C4H7orf73*	*C7orf73*	5.51835	1.83417	0.00262
ENSDARG00000099896	*si:ch211-261o3.3*		6.27645	2.09196	0.00270
ENSDARG00000104609	*crebbpa*	*CREBBP*	5.74357	1.93843	0.00305
ENSDARG00000053753	*mff*	*MFF*	5.53661	1.87577	0.00316
ENSDARG00000024092	*lmbr1*	*LMBR1*	6.33920	2.16069	0.00335
ENSDARG00000036394	*pop7*	*POP7*	6.12563	2.08800	0.00335
ENSDARG00000102379	*FQ378013.1*		6.14299	2.09509	0.00337
ENSDARG00000020926	*creb3l3l*	*CREB3*	5.54479	1.89895	0.00350
ENSDARG00000103134	*CKN149934.1H1orf52*	*C1orf52*	5.49387	1.88589	0.00358
ENSDARG00000042728	*plaa*	*PLAA*	6.24020	2.16254	0.00391
ENSDARG00000044183	*prkab1a*	*PRKAB1*	5.11126	1.77144	0.00391
ENSDARG00000103403	*sar1b*	*SAR1B*	3.72915	1.29578	0.00400
ENSDARG00000037012	*slc3a2b*	*SLC3A2*	5.99717	2.08652	0.00405
ENSDARG00000059886	*fam222ba*	*FAM222B*	5.95881	2.08447	0.00425
ENSDARG00000074759	*ccar1*	*CCAR1*	6.15526	2.15986	0.00437
ENSDARG00000098123	*GNG12*		6.14288	2.16116	0.00448
ENSDARG00000062315	*sik2b*	*SIK2*	5.31550	1.87361	0.00455

ENSDARG00000040463	*si:dkey-185e18.6*	*SLC25A51*	5.32278	1.87661	0.00456
ENSDARG00000043562	*zgc:65997*		5.52189	1.94801	0.00459
ENSDARG00000078042	*il10rb*	*IFNAR1*	4.36968	1.54589	0.00470
ENSDARG00000104216	*LOX*		4.70709	1.66623	0.00473
ENSDARG00000071395	*camk2g1*	*CAMK2G*	4.84306	1.72252	0.00493
ENSDARG00000031435	*zgc:56493*	*TXNDC8*	3.74212	1.35059	0.00559
ENSDARG00000076496	*frmd4bb*	*FRMD4B*	5.08490	1.83585	0.00561
ENSDARG00000098355	*atp5h*	*ATP5H*	3.50914	1.26770	0.00564
ENSDARG00000096091	*si:dkey-29j8.3*		5.94788	2.16083	0.00591
ENSDARG00000026072	*pdcd5*	*PDCD5*	5.32647	1.93716	0.00597
ENSDARG00000036482	*hexim1*	*HEXIM1*	5.39264	1.97163	0.00624
ENSDARG00000093538	*si:dkey-182g1.5*		5.90071	2.15838	0.00626
ENSDARG00000104373	*C4H11orf98*	*C11orf98*	5.89174	2.15932	0.00636
ENSDARG00000036625	*polr2f*	*POLR2F*	5.13030	1.90239	0.00700
ENSDARG00000040930	*deptor*	*DEPTOR*	4.71366	1.74995	0.00707
ENSDARG00000090054	*znf318*	*ZNF318*	4.59219	1.70692	0.00714
ENSDARG00000041533	*ccdc130*	*CCDC130*	5.79937	2.16610	0.00742
ENSDARG00000074677	*frem3*		5.76068	2.16168	0.00770
ENSDARG00000098317	*dync1li1*	*DYNC1LI1*	4.32309	1.62445	0.00778
ENSDARG00000004160	*reep3b*	*REEP3*	4.45408	1.68558	0.00823
ENSDARG00000024874	*dock4b*	*DOCK4*	4.63719	1.75702	0.00831
ENSDARG00000063466	*rab43*	*ISY1*	5.88273	2.23578	0.00851
ENSDARG00000061124	*srpr*	*SRPRA*	5.28088	2.00833	0.00855
ENSDARG00000051955	*brms1*	*BRMS1*	4.88853	1.86231	0.00867
ENSDARG00000058041	*ndufa8*	*NDUFA8*	4.03337	1.53853	0.00875
ENSDARG00000101828	*si:ch211-76m11.11*		4.27874	1.63690	0.00895
ENSDARG00000044601	*rtn4a*	*RTN4*	4.08898	1.56449	0.00896
ENSDARG00000034916	*hat1*	*HAT1*	4.82198	1.85254	0.00924
ENSDARG00000098466	*ube2a*	*UBE2A*	5.01667	1.92801	0.00927
ENSDARG00000100109	*trappc2l*	*TRAPPC2L*	5.00664	1.92431	0.00927
ENSDARG00000069135	*ppp1r15a*	*PPP1R15A*	5.15301	1.98975	0.00960
ENSDARG00000057826	*si:ch73-61d6.3*		3.92563	1.52312	0.00996

Table A5: Differentially upregulated genes in mutant monocytes. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000020711	*rrm2*	*RRM2*	-3.44624	0.86919	0.00007
ENSDARG00000060771	*map7d3*		-2.56457	0.65275	0.00009
ENSDARG00000102291	*eef1da*	*EEF1D*	-1.13373	0.30387	0.00019
ENSDARG00000101406	*rplp2*		-1.05110	0.28291	0.00020
ENSDARG00000103436	*BX548028.1*		-2.12402	0.59261	0.00034
ENSDARG00000030938	*fermt3b*	*FERMT3*	-2.20338	0.62982	0.00047

ENSDARG00000006691	*rpl12*	*RPL12*	-0.84192	0.24670	0.00064
ENSDARG00000101813	*nap1l1*	*NAP1L1*	-1.03719	0.30909	0.00079
ENSDARG00000058030	*hspa14*	*HSPA14*	-2.34247	0.70155	0.00084
ENSDARG00000019791	*prmt3*	*PRMT3*	-4.09064	1.23645	0.00094
ENSDARG00000030278	*idh3a*	*IDH3A*	-2.76649	0.87102	0.00149
ENSDARG00000101430	*si:dkeyp-3b12.7*		-2.83940	0.90711	0.00175
ENSDARG00000098385	*si:dkeyp-3b12.6*		-2.83940	0.90711	0.00175
ENSDARG00000099380	*rpl13*	*RPL13*	-1.24589	0.40324	0.00200
ENSDARG00000076858	*C10H8orf4*	*C8orf4*	-1.89253	0.61973	0.00226
ENSDARG00000070228	*cdk6*	*CDK6*	-1.44857	0.47720	0.00240
ENSDARG00000006818	*urod*	*UROD*	-1.70911	0.56356	0.00242
ENSDARG00000031795	*abcf1*	*ABCF1*	-1.53513	0.51231	0.00273
ENSDARG00000103846	*hspa5*		-2.48918	0.83298	0.00281
ENSDARG00000041811	*rps25*	*RPS25*	-0.87198	0.29267	0.00289
ENSDARG00000040440	*snrpd2*	*SNRPD2*	-1.92527	0.64669	0.00291
ENSDARG00000056600	*papss2b*	*PAPSS2*	-2.33489	0.81776	0.00430
ENSDARG00000021339	*cpa5*	*CPA1*	-2.05901	0.72302	0.00440
ENSDARG00000104353	*nop58*	*NOP58*	-1.36876	0.48244	0.00455
ENSDARG00000003098	*kdm5bb*	*KDM5B*	-2.18267	0.77047	0.00461
ENSDARG00000088959	*pdxka*	*PDXK*	-1.79416	0.63418	0.00467
ENSDARG00000037017	*ube4b*	*UBE4B*	-1.65157	0.58530	0.00478
ENSDARG00000074242	*serbp1a*	*SERBP1*	-0.87561	0.31057	0.00481
ENSDARG00000011405	*rps9*	*RPS9*	-0.84360	0.30042	0.00498
ENSDARG00000019810	*nfe2l3*	*NFE2L3*	-1.90103	0.67873	0.00510
ENSDARG00000061591	*abcb10*	*ABCB10*	-2.07206	0.74475	0.00540
ENSDARG00000099970	*malat1*		-0.87279	0.31457	0.00553
ENSDARG00000032013	*pafah1b1a*	*PAFAH1B1*	-1.67339	0.60985	0.00607
ENSDARG00000023532	*pinx1*	*PINX1*	-2.05005	0.74858	0.00617
ENSDARG00000055713	*fmnl1a*	*FMNL1*	-1.19307	0.43658	0.00628
ENSDARG00000096770	*si:ch73-21k16.6*		-1.86988	0.68531	0.00636
ENSDARG00000104674	*CABZ01075268.1*		-1.19534	0.44262	0.00692
ENSDARG00000101637	*ccnd1*	*CCND1*	-1.13444	0.42014	0.00693
ENSDARG00000104011	*rps17*		-0.86633	0.32114	0.00698
ENSDARG00000045297	*phb2a*	*PHB2*	-2.01947	0.75271	0.00730
ENSDARG00000100677	*tmem205*	*TMEM205*	-1.67572	0.62695	0.00752
ENSDARG00000074245	*spen*	*SPEN*	-1.59626	0.60519	0.00835
ENSDARG00000095556	*si:dkey-238c7.12*		-0.62718	0.23817	0.00845
ENSDARG00000061338	*ddx6*	*DDX6*	-2.00092	0.76353	0.00878
ENSDARG00000031136	*moxd1*	*MOXD1*	-1.85914	0.71249	0.00907
ENSDARG00000080009	*bahcc1*	*BAHCC1*	-1.49879	0.58046	0.00982
ENSDARG00000063626	*ddx21*	*DDX21*	-1.14781	0.44461	0.00983

Appendix

Table 6: A Differentially downregulated genes in mutant monocytes. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000075980	*tmem125b*	*TMEM125*	5.81446	1.10879	0.00000
ENSDARG00000078069	*rrm2*	*RRM2*	5.91399	1.14539	0.00000
ENSDARG00000029445	*EIF1B*	*EIF1B*	3.72738	0.75153	0.00000
ENSDARG00000068246	*plcb3*	*PLCB3*	3.58165	0.73243	0.00000
ENSDARG00000009961	*rundc3b*	*RUNDC3B*	3.82157	0.78281	0.00000
ENSDARG00000071601	*pvalb9*	*OCM*	5.25333	1.07779	0.00000
ENSDARG00000101840	*U3*		4.85935	1.00439	0.00000
ENSDARG00000097202	*si:ch211-141i4.3*		3.60741	0.75358	0.00000
ENSDARG00000068456	*tmem91*	*TMEM91*	4.32260	0.91603	0.00000
ENSDARG00000104635	*si:busm1-194e12.12*	*HLA-DRB1*	4.25329	0.91949	0.00000
ENSDARG00000087732	*Metazoa_SRP*	*RN7SL480P*	4.76687	1.04462	0.00001
ENSDARG00000008153	*serinc5*	*SERINC5*	2.87885	0.64300	0.00001
ENSDARG00000075421	*pttg1*		2.95234	0.66814	0.00001
ENSDARG00000096403	*si:dkey-153m14.1*		3.13641	0.71024	0.00001
ENSDARG00000007808	*zdhhc16a*	*ZDHHC16*	3.12222	0.70810	0.00001
ENSDARG00000079102	*PRRG3*	*PRRG3*	3.65787	0.83205	0.00001
ENSDARG00000013892	*sorl1*	*SORL1*	4.47785	1.05123	0.00002
ENSDARG00000092337	*gas5*		2.47073	0.58083	0.00002
ENSDARG00000099091	*si:ch73-158n7.2*		2.84406	0.67574	0.00003
ENSDARG00000081270	*rn7sk*	*RN7SKP90*	4.56857	1.09563	0.00003
ENSDARG00000091402	*eif2b1*	*EIF2B1*	3.10408	0.74991	0.00003
ENSDARG00000102311	*BX572103.3*		3.02770	0.73260	0.00004
ENSDARG00000083431	*U4*	*RNU4-2*	5.30289	1.28466	0.00004
ENSDARG00000099417	*CT027611.1*		2.89756	0.71034	0.00005
ENSDARG00000003017	*zgc:55512*		2.73266	0.67353	0.00005
ENSDARG00000071139	*zgc:64065*		4.54963	1.12680	0.00005
ENSDARG00000027807	*fynrk*	*FRK*	3.27901	0.81818	0.00006
ENSDARG00000040266	*sox19b*		2.76711	0.69175	0.00006
ENSDARG00000099359	*BX649497.2*		4.92607	1.23822	0.00007
ENSDARG00000099625	*si:dkey-97l11.1*		2.85659	0.72077	0.00007
ENSDARG00000100513	*rps27.2*	*RPS27L*	2.52503	0.63960	0.00008
ENSDARG00000104666	*BX511215.2*	*ZNF852*	4.44874	1.13166	0.00008
ENSDARG00000103976	*BX544877.1*		2.93445	0.74695	0.00009
ENSDARG00000102436	*WIZ*	*WIZ*	3.41309	0.89577	0.00014
ENSDARG00000028295	*mkrn4*		2.55026	0.67115	0.00014
ENSDARG00000102805	*cyp2aa12*		2.90624	0.77187	0.00017
ENSDARG00000092407	*TRIM35*		3.50623	0.93314	0.00017
ENSDARG00000021811	*calm1a*	*CALM1*	3.17258	0.84453	0.00017
ENSDARG00000093743	*si:ch73-55c23.2*		3.90244	1.04293	0.00018

ENSDARG00000019260	*dhrs9*		4.76988	1.27514	0.00018
ENSDARG00000101770	*ACEA_U3*	*SNORD3D*	4.88153	1.31394	0.00020
ENSDARG00000025436	*msrb1a*	*MSRB1*	3.01038	0.81042	0.00020
ENSDARG00000038754	*plk3*	*PLK3*	2.11695	0.57288	0.00022
ENSDARG00000015201	*pcmt*	*PCMT1*	2.70988	0.73497	0.00023
ENSDARG00000101777	*pdlim5a*	*PDLIM5*	2.10829	0.57320	0.00024
ENSDARG00000091137	*CABZ01079258.1*		2.33340	0.63461	0.00024
ENSDARG00000089213	*adam15*	*ADAM15*	2.65661	0.72295	0.00024
ENSDARG00000105109	*si:ch1073-329i9.1*		3.11309	0.84783	0.00024
ENSDARG00000058365	*hspb8*	*HSPB8*	2.46717	0.67491	0.00026
ENSDARG00000068787	*slc6a17*	*SLC6A17*	2.59182	0.70906	0.00026
ENSDARG00000031496	*ap1ar*	*AP1AR*	2.20766	0.60797	0.00028
ENSDARG00000042824	*nfe2l2a*	*NFE2L2*	2.23599	0.61790	0.00030
ENSDARG00000063436	*RPH3A*		3.83122	1.05892	0.00030
ENSDARG00000078966	*rbm15b*	*RBM15B*	3.13042	0.87118	0.00033
ENSDARG00000094308	*si:dkey-247i3.6*		3.30516	0.92015	0.00033
ENSDARG00000083389	*SNORA57*	*SNORA57*	4.06864	1.14254	0.00037
ENSDARG00000038235	*pkdccb*		2.97945	0.83746	0.00037
ENSDARG00000040190	*qdpra*		2.28414	0.64444	0.00039
ENSDARG00000038668	*gbp1*	*GBP1*	3.66357	1.03427	0.00040
ENSDARG00000071345	*mgst2*	*MGST2*	4.12755	1.16722	0.00041
ENSDARG00000020811	*efemp2b*	*EFEMP2*	2.17937	0.61643	0.00041
ENSDARG00000098315	*cyp1a*	*CYP1A1*	3.60583	1.02078	0.00041
ENSDARG00000097615	*si:ch211-108d22.2*		2.92815	0.82906	0.00041
ENSDARG00000043854	*ppil4*	*PPIL4*	3.67245	1.04372	0.00043
ENSDARG00000105196	*CU929447.2*		3.51304	0.99993	0.00044
ENSDARG00000071082	*p4ha1b*	*P4HA1*	3.10885	0.88642	0.00045
ENSDARG00000100776	*syt8*	*SYT8*	3.58966	1.03220	0.00051
ENSDARG00000070688	*ncalda*	*NCALD*	2.08606	0.60047	0.00051
ENSDARG00000061974	*grhl2b*	*GRHL2*	3.70947	1.07415	0.00055
ENSDARG00000030687	*phka2*	*PHKA2*	2.58391	0.74836	0.00055
ENSDARG00000028119	*sumo3a*	*SUMO3*	2.03891	0.59064	0.00056
ENSDARG00000095477	*si:ch1073-291l11.2*		2.96967	0.86226	0.00057
ENSDARG00000042621	*cryaba*	*CRYAB*	1.76291	0.51238	0.00058
ENSDARG00000102549	*CR450729.2*		2.25224	0.65955	0.00064
ENSDARG00000097844	*si:ch1073-488c15.2*		2.10526	0.61866	0.00067
ENSDARG00000061083	*igl4v8*	*IGKV2-40*	4.01903	1.18105	0.00067
ENSDARG00000087457	*ecscr*	*ECSCR*	2.54152	0.75094	0.00071
ENSDARG00000011312	*stk3*	*STK3*	1.68585	0.50051	0.00076
ENSDARG00000101337	*zgc:103700*	*HLA-DRB1*	2.39392	0.71807	0.00086
ENSDARG00000086495	*BX546500.1*		3.56326	1.06952	0.00086
ENSDARG00000040725	*zgc:114130*		2.12957	0.64248	0.00092
ENSDARG00000078024	*CU570881.1*		2.83124	0.85458	0.00092

ENSDARG00000101828	*si:ch211-76m11.11*		2.51334	0.75916	0.00093
ENSDARG00000104572	*si:dkey-35h6.1*		1.96480	0.59922	0.00104
ENSDARG00000100731	*slc27a2b*	*SLC27A2*	2.97019	0.91036	0.00110
ENSDARG00000039131	*atp1a1a.3*	*ATP1A1*	2.50603	0.76815	0.00110
ENSDARG00000008363	*mcl1b*	*MCL1*	1.89285	0.58162	0.00114
ENSDARG00000104715	*zgc:171422*		3.66249	1.12770	0.00116
ENSDARG00000098557	*BX323564.1*		1.85370	0.57157	0.00118
ENSDARG00000104636	*si:dkey-112a7.4*		3.22542	0.99946	0.00125
ENSDARG00000079742	*mcf2l2*	*MCF2L2*	3.13285	0.97448	0.00130
ENSDARG00000044254	*anxa3b*	*ANXA3*	2.97453	0.92592	0.00132
ENSDARG00000013990	*ube2q2*	*UBE2Q2*	1.95069	0.60749	0.00132
ENSDARG00000033138	*lyrm2*	*LYRM2*	1.63168	0.50865	0.00134
ENSDARG00000005841	*tnni2a.2*	*TNNI2*	1.49179	0.46700	0.00140
ENSDARG00000059442	*smtnb*	*SMTNL1*	2.98275	0.93579	0.00144
ENSDARG00000044339	*rp2*	*RP2*	2.81235	0.88378	0.00146
ENSDARG00000103639	*si:dkey-36i7.3*		2.32581	0.73134	0.00147
ENSDARG00000027017	*ppp2r5a*	*PPP2R5A*	2.91509	0.91685	0.00148
ENSDARG00000094110	*si:ch211-284e20.4*		3.10148	0.98634	0.00166
ENSDARG00000016213	*BX088711.1*		2.37347	0.75489	0.00167
ENSDARG00000018478	*agxtb*	*AGXT*	2.75209	0.87559	0.00167
ENSDARG00000096655	*si:ch211-184m19.3*		1.52186	0.48627	0.00175
ENSDARG00000103925	*5S_rRNA*		3.94722	1.26256	0.00177
ENSDARG00000025033	*stx5a*	*STX5*	3.12122	1.00402	0.00188
ENSDARG00000026990	*klhl6*	*KLHL6*	2.15842	0.69549	0.00191
ENSDARG00000062577	*arhgap35a*	*ARHGAP35*	2.05655	0.66282	0.00192
ENSDARG00000103430	*si:ch73-226i7.1*		3.40356	1.09759	0.00193
ENSDARG00000087762	*RAB27B*	*RAB27B*	1.71326	0.55264	0.00193
ENSDARG00000079834	*kdf1a*	*KDF1*	3.09499	0.99896	0.00195
ENSDARG00000092457	*si:ch211-207c6.6*		2.41437	0.78082	0.00199
ENSDARG00000089957	*lgi2a*	*LGI2*	2.45783	0.79495	0.00199
ENSDARG00000034700	*vegfab*	*VEGFA*	3.36542	1.09077	0.00203
ENSDARG00000041340	*mrpl51*	*MRPL51*	2.18989	0.71057	0.00206
ENSDARG00000005897	*dera*	*DERA*	2.96361	0.96247	0.00208
ENSDARG00000062565	*kcnh4a*	*KCNH4*	3.09797	1.00933	0.00215
ENSDARG00000070545	*top1l*	*TOP1*	1.14306	0.37517	0.00231
ENSDARG00000002897	*INPP1*	*INPP1*	2.07384	0.68283	0.00239
ENSDARG00000105100	*si:dkey-3h2.2*		1.97695	0.65136	0.00240
ENSDARG00000074201	*flna*	*FLNA*	1.25933	0.41531	0.00243
ENSDARG00000013014	*or101-1*	*OR2AT4*	2.15577	0.71766	0.00267
ENSDARG00000100446	*si:ch211-286o17.1*		2.72464	0.90852	0.00271
ENSDARG00000002305	*me3*	*ME3*	2.68337	0.89811	0.00281
ENSDARG00000104084	*GPRIN2*	*GPRIN2*	2.10592	0.70607	0.00286
ENSDARG00000054666	*pgpep1*		1.41178	0.47345	0.00286

ENSDARG00000089142	*zgc:113119*		3.73601	1.25407	0.00289
ENSDARG00000100633	*APH1A*	*APH1A*	1.23375	0.41530	0.00297
ENSDARG00000102750	*cdh1*	*CDH1*	2.17575	0.73288	0.00299
ENSDARG00000104282	*DCC*	*DCC*	3.11086	1.04980	0.00304
ENSDARG00000023712	*mao*	*MAOA*	2.82992	0.96019	0.00321
ENSDARG00000013144	*atp1b1a*	*ATP1B1*	1.08498	0.36936	0.00331
ENSDARG00000011065	*camk2b1*	*CAMK2B*	2.38594	0.81363	0.00336
ENSDARG00000098099	*zgc:64065*		2.09562	0.71490	0.00338
ENSDARG00000071409	*arl3*	*ARL3*	2.86959	0.98181	0.00347
ENSDARG00000053136	*b2m*	*B2M*	1.39791	0.47872	0.00350
ENSDARG00000011555	*spag7*	*SPAG7*	1.64065	0.56224	0.00352
ENSDARG00000068738	*cox5b2*	*COX5B*	2.62274	0.90115	0.00361
ENSDARG00000103380	*LPIN1*		1.53744	0.52853	0.00363
ENSDARG00000059461	*mepce*		1.92114	0.66046	0.00363
ENSDARG00000099546	*kynu*	*KYNU*	2.57049	0.88654	0.00374
ENSDARG00000035120	*GLDC*	*GLDC*	1.38382	0.48015	0.00395
ENSDARG00000067545	*adam19b*	*ADAM19*	1.95045	0.67831	0.00403
ENSDARG00000097300	*si:ch211-242h13.4*		3.17615	1.10676	0.00411
ENSDARG00000079034	*si:dkey-19b23.11*		1.91472	0.67157	0.00436
ENSDARG00000033046	*ccni2*	*CCNI2*	2.84090	0.99650	0.00436
ENSDARG00000099633	*si:dkey-186o21.1*	*SEC14L3*	2.20117	0.77270	0.00439
ENSDARG00000094133	*wu:fc21g02*		1.88009	0.66321	0.00458
ENSDARG00000036117	*mchr2*		1.78449	0.63216	0.00476
ENSDARG00000038577	*cox6c*	*COX6C*	1.63056	0.57819	0.00480
ENSDARG00000069632	*emp1*	*EMP1*	1.77853	0.63126	0.00484
ENSDARG00000038107	*sgcg*	*SGCG*	1.62507	0.57783	0.00492
ENSDARG00000087597	*si:dkey-51d8.3*		3.03127	1.07917	0.00497
ENSDARG00000075721	*zdhhc6*	*ZDHHC6*	2.65779	0.94692	0.00500
ENSDARG00000070951	*hmga1b*	*HMGA1*	1.91093	0.68457	0.00525
ENSDARG00000090386	*cd3eap*	*CD3EAP*	2.17912	0.78118	0.00528
ENSDARG00000102025	*znf644b*	*ZNF644*	1.91499	0.68683	0.00530
ENSDARG00000100265	*rhcgb*	*RHCG*	2.04416	0.73349	0.00532
ENSDARG00000078302	*BRINP1*	*BRINP1*	2.24778	0.80700	0.00535
ENSDARG00000095273	*cox8a*		1.39140	0.49954	0.00535
ENSDARG00000035781	*ctdnep1b*	*CTDNEP1*	1.43523	0.51553	0.00537
ENSDARG00000105300	*BX927081.1*		2.51524	0.90458	0.00543
ENSDARG00000007141	*psmc3*	*PSMC3*	2.55943	0.92434	0.00562
ENSDARG00000078309	*fam83b*	*FAM83B*	2.05988	0.74468	0.00567
ENSDARG00000068708	*ifrd1*	*IFRD1*	2.68506	0.97287	0.00578
ENSDARG00000052652	*fermt1*	*FERMT1*	2.05974	0.74808	0.00590
ENSDARG00000058476	*stc1l*	*STC1*	3.37701	1.22711	0.00592
ENSDARG00000060841	*pik3c2a*	*PIK3C2A*	2.01006	0.73189	0.00603
ENSDARG00000098171	*zgc:162193*	*TCAF1*	3.26928	1.19619	0.00627

ENSDARG00000093448	*mpc1*	*MPC1*	1.89887	0.69572	0.00635
ENSDARG00000100374	*txnrd1*	*TXNRD3*	2.36426	0.86662	0.00637
ENSDARG00000075054	*RASGRF2*		2.08037	0.76273	0.00638
ENSDARG00000005629	*smyd2b*		2.16454	0.79362	0.00638
ENSDARG00000043436	*si:dkey-5n18.1*		2.03206	0.74733	0.00655
ENSDARG00000020345	*clasp2*	*CLASP2*	2.03636	0.74959	0.00659
ENSDARG00000095824	*si:ch211-207c7.6*		3.70881	1.36635	0.00664
ENSDARG00000007219	*actn1*	*ACTN1*	2.61721	0.96445	0.00665
ENSDARG00000076230	*atp10b*	*ATP10B*	2.12192	0.78200	0.00666
ENSDARG00000001870	*atp1a1a.4*	*ATP1A1*	1.80662	0.66599	0.00667
ENSDARG00000092281	*FLNB*		1.60401	0.59191	0.00673
ENSDARG00000093549	*sepp1a*	*SEPP1*	2.16546	0.80113	0.00687
ENSDARG00000058256	*draxin*	*DRAXIN*	1.70148	0.63214	0.00711
ENSDARG00000027063	*arpc1b*	*ARPC1B*	1.19085	0.44255	0.00713
ENSDARG00000063916	*mt-nd4l*	*MT-ND4L*	1.70996	0.63553	0.00713
ENSDARG00000102911	*FO704569.1*	*IFFO2*	2.74555	1.02454	0.00737
ENSDARG00000019128	*tpm4b*		1.25233	0.46762	0.00740
ENSDARG00000036414	*CU019646.1*	*Irs3*	1.69193	0.63204	0.00743
ENSDARG00000063253	*hecw2b*	*HECW2*	1.74760	0.65442	0.00758
ENSDARG00000068127	*si:ch211-194e1.9*		1.92270	0.72004	0.00758
ENSDARG00000030224	*ppp2ca*		1.94734	0.72939	0.00759
ENSDARG00000105287	*mpp1*	*MPP1*	1.73671	0.65086	0.00762
ENSDARG00000081535	*U1*	*Gm25481*	3.64861	1.36949	0.00772
ENSDARG00000090323	*U1*	*Gm25481*	3.64861	1.36949	0.00772
ENSDARG00000074633	*gpr35.1*	*GPR35*	3.07977	1.15756	0.00780
ENSDARG00000103401	*CR318592.1*		1.32279	0.49729	0.00781
ENSDARG00000095627	*c1qc*	*C1QC*	1.73365	0.65331	0.00796
ENSDARG00000100324	*AL627305.1*		1.80653	0.68091	0.00798
ENSDARG00000003157	*nfkbil1*	*NFKBIL1*	1.96589	0.74211	0.00807
ENSDARG00000096072	*si:dkey-29j8.4*		1.68541	0.63636	0.00808
ENSDARG00000088141	*CU570684.4*		2.26242	0.85427	0.00809
ENSDARG00000102495	*CT573234.4*		2.69978	1.02324	0.00833
ENSDARG00000038812	*e2f5*	*E2F5*	1.27729	0.48417	0.00834
ENSDARG00000059483	*tead1b*	*TEAD1*	2.22345	0.84425	0.00845
ENSDARG00000016494	*ddc*	*DDC*	1.42753	0.54379	0.00866
ENSDARG00000008723	*prkcba*	*PRKCB*	2.84274	1.08329	0.00869
ENSDARG00000101394	*fam69b*	*FAM69B*	1.46925	0.56046	0.00875
ENSDARG00000058830	*zdhhc3b*	*ZDHHC3*	1.13491	0.43310	0.00878
ENSDARG00000014101	*pyroxd2*	*PYROXD2*	2.12772	0.81416	0.00897
ENSDARG00000097197	*si:ch211-129i21.3*		1.58916	0.60887	0.00905
ENSDARG00000098488	*CABZ01118154.1*		2.42216	0.92826	0.00907
ENSDARG00000089868		*TNFRSF25*	2.45031	0.93974	0.00912
ENSDARG00000078176	*PHYHIP*	*PHYHIP*	1.58381	0.60754	0.00914

ENSDARG00000035564	*dgcr8*	*DGCR8*	1.49635	0.57442	0.00919
ENSDARG00000097760	*si:dkey-7i4.8*		2.04308	0.78593	0.00933
ENSDARG00000096319	*si:ch211-239j9.1*		1.41906	0.54591	0.00934
ENSDARG00000102956	*mmp17b*		1.70881	0.65781	0.00938
ENSDARG00000102371	*CU570769.1*		2.07912	0.80079	0.00942
ENSDARG00000099177	*dok1b*	*DOK1*	2.88063	1.10998	0.00945
ENSDARG00000021398	*mul1a*		1.44188	0.55821	0.00979
ENSDARG00000045553	*hsd17b2*	*HSD17B2*	1.39776	0.54163	0.00986
ENSDARG00000079785	*si:dkeyp-73a2.2*		1.91236	0.74186	0.00994
ENSDARG00000044345	*cyfip1*	*CYFIP1*	2.03627	0.79000	0.00995

Table A7: Differentially upregulated genes in mutant neutrophils. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000042138	*selp*	*SELL*	-3.04896	0.59785	0.00000
ENSDARG00000033444	*map4k6*	*MAP4K2*	-4.01617	0.84634	0.00000
ERCC-00042	*ERCC-00042*	*NA*	-4.75380	1.08751	0.00001
ENSDARG00000104818	*CES3*	*CES3*	-2.75097	0.63641	0.00002
ENSDARG00000043334	*ccdc6a*	*CCDC6*	-3.61066	0.87577	0.00004
ENSDARG00000067741	*itpkcb*	*ITPKC*	-2.58164	0.63857	0.00005
ENSDARG00000098628	*si:dkey-9i5.1*		-4.16795	1.05068	0.00007
ENSDARG00000002956	*si:dkey-32n7.4*	*ITGAM*	-2.79746	0.72569	0.00012
ENSDARG00000077337	*si:ch211-171h4.5*		-1.87201	0.50600	0.00022
ENSDARG00000087564	*nek11*	*NEK11*	-4.06295	1.11601	0.00027
ENSDARG00000103924	*zgc:154055*		-1.67541	0.46136	0.00028
ENSDARG00000003216	*anxa2a*	*ANXA2*	-3.71167	1.02899	0.00031
ENSDARG00000100114	*sf3a3*	*SF3A3*	-2.97638	0.84447	0.00042
ENSDARG00000031683	*fosab*	*FOS*	-3.51773	1.03765	0.00070
ERCC-00046	*ERCC-00046*	*NA*	-1.46786	0.43507	0.00074
ENSDARG00000036094	*PIAS1*		-3.43700	1.03159	0.00086
ENSDARG00000104202	*CABZ01077978.1*		-2.66810	0.80729	0.00095
ENSDARG00000097135	*si:ch211-63i2.1*		-3.10475	0.94695	0.00104
ENSDARG00000045224	*glipr1a*	*GLIPR1L1*	-1.36832	0.41844	0.00108
ENSDARG00000070606	*ikbke*	*IKBKE*	-3.70728	1.14586	0.00121
ENSDARG00000105273	*si:ch73-295i22.2*		-2.10778	0.65916	0.00139
ENSDARG00000030110	*myod1*	*MYOD1*	-2.34943	0.73557	0.00140
ENSDARG00000090730	*zgc:158446*	*XXbac-BPG116M5.17*	-1.67997	0.52725	0.00144
ENSDARG00000060655	*BX005421.1*	*Gvin1*	-3.66100	1.16496	0.00167
ENSDARG00000078547	*si:ch211-264f5.2*		-2.62420	0.84418	0.00188
ENSDARG00000095699	*dicp1.9*		-3.56946	1.16318	0.00215
ERCC-00113	*ERCC-00113*	*NA*	-1.63787	0.53447	0.00218
ENSDARG00000094097	*si:ch211-209a2.1*		-2.00000	0.65453	0.00225

Appendix

ENSDARG00000007975	*fth1b*	*FTMT*	-2.68808	0.88747	0.00245
ENSDARG00000101553	*CABZ01111555.1*		-1.76871	0.59125	0.00278
ENSDARG00000010756	*nod2*	*NOD2*	-3.49984	1.17295	0.00285
ENSDARG00000105061	*si:ch73-261i21.5*		-3.47504	1.17061	0.00299
ENSDARG00000090767	*si:ch211-136m16.8*		-1.63015	0.54922	0.00300
ENSDARG00000096403	*si:dkey-153m14.1*		-1.28642	0.43377	0.00302
ENSDARG00000063670	*gtf2a1l*	*GTF2A1L*	-3.47369	1.17160	0.00303
ENSDARG00000056050	*kctd17*	*KCTD17*	-2.71101	0.92164	0.00327
ENSDARG00000103175	*CR626875.1*		-2.70467	0.92595	0.00349
ENSDARG00000029063	*clpxa*	*CLPX*	-3.17152	1.08927	0.00360
ENSDARG00000086173	*relb*	*RELB*	-2.00455	0.69434	0.00389
ENSDARG00000102043	*CABZ01040256.1*		-2.06733	0.72042	0.00411
ENSDARG00000099010	*CABZ01004882.1*		-2.20407	0.76900	0.00415
ENSDARG00000099404	*CR753902.1*		-2.08661	0.72828	0.00417
ENSDARG00000007601	*zmynd8*	*ZMYND8*	-3.07322	1.07482	0.00425
ENSDARG00000104968	*si:ch211-227e10.1*		-2.69580	0.94602	0.00438
ENSDARG00000101315	*CABZ01039820.1*		-1.78201	0.63195	0.00480
ENSDARG00000088309	*ENDOD1*		-2.66792	0.94787	0.00488
ERCC-00074	*ERCC-00074*	*NA*	-1.16507	0.41459	0.00495
ENSDARG00000054799	*rfc1*	*RFC1*	-3.11313	1.11136	0.00509
ENSDARG00000032373	*rnf145b*	*RNF145*	-2.21101	0.78962	0.00511
ENSDARG00000089382	*zgc:158463*		-1.53044	0.54657	0.00511
ENSDARG00000044235	*si:dkey-36h5.1*		-3.03557	1.08706	0.00523
ENSDARG00000076025	*dgkzb*	*DGKZ*	-1.31512	0.47126	0.00526
ENSDARG00000054616	*cldni*	*RP1-4G17.5*	-2.93782	1.05553	0.00538
ENSDARG00000061370	*tsen34*	*TSEN34*	-3.26057	1.18956	0.00613
ENSDARG00000006758	*FAM234B*	*FAM234B*	-2.34959	0.85727	0.00613
ENSDARG00000011824	*pbxip1b*		-1.55391	0.56761	0.00619
ENSDARG00000104308	*BX537263.2*		-1.57474	0.57796	0.00644
ENSDARG00000103816	*BX537263.1*		-1.57474	0.57796	0.00644
ENSDARG00000040571	*ube4a*	*UBE4A*	-2.99410	1.10350	0.00666
ENSDARG00000063097	*scube1*	*SCUBE1*	-1.72505	0.63592	0.00667
ENSDARG00000099973	*ctage5*	*CTAGE5*	-2.28491	0.84770	0.00703
ENSDARG00000077297	*nbr1*	*NBR1*	-2.71055	1.00819	0.00718
ENSDARG00000001220	*mycbp2*	*MYCBP2*	-1.29957	0.48419	0.00727
ENSDARG00000009953	*med14*	*MED14*	-2.84265	1.05967	0.00731
ENSDARG00000079104	*mfhas1*	*MFHAS1*	-2.94585	1.10042	0.00743
ENSDARG00000036168	*nfatc1*	*NFATC1*	-2.51203	0.93861	0.00744
ENSDARG00000068214	*ccni*	*CCNI*	-0.90480	0.33808	0.00744
ENSDARG00000079312	*kmt2ca*	*KMT2C*	-1.00799	0.37769	0.00761
ENSDARG00000040668	*lrtomt*	*Lrrc51*	-2.86069	1.07590	0.00784
ENSDARG00000008153	*serinc5*	*SERINC5*	-1.40503	0.52856	0.00785
ENSDARG00000103635	*si:dkeyp-3b12.10*		-1.34041	0.50442	0.00788

ENSDARG00000077918	*dnase2b*	*DNASE2B*	-2.62891	0.98970	0.00790
ENSDARG00000040439	*rsl24d1*	*RSL24D1*	-1.02622	0.38986	0.00848
ENSDARG00000017165	*slc3a1*	*SLC3A1*	-2.00396	0.76237	0.00857
ENSDARG00000062385	*flcn*	*FLCN*	-3.11766	1.19854	0.00929
ENSDARG00000074245	*spen*	*SPEN*	-2.35026	0.90411	0.00933
ENSDARG00000089706	*ANPEP*		-1.60632	0.61893	0.00945
ENSDARG00000020711	*rrm2*	*RRM2*	-2.87852	1.11340	0.00973

Table A8: Differentially downregulated genes in mutant neutrophils. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000040252	*atp1a1a.5*	*ATP1A1*	4.53260	0.77854	0.00000
ENSDARG00000098639	*pycr1b*	*PYCR1*	2.28789	0.54154	0.00002
ENSDARG00000018574	*sf3b4*	*SF3B4*	3.74856	0.92349	0.00005
ENSDARG00000070076	*actr2b*	*ACTR2*	3.19605	0.81186	0.00008
ENSDARG00000101534	*rab3db*	*RAB3D*	4.29220	1.14278	0.00017
ENSDARG00000100075	*abcg2a*	*ABCG2*	3.93941	1.05964	0.00020
ENSDARG00000008380	*brd7*	*BRD7*	3.62897	0.97548	0.00020
ENSDARG00000053609	*pddc1*	*PDDC1*	3.57902	0.94612	0.00016
ENSDARG00000098231	*HHEX*	*HHEX*	3.49944	0.93430	0.00018
ENSDARG00000074667	*akt1s1*	*AKT1S1*	3.33291	0.88736	0.00017
ENSDARG00000053136	*b2m*	*B2M*	1.92578	0.50926	0.00016
ENSDARG00000077473	*mych*		4.26896	1.15675	0.00022
ENSDARG00000104635	*si:busm1-194e12.12*	*HLA-DRB1*	3.61245	1.00462	0.00032
ENSDARG00000038587	*CU929150.1*		2.27314	0.65211	0.00049
ENSDARG00000075881	*si:ch211-39k3.2*	*SERF1B*	4.09407	1.18344	0.00054
ENSDARG00000071694	*ndc80*	*NDC80*	4.01394	1.18319	0.00069
ENSDARG00000074322	*si:ch211-194m7.3*		3.72320	1.08287	0.00059
ENSDARG00000100743	*si:dkey-190j3.4*		3.62591	1.06767	0.00068
ENSDARG00000055270	*si:ch1073-358c10.1*		3.59872	1.04720	0.00059
ENSDARG00000010625	*clic2*	*CLIC2*	3.21759	0.94367	0.00065
ENSDARG00000036700			3.20967	0.94815	0.00071
ENSDARG00000056085	*mob4*	*MOB4*	3.15209	0.92973	0.00070
ENSDARG00000091111	*TIFA*	*TIFA*	2.88701	0.85511	0.00073
ENSDARG00000092731	*mhc1uka*	*FCGRT*	1.80730	0.53178	0.00068
ENSDARG00000035853	*cdkn2aip*	*CDKN2AIP*	3.75028	1.12185	0.00083
ENSDARG00000092124	*cox14*	*COX14*	1.85608	0.55714	0.00086
ENSDARG00000059294	*marco*	*MARCO*	2.00016	0.60244	0.00090
ENSDARG00000017034	*sqrdl*	*SQRDL*	3.15307	0.95952	0.00102
ENSDARG00000038097	*pigq*	*PIGQ*	3.57942	1.09799	0.00111
ENSDARG00000074732	*BX682234.1*		2.62830	0.80647	0.00112
ENSDARG00000005536	*ubr7*	*UBR7*	3.42568	1.05489	0.00116

ENSDARG00000059020	*thap1*	*THAP1*	3.26196	1.01481	0.00131
ENSDARG00000018968	*acvr1ba*	*ACVR1B*	3.49882	1.09113	0.00134
ENSDARG00000095746			3.31616	1.03865	0.00141
ENSDARG00000005098	*zgc:86764*		3.78900	1.19852	0.00157
ENSDARG00000013771	*ctssb.2*	*CTSS*	2.32041	0.74009	0.00172
ENSDARG00000088745			2.42336	0.77531	0.00177
ENSDARG00000011146	*uqcrb*	*UQCRB*	1.36447	0.43696	0.00179
ENSDARG00000053990	*hmgb2b*		1.88566	0.60898	0.00196
ENSDARG00000055120	*ctsba*	*CTSB*	1.47301	0.47597	0.00197
ENSDARG00000097082	*im:7152348*		3.27231	1.06019	0.00203
ENSDARG00000094210	*zgc:109934*		1.58334	0.51601	0.00215
ENSDARG00000038667	*fggy*	*FGGY*	3.51797	1.15057	0.00223
ENSDARG00000033285	*gsto2*	*GSTO2*	2.75325	0.90407	0.00232
ENSDARG00000022303	*higd1a*	*HIGD1A*	2.19279	0.71917	0.00230
ENSDARG00000086712	*si:dkeyp-97b10.3*		3.39558	1.11952	0.00242
ENSDARG00000089667			3.33877	1.11770	0.00282
ENSDARG00000101337	*zgc:103700*	*HLA-DRB1*	2.45759	0.82070	0.00275
ENSDARG00000034817	*asah1b*	*ASAH1*	2.37546	0.79473	0.00280
ENSDARG00000056583	*ndufs6*	*NDUFS6*	1.78485	0.60099	0.00298
ENSDARG00000002165	*psme1*	*PSME1*	1.83852	0.62594	0.00331
ENSDARG00000055314	*mcmbp*	*MCMBP*	3.15962	1.08205	0.00350
ENSDARG00000020252	*btbd10b*	*BTBD10*	2.58145	0.88568	0.00356
ENSDARG00000086947	*si:ch211-147m6.1*		3.10102	1.07946	0.00407
ENSDARG00000100562	*sdc3*	*SDC3*	3.08242	1.06659	0.00385
ENSDARG00000037283	*plrg1*	*PLRG1*	3.01539	1.05339	0.00420
ENSDARG00000019861	*fgl2a*	*FGL2*	2.38800	0.83294	0.00414
ENSDARG00000039579	*cfd*	*CFD*	2.28427	0.79547	0.00408
ENSDARG00000090783	*mfap4*	*MFAP4*	2.20937	0.77103	0.00416
ENSDARG00000043154	*ucp2*	*UCP2*	1.89722	0.66259	0.00419
ENSDARG00000067975	*atpif1a*	*ATPIF1*	1.15120	0.39885	0.00390
ENSDARG00000099455	*ogt.2*	*OGT*	1.89602	0.66800	0.00453
ENSDARG00000028396	*fkbp5*	*FKBP5*	2.24094	0.79050	0.00458
ENSDARG00000104018	*mtpn*	*MTPN*	1.65711	0.58819	0.00484
ENSDARG00000070589	*mrpl35*	*MRPL35*	2.78353	0.99224	0.00503
ENSDARG00000096668	*MZT1*	*MZT1*	2.72871	0.97885	0.00531
ENSDARG00000018328	*rhoad*	*RHOC*	1.33618	0.48013	0.00539
ENSDARG00000057853	*atp6v0ca*	*ATP6V0C*	1.13943	0.40895	0.00533
ENSDARG00000101831	*irx1a*	*IRX1*	3.13593	1.12998	0.00552
ENSDARG00000089505	*bod1l1*	*BOD1L1*	2.82611	1.01848	0.00552
ENSDARG00000100825	*calm3a*		2.28073	0.82098	0.00547
ENSDARG00000101791	*ccbl2*	*CCBL2*	3.34393	1.21956	0.00611
ENSDARG00000061896	*slco2a1*	*SLCO2A1*	3.20746	1.17497	0.00634
ENSDARG00000095974	*si:ch73-211l13.2*		3.18486	1.16298	0.00617

ENSDARG00000069961	*il21r.1*	*IL21R*	3.03720	1.10633	0.00605
ENSDARG00000027803	*sbds*	*SBDS*	2.44424	0.88634	0.00582
ENSDARG00000014676	*bckdhb*	*BCKDHB*	2.14469	0.77680	0.00576
ENSDARG00000017049	*adsl*	*RP5-1042K10.14*	2.10812	0.77068	0.00623
ENSDARG00000004173	*copa*	*COPA*	1.97511	0.72232	0.00625
ENSDARG00000095627	*c1qc*	*C1QC*	1.91451	0.69376	0.00579
ENSDARG00000044267	*sumo1*	*SUMO1*	1.65065	0.60515	0.00638
ENSDARG00000098584	*ndufa13*	*NDUFA13*	1.46654	0.53816	0.00643
ENSDARG00000010149	*atp5a1*	*ATP5A1*	1.17689	0.42863	0.00604
ENSDARG00000005870	*mmadhc*	*MMADHC*	2.31712	0.85206	0.00654
ENSDARG00000037238	*smad5*		2.71075	1.00146	0.00679
ENSDARG00000036895	*dap1b*		2.45853	0.90799	0.00678
ENSDARG00000055592	*capn2b*	*CAPN2*	2.44983	0.90731	0.00693
ENSDARG00000100513	*rps27.2*	*RPS27L*	1.62017	0.60063	0.00699
ENSDARG00000014794	*uqcrc2a*	*UQCRC2*	1.74548	0.64869	0.00713
ENSDARG00000021346	*pdhb*	*PDHB*	1.03862	0.38963	0.00768
ENSDARG00000104537	*cox7c*	*COX7C*	1.05819	0.40205	0.00849
ENSDARG00000043687	*SAYSD1*	*SAYSD1*	2.45341	0.93739	0.00886
ENSDARG00000042793	*tpp1*	*TPP1*	2.08217	0.79876	0.00914
ENSDARG00000038938	*ssfa2*	*SSFA2*	2.83869	1.09293	0.00940
ENSDARG00000006260	*tuba8l4*		1.08866	0.42032	0.00960
ENSDARG00000056089	*nrbp1*	*NRBP1*	2.33396	0.90450	0.00987
ENSDARG00000039150	*lgmn*	*LGMN*	1.12194	0.43493	0.00989
ENSDARG00000042892	*paip1*	*PAIP1*	2.68064	1.04039	0.00998

Table A9: Differentially upregulated genes in mutant progenitors. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000004692	*nabp1a*	*NABP1*	-6.11321	1.26393	0.00000
ENSDARG00000100789	*plgrkt*	*PLGRKT*	-4.73613	1.00113	0.00000
ENSDARG00000091428	*FCHSD2*	*FCHSD2*	-2.32393	0.55272	0.00003
ENSDARG00000100854	*ago4*	*AGO4*	-3.28611	0.79013	0.00003
ENSDARG00000011094	*ccna2*	*CCNA2*	-5.45405	1.35635	0.00006
ENSDARG00000031795	*abcf1*	*ABCF1*	-2.27593	0.56599	0.00006
ENSDARG00000070959	*si:ch211-288g17.3*		-2.91850	0.76496	0.00014
ENSDARG00000015394	*thoc7*	*THOC7*	-5.09984	1.36451	0.00019
ENSDARG00000100623	*trrap*	*TRRAP*	-5.09684	1.37027	0.00020
ENSDARG00000099351	*igfbp1a*	*IGFBP1*	-3.43624	0.93008	0.00022
ENSDARG00000100894	*si:ch211-193k8.5*		-5.18729	1.40925	0.00023
ENSDARG00000058471	*plk1*	*PLK1*	-4.98256	1.36357	0.00026
ENSDARG00000020711	*rrm2*	*RRM2*	-3.58323	1.01127	0.00040
ENSDARG00000016783	*srsf6b*	*SRSF6*	-2.04507	0.58187	0.00044

Appendix

ENSDARG00000103028	*CABZ01052487.1*	*GIMAP8*	-4.92962	1.40317	0.00044
ENSDARG00000054007	*ppp1r2*	*PPP1R2*	-4.92928	1.41298	0.00049
ENSDARG00000075795	*nol7*	*NOL7*	-4.83986	1.39408	0.00052
ENSDARG00000031683	*fosab*	*FOS*	-4.58693	1.33680	0.00060
ENSDARG00000025859	*lmf2b*	*LMF2*	-4.72071	1.38236	0.00064
ENSDARG00000051923	*ccnb1*	*CCNB1*	-4.10270	1.21056	0.00070
ENSDARG00000026028	*ankrd44*	*ANKRD44*	-2.37154	0.70347	0.00075
ENSDARG00000008593	*nbas*	*NBAS*	-4.64063	1.37903	0.00077
ENSDARG00000005468	*irf2bp1*	*IRF2BP1*	-4.15585	1.24789	0.00087
ENSDARG00000051902	*prmt7*	*PRMT7*	-4.65158	1.39854	0.00088
ENSDARG00000095556	*si:dkey-238c7.12*		-1.39073	0.41958	0.00092
ENSDARG00000003526	*psma5*	*PSMA5*	-3.34648	1.01084	0.00093
ENSDARG00000035150	*si:dkey-261j4.4*		-4.52124	1.36972	0.00096
ENSDARG00000006240	*slc27a1a*	*SLC27A1*	-4.65038	1.41344	0.00100
ENSDARG00000060917	*anln*	*ANLN*	-4.65234	1.41432	0.00100
ENSDARG00000099161	*dyx1c1*	*DYX1C1*	-4.74873	1.44434	0.00101
ENSDARG00000041317	*rangap1a*	*RANGAP1*	-4.02224	1.22916	0.00107
ENSDARG00000046002	*necap2*	*NECAP2*	-1.76266	0.53903	0.00108
ENSDARG00000035634	*rfc5*	*RFC5*	-3.50785	1.08663	0.00125
ENSDARG00000007130	*mrto4*	*MRTO4*	-4.72355	1.46749	0.00129
ENSDARG00000101180	*mcm7*	*MCM7*	-1.93034	0.60000	0.00129
ENSDARG00000071341	*abch1*		-3.83865	1.19412	0.00131
ENSDARG00000059323	*rpn1*	*RPN1*	-2.94994	0.91950	0.00134
ENSDARG00000003527	*mak16*	*MAK16*	-4.51373	1.41019	0.00137
ENSDARG00000008502	*pno1*	*PNO1*	-1.74775	0.55207	0.00155
ENSDARG00000073850	*hdac7b*	*HDAC7*	-4.50233	1.42646	0.00160
ENSDARG00000036995	*lsm6*	*LSM6*	-3.63722	1.15902	0.00170
ENSDARG00000101521	*UBE2M*	*UBE2M*	-4.34394	1.39443	0.00184
ENSDARG00000078894	*hemk1*	*HEMK1*	-4.28099	1.37744	0.00188
ENSDARG00000021112	*c1d*	*C1D*	-4.34600	1.41480	0.00213
ENSDARG00000101935	*si:ch211-195b11.4*		-3.66765	1.19402	0.00213
ENSDARG00000008447	*fkbp4*	*FKBP4*	-4.09399	1.33292	0.00213
ENSDARG00000104708	*ddx24*	*DDX24*	-2.59427	0.84972	0.00226
ENSDARG00000102590	*polr3h*	*POLR3H*	-3.91086	1.28218	0.00229
ENSDARG00000005926	*ak2*	*AK2*	-1.98585	0.65827	0.00255
ENSDARG00000003599	*rpl3*	*RPL3*	-0.94842	0.31439	0.00256
ENSDARG00000101134	*CABZ01064859.2*		-4.42378	1.47488	0.00270
ENSDARG00000004372	*fmnl3*	*FMNL3*	-2.88507	0.96281	0.00273
ENSDARG00000076509	*polr2d*	*POLR2D*	-2.86996	0.95853	0.00275
ENSDARG00000087186	*ZHX3*		-2.95715	0.99073	0.00284
ENSDARG00000004713	*mad2l1*	*MAD2L1*	-4.30140	1.44423	0.00290
ENSDARG00000030022	*nup188*	*NUP188*	-4.03695	1.35589	0.00291
ENSDARG00000079605	*prmt5*	*PRMT5*	-3.34879	1.12803	0.00299

ENSDARG00000002403	*nusap1*	*NUSAP1*	-3.84951	1.30752	0.00324
ENSDARG00000074129	*edem3*	*EDEM3*	-4.18687	1.42218	0.00324
ENSDARG00000078315	*C5H12orf49*	*C12orf49*	-4.27258	1.45288	0.00327
ENSDARG00000098787	*si:dkey-10p5.7*		-2.18830	0.74754	0.00342
ENSDARG00000004527	*pin4*	*PIN4*	-3.52212	1.20992	0.00360
ENSDARG00000103492	*si:ch73-111k22.3*		-4.15822	1.44301	0.00396
ENSDARG00000101942	*si:dkey-260j18.2*		-4.14188	1.43832	0.00398
ENSDARG00000091902	*b3gnt2b*	*B3GNT2*	-2.65345	0.92294	0.00404
ENSDARG00000043960	*rpf2*	*RPF2*	-2.37113	0.82538	0.00407
ENSDARG00000032157	*grk6*	*GRK6*	-3.54695	1.23623	0.00412
ENSDARG00000043081	*ctsz*	*CTSZ*	-2.25601	0.78818	0.00421
ENSDARG00000029150	*hsp90ab1*	*HSP90AB1*	-0.97299	0.34141	0.00437
ENSDARG00000042854	*ephx1*	*EPHX1*	-4.13382	1.45549	0.00451
ENSDARG00000104837	*nudc*	*NUDC*	-2.16224	0.76204	0.00455
ENSDARG00000009743	*efhc1*	*EFHC1*	-4.16374	1.46799	0.00456
ENSDARG00000097421	*si:dkeyp-50b9.1*		-4.13353	1.46133	0.00468
ENSDARG00000011405	*rps9*	*RPS9*	-0.87277	0.30882	0.00471
ENSDARG00000069118	*ppp2r5eb*	*PPP2R5E*	-2.44501	0.86553	0.00473
ENSDARG00000019181	*rpsa*	*RPSA*	-0.94660	0.33510	0.00473
ENSDARG00000026489	*khsrp*	*KHSRP*	-2.20726	0.78464	0.00491
ENSDARG00000014329	*npm1a*	*NPM1*	-1.85758	0.66288	0.00507
ENSDARG00000035559	*tp53*	*TP53*	-2.73582	0.97792	0.00515
ENSDARG00000016393	*arf1l*	*Arf2*	-2.20588	0.79011	0.00524
ENSDARG00000022623	*fbxo44*	*FBXO44*	-3.81609	1.36873	0.00530
ENSDARG00000076889	*ralgapa1*	*RALGAPA1*	-2.51200	0.90193	0.00535
ENSDARG00000059711	*nol6*	*NOL6*	-3.84406	1.38421	0.00548
ENSDARG00000017445	*eif3i*	*EIF3I*	-1.58956	0.57371	0.00559
ENSDARG00000093708	*arhgap20b*	*ARHGAP20*	-4.19441	1.51419	0.00560
ENSDARG00000094646	*dyrk2*	*DYRK2*	-2.06285	0.74554	0.00566
ENSDARG00000045565	*noc4l*	*NOC4L*	-3.49729	1.26635	0.00575
ENSDARG00000092780	*si:ch1073-170o4.1*		-4.00759	1.45178	0.00577
ENSDARG00000052856	*khdrbs1a*	*KHDRBS1*	-1.12631	0.40868	0.00585
ENSDARG00000056888		*DNAH8*	-3.58295	1.30082	0.00588
ENSDARG00000029533	*rpl18*	*RPL18*	-0.82967	0.30137	0.00590
ENSDARG00000059360	*srsf3b*	*SRSF3*	-2.17223	0.78919	0.00591
ENSDARG00000060695	*znf346*	*ZNF346*	-3.36929	1.22785	0.00607
ENSDARG00000057026	*ran*	*RAN*	-1.03358	0.37670	0.00607
ENSDARG00000032430	*ppp2r1b*	*PPP2R1B*	-3.18154	1.16383	0.00626
ENSDARG00000076228	*kif2c*	*KIF2C*	-4.07476	1.49078	0.00627
ENSDARG00000012073	*kif15*	*KIF15*	-4.08327	1.49576	0.00634
ENSDARG00000045843	*apex1*	*APEX1*	-2.72842	1.00194	0.00647
ENSDARG00000079783	*isg20*	*AEN*	-3.25090	1.19496	0.00652
ENSDARG00000030665	*TCERG1*		-4.05933	1.49305	0.00655

Appendix

ENSDARG00000099358	*hccsb*	*HCCS*	-4.00750	1.47419	0.00656
ENSDARG00000041350	*ankrd39*	*ANKRD39*	-4.06922	1.49692	0.00656
ENSDARG00000061100	*nars*	*NARS*	-1.95386	0.71927	0.00660
ENSDARG00000094719	*si:dkeyp-1h4.9*		-1.68042	0.62122	0.00683
ENSDARG00000075650	*dscr3*	*DSCR3*	-2.34635	0.86872	0.00691
ENSDARG00000055585	*c1galt1a*		-4.02413	1.49062	0.00694
ENSDARG00000031108	*lrba*	*LRBA*	-2.57146	0.95263	0.00695
ENSDARG00000016548	*eif5b*	*EIF5B*	-2.00373	0.74302	0.00700
ENSDARG00000005791	*rpl28*	*RPL28*	-0.83371	0.30959	0.00708
ENSDARG00000058328	*lsm7*	*LSM7*	-2.02065	0.75186	0.00720
ENSDARG00000089602	*si:dkey-217f16.1*	*Gvin1*	-1.77492	0.66119	0.00727
ENSDARG00000070606	*ikbke*	*IKBKE*	-3.45540	1.28996	0.00739
ENSDARG00000100427	*BX547934.1*		-1.94059	0.72476	0.00742
ENSDARG00000093647	*si:dkey-48g21.7*		-3.93777	1.47371	0.00754
ENSDARG00000099243	*CR855389.1*		-3.38027	1.26747	0.00765
ENSDARG00000002213	*invs*	*INVS*	-4.00168	1.50186	0.00771
ENSDARG00000087950	*IWS1*	*IWS1*	-2.46194	0.92437	0.00774
ENSDARG00000100292	*fam166b*	*FAM166B*	-3.93517	1.47793	0.00775
ENSDARG00000004525	*C2H1orf35*	*C1orf35*	-3.03064	1.13854	0.00777
ENSDARG00000078751	*TPP2*	*TPP2*	-1.69170	0.63569	0.00779
ENSDARG00000016871	*smarce1*	*SMARCE1*	-1.77778	0.66941	0.00791
ENSDARG00000098398	*CABZ01079480.1*		-3.90717	1.47306	0.00799
ENSDARG00000101766	*ptmab*		-1.13536	0.42855	0.00807
ENSDARG00000055502	*cicb*	*CIC*	-2.56901	0.96991	0.00808
ENSDARG00000055996	*rps8a*	*RPS8*	-1.05080	0.39708	0.00814
ENSDARG00000102640	*pdia3*	*PDIA3*	-1.59758	0.60574	0.00835
ENSDARG00000100514	*cfap20*	*CFAP20*	-3.41672	1.29683	0.00842
ENSDARG00000070617	*vhl*	*VHL*	-3.89672	1.47957	0.00845
ENSDARG00000088641	*grn2*		-3.26229	1.23968	0.00850
ENSDARG00000060089	*btaf1*	*BTAF1*	-3.53527	1.34422	0.00854
ENSDARG00000014634	*mbtps1*	*MBTPS1*	-3.33413	1.26796	0.00855
ENSDARG00000035692	*rps3a*	*RPS3A*	-1.05131	0.40020	0.00861
ENSDARG00000077584	*cln6a*	*CLN6*	-3.95042	1.50601	0.00871
ENSDARG00000010279	*scamp2*	*SCAMP2*	-3.00204	1.14562	0.00878
ENSDARG00000019117	*parvb*	*PARVB*	-3.38174	1.29520	0.00903
ENSDARG00000077092	*elk4*	*ELK4*	-3.90162	1.49554	0.00909
ENSDARG00000076729	*myo9aa*	*MYO9A*	-3.27436	1.25586	0.00913
ENSDARG00000057683	*mcm6*	*MCM6*	-2.48226	0.95358	0.00924
ENSDARG00000015862	*rpl5b*	*RPL5*	-0.92330	0.35486	0.00927
ENSDARG00000053990	*hmgb2b*		-1.92119	0.73851	0.00928
ENSDARG00000056679	*morc3a*	*MORC3*	-2.83339	1.09068	0.00938
ENSDARG00000016789	*zgc:152891*	*ALOXE3*	-3.89583	1.50064	0.00943
ENSDARG00000074242	*serbp1a*	*SERBP1*	-1.01029	0.38917	0.00943

ENSDARG00000062116	*ctdsplb*	*CTDSPL*	-2.83024	1.09248	0.00958
ENSDARG00000068681	*crfb1*		-3.67630	1.42241	0.00975
ENSDARG00000021753	*ccdc25*	*CCDC25*	-2.85942	1.10738	0.00982
ENSDARG00000097102	*si:ch73-281n10.2*		-1.93791	0.75088	0.00986
ENSDARG00000041875	*ube2v1*		-1.82376	0.70678	0.00987
ENSDARG00000074688	*fbrs*	*FBRS*	-3.42459	1.32725	0.00987
ENSDARG00000006307	*shisa4*	*SHISA4*	-3.12270	1.21062	0.00990
ENSDARG00000096885	*si:dkey-74h17.4*		-3.93961	1.52801	0.00993
ENSDARG00000007320	*rpl7*	*RPL7*	-0.90308	0.35051	0.00998

Table A10: Differentially downregulated genes in mutant progenitors. Only genes with a $P < .001$ are shown. FC SE = fold change standard error. ENSEMBL IDs correspond to the zebrafish gene.

ENSEMBL ID	Zebrafish	Human	Fold change (log2)	FC SE	*P*
ENSDARG00000055523	*slc22a6l*	*SLC22A8*	4.66096	0.96073	0.00000
ENSDARG00000101914	*metrnl*	*METRNL*	4.80443	1.03917	0.00000
ENSDARG00000040252	*atp1a1a.5*	*ATP1A1*	4.42554	1.02970	0.00002
ENSDARG00000053535	*lmo7b*	*LMO7*	5.23446	1.24815	0.00003
ENSDARG00000089917	*sh3tc2*	*SH3TC2*	4.95516	1.20801	0.00004
ENSDARG00000058094	*ciarta*	*CIART*	4.40436	1.08706	0.00005
ENSDARG00000099448	*sh3d21*	*SH3D21*	5.69812	1.41365	0.00006
ENSDARG00000006545	*rgp1*	*RGP1*	4.79622	1.19113	0.00006
ENSDARG00000056108	*ndufa4*		4.24860	1.05673	0.00006
ENSDARG00000075980	*tmem125b*	*TMEM125*	4.23016	1.07369	0.00008
ENSDARG00000037739	*zgc:112980*		3.77631	0.96407	0.00009
ENSDARG00000035540	*brwd3*	*BRWD3*	5.13085	1.32140	0.00010
ENSDARG00000028780	*gzma*	*GZMK*	4.65030	1.19795	0.00010
ENSDARG00000052170	*uap1*	*UAP1*	4.80270	1.26544	0.00015
ENSDARG00000097300	*si:ch211-242h13.4*		4.12655	1.08744	0.00015
ENSDARG00000003219	*bin2a*	*BIN2*	4.09537	1.08207	0.00015
ENSDARG00000095717	*si:dkey-172k15.6*		5.51321	1.46558	0.00017
ENSDARG00000098686	*tmprss2*	*TMPRSS2*	3.93552	1.04955	0.00018
ENSDARG00000100198	*SNCAIP*	*SNCAIP*	5.46773	1.46045	0.00018
ENSDARG00000016519	*ctdnep1a*	*CTDNEP1*	4.73497	1.27116	0.00020
ENSDARG00000055648	*cpda*	*CPD*	3.63825	0.99010	0.00024
ENSDARG00000102870	*si:dkeyp-123a12.2*		5.29655	1.44912	0.00026
ENSDARG00000082142	*CR753862.1*		5.21744	1.43110	0.00027
ENSDARG00000042090	*PLA2G4C*		2.31861	0.63996	0.00029
ENSDARG00000038475	*acy1*	*ACY1*	3.44344	0.95284	0.00030
ENSDARG00000103144	*ppp1r21*	*PPP1R21*	5.14704	1.42489	0.00030
ENSDARG00000017168	*nr2f1b*		5.17033	1.43534	0.00032
ENSDARG00000061180	*vps45*	*VPS45*	5.26262	1.46285	0.00032
ENSDARG00000092346	*nfil3-4*		5.24512	1.46076	0.00033

Appendix

ENSDARG00000091539	*ptprjb.1*	*PTPRJ*	5.23128	1.46032	0.00034
ENSDARG00000087390	*si:ch211-5k11.12*	*HBD*	5.28762	1.49074	0.00039
ENSDARG00000098817	*KCTD5*		3.06037	0.87144	0.00044
ENSDARG00000062154	*dip2ca*	*DIP2C*	5.09850	1.46084	0.00048
ENSDARG00000102090	*CABZ01021453.1*		3.22840	0.92612	0.00049
ENSDARG00000057681	*ZMYM4*	*RP11-244H3.4*	4.68571	1.34969	0.00052
ENSDARG00000021149	*cbr1l*		4.09786	1.18203	0.00053
ENSDARG00000094912	*si:dkey-270m17.3*		4.16127	1.20372	0.00055
ENSDARG00000077434	*mon1bb*	*MON1B*	5.12537	1.49180	0.00059
ENSDARG00000002339	*ccdc101*	*RP11-347C12.3*	5.05534	1.47673	0.00062
ENSDARG00000101074	*glud1b*		2.81900	0.82930	0.00068
ENSDARG00000022660	*armc2*	*ARMC2*	2.29186	0.67494	0.00068
ENSDARG00000033413	*si:dkey-183c16.7*		4.57798	1.35436	0.00072
ENSDARG00000043497	*scrn2*		5.03727	1.49967	0.00078
ENSDARG00000098631	*DST*		5.04248	1.50300	0.00079
ENSDARG00000087394	*tshz3a*		5.01261	1.50647	0.00088
ENSDARG00000100776	*syt8*	*SYT8*	3.65660	1.10113	0.00090
ENSDARG00000015224	*cd2ap*	*CD2AP*	4.92979	1.48781	0.00092
ENSDARG00000078966	*rbm15b*	*RBM15B*	3.24245	0.97886	0.00092
ENSDARG00000009018	*rhbg*	*RHBG*	2.90792	0.88176	0.00097
ENSDARG00000105109	*si:ch1073-329i9.1*		2.82497	0.85698	0.00098
ENSDARG00000025593	*plekhj1*	*PLEKHJ1*	1.64486	0.50070	0.00102
ENSDARG00000036569	*bach2a*		4.09674	1.24942	0.00104
ENSDARG00000041644	*TMEM27*	*TMEM27*	2.92272	0.89876	0.00115
ENSDARG00000073999	*tapt1a*	*TAPT1*	3.88371	1.20442	0.00126
ENSDARG00000041078	*chka*	*CHKA*	3.69334	1.14935	0.00131
ENSDARG00000061256	*FAM195A*	*FAM195A*	4.19439	1.30578	0.00132
ENSDARG00000079102	*PRRG3*	*PRRG3*	2.70368	0.84294	0.00134
ENSDARG00000081270	*rn7sk*	*RN7SKP90*	3.82521	1.19341	0.00135
ENSDARG00000070116	*nit1*	*NIT1*	3.97214	1.24483	0.00142
ENSDARG00000052207	*c3a.3*	*C3*	4.83490	1.51772	0.00144
ENSDARG00000068246	*plcb3*	*PLCB3*	1.66212	0.52218	0.00146
ENSDARG00000103716	*si:busm1-48c11.3*	*HLA-DMA*	3.41185	1.07293	0.00147
ENSDARG00000077410	*myo9b*	*MYO9B*	2.94579	0.93167	0.00157
ENSDARG00000069339	*tbc1d24*	*TBC1D24*	4.85084	1.53465	0.00157
ENSDARG00000056627	*cxcl14*	*CXCL14*	4.79710	1.51903	0.00159
ENSDARG00000069292	*si:dkeyp-89c11.1*		4.84071	1.53672	0.00163
ENSDARG00000052099	*agxta*	*AGXT*	4.06820	1.29230	0.00164
ENSDARG00000004141	*zgc:92630*		4.82741	1.53742	0.00169
ENSDARG00000062634	*kat2b*	*KAT2B*	4.75800	1.51569	0.00169
ENSDARG00000095459	*si:ch211-191j22.3*		3.24451	1.03372	0.00170
ENSDARG00000070571	*C11H1orf106*		4.75955	1.51644	0.00170
ENSDARG00000100677	*tmem205*	*TMEM205*	2.39036	0.76337	0.00174

ENSDARG00000015201	*pcmt*	*PCMT1*	2.71712	0.87260	0.00185
ENSDARG00000032849	*ndrg1a*	*NDRG1*	1.79975	0.57849	0.00186
ENSDARG00000042484	*tle2*		3.34564	1.07918	0.00193
ENSDARG00000079497	*C5H8orf4*	*C8orf4*	3.82697	1.23668	0.00197
ENSDARG00000096758	*si:dkey-117p18.2*		4.71549	1.52645	0.00201
ENSDARG00000094561	*si:ch211-139a5.9*		2.15316	0.69866	0.00206
ENSDARG00000103354	*U1*	*RNU1-108P*	4.42013	1.43845	0.00212
ENSDARG00000061543	*ccdc85b*	*CCDC85B*	3.84828	1.25495	0.00217
ENSDARG00000040812	*ncf4*	*NCF4*	2.50487	0.81723	0.00218
ENSDARG00000010571	*ezh2*	*EZH2*	2.82572	0.92213	0.00218
ENSDARG00000104084	*GPRIN2*	*GPRIN2*	2.86751	0.93998	0.00228
ENSDARG00000015915	*slc25a36b*	*SLC25A36*	2.45524	0.80490	0.00229
ENSDARG00000002670	*ATG14*	*ATG14*	3.95746	1.30007	0.00233
ENSDARG00000074633	*gpr35.1*	*GPR35*	3.83375	1.26477	0.00244
ENSDARG00000039997	*ptp4a3*	*PTP4A3*	3.92112	1.29688	0.00250
ENSDARG00000105017	*CT737138.1*		3.68439	1.21903	0.00251
ENSDARG00000036252	*rras2*	*RRAS2*	2.60459	0.86284	0.00254
ENSDARG00000055043	*depdc7*	*DEPDC7*	3.82276	1.26830	0.00258
ENSDARG00000102805	*cyp2aa12*		2.54246	0.84782	0.00271
ENSDARG00000078069	*rrm2*	*RRM2*	4.25874	1.42055	0.00272
ENSDARG00000018073	*mrps22*	*MRPS22*	3.49986	1.17109	0.00280
ENSDARG00000104635	*si:busm1-194e12.12*	*HLA-DRB1*	3.00193	1.00511	0.00282
ENSDARG00000078326	*arhgap10*	*ARHGAP10*	3.84371	1.29125	0.00291
ENSDARG00000041691	*bhlhe41*	*BHLHE41*	3.81432	1.28164	0.00292
ENSDARG00000022261	*pdzk1*	*PDZK1*	3.14777	1.05969	0.00297
ENSDARG00000099139	*fbxw5*	*FBXW5*	3.08145	1.03820	0.00300
ENSDARG00000104178	*C3H19orf43*	*C19orf43*	3.11087	1.05047	0.00306
ENSDARG00000030839	*tescb*	*TESC*	4.34226	1.46683	0.00307
ENSDARG00000036041	*f2*	*F2*	3.44598	1.16455	0.00309
ENSDARG00000016875	*gys1*	*GYS1*	2.88009	0.97532	0.00315
ENSDARG00000017565	*itk*	*ITK*	3.91112	1.32495	0.00316
ENSDARG00000093549	*sepp1a*	*SEPP1*	1.95063	0.66150	0.00319
ENSDARG00000037855	*taf11*	*TAF11*	2.33529	0.79356	0.00325
ENSDARG00000041787	*cx32.3*		3.32156	1.13020	0.00329
ENSDARG00000103340	*clic1*	*CLIC1*	1.45145	0.49450	0.00333
ENSDARG00000024090	*dnajc4*	*DNAJC4*	2.71008	0.92729	0.00347
ENSDARG00000086808	*ddhd1a*	*DDHD1*	3.36063	1.14991	0.00347
ENSDARG00000055647	*ftr82*	*TRIM29*	4.53264	1.55380	0.00353
ENSDARG00000036232	*trpm7*	*TRPM1*	2.48939	0.85363	0.00354
ENSDARG00000012672	*gtf2e2*	*GTF2E2*	3.06320	1.05041	0.00354
ENSDARG00000092170	*apoc1l*		4.52724	1.55291	0.00355
ENSDARG00000090159	*si:dkey-37m8.8*		2.94296	1.00990	0.00357
ENSDARG00000102566	*fam102aa*	*FAM102A*	3.44785	1.18980	0.00376

Appendix

ENSDARG00000069056	*ranbp3b*	*RANBP3*	2.41843	0.83580	0.00381
ENSDARG00000087981	*SLC6A11*	*SLC6A11*	1.80897	0.62710	0.00392
ENSDARG00000021494	*hnf4a*	*HNF4A*	3.04403	1.05653	0.00396
ENSDARG00000089930	*iqce*	*IQCE*	3.98109	1.38347	0.00401
ENSDARG00000058865	*endog*	*ENDOG*	4.47174	1.55714	0.00408
ENSDARG00000073726	*zbtb32*	*ZBTB32*	2.91736	1.01649	0.00410
ENSDARG00000007906	*lrp2b*		4.47300	1.56071	0.00416
ENSDARG00000079745	*si:ch211-166a6.5*		3.25856	1.14294	0.00436
ENSDARG00000042953	*cyp2n13*	*CYP2J2*	4.44811	1.56077	0.00437
ENSDARG00000063436	*RPH3A*		2.22889	0.78272	0.00440
ENSDARG00000037810	*zgc:112175*		3.37278	1.18458	0.00441
ENSDARG00000008703	*apeh*	*APEH*	2.66590	0.93651	0.00442
ENSDARG00000105183	*si:ch211-198i6.4*		2.65541	0.93517	0.00452
ENSDARG00000095082	*BX571811.1*	*ITLN1*	4.39884	1.54967	0.00453
ENSDARG00000053257	*zgc:153733*		4.07124	1.43607	0.00458
ENSDARG00000043457	*gapdh*	*GAPDH*	2.41136	0.85079	0.00459
ENSDARG00000028389	*rab44*	*RAB44*	2.41813	0.85506	0.00468
ENSDARG00000088048	*fgf18a*	*FGF18*	4.04314	1.43019	0.00470
ENSDARG00000098661	*AL935300.1*		3.94111	1.39568	0.00475
ENSDARG00000103573	*si:dkey-48g21.5*		1.92027	0.68239	0.00489
ENSDARG00000044339	*rp2*	*RP2*	2.60806	0.92839	0.00497
ENSDARG00000089885	*slc16a12b*	*SLC16A12*	4.07096	1.46042	0.00531
ENSDARG00000038577	*cox6c*	*COX6C*	1.40727	0.50667	0.00548
ENSDARG00000057433	*st6galnac5b*		4.33835	1.56540	0.00558
ENSDARG00000100731	*slc27a2b*	*SLC27A2*	2.36937	0.85513	0.00559
ENSDARG00000059483	*tead1b*	*TEAD1*	2.17327	0.78505	0.00563
ENSDARG00000034714	*esyt1a*	*ESYT1*	1.91368	0.69277	0.00574
ENSDARG00000019579	*ldb2a*	*LDB2*	2.37136	0.85868	0.00575
ERCC-00053	*ERCC-00053*	*NA*	4.28651	1.55429	0.00582
ENSDARG00000056252	*sort1b*	*SORT1*	2.51930	0.91636	0.00597
ENSDARG00000104666	*BX511215.2*	*ZNF852*	2.99792	1.09143	0.00602
ENSDARG00000103388	*znf609a*	*ZNF609*	2.36098	0.86020	0.00606
ENSDARG00000020364	*fbp1b*	*FBP1*	2.98153	1.08726	0.00610
ENSDARG00000061985	*rbm47*	*RBM47*	2.08934	0.76235	0.00613
ENSDARG00000087403	*si:ch211-214p13.3*		3.62829	1.32446	0.00615
ENSDARG00000017143	*brd9*	*BRD9*	2.01004	0.73564	0.00629
ENSDARG00000023362	*nr5a1b*	*NR5A1*	2.93039	1.07251	0.00629
ENSDARG00000027595	*selt1b*		3.18588	1.16663	0.00632
ENSDARG00000096003	*USMG5*	*USMG5*	1.88522	0.69071	0.00635
ENSDARG00000033056	*zgc:171927*		4.26492	1.56959	0.00658
ENSDARG00000093354	*si:ch211-57i17.2*		3.93303	1.44821	0.00661
ENSDARG00000055253	*slc12a10.3*		1.85091	0.68210	0.00666
ENSDARG00000054060	*pof1b*	*POF1B*	2.52600	0.93096	0.00666

ENSDARG00000070688	*ncalda*	*NCALD*	1.02230	0.37710	0.00671
ENSDARG00000037476	*sorbs3*	*SORBS3*	3.64437	1.34791	0.00686
ENSDARG00000068474	*si:dkey-94e7.2*		4.04864	1.49892	0.00691
ENSDARG00000053563	*si:ch73-56d11.4*	*MS4A18*	3.84591	1.42492	0.00695
ENSDARG00000013087	*ndrg3a*	*NDRG3*	3.52161	1.30658	0.00703
ENSDARG00000034600	*tmem165*	*TMEM165*	3.96733	1.47496	0.00715
ENSDARG00000009961	*rundc3b*	*RUNDC3B*	1.73104	0.64383	0.00717
ENSDARG00000053636	*cracr2b*	*CRACR2B*	3.36234	1.25068	0.00718
ENSDARG00000056331	*ahcyl1*	*AHCYL1*	2.19484	0.81691	0.00721
ENSDARG00000086107	*MTERF1*	*MTERF1*	3.85667	1.43568	0.00722
ENSDARG00000033231	*mcm6l*		4.22222	1.57297	0.00727
ENSDARG00000099155	*tpst2*	*TPST2*	2.14378	0.79925	0.00731
ENSDARG00000008884	*hprt1*	*HPRT1*	2.01657	0.75188	0.00732
ENSDARG00000096809	*si:ch211-168b3.2*		1.91281	0.71344	0.00734
ENSDARG00000017165	*slc3a1*	*SLC3A1*	2.68717	1.00244	0.00735
ENSDARG00000098497	*hrasa*	*HRAS*	3.33565	1.24484	0.00737
ENSDARG00000077737	*spsb3a*	*SPSB3*	1.70957	0.63807	0.00738
ENSDARG00000098258	*SLC16A7*	*SLC16A7*	2.37630	0.88872	0.00750
ENSDARG00000094215	*si:dkey-240n22.2*		3.45112	1.29408	0.00766
ENSDARG00000095824	*si:ch211-207c7.6*		3.96139	1.48582	0.00767
ENSDARG00000102762	*pik3r5*	*PIK3R5*	3.90849	1.46716	0.00772
ENSDARG00000101547	*CR354395.1*		3.95906	1.48953	0.00786
ENSDARG00000033046	*ccni2*	*CCNI2*	2.17099	0.81756	0.00792
ENSDARG00000074468	*prdm11*	*PRDM11*	2.55153	0.96220	0.00801
ENSDARG00000073843	*myo9ab*	*MYO9A*	2.77831	1.04793	0.00802
ENSDARG00000097082	*im:7152348*		3.81288	1.43887	0.00805
ENSDARG00000031434	*rcor1*	*RCOR1*	2.63775	0.99633	0.00811
ENSDARG00000099902	*IL17RC*	*IL17RC*	2.70142	1.02061	0.00812
ENSDARG00000043102	*lxn*	*RARRES1*	1.94432	0.73601	0.00825
ENSDARG00000010042	*dnm1a*	*DNM1*	3.04935	1.15554	0.00832
ENSDARG00000097721	*si:dkey-234i14.21*		2.28694	0.86694	0.00834
ENSDARG00000010423	*npsn*	*ASTL*	2.56409	0.97363	0.00845
ENSDARG00000091003	*il34*	*IL34*	3.37825	1.28368	0.00850
ENSDARG00000102153	*nrp1a*	*NRP1*	2.70626	1.02860	0.00851
ENSDARG00000101979	*eps8l3a*	*EPS8L3*	2.99633	1.13961	0.00856
ENSDARG00000012610	*saga*	*SAG*	1.73437	0.66174	0.00877
ENSDARG00000017115	*kctd10*	*KCTD10*	2.42139	0.92620	0.00894
ENSDARG00000053129	*carhsp1*	*CARHSP1*	3.04557	1.16997	0.00924
ENSDARG00000069271	*kbtbd4*	*KBTBD4*	1.69588	0.65242	0.00934
ENSDARG00000099546	*kynu*	*KYNU*	1.81459	0.69864	0.00940
ENSDARG00000044254	*anxa3b*	*ANXA3*	2.74201	1.05601	0.00942
ENSDARG00000003820	*nr1d2a*		1.92075	0.74014	0.00946
ENSDARG00000075664	*si:ch1073-429i10.1*		2.67725	1.03172	0.00946

ENSDARG00000075024	*KIF1C*		2.41031	0.92926	0.00949
ENSDARG00000103714	*rab6a*		1.28745	0.49659	0.00953
ENSDARG00000021147	*nipal3*	*NIPAL3*	2.91240	1.12451	0.00960
ENSDARG00000078508	*KCP*	*KCP*	2.74329	1.05958	0.00962
ENSDARG00000061544	*ano6*	*ANO6*	1.42371	0.55018	0.00966
ENSDARG00000095653	*si:ch211-32p8.2*		1.94806	0.75289	0.00967
ENSDARG00000070487	*MARC2*		2.91604	1.12709	0.00967
ENSDARG00000104387	*slc4a5*	*SLC4A5*	2.79011	1.07991	0.00978
ENSDARG00000053961	*slc2a11a*	*AP000350.10*	2.86629	1.11143	0.00991

Table A11: GO terms from EnrichR KEGG for genes upregulated in the *Tg(itga2b:EGFP)* GFP^{low} population in mutants.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Apoptosis *Homo sapiens* hsa04210	3/140	0.00455	0.18208	-1.96126	3.34064	*PARP1, EIF2AK3, CSF2RB*
Jak-STAT signaling pathway *Homo sapiens* hsa04630	2/158	0.05603	0.41614	-1.87990	1.64818	*CSF2RB, PTPN2*
Wnt signaling pathway *Homo sapiens* hsa04310	2/142	0.04639	0.41614	-1.85411	1.62557	*FZD3, CACYBP*
Protein processing in endoplasmic reticulum *Homo sapiens* hsa04141	2/169	0.06304	0.41614	-1.71881	1.50694	*STT3A, EIF2AK3*
DNA replication *Homo sapiens* hsa03030	1/36	0.08593	0.41614	-1.60585	1.40791	*RFC4*
Regulation of autophagy *Homo sapiens* hsa04140	1/39	0.09259	0.41614	-1.53886	1.34917	*ATG14*
Sphingolipid metabolism *Homo sapiens* hsa00600	1/47	0.11011	0.41614	-1.52434	1.33644	*SMPD4*
Nucleotide excision repair *Homo sapiens* hsa03420	1/47	0.11011	0.41614	-1.51266	1.32620	*RFC4*
Mismatch repair *Homo sapiens* hsa03430	1/23	0.05656	0.41614	-1.41293	1.23876	*RFC4*
Basal cell carcinoma *Homo sapiens* hsa05217	1/55	0.12731	0.42437	-1.43700	1.23173	*FZD3*
Base excision repair *Homo sapiens* hsa03410	1/33	0.07923	0.41614	-1.37029	1.20138	*PARP1*
N-Glycan biosynthesis *Homo sapiens* hsa00510	1/49	0.11444	0.41614	-1.32294	1.15987	*STT3A*
Ribosome biogenesis in eukaryotes *Homo sapiens* hsa03008	1/89	0.19699	0.48197	-1.55089	1.13197	*MDN1*
Taste transduction *Homo sapiens* hsa04742	1/83	0.18509	0.48197	-1.52293	1.11156	*P2RX3*
Melanogenesis *Homo sapiens* hsa04916	1/100	0.21839	0.48531	-1.45735	1.05361	*FZD3*
Morphine addiction *Homo sapiens* hsa05032	1/91	0.20092	0.48197	-1.42639	1.04110	*PDE10A*
PPAR signaling pathway *Homo sapiens* hsa03320	1/69	0.15666	0.48197	-1.32034	0.96369	*UBC*
NF-kappa B signaling pathway *Homo sapiens* hsa04064	1/93	0.20484	0.48197	-1.31241	0.95790	*PARP1*
Non-alcoholic fatty liver disease (NAFLD) *Homo sapiens* hsa04932	1/151	0.31078	0.50558	-1.39485	0.95136	*EIF2AK3*
Measles *Homo sapiens* hsa05162	1/136	0.28473	0.50558	-1.35358	0.92321	*EIF2AK3*
Hepatitis C *Homo sapiens* hsa05160	1/133	0.27941	0.50558	-1.35028	0.92095	*EIF2AK3*
Lysosome *Homo sapiens* hsa04142	1/123	0.26141	0.50558	-1.32242	0.90195	*ACP2*
Influenza A *Homo sapiens* hsa05164	1/175	0.35060	0.50558	-1.32208	0.90172	*EIF2AK3*

Signaling pathways regulating pluripotency of stem cells *Homo sapiens* hsa04550	1/142	0.29526	0.50558	-1.30498	0.89006	*FZD3*
Calcium signaling pathway *Homo sapiens* hsa04020	1/180	0.35862	0.50558	-1.27727	0.87116	*P2RX3*
Alzheimer's disease *Homo sapiens* hsa05010	1/168	0.33921	0.50558	-1.24114	0.84651	*EIF2AK3*
Proteoglycans in cancer *Homo sapiens* hsa05205	1/203	0.39432	0.50881	-1.23684	0.83572	*FZD3*
Epstein-Barr virus infection *Homo sapiens* hsa05169	1/202	0.39281	0.50881	-1.23172	0.83226	*EIF2AK3*
Hippo signaling pathway *Homo sapiens* hsa04390	1/153	0.31418	0.50558	-1.18322	0.80701	*FZD3*
Purine metabolism *Homo sapiens* hsa00230	1/176	0.35221	0.50558	-1.15989	0.79110	*PDE10A*
Ras signaling pathway *Homo sapiens* hsa04014	1/227	0.42958	0.52070	-1.18088	0.77061	*NF1*
Regulation of actin cytoskeleton *Homo sapiens* hsa04810	1/214	0.41073	0.51341	-1.15421	0.76948	*ITGAE*
Herpes simplex infection *Homo sapiens* hsa05168	1/185	0.36655	0.50558	-1.10988	0.75699	*EIF2AK3*
MAPK signaling pathway *Homo sapiens* hsa04010	1/255	0.46827	0.53498	-1.15762	0.72412	*NF1*
HTLV-I infection *Homo sapiens* hsa05166	1/258	0.47227	0.53498	-1.09368	0.68413	*FZD3*
Cytokine-cytokine receptor interaction *Homo sapiens* hsa04060	1/265	0.48148	0.53498	-1.03403	0.64681	*CSF2RB*
Neuroactive ligand-receptor interaction *Homo sapiens* hsa04080	1/277	0.49692	0.53721	-1.02367	0.63607	*P2RX3*
MicroRNAs in cancer *Homo sapiens* hsa05206	1/297	0.52170	0.54916	-1.01481	0.60824	*FZD3*
Pathways in cancer *Homo sapiens* hsa05200	1/397	0.62930	0.64544	-1.09543	0.47961	*FZD3*
Metabolic pathways *Homo sapiens* hsa01100	2/1239	0.82956	0.82956	-1.04311	0.19492	*SMPD4, STT3A*

Table A12: GO terms from EnrichR KEGG for genes down regulated in the *Tg(itga2b:EGFP)* GFP^{low} population in mutants.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Epithelial cell signaling in Helicobacter pylori infection *Homo sapiens* hsa05120	3/68	0.01221	0.48782	-1.92319	1.38046	*TJP1, MAPK14, ATP6V0C*
Dopaminergic synapse *Homo sapiens* hsa04728	4/129	0.01236	0.48782	-1.84119	1.32161	*PPP3CC, PPP2R5C, MAPK14, ATF4*
Tuberculosis *Homo sapiens* hsa05152	5/178	0.00770	0.48782	-1.80427	1.29510	*PPP3CC, FCER1G, MAPK14, ATP6V0C, CTSS*
Wnt signaling pathway *Homo sapiens* hsa04310	4/142	0.01690	0.48782	-1.80333	1.29442	*SFRP1, PPP3CC, DAAM1, WNT7B*
Vibrio cholerae infection *Homo sapiens* hsa05110	2/51	0.05011	0.48782	-1.61188	1.15700	*TJP1, ATP6V0C*
Salmonella infection *Homo sapiens* hsa05132	3/86	0.02229	0.48782	-1.60581	1.15265	*TJP1, ARPC4, MAPK14*
Neurotrophin signaling pathway *Homo sapiens* hsa04722	3/120	0.05065	0.48782	-1.59297	1.14343	*ARHGDIA, MAPK14, ATF4*
mRNA surveillance pathway *Homo sapiens* hsa03015	3/91	0.02570	0.48782	-1.58746	1.13948	*PPP2R5C, MSI2, WDR33*
Arginine and proline metabolism *Homo sapiens* hsa00330	2/50	0.04842	0.48782	-1.57896	1.13337	*GOT2, CKB*
Oocyte meiosis *Homo sapiens* hsa04114	3/123	0.05373	0.48782	-1.56583	1.12395	*PPP3CC, PPP2R5C, YWHAH*
Osteoclast differentiation *Homo sapiens* hsa04380	3/132	0.06348	0.48782	-1.52768	1.09657	*PPP3CC, NCF4, MAPK14*
Sphingolipid signaling pathway *Homo sapiens* hsa04071	3/120	0.05065	0.48782	-1.51833	1.08985	*FCER1G, PPP2R5C, MAPK14*

Appendix

VEGF signaling pathway *Homo sapiens* hsa04370	2/61	0.06807	0.48782	-1.48416	1.06533	*PPP3CC, MAPK14*
Amyotrophic lateral sclerosis (ALS) *Homo sapiens* hsa05014	2/51	0.05011	0.48782	-1.46129	1.04891	*PPP3CC, MAPK14*
Fc epsilon RI signaling pathway *Homo sapiens* hsa04664	2/68	0.08171	0.48782	-1.38230	0.99222	*FCER1G, MAPK14*
MAPK signaling pathway *Homo sapiens* hsa04010	4/255	0.09825	0.48782	-1.36129	0.97714	*DUSP2, PPP3CC, MAPK14, ATF4*
Adrenergic signaling in cardiomyocytes *Homo sapiens* hsa04261	3/148	0.08268	0.48782	-1.35958	0.97590	*PPP2R5C, MAPK14, ATF4*
Amphetamine addiction *Homo sapiens* hsa05031	2/67	0.07971	0.48782	-1.35265	0.97093	*PPP3CC, ATF4*
Endocytosis *Homo sapiens* hsa04144	4/259	0.10250	0.48782	-1.29895	0.93239	*SH3KBP1, CAV2, ARPC4, VPS28*
HTLV-I infection *Homo sapiens* hsa05166	4/258	0.10143	0.48782	-1.29352	0.92849	*ZFP36, PPP3CC, WNT7B, ATF4*
Long-term potentiation *Homo sapiens* hsa04720	2/66	0.07773	0.48782	-1.29051	0.92632	*PPP3CC, ATF4*
Hippo signaling pathway *Homo sapiens* hsa04390	3/153	0.08914	0.48782	-1.23133	0.88384	*WNT7B, AREG, YWHAH*
Phagosome *Homo sapiens* hsa04145	3/154	0.09046	0.48782	-1.17846	0.84590	*NCF4, ATP6V0C, CTSS*
Shigellosis *Homo sapiens* hsa05131	2/65	0.07576	0.48782	-1.16493	0.83619	*ARPC4, MAPK14*
Synaptic vesicle cycle *Homo sapiens* hsa04721	2/63	0.07188	0.48782	-1.13461	0.81442	*STX1B, ATP6V0C*
Leishmaniasis *Homo sapiens* hsa05140	2/73	0.09194	0.48782	-1.11381	0.79949	*NCF4, MAPK14*
Adherens junction *Homo sapiens* hsa04520	2/74	0.09402	0.48782	-1.06382	0.76361	*TJP1, PTPN1*
Bacterial invasion of epithelial cells *Homo sapiens* hsa05100	2/78	0.10251	0.48782	-0.91517	0.65691	*CAV2, ARPC4*
GnRH signaling pathway *Homo sapiens* hsa04912	2/91	0.13144	0.54261	-1.03899	0.63520	*MAPK14, ATF4*
Epstein-Barr virus infection *Homo sapiens* hsa05169	3/202	0.16238	0.55787	-1.00941	0.58913	*PSMD3, MAPK14, YWHAH*
Proteoglycans in cancer *Homo sapiens* hsa05205	3/203	0.16403	0.55787	-1.00389	0.58591	*CAV2, WNT7B, MAPK14*
Glucagon signaling pathway *Homo sapiens* hsa04922	2/101	0.15482	0.55787	-0.90802	0.52995	*PPP3CC, ATF4*
T cell receptor signaling pathway *Homo sapiens* hsa04660	2/104	0.16199	0.55787	-0.85738	0.50040	*PPP3CC, MAPK14*
PI3K-Akt signaling pathway *Homo sapiens* hsa04151	4/341	0.20615	0.55787	-0.82256	0.48008	*TNR, PPP2R5C, ATF4, YWHAH*
TNF signaling pathway *Homo sapiens* hsa04668	2/110	0.17650	0.55787	-0.81075	0.47319	*MAPK14, ATF4*
Glycerophospholipid metabolism *Homo sapiens* hsa00564	2/95	0.14069	0.55471	-0.79319	0.46744	*LYPLA2, CHKA*
Platelet activation *Homo sapiens* hsa04611	2/122	0.20611	0.55787	-0.67601	0.39454	*FCER1G, MAPK14*
Leukocyte transendothelial migration *Homo sapiens* hsa04670	2/118	0.19617	0.55787	-0.57246	0.33411	*NCF4, MAPK14*
Lysosome *Homo sapiens* hsa04142	2/123	0.20861	0.55787	-0.54364	0.31729	*ATP6V0C, CTSS*
Cell cycle *Homo sapiens* hsa04110	2/124	0.21111	0.55787	-0.46921	0.27385	*STAG2, YWHAH*
Apoptosis *Homo sapiens* hsa04210	2/140	0.25142	0.57355	-0.44476	0.24725	*CTSS, ATF4*
FoxO signaling pathway *Homo sapiens* hsa04068	2/133	0.23372	0.56974	-0.42847	0.24105	*MAPK14, GABARAP*
Natural killer cell mediated cytotoxicity *Homo sapiens* hsa04650	2/135	0.23877	0.56974	-0.42325	0.23811	*PPP3CC, FCER1G*
Signaling pathways regulating pluripotency of stem cells *Homo sapiens* hsa04550	2/142	0.25649	0.57355	-0.33216	0.18465	*WNT7B, MAPK14*
Protein processing in endoplasmic reticulum *Homo sapiens* hsa04141	2/169	0.32483	0.59547	-0.04771	0.02473	*MBTPS2, ATF4*
cGMP-PKG signaling pathway *Homo sapiens* hsa04022	2/167	0.31979	0.59547	-0.04469	0.02317	*PPP3CC, ATF4*
Phenylalanine, tyrosine and tryptophan biosynthesis *Homo sapiens* hsa00400	1/5	0.04013	0.48782	0.05551	-0.03985	*GOT2*

Focal adhesion *Homo sapiens* hsa04510	2/202	0.40619	0.64994	0.16507	-0.07113	*CAV2, TNR*
Viral carcinogenesis *Homo sapiens* hsa05203	2/205	0.41338	0.64994	0.17990	-0.07752	*ATF4, YWHAH*
Pathways in cancer *Homo sapiens* hsa05200	1/397	0.93892	0.93892	1.54722	-0.09751	*WNT7B*
Metabolic pathways *Homo sapiens* hsa01100	6/1239	0.87351	0.87989	1.36568	-0.17476	*CHKA, GOT2, CKB, B4GALT6, ATP6V0C, NUDT12*
B cell receptor signaling pathway *Homo sapiens* hsa04662	1/73	0.39807	0.64994	0.73009	-0.31458	*PPP3CC*
Prolactin signaling pathway *Homo sapiens* hsa04917	1/72	0.39391	0.64994	0.73334	-0.31598	*MAPK14*
MicroRNAs in cancer *Homo sapiens* hsa05206	2/297	0.60984	0.70017	0.91785	-0.32715	*CCNG1, TNR*
Rap1 signaling pathway *Homo sapiens* hsa04015	1/211	0.76981	0.78601	1.40573	-0.33849	*MAPK14*
Regulation of actin cytoskeleton *Homo sapiens* hsa04810	1/214	0.77462	0.78601	1.41796	-0.34143	*ARPC4*
ErbB signaling pathway *Homo sapiens* hsa04012	1/87	0.45352	0.65610	0.83236	-0.35079	*AREG*
Progesterone-mediated oocyte maturation *Homo sapiens* hsa04914	1/98	0.49355	0.65610	0.85758	-0.36142	*MAPK14*
Aldosterone synthesis and secretion *Homo sapiens* hsa04925	1/81	0.43040	0.65610	0.86422	-0.36422	*ATF4*
Inflammatory mediator regulation of TRP channels *Homo sapiens* hsa04750	1/98	0.49355	0.65610	0.88226	-0.37183	*MAPK14*
AGE-RAGE signaling pathway in diabetic complications *Homo sapiens* hsa04933	1/101	0.50396	0.65610	0.88454	-0.37279	*MAPK14*
Biosynthesis of amino acids *Homo sapiens* hsa01230	1/74	0.40221	0.64994	0.86869	-0.37430	*GOT2*
Fc gamma R-mediated phagocytosis *Homo sapiens* hsa04666	1/93	0.47572	0.65610	0.89261	-0.37619	*ARPC4*
Purine metabolism *Homo sapiens* hsa00230	1/176	0.70572	0.74068	1.27802	-0.38364	*GUCY2C*
Thyroid hormone synthesis *Homo sapiens* hsa04918	1/71	0.38971	0.64994	0.89088	-0.38386	*ATF4*
Prostate cancer *Homo sapiens* hsa05215	1/89	0.46102	0.65610	0.91310	-0.38482	*ATF4*
p53 signaling pathway *Homo sapiens* hsa04115	1/69	0.38124	0.64994	0.89548	-0.38584	*CCNG1*
RNA degradation *Homo sapiens* hsa03018	1/77	0.41445	0.64994	0.89694	-0.38647	*TOB1*
Estrogen signaling pathway *Homo sapiens* hsa04915	1/99	0.49704	0.65610	0.92137	-0.38831	*ATF4*
Insulin secretion *Homo sapiens* hsa04911	1/85	0.44592	0.65610	0.92184	-0.38851	*ATF4*
Longevity regulating pathway - mammal *Homo sapiens* hsa04211	1/94	0.47934	0.65610	0.92617	-0.39033	*ATF4*
RIG-I-like receptor signaling pathway *Homo sapiens* hsa04622	1/70	0.38549	0.64994	0.90794	-0.39121	*MAPK14*
AMPK signaling pathway *Homo sapiens* hsa04152	1/124	0.57710	0.69859	1.09966	-0.39444	*PPP2R5C*
Hepatitis C *Homo sapiens* hsa05160	1/133	0.60274	0.70017	1.11265	-0.39659	*MAPK14*
Melanogenesis *Homo sapiens* hsa04916	1/100	0.50051	0.65610	0.94199	-0.39700	*WNT7B*
Insulin resistance *Homo sapiens* hsa04931	1/109	0.53070	0.67190	1.00130	-0.39816	*PTPN1*
GABAergic synapse *Homo sapiens* hsa04727	1/88	0.45728	0.65610	0.95203	-0.40123	*GABARAP*
Alzheimer's disease *Homo sapiens* hsa05010	1/168	0.68878	0.74068	1.33705	-0.40136	*PPP3CC*
Jak-STAT signaling pathway *Homo sapiens* hsa04630	1/158	0.66626	0.72397	1.24701	-0.40280	*FHL1*
Calcium signaling pathway *Homo sapiens* hsa04020	1/180	0.71385	0.74068	1.34427	-0.40353	*PPP3CC*
Influenza A *Homo sapiens* hsa05164	1/175	0.70365	0.74068	1.34444	-0.40358	*MAPK14*
Ribosome *Homo sapiens* hsa03010	1/137	0.61365	0.70017	1.13846	-0.40578	*RPS27L*

Appendix

Alcoholism *Homo sapiens* hsa05034	1/179	0.71184	0.74068	1.35593	-0.40703	*ATF4*
Transcriptional misregulation in cancer *Homo sapiens* hsa05202	1/180	0.71385	0.74068	1.35959	-0.40813	*ETV5*
Herpes simplex infection *Homo sapiens* hsa05168	1/185	0.72370	0.74530	1.39315	-0.40954	*TAF6*
Toll-like receptor signaling pathway *Homo sapiens* hsa04620	1/106	0.52084	0.66552	1.00709	-0.41007	*MAPK14*
Rheumatoid arthritis *Homo sapiens* hsa05323	1/90	0.46474	0.65610	0.97429	-0.41061	*ATP6V0C*
Axon guidance *Homo sapiens* hsa04360	1/127	0.58582	0.70017	1.15254	-0.41080	*PPP3CC*
Renin secretion *Homo sapiens* hsa04924	1/146	0.35955	0.64440	0.93550	-0.41110	*PPP3CC*
Hepatitis B *Homo sapiens* hsa05161	1/146	0.63712	0.70906	1.19878	-0.41217	*ATF4*
Non-alcoholic fatty liver disease (NAFLD) *Homo sapiens* hsa04932	1/151	0.64955	0.71711	1.24017	-0.41239	*ATF4*
Tight junction *Homo sapiens* hsa04530	1/139	0.61899	0.70017	1.16007	-0.41349	*TJP1*
Ubiquitin mediated proteolysis *Homo sapiens* hsa04120	1/137	0.61365	0.70017	1.16125	-0.41391	*UBE2K*
Cholinergic synapse *Homo sapiens* hsa04725	1/111	0.53717	0.67361	1.05122	-0.41535	*ATF4*
Pertussis *Homo sapiens* hsa05133	1/75	0.40632	0.64994	0.96683	-0.41659	*MAPK14*
Phospholipase D signaling pathway *Homo sapiens* hsa04072	1/144	0.63203	0.70906	1.21268	-0.41695	*FCER1G*
Gap junction *Homo sapiens* hsa04540	1/88	0.45728	0.65610	0.98966	-0.41709	*TJP1*
ECM-receptor interaction *Homo sapiens* hsa04512	1/82	0.43432	0.65610	0.99148	-0.41786	*TNR*
Oxidative phosphorylation *Homo sapiens* hsa00190	1/133	0.60274	0.70017	1.17610	-0.41920	*ATP6V0C*
Choline metabolism in cancer *Homo sapiens* hsa05231	1/101	0.50396	0.65610	1.00503	-0.42356	*CHKA*
Oxytocin signaling pathway *Homo sapiens* hsa04921	1/158	0.66626	0.72397	1.32204	-0.42703	*PPP3CC*
Insulin signaling pathway *Homo sapiens* hsa04910	1/139	0.61899	0.70017	1.20342	-0.42894	*PTPN1*
Chagas disease (American trypanosomiasis) *Homo sapiens* hsa05142	1/104	0.51416	0.66312	1.04955	-0.43115	*MAPK14*
Peroxisome *Homo sapiens* hsa04146	1/83	0.43821	0.65610	1.03434	-0.43592	*NUDT12*
Glutamatergic synapse *Homo sapiens* hsa04724	1/114	0.54670	0.67361	1.10687	-0.43733	*PPP3CC*
Carbon metabolism *Homo sapiens* hsa01200	1/113	0.54354	0.67361	1.13376	-0.44796	*GOT2*
Toxoplasmosis *Homo sapiens* hsa05145	1/118	0.55910	0.68280	1.18788	-0.45324	*MAPK14*
Antigen processing and presentation *Homo sapiens* hsa04612	1/77	0.41445	0.64994	1.05358	-0.45397	*CTSS*
Retrograde endocannabinoid signaling *Homo sapiens* hsa04723	1/101	0.50396	0.65610	1.09017	-0.45945	*MAPK14*
Pathogenic Escherichia coli infection *Homo sapiens* hsa05130	1/55	0.31863	0.59547	1.12310	-0.58222	*ARPC4*
NOD-like receptor signaling pathway *Homo sapiens* hsa04621	1/57	0.32794	0.59547	1.17225	-0.60771	*MAPK14*
Basal cell carcinoma *Homo sapiens* hsa05217	1/55	0.31863	0.59547	1.22080	-0.63287	*WNT7B*
Cocaine addiction *Homo sapiens* hsa05030	1/49	0.28993	0.58839	1.21732	-0.64562	*ATF4*
Vasopressin-regulated water reabsorption *Homo sapiens* hsa04962	1/44	0.26512	0.57355	1.19236	-0.66285	*ARHGDIA*
Asthma *Homo sapiens* hsa05310	1/31	0.19656	0.55787	1.15504	-0.67413	*FCER1G*
Glutathione metabolism *Homo sapiens* hsa00480	1/52	0.30443	0.59547	1.30612	-0.67710	*GPX4*
Fanconi anemia pathway *Homo sapiens* hsa03460	1/53	0.30919	0.59547	1.33088	-0.68994	*RPA1*

Hedgehog signaling pathway *Homo sapiens* hsa04340	1/50	0.29480	0.58959	1.31007	-0.69214	*WNT7B*
Nicotinate and nicotinamide metabolism *Homo sapiens* hsa00760	1/29	0.18547	0.55787	1.19801	-0.69921	*NUDT12*
Basal transcription factors *Homo sapiens* hsa03022	1/45	0.27015	0.57355	1.25830	-0.69951	*TAF6*
Nucleotide excision repair *Homo sapiens* hsa03420	1/47	0.28011	0.57694	1.32561	-0.72912	*RPA1*
Fat digestion and absorption *Homo sapiens* hsa04975	1/41	0.24982	0.57355	1.33342	-0.74127	*GOT2*
Sphingolipid metabolism *Homo sapiens* hsa00600	1/47	0.28011	0.57694	1.42959	-0.78631	*B4GALT6*
Homologous recombination *Homo sapiens* hsa03440	1/29	0.18547	0.55787	1.36653	-0.79756	*RPA1*
Mismatch repair *Homo sapiens* hsa03430	1/23	0.15130	0.55787	1.37242	-0.80100	*RPA1*
Proteasome *Homo sapiens* hsa03050	1/44	0.26512	0.57355	1.47027	-0.81735	*PSMD3*
Cysteine and methionine metabolism *Homo sapiens* hsa00270	1/45	0.27015	0.57355	1.47255	-0.81862	*GOT2*
Collecting duct acid secretion *Homo sapiens* hsa04966	1/27	0.17423	0.55787	1.45342	-0.84827	*ATP6V0C*
Alanine, aspartate and glutamate metabolism *Homo sapiens* hsa00250	1/35	0.21830	0.55787	1.45576	-0.84964	*GOT2*
Tyrosine metabolism *Homo sapiens* hsa00350	1/35	0.21830	0.55787	1.47411	-0.86034	*GOT2*
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate *Homo sapiens* hsa00532	1/20	0.13369	0.54261	1.40982	-0.86191	*CHST15*
Regulation of autophagy *Homo sapiens* hsa04140	1/397	0.23946	0.56974	1.54396	-0.86859	*GABARAP*
Circadian rhythm *Homo sapiens* hsa04710	1/30	0.19103	0.55787	1.48911	-0.86910	*BHLHE40*
Phenylalanine metabolism *Homo sapiens* hsa00360	1/17	0.11572	0.51514	1.32865	-0.88132	*GOT2*
DNA replication *Homo sapiens* hsa03030	1/36	0.22364	0.56113	1.55622	-0.89919	*RPA1*
SNARE interactions in vesicular transport *Homo sapiens* hsa04130	1/34	0.21292	0.55787	1.58438	-0.92470	*STX1B*
2-Oxocarboxylic acid metabolism *Homo sapiens* hsa01210	1/17	0.11572	0.51514	1.40090	-0.92924	*GOT2*
Arginine biosynthesis *Homo sapiens* hsa00220	1/20	0.13369	0.54261	1.54162	-0.94249	*GOT2*

Table A13: GO terms from EnrichR KEGG for genes upregulated in mutant erythrocytes.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Vasopressin-regulated water reabsorption *Homo sapiens* hsa04962	1/44	0.05018	0.21163	-1.84738	2.86885	*RAB11B*
Regulation of autophagy *Homo sapiens* hsa04140	1/39	0.04472	0.21163	-1.84714	2.86848	*GABARAP*
Collecting duct acid secretion *Homo sapiens* hsa04966	1/27	0.03149	0.21163	-1.84280	2.86175	*SLC12A7*
GABAergic synapse *Homo sapiens* hsa04727	1/88	0.09711	0.21163	-1.70520	2.64807	*GABARAP*
Neurotrophin signaling pathway *Homo sapiens* hsa04722	1/120	0.12995	0.21163	-1.69916	2.63868	*PRDM4*
Cardiac muscle contraction *Homo sapiens* hsa04260	1/78	0.08663	0.21163	-1.66888	2.59166	*COX7A1*
Mucin type O-Glycan biosynthesis *Homo sapiens* hsa00512	1/31	0.03592	0.21163	-1.66035	2.57841	*GALNT11*
Non-alcoholic fatty liver disease (NAFLD) *Homo sapiens* hsa04932	1/151	0.16076	0.21163	-1.62736	2.52718	*COX7A1*
Oxidative phosphorylation *Homo sapiens* hsa00190	1/133	0.14299	0.21163	-1.59458	2.47628	*COX7A1*

AMPK signaling pathway *Homo sapiens* hsa04152	1/124	0.13398	0.21163	-1.58752	2.46531	*RAB11B*
FoxO signaling pathway *Homo sapiens* hsa04068	1/133	0.14299	0.21163	-1.57211	2.44139	*GABARAP*
Glycerophospholipid metabolism *Homo sapiens* hsa00564	1/95	0.10438	0.21163	-1.56249	2.42645	*ETNK1*
Alzheimer's disease *Homo sapiens* hsa05010	1/168	0.17725	0.21163	-1.49145	2.31612	*COX7A1*
Parkinson's disease *Homo sapiens* hsa05012	1/142	0.15192	0.21163	-1.48113	2.30009	*COX7A1*
Protein processing in endoplasmic reticulum *Homo sapiens* hsa04141	1/169	0.17821	0.21163	-1.45621	2.26139	*UBE2D1*
Huntington's disease *Homo sapiens* hsa05016	1/193	0.20098	0.22463	-1.50423	2.24627	*COX7A1*
Ubiquitin mediated proteolysis *Homo sapiens* hsa04120	1/137	0.14697	0.21163	-1.38342	2.14837	*UBE2D1*
Endocytosis *Homo sapiens* hsa04144	1/259	0.26078	0.27527	-1.54434	1.99220	*RAB11B*
Metabolic pathways *Homo sapiens* hsa01100	2/1239	0.42693	0.42693	-1.56460	1.33168	*GALNT11, ETNK1*

Table A14 GO terms from EnrichR KEGG for genes downregulated in mutant erythrocytes.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Glucagon signaling pathway *Homo sapiens* hsa04922	5/101	0.00034	0.03676	-1.91641	6.33064	*CREB3, CREBBP, SIK2, PRKAB1, CAMK2G*
Proteasome *Homo sapiens* hsa03050	3/44	0.00248	0.13381	-1.68762	3.39435	*PSMA3, PSME4, PSMC2*
Epstein-Barr virus infection *Homo sapiens* hsa05169	5/202	0.00659	0.17751	-1.87394	3.23952	*CREBBP, XPO1, PSMC2, CSNK2A2, POLR2F*
Huntington's disease *Homo sapiens* hsa05016	5/193	0.00547	0.17751	-1.85137	3.20050	*CREB3, NDUFA8, CREBBP, POLR2F, ATP5H*
Parkinson's disease *Homo sapiens* hsa05012	4/142	0.00986	0.17751	-1.64854	2.84986	*NDUFA8, ATP5H, UBE2L3, UBE2J1*
Ubiquitin mediated proteolysis *Homo sapiens* hsa04120	4/137	0.00875	0.17751	-1.54990	2.67934	*UBE2A, UBE2L3, BIRC3, UBE2J1*
Ribosome biogenesis in eukaryotes *Homo sapiens* hsa03008	3/89	0.01603	0.23328	-1.80410	2.62588	*POP7, XPO1, CSNK2A2*
Influenza A *Homo sapiens* hsa05164	3/89	0.01958	0.23328	-1.76218	2.56487	*IVNS1ABP, CREBBP, XPO1, IFNAR1*
Melanogenesis *Homo sapiens* hsa04916	4/175	0.02160	0.23328	-1.68317	2.44986	*CREB3, CREBBP, CAMK2G*
Protein processing in endoplasmic reticulum *Homo sapiens* hsa04141	3/100	0.01749	0.23328	-1.64719	2.39749	*PPP1R15A, SAR1B, PLAA, UBE2J1*
Insulin resistance *Homo sapiens* hsa04931	4/169	0.02686	0.26153	-1.59642	2.14111	*CREB3, SLC27A1, PRKAB1*
Vasopressin-regulated water reabsorption *Homo sapiens* hsa04962	3/109	0.02906	0.26153	-1.42810	1.91536	*CREB3, DYNC1LI1*
Hepatitis B *Homo sapiens* hsa05161	3/146	0.05499	0.40909	-1.60501	1.43458	*CREB3, CREBBP, IFNAR1*
Wnt signaling pathway *Homo sapiens* hsa04310	3/146	0.05146	0.40909	-1.57478	1.40756	*CREBBP, CSNK2A2, CAMK2G*
Amphetamine addiction *Homo sapiens* hsa05031	3/142	0.06061	0.40909	-1.46367	1.30825	*CREB3, CAMK2G*
Long-term potentiation *Homo sapiens* hsa04720	2/67	0.05906	0.40909	-1.39703	1.24869	*CREBBP, CAMK2G*
Adherens junction *Homo sapiens* hsa04520	2/66	0.07180	0.45611	-1.33283	1.04629	*CREBBP, CSNK2A2*
Aldosterone synthesis and secretion *Homo sapiens* hsa04925	2/74	0.08360	0.50160	-1.47507	1.01773	*CREB3, CAMK2G*
Prostate cancer *Homo sapiens* hsa05215	2/81	0.09776	0.50215	-1.37705	0.94859	*CREB3, CREBBP*
Insulin secretion *Homo sapiens* hsa04911	2/89	0.09060	0.50215	-1.34608	0.92725	*CREB3, CAMK2G*
cAMP signaling pathway *Homo sapiens* hsa04024	3/199	0.11192	0.50366	-1.35006	0.92594	*CREB3, CREBBP, CAMK2G*

Herpes simplex infection *Homo sapiens* hsa05168	3/185	0.09519	0.50215	-1.32443	0.91234	*CREBBP, CSNK2A2, IFNAR1*
Longevity regulating pathway - mammal *Homo sapiens* hsa04211	2/94	0.10694	0.50215	-1.28346	0.88412	*CREB3, PRKAB1*
Cholinergic synapse *Homo sapiens* hsa04725	2/111	0.13971	0.53620	-1.31869	0.82187	*CREB3, CAMK2G*
NF-kappa B signaling pathway *Homo sapiens* hsa04064	2/93	0.10509	0.50215	-1.17033	0.80619	*CSNK2A2, BIRC3*
TNF signaling pathway *Homo sapiens* hsa04668	2/110	0.13772	0.53620	-1.24154	0.77379	*CREB3, BIRC3*
HIF-1 signaling pathway *Homo sapiens* hsa04066	2/103	0.12401	0.53573	-1.19770	0.74751	*CREBBP, CAMK2G*
Dopaminergic synapse *Homo sapiens* hsa04728	2/129	0.17644	0.53620	-1.09346	0.68150	*CREB3, CAMK2G*
AMPK signaling pathway *Homo sapiens* hsa04152	2/124	0.16607	0.53620	-1.07568	0.67041	*CREB3, PRKAB1*
HTLV-I infection *Homo sapiens* hsa05166	3/258	0.19285	0.53620	-1.04372	0.65050	*CREBBP, XPO1, TCF3*
Osteoclast differentiation *Homo sapiens* hsa04380	2/132	0.18272	0.53620	-1.03592	0.64564	*SQSTM1, IFNAR1*
FoxO signaling pathway *Homo sapiens* hsa04068	2/133	0.18481	0.53620	-0.98759	0.61551	*CREBBP, PRKAB1*
Measles *Homo sapiens* hsa05162	2/136	0.19113	0.53620	-0.97521	0.60780	*CSNK2A2, IFNAR1*
Oxidative phosphorylation *Homo sapiens* hsa00190	2/133	0.18481	0.53620	-0.94723	0.59036	*NDUFA8, ATP5H*
Non-alcoholic fatty liver disease (NAFLD) *Homo sapiens* hsa04932	2/151	0.22313	0.53620	-0.90402	0.56343	*NDUFA8, PRKAB1*
Oxytocin signaling pathway *Homo sapiens* hsa04921	2/158	0.23823	0.53620	-0.87330	0.54429	*PRKAB1, CAMK2G*
Adrenergic signaling in cardiomyocytes *Homo sapiens* hsa04261	2/148	0.21669	0.53620	-0.85994	0.53595	*CREB3, CAMK2G*
Jak-STAT signaling pathway *Homo sapiens* hsa04630	2/158	0.23823	0.53620	-0.83760	0.52203	*CREBBP, IFNAR1*
Tight junction *Homo sapiens* hsa04530	2/139	0.19748	0.53620	-0.81618	0.50868	*CSNK2A2, EXOC3*
cGMP-PKG signaling pathway *Homo sapiens* hsa04022	2/167	0.25773	0.53620	-0.74667	0.46536	*CREB3, ADRB3*
RNA transport *Homo sapiens* hsa03013	2/172	0.26859	0.53620	-0.67870	0.42300	*POP7, XPO1*
Alzheimer's disease *Homo sapiens* hsa05010	2/168	0.25991	0.53620	-0.66542	0.41472	*NDUFA8, ATP5H*
Purine metabolism *Homo sapiens* hsa00230	2/176	0.27728	0.53620	-0.59183	0.36886	*POLR2F, AMPD3*
Calcium signaling pathway *Homo sapiens* hsa04020	2/180	0.28597	0.53620	-0.57147	0.35617	*ADRB3, CAMK2G*
Alcoholism *Homo sapiens* hsa05034	2/179	0.28380	0.53620	-0.54928	0.34234	*CREB3, HAT1*
Tuberculosis *Homo sapiens* hsa05152	2/178	0.28163	0.53620	-0.53751	0.33500	*CREBBP, CAMK2G*
Transcriptional misregulation in cancer *Homo sapiens* hsa05202	2/180	0.28597	0.53620	-0.43063	0.26839	*TCF3, BIRC3*
Viral carcinogenesis *Homo sapiens* hsa05203	2/205	0.33996	0.55496	-0.35430	0.20863	*CREB3, CREBBP*
Endocytosis *Homo sapiens* hsa04144	2/259	0.45179	0.58787	0.05069	-0.02693	*ADRB3, VPS28*
Metabolic pathways *Homo sapiens* hsa01100	4/1239	0.94701	0.94701	0.64553	-0.03515	*NDUFA8, POLR2F, ATP5H, AMPD3*
Olfactory transduction *Homo sapiens* hsa04740	1/415	0.91796	0.92654	0.60400	-0.04608	*CAMK2G*
Regulation of lipolysis in adipocytes *Homo sapiens* hsa04923	1/56	0.28376	0.53620	0.12920	-0.08052	*ADRB3*
RNA polymerase *Homo sapiens* hsa03020	1/32	0.17542	0.53620	0.14636	-0.09122	*POLR2F*
Circadian rhythm *Homo sapiens* hsa04710	1/30	0.16571	0.53620	0.15158	-0.09447	*PRKAB1*
MicroRNAs in cancer *Homo sapiens* hsa05206	1/297	0.83062	0.84630	0.62795	-0.10480	*CREBBP*
Neuroactive ligand-receptor interaction *Homo sapiens* hsa04080	1/277	0.80872	0.83183	0.61911	-0.11400	*ADRB3*

Appendix

Cytokine-cytokine receptor interaction *Homo sapiens* hsa04060	1/265	0.79428	0.82483	0.60411	-0.11634	*IFNAR1*
Cocaine addiction *Homo sapiens* hsa05030	1/49	0.25369	0.53620	0.19589	-0.12209	*CREB3*
Longevity regulating pathway - multiple species *Homo sapiens* hsa04213	1/64	0.31669	0.55496	0.22086	-0.13005	*PRKAB1*
TGF-beta signaling pathway *Homo sapiens* hsa04350	1/84	0.39267	0.58526	0.24659	-0.13210	*CREBBP*
Notch signaling pathway *Homo sapiens* hsa04330	1/48	0.24929	0.53620	0.22075	-0.13758	*CREBBP*
Legionellosis *Homo sapiens* hsa05134	1/55	0.27954	0.53620	0.22745	-0.14176	*SAR1B*
Glioma *Homo sapiens* hsa05214	1/65	0.32070	0.55496	0.24086	-0.14183	*CAMK2G*
ErbB signaling pathway *Homo sapiens* hsa04012	1/87	0.40333	0.58526	0.26810	-0.14362	*CAMK2G*
Inflammatory mediator regulation of TRP channels *Homo sapiens* hsa04750	1/98	0.44089	0.58787	0.29701	-0.15779	*CAMK2G*
GnRH signaling pathway *Homo sapiens* hsa04912	1/91	0.41727	0.58526	0.29505	-0.15806	*CAMK2G*
Gastric acid secretion *Homo sapiens* hsa04971	1/74	0.35577	0.56505	0.28313	-0.16163	*CAMK2G*
Pathways in cancer *Homo sapiens* hsa05200	2/397	0.68336	0.73803	0.53287	-0.16187	*CREBBP, BIRC3*
Circadian entrainment *Homo sapiens* hsa04713	1/95	0.43088	0.58787	0.31712	-0.16847	*CAMK2G*
Rap1 signaling pathway *Homo sapiens* hsa04015	1/211	0.71500	0.74971	0.60165	-0.17332	*DOCK4*
Estrogen signaling pathway *Homo sapiens* hsa04915	1/99	0.44419	0.58787	0.32971	-0.17516	*CREB3*
Focal adhesion *Homo sapiens* hsa04510	1/202	0.69916	0.74220	0.59254	-0.17666	*BIRC3*
Renal cell carcinoma *Homo sapiens* hsa05211	1/66	0.32468	0.55496	0.30055	-0.17698	*CREBBP*
Proteoglycans in cancer *Homo sapiens* hsa05205	1/203	0.70096	0.74220	0.60086	-0.17914	*CAMK2G*
Base excision repair *Homo sapiens* hsa03410	1/33	0.18024	0.53620	0.29537	-0.18409	*LIG3*
Salivary secretion *Homo sapiens* hsa04970	1/89	0.41034	0.58526	0.34851	-0.18670	*ADRB3*
Choline metabolism in cancer *Homo sapiens* hsa05231	1/101	0.45072	0.58787	0.35270	-0.18737	*GPCPD1*
Small cell lung cancer *Homo sapiens* hsa05222	1/86	0.39980	0.58526	0.35745	-0.19149	*BIRC3*
Peroxisome *Homo sapiens* hsa04146	1/83	0.38908	0.58526	0.36089	-0.19333	*PEX13*
Toll-like receptor signaling pathway *Homo sapiens* hsa04620	1/106	0.46674	0.59304	0.38164	-0.19941	*IFNAR1*
Renin secretion *Homo sapiens* hsa04924	1/64	0.31669	0.55496	0.33950	-0.19992	*ADRB3*
PI3K-Akt signaling pathway *Homo sapiens* hsa04151	2/341	0.59999	0.65454	0.47378	-0.20080	*CREB3, IFNAR1*
Thyroid hormone signaling pathway *Homo sapiens* hsa04919	1/118	0.50335	0.61800	0.41932	-0.20181	*CREBBP*
Neurotrophin signaling pathway *Homo sapiens* hsa04722	1/120	0.50921	0.61800	0.42468	-0.20438	*CAMK2G*
Salmonella infection *Homo sapiens* hsa05132	1/86	0.39980	0.58526	0.38909	-0.20844	*DYNC1LI1*
Lysosome *Homo sapiens* hsa04142	1/123	0.51787	0.61800	0.43611	-0.20988	*SLC11A2*
Adipocytokine signaling pathway *Homo sapiens* hsa04920	1/70	0.34041	0.55496	0.36521	-0.21506	*PRKAB1*
Spliceosome *Homo sapiens* hsa03040	1/134	0.54837	0.63312	0.47115	-0.21537	*ISY1*
Hypertrophic cardiomyopathy (HCM) *Homo sapiens* hsa05410	1/83	0.38908	0.58526	0.40863	-0.21891	*PRKAB1*
Apoptosis *Homo sapiens* hsa04210	1/140	0.56420	0.63392	0.48055	-0.21905	*BIRC3*
Glycerophospholipid metabolism *Homo sapiens* hsa00564	1/95	0.43088	0.58787	0.41572	-0.22085	*GPCPD1*
Toxoplasmosis *Homo sapiens* hsa05145	1/118	0.50335	0.61800	0.46390	-0.22326	*BIRC3*

Hepatitis C *Homo sapiens* hsa05160	1/133	0.54567	0.63312	0.49692	-0.22714	*IFNAR1*
Oocyte meiosis *Homo sapiens* hsa04114	1/123	0.51787	0.61800	0.47418	-0.22821	*CAMK2G*
Regulation of autophagy *Homo sapiens* hsa04140	1/39	0.20857	0.53620	0.37249	-0.23215	*ATG4C*
Mineral absorption *Homo sapiens* hsa04978	1/51	0.26240	0.53620	0.37635	-0.23456	*SLC11A2*
Cytosolic DNA-sensing pathway *Homo sapiens* hsa04623	1/64	0.31669	0.55496	0.39971	-0.23538	*POLR2F*
Natural killer cell mediated cytotoxicity *Homo sapiens* hsa04650	1/135	0.55104	0.63312	0.51746	-0.23653	*IFNAR1*
Insulin signaling pathway *Homo sapiens* hsa04910	1/139	0.56160	0.63392	0.53938	-0.24586	*PRKAB1*
PPAR signaling pathway *Homo sapiens* hsa03320	1/69	0.33651	0.55496	0.41849	-0.24644	*SLC27A1*
Pyruvate metabolism *Homo sapiens* hsa00620	1/40	0.21320	0.53620	0.39797	-0.24803	*HAGH*
Pyrimidine metabolism *Homo sapiens* hsa00240	1/105	0.46357	0.59304	0.47608	-0.24875	*POLR2F*
Thyroid hormone synthesis *Homo sapiens* hsa04918	1/71	0.34428	0.55496	0.43361	-0.25534	*CREB3*
Signaling pathways regulating pluripotency of stem cells *Homo sapiens* hsa04550	1/142	0.56936	0.63392	0.56582	-0.25792	*TCF3*
Cell cycle *Homo sapiens* hsa04110	1/124	0.52072	0.61800	0.54199	-0.26084	*CREBBP*
Phagosome *Homo sapiens* hsa04145	1/154	0.59907	0.65454	0.62199	-0.26362	*DYNC1LI1*
Protein digestion and absorption *Homo sapiens* hsa04974	1/90	0.41381	0.58526	0.52272	-0.28002	*SLC3A2*
NOD-like receptor signaling pathway *Homo sapiens* hsa04621	1/57	0.28796	0.53620	0.45680	-0.28470	*BIRC3*

Table A4: GO terms from EnrichR KEGG for genes upregulated in mutant monocytes.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Ribosome *Homo sapiens* hsa03010	4/137	0.00059	0.02954	-1.74614	6.15014	*RPS25, RPS9, RPL12, RPL13*
p53 signaling pathway *Homo sapiens* hsa04115	3/69	0.00104	0.02954	-1.73739	6.11932	*CDK6, RRM2, CCND1*
Non-small cell lung cancer *Homo sapiens* hsa05223	2/56	0.01156	0.14075	-2.02977	3.97997	*CDK6, CCND1*
Glioma *Homo sapiens* hsa05214	2/65	0.01522	0.14075	-2.02243	3.96558	*CDK6, CCND1*
Pancreatic cancer *Homo sapiens* hsa05212	2/66	0.01566	0.14075	-1.86302	3.65301	*CDK6, CCND1*
Melanoma *Homo sapiens* hsa05218	2/71	0.01791	0.14075	-1.78760	3.50513	*CDK6, CCND1*
Chronic myeloid leukemia *Homo sapiens* hsa05220	2/73	0.01885	0.14075	-1.74118	3.41411	*CDK6, CCND1*
Small cell lung cancer *Homo sapiens* hsa05222	2/86	0.02543	0.16107	-1.63486	2.98511	*CDK6, CCND1*
Cell cycle *Homo sapiens* hsa04110	2/124	0.04907	0.21944	-1.48040	2.24527	*CDK6, CCND1*
Measles *Homo sapiens* hsa05162	2/136	0.05772	0.23501	-1.53016	2.21583	*CDK6, CCND1*
Hepatitis B *Homo sapiens* hsa05161	2/146	0.06532	0.24821	-1.57952	2.20105	*CDK6, CCND1*
Viral carcinogenesis *Homo sapiens* hsa05203	2/205	0.11599	0.30279	-1.47358	1.76053	*CDK6, CCND1*
Purine metabolism *Homo sapiens* hsa00230	2/176	0.08995	0.28485	-1.37291	1.72411	*RRM2, PAPSS2*
Metabolic pathways *Homo sapiens* hsa01100	6/1239	0.12749	0.30279	-1.44043	1.72093	*PDXK, RRM2, UROD, PAPSS2, PAFAH1B1, IDH3A*
Bladder cancer *Homo sapiens* hsa05219	1/41	0.11309	0.30279	-0.99388	1.18741	*CCND1*
Thyroid cancer *Homo sapiens* hsa05216	1/29	0.08208	0.28401	-0.92694	1.16678	*CCND1*

Appendix

MicroRNAs in cancer *Homo sapiens* hsa05206	2/297	0.20815	0.35954	-1.13560	1.16165	*CDK6, CCND1*
Acute myeloid leukemia *Homo sapiens* hsa05221	1/57	0.15288	0.32111	-0.96977	1.10164	*CCND1*
Endometrial cancer *Homo sapiens* hsa05213	1/52	0.14063	0.30831	-0.93250	1.09722	*CCND1*
PI3K-Akt signaling pathway *Homo sapiens* hsa04151	2/341	0.25500	0.37270	-1.08182	1.06775	*CDK6, CCND1*
Prolactin signaling pathway *Homo sapiens* hsa04917	1/72	0.18865	0.35543	-0.99399	1.02820	*CCND1*
Colorectal cancer *Homo sapiens* hsa05210	1/62	0.16497	0.32425	-0.90190	1.01577	*CCND1*
Prostate cancer *Homo sapiens* hsa05215	1/89	0.22744	0.36364	-0.98009	0.99144	*CCND1*
AGE-RAGE signaling pathway in diabetic complications *Homo sapiens* hsa04933	1/101	0.25376	0.37270	-1.00367	0.99061	*CCND1*
Porphyrin and chlorophyll metabolism *Homo sapiens* hsa00860	1/42	0.11563	0.30279	-0.82529	0.98600	*UROD*
Viral myocarditis *Homo sapiens* hsa05416	1/59	0.15774	0.32111	-0.81879	0.93013	*CCND1*
Ribosome biogenesis in eukaryotes *Homo sapiens* hsa03008	1/89	0.22744	0.36364	-0.91788	0.92851	*NOP58*
Glutathione metabolism *Homo sapiens* hsa00480	1/52	0.14063	0.30831	-0.78077	0.91870	*RRM2*
RNA degradation *Homo sapiens* hsa03018	1/77	0.20025	0.35669	-0.87351	0.90049	*DDX6*
Pathways in cancer *Homo sapiens* hsa05200	2/397	0.31522	0.38914	-0.93527	0.88271	*CDK6, CCND1*
ABC transporters *Homo sapiens* hsa02010	1/44	0.12068	0.30279	-0.71895	0.85895	*ABCB10*
Ether lipid metabolism *Homo sapiens* hsa00565	1/45	0.12320	0.30279	-0.71146	0.85001	*PAFAH1B1*
Biosynthesis of amino acids *Homo sapiens* hsa01230	1/74	0.19330	0.35543	-0.80559	0.83332	*IDH3A*
Selenocompound metabolism *Homo sapiens* hsa00450	1/17	0.05005	0.21944	-0.52446	0.79543	*PAPSS2*
Platelet activation *Homo sapiens* hsa04611	1/122	0.29778	0.38914	-0.80795	0.76254	*FERMT3*
Pancreatic secretion *Homo sapiens* hsa04972	1/96	0.24290	0.37270	-0.73547	0.72590	*CPA1*
Thyroid hormone signaling pathway *Homo sapiens* hsa04919	1/118	0.28959	0.38914	-0.76140	0.71862	*CCND1*
Citrate cycle (TCA cycle) *Homo sapiens* hsa00020	1/30	0.08470	0.28401	-0.56962	0.71700	*IDH3A*
Pyrimidine metabolism *Homo sapiens* hsa00240	1/105	0.26234	0.37384	-0.72489	0.71324	*RRM2*
Protein digestion and absorption *Homo sapiens* hsa04974	1/90	0.22967	0.36364	-0.67328	0.68108	*CPA1*
AMPK signaling pathway *Homo sapiens* hsa04152	1/124	0.30184	0.38914	-0.69179	0.65292	*CCND1*
FoxO signaling pathway *Homo sapiens* hsa04068	1/133	0.31984	0.38914	-0.68262	0.64426	*CCND1*
Carbon metabolism *Homo sapiens* hsa01200	1/113	0.27923	0.38819	-0.67733	0.64093	*IDH3A*
Spliceosome *Homo sapiens* hsa03040	1/134	0.32181	0.38914	-0.66273	0.62549	*SNRPD2*
Wnt signaling pathway *Homo sapiens* hsa04310	1/142	0.33740	0.39249	-0.66058	0.61781	*CCND1*
Oxytocin signaling pathway *Homo sapiens* hsa04921	1/158	0.36756	0.40290	-0.66422	0.60382	*CCND1*
Jak-STAT signaling pathway *Homo sapiens* hsa04630	1/158	0.36756	0.40290	-0.63422	0.57655	*CCND1*
Hippo signaling pathway *Homo sapiens* hsa04390	1/153	0.35828	0.40290	-0.55780	0.50707	*CCND1*
Protein processing in endoplasmic reticulum *Homo sapiens* hsa04141	1/169	0.38753	0.41678	-0.57291	0.50141	*UBE4B*
Ubiquitin mediated proteolysis *Homo sapiens* hsa04120	1/137	0.32770	0.38914	-0.52727	0.49764	*UBE4B*

2-Oxocarboxylic acid metabolism *Homo sapiens* hsa01210	1/17	0.05005	0.21944	-0.31871	0.48338	*IDH3A*
Proteoglycans in cancer *Homo sapiens* hsa05205	1/203	0.44554	0.45350	-0.58976	0.46636	*CCND1*
Focal adhesion *Homo sapiens* hsa04510	1/202	0.44391	0.45350	-0.56415	0.44611	*CCND1*
Herpes simplex infection *Homo sapiens* hsa05168	1/185	0.41551	0.43860	-0.51392	0.42356	*EEF1D*
HTLV-I infection *Homo sapiens* hsa05166	1/258	0.52848	0.52848	-0.54412	0.34701	*CCND1*
Sulfur metabolism *Homo sapiens* hsa00920	1/10	0.03087	0.17599	0.07236	-0.12571	*PAPSS2*
Vitamin B6 metabolism *Homo sapiens* hsa00750	1/6	0.01975	0.14075	0.17640	-0.34588	*PDXK*

Table A16: GO terms from EnrichR KEGG for genes downregulated in mutant monocytes.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Gastric acid secretion *Homo sapiens* hsa04971	6/74	0.00005	0.00840	-2.02272	9.66656	*CAMK2B, PLCB3, PRKCB, ATP1A1, ATP1B1, CALM1*
Dopaminergic synapse *Homo sapiens* hsa04728	7/129	0.00012	0.01071	-1.86790	8.47353	*CAMK2B, PLCB3, DDC, PRKCB, MAOA, PPP2R5A, CALM1*
Amphetamine addiction *Homo sapiens* hsa05031	5/67	0.00031	0.01794	-1.94473	7.81935	*CAMK2B, DDC, MAOA, PRKCB, CALM1*
Tryptophan metabolism *Homo sapiens* hsa00380	4/40	0.00046	0.02004	-1.84865	7.22868	*DDC, MAOA, KYNU, CYP1A1*
Endocrine and other factor-regulated calcium reabsorption *Homo sapiens* hsa04961	4/47	0.00082	0.02513	-1.81165	6.67366	*PLCB3, PRKCB, ATP1A1, ATP1B1*
Insulin secretion *Homo sapiens* hsa04911	5/85	0.00087	0.02513	-1.74995	6.44637	*CAMK2B, PLCB3, PRKCB, ATP1A1, ATP1B1*
Salivary secretion *Homo sapiens* hsa04970	5/89	0.00106	0.02623	-1.67567	6.10108	*PLCB3, PRKCB, ATP1A1, ATP1B1, CALM1*
Phosphatidylinositol signaling system *Homo sapiens* hsa04070	5/98	0.00160	0.02916	-1.63960	5.79573	*PLCB3, INPP1, PRKCB, CALM1, PIK3C2A*
Adrenergic signaling in cardiomyocytes *Homo sapiens* hsa04261	6/148	0.00169	0.02916	-1.63438	5.77728	*CAMK2B, PLCB3, PPP2R5A, ATP1A1, ATP1B1, CALM1*
Pancreatic secretion *Homo sapiens* hsa04972	5/96	0.00146	0.02916	-1.53503	5.42610	*PLCB3, PRKCB, RAB27B, ATP1A1, ATP1B1*
Long-term potentiation *Homo sapiens* hsa04720	4/66	0.00266	0.04189	-1.53906	4.88311	*CAMK2B, PLCB3, PRKCB, CALM1*
Thyroid hormone synthesis *Homo sapiens* hsa04918	4/71	0.00342	0.04935	-1.50101	4.51624	*PLCB3, PRKCB, ATP1A1, ATP1B1*
Aldosterone synthesis and secretion *Homo sapiens* hsa04925	4/81	0.00536	0.05815	-1.57392	4.47728	*CAMK2B, PLCB3, PRKCB, CALM1*
Cardiac muscle contraction *Homo sapiens* hsa04260	4/78	0.00472	0.05815	-1.41091	4.01357	*ATP1A1, ATP1B1, COX6C, COX5B*
GnRH signaling pathway *Homo sapiens* hsa04912	4/91	0.00793	0.07618	-1.45402	3.74365	*CAMK2B, PLCB3, PRKCB, CALM1*
Inflammatory mediator regulation of TRP channels *Homo sapiens* hsa04750	4/98	0.01013	0.08418	-1.45874	3.61009	*CAMK2B, PLCB3, PRKCB, CALM1*
Circadian entrainment *Homo sapiens* hsa04713	4/95	0.00914	0.08326	-1.41906	3.52748	*CAMK2B, PLCB3, PRKCB, CALM1*
Aldosterone-regulated sodium reabsorption *Homo sapiens* hsa04960	3/39	0.00503	0.05815	-1.19737	3.40612	*PRKCB, ATP1A1, ATP1B1*
Melanogenesis *Homo sapiens* hsa04916	4/100	0.01083	0.08418	-1.34444	3.32720	*CAMK2B, PLCB3, PRKCB, CALM1*
Glucagon signaling pathway *Homo sapiens* hsa04922	4/101	0.01119	0.08418	-1.30017	3.21766	*CAMK2B, PLCB3, CALM1, PHKA2*
Glycine, serine and threonine metabolism *Homo sapiens* hsa00260	3/40	0.00538	0.05815	-1.05552	3.00260	*GLDC, MAOA, AGXT*
Alzheimer's disease *Homo sapiens* hsa05010	5/168	0.01429	0.10303	-1.26617	2.87761	*APH1A, PLCB3, CALM1, COX6C, COX5B*
Carbohydrate digestion and absorption *Homo sapiens* hsa04973	3/45	0.00732	0.07448	-1.08935	2.82924	*PRKCB, ATP1A1, ATP1B1*

Appendix

Calcium signaling pathway *Homo sapiens* hsa04020	5/180	0.01861	0.11252	-1.27727	2.79037	*CAMK2B, PLCB3, PRKCB, CALM1, PHKA2*
Serotonergic synapse *Homo sapiens* hsa04726	4/112	0.01567	0.10844	-1.22251	2.71594	*PLCB3, DDC, PRKCB, MAOA*
Thyroid hormone signaling pathway *Homo sapiens* hsa04919	4/118	0.01853	0.11252	-1.16392	2.54275	*PLCB3, PRKCB, ATP1A1, ATP1B1*
Glioma *Homo sapiens* hsa05214	3/65	0.01886	0.11252	-1.04687	2.28704	*CAMK2B, PRKCB, CALM1*
Focal adhesion *Homo sapiens* hsa04510	5/202	0.02859	0.14989	-1.11735	2.12054	*PRKCB, ACTN1, FLNA, ARHGAP35, VEGFA*
Rap1 signaling pathway *Homo sapiens* hsa04015	5/211	0.03350	0.17047	-1.13215	2.00298	*PLCB3, PRKCB, CDH1, CALM1, VEGFA*
Inositol phosphate metabolism *Homo sapiens* hsa00562	5/71	0.02356	0.13146	-0.90590	1.83809	*PLCB3, INPP1, PIK3C2A*
Oxytocin signaling pathway *Homo sapiens* hsa04921	4/158	0.04571	0.21371	-1.03011	1.58963	*CAMK2B, PLCB3, PRKCB, CALM1*
cGMP-PKG signaling pathway *Homo sapiens* hsa04022	4/167	0.05384	0.23465	-0.95466	1.38395	*PLCB3, ATP1A1, ATP1B1, CALM1*
AGE-RAGE signaling pathway in diabetic complications *Homo sapiens* hsa04933	3/101	0.05561	0.23465	-0.91784	1.33058	*PLCB3, PRKCB, VEGFA*
RNA transport *Homo sapiens* hsa03013	4/172	0.05867	0.23606	-0.84848	1.22493	*CYFIP1, SUMO3, EIF2B1, EIF1B*
Amoebiasis *Homo sapiens* hsa05146	3/100	0.05432	0.23465	-0.77741	1.12700	*PLCB3, PRKCB, ACTN1*
Cholinergic synapse *Homo sapiens* hsa04725	3/111	0.06932	0.24993	-0.77618	1.07622	*CAMK2B, PLCB3, PRKCB*
HIF-1 signaling pathway *Homo sapiens* hsa04066	3/103	0.05824	0.23606	-0.74035	1.06882	*CAMK2B, PRKCB, VEGFA*
Tuberculosis *Homo sapiens* hsa05152	4/178	0.06478	0.24993	-0.75676	1.04929	*CAMK2B, PLK3, CALM1, HLA-DRB1*
Metabolic pathways *Homo sapiens* hsa01100	15/1239	0.08772	0.25477	-0.59611	0.81513	*DDC, RRM2, INPP1, GLDC, MAOA, COX6C, COX5B, PIK3C2A, PLCB3, P4HA1, KYNU, HSD17B2, CYP1A1, ME3, AGXT*
Carbon metabolism *Homo sapiens* hsa01200	3/113	0.07224	0.24993	-0.44448	0.61629	*GLDC, ME3, AGXT*
cAMP signaling pathway *Homo sapiens* hsa04024	4/199	0.08866	0.25477	-0.44539	0.60902	*CAMK2B, ATP1A1, CALM1, ATP1B1*
Proteoglycans in cancer *Homo sapiens* hsa05205	4/203	0.09363	0.25477	-0.43446	0.59409	*CAMK2B, PRKCB, FLNA, VEGFA*
Vascular smooth muscle contraction *Homo sapiens* hsa04270	3/120	0.08285	0.25477	-0.43263	0.59158	*PLCB3, PRKCB, CALM1*
Leukocyte transendothelial migration *Homo sapiens* hsa04670	3/118	0.07975	0.25477	-0.41863	0.57244	*PRKCB, ACTN1, ARHGAP35*
Oocyte meiosis *Homo sapiens* hsa04114	3/123	0.08760	0.25477	-0.40011	0.54711	*CAMK2B, PPP2R5A, CALM1*
Sphingolipid signaling pathway *Homo sapiens* hsa04071	3/120	0.08285	0.25477	-0.38943	0.53251	*PLCB3, PRKCB, PPP2R5A*
Regulation of actin cytoskeleton *Homo sapiens* hsa04810	4/214	0.10800	0.27476	-0.26384	0.34085	*CYFIP1, ACTN1, ARPC1B, ARHGAP35*
Wnt signaling pathway *Homo sapiens* hsa04310	3/142	0.12019	0.29703	-0.12729	0.15453	*CAMK2B, PLCB3, PRKCB*
Systemic lupus erythematosus *Homo sapiens* hsa05322	3/135	0.10770	0.27476	-0.09852	0.12728	*ACTN1, HLA-DRB1, C1QC*
Bladder cancer *Homo sapiens* hsa05219	2/41	0.04971	0.22631	-0.01682	0.02499	*CDH1, VEGFA*
Hippo signaling pathway *Homo sapiens* hsa04390	3/153	0.14084	0.30841	0.13979	-0.16444	*CDH1, TEAD1, STK3*
Ovarian steroidogenesis *Homo sapiens* hsa04913	2/50	0.06963	0.24993	0.13080	-0.18136	*HSD17B2, CYP1A1*
Cocaine addiction *Homo sapiens* hsa05030	2/49	0.06730	0.24993	0.14946	-0.20724	*DDC, MAOA*
Neuroactive ligand-receptor interaction *Homo sapiens* hsa04080	1/277	0.90718	0.90718	2.26189	-0.22034	*GPR35*
Endocytosis *Homo sapiens* hsa04144	1/259	0.89142	0.89660	2.23466	-0.24390	*ARPC1B*
HTLV-I infection *Homo sapiens* hsa05166	1/258	0.89047	0.89660	2.30360	-0.25143	*HLA-DRB1*
African trypanosomiasis *Homo sapiens* hsa05143	2/35	0.03784	0.18184	0.16861	-0.28741	*PLCB3, PRKCB*

Alcoholism *Homo sapiens* hsa05034	3/179	0.19379	0.37669	0.30305	-0.29588	*DDC, MAOA, CALM1*
Arginine and proline metabolism *Homo sapiens* hsa00330	2/50	0.06963	0.24993	0.26149	-0.36258	*P4HA1, MAOA*
Tyrosine metabolism *Homo sapiens* hsa00350	2/35	0.03784	0.18184	0.22707	-0.38706	*DDC, MAOA*
Viral carcinogenesis *Homo sapiens* hsa05203	1/205	0.82661	0.84119	2.31671	-0.40064	*ACTN1*
Huntington's disease *Homo sapiens* hsa05016	3/193	0.22415	0.40036	0.45466	-0.41619	*PLCB3, COX6C, COX5B*
Pathogenic Escherichia coli infection *Homo sapiens* hsa05130	2/55	0.08164	0.25477	0.32319	-0.44193	*CDH1, ARPC1B*
Pathways in cancer *Homo sapiens* hsa05200	5/397	0.24338	0.42106	0.53287	-0.46093	*PLCB3, PRKCB, CDH1, DCC, VEGFA*
Herpes simplex infection *Homo sapiens* hsa05168	1/185	0.79398	0.81277	2.22749	-0.46177	*HLA-DRB1*
Purine metabolism *Homo sapiens* hsa00230	1/176	0.77739	0.80053	2.17745	-0.48444	*RRM2*
Mineral absorption *Homo sapiens* hsa04978	2/51	0.07198	0.24993	0.37635	-0.52184	*ATP1A1, ATP1B1*
mTOR signaling pathway *Homo sapiens* hsa04150	2/60	0.09425	0.25477	0.38661	-0.52865	*PRKCB, VEGFA*
Estrogen signaling pathway *Homo sapiens* hsa04915	2/99	0.20605	0.39607	0.58328	-0.54021	*PLCB3, CALM1*
Staphylococcus aureus infection *Homo sapiens* hsa05150	2/56	0.08412	0.25477	0.39649	-0.54217	*HLA-DRB1, C1QC*
Leishmaniasis *Homo sapiens* hsa05140	2/73	0.12931	0.30476	0.47892	-0.56907	*PRKCB, HLA-DRB1*
Glutathione metabolism *Homo sapiens* hsa00480	2/52	0.07436	0.25224	0.42499	-0.58537	*RRM2, MGST2*
Cytokine-cytokine receptor interaction *Homo sapiens* hsa04060	2/265	0.65920	0.72868	1.85680	-0.58771	*TNFRSF25, VEGFA*
ErbB signaling pathway *Homo sapiens* hsa04012	2/87	0.16983	0.34977	0.56508	-0.59361	*CAMK2B, PRKCB*
VEGF signaling pathway *Homo sapiens* hsa04370	2/61	0.09683	0.25773	0.44214	-0.59948	*PRKCB, VEGFA*
Ras signaling pathway *Homo sapiens* hsa04014	3/227	0.30107	0.47350	0.81076	-0.60613	*PRKCB, CALM1, VEGFA*
Fc gamma R-mediated phagocytosis *Homo sapiens* hsa04666	2/93	0.18781	0.36922	0.60879	-0.60658	*PRKCB, ARPC1B*
Jak-STAT signaling pathway *Homo sapiens* hsa04630	1/158	0.74019	0.76679	2.28931	-0.60792	*MCL1*
Chagas disease (American trypanosomiasis) *Homo sapiens* hsa05142	2/104	0.22140	0.40036	0.66879	-0.61221	*PLCB3, C1QC*
Phagosome *Homo sapiens* hsa04145	1/154	0.73113	0.76196	2.25440	-0.61288	*HLA-DRB1*
Proximal tubule bicarbonate reclamation *Homo sapiens* hsa04964	2/23	0.01816	0.11252	0.28545	-0.62361	*ATP1A1, ATP1B1*
Retrograde endocannabinoid signaling *Homo sapiens* hsa04723	2/101	0.21217	0.39898	0.68590	-0.63024	*PLCB3, PRKCB*
Long-term depression *Homo sapiens* hsa04730	2/60	0.09425	0.25477	0.46636	-0.63771	*PLCB3, PRKCB*
Insulin resistance *Homo sapiens* hsa04931	2/109	0.23684	0.41387	0.73622	-0.64949	*PRKCB, SLC27A2*
Hepatitis B *Homo sapiens* hsa05161	1/146	0.71206	0.74659	2.24382	-0.65574	*PRKCB*
Measles *Homo sapiens* hsa05162	1/136	0.68635	0.73581	2.15267	-0.66042	*DOK1*
Ribosome *Homo sapiens* hsa03010	1/137	0.68902	0.73581	2.15655	-0.66161	*RPS27L*
Phospholipase D signaling pathway *Homo sapiens* hsa04072	1/144	0.70709	0.74590	2.25765	-0.66188	*PLCB3*
Ubiquitin mediated proteolysis *Homo sapiens* hsa04120	1/137	0.68902	0.73581	2.16009	-0.66269	*UBE2Q2*
FoxO signaling pathway *Homo sapiens* hsa04068	1/133	0.67821	0.73581	2.16377	-0.66382	*PLK3*
Apoptosis *Homo sapiens* hsa04210	1/140	0.69690	0.73965	2.20268	-0.66428	*MCL1*
Natural killer cell mediated cytotoxicity *Homo sapiens* hsa04650	1/135	0.68366	0.73581	2.17007	-0.66575	*PRKCB*

Appendix

Protein processing in endoplasmic reticulum *Homo sapiens* hsa04141	2/169	0.41990	0.56511	1.16980	-0.66764	*CRYAB, NFE2L2*
AMPK signaling pathway *Homo sapiens* hsa04152	1/124	0.65250	0.72828	2.12334	-0.67325	*PPP2R5A*
Glutamatergic synapse *Homo sapiens* hsa04724	2/114	0.25235	0.43210	0.80501	-0.67547	*PLCB3, PRKCB*
Platelet activation *Homo sapiens* hsa04611	2/122	0.27722	0.45184	0.85448	-0.67882	*PLCB3, ARHGAP35*
Renin secretion *Homo sapiens* hsa04924	2/64	0.10471	0.27447	0.52575	-0.67975	*PLCB3, CALM1*
Cell cycle *Homo sapiens* hsa04110	1/124	0.65250	0.72828	2.15989	-0.68484	*E2F5*
Drug metabolism - cytochrome P450 *Homo sapiens* hsa00982	2/69	0.11821	0.29638	0.56944	-0.69250	*MAOA, MGST2*
Adherens junction *Homo sapiens* hsa04520	2/74	0.13212	0.30476	0.58388	-0.69379	*CDH1, ACTN1*
Tight junction *Homo sapiens* hsa04530	2/139	0.32986	0.49293	0.99340	-0.70272	*PRKCB, ACTN1*
Neurotrophin signaling pathway *Homo sapiens* hsa04722	2/120	0.27100	0.44651	0.87599	-0.70631	*CAMK2B, CALM1*
Axon guidance *Homo sapiens* hsa04360	1/127	0.66129	0.72868	2.23170	-0.70638	*DCC*
Metabolism of xenobiotics by cytochrome P450 *Homo sapiens* hsa00980	2/73	0.12931	0.30476	0.59447	-0.70638	*CYP1A1, MGST2*
Bile secretion *Homo sapiens* hsa04976	2/71	0.12373	0.30148	0.59089	-0.70851	*ATP1A1, ATP1B1*
Rheumatoid arthritis *Homo sapiens* hsa05323	2/90	0.17878	0.35551	0.68793	-0.71146	*HLA-DRB1, VEGFA*
MAPK signaling pathway *Homo sapiens* hsa04010	3/255	0.36563	0.51543	1.08284	-0.71766	*PRKCB, FLNA, STK3*
Gap junction *Homo sapiens* hsa04540	2/88	0.17281	0.35171	0.69273	-0.72386	*PLCB3, PRKCB*
Insulin signaling pathway *Homo sapiens* hsa04910	2/139	0.32986	0.49293	1.03126	-0.72950	*CALM1, PHKA2*
Olfactory transduction *Homo sapiens* hsa04740	4/415	0.46642	0.58050	1.34406	-0.73098	*CAMK2B, OR2AT4, CALM1, NCALD*
Viral myocarditis *Homo sapiens* hsa05416	2/59	0.09168	0.25477	0.54179	-0.74085	*HLA-DRB1, SGCG*
Parkinson's disease *Homo sapiens* hsa05012	2/142	0.33907	0.49293	1.05398	-0.74557	*COX6C, COX5B*
MicroRNAs in cancer *Homo sapiens* hsa05206	3/297	0.46049	0.58050	1.37686	-0.74882	*PRKCB, VEGFA, MCL1*
Salmonella infection *Homo sapiens* hsa05132	2/86	0.16686	0.34780	0.71661	-0.75683	*ARPC1B, FLNA*
Pyrimidine metabolism *Homo sapiens* hsa00240	2/105	0.22448	0.40036	0.83092	-0.76062	*RRM2, TXNRD3*
Toxoplasmosis *Homo sapiens* hsa05145	1/118	0.63426	0.71717	2.30170	-0.76518	*HLA-DRB1*
Oxidative phosphorylation *Homo sapiens* hsa00190	2/133	0.31136	0.48093	1.04663	-0.76616	*COX6C, COX5B*
Epstein-Barr virus infection *Homo sapiens* hsa05169	2/202	0.51166	0.61470	1.58417	-0.77089	*PSMC3, HLA-DRB1*
Cell adhesion molecules (CAMs) *Homo sapiens* hsa04514	2/142	0.33907	0.49293	1.09076	-0.77159	*CDH1, HLA-DRB1*
Chemical carcinogenesis *Homo sapiens* hsa05204	2/82	0.15510	0.33126	0.71261	-0.78733	*CYP1A1, MGST2*
Non-alcoholic fatty liver disease (NAFLD) *Homo sapiens* hsa04932	2/151	0.36646	0.51543	1.18850	-0.78769	*COX6C, COX5B*
Pertussis *Homo sapiens* hsa05133	2/75	0.13495	0.30718	0.67711	-0.79921	*CALM1, C1QC*
Bacterial invasion of epithelial cells *Homo sapiens* hsa05100	2/78	0.14351	0.31033	0.68372	-0.80003	*CDH1, ARPC1B*
Influenza A *Homo sapiens* hsa05164	2/175	0.43723	0.56873	1.44800	-0.81717	*PRKCB, HLA-DRB1*
Steroid hormone biosynthesis *Homo sapiens* hsa00140	2/58	0.08914	0.25477	0.60624	-0.82898	*HSD17B2, CYP1A1*
PI3K-Akt signaling pathway *Homo sapiens* hsa04151	3/341	0.55323	0.63383	1.82197	-0.83076	*PPP2R5A, VEGFA, MCL1*
Peroxisome *Homo sapiens* hsa04146	2/83	0.15802	0.33339	0.75883	-0.83354	*AGXT, SLC27A2*
Protein digestion and absorption *Homo sapiens* hsa04974	2/90	0.17878	0.35551	0.82172	-0.84982	*ATP1A1, ATP1B1*

Antigen processing and presentation *Homo sapiens* hsa04612	2/77	0.14064	0.30841	0.72633	-0.85439	*B2M, HLA-DRB1*
Arrhythmogenic right ventricular cardiomyopathy (ARVC) *Homo sapiens* hsa05412	2/74	0.13212	0.30476	0.73568	-0.87417	*ACTN1, SGCG*
Chemokine signaling pathway *Homo sapiens* hsa04062	2/187	0.47107	0.58210	1.67726	-0.90758	*PLCB3, PRKCB*
Phenylalanine metabolism *Homo sapiens* hsa00360	2/17	0.01069	0.08418	0.38353	-0.94916	*DDC, MAOA*
Choline metabolism in cancer *Homo sapiens* hsa05231	1/101	0.57730	0.65706	2.33805	-0.98195	*PRKCB*
Glyoxylate and dicarboxylate metabolism *Homo sapiens* hsa00630	2/28	0.02565	0.13866	0.51500	-1.01749	*GLDC, AGXT*
Morphine addiction *Homo sapiens* hsa05032	1/91	0.53981	0.62676	2.40071	-1.12161	*PRKCB*
mRNA surveillance pathway *Homo sapiens* hsa03015	1/91	0.53981	0.62676	2.40336	-1.12284	*PPP2R5A*
NF-kappa B signaling pathway *Homo sapiens* hsa04064	1/93	0.54756	0.63152	2.46684	-1.13384	*PRKCB*
Hematopoietic cell lineage *Homo sapiens* hsa04640	1/88	0.52794	0.62557	2.45225	-1.15034	*HLA-DRB1*
GABAergic synapse *Homo sapiens* hsa04727	1/88	0.52794	0.62557	2.49202	-1.16900	*PRKCB*
Dilated cardiomyopathy *Homo sapiens* hsa05414	1/90	0.53588	0.62676	2.52122	-1.17791	*SGCG*
TGF-beta signaling pathway *Homo sapiens* hsa04350	1/84	0.51164	0.61470	2.42831	-1.18166	*E2F5*
Hypertrophic cardiomyopathy (HCM) *Homo sapiens* hsa05410	1/83	0.50748	0.61470	2.76593	-1.34596	*SGCG*
Complement and coagulation cascades *Homo sapiens* hsa04610	1/79	0.49049	0.60180	2.74107	-1.39199	*C1QC*
B cell receptor signaling pathway *Homo sapiens* hsa04662	1/73	0.46391	0.58050	2.61589	-1.42268	*PRKCB*
Melanoma *Homo sapiens* hsa05218	1/71	0.45475	0.57847	2.63492	-1.44225	*CDH1*
Histidine metabolism *Homo sapiens* hsa00340	2/24	0.01957	0.11288	0.67195	-1.46582	*DDC, MAOA*
Longevity regulating pathway - multiple species *Homo sapiens* hsa04213	1/64	0.42148	0.56511	2.76325	-1.57707	*CRYAB*
Renal cell carcinoma *Homo sapiens* hsa05211	1/66	0.43118	0.56511	2.77114	-1.58158	*VEGFA*
p53 signaling pathway *Homo sapiens* hsa04115	1/69	0.44544	0.57083	2.83549	-1.58977	*RRM2*
PPAR signaling pathway *Homo sapiens* hsa03320	1/69	0.44544	0.57083	2.83923	-1.59186	*SLC27A2*
Pancreatic cancer *Homo sapiens* hsa05212	1/66	0.43118	0.56511	2.79242	-1.59372	*VEGFA*
Colorectal cancer *Homo sapiens* hsa05210	1/62	0.41161	0.56511	2.88049	-1.64399	*DCC*
Shigellosis *Homo sapiens* hsa05131	1/65	0.42635	0.56511	2.88353	-1.64572	*ARPC1B*
Retinol metabolism *Homo sapiens* hsa00830	1/65	0.42635	0.56511	2.90737	-1.65933	*CYP1A1*
Inflammatory bowel disease (IBD) *Homo sapiens* hsa05321	1/65	0.42635	0.56511	2.96544	-1.69247	*HLA-DRB1*
Non-small cell lung cancer *Homo sapiens* hsa05223	1/56	0.38100	0.53155	3.17328	-2.00536	*PRKCB*
Endometrial cancer *Homo sapiens* hsa05213	1/52	0.35973	0.51543	3.10137	-2.05547	*CDH1*
Autoimmune thyroid disease *Homo sapiens* hsa05320	1/53	0.36511	0.51543	3.50050	-2.31999	*HLA-DRB1*
Notch signaling pathway *Homo sapiens* hsa04330	1/48	0.33774	0.49293	3.47977	-2.46154	*APH1A*
Intestinal immune network for IgA production *Homo sapiens* hsa04672	1/48	0.33774	0.49293	3.68536	-2.60697	*HLA-DRB1*
Type I diabetes mellitus *Homo sapiens* hsa04940	1/43	0.30920	0.48093	4.01150	-2.93652	*HLA-DRB1*
Proteasome *Homo sapiens* hsa03050	1/44	0.31500	0.48226	4.05871	-2.95987	*PSMC3*

Graft-versus-host disease *Homo sapiens* hsa05332	1/41	0.29745	0.47211	4.10757	-3.08294	*HLA-DRB1*
Prion diseases *Homo sapiens* hsa05020	1/35	0.26102	0.43419	4.04942	-3.37831	*C1QC*
Pyruvate metabolism *Homo sapiens* hsa00620	1/40	0.29151	0.46695	4.51685	-3.43972	*ME3*
Allograft rejection *Homo sapiens* hsa05330	1/38	0.27946	0.45184	4.34520	-3.45194	*HLA-DRB1*
Thyroid cancer *Homo sapiens* hsa05216	1/29	0.22272	0.40036	3.92208	-3.59025	*CDH1*
Pentose phosphate pathway *Homo sapiens* hsa00030	1/29	0.22272	0.40036	3.96137	-3.62622	*DERA*
Asthma *Homo sapiens* hsa05310	1/31	0.23570	0.41387	4.23748	-3.73828	*HLA-DRB1*
Alanine, aspartate and glutamate metabolism *Homo sapiens* hsa00250	1/35	0.26102	0.43419	4.65382	-3.88254	*AGXT*
SNARE interactions in vesicular transport *Homo sapiens* hsa04130	1/34	0.25477	0.43210	4.99315	-4.18971	*STX5*
Phototransduction *Homo sapiens* hsa04744	1/27	0.20953	0.39834	5.04511	-4.64379	*CALM1*
Selenocompound metabolism *Homo sapiens* hsa00450	1/17	0.14019	0.30841	7.67439	-9.02749	*TXNRD3*

Table A17: GO terms from EnrichR KEGG for genes upregulated in mutant neutrophils.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
MAPK signaling pathway *Homo sapiens* hsa04010	4/255	0.00779	0.26103	-1.99779	2.68328	*MAP4K2, NFATC1, FOS, RELB*
Hepatitis B *Homo sapiens* hsa05161	3/146	0.01069	0.26103	-1.88539	2.53231	*NFATC1, FOS, IKBKE*
B cell receptor signaling pathway *Homo sapiens* hsa04662	2/73	0.02259	0.26103	-1.86290	2.50211	*NFATC1, FOS*
Inflammatory bowel disease (IBD) *Homo sapiens* hsa05321	2/65	0.01827	0.26103	-1.85842	2.49610	*NFATC1, NOD2*
Osteoclast differentiation *Homo sapiens* hsa04380	3/132	0.00816	0.26103	-1.83827	2.46903	*NFATC1, FOS, RELB*
Leishmaniasis *Homo sapiens* hsa05140	2/73	0.02259	0.26103	-1.71108	2.29820	*ITGAM, FOS*
Choline metabolism in cancer *Homo sapiens* hsa05231	2/101	0.04057	0.27964	-1.71774	2.18885	*FOS, DGKZ*
HTLV-I infection *Homo sapiens* hsa05166	3/258	0.04623	0.27964	-1.66822	2.12576	*NFATC1, FOS, RELB*
Phosphatidylinositol signaling system *Homo sapiens* hsa04070	2/98	0.03845	0.27964	-1.66814	2.12565	*ITPKC, DGKZ*
T cell receptor signaling pathway *Homo sapiens* hsa04660	2/104	0.04274	0.27964	-1.63202	2.07963	*NFATC1, FOS*
Toll-like receptor signaling pathway *Homo sapiens* hsa04620	2/106	0.04421	0.27964	-1.63066	2.07790	*FOS, IKBKE*
Pertussis *Homo sapiens* hsa05133	2/75	0.02373	0.26103	-1.54410	2.07392	*ITGAM, FOS*
TNF signaling pathway *Homo sapiens* hsa04668	2/110	0.04721	0.27964	-1.59156	2.02807	*NOD2, FOS*
Oxytocin signaling pathway *Homo sapiens* hsa04921	2/158	0.08835	0.40718	-1.57895	1.41867	*NFATC1, FOS*
Cell adhesion molecules (CAMs) *Homo sapiens* hsa04514	2/142	0.07367	0.37817	-1.38877	1.35044	*ITGAM, SELL*
cAMP signaling pathway *Homo sapiens* hsa04024	2/199	0.12942	0.41622	-1.42760	1.25136	*NFATC1, FOS*
Tuberculosis *Homo sapiens* hsa05152	2/178	0.10783	0.41622	-1.39014	1.21852	*ITGAM, NOD2*
Herpes simplex infection *Homo sapiens* hsa05168	2/185	0.11491	0.41622	-1.32443	1.16092	*FOS, IKBKE*
Thyroid cancer *Homo sapiens* hsa05216	1/29	0.08990	0.40718	-0.86556	0.77769	*CCDC6*
Lysine degradation *Homo sapiens* hsa00310	1/52	0.15354	0.41622	-0.87910	0.77058	*KMT2C*
Legionellosis *Homo sapiens* hsa05134	1/55	0.16152	0.41622	-0.86736	0.76028	*ITGAM*

Colorectal cancer *Homo sapiens* hsa05210	1/62	0.17986	0.41622	-0.86250	0.75603	*FOS*
Basal transcription factors *Homo sapiens* hsa03022	1/45	0.13463	0.41622	-0.83935	0.73573	*GTF2A1L*
Glutathione metabolism *Homo sapiens* hsa00480	1/52	0.15354	0.41622	-0.82715	0.72504	*RRM2*
Nucleotide excision repair *Homo sapiens* hsa03420	1/47	0.14008	0.41622	-0.81554	0.71486	*RFC1*
DNA replication *Homo sapiens* hsa03030	1/36	0.10974	0.41622	-0.81534	0.71468	*RFC1*
Glycerolipid metabolism *Homo sapiens* hsa00561	1/59	0.17205	0.41622	-0.80999	0.71000	*DGKZ*
Amphetamine addiction *Homo sapiens* hsa05031	1/67	0.19273	0.41622	-0.79759	0.69912	*FOS*
Staphylococcus aureus infection *Homo sapiens* hsa05150	1/56	0.16416	0.41622	-0.77751	0.68152	*ITGAM*
Renal cell carcinoma *Homo sapiens* hsa05211	1/66	0.19017	0.41622	-0.77363	0.67812	*FLCN*
NOD-like receptor signaling pathway *Homo sapiens* hsa04621	1/57	0.16680	0.41622	-0.73562	0.64480	*NOD2*
Prolactin signaling pathway *Homo sapiens* hsa04917	1/72	0.20541	0.41622	-0.72303	0.63377	*FOS*
Inositol phosphate metabolism *Homo sapiens* hsa00562	1/71	0.20289	0.41622	-0.69605	0.61012	*ITPKC*
AGE-RAGE signaling pathway in diabetic complications *Homo sapiens* hsa04933	1/101	0.27525	0.42941	-0.68897	0.58241	*NFATC1*
Cytosolic DNA-sensing pathway *Homo sapiens* hsa04623	1/64	0.18503	0.41622	-0.63227	0.55421	*IKBKE*
Circadian entrainment *Homo sapiens* hsa04713	1/95	0.26130	0.42941	-0.65397	0.55283	*FOS*
Shigellosis *Homo sapiens* hsa05131	1/65	0.18761	0.41622	-0.62273	0.54585	*NOD2*
Estrogen signaling pathway *Homo sapiens* hsa04915	1/99	0.27063	0.42941	-0.62822	0.53106	*FOS*
Mismatch repair *Homo sapiens* hsa03430	1/23	0.07256	0.37817	-0.54251	0.52753	*RFC1*
Hematopoietic cell lineage *Homo sapiens* hsa04640	1/88	0.24470	0.42941	-0.61749	0.52199	*ITGAM*
p53 signaling pathway *Homo sapiens* hsa04115	1/69	0.19782	0.41622	-0.59417	0.52082	*RRM2*
Cholinergic synapse *Homo sapiens* hsa04725	1/111	0.29795	0.42949	-0.60486	0.51120	*FOS*
Complement and coagulation cascades *Homo sapiens* hsa04610	1/79	0.22283	0.42941	-0.58501	0.49453	*ITGAM*
Rheumatoid arthritis *Homo sapiens* hsa05323	1/90	0.24948	0.42941	-0.57203	0.48356	*FOS*
RIG-I-like receptor signaling pathway *Homo sapiens* hsa04622	1/70	0.20036	0.41622	-0.54805	0.48039	*IKBKE*
Salmonella infection *Homo sapiens* hsa05132	1/86	0.23989	0.42941	-0.56370	0.47652	*FOS*
Amoebiasis *Homo sapiens* hsa05146	1/100	0.27295	0.42941	-0.55237	0.46694	*ITGAM*
NF-kappa B signaling pathway *Homo sapiens* hsa04064	1/93	0.25659	0.42941	-0.54519	0.46087	*RELB*
Chagas disease (American trypanosomiasis) *Homo sapiens* hsa05142	1/104	0.28214	0.42941	-0.50068	0.42324	*FOS*
Protein digestion and absorption *Homo sapiens* hsa04974	1/90	0.24948	0.42941	-0.49388	0.41750	*SLC3A1*
Glycerophospholipid metabolism *Homo sapiens* hsa00564	1/95	0.26130	0.42941	-0.49096	0.41503	*DGKZ*
Thyroid hormone signaling pathway *Homo sapiens* hsa04919	1/118	0.31343	0.42949	-0.46622	0.39403	*MED14*
Pathways in cancer *Homo sapiens* hsa05200	2/397	0.35741	0.42949	-0.45479	0.38437	*CCDC6, FOS*
Dopaminergic synapse *Homo sapiens* hsa04728	1/129	0.33710	0.42949	-0.42585	0.35991	*FOS*

Pyrimidine metabolism *Homo sapiens* hsa00240	1/105	0.28442	0.42941	-0.42464	0.35897	*RRM2*
Lysosome *Homo sapiens* hsa04142	1/123	0.32429	0.42949	-0.41803	0.35331	*DNASE2B*
Hepatitis C *Homo sapiens* hsa05160	1/133	0.34552	0.42949	-0.40102	0.33893	*IKBKE*
Spliceosome *Homo sapiens* hsa03040	1/134	0.34761	0.42949	-0.39159	0.33095	*SF3A3*
Natural killer cell mediated cytotoxicity *Homo sapiens* hsa04650	1/135	0.34969	0.42949	-0.37240	0.31474	*NFATC1*
Measles *Homo sapiens* hsa05162	1/136	0.35176	0.42949	-0.36982	0.31255	*IKBKE*
Leukocyte transendothelial migration *Homo sapiens* hsa04670	1/118	0.31343	0.42949	-0.36736	0.31048	*ITGAM*
Apoptosis *Homo sapiens* hsa04210	1/140	0.36000	0.42949	-0.34195	0.28900	*FOS*
Ribosome *Homo sapiens* hsa03010	1/137	0.35383	0.42949	-0.26748	0.22606	*RSL24D1*
Wnt signaling pathway *Homo sapiens* hsa04310	1/142	0.36408	0.42949	-0.25427	0.21490	*NFATC1*
Phospholipase D signaling pathway *Homo sapiens* hsa04072	1/144	0.36813	0.42949	-0.24007	0.20290	*DGKZ*
Ubiquitin mediated proteolysis *Homo sapiens* hsa04120	1/137	0.35383	0.42949	-0.21811	0.18433	*UBE4A*
Influenza A *Homo sapiens* hsa05164	1/175	0.42792	0.46739	-0.20887	0.15886	*IKBKE*
cGMP-PKG signaling pathway *Homo sapiens* hsa04022	1/167	0.41303	0.46739	-0.20068	0.15263	*NFATC1*
Calcium signaling pathway *Homo sapiens* hsa04020	1/180	0.43704	0.46739	-0.16815	0.12790	*ITPKC*
Epstein-Barr virus infection *Homo sapiens* hsa05169	1/202	0.47556	0.50010	-0.16959	0.11752	*RELB*
Purine metabolism *Homo sapiens* hsa00230	1/176	0.42976	0.46739	-0.14212	0.10809	*RRM2*
Phagosome *Homo sapiens* hsa04145	1/154	0.38804	0.44596	-0.12220	0.09868	*ITGAM*
Viral carcinogenesis *Homo sapiens* hsa05203	1/205	0.48061	0.50010	-0.12536	0.08687	*GTF2A1L*
Transcriptional misregulation in cancer *Homo sapiens* hsa05202	1/180	0.43704	0.46739	-0.10085	0.07671	*ITGAM*
Rap1 signaling pathway *Homo sapiens* hsa04015	1/211	0.49058	0.50201	-0.10192	0.07024	*ITGAM*
Regulation of actin cytoskeleton *Homo sapiens* hsa04810	1/214	0.49549	0.50201	-0.06598	0.04547	*ITGAM*
Metabolic pathways *Homo sapiens* hsa01100	3/1239	0.77310	0.77310	-0.12429	0.03198	*ITPKC, RRM2, DGKZ*

Table A18: GO terms from EnrichR KEGG for genes downregulated in mutant neutrophils.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Oxidative phosphorylation *Homo sapiens* hsa00190	7/133	0.00002	0.00091	-1.80174	12.61986	*NDUFA13, UQCRB, NDUFS6, ATP5A1, UQCRC2, ATP6V0C, COX7C*
Antigen processing and presentation *Homo sapiens* hsa04612	6/77	0.00001	0.00091	-1.79354	12.56244	*PSME1, B2M, CTSS, HLA-DRB1, CTSB, LGMN*
Alzheimer's disease *Homo sapiens* hsa05010	7/168	0.00008	0.00254	-1.79182	10.70583	*NDUFA13, UQCRB, NDUFS6, ATP5A1, CAPN2, UQCRC2, COX7C*
Lysosome *Homo sapiens* hsa04142	6/123	0.00012	0.00286	-1.69925	9.95417	*ASAH1, TPP1, ATP6V0C, CTSS, LGMN, CTSB*
Parkinson's disease *Homo sapiens* hsa05012	6/142	0.00025	0.00486	-1.67246	8.90974	*NDUFA13, UQCRB, NDUFS6, ATP5A1, UQCRC2, COX7C*
Huntington's disease *Homo sapiens* hsa05016	6/193	0.00119	0.01950	-1.77698	6.99661	*NDUFA13, UQCRB, NDUFS6, ATP5A1, UQCRC2, COX7C*
Non-alcoholic fatty liver disease (NAFLD) *Homo sapiens* hsa04932	5/151	0.00243	0.02971	-1.78236	6.26729	*NDUFA13, UQCRB, NDUFS6, UQCRC2, COX7C*
Cardiac muscle contraction *Homo sapiens* hsa04260	4/78	0.00148	0.02077	-1.60439	6.21545	*UQCRB, UQCRC2, ATP1A1, COX7C*

Staphylococcus aureus infection *Homo sapiens* hsa05150	4/154	0.00541	0.05893	-1.50620	4.26461	*CFD, HLA-DRB1, C1QC*
Phagosome *Homo sapiens* hsa04145	4/154	0.01518	0.14881	-1.49054	2.83964	*MARCO, ATP6V0C, HLA-DRB1, CTSS*
Metabolic pathways *Homo sapiens* hsa01100	13/1239	0.03146	0.28027	-1.76326	2.24288	*ASAH1, NDUFA13, UQCRB, PIGQ, ATP5A1, BCKDHB, CCBL2, PYCR1, PDHB, COX7C, NDUFS6, UQCRC2, ATP6V0C*
Apoptosis *Homo sapiens* hsa04210	3/140	0.05590	0.45571	-1.67852	1.31916	*CAPN2, CTSS, CTSB*
Inflammatory bowel disease (IBD) *Homo sapiens* hsa05321	2/65	0.06257	0.45571	-1.50830	1.18538	*IL21R, HLA-DRB1*
Bile secretion *Homo sapiens* hsa04976	2/71	0.07270	0.47499	-1.29529	0.96430	*ATP1A1, ABCG2*
Tuberculosis *Homo sapiens* hsa05152	3/178	0.09734	0.52833	-1.39014	0.88696	*ATP6V0C, HLA-DRB1, CTSS*
Complement and coagulation cascades *Homo sapiens* hsa04610	2/79	0.08697	0.52833	-1.33501	0.85178	*CFD, C1QC*
Rheumatoid arthritis *Homo sapiens* hsa05323	2/90	0.10782	0.52833	-1.23065	0.78520	*ATP6V0C, HLA-DRB1*
Chemical carcinogenesis *Homo sapiens* hsa05204	2/82	0.09253	0.52833	-1.21160	0.77304	*GSTO2, CCBL2*
Pancreatic secretion *Homo sapiens* hsa04972	2/96	0.11971	0.53051	-1.17661	0.74588	*RAB3D, ATP1A1*
Spliceosome *Homo sapiens* hsa03040	2/134	0.20089	0.53051	-1.05713	0.67014	*SF3B4, PLRG1*
Systemic lupus erythematosus *Homo sapiens* hsa05322	2/135	0.20312	0.53051	-0.97280	0.61668	*HLA-DRB1, C1QC*
Ribosome *Homo sapiens* hsa03010	2/137	0.20759	0.53051	-0.97045	0.61519	*RPS27L, MRPL35*
Cell adhesion molecules (CAMs) *Homo sapiens* hsa04514	2/142	0.21881	0.53051	-0.88323	0.55990	*SDC3, HLA-DRB1*
RNA transport *Homo sapiens* hsa03013	2/172	0.28691	0.53051	-0.60593	0.38411	*SUMO1, PAIP1*
Prion diseases *Homo sapiens* hsa05020	1/35	0.19804	0.53051	-0.17510	0.11100	*C1QC*
Aldosterone-regulated sodium reabsorption *Homo sapiens* hsa04960	1/39	0.21753	0.53051	-0.09396	0.05957	*ATP1A1*
Asthma *Homo sapiens* hsa05310	1/31	0.17809	0.53051	-0.05376	0.03408	*HLA-DRB1*
Proximal tubule bicarbonate reclamation *Homo sapiens* hsa04964	1/23	0.13671	0.53051	-0.02721	0.01725	*ATP1A1*
Propanoate metabolism *Homo sapiens* hsa00640	1/32	0.18312	0.53051	0.00892	-0.00566	*BCKDHB*
Cytokine-cytokine receptor interaction *Homo sapiens* hsa04060	2/265	0.48847	0.59568	0.02594	-0.01344	*IL21R, ACVR1B*
Citrate cycle (TCA cycle) *Homo sapiens* hsa00020	1/30	0.17303	0.53051	0.04938	-0.03130	*PDHB*
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis *Homo sapiens* hsa00563	1/25	0.14724	0.53051	0.05272	-0.03342	*PIGQ*
Other types of O-glycan biosynthesis *Homo sapiens* hsa00514	1/31	0.17809	0.53051	0.06049	-0.03835	*OGT*
Allograft rejection *Homo sapiens* hsa05330	1/38	0.21270	0.53051	0.06117	-0.03878	*HLA-DRB1*
Maturity onset diabetes of the young *Homo sapiens* hsa04950	1/26	0.15246	0.53051	0.07479	-0.04741	*HHEX*
mTOR signaling pathway *Homo sapiens* hsa04150	1/60	0.31246	0.53797	0.08047	-0.04989	*AKT1S1*
Collecting duct acid secretion *Homo sapiens* hsa04966	1/27	0.15765	0.53051	0.13493	-0.08554	*ATP6V0C*
Tryptophan metabolism *Homo sapiens* hsa00380	1/40	0.22232	0.53051	0.13998	-0.08874	*CCBL2*
Longevity regulating pathway - multiple species *Homo sapiens* hsa04213	1/64	0.32922	0.53797	0.14497	-0.08987	*AKT1S1*
Central carbon metabolism in cancer *Homo sapiens* hsa05230	1/67	0.34153	0.53797	0.14544	-0.09017	*PDHB*
Endocrine and other factor-regulated calcium reabsorption *Homo sapiens* hsa04961	1/47	0.25511	0.53051	0.15233	-0.09657	*ATP1A1*
Estrogen signaling pathway *Homo sapiens* hsa04915	1/99	0.45984	0.59090	0.18884	-0.09935	*FKBP5*

Appendix

TGF-beta signaling pathway *Homo sapiens* hsa04350	1/84	0.40722	0.56793	0.18599	-0.10523	*ACVR1B*
HTLV-I infection *Homo sapiens* hsa05166	1/258	0.80084	0.80084	0.48006	-0.10662	*HLA-DRB1*
Glucagon signaling pathway *Homo sapiens* hsa04922	1/101	0.46650	0.59090	0.21242	-0.11175	*PDHB*
Carbohydrate digestion and absorption *Homo sapiens* hsa04973	1/45	0.24589	0.53051	0.17665	-0.11198	*ATP1A1*
Jak-STAT signaling pathway *Homo sapiens* hsa04630	1/158	0.62598	0.68928	0.30639	-0.11401	*IL21R*
Renin secretion *Homo sapiens* hsa04924	1/64	0.32922	0.53797	0.19050	-0.11810	*CTSB*
Protein processing in endoplasmic reticulum *Homo sapiens* hsa04141	1/169	0.65087	0.70094	0.33425	-0.11877	*CAPN2*
HIF-1 signaling pathway *Homo sapiens* hsa04066	1/103	0.47308	0.59090	0.22818	-0.12005	*PDHB*
Type I diabetes mellitus *Homo sapiens* hsa04940	1/43	0.23655	0.53051	0.19371	-0.12279	*HLA-DRB1*
Longevity regulating pathway - mammal *Homo sapiens* hsa04211	1/94	0.44283	0.57863	0.22839	-0.12495	*AKT1S1*
Adrenergic signaling in cardiomyocytes *Homo sapiens* hsa04261	1/148	0.60185	0.67024	0.31422	-0.12573	*ATP1A1*
Vibrio cholerae infection *Homo sapiens* hsa05110	1/51	0.27324	0.53051	0.20007	-0.12683	*ATP6V0C*
Epstein-Barr virus infection *Homo sapiens* hsa05169	1/202	0.71623	0.72362	0.39853	-0.12892	*HLA-DRB1*
Signaling pathways regulating pluripotency of stem cells *Homo sapiens* hsa04550	1/142	0.58664	0.66082	0.31638	-0.13107	*ACVR1B*
cGMP-PKG signaling pathway *Homo sapiens* hsa04022	1/167	0.64647	0.70094	0.37130	-0.13194	*ATP1A1*
Leishmaniasis *Homo sapiens* hsa05140	1/73	0.36549	0.53797	0.21346	-0.13234	*HLA-DRB1*
Influenza A *Homo sapiens* hsa05164	1/175	0.66376	0.70705	0.38657	-0.13401	*HLA-DRB1*
Insulin resistance *Homo sapiens* hsa04931	1/109	0.49235	0.59568	0.25909	-0.13422	*OGT*
Insulin secretion *Homo sapiens* hsa04911	1/85	0.41088	0.56793	0.23836	-0.13485	*ATP1A1*
Pyruvate metabolism *Homo sapiens* hsa00620	1/40	0.22232	0.53051	0.21626	-0.13709	*PDHB*
Valine, leucine and isoleucine degradation *Homo sapiens* hsa00280	1/48	0.25969	0.53051	0.21862	-0.13859	*BCKDHB*
Glycolysis / Gluconeogenesis *Homo sapiens* hsa00010	1/67	0.34153	0.53797	0.22769	-0.14116	*PDHB*
Transcriptional misregulation in cancer *Homo sapiens* hsa05202	1/180	0.67414	0.71039	0.41737	-0.14272	*HHEX*
Graft-versus-host disease *Homo sapiens* hsa05332	1/41	0.22709	0.53051	0.22542	-0.14290	*HLA-DRB1*
Ribosome biogenesis in eukaryotes *Homo sapiens* hsa03008	1/89	0.42529	0.56793	0.25321	-0.14326	*SBDS*
Hematopoietic cell lineage *Homo sapiens* hsa04640	1/88	0.42172	0.56793	0.25958	-0.14686	*HLA-DRB1*
Herpes simplex infection *Homo sapiens* hsa05168	1/185	0.68421	0.71333	0.43961	-0.14851	*HLA-DRB1*
Sphingolipid metabolism *Homo sapiens* hsa00600	1/47	0.25511	0.53051	0.23765	-0.15065	*ASAH1*
Chagas disease (American trypanosomiasis) *Homo sapiens* hsa05142	1/104	0.47634	0.59090	0.28803	-0.15154	*C1QC*
Gastric acid secretion *Homo sapiens* hsa04971	1/74	0.36940	0.53797	0.24872	-0.15419	*ATP1A1*
cAMP signaling pathway *Homo sapiens* hsa04024	1/199	0.71082	0.72362	0.48514	-0.15694	*ATP1A1*
Viral myocarditis *Homo sapiens* hsa05416	1/59	0.30821	0.53797	0.25318	-0.15696	*HLA-DRB1*
Focal adhesion *Homo sapiens* hsa04510	1/202	0.71623	0.72362	0.49196	-0.15915	*CAPN2*
Salivary secretion *Homo sapiens* hsa04970	1/89	0.42529	0.56793	0.28898	-0.16349	*ATP1A1*

Arginine and proline metabolism *Homo sapiens* hsa00330	1/50	0.26875	0.53051	0.26149	-0.16577	*PYCR1*
Epithelial cell signaling in Helicobacter pylori infection *Homo sapiens* hsa05120	1/68	0.34558	0.53797	0.26754	-0.16586	*ATP6V0C*
Thyroid hormone signaling pathway *Homo sapiens* hsa04919	1/118	0.51997	0.60634	0.33882	-0.16952	*ATP1A1*
Thyroid hormone synthesis *Homo sapiens* hsa04918	1/71	0.35760	0.53797	0.29291	-0.18159	*ATP1A1*
AMPK signaling pathway *Homo sapiens* hsa04152	1/124	0.53756	0.61257	0.38308	-0.18774	*AKT1S1*
Sphingolipid signaling pathway *Homo sapiens* hsa04071	1/120	0.52590	0.60634	0.38027	-0.19026	*ASAH1*
Toxoplasmosis *Homo sapiens* hsa05145	1/118	0.51997	0.60634	0.38036	-0.19030	*HLA-DRB1*
Carbon metabolism *Homo sapiens* hsa01200	1/113	0.50481	0.60331	0.38345	-0.19377	*PDHB*
Drug metabolism - cytochrome P450 *Homo sapiens* hsa00982	1/69	0.34961	0.53797	0.31631	-0.19610	*GSTO2*
Biosynthesis of amino acids *Homo sapiens* hsa01230	1/74	0.36940	0.53797	0.32133	-0.19921	*PYCR1*
ABC transporters *Homo sapiens* hsa02010	1/44	0.24123	0.53051	0.31745	-0.20124	*ABCG2*
Metabolism of xenobiotics by cytochrome P450 *Homo sapiens* hsa00980	1/73	0.36549	0.53797	0.32493	-0.20144	*GSTO2*
Proteasome *Homo sapiens* hsa03050	1/44	0.24123	0.53051	0.33136	-0.21006	*PSME1*
Intestinal immune network for IgA production *Homo sapiens* hsa04672	1/48	0.25969	0.53051	0.33896	-0.21488	*HLA-DRB1*
Selenocompound metabolism *Homo sapiens* hsa00450	1/17	0.10435	0.52833	0.35849	-0.22873	*CCBL2*
Synaptic vesicle cycle *Homo sapiens* hsa04721	1/63	0.32507	0.53797	0.39277	-0.24350	*ATP6V0C*
Autoimmune thyroid disease *Homo sapiens* hsa05320	1/53	0.28214	0.53051	0.38866	-0.24638	*HLA-DRB1*
Pertussis *Homo sapiens* hsa05133	1/75	0.37329	0.53797	0.41958	-0.26012	*C1QC*
Protein digestion and absorption *Homo sapiens* hsa04974	1/90	0.42884	0.56793	0.46292	-0.26190	*ATP1A1*
Mineral absorption *Homo sapiens* hsa04978	1/51	0.27324	0.53051	0.42188	-0.26744	*ATP1A1*
Glutathione metabolism *Homo sapiens* hsa00480	1/52	0.27770	0.53051	0.42499	-0.26941	*GSTO2*
Sulfur metabolism *Homo sapiens* hsa00920	1/10	0.06510	0.45571	0.77183	-0.60659	*SQRDL*

Table A19: GO terms from EnrichR KEGG for genes upregulated in mutant progenitors.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Ribosome *Homo sapiens* hsa03010	9/137	0.00000	0.00039	-1.74614	13.71052	*RPL5, RPS9, RPL3, RPS8, RPSA, RPS3A, RPL18, RPL28, RPL7*
Cell cycle *Homo sapiens* hsa04110	7/124	0.00012	0.00636	-1.70792	8.63898	*CCNA2, CCNB1, MCM7, PLK1, MCM6, TP53, MAD2L1*
RNA transport *Homo sapiens* hsa03013	7/172	0.00082	0.02854	-1.81865	6.46798	*EIF5B, PRMT5, NUP188, EIF3I, RANGAP1, RAN, THOC7*
Progesterone-mediated oocyte maturation *Homo sapiens* hsa04914	5/98	0.00185	0.04867	-1.85638	5.61125	*CCNA2, CCNB1, HSP90AB1, PLK1, MAD2L1*
Oocyte meiosis *Homo sapiens* hsa04114	5/123	0.00476	0.08323	-1.75128	4.35402	*CCNB1, PPP2R1B, PPP2R5E, PLK1, MAD2L1*
DNA replication *Homo sapiens* hsa03030	3/36	0.00448	0.08323	-1.73761	4.32002	*RFC5, MCM7, MCM6*
HTLV-I infection *Homo sapiens* hsa05166	6/258	0.02552	0.25810	-1.71818	2.32711	*ELK4, TRRAP, FOS, TP53, RAN, MAD2L1*
Protein processing in endoplasmic reticulum *Homo sapiens* hsa04141	5/169	0.01669	0.25041	-1.67106	2.31386	*PDIA3, EDEM3, MBTPS1, HSP90AB1, RPN1*

Epstein-Barr virus infection *Homo sapiens* hsa05169	5/202	0.03240	0.25810	-1.65163	2.23698	*CCNA2, POLR2D, POLR3H, TP53, RAN*
Herpes simplex infection *Homo sapiens* hsa05168	5/185	0.02347	0.25810	-1.61049	2.18125	*SRSF3, FOS, SRSF6, TP53, IKBKE*
Spliceosome *Homo sapiens* hsa03040	4/134	0.03077	0.25810	-1.55012	2.09949	*LSM7, LSM6, SRSF3, SRSF6*
RNA degradation *Homo sapiens* hsa03018	3/77	0.03139	0.25810	-1.51730	2.05504	*LSM7, LSM6, C1D*
p53 signaling pathway *Homo sapiens* hsa04115	3/69	0.02391	0.25810	-1.49489	2.02468	*CCNB1, RRM2, TP53*
Hepatitis B *Homo sapiens* hsa05161	4/146	0.03997	0.27979	-1.57952	2.01189	*CCNA2, FOS, TP53, IKBKE*
Estrogen signaling pathway *Homo sapiens* hsa04915	3/99	0.05750	0.37734	-1.50163	1.46352	*HSP90AB1, FOS, FKBP4*
RNA polymerase *Homo sapiens* hsa03020	2/32	0.03441	0.25810	-1.01856	1.37954	*POLR2D, POLR3H*
Purine metabolism *Homo sapiens* hsa00230	4/176	0.06914	0.40331	-1.37291	1.24666	*RRM2, POLR2D, AK2, POLR3H*
Pyrimidine metabolism *Homo sapiens* hsa00240	3/105	0.06597	0.40331	-1.35267	1.22828	*RRM2, POLR2D, POLR3H*
Dopaminergic synapse *Homo sapiens* hsa04728	3/129	0.10501	0.49112	-1.36051	0.96742	*PPP2R1B, PPP2R5E, FOS*
Sphingolipid signaling pathway *Homo sapiens* hsa04071	3/120	0.08945	0.49112	-1.31307	0.93369	*PPP2R1B, PPP2R5E, TP53*
AMPK signaling pathway *Homo sapiens* hsa04152	3/124	0.09624	0.49112	-1.30601	0.92866	*CCNA2, PPP2R1B, PPP2R5E*
Hepatitis C *Homo sapiens* hsa05160	3/133	0.11226	0.49112	-1.24765	0.88717	*PPP2R1B, TP53, IKBKE*
Apoptosis *Homo sapiens* hsa04210	3/140	0.12538	0.52329	-1.34438	0.87065	*CTSZ, FOS, TP53*
Colorectal cancer *Homo sapiens* hsa05210	2/62	0.10511	0.49112	-1.17770	0.83743	*FOS, TP53*
Cytosolic DNA-sensing pathway *Homo sapiens* hsa04623	2/64	0.11065	0.49112	-0.92712	0.65925	*POLR3H, IKBKE*
Antigen processing and presentation *Homo sapiens* hsa04612	2/77	0.14828	0.57666	-0.94268	0.51894	*PDIA3, HSP90AB1*
Ribosome biogenesis in eukaryotes *Homo sapiens* hsa03008	2/89	0.18497	0.64768	-1.13943	0.49493	*RAN, NOL6*
Prostate cancer *Homo sapiens* hsa05215	2/89	0.18497	0.64768	-1.13277	0.49203	*HSP90AB1, TP53*
Viral carcinogenesis *Homo sapiens* hsa05203	3/205	0.26674	0.67409	-1.04113	0.41061	*CCNA2, TP53, HDAC7*
Huntington's disease *Homo sapiens* hsa05016	3/193	0.23880	0.67409	-1.03310	0.40744	*DNAH8, POLR2D, TP53*
mRNA surveillance pathway *Homo sapiens* hsa03015	2/91	0.19122	0.64768	-0.88977	0.38648	*PPP2R1B, PPP2R5E*
Toll-like receptor signaling pathway *Homo sapiens* hsa04620	2/106	0.23879	0.67409	-0.97802	0.38572	*FOS, IKBKE*
Chagas disease (American trypanosomiasis) *Homo sapiens* hsa05142	2/104	0.23240	0.67409	-0.93583	0.36908	*PPP2R1B, FOS*
PI3K-Akt signaling pathway *Homo sapiens* hsa04151	4/341	0.34719	0.67409	-0.90034	0.35508	*HSP90AB1, PPP2R1B, PPP2R5E, TP53*
Measles *Homo sapiens* hsa05162	2/136	0.33488	0.67409	-0.79864	0.31497	*TP53, IKBKE*
FoxO signaling pathway *Homo sapiens* hsa04068	2/133	0.32535	0.67409	-0.78427	0.30931	*CCNB1, PLK1*
Wnt signaling pathway *Homo sapiens* hsa04310	2/142	0.35382	0.67409	-0.68597	0.27054	*INVS, TP53*
MAPK signaling pathway *Homo sapiens* hsa04010	3/255	0.38550	0.67409	-0.67388	0.26577	*ELK4, FOS, TP53*
Ubiquitin mediated proteolysis *Homo sapiens* hsa04120	2/137	0.33805	0.67409	-0.62240	0.24547	*VHL, UBE2M*
Adrenergic signaling in cardiomyocytes *Homo sapiens* hsa04261	2/148	0.37259	0.67409	-0.58513	0.23077	*PPP2R1B, PPP2R5E*
Pathways in cancer *Homo sapiens* hsa05200	4/397	0.45785	0.67409	-0.40140	0.15831	*HSP90AB1, FOS, VHL, TP53*
Transcriptional misregulation in cancer *Homo sapiens* hsa05202	2/180	0.46856	0.67409	-0.14796	0.05836	*ELK4, TP53*
Bladder cancer *Homo sapiens* hsa05219	1/41	0.30578	0.67409	0.03460	-0.01365	*TP53*

Metabolic pathways *Homo sapiens* hsa01100	6/1239	0.96953	0.96953	0.57103	-0.01767	*RRM2, RPN1, POLR2D, AK2, B3GNT2, POLR3H*
Glioma *Homo sapiens* hsa05214	1/65	0.43701	0.67409	0.08477	-0.03343	*TP53*
MicroRNAs in cancer *Homo sapiens* hsa05206	1/297	0.92849	0.93742	0.57964	-0.03746	*TP53*
Endometrial cancer *Homo sapiens* hsa05213	1/52	0.36929	0.67409	0.09743	-0.03842	*TP53*
Thyroid cancer *Homo sapiens* hsa05216	1/297	0.22931	0.67409	0.11652	-0.04596	*TP53*
Endocytosis *Homo sapiens* hsa04144	1/259	0.89926	0.91672	0.54147	-0.04708	*GRK6*
Renal cell carcinoma *Homo sapiens* hsa05211	1/66	0.44191	0.67409	0.19313	-0.07617	*VHL*
Long-term depression *Homo sapiens* hsa04730	1/60	0.41186	0.67409	0.19736	-0.07784	*PPP2R1B*
Pancreatic cancer *Homo sapiens* hsa05212	1/66	0.44191	0.67409	0.20606	-0.08127	*TP53*
Non-small cell lung cancer *Homo sapiens* hsa05223	1/56	0.39094	0.67409	0.20625	-0.08134	*TP53*
Influenza A *Homo sapiens* hsa05164	1/175	0.78651	0.85068	0.51601	-0.08345	*IKBKE*
Central carbon metabolism in cancer *Homo sapiens* hsa05230	1/67	0.44677	0.67409	0.21445	-0.08458	*TP53*
Alcoholism *Homo sapiens* hsa05034	1/179	0.79397	0.85068	0.52867	-0.08550	*HDAC7*
Focal adhesion *Homo sapiens* hsa04510	1/202	0.83213	0.85814	0.59254	-0.09065	*PARVB*
Proteoglycans in cancer *Homo sapiens* hsa05205	1/203	0.83362	0.85814	0.60086	-0.09193	*TP53*
cAMP signaling pathway *Homo sapiens* hsa04024	1/199	0.82758	0.85814	0.61437	-0.09399	*FOS*
Chemokine signaling pathway *Homo sapiens* hsa04062	1/187	0.80812	0.85710	0.61379	-0.09465	*GRK6*
Oxytocin signaling pathway *Homo sapiens* hsa04921	1/158	0.75175	0.82222	0.51185	-0.10019	*FOS*
Hippo signaling pathway *Homo sapiens* hsa04390	1/153	0.74050	0.81845	0.52467	-0.10511	*PPP2R1B*
Amphetamine addiction *Homo sapiens* hsa05031	1/67	0.44677	0.67409	0.27555	-0.10867	*FOS*
Nucleotide excision repair *Homo sapiens* hsa03420	1/47	0.34116	0.67409	0.27993	-0.11040	*RFC5*
Amyotrophic lateral sclerosis (ALS) *Homo sapiens* hsa05014	1/51	0.36376	0.67409	0.28710	-0.11323	*TP53*
Aminoacyl-tRNA biosynthesis *Homo sapiens* hsa00970	1/66	0.44191	0.67409	0.29128	-0.11488	*NARS*
Porphyrin and chlorophyll metabolism *Homo sapiens* hsa00860	1/42	0.31180	0.67409	0.29680	-0.11705	*HCCS*
Tight junction *Homo sapiens* hsa04530	1/139	0.70628	0.78893	0.51719	-0.12262	*PPP2R1B*
Melanoma *Homo sapiens* hsa05218	1/71	0.46580	0.67409	0.32160	-0.12683	*TP53*
Choline metabolism in cancer *Homo sapiens* hsa05231	1/101	0.58941	0.73267	0.40942	-0.12735	*FOS*
Insulin resistance *Homo sapiens* hsa04931	1/109	0.61731	0.73616	0.41813	-0.12808	*SLC27A1*
HIF-1 signaling pathway *Homo sapiens* hsa04066	1/103	0.59656	0.73267	0.41650	-0.12955	*VHL*
Proteasome *Homo sapiens* hsa03050	1/44	0.32370	0.67409	0.33136	-0.13068	*PSMA5*
Osteoclast differentiation *Homo sapiens* hsa04380	1/132	0.68754	0.77626	0.51700	-0.13094	*FOS*
TNF signaling pathway *Homo sapiens* hsa04668	1/110	0.62066	0.73616	0.42776	-0.13103	*FOS*
Cholinergic synapse *Homo sapiens* hsa04725	1/111	0.62398	0.73616	0.45161	-0.13833	*FOS*
N-Glycan biosynthesis *Homo sapiens* hsa00510	1/49	0.35256	0.67409	0.35326	-0.13932	*RPN1*
Chronic myeloid leukemia *Homo sapiens* hsa05220	1/73	0.47507	0.67409	0.35467	-0.13988	*TP53*
Prolactin signaling pathway *Homo sapiens* hsa04917	1/72	0.47046	0.67409	0.36078	-0.14229	*FOS*

Neurotrophin signaling pathway *Homo sapiens* hsa04722	1/120	0.65264	0.75305	0.50432	-0.14304	*TP53*
Longevity regulating pathway - mammal *Homo sapiens* hsa04211	1/94	0.56336	0.71751	0.43191	-0.14338	*TP53*
Lysosome *Homo sapiens* hsa04142	1/123	0.66171	0.75522	0.51147	-0.14360	*CTSZ*
T cell receptor signaling pathway *Homo sapiens* hsa04660	1/104	0.60010	0.73267	0.47058	-0.14637	*FOS*
NOD-like receptor signaling pathway *Homo sapiens* hsa04621	1/57	0.39623	0.67409	0.37731	-0.14881	*HSP90AB1*
TGF-beta signaling pathway *Homo sapiens* hsa04350	1/84	0.52331	0.70655	0.42840	-0.14881	*PPP2R1B*
Thyroid hormone signaling pathway *Homo sapiens* hsa04919	1/118	0.64647	0.75305	0.52666	-0.14937	*TP53*
Morphine addiction *Homo sapiens* hsa05032	1/91	0.55171	0.71517	0.45817	-0.15359	*GRK6*
Circadian entrainment *Homo sapiens* hsa04713	1/95	0.56718	0.71751	0.46425	-0.15411	*FOS*
Basal cell carcinoma *Homo sapiens* hsa05217	1/55	0.38560	0.67409	0.39296	-0.15498	*TP53*
Rheumatoid arthritis *Homo sapiens* hsa05323	1/90	0.54775	0.71517	0.48749	-0.16342	*FOS*
PPAR signaling pathway *Homo sapiens* hsa03320	1/69	0.45637	0.67409	0.41849	-0.16505	*SLC27A1*
Glutathione metabolism *Homo sapiens* hsa00480	1/52	0.36929	0.67409	0.42499	-0.16761	*RRM2*
B cell receptor signaling pathway *Homo sapiens* hsa04662	1/73	0.47507	0.67409	0.42701	-0.16841	*FOS*
Small cell lung cancer *Homo sapiens* hsa05222	1/86	0.53160	0.70655	0.51310	-0.17823	*TP53*
Bile secretion *Homo sapiens* hsa04976	1/71	0.46580	0.67409	0.45616	-0.17990	*EPHX1*
RIG-I-like receptor signaling pathway *Homo sapiens* hsa04622	1/70	0.46111	0.67409	0.46776	-0.18448	*IKBKE*
Leishmaniasis *Homo sapiens* hsa05140	1/73	0.47507	0.67409	0.47892	-0.18888	*FOS*
Chemical carcinogenesis *Homo sapiens* hsa05204	1/82	0.51488	0.70655	0.56228	-0.19531	*EPHX1*
Salmonella infection *Homo sapiens* hsa05132	1/86	0.53160	0.70655	0.59751	-0.20755	*FOS*
Base excision repair *Homo sapiens* hsa03410	1/33	0.25567	0.67409	0.61569	-0.24282	*APEX1*
Metabolism of xenobiotics by cytochrome P450 *Homo sapiens* hsa00980	1/73	0.47507	0.67409	0.62816	-0.24774	*EPHX1*
Pertussis *Homo sapiens* hsa05133	1/75	0.48419	0.67786	0.64492	-0.25075	*FOS*
Glycosphingolipid biosynthesis - lacto and neolacto series *Homo sapiens* hsa00601	1/26	0.20893	0.67409	0.67754	-0.26722	*B3GNT2*
Mismatch repair *Homo sapiens* hsa03430	1/23	0.18802	0.64768	0.85017	-0.36928	*RFC5*
Glycosaminoglycan biosynthesis - keratan sulfate *Homo sapiens* hsa00533	1/15	0.12958	0.52329	1.44988	-0.93897	*B3GNT2*

Table A20: GO terms from EnrichR KEGG for genes downregulated in mutant progenitors.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
HTLV-I infection *Homo sapiens* hsa05166	8/258	0.00095	0.16325	-1.94300	3.52169	*KAT2B, NRP1, HLA-DMA, RANBP3, RRAS2, HRAS, HLA-DRB1, PIK3R5*
Thyroid hormone signaling pathway *Homo sapiens* hsa04919	5/118	0.00256	0.16325	-1.83479	3.32556	*KAT2B, PLCB3, ATP1A1, HRAS, PIK3R5*
Leishmaniasis *Homo sapiens* hsa05140	4/73	0.00292	0.16325	-1.81063	3.28176	*C3, HLA-DMA, NCF4, HLA-DRB1*
Bacterial invasion of epithelial cells *Homo sapiens* hsa05100	4/78	0.00367	0.16325	-1.69894	3.07933	*ARHGAP10, DNM1, PIK3R5, CD2AP*

Phospholipase D signaling pathway *Homo sapiens* hsa04072	5/144	0.00581	0.20048	-1.79478	2.88431	*PLCB3, RRAS2, HRAS, DNM1, PIK3R5*
Endocrine and other factor-regulated calcium reabsorption *Homo sapiens* hsa04961	3/47	0.00676	0.20048	-1.76255	2.83252	*PLCB3, ATP1A1, DNM1*
Chemokine signaling pathway *Homo sapiens* hsa04062	5/187	0.01631	0.30527	-1.74660	2.07246	*ITK, PLCB3, CXCL14, HRAS, PIK3R5*
Staphylococcus aureus infection *Homo sapiens* hsa05150	3/56	0.01067	0.27134	-1.58716	2.07024	*C3, HLA-DMA, HLA-DRB1*
Viral carcinogenesis *Homo sapiens* hsa05203	5/205	0.02309	0.30527	-1.70252	2.02016	*C3, KAT2B, HRAS, GTF2E2, PIK3R5*
Melanoma *Homo sapiens* hsa05218	3/71	0.01961	0.30527	-1.61751	1.91928	*FGF18, HRAS, PIK3R5*
AMPK signaling pathway *Homo sapiens* hsa04152	4/124	0.01721	0.30527	-1.58752	1.88370	*GYS1, HNF4A, FBP1, PIK3R5*
Regulation of actin cytoskeleton *Homo sapiens* hsa04810	5/214	0.02708	0.30527	-1.57466	1.86844	*FGF18, RRAS2, F2, HRAS, PIK3R5*
Inflammatory mediator regulation of TRP channels *Homo sapiens* hsa04750	3/98	0.04357	0.30527	-1.48801	1.76562	*CYP2J2, PLCB3, PIK3R5*
Insulin signaling pathway *Homo sapiens* hsa04910	4/139	0.02474	0.30527	-1.47736	1.75298	*GYS1, FBP1, HRAS, PIK3R5*
Bile secretion *Homo sapiens* hsa04976	3/71	0.01961	0.30527	-1.46370	1.73678	*ATP1A1, SLC22A8, SLC4A5*
AGE-RAGE signaling pathway in diabetic complications *Homo sapiens* hsa04933	3/101	0.04685	0.30527	-1.46142	1.73407	*PLCB3, HRAS, PIK3R5*
Estrogen signaling pathway *Homo sapiens* hsa04915	3/99	0.04465	0.30527	-1.36076	1.61463	*PLCB3, HRAS, PIK3R5*
Phagosome *Homo sapiens* hsa04145	4/154	0.03401	0.30527	-1.34650	1.59772	*C3, HLA-DMA, NCF4, HLA-DRB1*
Influenza A *Homo sapiens* hsa05164	4/175	0.04999	0.30527	-1.32208	1.56873	*HLA-DMA, TMPRSS2, HLA-DRB1, PIK3R5*
Choline metabolism in cancer *Homo sapiens* hsa05231	3/101	0.04685	0.30527	-1.32067	1.56706	*CHKA, HRAS, PIK3R5*
Glucagon signaling pathway *Homo sapiens* hsa04922	3/101	0.04685	0.30527	-1.27216	1.50951	*GYS1, PLCB3, FBP1*
T cell receptor signaling pathway *Homo sapiens* hsa04660	3/104	0.05024	0.30527	-1.16170	1.37844	*ITK, HRAS, PIK3R5*
Cholinergic synapse *Homo sapiens* hsa04725	3/111	0.05863	0.30527	-1.14737	1.36143	*PLCB3, HRAS, PIK3R5*
Chagas disease (American trypanosomiasis) *Homo sapiens* hsa05142	3/104	0.05024	0.30527	-1.09901	1.30405	*C3, PLCB3, PIK3R5*
Insulin resistance *Homo sapiens* hsa04931	3/109	0.05617	0.30527	-1.06628	1.26521	*GYS1, SLC27A2, PIK3R5*
Aldosterone-regulated sodium reabsorption *Homo sapiens* hsa04960	3/39	0.04021	0.30527	-0.98720	1.17137	*ATP1A1, PIK3R5*
Serotonergic synapse *Homo sapiens* hsa04726	3/112	0.05988	0.30527	-0.97446	1.15627	*CYP2J2, PLCB3, HRAS*
Allograft rejection *Homo sapiens* hsa05330	3/38	0.03844	0.30527	-0.93648	1.11120	*HLA-DMA, HLA-DRB1*
Toxoplasmosis *Homo sapiens* hsa05145	3/118	0.06763	0.30527	-0.92837	1.10158	*HLA-DMA, HLA-DRB1, PIK3R5*
Asthma *Homo sapiens* hsa05310	3/31	0.02693	0.30527	-0.83948	0.99609	*HLA-DMA, HLA-DRB1*
Carbon metabolism *Homo sapiens* hsa01200	3/113	0.06114	0.30527	-0.83257	0.98790	*AGXT, FBP1, GAPDH*
Graft-versus-host disease *Homo sapiens* hsa05332	2/41	0.04385	0.30527	-0.82831	0.98285	*HLA-DMA, HLA-DRB1*
Neurotrophin signaling pathway *Homo sapiens* hsa04722	3/120	0.07031	0.30527	-0.82308	0.97663	*SORT1, HRAS, PIK3R5*
Rap1 signaling pathway *Homo sapiens* hsa04015	4/211	0.08550	0.31794	-0.83062	0.95180	*PLCB3, FGF18, HRAS, PIK3R5*
Leukocyte transendothelial migration *Homo sapiens* hsa04670	3/118	0.06763	0.30527	-0.77757	0.92264	*ITK, NCF4, PIK3R5*
Sphingolipid signaling pathway *Homo sapiens* hsa04071	3/120	0.07031	0.30527	-0.77428	0.91874	*PLCB3, HRAS, PIK3R5*
Type I diabetes mellitus *Homo sapiens* hsa04940	2/43	0.04760	0.30527	-0.61287	0.72721	*HLA-DMA, HLA-DRB1*
Apoptosis *Homo sapiens* hsa04210	3/140	0.09968	0.31794	-0.59898	0.68637	*ENDOG, HRAS, PIK3R5*

Appendix

Systemic lupus erythematosus *Homo sapiens* hsa05322	3/135	0.09192	0.31794	-0.57313	0.65674	*C3, HLA-DMA, HLA-DRB1*
Ras signaling pathway *Homo sapiens* hsa04014	4/227	0.10443	0.31794	-0.53425	0.61218	*FGF18, RRAS2, HRAS, PIK3R5*
Endometrial cancer *Homo sapiens* hsa05213	2/52	0.06585	0.30527	-0.50336	0.59728	*HRAS, PIK3R5*
Basal transcription factors *Homo sapiens* hsa03022	2/45	0.05147	0.30527	-0.48974	0.58112	*TAF11, GTF2E2*
Carbohydrate digestion and absorption *Homo sapiens* hsa04973	2/45	0.05147	0.30527	-0.45635	0.54149	*ATP1A1, PIK3R5*
Intestinal immune network for IgA production *Homo sapiens* hsa04672	2/48	0.05748	0.30527	-0.38851	0.46100	*HLA-DMA, HLA-DRB1*
Adrenergic signaling in cardiomyocytes *Homo sapiens* hsa04261	3/148	0.11262	0.31794	-0.38528	0.44148	*PLCB3, ATP1A1, PIK3R5*
Non-small cell lung cancer *Homo sapiens* hsa05223	2/56	0.07459	0.31614	-0.35276	0.40622	*HRAS, PIK3R5*
Acute myeloid leukemia *Homo sapiens* hsa05221	2/57	0.07684	0.31794	-0.31015	0.35540	*HRAS, PIK3R5*
Long-term depression *Homo sapiens* hsa04730	2/60	0.08369	0.31794	-0.26378	0.30226	*PLCB3, HRAS*
VEGF signaling pathway *Homo sapiens* hsa04370	2/61	0.08602	0.31794	-0.25133	0.28799	*HRAS, PIK3R5*
Autoimmune thyroid disease *Homo sapiens* hsa05320	2/53	0.06800	0.30527	-0.19763	0.23450	*HLA-DMA, HLA-DRB1*
Longevity regulating pathway - multiple species *Homo sapiens* hsa04213	2/64	0.09311	0.31794	-0.19655	0.22522	*HRAS, PIK3R5*
Glioma *Homo sapiens* hsa05214	2/65	0.09551	0.31794	-0.18838	0.21587	*HRAS, PIK3R5*
Oxytocin signaling pathway *Homo sapiens* hsa04921	3/158	0.12965	0.32937	-0.19379	0.21522	*PLCB3, HRAS, PIK3R5*
Viral myocarditis *Homo sapiens* hsa05416	2/59	0.08139	0.31794	-0.15911	0.18233	*HLA-DMA, HLA-DRB1*
Long-term potentiation *Homo sapiens* hsa04720	2/66	0.09793	0.31794	-0.08323	0.09538	*PLCB3, HRAS*
Inflammatory bowel disease (IBD) *Homo sapiens* hsa05321	2/65	0.09551	0.31794	-0.06893	0.07898	*HLA-DMA, HLA-DRB1*
Fc epsilon RI signaling pathway *Homo sapiens* hsa04664	2/68	0.10283	0.31794	-0.06652	0.07623	*HRAS, PIK3R5*
cGMP-PKG signaling pathway *Homo sapiens* hsa04022	3/167	0.14573	0.34671	-0.07068	0.07487	*PLCB3, ATP1A1, PIK3R5*
Renal cell carcinoma *Homo sapiens* hsa05211	2/66	0.09793	0.31794	-0.05751	0.06590	*HRAS, PIK3R5*
Central carbon metabolism in cancer *Homo sapiens* hsa05230	2/67	0.10037	0.31794	0.00743	-0.00851	*HRAS, PIK3R5*
Alzheimer's disease *Homo sapiens* hsa05010	3/168	0.14755	0.34671	0.01041	-0.01103	*PLCB3, COX6C, GAPDH*
Metabolic pathways *Homo sapiens* hsa01100	13/1239	0.16361	0.34671	0.04954	-0.05248	*CYP2J2, AHCYL1, RRM2, ACY1, CHKA, UAP1, COX6C, PLCB3, KYNU, HPRT1, AGXT, GAPDH, FBP1*
Prolactin signaling pathway *Homo sapiens* hsa04917	2/72	0.11281	0.31794	0.05596	-0.06412	*HRAS, PIK3R5*
Chronic myeloid leukemia *Homo sapiens* hsa05220	2/73	0.11534	0.31794	0.05994	-0.06869	*HRAS, PIK3R5*
Olfactory transduction *Homo sapiens* hsa04740	1/415	0.96495	0.96495	2.24601	-0.08014	*NCALD*
Glycolysis / Gluconeogenesis *Homo sapiens* hsa00010	2/67	0.10037	0.31794	0.07128	-0.08168	*FBP1, GAPDH*
B cell receptor signaling pathway *Homo sapiens* hsa04662	2/73	0.11534	0.31794	0.09026	-0.10343	*HRAS, PIK3R5*
Thyroid hormone synthesis *Homo sapiens* hsa04918	2/71	0.11029	0.31794	0.15221	-0.17442	*PLCB3, ATP1A1*
Gastric acid secretion *Homo sapiens* hsa04971	2/74	0.11789	0.31794	0.21430	-0.24556	*PLCB3, ATP1A1*
Tuberculosis *Homo sapiens* hsa05152	3/178	0.16622	0.34808	0.24203	-0.25542	*C3, HLA-DMA, HLA-DRB1*
Cytokine-cytokine receptor interaction *Homo sapiens* hsa04060	1/265	0.87985	0.88482	2.36269	-0.28912	*CXCL14*
Biosynthesis of amino acids *Homo sapiens* hsa01230	2/74	0.11789	0.31794	0.25693	-0.29442	*ACY1, GAPDH*

Adherens junction *Homo sapiens* hsa04520	2/74	0.11789	0.31794	0.28124	-0.32227	*LMO7, PTPRJ*
Herpes simplex infection *Homo sapiens* hsa05168	3/185	0.17969	0.35937	0.32042	-0.32792	*C3, HLA-DMA, HLA-DRB1*
ErbB signaling pathway *Homo sapiens* hsa04012	2/87	0.15220	0.34671	0.35719	-0.37837	*HRAS, PIK3R5*
Insulin secretion *Homo sapiens* hsa04911	2/85	0.14680	0.34671	0.36263	-0.38412	*PLCB3, ATP1A1*
Cardiac muscle contraction *Homo sapiens* hsa04260	2/78	0.12822	0.32937	0.36268	-0.40278	*ATP1A1, COX6C*
Antigen processing and presentation *Homo sapiens* hsa04612	2/77	0.12562	0.32937	0.36635	-0.40686	*HLA-DMA, HLA-DRB1*
Huntington's disease *Homo sapiens* hsa05016	3/193	0.19544	0.36505	0.40507	-0.40820	*PLCB3, COX6C, RCOR1*
Prostate cancer *Homo sapiens* hsa05215	2/89	0.15765	0.34671	0.39400	-0.41736	*HRAS, PIK3R5*
Complement and coagulation cascades *Homo sapiens* hsa04610	2/79	0.13084	0.32937	0.42586	-0.47295	*C3, F2*
Peroxisome *Homo sapiens* hsa04146	2/83	0.14143	0.34671	0.45272	-0.47956	*AGXT, SLC27A2*
GnRH signaling pathway *Homo sapiens* hsa04912	2/91	0.16313	0.34671	0.47292	-0.50095	*PLCB3, HRAS*
Gap junction *Homo sapiens* hsa04540	2/88	0.15492	0.34671	0.48487	-0.51361	*PLCB3, HRAS*
Salivary secretion *Homo sapiens* hsa04970	2/89	0.15765	0.34671	0.49735	-0.52683	*PLCB3, ATP1A1*
Epstein-Barr virus infection *Homo sapiens* hsa05169	3/202	0.21357	0.36902	0.54673	-0.54505	*HLA-DRB1, GTF2E2, PIK3R5*
cAMP signaling pathway *Homo sapiens* hsa04024	3/199	0.20748	0.36901	0.56268	-0.56094	*RRAS2, ATP1A1, PIK3R5*
Rheumatoid arthritis *Homo sapiens* hsa05323	2/90	0.16039	0.34671	0.54476	-0.57705	*HLA-DMA, HLA-DRB1*
Longevity regulating pathway - mammal *Homo sapiens* hsa04211	2/94	0.17142	0.35073	0.57728	-0.60484	*HRAS, PIK3R5*
Transcriptional misregulation in cancer *Homo sapiens* hsa05202	1/180	0.76140	0.77005	2.37248	-0.61991	*TMPRSS2*
Calcium signaling pathway *Homo sapiens* hsa04020	1/180	0.76140	0.77005	2.42818	-0.63447	*PLCB3*
Alcoholism *Homo sapiens* hsa05034	1/179	0.75948	0.77005	2.43387	-0.63595	*HRAS*
Melanogenesis *Homo sapiens* hsa04916	2/100	0.18821	0.36415	0.63149	-0.63792	*PLCB3, HRAS*
Proteoglycans in cancer *Homo sapiens* hsa05205	3/203	0.21560	0.36902	0.65263	-0.65061	*RRAS2, HRAS, PIK3R5*
Amoebiasis *Homo sapiens* hsa05146	2/100	0.18821	0.36415	0.65727	-0.66397	*PLCB3, PIK3R5*
Phosphatidylinositol signaling system *Homo sapiens* hsa04070	2/98	0.18259	0.36112	0.67201	-0.68448	*PLCB3, PIK3R5*
Pancreatic secretion *Homo sapiens* hsa04972	2/96	0.17699	0.35801	0.67066	-0.68890	*PLCB3, ATP1A1*
HIF-1 signaling pathway *Homo sapiens* hsa04066	2/103	0.19669	0.36505	0.68553	-0.69084	*GAPDH, PIK3R5*
Protein digestion and absorption *Homo sapiens* hsa04974	2/90	0.16039	0.34671	0.67222	-0.71207	*SLC3A1, ATP1A1*
PI3K-Akt signaling pathway *Homo sapiens* hsa04151	4/341	0.28065	0.43084	0.88860	-0.74822	*GYS1, FGF18, HRAS, PIK3R5*
MicroRNAs in cancer *Homo sapiens* hsa05206	2/297	0.68314	0.70697	2.22240	-0.77065	*HRAS, EZH2*
Hippo signaling pathway *Homo sapiens* hsa04390	1/153	0.70383	0.72418	2.40093	-0.77483	*TEAD1*
Platelet activation *Homo sapiens* hsa04611	2/122	0.25129	0.41416	0.90725	-0.79976	*PLCB3, PIK3R5*
Neuroactive ligand-receptor interaction *Homo sapiens* hsa04080	2/277	0.64550	0.68392	2.14110	-0.81343	*GPR35, F2*
Axon guidance *Homo sapiens* hsa04360	2/127	0.26579	0.43084	0.99837	-0.84064	*NRP1, HRAS*
Wnt signaling pathway *Homo sapiens* hsa04310	1/142	0.67665	0.70435	2.43754	-0.85431	*PLCB3*
Osteoclast differentiation *Homo sapiens* hsa04380	2/132	0.28029	0.43084	1.03464	-0.87118	*NCF4, PIK3R5*
Measles *Homo sapiens* hsa05162	1/136	0.66080	0.69190	2.40491	-0.88576	*PIK3R5*

Appendix

Tight junction *Homo sapiens* hsa04530	2/139	0.30058	0.44392	1.11245	-0.90343	*RRAS2, HRAS*
Parkinson's disease *Homo sapiens* hsa05012	2/142	0.30925	0.44392	1.12573	-0.91421	*COX6C, SNCAIP*
FoxO signaling pathway *Homo sapiens* hsa04068	2/133	0.28319	0.43084	1.09638	-0.92317	*HRAS, PIK3R5*
Natural killer cell mediated cytotoxicity *Homo sapiens* hsa04650	2/135	0.28899	0.43594	1.12765	-0.93624	*HRAS, PIK3R5*
Oxidative phosphorylation *Homo sapiens* hsa00190	1/133	0.65260	0.68735	2.52260	-0.94574	*COX6C*
Lysosome *Homo sapiens* hsa04142	1/123	0.62381	0.66891	2.37049	-0.95320	*SORT1*
Hepatitis C *Homo sapiens* hsa05160	2/133	0.28319	0.43084	1.13831	-0.95847	*HRAS, PIK3R5*
Dopaminergic synapse *Homo sapiens* hsa04728	1/129	0.64135	0.68360	2.53836	-0.96555	*PLCB3*
Cell adhesion molecules (CAMs) *Homo sapiens* hsa04514	2/142	0.30925	0.44392	1.21113	-0.98357	*HLA-DMA, HLA-DRB1*
Endocytosis *Homo sapiens* hsa04144	3/259	0.33422	0.44853	1.22856	-0.98503	*VPS45, HRAS, DNM1*
Signaling pathways regulating pluripotency of stem cells *Homo sapiens* hsa04550	2/142	0.30925	0.44392	1.23931	-1.00645	*HRAS, PIK3R5*
MAPK signaling pathway *Homo sapiens* hsa04010	3/255	0.32560	0.44853	1.26106	-1.01109	*FGF18, RRAS2, HRAS*
Hepatitis B *Homo sapiens* hsa05161	2/146	0.32078	0.44609	1.30073	-1.05000	*HRAS, PIK3R5*
Vascular smooth muscle contraction *Homo sapiens* hsa04270	1/120	0.61473	0.66316	2.57930	-1.05941	*PLCB3*
Focal adhesion *Homo sapiens* hsa04510	2/202	0.47474	0.55638	1.84982	-1.08455	*HRAS, PIK3R5*
Glutamatergic synapse *Homo sapiens* hsa04724	1/114	0.59591	0.64678	2.53383	-1.10411	*PLCB3*
TNF signaling pathway *Homo sapiens* hsa04668	1/110	0.58287	0.63651	2.44708	-1.10550	*PIK3R5*
Jak-STAT signaling pathway *Homo sapiens* hsa04630	2/158	0.35510	0.46515	1.47581	-1.12959	*HRAS, PIK3R5*
Purine metabolism *Homo sapiens* hsa00230	2/176	0.40542	0.49769	1.63306	-1.13953	*RRM2, HPRT1*
Non-alcoholic fatty liver disease (NAFLD) *Homo sapiens* hsa04932	2/151	0.33514	0.44853	1.44684	-1.16004	*COX6C, PIK3R5*
Pathways in cancer *Homo sapiens* hsa05200	4/397	0.38082	0.48419	1.60060	-1.16089	*PLCB3, FGF18, HRAS, PIK3R5*
Toll-like receptor signaling pathway *Homo sapiens* hsa04620	1/106	0.56941	0.62565	2.47553	-1.16094	*PIK3R5*
Pyrimidine metabolism *Homo sapiens* hsa00240	1/105	0.56598	0.62565	2.57778	-1.20889	*RRM2*
Progesterone-mediated oocyte maturation *Homo sapiens* hsa04914	1/98	0.54121	0.60588	2.48037	-1.24285	*PIK3R5*
Retrograde endocannabinoid signaling *Homo sapiens* hsa04723	1/101	0.55199	0.61409	2.64948	-1.29193	*PLCB3*
Fc gamma R-mediated phagocytosis *Homo sapiens* hsa04666	1/93	0.52267	0.59637	2.53877	-1.31225	*PIK3R5*
Glycerophospholipid metabolism *Homo sapiens* hsa00564	1/95	0.53017	0.59728	2.58625	-1.33288	*CHKA*
Circadian entrainment *Homo sapiens* hsa04713	1/95	0.53017	0.59728	2.64183	-1.36153	*PLCB3*
Hematopoietic cell lineage *Homo sapiens* hsa04640	1/88	0.50339	0.57809	2.68613	-1.47209	*HLA-DRB1*
GABAergic synapse *Homo sapiens* hsa04727	1/88	0.50339	0.57809	2.79398	-1.53119	*SLC6A11*
Small cell lung cancer *Homo sapiens* hsa05222	1/86	0.49547	0.57643	2.84785	-1.56891	*PIK3R5*
Aldosterone synthesis and secretion *Homo sapiens* hsa04925	1/81	0.47511	0.55638	2.93992	-1.72367	*PLCB3*
Pertussis *Homo sapiens* hsa05133	1/75	0.44962	0.53355	3.05927	-1.92183	*C3*
Inositol phosphate metabolism *Homo sapiens* hsa00562	1/71	0.43196	0.51603	3.22123	-2.13113	*PLCB3*
PPAR signaling pathway *Homo sapiens* hsa03320	1/69	0.42292	0.50865	3.24836	-2.19591	*SLC27A2*

p53 signaling pathway *Homo sapiens* hsa04115	1/69	0.42292	0.50865	3.32049	-2.24467	*RRM2*
Pancreatic cancer *Homo sapiens* hsa05212	1/66	0.40909	0.49876	3.34664	-2.32804	*PIK3R5*
mTOR signaling pathway *Homo sapiens* hsa04150	1/60	0.38046	0.48419	3.29492	-2.38976	*PIK3R5*
Renin secretion *Homo sapiens* hsa04924	1/64	0.39970	0.49407	3.43124	-2.41930	*PLCB3*
Colorectal cancer *Homo sapiens* hsa05210	1/62	0.39015	0.48906	3.55029	-2.53938	*PIK3R5*
Regulation of lipolysis in adipocytes *Homo sapiens* hsa04923	1/56	0.36062	0.46515	3.38324	-2.58954	*PIK3R5*
Synaptic vesicle cycle *Homo sapiens* hsa04721	1/63	0.39494	0.49161	3.71483	-2.63781	*DNM1*
Arachidonic acid metabolism *Homo sapiens* hsa00590	1/62	0.39015	0.48906	3.93436	-2.81409	*CYP2J2*
Legionellosis *Homo sapiens* hsa05134	1/55	0.35556	0.46515	3.76453	-2.88138	*C3*
Starch and sucrose metabolism *Homo sapiens* hsa00500	1/56	0.36062	0.46515	4.08474	-3.12647	*GYS1*
Notch signaling pathway *Homo sapiens* hsa04330	1/48	0.31905	0.44609	3.88715	-3.13786	*KAT2B*
Ovarian steroidogenesis *Homo sapiens* hsa04913	1/50	0.32968	0.44853	4.03806	-3.23763	*CYP2J2*
Amino sugar and nucleotide sugar metabolism *Homo sapiens* hsa00520	1/48	0.31905	0.44609	4.01121	-3.23801	*UAP1*
Bladder cancer *Homo sapiens* hsa05219	1/41	0.28050	0.43084	3.89141	-3.27662	*HRAS*
Type II diabetes mellitus *Homo sapiens* hsa04930	1/48	0.31905	0.44609	4.12303	-3.32828	*PIK3R5*
Mineral absorption *Homo sapiens* hsa04978	1/51	0.33494	0.44853	4.15519	-3.33155	*ATP1A1*
Glutathione metabolism *Homo sapiens* hsa00480	1/52	0.34015	0.45185	4.27414	-3.39543	*RRM2*
Drug metabolism - other enzymes *Homo sapiens* hsa00983	1/46	0.30825	0.44392	4.25038	-3.45176	*HPRT1*
Cysteine and methionine metabolism *Homo sapiens* hsa00270	1/45	0.30278	0.44392	4.38106	-3.55788	*AHCYL1*
Tryptophan metabolism *Homo sapiens* hsa00380	1/40	0.27482	0.43084	4.46818	-3.76227	*KYNU*
Glycine, serine and threonine metabolism *Homo sapiens* hsa00260	1/40	0.27482	0.43084	4.62082	-3.89079	*AGXT*
African trypanosomiasis *Homo sapiens* hsa05143	1/35	0.24577	0.40885	4.35659	-3.89661	*PLCB3*
Regulation of autophagy *Homo sapiens* hsa04140	1/39	0.26910	0.43084	4.75008	-3.99963	*ATG14*
Thyroid cancer *Homo sapiens* hsa05216	1/29	0.20938	0.36901	4.22898	-4.21595	*HRAS*
Pentose phosphate pathway *Homo sapiens* hsa00030	1/29	0.20938	0.36901	4.25791	-4.24479	*FBP1*
Alanine, aspartate and glutamate metabolism *Homo sapiens* hsa00250	1/35	0.24577	0.40885	4.77923	-4.27464	*AGXT*
Fructose and mannose metabolism *Homo sapiens* hsa00051	1/32	0.22779	0.38615	4.99270	-4.75068	*FBP1*
Linoleic acid metabolism *Homo sapiens* hsa00591	1/29	0.20938	0.36901	4.89436	-4.87928	*CYP2J2*
Glyoxylate and dicarboxylate metabolism *Homo sapiens* hsa00630	1/28	0.20316	0.36901	4.93064	-4.91545	*AGXT*
Proximal tubule bicarbonate reclamation *Homo sapiens* hsa04964	1/23	0.17128	0.35073	4.97539	-5.21289	*ATP1A1*
Phototransduction *Homo sapiens* hsa04744	1/27	0.19688	0.36505	5.43863	-5.48068	*SAG*
2-Oxocarboxylic acid metabolism *Homo sapiens* hsa01210	1/17	0.13138	0.32937	5.22227	-5.79976	*ACY1*
Maturity onset diabetes of the young *Homo sapiens* hsa04950	1/26	0.19055	0.36471	5.81982	-5.87010	*HNF4A*
Circadian rhythm *Homo sapiens* hsa04710	1/30	0.21557	0.36902	6.05243	-6.03376	*BHLHE41*

Appendix

Arginine biosynthesis *Homo sapiens* hsa00220	1/20	0.15156	0.34671	6.17173	-6.53756	*ACY1*