PhD degree in Molecular Medicine (curriculum in Molecular Oncology)

European School Of Molecular Medicine (SEMM)

University of Milan and University of Naples "Federico II"

Settore disciplinare: BIO/11

Tumor suppressor mutagenesis driven by DNA deaminase

Lara Sicouri

IFOM, Milan

Matricola n. R09850

Supervisor: Dr. Svend K. Petersen-Mahrt

IFOM, Milan

Anno Accademico: 2014-2015

A Isabel,

Insegui sempre i tuoi sogni.

INDEX

LIST OF ABBREVIATIONS

3MeA	3-methyladenine
5'-dRP	5'-deoxyribose phosphate moiety
5mC	5-methylcytosine
6-4PPs	pyrimidine (6-4) pyrimidone photoproducts
AID	activation-induced deaminase
AP	abasic sites
APC	adenomatous poliposis coli
APE1	AP endonuclease 1
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
ATM	ataxia telangiectasia mutated
ATR	ataxia-telangiectasia and Rad3- related
BER	base-excision repair
bp	base pairs
BU	5 – bromouracil
C	immunoglobulin constant region
Cas	CRISPR-Associated genes
cDNA	complement DNA
CHK	checkpoint kinase
CPD	*cis – syn* cyclobutane pyrimidine dimers
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	short CRISPR RNA
CSR	class switch recombination
DBD	DNA binding domain
dC	cytosine
DDR	DNA damage response
dG	guanine
DLBCL	diffuse large B-cell lymphoma
DNA-PK	DNA-dependent protein kinase
dpf	days post fertilization
DSB	double strand breaks
dT	thymine
dU	uracil
EBV	Epstein-Bar virus

EGFP-T-AIDmt	EGFP-TALE-AIDmt
EGFP-T-AIDwt	EGFP-TALE-AIDwt
ENU	N-ethyl-N-nitrosourea
ERD	ERF Repressor Domain
ES cells	embryonic stem cells
EXO1	exonuclease 1
FEN1	flap endonuclease 1
GC	germinal center
gDNA	genomic DNA
GEM	genetically engineered model
GG-NER	global-genome nucleotide excision repair
gRNA	guide RNA
h	hours
HBV	Hepatitis B Virus
hpf	hours post fertilization
HPV	Human Papillomavirus
HR	homologous recombination
Ig	immunoglobulin
iGC	immunoglobulin gene conversion
IgH	immunoglobulin heavy chain *locus*
IgL	immunoglobulin light chain *locus*
iPS cells	induced pluripotent stem cells
IR	Ionizing radiation
kb	kilobases
kD	kilodaltons
KRAB	Krüppel-associated box
LOH	loss of heterozygosity
MCR	mutation cluster region
MDM2	murine double mutant 2
Min	multiple intestinal neoplasia
min	minutes
MMC	mitomycin C
MMR	mismatch repair
MO	microorganisms

mpf	months post fertilization
mRNA	messenger RNA
mt	mutant
ncMMR	non-canonical mismatch repair
NER	nucleotide excision repair
NES	nuclear export signal
NGS	next-generation sequencing
NHEJ	non-homologous end joining
NLS	nuclear localization signal
Oligo	oligonucleotide
PAM	protospacer adjacent motif
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PGC	primordial germ cells
PIKK	phosphatidylinositol 3-kinase-like protein kinase family
pMEFs	primary mouse embryonic fibroblasts
polβ	polymerase β
pRB	retinoblastoma protein
PTM	posttranslational modification
RNAPII	RNA polymerase II
ROS	reactive oxygen species
RT	room temperature
RVD	repeat variable diresidue
SAM	S-adenosylmethionine
sec	seconds
SHM	somatic hypermutation
SID	mSIN3 Interaction domain
SNP	single nucleotide polymorphism
SSB	single-strand breaks
ssDNA	single strand DNA
SV40	simian virus 40
T-Aid-mt	TALE-Aid-mt-FL
T-Aid-wt	TALE-Aid-wt-FL
TALE	transcription activator-like effector

TALEN	TALE nuclease
TC-NER	transcription coupled nucleotide excision repair
Tet	tetracyclin-regulated system
TP53	tumor protein p53
tracrRNA	trans-activating crRNA
UAS	upstream activation sequence
UNG	uracil-DNA glycosylase
UV	ultraviolet
V	immunoglobulin variable domain
wt	wild type
XRCC1	x-ray repair cross-complementing protein 1
ZFN	zing finger nuclease
ZFP	zinc finger protein

FIGURES INDEX

TABLES

ABSTRACT

Genomic instability is commonly associated with pathological disorders including cancer. The progressive accumulation of genetic abnormalities in cancer-associated genes can confer cellular autonomous proliferation, contributing to the oncogenic transformation. Although animal models have been instrumental to the understanding of the molecular mechanisms responsible for tumor progression, there are severe limitations for success. This thesis work aimed at the development of an innovative animal model that could recapitulate the ongoing lifelong accumulation of DNA lesions, leading to neoplastic transformation. To this end, the natural DNA mutating enzyme AID was fused to sequence specific DNA binding proteins. They were engineered to target tumor suppressor genes and to induce low-frequency mutagenesis in cell lines and in zebrafish. A TALE-Aid fusion protein was targeted to the *p53 locus* of mouse cell lines, and the Aid-dependent mutations were monitored by next-generation sequencing. The induced mutations occurred at a comparable frequency to those observed in Aid-induced non-immunoglobulin gene targets in B cells. Mutations were found mostly in the DNA binding domain of *p53*, possibly reflecting AID- hotspot residues in *p53*. Our approach also induced mutations that have not been characterized previously, and could provide further insight into p53 dependent oncogenesis. When TALE-AID were expressed and targeted to the *p53 locus* of zebrafish embryos, the activity induced developmental abnormalities with variable severity, leading to both an increased mortality, and impaired ovarian maturation and fertility. We have developed and initially characterized a novel tool for the *in* vitro / *in vivo* study of the accumulation of mutations in cancer-associated genes.

CHAPTER 1 – INTRODUCTION

1.1 GENOME INSTABILITY AND DNA DAMAGE RESPONSE

Eukaryotic cells are constantly subjected to the effects of physical and chemical stress from endogenous and exogenous origins. The chemical and biochemical reactions in response to this stress damage macromolecules including DNA.

The cellular DNA damage that occurs every day results in the accumulation of thousands of lesions (Table 1). If unrepaired, these lesions can lead to mutations and chromosomal aberrations, turning on oncogenes or inactivating tumor suppressor genes. Either mechanism increases the risk of human diseases such as cancer. Hence, the integrity of the genetic information is a requirement for the survival of the cell itself. To ensure genome integrity and faithful genome propagation, cells have evolved several mechanisms of damage tolerance, surveillance and repair (Figure 1).

Table 1: DNA lesions generated by endogenous and environmental DNA damage.

Spontaneous DNA damage	DNA lesion generated	Number of lesions/cell/day	Reference
Depurination	AP sites	9,000	(Lindahl, 1972)
Deamination	dU	100 - 500	(Frederico et al., 1990; Lindahl and Nyberg, 1974)
Alkylation	3MeA	600	(Lindahl and Barnes, 2000)
Oxidation	8-oxo-G	20,000	(Beckman and Ames, 1997)
UV radiation	Pyirimidine dimers, (6-4) photoproducts	10^5	(Hoeijmakers, 2009)

1.1.1 Origin of endogenous DNA damages

Although nature has chosen DNA as the predominant carrier of genetic information, it can be considered fragile due to some chemical and biological instabilities (Figure 1). N-glycosyl bonds joining the base and sugar moiety are susceptible to spontaneous hydrolysis, resulting in the formation of mutagenic abasic sites (AP). Purines are more susceptible than pyrimidines, and it has been estimated that each human cell undergoes 9,000 – 10,000 depurinations per day (Table 1) (Lindahl, 1972; Nakamura and Swenberg, 1999). DNA bases are also susceptible to spontaneous deamination, with the N4 of cytosine (dC) and 5-methylcytosine (5mC) being the main target, leading to uracil (dU) and dT, respectively (Krokan et al., 2002). This hydrolytic deamination occurs 100 – 500 times per day in human cells (Frederico et al., 1990; Lindahl and Nyberg, 1974) (Table 1), and 5-methylcytosine being about 5 – 10 times more reactive.

Reactive oxygen species (ROS), exogenous or generated during normal cellular metabolism, also cause endogenous DNA damage. In aerobic cells, oxygen is normally converted to water without accumulation of reactive intermediates, yet 0.2 – 2 % is converted into reactive molecules, such as free radicals and peroxide (Maynard et al., 2009). It has been estimated that approximately 20,000 bases in the DNA of each human cell are damaged by ROS every day (Beckman and Ames, 1997).

Alkylation of endogenous DNA can be derived from the reactive methyl donor S-adenosylmethionine (SAM), predominantly resulting in the formation of 3-methyladenine (3MeA), with a frequency of ~ 600 3MeA per day per cell (Lindahl and Barnes, 2000).

The DNA replication process itself can be a source of genomic instability, including errors in DNA synthesis occurring during copying of the DNA template. The error rate of DNA polymerases is estimated to be one in 10^7 nucleotides (Kunkel and Bebenek, 2000), with specific regions of the genome called hotspots, having higher susceptibility to mutation due to their base composition. These hotspots can be encountered in repeat elements,

fragile sites of nucleotide repeats, and telomeric regions (Aguilera and Gomez-Gonzalez, 2008). For example, fragile sites are characterized by nucleotide repeats (G-rich or AT rich) and are susceptible to the formation of hairpin structures that are prone to breakage and DNA damage or mutation (Durkin and Glover, 2007).

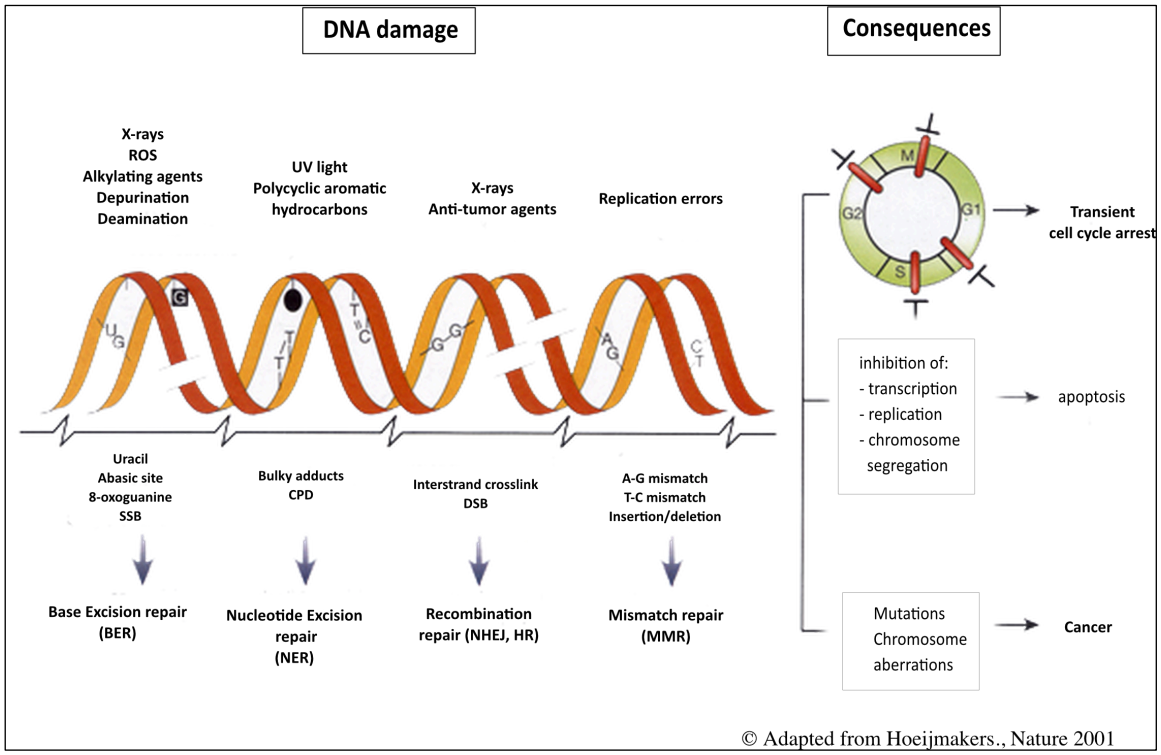


Figure 1: DNA damage, DNA repair and consequences.

DNA damage is induced by endogenous and exogenous agents and results in the insertion of modified bases, SSBs, DSBs and bulky adducts. Upon DNA damage, the cell can undergo transient cell cycle arrest to repair the lesions or, in case of extensive damage, can also undergo apoptosis. Repeated or persisted DNA damage causes permanent changes in the DNA sequence (mutation) or rearrangements, possibly leading to neoplasia.

1.1.2 Environmental DNA damage

Aside from endogenous agents, environmental physical and chemical sources and biological agents, such as microorganisms (MO), can be a source of DNA damage (Figure 1). Examples of physical genotoxic agents are ultraviolet radiation (UV) and ionizing radiation (IR). UV causes the formation of dimeric photoproducts between adjacent pyrimidine bases, *cis – syn* cyclobutane pyrimidine dimers (CPD) and pyrimidine (6-4) pyrimidone photoproducts (6-4PPs) (Pfeifer et al., 2005), with an estimated frequency of 10^5 DNA lesions per cell per day (Hoeijmakers, 2009). The main genotoxic intermediate from IR are free radicals from the radiolysis of water, which lead to single-strand and double strand breaks (SSBs and DSBs respectively), as well as other DNA lesions (Wallace, 1998).

Chemical compounds can be distinguished by their mode of actions on DNA. Base analogs, e.g. 5 - bromouracil (BU), that structurally resemble both purines and pyrimidines can be incorporated into DNA during replication and cause base pair changes upon strand recognition (Lasken and Goodman, 1984). Intercalating agents insert themselves in between the DNA bases, causing frameshift mutations (Dasgupta et al., 2012). Crosslinking agents such as psoralen, cisplatin, and mitomycin C (MMC) introduce covalent intra- and inter-strand bonds (Ciccia and Elledge, 2010; Rajski and Williams, 1998). Although the direct mechanisms of action are not understood, it is thought that air pollution, tobacco smoke, pesticides, asbestos, and other toxins, are also source of exogenous DNA damage, either direct or indirect (Msiska et al., 2010; Phillips et al., 1988; Pope et al., 2002).

Viral infections can also induce DNA damage. For example, chronic exposure to Hepatitis B Virus (HBV), Human Papillomavirus (HPV) and Epstein-Bar Virus (EBV), can lead to direct inactivation of the tumor protein p53 (TP53) and the retinoblastoma protein (pRB), allowing for the accumulation of genomic instability (Chen et al., 2014).

1.1.3 DNA damage response

To counteract the accumulation of DNA lesions in a cell, the DNA damage response (DDR) is a network of interacting pathways that senses DNA damage and transduces this information to the cells to initiate a DNA repair response, in order to protect against transformation (Branzei and Foiani, 2005; Zhou and Elledge, 2000). Key components of the DDR machinery are proteins of the phosphatidylinositol 3-kinase-like protein kinase (PIKKs) family, ataxia telangiectasia mutated (ATM), ataxia-telangiectasia and Rad3-related (ATR), and the DNA-dependent protein kinase (DNA-PK).

Upon lesion recognition by sensor proteins (e.g. MRN complex and RPA / 9-1-1 complex) (Ciccia and Elledge, 2010; Zhou and Elledge, 2000), the transducers (either ATM or ATR) are recruited at the lesion and undergo auto-phosphorylation. In turn, they phosphorylate mediator proteins (e.g. BRCA1, 53BP1, MDC1, H2AX, TOPBP1). Posttranslational modifications (PTMs) of mediators induce the amplification of DDR signals by recruiting effector proteins to the DNA lesion (e.g. TP53, CDC25). The outcome of activating this network includes DNA repair, checkpoint signaling, and cell fate decision (Jackson and Bartek, 2009). Two major ATM/ATR targets are checkpoint kinase 1 and 2 (CHK1 and CHK2). CHK1 is activated primarily by ATR in response to SSBs, while CHK2 activation is downstream of ATM in response to DNA damaging agents that creates DSBs (Liu et al., 2000; Stracker et al., 2009). CHK1 can then target the CDC25 family of phosphatases with consequent slow down of cell cycle progression (Mailand et al., 2000), while CHK2 phosphorylates TP53 that, depending on the severity of the damage, triggers cell-cycle arrest, apoptosis or senescence (Chehab et al., 2000).

1.2 DNA REPAIR MECHANISMS

Ensuring genome integrity, the DDR, upon DNA damage and replication stress, initiates either repair or processing of the stalled-replication forks to avoid fork-collapse. Evolution has developed multiple DNA repair mechanisms able to deal with every specific type of lesion (Figure 1). Five main repair mechanisms are present in mammals, with some overlapping specificities: base-excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER), homologous recombination (HR), and non-homologous end joining (NHEJ).

1.2.1 Base excision repair

BER is the predominant mechanism active against damage caused by cellular metabolism that result in single base lesions, including oxidation, alkylation, deamination and SSBs (reviewed by (Zharkov, 2008)). The first step is the recognition and excision of the damaged base by lesion-specific DNA glycosylases cleaving the N-glycosidic bond and producing an AP site (Lindahl, 1974). Eleven different DNA glycosylases are known to act on single or overlapping substrates. Although DNA glycosylases are subdivided into four structurally distinct super families, they share a common mode of action: flipping the base out of the DNA helix into a catalytic pocket for excision (Jacobs and Schar, 2012). The resultant AP site is processed by AP endonuclease 1 (APE1), that hydrolyzing the phosphodiester backbone 5' to the AP site, creates a 3'OH and a 5'-deoxyribose phosphate moiety (5'-dRP) (Lindahl, 1974). APE1 interacts with DNA polymerase β (polβ), which in turn hydrolyzes the 5'-dRP and fills in the single-nucleotide gap. The gap can be sealed by the protein complex DNA ligase III / XRCC1 (X-ray repair cross-complementing protein 1) (Kubota et al., 1996). In this process only the site of the lesion is repaired, is termed short-patch BER, and represents 80 – 90 % of all BER in a cell. An alternative synthesis/ligation step, called long-patch BER, is initiated when DNA polβ is unable to

remove the 5'-dRP and several nucleotides are needed to fill the gap. The resultant displaced overhang is cleaved by Flap endonuclease 1 (FEN1) prior to completion by another ligase complex (DNA Ligase I) (Frosina et al., 1996).

1.2.2 Mismatch repair

MMR plays an important role in post-replication repair of misincorporated bases and insertion/deletion loops, both of which can result from errors during DNA polymerase activity. In mammals there are two different MMR pathways, canonical (classical) MMR and non-canonical MMR (ncMMR) (reviewed by (Bak et al., 2014)).

The MMR pathway can be divided into three steps: 1) recognition, 2) excision of the region surrounding the mismatch, and 3) re-synthesis and ligation of the gap. The canonical MMR pathway is a high fidelity process involving two major protein complexes, hMutS and hMutL. The initial recognition step is carried out by one of two hMutS orthologous complexes: MSH2/6 (hMutSα) recognizes mismatches and single-base loops, while MSH2/3 (hMutSβ) recognizes larger insertion/deletion loops. The bound proteins recruit MLH1/PMS2 (hMutLα) heterodimeric complex that activates exonuclease 1 (EXO1) - mediated degradation of the damaged DNA strand (Kunkel and Erie, 2005; Tran et al., 2004). The gap left by EXO1 is resynthesized by DNA polymerase δ and ε and sealed by DNA ligase I (Gu et al., 1998).

In contrast to replication-linked classical MMR, non-canonical MMR can be highly mutagenic and is activated by a variety of lesions (Pena-Diaz et al., 2012). In B cells, MMR is recruited during the process of somatic hypermutation (SHM) to generate antibody diversity (see section 1.3.1.1) (Cascalho et al., 1998; Rada et al., 2004). Following the mismatch recognition by hMutSα and hMutLα, PCNA is recruited and monoubiquitinated at lysine 164 (mUb-PCNA) (Langerak et al., 2007). mUb-PCNA allows

for the binding of DNA-polymerase η, a translesion polymerase that mediates error-prone DNA synthesis, thus leading to the incorporation of mutations (Chahwan et al., 2012).

1.2.3 Further repair pathways

Aside from BER and MMR, other DNA repair pathways recognize specific DNA lesions, such as pyrimidine dimers and DSBs, such as NER, HR, and NHEJ.

NER can recognize and remove a wide range of helix distorting base lesions, including pyrimidine dimers from UV light. NER has two sub-pathways that differ with regards to lesion recognition: a) transcription-coupled NER (TC-NER) takes place at DNA damage sites blocking transcription, with the stalled polymerase serving for lesion recognition (Fousteri and Mullenders, 2008); b) global genome NER (GG-NER) is recruited to sites of damage by DNA damage-binding proteins and eliminates lesions throughout the genome (Gillet and Scharer, 2006).

HR together with NHEJ is the main pathway involved in DSBs repair. HR occurs specifically in late S and G2 phases of the cell cycle. It is an error-free mechanism as it uses the genetic information contained in the undamaged sister chromatid as a template (Li and Heyer, 2008). In contrast to HR, NHEJ functions throughout the cell cycle and involves the elimination of DSBs by direct re-ligation of the broken ends. Small insertions and deletions at the breakage site account for the error prone nature of NHEJ pathway (Lieber, 2010).

1.3 BALANCE BETWEEN DNA DAMAGE AND DNA REPAIR

As shown in Table 1, as a result of exposure to various damaging agents, a cell experiences up to 10^5 spontaneous DNA lesions per day. The fine crosstalk among DNA repair pathways ensures that the lesions are repaired efficiently. Furthermore, in case of excessive damage, the cell can also undergo programmed-cell death to avoid tumorigenic transformation and to ensure the survival of the organism. On the other hand, mutations themselves are a fundamental process without which evolution could not occur. Mutations that arise from inaccurate DNA repair mechanisms, if controlled, can lead to genetic variation, thus creating biological novelty that can be positively selected during evolution. A fine balance between genome stability and instability allows just enough genetic variation for evolution to take place.

1.3.1 Disturbing the balance toward evolution: Activation-induced deaminase

Activation-induced deaminase (AID) is a natural DNA mutator finely regulated and expressed in our body. This enzyme introduces DNA lesions by deaminating cytosine (dC) residues to uracils (dU) in single strand DNA (ssDNA) (Figure 2A) (Dickerson et al., 2003). dUs, upon double strand DNA (dsDNA) reformation can activate various DNA repair mechanisms, or if ignored lead to transitions during replication. Given its ability to mutate DNA, AID is a key player in the homeostasis between survival and evolution.

AID is a 24 kilodaltons (kD) protein discovered in 1999 in a subtractive hybridization screen, comparing unstimulated and stimulated mouse B cells. It was shown to be expressed in activated B cells in the germinal center (GC) (Muramatsu et al., 1999). Subsequent studies have highlighted the importance of AID in mediating two processes in acquired immunity, somatic hypermutation (SHM) and class switch recombination (CSR); both are part of antigen-dependent antibody diversification (Muramatsu et al., 2000). Furthermore, in chicken and some ruminants, AID mediates immunoglobulin gene

conversion (iGC), the primary mechanism used to generate immunoglobulin (Ig) gene diversity (Arakawa et al., 2002; Harris et al., 2002). Initially, AID was thought to be an RNA editing enzyme, but it was later on discovered that AID could mutate *Escherichia coli* (*E. coli*) DNA, providing substantial insight into its mechanism of action (Petersen-Mahrt et al., 2002).

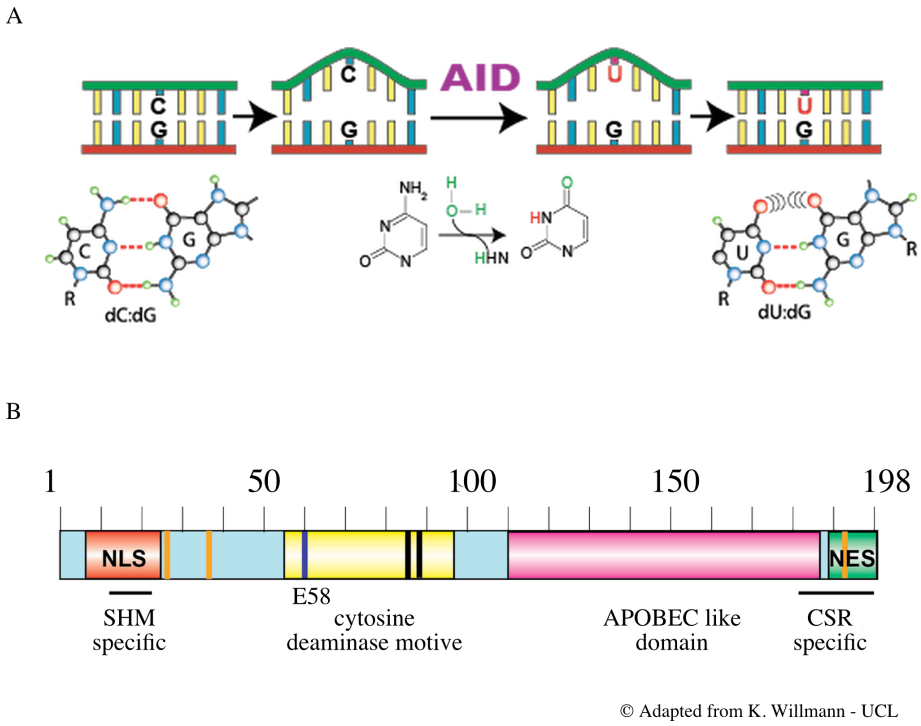


Figure 2: AID properties and structure.

A) The deamination reaction catalyzed by AID on ssDNA produces a uracil from a cytosine, resulting in a dU:dG mismatch in dsDNA. B) AID protein structure is depicted in light blue, with the functional domains in different colors. The catalytic active site is highlighted in blue, the zinc-coordinating amino acid residues in black and the critical sites of phosphorylation in orange.

AID is the founding member of the APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of cytosine deaminases, with the human genome comprising 10 members, each playing a potential different physiological function, but sharing a common zinc-binding catalytic domain (Conticello et al., 2005). AID is

composed of an N-terminal domain containing a nuclear localization signal (NLS), essential for SHM (Brar et al., 2004); a catalytic domain where a histidine and two cysteines coordinate a zinc atom and form a pocket, with glutamic acid (E) providing the proton donor; and a C-terminus with a nuclear export signal (NES), essential for CSR (Barreto et al., 2003; Prochnow et al., 2007) (Figure 2B).

For efficient deamination, AID requires ssDNA, which during Ig diversification can be provided by the concomitant transcription elongation process (Basu et al., 2011; Pavri et al., 2010; Vaidyanathan et al., 2014; Willmann et al., 2012). Moreover, biochemically AID has a preference for dC in the context of WRC (W= dA or dT; R= dA or dG) (Beale et al., 2004; Morgan et al., 2004; Pham et al., 2003). Aside from its involvement in antibody affinity maturation (see section 1.3.1.1), several studies revealed that AID is expressed outside of the immune system, such as primordial germ cells (PGC), embryonic stem cells, oocytes, and spermatocytes (Morgan et al., 2004; Schreck et al., 2006), and hence can be regulated by non-B cell specific factors such as hormones (estrogen, progesterone, and androgen) (Lin et al., 2009; Pauklin and Petersen-Mahrt, 2009; Pauklin et al., 2009). This suggested additional functions of AID outside the immune system, such as epigenetic reprogramming and active demethylation, which is supported by AID's ability to deaminate 5-methylcytosine (Bhutani et al., 2010; Franchini et al., 2014; Franchini et al., 2012; Morgan et al., 2004; Popp et al., 2010; Rai et al., 2008).

1.3.1.1 AID in SHM and CSR

In its daily fight against pathogens, the immune system creates a large repertoire of low affinity antibodies, and upon antigen encounter produces high affinity antibodies for each target. Immunoglobulins are heterodimeric proteins composed of two heavy (H) and two light (L) chains linked via disulfide bridges. Each component chain contains one N-terminal variable (V) domain, which contributes to the antigen-binding site, and one or

more C-terminal constant (C) domains that determine the isotype (Schroeder and Cavacini, 2010). In mammals there are two classes of light chain *loci* (IgL), termed Igκ and Igλ, and a single heavy chain *locus* (IgH).

Upon RAG1/2 dependent V(D)J recombination in the bone marrow (Rajewsky, 1996), IgM molecules are expressed on the surface of B cells as part of the diverse naïve Ig repertoire. B cells expressing fully reactive IgM can colonize secondary lymphoid organs, where they undergo a physiological process of alterations of the rearranged *Ig locus*, termed *Ig locus* diversification, with AID playing a fundamental role. Somatic hypermutation introduces point mutations at high rates into variable region exons of IgH and IgL genes, which upon antigen selection generates antibodies with higher affinity for antigens (Neuberger et al., 2003). After AID initiates SHM by the deamination of dC, the resulting dU:dG mismatch becomes a substrate for different DNA repair pathways (Petersen-Mahrt, 2005). If DNA replication takes place before mismatch recognition, a transition mutation of dC to dT will occur (Neuberger et al., 2003). If uracil is recognized and removed by uracil-DNA glycosylase (UNG), an AP site can become a template for DNA replication, leading to the incorporation of any nucleotide and resulting in both transitions and transversions (by the action of low fidelity polymerases) (Di Noia and Neuberger, 2007; Neuberger et al., 2003). In addition to be partially processed by BER, the dU:dG mismatch or the AP site can be recognized by the MMR machinery, which allows for mutations at dA:dT near the initiating dU:dG lesion through an error-prone repair process (Cascalho et al., 1998; Rada et al., 2004) (see section 1.2.2).

The IgH is organized in sub-classes (μ, δ, γ, α, ε), with each defining the effector function of the antibodies upon class switch recombination (CSR). CSR is a DNA recombination event between the initially expressed C_H exons (Cμ) and downstream exons, such as Cγ, Cα and Cε, with the DNA in between being deleted, leading to the expression of IgG, IgA and IgE antibodies (reviewed in (Xu et al., 2012)). Aside from the deamination activity,

CSR also requires AID's C-terminus, a region seemingly unnecessary for deamination and SHM (Barreto et al., 2003; Shinkura et al., 2004; Ta et al., 2003). AID deaminates dC residues in the switch region of C_H. The removal of uracil by either UNG or components of ncMMR pathway introduces nicks or gaps on opposite strands of the switch region, leading to DSBs formation (Begum et al., 2004; Zan and Casali, 2008).

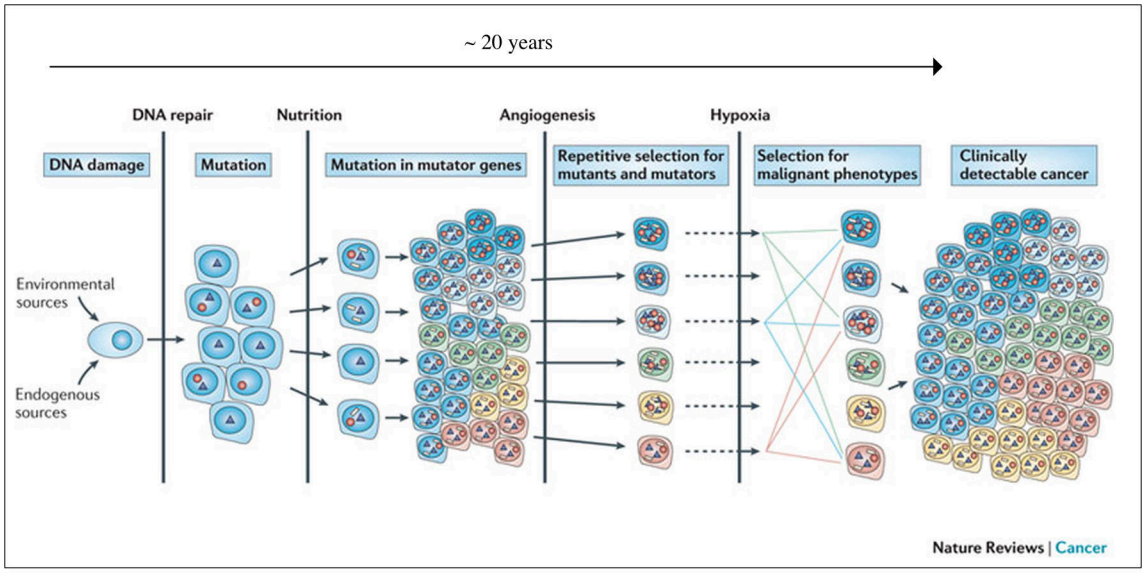
It is now clear how mutations, if controlled and programmed, contribute to genetic diversity. AID, by introducing mutations in the *Ig locus*, plays an important role in the positive selection and evolution of antibodies, a process required for an optimal immune response. On the other hand, uncontrolled expression of AID contributes to aberrant genetic changes and cancer.

1.3.2 Disturbing the balance towards cancer

The alteration of the balance between genomic stability and DNA lesions can cause an accumulation of somatic mutations, leading to genome instability, the hallmark of all human cancers (Hanahan and Weinberg, 2011). Cancer is an evolutionary process: among the cells that continuously acquire genetic variation by mutations and selection, occasionally a single cell acquires a set of sufficiently advantageous mutations that confers autonomous proliferation and transformation (Stratton et al., 2009). Tumors are heterogeneous and are composed of multiple cell types that participate in heterotypic interactions with one another. They are characterized by biological capabilities acquired during the multistep transformation process, which include: a) sustaining proliferative signaling, b) evading growth suppressors, c) resisting cell death, d) enabling replicative immortality, e) inducing angiogenesis, f) activating invasion and metastasis, g) avoiding immune destruction, h) deregulating cellular metabolism, i) promoting inflammation and, as already mentioned above, genome instability and mutations (Hanahan and Weinberg,

2011). Importantly even among the individual cell types, cellular heterogeneity derives from functional heterogeneity encoded by multiple aberrations in the cancer genome.

Considering that the spontaneous mutation rate in normal cells is not sufficient to account for the number of mutations found in human cancers (Loeb et al., 1974), somatic mutations found in cancer genomes are the consequence of exogenous or endogenous exposure to mutagens (section 1.1), defective or error-prone DNA repair mechanisms (section 1.2), intrinsic infidelity of the DNA replication machinery, and enzymatic modification of the DNA (section 1.3.1). As a result of the action/deregulation of the aforementioned pathways, cancer cells express the so called mutator phenotype, which allows a progressive accumulation of mutations during tumor progression (Loeb et al., 1974). Indeed, by the time cancer is detected clinically, cancer cells already contain thousands of mutations (Figure 3).



© Adapted from Loeb., Nat Rev Cancer. 2011

Figure 3: Mutation accumulation during cancer evolution.

Random mutations occur when DNA damage exceed the cell's capacity for DNA repair. For a solid tumor, more than 20 years pass from the time of exposure to a damaging source to the clinical detection of a tumor. This process starts with mutations in genes involved in genome stability (red circle), resulting in enhanced mutagenesis. Driver (blue triangle) and passenger (white rectangles) mutations accumulate after each round of selection in the tumor microenvironment, leading to cellular heterogeneity.

Mutations can be classified according to their consequence for cancer development. Driver mutations confer growth advantage on the cell and have been positively selected during cancer evolution, while passenger mutations do not confer any selective advantage and do not contribute to cancer development (Stratton et al., 2009). Driver mutations usually cluster in cancer genes, whereas passenger mutations are more or less randomly distributed and represent a rich source of historical information of the cancer genome, also known as mutational signature (Alexandrov et al., 2013; Helleday et al., 2014). This signature contains the scar of different mutational events that have occurred during the lifetime of the tumor and within a tumor, with different signatures as a result of heterogeneous processes, increasing the complexity of the final landscape (Alexandrov et al., 2013; Nik-Zainal et al., 2012). Recently, using next-generation sequencing (NGS), 30 primary human tumors have been analyzed and 21 different mutational signatures identified: some characterized by prominence of dC>dT substitutions, while others with primarily dC>dG transversions (Alexandrov et al., 2013; Nik-Zainal et al., 2012).

Beyond point mutations, changes in the DNA sequence of cancer genomes include genomic rearrangements such as insertions/deletions, chromosomal translocations and amplifications, that occur at much lower frequency than single base substitutions do (Stratton et al., 2009).

1.3.3 Cancer genes

In the last four decades, many important genes responsible for tumor progression have been discovered. Cancer arises from the accumulation of mutations in three types of so called cancer genes: oncogenes, tumor suppressor genes, and genes that maintain genomic integrity.

Alterations of an oncogene can render the gene constitutively active or active under non-physiological conditions. One mutate allele is often sufficient to confer growth advantage of a cell, with oncogene activation being the result of chromosomal translocation, gene amplification, or point mutations in residues critical for its activity (gain of function mutation). Key transforming oncogenes include *RAS*, *MYC* and *BRAF*, which under normal conditions govern the processes of growth, proliferation, and survival (Hanahan and Weinberg, 2011).

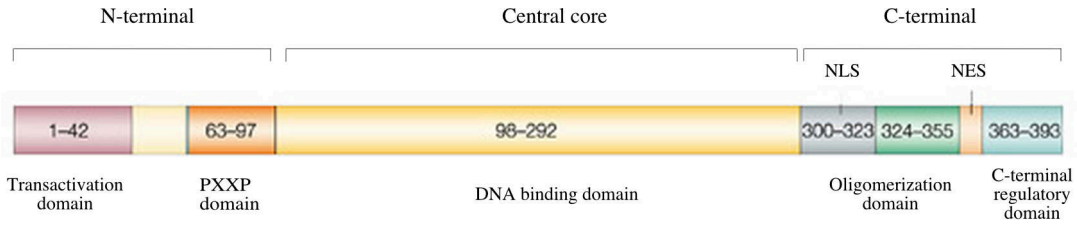
For oncogenesis, tumor suppressor genes are targeted in the opposite way and mutations impair the activity of the gene product. Here, changes arise from: missense mutations at essential residues for the activity, truncations to inactivate the protein, insertions or deletions in the gene, or epigenetic silencing of the expression (Vogelstein and Kinzler, 2004). Inactivation of both alleles is generally required to confer a selective advantage to the cell, known as the two-hit hypothesis (Knudson, 1971). During this process, a cell with a single inactive allele (either somatically acquired or genetically inherited) receives a second hit in its remaining functional copy of the tumor suppressor gene, thus acquiring a homozygous loss genotype – also known as loss of heterozygosity (LOH) (Knudson, 1971). In their physiological context, tumor suppressor genes operate in various ways to limit cell growth and proliferation, such as by promoting apoptosis or by blocking the cell cycle (Weinberg, 1995). Two of the most studied tumor suppressor genes are *TP53* and *APC* (see section 1.3.3.1).

The third class of cancer genes, called genome stability genes, promotes tumorigenesis when mutations occur in MMR, NER, BER and HR genes, or in critical DNA replication checkpoint genes (Cleaver, 2005; Friedberg, 2003; Helleday, 2010; Peltomaki, 1997). Stability genes keep genetic alterations at minimum, and thus their inactivation lead to the mutator phenotype (Loeb et al., 1974; Peltomaki, 2003). Similar to tumor suppressor genes, usually both alleles of stability genes are inactivated during cancer progression.

Mutations in cancer genes can occur both in the germline, resulting in hereditary predisposition to cancer, or in somatic cells leading to sporadic tumorigenesis. The Cancer Gene Census, a database of genes mutated in human cancers, contains 522 genes (Forbes et al., 2015). Of these, ~ 90 % are altered by somatic mutations and ~ 20 % by germline mutations. Moreover, approximately 90 % of somatic mutated genes are acting in a dominant manner, while 10 % act in a recessive manner (Stratton et al., 2009).

1.3.3.1 Tumor suppressors TP53 and APC

The tumor suppressor TP53, known also as "guardian of the genome", is a transcription factor essential for preventing inappropriate cell proliferation and maintaining genome integrity after genotoxic stress (Lane, 1992). TP53 mutations are the most common genetic alterations found in human cancer, ranging from 5 % to 80 % depending on the type, stage, and etiology of the tumor (Petitjean et al., 2007a). Germline mutations are associated with Li-Fraumeni syndrome, a disorder that greatly increases the risk of developing cancer via LOH (Li et al., 1988; Nichols et al., 2001). TP53 responds to oncogene activation, hypoxia, loss of normal cell contacts, and telomere shortening, each of which can lead to TP53-dependent transcriptional regulation of hundreds of genes. The three main consequences of TP53 activation are cell cycle arrest, senescence, and apoptosis. Aside from these outputs, other TP53 activities impacting tumor suppression include autophagy, regulation of metabolism, and angiogenesis (Sablina et al., 2005; Teodoro et al., 2007; Vousden and Ryan, 2009). Importantly, the choice between these responses is influenced by the type of cell, type of stress, and the action of TP53 co-activators. For example, in response to DNA damage, some cell types undergo TP53 mediated apoptosis (e.g. T lymphocytes) (Strasser et al., 1994), whereas others undergo senescence (e.g. mouse embryonic fibroblasts, MEFs).



© Adapted from Bode & Dong., Nat Rev Cancer. 2004

Figure 4: TP53 protein structure.

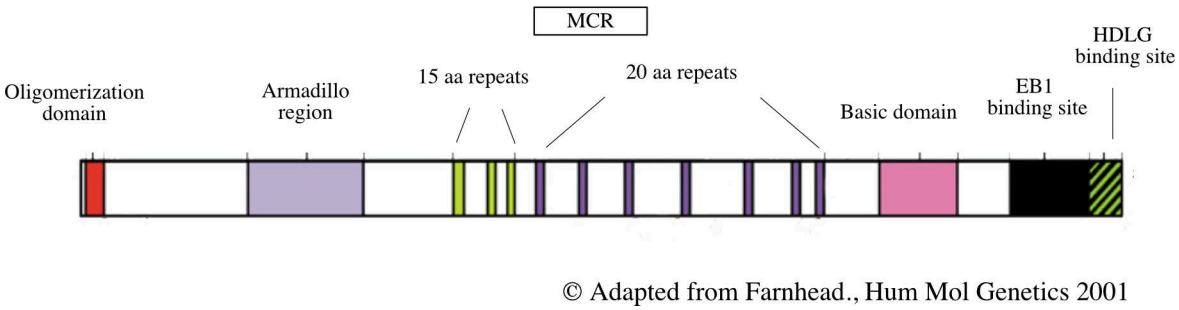
The TP53 protein consists of 393 amino acids. The functional domains are depicted in different colors. The N-terminal domain contains the transactivation domain and the poly-proline domain (PXXP) is important for the interaction with MDM2. The central region includes the DNA binding domain, necessary for sequence-specific DNA targeting. The C-terminal domain contains a nuclear localization signal (NLS), an oligomerization domain, a nuclear export signal (NES), and a regulatory domain.

The *TP53* gene contains 11 exons and encodes for a 393 amino acids containing protein (53 kD), whose structure is depicted in Figure 4. TP53 protein is composed of several structural and functional domains: an amino-terminal domain (residues 1 – 42), a proline-rich region with repetitive PXXP sequences (residue 63 – 97 where X is any amino acid), a highly conserved central core domain (residues 98 – 292), and a carboxyl-terminal region containing an oligomerization domain (residues 324 – 355), a regulatory domain (residues 363 – 393), a nuclear localization signal, and three nuclear export signals sequence. The N-terminal domain is required for transactivation activity and interaction with various factors including its regulator MDM2 (murine double mutant 2) (Fields and Jang, 1990; Lin et al., 1994). The poly-proline region is important for inducing cell death, but not growth arrest (Sakamuro et al., 1997), while the central core is composed mainly of the DNA binding domain (DBD), required for sequence-specific DNA binding (Kern et al., 1991). The vast majority (80 %) of *TP53* mutations found in human cancers are located in the DNA binding domain (Petitjean et al., 2007b). The oligomerization domain downstream of the nuclear localization domain is necessary for the tetramerization of TP53, enhancing DNA binding (McLure and Lee, 1998). Finally, the C-terminal tail is composed of a regulatory

domain essential for transactivation and binding (Kim et al., 2012), that undergoes PTMs, including phosphorylation and acetylation (reviewed in (Meek and Anderson, 2009)).

TP53 is conserved across species, with the highest homology found in the DBD. Many animal models, including knock-in and knockout mice, have been developed to study TP53 function and its implication in tumorigenesis (reviewed in (Lozano, 2010)). For example, mice lacking *trp53* (*p53*) are viable and prone to spontaneous tumor development by 6 months of age, mainly sarcomas and lymphomas (Donehower et al., 1992). Similarly, three p53 zebrafish models ($tp53^{I166T}$ and $tp53^{N168K}$ in exon 6, $tp53^{M214K}$ in exon 7) generated via target-selected mutagenesis strategy, lack apoptosis and cell-cycle arrest responses to DNA damage. After 8.5 months of age, these animals spontaneously develop malignant peripheral nerve sheath tumors (MPNSTs) with an incidence of 25 % (Berghmans et al., 2005) and sarcomas with an incidence of 50 % (Parant et al., 2010).

APC (Adenomatous Poliposis Coli) is a tumor suppressor gene found mutated in the majority of sporadic colorectal cancers (80 %) (Cottrell et al., 1992; Miyoshi et al., 1992). Germline mutations are responsible for familial adenomatous polyposis (FAP) syndrome, characterized by the development of many colonic polyps early in the adulthood that can systematically evolve into malignant tumors (Bodmer et al., 1987; Kinzler et al., 1991; Nakamura et al., 1991). Unlike *TP53*, mutations in *APC*, both germline and somatic mutations, predominantly result in protein truncations (Cottrell et al., 1992; Miyoshi et al., 1992; Nagase and Nakamura, 1993).



© Adapted from Farnhead., Hum Mol Genetics 2001

Figure 5: Functional domains of APC protein.

APC full-length protein consists of 2843 amino acids and includes several domains, indicated here with different colors. For details see text.

The *APC* gene spans 107 kilobases (kb) and contains 21 exons. Many transcript variants are present due to alternative splicing, with the most common isoform consisting of 2843 amino acids encoded by exons 2 – 16 (Horii et al., 1993). Exon 16 includes more than 75 % of the coding sequence of *APC* and is the most frequently targeted exon for cancer mutations (containing the Mutation Cluster Region - MCR). The APC protein consists of numerous domains that allow its interaction with a variety of cytoplasmic proteins (Figure 5). The N-terminus is composed of an oligomerization domain, important for homodimerization (Su et al., 1993), and an armadillo region that allows the binding to its partners. The central region is made of multiple 15- or 20-amino acid repeats required for binding of β-catenin and contains the MCR (Rubinfeld et al., 1993). The C-terminus contains a basic domain and binding sites for two different proteins (EB1 and HDLG) (Fearnhead et al., 2001). APC is a crucial component of the β-catenin destruction complex that catalyzes the phosphorylation of β-catenin (Munemitsu et al., 1995). Moreover, APC has a broad spectrum of functions, including control of the WNT signal transduction pathway, cell adhesion and migration, apoptosis, and chromosomal segregation during mitosis (Fearnhead et al., 2001).

Mice carrying a homozygous deletion of *Apc* are embryonic lethal (Moser et al., 1995), but a number of mouse models have been created that mimic the human pathology, with the

development of multiple intestinal polyps (reviewed in (McCart et al., 2008)); e.g. N-ethyl-N-nitrosourea (ENU) induced Multiple intestinal neoplasia (Min) mouse, which carries a point mutation at codon 850 (Moser et al., 1990).

1.3.4 AID and cancer

As discussed in section 1.3.1.1, AID is the key protein for antibody diversification. A number of mechanisms are responsible for confining AID activity to the *Ig locus* in B cells, such as target specificity controlled by the interaction with several proteins (Pavri et al., 2010; Sayegh et al., 2003; Willmann et al., 2012), subcellular localization (Hu et al., 2013; Maeda et al., 2010), post-transcriptional regulation (de Yebenes et al., 2008; Teng et al., 2008; Tran et al., 2010), and PTMs (Basu et al., 2009). In addition to *Ig* genes, AID is able to target non-*Ig* genes in B cells, including oncogenes such as *MYC, PAX5, BCL-6* and *PIM1* (Pasqualucci et al., 1998; Pasqualucci et al., 2001; Shen et al., 1998; Wang et al., 2004). These genes normally escape acquisition of mutations due to the combined action of MMR and BER. However, in response to cellular stress including viral infection (Epeldegui et al., 2007; He et al., 2003), the actions of DNA repair mechanisms can be altered, leading to the accumulation of AID-dependent mutations and possible tumor induction/progression (Liu et al., 2008).

B cell malignancies can bear chromosomal translocations involving *Ig* switch regions and oncogenes as a result of aberrant CSR. In particular, in Burkitt's lymphoma, AID drives the chromosomal translocation between *IgH* and *C-MYC* (Ramiro et al., 2004; Robbiani et al., 2008). Furthermore, AID is expressed at different levels in all phenotypic subtypes of diffuse large B-cell lymphomas (DLBCL) (Pasqualucci et al., 2004).

Aside from B cell malignancies, AID can be genotoxic in other tissues, since it can be induced by pro-inflammatory cytokines (Endo et al., 2007; Endo et al., 2008; Komori et al., 2008), microbial infections (Kim et al., 2013; Matsumoto et al., 2007), and hormones

(Lin et al., 2009; Pauklin and Petersen-Mahrt, 2009; Pauklin et al., 2009). Inflammation of the colon (known as inflammatory Bowel disease) induces aberrant AID expression via NF-kB, frequently leading to colitis-associated cancer (Takai et al., 2012), while *Helicobacter pilori* infection triggers AID up regulation resulting in mutations in *TP53* and gastric cancer progression (Matsumoto et al., 2007; Shimizu et al., 2014). Furthermore, activation of AID by androgens in prostate cancer lead to chromosomal translocation (Lin et al., 2009), pointing the potential link between hormone exposure and AID/APOBEC mediated oncogenesis.

Initial bioinformatics analysis uncovered that dC to dT transitions occurring in *TP53* and *APC* tumor suppressor genes are embedded within DNA deaminases footprints (Beale et al., 2004; Morgan et al., 2004). The role of DNA deaminases as mutators in oncogenesis was supported by uncovering *TP53* mutations in AID footprints after *H. pilori* infection (Shimizu et al., 2014), and *TP53* mutations from breast cancer samples showing characteristic AID/APOBEC *Ig-like* mutation signature (Lindley, 2013). More recently, NGS approaches of human cancers have provided further support for AID and other DNA deaminases involvement in the generation of oncogenic mutations. First, C>T DNA deaminase mutational signature was identified in breast cancer sequence data (Burns et al., 2013a; Nik-Zainal et al., 2012), then it was extended to different type of tumors, where AID/APOBEC mutagenic activity was linked to clusters of unstable genomic regions called kataegis (Alexandrov et al., 2013; Burns et al., 2013b; Roberts et al., 2013; Zhang et al., 2015).

1.3.5 Animal models of cancer

Animal models have been instrumental in the study of genes involved in human cancer initiation and progression. Both spontaneous and carcinogen-induced malignancies have been studied in different species, from worms to drosophila, to zebrafish, to mouse and

dog. For over a century rodents and especially mice are the model of choice for several reasons: we share 95 % of the same genes and pathways; mice are inbreed, easy to be genetically manipulated, cheap to maintain, small in size, and they reproduce quickly (Cheon and Orsulic, 2011).

Although initially used as a classical developmental and embryological model, the identification of thousands of zebrafish mutants through genetic screens in the 1990s established zebrafish as a model for human diseases, including cancer. Despite the obvious differences in the physiology of fish and humans that can alter the phenotypic outcome of the disease in this model, zebrafish has a number of advantages over the use of mice: 1) the large number of offspring, 2) transparent embryos, 3) small size, 4) ease of daily upkeep, and 5) fertilization *ex utero* that allows easy genetic manipulation and continuous analysis of the developing embryo (Lieschke and Currie, 2007).

Animal models of cancer are very heterogeneous and include four main classes: 1) xenograft models with established human/mouse tumor cell lines; 2) primary tumorgraft models; 3) germline transgenic and conditional transgenic models (GEMs); and 4) carcinogen-induced models (reviewed in (Cheon and Orsulic, 2011; Ruggeri et al., 2014)). In the first class, tumor derived cell lines or tissue pieces are implanted into syngeneic or immune-compromised mice allowing a reproducible evaluation of pharmacological screening. The main disadvantages are the weak histological and phenotypic similarities to the primary tumor, loss of heterogeneity, low metastatic rates, and loss of the primary tumor microenvironment. The model can be improved, if the tumor cells are implanted into the organ of origin (orthotopic models), hence the correct microenvironment is established and a higher metastatic rate is observed (Cheon and Orsulic, 2011; Ruggeri et al., 2014).

Primary tumorgraft models, also referred to as *avatar*, involve the direct implantation of fresh tumor biopsies into immune-deficient mice, preserving genotypic and phenotypic

features of the original tumor (Kopetz et al., 2012). Weaknesses of this model are the limited engraftment rates, long latency for tumor development, high front-end costs, and reduced accessibility to fresh tumor samples (Ruggeri et al., 2014). In contrast to mouse, transplantation of human cancer cells into zebrafish embryos or irradiated transparent zebrafish (Casper fish (White et al., 2008)) allows to follow cancer progression in real-time. Importantly, this model also provides a means to monitor drug response of patient-derived biopsy in a cost-effective and time-efficient manner (Jung et al., 2012; Lee et al., 2005; Veinotte et al., 2014).

The third class includes animals in which a tumor suppressor gene or oncogene is expressed/inhibited ubiquitously or in a tissue-temporal specific manner, thus recapitulating how an important mutation induces oncogenesis. This can help in elucidating deregulated molecular mechanisms linked to the gene itself. Despite the genetic and histologic similarities to human cancers, the main limitations are a low and heterogeneous penetrance, heterogeneous genetic background, as well as other technical difficulties such as targeting of mouse embryonic stem (ES) cells by homologous recombination (Thomas and Capecchi, 1987), the generation of a germline transmitted chimera, and subsequent complex breeding (Cheon and Orsulic, 2011; Politi and Pao, 2011; Ruggeri et al., 2014). Transgenic zebrafish lines expressing human oncogenes ubiquitously or in a tissue-temporal specific manner have been generated using the Gal4/UAS systems (Scheer and Campos-Ortega, 1999) and the Tol2 mediated approach (Kawakami, 2007; Urasaki et al., 2006), with the advantage of following cancer progression in live animals. New genome editing techniques using engineered nucleases (see section 1.4), such as zinc-finger nuclease (ZFNs), transcription activator-like effector (TALEN) nuclease, and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system, provide us with alternative methods of generating GEMs, accelerating the generation time (Reviewed in (Gaj et al., 2013)).

Carcinogen-induced models recapitulate multistep tumor progression in response to environmental DNA damaging agents (tobacco smoke associated carcinogens) and tumor-promoting agents, such as and N-ethyl-N-nitrosourea (ENU). The phenotypic outcome is very similar to certain human cancers, and allows one to also investigate the contribution of inflammatory processes in tumor development (Cheon and Orsulic, 2011; Ruggeri et al., 2014). However, this approach has several limitations, such as the inability to make specific targeted mutations, prolonged time frames, and costs required for dose administration. In zebrafish, chemical mutagens (ENU), psoralen and radiation can be used to create stable mutant lines with knock down gene function, with the limitation of random mutations being inserted (Wienholds et al., 2003).

1.4 NOVEL APPROACHES FOR GENOME EDITING

New technologies have emerged that enable targeted editing of genomes in diverse systems. These approaches include the generation of different modular proteins in which the DNA binding domain is fused to an effector domain to alter genomic structure and function. The DNA recognition domain can be customized to recognize virtually any sequence and is typically based on the structure of natural DNA binding proteins, such as zinc finger proteins (ZFP) and transcription activator-like effectors (TALE). The combination of the DNA-binding protein with effector domains create functional enzymes including nucleases (Christian et al., 2010; Kim et al., 1996), transcription factors (Beerli et al., 1998; Zhang et al., 2011), and methyltransferases (Meister et al., 2010).

1.4.1 Zinc finger proteins

The Cys_2-His_2 zinc finger domain is among the most common DNA binding motif found in eukaryotes and consists of 30 amino acids in a ββα conformation, where the α-helix projects into the major groove of the DNA and recognizes 3 – 4 adjacent nucleotides (Pavletich and Pabo, 1991). This modular structure of ZFP has been extensively engineered through site-specific mutagenesis and the generation of large combinatorial libraries to create a unique zinc finger domain with specificity for all 64 possible trinucleotide combinations (Pabo et al., 2001). In addition, these preselected modules can be linked together in tandem to target DNA sequences of 9 – 18 nucleotides. For many years, ZFP technology has been the only approach available to create custom DNA binding proteins and enzymes (Kim et al., 2009; Mandell and Barbas, 2006). However, engineering ZFPs with high activity and specificity remains technically challenging, and requires high-throughput selection of large libraries to find the functional protein. This is predominantly due to difficulty in understanding the prediction of ZFP DNA-binding specificity and affinity (Maeder et al., 2008).

1.4.2 Transcription activator-like effectors

In 2009, the discovery of the modular DNA recognition code by TALE proteins (Boch et al., 2009; Moscou and Bogdanove, 2009) has led to the expansion of an alternative option for customable engineered DNA binding proteins. TALEs are naturally occurring DNA-binding proteins produced by plant pathogenic bacteria to regulate host gene expression. They bind effector-specific DNA sequences and transcriptionally activate the expression of genes that may contribute to bacterial colonization (Kay et al., 2007; Romer et al., 2007). TALEs contain N- and C-termini important for localization and activation respectively, and a central region for DNA recognition. The central domain comprises as many as 30 tandem repeats of a 33- to 35-amino acids sequence motif. This sequence is largely

conserved, with the exception of two adjacent amino acids at position 12 and 13, termed repeat variable diresidue (RVD), and it determines the nucleotide binding specificity of each TALE. Repeats with different RVDs recognize different DNA base pairs (bp) with one-to-one correspondence, constituting a straightforward cipher repeat, called TALE DNA binding code (Boch et al., 2009; Meckler et al., 2013; Moscou and Bogdanove, 2009). Importantly, each repeat consists of two α helices connected by a three-residue loop that contains the RVDs (RVD loop). Sequential repeats bind to one DNA strand with 5' to 3' directionality. Because of their modularity, TALE proteins have become an attractive tool for genome editing, with many protocols available for an easy and rapid assembly of custom engineered binding proteins.

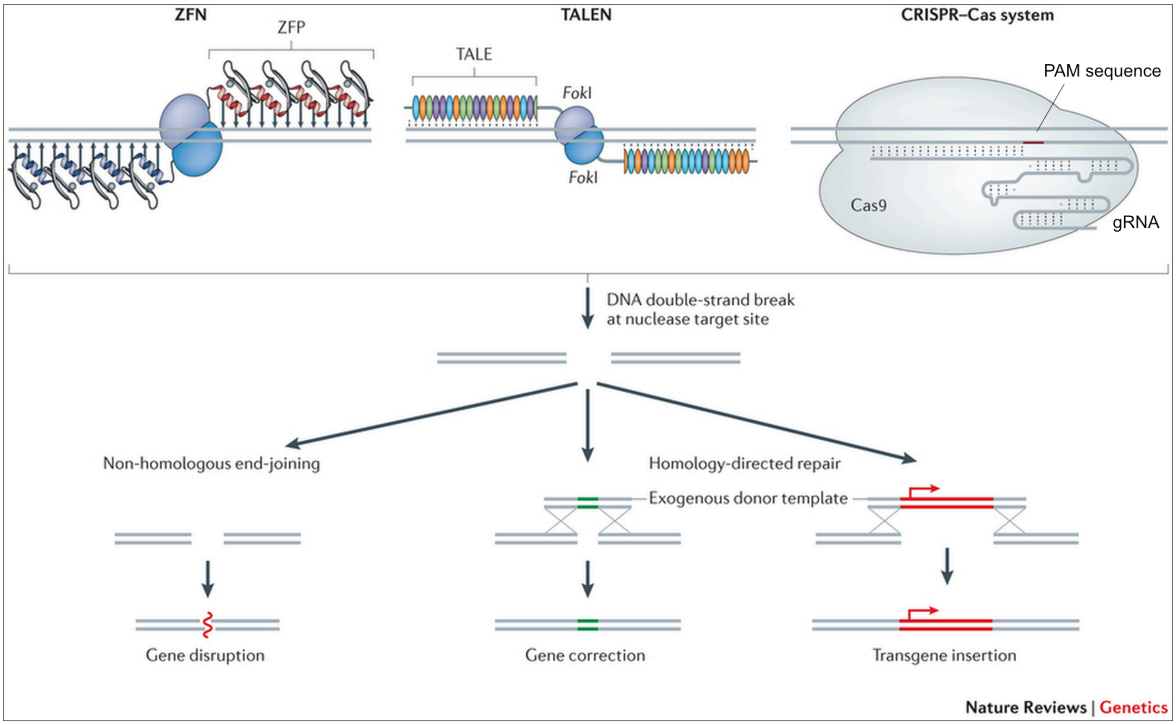
1.4.3 Clustered regularly interspaced short palindromic repeat-Cas system

Distinct from the site-specific targeting tool described above, the CRISPR-Cas system is a naturally occurring, adaptive microbial immune system for defense against foreign nucleic acids, such as phages and plasmids (Barrangou et al., 2007; Marraffini and Sontheimer, 2010). CRISPRs were initially found in the *E. coli* genome in 1987 (Ishino et al., 1987), and this family of DNA repeats is present in 90 % of archaeal and 40 % of bacterial genomes, with common evolutionary origin (Jansen et al., 2002; Sorek et al., 2008). The CRISPR loci are surrounded by the *Cas* genes, which encode a heterogeneous family of proteins with functional domains typical of nucleases, helicases, polymerases and polynucleotide-binding proteins (Haft et al., 2005; Jansen et al., 2002). When the CRISPR array is transcribed and processed into short CRISPR RNAs (crRNAs), the Cas proteins are able to utilize the crRNA molecules to target and destroy the foreign DNA (Garneau et al., 2010). In particular, type II CRISPR system recognizes and cleaves the target DNA via the RNA-guided endonuclease Cas9, that consists of two noncoding RNAs, the crRNA and the trans-activating crRNA (tracrRNA) (reviewed in (Makarova et al., 2015)). Here, each

crRNA hybridizes with the tracrRNA forming a complex with the Cas9 nuclease. The protospacer-encoded portion of the crRNA guides Cas9 to cleave complementary target-DNA sequences if adjacent to a short sequence - the protospacer adjacent motif (PAM). Cas9 protein includes two nuclease domains, a RuvC-like nuclease domain located at the N-terminus and a HNH-like nuclease domain located in the central region (Sapranauskas et al., 2011). Recently, the type II CRISPR system was programmed to cut various DNA sites *in vitro* (Jinek et al., 2012), and subsequently it was adapted for facilitating genome editing in eukaryotic cells and organisms (DiCarlo et al., 2013; Hwang et al., 2013; Mali et al., 2013).

1.4.4 Genome editing with site-specific nucleases

Historically, targeted gene inactivation, replacement or addition have been achieved by homologous recombination (Thomas and Capecchi, 1987). However, the low efficiency of recombination in model organisms, the technical difficulties as well as the high costs limit the use of this approach. With the discovery and the development of synthetic DNA binding proteins and the CRISPR/Cas9 system, new efficient and less expensive approaches emerged, and became the methods of choice for genome editing (Figure 6).



© Adapted from Yin., Nat Rev Genetics. 2014

Figure 6: Mechanisms of ZFN, TALEN and CRISPR/Cas9 systems.

ZFN, TALEN and CRISPR/Cas9 systems induce DSBs with high precision and affinity in the targeted genomic sequence. DSB can be repaired by NHEJ pathway resulting in gene disruption, or by HR if the donor template is provided, leading to gene correction or replacement.

DNA binding proteins, such as ZFP and TALE, can be fused to the catalytic domain of a restriction endonuclease to generate respectively zinc finger nucleases (ZFN) (Kim et al., 1996) and TALE nucleases (TALENs) (Cermak et al., 2011; Christian et al., 2010; Li et al., 2011) that create DSB at the targeted *locus*. The endonuclease *Fok* I - modified to increase its specificity and activity - acting as a dimer on two target sites, is the nuclease of choice. These DSB are usually resolved by NHEJ-mediated repair. Occasionally NHEJ introduces small insertions or deletions, resulting in the knockout of the gene function via frameshift mutations. Both ZFN and TALENs have been widely used for precise genome editing in diverse organisms, including plants (Christian et al., 2010), fish (Huang et al., 2011), frogs (Lei et al., 2012), flies (Liu et al., 2012), mice (Sung et al., 2013), as well as

critical mammalian cells, e.g. human ES cells and induced pluripotent stem (iPS) cells (Hockemeyer et al., 2009; Hockemeyer et al., 2011; Lombardo et al., 2007).

In contrast to ZFN and TALEN methods, which use protein-DNA interactions, Cas9 is RNA-guided and uses the base-pairs rules between the engineered RNA and the target DNA site. CRISPR/Cas9 requires two components to be introduced and expressed in the same cells: the Cas9 nuclease and a guide RNA (gRNA), consisting of crRNA fused to a fixed tracrRNA. Two nuclease domains of the Cas9 cut both DNA strands, generating DSBs, which are processed by NHEJ mechanisms as described for ZFN and TALENs. New advances in the Cas9 system were made, with the creation of a mutant form, known as Cas9D10A, harboring only nickase activity (Cong et al., 2013). On the other hand, the double mutant of Cas9, D10A and H840A (dCas9), was catalytically dead, creating a sequence specific DNA binding protein without any nuclease activity (Jinek et al., 2012).

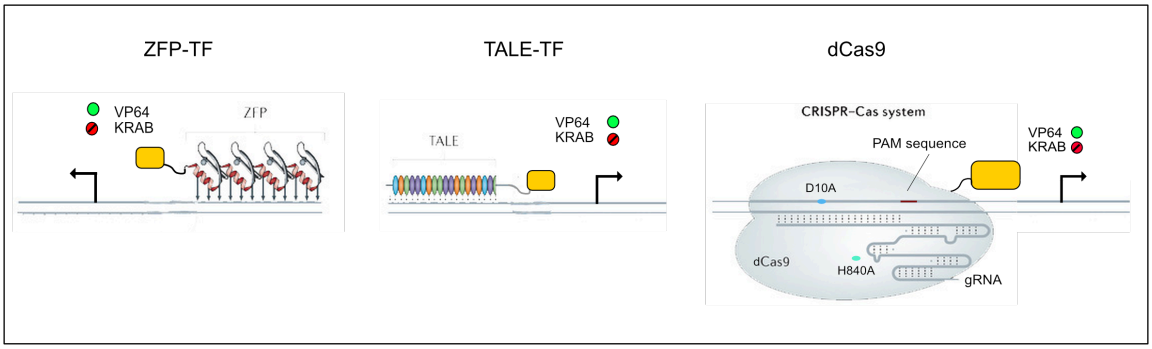
The simplicity of the CRISPR/Cas9 system led to the creation of large gRNA libraries using array-based oligonucleotide synthesis, engineered to target almost every gene in the cell or organism of choice, without the need of creating novel binding proteins (Chen et al., 2015; Wang et al., 2014).

1.4.5 Regulation of gene expression

Gene regulation in bacteria and yeast is driven by sequence specific DNA binding proteins. In mammals this specificity is lost and replaced with a pleiotropic set of complexes and molecular machineries. In the past, researchers have adapted the prokaryotic (e.g. lacO and Tet) and yeast (e.g. Gal4) regulatory system to study gene expression in eukaryotes. For example, the Tetracyclin-regulated (Tet) system is based on a chimeric transcription factor consisting of the bacterial Tet repressor (TetR) fused to an eukaryotic transcriptional regulatory domain, which activates (tTA) or silences (tTS) expression from a Tet-dependent promoter. The Tet-dependent promoter contains several TetR binding sites

(tetO), placed upstream of the transcription initiation site and the binding to TetO is regulated by tetracycline and its derivative (Berens and Hillen, 2003). The *Tet* systems have been used *in vivo* for the inducible expression of several transgenes, encoding for example reporter genes and oncogenes (Furth et al., 1994).

The yeast Gal4 system consists in the transcription regulator protein Gal4 that binds specifically to UAS (Upstream Activation Sequence) to activate transcription of target genes. This tool was widely used in fruit flies (Duffy, 2002), zebrafish (Scheer and Campos-Ortega, 1999) and mouse (Ornitz et al., 1991) to study gene functions *in vivo*. Two transgenic lines are needed, one expressing the Gal4 protein under a specific promoter, and the other carrying the UAS upstream of the gene of interest. To achieve a stronger induction of expression, the Gal4 protein was fused to transcriptional activation domains such as herpes simplex VP16 transactivation domain (VP16) (Koster and Fraser, 2001). Although these natural transcription factors continue to be used in specific applications, they are being replaced by the programmable DNA binding proteins ZFP, TALE and CRISPR/Cas9, that can be used to control gene regulation at targeted promoter or enhancer elements in eukaryotic systems (Beerli et al., 1998; Qi et al., 2013; Tan et al., 2003; Zhang et al., 2011) (Figure 7).



© Adapted from Yin., Nat Rev Genetics. 2014

Figure 7: Gene regulation using ZFP-TF, TALE-TF and dCas9-TF.

ZFP, TALE DNA binding domains, and dCas9 can be fused to transcription activator (VP64) or repressor (KRAB) domains and targeted to promoter regions to control gene regulation. The effector domain is depicted in yellow, while the green circle represents activation and red circle indicates repression of transcription.

For transcriptional activation, ZFP were fused to VP16 or its derivative, VP64 (Beerli et al., 1998), while TALEs had their endogenous NLS and acidic transcription activation domain replaced with a mammalian SV-40 NLS and VP64 (Zhang et al., 2011), or with the mammalian activator domains NF-kB (Maeder et al., 2013b). To achieve transcriptional repression, ZFP or TALEs were fused to the Krüppel-Associated Box (KRAB), ERF Repressor Domain (ERD), or mSIN3 Interaction Domain (SID) (Beerli et al., 1998; Cong et al., 2012; Mahfouz et al., 2012). In addition, TALE proteins can induce transcriptional repressors at their target site through steric hindrance of key transcriptional initiation elements (Li et al., 2015).

Because of its large size (190kD), the use of the dCas9 allowed for gene repression by sterically blocking RNA polymerase binding or elongation, both in bacteria and mammalian cells (Bikard et al., 2013; Qi et al., 2013). Alternatively, dCas9 can be fused to transcriptional activator (VP64) (Maeder et al., 2013a) or repressor domains (KRAB) (Gilbert et al., 2013), as already described for ZF and TALEs.

1.5 AIM

In somatic cells, DNA damage can be introduced both by endogenous and exogenous agents, leading to mutations and DNA recombination. The progressive accumulation of DNA lesions in a cell can generate genetic variation by natural selection, or can confer autonomous proliferation eventually leading to oncogenic transformation. AID, through its mutator activity, is functioning at the balance between DNA stability and instability. Mutations introduced at the *Ig locus* confer positive selection for antibodies, while mutations introduced at non-*Ig* genes contributes to aberrant genetic changes and cancer. Although animal models have contributed significantly to the understanding of oncogenesis, tumor development is more rapid and homogeneous in animal models as compared to the heterogeneity and slow progressive accumulation of mutations during human tumor progression. For these reasons, innovative models are needed to better understand mechanisms of cancer development.

The aim of my PhD thesis was to create a new model of neoplasia by inducing low-frequency mutagenesis in tumor suppressor gene *loci* over time, to mimic human cancer progression.

To this end, I used the natural DNA damaging protein AID as a tool to induce continuously mutations in *p53* and *Apc* tumor suppressor genes, both in cell lines and in laboratory animals (zebrafish and mouse). I used different approaches to target AID to the desired genomic *locus*, including TALE and Gal4 DNA binding proteins. TALE-AID fusion proteins were used to mutate at low frequency the endogenous *p53 locus* of mammalian cell lines (NIH3T3 and MEFs) and zebrafish, while Gal4-Aid was expressed in mice to target *p53* and *Apc* genes (Gal4-Aid/$p53^{UAS/+}$ and Gal4-Aid/$Apc^{UAS/+}$).

The working hypothesis is that by targeting AID to the genomic *loci* of tumor suppressors during the lifetime of an organism, the majority of the mutations will be repaired by the

DNA repair machinery, but some will escape these mechanisms, thus accumulating at the *locus*, leading to gene inactivation and cellular transformation.

CHAPTER 2 – MATERIALS AND METHODS

2.1 BUFFERS AND REAGENTS

Items with "*" were prepared by IFOM Kitchen, while items with "**" were prepared by IFOM Zebrafish Unit.

CALBIOCHEM® EDTA-free protease inhibitors *

Denaturing solution*: 1.5 M NaCl, 0.5 M NaOH.

Danieau solution:** 50 mM NaCl, 0.7 mM KCl, 0.4 mM $MgSO_4$, 0.6 mM $Ca(NO_3)_2$, 5 mM HEPES, pH 7.5.

E3 water:** 5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$, 10^{-5} % Methylene Blue (Sigma).

FACS buffer: PBS 1X + 2 % Fetal Bovine Serum South Americans (FBS SA, BioWest).

FACS buffer II: DMEM, 2 % FBS SA, 100 Units/mL penicillin (Microtech®) and 100 Units/mL streptomycin (Microtech®).

Hypotonic lysis buffer: 150 mM NH_4Cl, 10 mM $KHCO_3$, 100 μM EDTA in dH_2O (pH 7.3).

I.E.D: Ion-Exchange Decal Unit (Biocare Medical).

Laemmi buffer 5x: 10 % SDS, 50 % glycerol, 0.1 % bromophenol blue, 50 mM DTT, 150 mM Tris-HCl pH 6.8.

Lysis buffer: 0.1 M Tris-HCl pH 8.0, 0.5 mM EDTA, 200 mM NaCl. Freshly added 0.2 % SDS and 0.1 mg/mL of proteinase K (Promega).

MESAB 1X:** 0.4 % Tricaine (3-amino benzoic acid ethyl ester; Sigma), 1 % $Na_2HPO_4x2H_2O$.

MOPS 10X: 400 mM Morpholinopropanesulfonic acid (free acid), 100 mM Na-acetate-$3H_2O$, 10 mM EDTA, adjusts to pH 7.2.

PBS* 1X: 10 mM phosphate, 137 mM NaCl, and 2.7 mM KCl pH 7.0.

RIPA-02: 0.1 % NP40, 0.2 % DOC, 200 mM NaCl, 50 mM KCl, 10 mM Tris-HCl pH7.4, 0.5 mM EDTA, 5 mM β-Mercaptoethanol, freshly added 1X EDTA-free protease inhibitors.

Running Buffer*: 2.5 mM Tris-HCl p.H 8.6, 190 mM glycine, 0.1 % SDS.

Sorting buffer: PBS 1X, 33 % FBS SA, 300 Units/mL penicillin (Microtech®) and 300 Units/mL streptomycin (Microtech®).

Sorting buffer II: DMEM, 33 % FBS, 300 Units/mL penicillin (Microtech®) and 300 Units/mL streptomycin (Microtech®), 0.3 % gentamicin (50 mg/mL Life Tehcnologies).

TBE* 1X: 89 mM Tris-HCl pH8, 89 mM boric acid, 2 mM EDTA.

TBS* 1X: 50 mM Tris-HCl pH 7.5, 150 mM NaCl

TBST 1X: 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 % Tween 20.

TE* 1X: 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA.

Washing buffer: 0.1X SSC, 0.1 % SDS.

2.2 OLIGONUCLEOTIDES

Table 2: Oligonucleotides used in this thesis.

Name	Sequence 5' to 3'	Length	Orientation	Restriction site	Application
SPM431	CTAGTGACTACAAGGACCACGA TGGCGACTACAAGG	36	For	Spe I EcoRI	generation of 3X flag
SPM432	ATCACGACATCGATTACAAGGA CGATGACGACAAGG	37	For	Spe I EcoRI	generation of 3X flag
SPM433	AATTCACTTGTCGTCATCGTCCT TGTAATCGATGTCG	37	Rev	Spe I EcoRI	generation of 3X flag
SPM434	TGATCCTTGTAGTCGCCATCGT GGTCCTTGTAGTCA	36	Rev	Spe I EcoRI	generation of 3X flag
DMF1828	AGTGCGACATCATCATCGG	19	For	-	Gal4-Aid amplification
DMF1829	TCTCTCCTCTTTCACCACG	19	Rev	-	Gal4-Aid amplification
SP3018	TCCACCACCCTGTTGCTGTA	20	Rev	-	Mouse *Gapdh* mRNA
SP3033	GCACAGTCAAGGCCGAGAAT	20	For	-	Mouse *Gapdh* mRNA
SP3034	GCCTTCTCCATGGTGGTGAA	20	Rev	-	Mouse *Gapdh* mRNA
LS6917	AGCATGTGTGTCAGAGGTCAG	21	For	Kpn I	*p53* 5'arm (pLS7070)
LS6920	CACGTCGACCCTTAAGCAATCC TGCCTCA	29	Rev	Sal I	*p53* 5'arm (pLS7070) (p53 Br)
LS6921	CCTACTAGTCAGCCTGAGCATG GAAGTAAG	30	For	Spe I	*p53* 3'arm (pLS7070)
LS6924	CTATGCGGCCGCAGGCCCCACT TTCTTGAC	30	Rev	Not I	*p53* 3'arm (pLS7070)
LS6932	GTTGGGTGTCTGTAAATCCTG	21	For	-	Mouse *p53* gDNA (Bf)
LS6934	GAGGTGATGGCTGTGGATG	19	Rev	-	Mouse *p53* gDNA (Ar)
LS6945	CTTGTGCTGGTCCTTTTCTTG	21	For	-	Mouse *p53* gDNA (Ef)
LS6946	CAGGTTGCCAGGTCTTTTTC	20	For	-	*p53* ^{32}P 3' probe
LS6951	CCTTGGGTTACTTTTTCTTGC	20	Rev	-	*p53* ^{32}P 3' probe (p53 Er)
LS6988	TCGACGTCGGAGGACTGTCCTC CGAAGCGGAGGACTGTCCTCCG ACTTCGGAGGACTGTCCTCCGA TAC	69	For	Sal I	UAS sequence (pLS7070 and pLS7106)
LS6989	TCGAGTATCGGAGGACAGTCCT CCGAAGTCGGAGGACAGTCCTC CGCTTCGGAGGACAGTCCTCCG ACG	69	Rev	Xho I	UAS sequence (pLS7070 and pLS7106)
LS7011	CAAGTCGACGACACCA AGTCCAAAGCACAC	30	Rev	Sal I	*Apc* 5'arm (pLS7106)
LS7014	CAAAGCGGCCGCACTG AGACAGAGACCCCGTA	32	For	Not I	*Apc* 3'arm (pLS7106)
LS7016	CTTGAGCTCGTGGCTTT TTCCTTTGGGC	28	Rev	Sac I	*Apc* 3'arm (pLS7106)
LS7017	GTGGTACCCAACAAGT TATGGCGGAGTTC	29	For	Kpn I	*Apc* 5'arm (pLS7106)
LS7021	GGAGGCATGTGAAGAA ACCTTG	22	For	-	*Apc* ^{32}P 5' probe
LS7029	CACGGCACTAATTTCC TGGT	20	Rev	-	*Apc* ^{32}P 5' probe

LS7095	CGTCTTGGTTGGTTGGT TTT	20	For	-	*p53* ^{32}P 5' probe
LS7096	CGGCTACAATGGCACT TCTT	20	Rev	-	*p53* ^{32}P 5' probe
LS7124	CCGCTGCAGTGTACAGGCGCGC CATGGACAGCCTTCTGATGAAG	44	For	BsrGI	Mouse *Aid* mRNA
LS7125	ACACACTAGTAAATCCCAACAT ACGAAATGC	31	Rev	Spe I	Mouse *Aid* mRNA
LS7159	CGTGTACAACATGGACAGCCTC TTGATGAAC	31	For	BsrGI	Human *AID* mRNA
LS7160	GGTGATCGATTTTCAAAGTCCC AAAGTACGAAATG	35	Rev	Cla I	Human *AID* mRNA
LS7169	CCCAGAACCAAACTCAAGCC	20	For	-	*Apc* ^{32}P 3' probe
LS7170	AGTTGCTGCCCCTCTATCTG	20	Rev	-	*Apc* ^{32}P 3' probe
LS7251	CATCCTGGTCGAGCTGGACGGC GACGT	27	For	-	EGFP amplification
LS7252	GTACAGCTCGTCCATGCCGAGA GTGAT	27	Rev	-	EGFP amplification
LS7253	ACCTCATGAAGATCCTGACC	20	For	-	Zebrafish *Actin* mRNA
LS7254	TAATCCACATCTGCTGCTGG	20	Rev	-	Zebrafish *Actin* mRNA
LS7282	GCCACCACATCAAACCTGC	19	For	-	Zebrafish *p53* gDNA
LS7283	GGTCCTACAAAAAGGCTGTGA	21	Rev	-	Zebrafish *p53* gDNA
LS7295	CAGCCAAGTCTGTTATGTGCA	21	For	-	Mouse *p53* gDNA (Cf)
LS7296	CCTCCGTCATGTGCTGTGA	19	Rev	-	Mouse *p53* gDNA (Cr)
LS7318	GGATGGTGGTATACTCAGAGC	21	Rev	-	Mouse *p53* gDNA (p53 Dr)
LS7319	AGCGTGGTGGTACCTTATGAGC	22	For	-	Mouse *p53* gDNA (p53 Df)
LS7320	GCTATCAGGACATAGCGTTGG	21	For	-	Mouse *p53* gDNA
LS7328	ACAGTCGGATATCAGCCTCG	20	For	-	Mouse *p53* mRNA
LS7329	AGACTCCTCTGTAGCATGGG	20	Rev	-	Mouse *p53* mRNA
LS7341	CAGCCAAGAGTTCGCGGAG	19	For	-	Zebrafish *p53* mRNA
LS7342	TCGTCCTTCACCATCAGCTT	20	Rev	-	Zebrafish *p53* mRNA

2.3 PLASMIDS

Table 3: Plasmids created and used in this thesis.

ID	Insert	Application	Abbreviation
pLS7070	5'arm-3xUAS-Neo-3'arm	*p53*$^{UAS/+}$ transgenic mouse	
pLS7106	5'arm-3xUAS-Neo-3'arm	*Apc*$^{UAS/+}$ transgenic mouse	
pLS7121	TALEa-Aid-wt-FL	mouse *p53* targeting *in vitro*	Ta-Aid-wt
pLS7122	TALEc-Aid-wt-FL	mouse *p53* targeting *in vitro*	Tc-Aid-wt
pLS7123	TALEe-Aid-wt-FL	mouse *p53* targeting *in vitro*	Te-Aid-wt
pLS7154	TALEa-Aid-mt-FL	mouse *p53* targeting *in vitro*	Ta-Aid-mt
pLS7155	TALEc-Aid-mt-FL	mouse *p53* targeting *in vitro*	Tc-Aid-mt
pLS7156	TALEe-Aid-mt-FL	mouse *p53* targeting *in vitro*	Te-Aid-mt
pLS7127	EGFP-AIDwt	zebrafish *p53* targeting *in vivo*	EGFP-AIDwt
pLS7129	EGFP-TALEg-AIDwt	zebrafish *p53* targeting *in vivo*	EGFP-Tg-AIDwt
pLS7130	EGFP-TALEh-AIDwt	zebrafish *p53* targeting *in vivo*	EGFP-Th-AIDwt
pLS7133	EGFP-TALEg-AIDmt	zebrafish *p53* targeting *in vivo*	EGFP-Tg-AIDmt
pLS7134	EGFP-TALEh-AIDmt	zebrafish *p53* targeting *in vivo*	EGFP-Th-AIDmt
pCDNA3-EGFP	EGFP (Addgene)	mouse *p53* targeting *in vitro*	EGFP
pT2A200R150G	Tol2-EGFP-Tol2 (K. Kawakami)	zebrafish *p53* targeting *in vivo*	EGFP

Intermediates plasmids are not shown.

2.4 ANTIBODIES

Table 4: Antibodies used in this thesis.

Antibody	Manufacturer	Dilution
Mouse monoclonal anti-AID STR9	Bernardo Reina-S-Martin	1:10,000
Mouse monoclonal anti-Flag HRP conjugated	A8592 – Abcam	1:10,000
Mouse monoclonal anti-GFP	1181446001 – Roche	1:5,000
Mouse monoclonal anti-α tubulin	Sigma*	1:5,000
Mouse monoclonal anti-vinculin	Sigma*	1:10,000
Polyclonal goat anti mouse immunoglobulins HRP	Dako	1:10,000

2.5 CELL CULTURE

2.5.1 Cell thawing and freezing

A cryotube containing eukaryotic cells in freezing medium was thawed quickly in a water-bath at 37 ˚C. Cells were slowly resuspended in 10 mL of warm culture medium and transferred to a 15 mL Falcon Tube. Cells were collected by centrifugation at 1,200 revolutions per minute (rpm) for 5 minutes (min) and then the supernatant was removed. The cell pellet was resuspended in fresh medium in a volume fitting the desired cell density and the resuspension was transferred in a culture dish.

For freezing, growing cells were resuspended in culture medium and collected by centrifugation at 1,200 rpm for 5 min and the supernatant aspired. The pellet was gently resuspended in 1 mL of freezing medium (10 % DMSO in complete culture medium) and transferred to a cryotube, which was then placed into a Nalgene® cryogenic vial freezing Container containing isopropanol and transferred at – 80 ˚C for 72 hours (h). Cryotubes were then transferred to a liquid nitrogen tank.

2.5.2 Cell lines

Cell lines used in this thesis were purchased by IFOM Cell culture facility and were monitored daily for morphology and cell density. Every two days, adherent cells were treated with 0.05 % Trypsin / 0.02 % EDTA (Try/E, Microtech®) and resuspended in PBS containing 0.1 % Erythrosin B and cell number was counted in a hemocytometer. Cells were cultured in a humidified cell culture incubator at 37 ˚C with 5 % of CO_2.

NIH3T3 cells (Mouse embryonic fibroblasts, ATCC® CRL-1658™) at passage 131 were maintained in Dulbecco modified Eagle's medium high glucose (DMEM) (Lonza) supplemented with 10 % Calf Serum (CS) (US origin-BioWhittaker), 2 mM L-Glutamine

(Microtech®), 100 Units/mL Penicillin and 100 Units/mL Streptomycin (1 % P/S, Microtech®).

HEK 293T (Human embryonic kidney 293 with large T antigen) cells were maintained in DMEM (Lonza) supplemented with 10 % Fetal Bovine Serum South American (FBS SA) (Biowest), 2 mM L-Glutamine (Microtech®), 1 % P/S.

E14tg2α mouse ES cells were thawed at passage 24 on Mitomycin C-inactivated feeder MEFs (MMC – MEFs) and maintained in DMEM Knock out (Lonza) supplemented with 15 % FBS ES tested (Pan Biotech), 2 mM L-Glutamine (Microtech®), 1 mM Sodium-pyruvate (Euroclone®), 1X MEM non essential amino acids (Microtech®), homemade LIF (Cogentech S.c.a.r.l Transgenic Unit), 50 μM β-mercaptoethanol (Life Technologies).

E14tg2α ES cells and MMC-MEFs were provided by the Cogentech S.c.a.r.l Transgenic Unit.

2.5.3 Primary mouse embryonic fibroblasts generation

TRP53/C57 heterozygous mice (Jacks et al., 1994) were bred in the IFOM animal house and females were checked every day for presence of plug. Females being pregnant at day 12.5 or 13.5 were terminally anaesthetized with gradually increasing concentration of CO_2. The entire uterus was exported and placed in PBS 1X. Next steps were performed under the cell culture hood with sterile surgical instruments. The uterus was washed several times with PBS 1X and opened from the ovary edge in order to take one embryo at the time. To avoid contamination within different genotypes, each embryo was treated independently. The placenta was opened, the yolk sac discarded and the embryo was washed 2 times in PBS 1X supplemented with 1 % P/S. Fetal head and liver were collected separately and were used to perform genotyping of the embryo. The remaining embryonic parts were finely minced in PBS 1X mixed with Try/E and then placed at 37 °C for 10 min to further accelerate the process of dissociation. Try/E was inactivated by adding 2 volumes of

complete culture medium and the suspension was transferred into a 10-cm dish supplemented with 10 mL of complete culture medium and grown for 1 week at 37 ˚C in a humidified incubator. The resulting primary mouse embryonic fibroblasts were considered at passage 0 and named pMEFs p0. pMEFs were cultured in a humidified cell culture incubator at 37 ˚C with 5 % of CO_2 and 21 % of O_2 and were maintained in DMEM (Lonza) supplemented with 10 % FBS SA (Biowest), 2 mM L-Glutamine (Microtech®), 1 % P/S.

2.5.4 Transient transfections

NIH3T3 cells, seeded at a density of $1.1x10^6$ cells in a 10-cm dish, were transfected the day after with 10 μg of plasmid diluted in Opti-MEM® (Life Technologies) mixed with Lipofectamine2000® (Life Technologies) transfection reagent diluted in Opti-MEM®. The transfection was done following manufacturer's protocol.

Primary *p53*$^{+/-}$ MEFs at passage 4 were seeded at the density of $3x10^5$ cells/well in a 6-well plate and transfected the day after with 2 μg of plasmid DNA and FuGENE®6 (Promega) transfection reagents following manufacturer's instructions.

When two plasmids were used for co-transfection, the DNA were mixed in a ratio suitable for the subsequent experiment. For the sorting of GFP^+ cells, the reporter plasmid pCDNA3-EGFP was used at 1/10 of the concentration of the other plasmids (pLS7154 and pLS7121).

HEK 293T cells, seeded at the density of $2.5x10^5$ cells/well in a 6-well plate, were transfected with 2 μg of plasmid DNA using the Calcium phosphate based standard protocol for transfection.

2.5.5 Soft agar assay

Transfected NIH3T3 cells were counted after 24 h post transfection and transferred to 1.5 mL Eppendorf tube in order to reach the density of 10^3 cells in 0.9 mL of DMEM supplemented with 20 % FBS SA. Each tube contained the cell number optimized for one well of a 6-well plate. Bacto™ Agar (DIFCO™ BD) solution 3 % (stored at 60 ˚C to avoid solidification) was mixed to DMEM + FBS SA (stored at 45 ˚C) to reach a final concentration of 0.6 % agar and 20 % FBS SA. 2 mL of this solution were placed in each well of a 6-well plate. Once solidified, 0.9 mL of cell suspensions were mixed very rapidly with 100 μL of agar 3 % and placed on top of the 0.6 % agar layer. The 6-well plate was moved carefully to a humidified incubator at 37 ˚C and cells were cultured at 37 ˚C for 4 weeks.

2.5.6 Electroporation of ES cells with pLS7070 or pLS7106

E14tg2α ES cells at passage 25 were collected in PBS 1X. 10^7 cells / electroporation were resuspended in 400 μL of PBS 1X. Linearized plasmid DNA pLS7070 or pLS7106 was resuspended in 100 μL of PBS 1X and added to the ES cell suspensions by mixing well and avoiding bubbles formation. For each electroporation, the mix was transferred to a electroporation cuvette 4 mm (Bio-Rad) and electroporation was performed by Cogentech S.c.a.r.l Transgenic Unit with Gene Pulser Xcell™ Electroporation System (Bio-Rad) following the conditions listed in Table 5.

Table 5: Electroporation conditions used for ES cells.

Input Voltage	230 V
Output Voltage	227 V
Capacitance	500 μF
Resistance	∞
Time	7.9 msec

Cell suspensions were incubated at room temperature (RT) for 5 min and then transferred to a 50 ml Falcon Tube containing complete culture medium. Electroporated ES cells suspension was spread on pre-seeded MMC - MEFs in a 10-cm dish and maintained in ES completed medium at 37 ˚C. After 24 h (day 1) the medium was supplemented with 175 μg/mL G418 (Life Technologies) to allow the selection of neomycin-resistant cells. Once all the cells from the negative control (electroporation without DNA) died upon G418 treatment (day 7), G418-resistant colonies were picked and moved to one well per colony in a 96-well plate. Selected colonies were expanded and cells were collected for DNA extraction and southern blot analysis of the recombined allele (see sections 2.6.2.1 and 2.6.6). The first generation of clones in the 96-well plates were frozen as follow: each well was washed once with PBS 1X, filled with 30 μL of 0.25 % Try/E and incubated at 37 ˚C for 5 min. Try/E was inactivated by adding 70 μL of complete medium supplemented with 175 μg/mL G418 and 1 % P/S. 100 μL of 2X freezing solution (20 % DMSO, 80 % FBS ES tested) was added by gentle mixing. The 96-well plate was covered with aluminum foil, placed on dry ice and stored at -80 ˚C.

2.5.7 Cytogenetic analysis of ES cells selected clones for transgenic mouse generation

Cytogenetic analysis of ES cells with the correctly homologous recombined allele was performed by IFOM Cell Biology Unit with Giemsa staining following standard protocols.

2.6 MOLECULAR BIOLOGY

2.6.1 Cloning

2.6.1.1 Mouse TALE-Aid-FL vectors

Vectors pLS7154, pLS7155, pLS7156 and pLS7121, pLS7122, pLS7123 are based on the pBK-CMV vector. Mouse Aid wild type (wt) or Aid mutant (mt), carrying a point mutation in the catalytic domain (E58Q) complement DNA (cDNA) were amplified by polymerase chain reaction (PCR) (oligos: LS7124 - LS7125), using vectors containing either Aid-wt or Aid-mt (pDMF7006, pDMF7009) (see section 2.6.1.4). Three tandem flag sequences were generated annealing two complementary oligos (SPM431, SPM432, SPM433, SPM434) and further ligation of the paired oligos and inserted into pBK-CMV vector (resulting construct: pLS7107). pLS7107 served as the backbone for the cloning of the TALE-Aid fusion cDNA expression plasmids as follow: initially three TALE containing vectors (TALEa, TALEc, TALEe) were designed using the online tool TALE-NT (Cermak et al., 2011; Doyle et al., 2012) and created using the Golden gate kit (Addgene) (Cermak et al., 2011) in collaboration with Dr. Riccardo Pecori (Dr. Silvestro Conticello Lab. Istituto Toscano Tumori, Italy), to recognize three different *loci* within intron 4 of mouse *p53* (Table 6). For the recognition of the nucleotides A, T, G and C, the repeats NI, NN, NG and HD were used, respectively. Each TALE insert was ligated into pLS7107 and the resulting plasmids were used for insertion of the mouse Aid cDNA, and led to the vector construction listed in Table 3.

2.6.1.2 Zebrafish EGFP-TALE-AID vectors

Vectors pLS7129, pLS7130 and pLS7133, pLS7134 containing EGFP-TALE-AID inserts, were generated modifying the Tol2 vector pT2AL200R150G (Urasaki et al., 2006), which contains a EGFP cassette driven by an EF1α promoter, flanked by two Tol2 sites. Human

AID wt or AID mt (E58Q) cDNA were amplified by PCR (LS7159 - LS7160), using vectors containing either AID wt or AID mt cDNA (pKMS7010, pKMS7011) (see section 2.6.1.4) and inserted into pT2AL200R150G, resulting in pLS7127 and pLS7131 respectively. Two TALE containing vectors (TALEg, TALEh) were engineered as described above. The TALEs cDNA were designed to recognize two different *loci* within intron 6 of zebrafish *p53* (Table 6). TALEs fragments were inserted into pLS7127 and pLS7131 to create the final plasmids listed in Table 3.

Table 6: TALE design.

TALE ID	Species	Target sequence	Strand	Length (nt)	RVD sequence
TALEa	Mouse	T GTCCCTCCCACTTTG GCCT	Minus	19	NN NG HD HD HD NG HD HD HD NI HD NG NG NG NN NN HD HD NG
TALEc	Mouse	T CCCAGTCCTCTCTTTG CT	Plus	18	HD HD HD NI NN NG HD HD NG HD NG HD NG NG NG NN HD NG
TALEe	Mouse	T GGAAGTAAGACCCCT TCT	Plus	18	NN NN NI NI NN NG NI NI NN NI HD HD HD HD NG NG HD NG
TALEg	Zebrafish	T ACGGATAGTGTAACT CAAT	Plus	19	NI HD NN NN NI NG NI NN NG NN NG NI NI HD NG HD NI NI NG
TALEh	Zebrafish	T GCAACTATACTTCAC TCAT	Minus	19	NN HD NI NI HD NG NI NG NI HD NG NG HD NI HD NG HD NI NG

2.6.1.3 *p53*^{UAS} and *Apc*^{UAS} targeting vectors

A targeting vector containing a neomycin cassette flanked by two *loxP* sites was used as a backbone (pBML3) to create vectors pLS7070 and pLS7106 (Table 3). Three UAS sites were made by annealing sense (LS6988) and anti-sense (LS6989) UAS encoding oligos, and ligated into bML3 vector (pLS7063). For *p53* targeting, 5' homology arm of 3.9 kb and 3' homology arm of 3.5 kb were amplified by PCR using gDNA from E14tg2α ES cells as template (5' arm: LS6917 - LS6920; 3' arm: LS6921 - LS7924) and TaKaRa LA

PCR kit for amplification (TaKaRa Clontech) with standard amplification program (see section 2.6.1.4). PCR products of 5' and 3' arms were ligated into pLS7063 sequentially, generating the final vector pLS7070. For *Apc* targeting, 5' homology arm of 4.3 kb and 3' homology arm of 3.7 kb were amplified following the same protocol described above for *p53* (5' arm: LS7017 - LS7011; 3' arm: LS7014 – LS7016) and ligated sequentially into pLS7063 to generate the final vector pLS7106.

2.6.1.4 PCR amplification

a) TALE-AID plasmids

PCR from template plasmids (pDMF7006, pDMF7009, pKMS7010, pKMS7011) was performed in a total volume of 50 μL as follows: 50 ng of plasmid DNA were amplified in a reaction buffer composed of 1X Pfu Buffer, 0.25 mM dNTPs (Promega), 0.25 pmol/μL of each oligo and 1.25 Units of Pfu Polymerase (Promega). The amplification program consisted in: initial denaturation at 95 ˚C for 2 min, 15 cycles of 95 ˚C for 30 seconds (sec), 57 ˚C for 30 sec and 72 ˚C for 30 sec, and final extension step at 72 ˚C for 5 min. PCR was carried out using the thermal cycler GeneAmp® PCR system 9700 from Applied Biosystems®.

b) *p53*^{UAS} and *Apc*^{UAS} targeting

PCR using ES cell genomic DNA as template was performed in a total volume of 50 μL. For each reaction 500 ng of template DNA was diluted in a reaction buffer composed of 1X LA PCR Buffer II (TaKaRa Clontech), 0.25 mM dNTPs (Promega), 0.2 pmol/μL of each oligo and 2.5 Units of TaKaRa La Taq polymerase (TaKaRa Clontech). Amplification program: initial denaturation at 94 ˚C for 1 min, 10 cycles of: denaturation at 98 ˚C for 10 sec, annealing with decreasing temperature (touchdown from 64 ˚C down by 1 ˚C per cycle) and amplification at 68 ˚C for 6 min. Then 25 cycles at 98 ˚C for 10 sec,

gradient from 55 ˚C to 59 ˚C for 1 min and 68 ˚C for 6 min was applied. PCR was carried out using an Eppendorf Mastercycler gradient PCR machine.

2.6.1.5 Gel purification

For purification, PCR products were loaded onto agarose gels and after electrophoresis, corresponding bands were excised and purified with Wizard® SV Gel and PCR Clean-Up System (Promega), according to the manufacturer's instructions. Elution from columns was performed in 50 μL of ddH_2O.

2.6.1.6 Fragment preparation by enzymatic digestion and gel purification

Samples were digested in a total reaction volume of 30 μL, by diluting 1 μg of template plasmid DNA in a restriction enzyme compatible 1X NEB buffer and 10 Units of restriction enzyme (NEB: NEW ENGLAND Biolabs® Inc). Digestions were incubated at the enzyme recommended temperature for 3 h. After incubation the enzymes were heat inactivated at 65 – 80 ˚C for 20 min and the DNA was treated with 0.5 Units of Calf Alkaline Phosphatase (CIP) from NEB at 37 ˚C for 1 hour (h). Digested fragments were separated by agarose gel electrophoresis and purified as described above.

2.6.1.7 Single strand oligo annealing

Single strand oligonucleotides were resuspended in dH_2O to reach a final concentration of 100 pM. Each oligo was subjected to phosphorylation reactions. 100 pmol of each oligo were diluted in 1X T4 DNA ligase buffer (NEB) and 10 Units of T4 polynucleotide kinase (PNK) (NEB) in 20 μL total volume. Reactions were incubated at 37 ˚C for 1 h, complementary phosphorylated oligos were mixed (1:1) and 10 mM Tris-HCl pH 8 solution was added to reach a final volume of 100 μL. Oligos were annealed as follows:

denaturation: incubation of the mix in a heat block at 95 ˚C for 1 min; annealing: the oligo mix still placed into the heat block was allowed to cool down slowly to RT by placing the heat block on the lab bench after denaturation.

2.6.1.8 DNA ligation

Ligation reaction was performed in 20 μL as follows: 2 μL of 10X T4 DNA ligase buffer (NEB), approximately 50 ng of digested backbone vector and DNA insert in molar excess of 3 – 5 times compared to backbone. 400 Units of T4 DNA ligase (NEB) were added to the reaction and samples were incubated for 1 h at RT. Alternatively, for blunt end ligation, the ligation reaction was performed at 16 ˚C overnight (o/n).

2.6.1.9 Bacterial transformation and Midi preparation of plasmid DNA

E. coli transformation was done as follow: 5 μL of ligation product (plasmids) was gently mixed with 50 μL of One Shot® TOP10 Chemically (Life Technologies) or Subcloning Efficiency™ DH5α™ Competent *E. coli* and transformations were performed following the manufacturer's instructions.

For a Midi-preparation of plasmid DNA 100 mL of LB medium containing antibiotics was inoculated with a single *E. coli* colony and bacteria were grown for 12 h at 37 ˚C upon shaking. Bacterial cells were harvested by centrifugation of the suspension culture at 5,000 rpm for 15 min at RT in a Beckman JA-14 rotor (Beckman Coulter, Inc), cells were resuspended and DNA was extracted and purified using PureYeld™ Plasmid Midiprep System (Promega) following manufacturer's instructions.

2.6.2 Nucleic acid extractions and amplification

2.6.2.1 DNA extraction

a) ES cells

Complete culture medium from ES cell clones grown in 96-well plates (after electroporation with pLS7070 and pLS7106 plasmids) was removed and cells were covered and incubated with 50 μL of Lysis buffer for 3 h in a humidified chamber. Plates were transferred to RT and incubated for 1 h, and then NaCl/EtOH precipitation was performed. Briefly, 1.5 μL of NaCl 5 M and 100 μL of cold EtOH 100 % were added to each well and the 96-well plate was incubated for 2 – 3 h at RT upon shaking. The supernatant was discarded inverting carefully the plate and each well was washed two times with 100 μL of EtOH 70 %. After centrifugation at 1,200 rpm for 5 min, the supernatant was discarded and the resulting DNA pellet, upon air-drying for 20 min, was resuspended in 30 μL of Digestion mix (1X restriction enzyme buffer, 1X BSA, 1 mM spermidine, NEB restriction enzyme 1 Unit/μg of DNA) and incubated o/n at 37 ˚C for subsequent southern blot analysis (see section 2.6.6).

b) Zebrafish caudal fin clip, mouse tail biopsy, mouse embryo tissues

Zebrafish caudal fin biopsies or mouse tissue biopsies were digested in 500 μL of lysis buffer at 56 ˚C in Thermomixer compact (Eppendorf) o/n upon vigorously shaking. DNA was precipitated adding an equal volume of isopropanol and collected by centrifugation at 13,000 rpm for 10 min. Supernatant was removed and the DNA pellet, after air-drying for few minutes, was resuspended in TE 1X.

c) pMEFs

Transiently transfected pMEFs clones expanded to a 6-well plate were treated with 500 μL Try/E and placed at 37 ˚C for 5 min to allow the process of dissociation. Try/E was inactivated adding 9 volumes of complete culture medium and the suspension was harvested by centrifugation at 1,200 rpm at 4˚C for 5 min. The cell pellet was digested in

500 μL of Lysis buffer at 56 ˚C as described above. The DNA was extracted with Phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma-Aldrich) following standard protocols. The resulting pellet DNA was resuspended in 100 μL of TE 1X.

2.6.2.2 RNA extraction

a) RNA extraction of GFP^+ sorted NIH3T3 cells and zebrafish larvae for Ion Torrent™ analysis

After FACS sorting (see section 2.8.1), 6 – 8x10^5 GFP^+ NIH3T3 cells and mock control sample were collected by centrifugation at 1,200 rpm at 4 ˚C for 5 min. The supernatant was discarded and RNA from cells was isolated using All Prep DNA/RNA Mini kit (QIAGEN) following manufacturer's instructions, with the centrifugation steps performed at 13,000 rpm for 1 min at RT. RNA was eluted in 30 μL of H_2O and quantified by microvolume UV spectrophotometer (NanoDrop® ND-1000).

20 injected zebrafish embryos at 24 hours post fertilization (hpf) were lysed and homogenized in RLT plus buffer (QIAGEN), by passing at least 10 times through a 1 mL RNAse-free syringe and RNA was isolated using All Prep DNA/RNA mini Kit (QIAGEN) following manufacturer's protocol and centrifugation steps as described above.

b) Others (zebrafish larva, mouse organs and pMEFs)

Total RNA from one zebrafish larva at 5 days post fertilization (dpf), mouse organs or 6x10^5 of pMEFs cells were isolated using TRIzol® reagent (Life Technologies), following manufacturer's instructions. Where necessary, homogenization of the sample was facilitated by passing the lysate at least 5 times through a 1 mL RNAse-free syringe. The RNA pellet was resuspended in 30 μL of 10 mM Tris-HCl pH 7.4 and incubated 20 min at 55 ˚C on a Thermomixer compact (Eppendorf) prior to DNAse I treatment. Each sample was mixed with 1X DNAse I buffer (Roche), RNasin® 60 Units (Promega), 4.5 Units DNAse I (Roche) and incubated for 30 min at 37 ˚C in a Thermomixer compact. After

DNAse I digestion samples were purified with TRIzol® reagent according to manufacturers protocol, and RNA was dissolved in 30 μL of 10 mM Tris-HCl pH 7.4. RNA amount was measured with microvolume UV spectrophotometer (NanoDrop® ND-1000) and RNA quality was assessed by agarose gel electrophoresis (see section 2.6.4).

2.6.2.3 cDNA synthesis

The cDNA was obtained by reverse transcription (RT) of RNA using the reverse transcriptase SuperScript II® kit (Life Technologies) and oligo dT (Life Technologies). RNA for RT+ and RT- reactions (1 μg or 500 ng of RNA each) was added to a mastermix of 1X Buffer FS (Life Technologies), 5 mM dithiothreitol (DTT) (Life Technologies), 8 Units RNAsin® (Promega), 0.4 mM of each dNTPs, 2.5 μM oligo dT and distilled water DNAse/RNAse free (Life Technologies). 200 Units of SuperScript II® reverse transcriptase were added to RT+ samples, while the equivalent amount of ddH_2O was added to the RT- samples. cDNA synthesis was performed in the thermal cycler GeneAmp® PCR system 9700 (Applied Biosystems®) according to the following protocol: 37 ˚C for 10 min, 50 ˚C for 40 min and 55 ˚C for 10 min. cDNA was used for subsequent PCR reactions (RT-PCR).

2.6.3 PCR reactions (genotyping, gDNA analysis, RT-PCR and qRT-PCR)

a) Genotyping of adult mice and embryo, adult zebrafish and larvae, and gDNA analysis of pMEFs

Genotyping of animals and gDNA analysis of *p53* on pMEFs was performed by PCR using gDNA extracted from tissue biopsies or cells as template. PCR were performed using the GoTaq G2 kit (Promega). Each PCR reaction was assembled in a total volume of 25 μL: 100 ng of DNA were amplified in a reaction mix composed of 1X GoTaq Buffer, 0.2 mM

dNTPs, 0.2 pmol/μL of each oligo and 0.5 Units of GoTaq G2 polymerase. PCR reaction was performed in the thermal cycler GeneAmp® PCR system 9700 (Applied Biosystems®) and the amplification programs and oligos used are listed in Table 7 according to the species and the gene amplified.

Table 7: PCR conditions and oligos used for genotyping of animals gDNA and analysis of pMEFs

	Gal4-Aid (transgenic mouse)	**Human AID (adult zebrafish and larval embryos)**	***p53* (*p53*$^{+/-}$ mouse)**
Oligos	DMF1828 – DMF 1829	LS7159 – LS7160	LS7318 – LS7139 – LS7320
PCR program	95 °C 2 min 95 °C 45 sec ǀ 56 °C 45 sec ǀ x 32 cycles 72 °C 30 sec ǀ 72 °C 5 min	95 °C 2 min 95 °C 30 sec ǀ 60 °C 30 sec ǀ x 30 cycles 72 °C 1 min ǀ 72 °C 5 min	95 °C 2 min 95 °C 30 sec ǀ 55 °C 30 sec ǀ x 35 cycles 72 °C 1 min ǀ 72 °C 5 min
	***p53* (adult zebrafish and larval embryos)**		
Oligos		LS7282 – LS7283	
PCR program		95 °C 2 min 95 °C 30 sec ǀ 60 °C 30 sec ǀ x 30 cycles 72 °C 1 min ǀ 72 °C 5 min	
	***p53* for pMEFs analysis**		
Oligos	LS7328 – LS6934	LS6932 – LS6920	LS7295 – LS7296
PCR program	95 °C 2 min 95 °C 30 sec ǀ 55.5 °C 30 sec ǀ x 35 cycles 72 °C 40 sec ǀ 72 °C 5 min	95 °C 2 min 95 °C 30 sec ǀ 55.5 °C 30 sec ǀ x 35 cycles 72 °C 1 min ǀ 72 °C 5 min	95 °C 2 min 95 °C 30 sec ǀ 55.5 °C 30 sec ǀ x 35 cycles 72 °C 1 min ǀ 72 °C 5 min
Oligos	LS7319- LS7318	LS6945 – LS6951	
PCR program	95 °C 2 min 95 °C 30 sec ǀ 55.5 °C 30 sec ǀ x 35 cycles 72 °C 1 min ǀ 72 °C 5 min	95 °C 2 min 95 °C 30 sec ǀ 55.5 °C 30 sec ǀ x 35 cycles 72 °C 1 min ǀ 72 °C 5 min	

b) RT-PCR

Each RT-PCR reaction was performed using the cDNA retro-transcribed from total RNA as template. The reaction was performed in a volume of 50 μL as follow: 100 ng of cDNA were diluted in a reaction buffer composed of 1X Pfu Buffer, 0.25 mM dNTPs (Promega), 0.25 pmol/μL of each primer and 1.25 Units of Pfu Polymerase (Promega). RT-PCR reaction was performed in the thermal cycler GeneAmp® PCR system 9700 (Applied Biosystems®) and the amplification programs and oligos used are listed in Table 8 according to the species and the gene amplified.

Table 8: PCR conditions and oligos used for mice and zebrafish genes expression.

	***p53* (mouse NGS analysis)**	***Gapdh* (mouse NGS analysis)**	**Gal4-AID (tissue expression)**	***Gapdh* (tissue expression)**
Oligos	LS7328 – LS7329	SP3033 - SP3018	DMF1828 – DMF 1829	SP3033 – SP3034
PCR program	95 °C 2 min 95 °C 30 sec ⌋ 57 °C 30 sec ⌋ x 30 cycles 72 °C 1 min ⌋ 72 °C 5 min	95 °C 2 min 95 °C 30 sec ⌋ 55 °C 30 sec ⌋ x 30 cycles 72 °C 30 sec ⌋ 72 °C 5 min	95 °C 2 min 95 °C 45 sec ⌋ 56 °C 45 sec ⌋ x 30 cycles 72 °C 30 sec ⌋ 72 °C 5 min	95 °C 2 min 95 °C 30 sec ⌋ 55 °C 30 sec ⌋ x 30 cycles 72 °C 30 sec ⌋ 72 °C 5 min
	Human AID (zebrafish)	**EGFP (zebrafish tissue expression)**	***β-actin* (NGS analysis and tissue expression)**	***p53* (NGS analysis)**
Oligos	Same as Table 7	LS7251 – LS7252	LS7253 – LS7254	LS7341 – LS7342
PCR program		95 °C 2 min 95 °C 30 sec ⌋ 60 °C 30 sec ⌋ x 30 cycles 72 °C 1 min ⌋ 72 °C 5 min	95 °C 2 min 95 °C 30 sec ⌋ 58 °C 40 sec ⌋ x 30 cycles 72 °C 30 sec ⌋ 72 °C 5 min	95 °C 2 min 95 °C 30 sec ⌋ 55 °C 30 sec ⌋ x 30 cycles 72 °C 1 min ⌋ 72 °C 5 min

c) Quantitative Real-Time PCR (qRT-PCR)

Quantitative Real-Time PCR reactions were performed in triplicates in a final volume of 10 μl per reaction in 96 well plates (Euroclone®). 50 ng of cDNA template were diluted in a reaction buffer composed of 5 μl of GoTaq qPCR Master Mix (Promega) and 0.2 pmol/μL of each primer. Standard curves were prepared by making serial dilutions of cDNA from mock transfected cells. Amplification programs and oligos used are listed in

Table 9. The reaction was monitored in the LightCycler® 480 Real-Time PCR System (Roche) and the data were acquired using the LightCycler® 480 software version 1.5 (Roche). Gene expression levels were normalized to *Gapdh* mRNA. Transcript fold enrichment was calculated by $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Table 9: PCR conditions and oligos used for qRT-PCR.

	p53	*Gapdh*
Oligos	LS7295-7296	SP3033-3034
qRT-PCR program	95 °C 10 min 95 °C 10 sec | 62 °C 10 sec | x 45 cycles 72 °C 15 sec | 95°C 5 sec 60 – 97 °C 1.2°C/min	

2.6.4 Agarose gel electrophoresis

Separation of PCR fragments by size was performed by agarose gel electrophoresis in TBE 1X buffer. The agarose gel was obtained by dissolving agarose (0.7 -1.5 %) (GellyPhor^{LE}, Euroclone®) in TBE 1X and 1X GelRed™ nucleic acid gel stain (Biotium). Size of fragments was determined by a 100 bp or 1 kb DNA ladder (Promega) and visualized using Molecular Imager® ChemiDoc™ XRS+ Imaging system (Bio-Rad).

Quality of isolated RNA was monitored by agarose gel electrophoresis in MOPS 1X buffer. An aliquot of RNA was mixed 1:1 with 2X RNA loading dye (50 % Formamide, 20 % formaldehyde, MOPS 2X, 0.05 mg/mL ethidium-bromide) and denatured at 65 °C for 15 min. Reactions were quenched by placing the sample on ice for 5 min and then RNA samples were loaded on a denaturing 1 % agarose MOPS gel.

2.6.5 DNA sequencing

2.6.5.1 Sanger sequencing

DNA Sanger sequencing reactions were performed by Cogentech S.c.a.r.l Sequencing Unit. The dRhodamine Dye Terminator Ready Reaction Cycle Sequencing Kit (Thermofisher) was used and the samples were processed on a 16 capillary 3130xl Genetic Analyzers (Applied Biosystem).

2.6.5.2 Next-generation sequencing (NGS)

Ion Torrent™ NGS on PCR fragments was performed by the IFOM Genomic Unit.

a) Ion Proton™ System

Libraries for *p53* and *Gapdh* cDNA amplicons derived from GFP^+ NIH3T3 cells or *p53* and *β-actin* amplicons derived from 20 zebrafish embryos were prepared following the manufacturer protocols for genome sequencing with the Ion Proton™ sequencer (Life Technologies). Briefly ~ 480 - 680 ng of mouse cDNA and ~ 450 – 3000 ng of zebrafish cDNA were mechanically sheared to ~ 200 bp with Covaris S220/E220 Focused-ultrasonicator (Covaris, Inc) for 6 min, end-repaired and adapter-ligated using the NEBNext® Fast DNA Library Prep Set for Ion Torrent™ (NEW ENGLAND Biolabs® Inc) and adapter barcode Ion Xpress Barcode Adaptors 1-16 (Life Technologies). After adapter ligation, each sample was size-selected using AMPure XP Bead (Beckman Coulter, Inc), and enriched by PCR reaction. A SPRI cleanup with a 1.5X Bead:DNA ratio was performed post amplification and final libraries were eluted in 25 μL. Libraries were quantified and checked for size on an Agilent Bioanalyzer with HS DNA kit (Agilent, Santa Clara, CA). Each size selected library was diluted to a final concentration of 11 pM and clonally amplified using the Ion Proton™ 200™ Template Kit v3 (Life Technologies). After emulsion PCR, DNA positive Ion Sphere Particles (ISPs) were recovered and

enriched according to standard protocols with the Ion OneTouch™ 2 Instrument (Life Technologies). A sequencing oligo was annealed to DNA positive ISPs and the sequencing polymerase bound, prior to loading of ISPs into Ion P1 v2 sequencing chips. Six libraries were loaded onto one P1 sequencing chips and run on a Ion Proton™ Machine according to the Ion Proton™ Sequencing 200 Kit v3 protocol.

The second Ion Proton™ run of mouse *p53* cDNA, differs from the one already described in the kit used for clonal amplification of the size-selected libraries, the chip used and the protocol applied for sequencing. Each size-selected library was diluted to a final concentration of 11 pM and clonally amplified using the Ion Proton™ Hi-Q™ Template Kit (Life Technologies) with Ion OneTouch™ 2 instrument (Life Technologies). Sequencing of the samples was conducted on one P1 sequencing chip loaded with four libraries, according to the Ion Proton™ Hi-Q Sequencing Kit Protocol. Samples were run on an Ion Proton™ sequencer.

b) Ion PGM™ System

Libraries for *p53* cDNA amplicons derived from GFP^+ NIH3T3 cells, were prepared following the manufacturer protocols for genome sequencing with the Ion PGM™ sequencer (Life Technologies) as already described for Ion Proton™ System. Each size-selected library was diluted to a final concentration of 25 pM and clonally amplified using the Ion PGM™ Hi-Q™ Template Kit (Life Technologies). After emulsion PCR, DNA positive ISPs were recovered and enriched according to standard protocols with the Ion OneTouch™ ES Instrument (Technologies). A sequencing oligo was annealed to DNA positive ISPs and the sequencing polymerase bound, prior to loading of ISPs into Ion 316 sequencing chips. Sequencing of the samples was conducted on one 316 sequencing chip loaded with six libraries, according to the Ion PGM™ Hi-Q Sequencing Kit Protocol. The chip was run on an Ion Personal Genome Machine.

2.6.6 Southern blot analysis

Southern blot analysis was performed on ES cells previously electroporated with linearized pLS7154 or pLS7106, to assess if the homologous recombination occurred at the desired genomic *locus*. DNA (3 μg) extracted from ES cells was digested with Hind III or EcoRV (*p53* 5' and 3' arms respectively), or with Psi I – Cla I and Bgl II (*Apc* 5' and 3' arms respectively) in 96-well plates. Digested DNA was loaded onto a 0.8 % agarose gel and electrophoresis was performed o/n at 4 ˚C at 80 Volts. After the run, incubation of the gel for 1 h at RT in denaturing solution upon slow agitation allowed for the denaturation of the DNA in the gel. Subsequently, capillary transfer of the DNA from the gel to a Hybond-N+ (Amersham Biosciences) membrane was performed o/n in SSC 20X (EuroClone) buffer according to standard protocols. Membranes were then exposed twice to 0.12 joules UV light using the UV-stratalinker 2400 (Stratagene) and then the membrane was placed in a roller bottle, pre-incubated with pre-warmed PerfectHyb Plus (Sigma) hybridization solution for 1 h at 65 ˚C upon agitation on a wheel inside in the hybridization oven. A radiolabelled probe was added to the roller bottle and hybridization was allowed o/n at 65 ˚C. The membrane was washed 3 times in washing buffer at 65 ˚C and the radioactive signal was detected with Thypoon TRIO (GE Healthcare) phospho-imager. Images were acquired with ImageQuantTL Software (GE Healthcare).

The radiolabelled probe was prepared by PCR using mouse ES cells genomic DNA as a template. Table 10 reports the oligos used for the creation of *p53* and *Apc* specific probes and the PCR program applied. Probes were labeled with dCTP [α-^{32}P] 6000 Ci/mMol (PerkinElmer) using the Prime-a-Gene® Labeling system kit (Promega) following manufacturer's instructions. Radioactive probes were purified from free ^{32}P dCTP with GE Illustra Probequant G-50 Micro columns (GE Healthcare), following manufacturer's protocol.

Table 10: PCR conditions and oligos used for southern blot probe generation.

	p53 5' probe	*p53* 3' probe	*Apc* 5' probe	*Apc* 3' probe
Oligos	LS7095 – LS7096	LS6946 – LS6951	LS7021 – LS7029	LS7169 – LS7170
PCR program	94 °C 2 min 94 °C 30 sec | 55 °C 30 sec | x 25 cycles 72 °C 30 sec | 72 °C 5 min	94 °C 2 min 94 °C 30 sec | 55.5 °C 30 sec | x 35 cycles 72 °C 30 sec | 72 °C 5 min	94 °C 2 min 94 °C 30 sec | 59 °C 30 sec | x 25 cycles 72 °C 30 sec | 72 °C 5 min	94 °C 2 min 94 °C 30 sec | 56 °C 45 sec | x 25 cycles 72 °C 30 sec | 72 °C 5 min

2.7 BIOCHEMISTRY

2.7.1 Protein lysates

Transfected cells (see section 2.5.4) were harvested 24 h post transfection and lysed for immunoblot assay. NIH3T3 cells were harvested at the density of $2x10^6$ cells in a 10-cm dish, while HEK 293T cells and pMEFs were harvested at the density of $6x10^5$ in a 6-well plate. Cells were washed on the plate with PBS 1X and then placed on ice covered with 1 mL of PBS 1X. Cells were scraped and collected in a 1.5 mL Eppendorf tube. After centrifugation at 5,000 rpm for 5 min at 4 °C, supernatant was removed and the pellet was resuspended in 60 – 100 μl of RIPA-02 buffer and incubated on ice for 10 min. After incubation, samples were sonicated for 5 cycles of 15 sec each using the Bioruptor™ Next Gen (Diagenode) and lysates were stored at - 80 °C.

2.7.2 Immunoblot

Sonicated protein lysates (20 μl) were mixed with Laemmli buffer to a final concentration of 1X Laemmli supplemented with EDTA-free protease inhibitors 1X, denatured at 95 °C for 5 min and loaded on a Mini-PROTEAN® TGX™ Precast Gel 4 – 20 % (Bio-Rad). Electrophoresis was performed for 30 min at 200 Volts in 1X Running Buffer. Proteins

were transferred from the gel to a 0.2 μm nitrocellulose membrane by Trans-Blot® Turbo™ Transfer System (Bio-Rad) for 7 min. The membrane was blocked with 5 % milk in TBS-T for 1 h at RT shaking to saturate the non-protein containing areas and to prevent unspecific binding of the antibody (indicated in Table 4). After blocking, milk was removed and the membrane was incubated with the specific antibody o/n at 4 ˚C upon agitation. After 3 washes for 10 min with TBS-T 1X shaking, the membrane was incubated with the specific secondary antibody for 1 h at RT upon agitation. This step was not performed if the membrane was incubated with an HRP-linked primary antibody. After 3 washes for 10 min in TBS-T 1X, the membrane was incubated with Amersham™ ECL™ RPN2016 (GE Healthcare) according to manufacturer's instructions and the signals were detected by different exposure using Molecular Imager® ChemiDoc™ XRS+ Imaging system.

2.8 IMAGING

2.8.1 Fluorescence-activated cell sorting (FACS)

NIH3T3 cells (6 – 8 x10^6 cells/transfection) were collected 48 h after transfection and washed in PBS 1X. Cells were resuspended to reach the density of 5x10^6/mL in FACS buffer on ice. Cells were sorted by IFOM Imaging Unit with FACSAria™ III sorter (BD Bioscience) and collected in Sorting buffer for subsequent RNA analysis.

pMEFs *p53*$^{+/-}$ at passage 5 (9x10^6 cells/transfection) were harvested 72 h post transfection and washed with PBS 1X. Cells were resuspended to reach the density of 5x10^6/mL in FACS buffer II on ice. Isolation of GFP$^+$ cells was performed by IFOM Imaging Unit with FACSAria™ III sorter (BD Bioscience) and cells were collected in Sorting buffer II. GFP$^+$ cells were re-seeded at the density of 300 cells/well in a 96-well plate and maintained in 20 % conditioned medium to allow for complete recovery. Conditioned medium was prepared

as follows: complete medium from *p53*$^{-/-}$ pMEFs was harvested every 48 h and filter with 0.22 μm sterile filters.

2.8.2 Bright field microscopy

a) Zebrafish embryos

Injected embryos at 24 hpf (see section 2.9.4.1) were anaesthetized in 15 mg/L Tricaine (Sigma) in a 10-cm dish and decoryonated manually, by using two 1 mL insulin syringes. Embryos were placed on a layer of methylcellulose 3 % in E3 water. Morphological analysis, EGFP and Dextran Rhodamine B (Molecular Probes™) (Rhddext) expression were evaluated using Olympus SZX12 stereomicroscope. Image acquisitions were performed at 22.5X objective for GFP and Rhddext expression or 40X objective to evaluate developmental defects.

b) Adult zebrafish

Images of haematoxylin and eosin (H&E) stains of fish paraffin sections (see section 2.9.7) were acquired with an Olympus SZX12 stereomicroscope at 7X objective, while images at 10X, 20X of specific organs were acquired with a NIKON Eclypse 90i microscope.

Image analysis was performed with NIS-Element image software D 3.1 (Nikon).

Both a) and b) procedures were performed in collaboration with the IFOM Zebrafish Unit.

c) HEK 293T transfected cells

EGFP expression in transfected HEK 293T cells was evaluated using EVOS™ fl Digital Inverted Fluorescence Microscope (AMG). Image acquisitions were performed at 10X.

2.8.3 Confocal microscopy

Embryos at 24 hpf were anaesthetized in 15 mg/L Tricaine (Sigma) in a petri dish, decoryonated manually with two 1 mL insulin syringes and mounted in 1.2 % low-melting

agarose in E3 water. Image acquisition was performed at 40X water immersion objective on a Leica TCS SP2 confocal microscope. Image analysis was done with Leica LCS software in collaboration with the IFOM Zebrafish Unit.

2.9 *IN VIVO* TECHNIQUES

All the procedures performed on live animals were performed according to the Italian Government rules and were authorized by the Italian Minister of Health (project 71/2015 PR, project 02/13).

2.9.1 Mouse strains

The CMV-Gal4-Aid mouse is described in (Franchini et al., 2014) while TRP53/C57 mouse strain (Jacks et al., 1994) was a gift from Dr. Pier Giuseppe Pelicci (IEO, Milan). The Deleter-Cre mouse strain (Schwenk et al., 1995) was a gift from Dr. Stefano Casola (IFOM, Milan). C57BL/6J and CD1 mouse strains were purchased from Charles River Laboratories Italia, SRL.

2.9.2 Organ collection from CMV-Gal4-Aid mouse strains

Organs were obtained from 16-weeks old CMV-Gal4-Aid transgenic mice. Mice were terminally anesthetized with increasing concentrations of CO_2 in the IFOM animal house. Organs were exported with sterile surgical instruments and placed in 10 mL of PBS 1X. Organs were then smashed and passed through a Falcon 70 μm cell strainer filter (Becton Dickinson, USA) with 5 mL of cold PBS 1X and suspensions were collected in a 50 mL Falcon tube. Cell suspensions were centrifuged at 1,200 rpm at 4 ˚C for 10 min and the pellet resuspended in 1 mL of hypotonic lysis buffer to achieve erythrocytes lysis. After 10

min of incubation on ice, the reaction was inactivated by adding 9 volumes of cold PBS 1X. After centrifugation at 1,500 rpm for 10 min at 4 °C, cells were resuspended in PBS 1X, divided into aliquots, collected by centrifugation and stored at -80 °C.

2.9.3 Injection of transgenic ES cells into C57BL/6J female blastocyst

Microinjection of the selected homologously recombined ES cells into C57BL/6J female blastocysts was performed by Cogentech S.c.a.r.l Transgenic Unit using a NT88 Micromanipulator System (Nikon). Blastocysts were implanted into pseudo-pregnant CD1 females and pregnancy monitored over-time.

2.9.4 Zebrafish embryo microinjection

Injection procedures and zebrafish handling was done in collaboration with the IFOM Zebrafish Unit.

2.9.4.1 Transient injections

Vectors (40 pg) and Rhddext (5 pg) in Danieau solution were co-injected into wild type (AB) fertilized one cell-stage embryos. 200 – 300 fertilized eggs were injected for each construct and were maintained in a petri dish with water E3 at 28.5 °C. Analysis of developmental problems or of EGFP and Rhddext expression was performed at 24 hpf. The micro injector used was composed of the injection pump PICOSPRITZER® III (Parker Hannifin Corporation) and the MN-153 micromanipulator (NARISHIGE).

2.9.4.2 Generation of transgenic lines

The transposase messenger RNA (mRNA) (70 pg) and zebrafish specific plasmids (40 pg) were co-injected into wt (AB) fertilized one cell-stage embryos. 200 – 300 embryos were

injected for each construct and maintained in a petri dish with water E3 at 28.5 ˚C until 5 dpf. At 6 dpf the larvae were transferred into aquarium tanks of the IFOM Zebrafish Unit with water and room temperature of 28.5 ˚C. The micro injector was used as described above.

2.9.5 Fin clip of larval embryos

Genotyping of zebrafish larvae at 72 hpf was performed via transection of the embryonic caudal fin, followed by rapid genomic DNA extraction and genotyping by PCR (Wilkinson et al., 2013). Briefly, larvae were anaesthetized in 15 mg/L Tricaine (Sigma) for 2 min and transferred on a lid of a petri dish that had been lined with a strip of autoclave tape. Under the Olympus SZX12 stereomicroscope, the caudal fin within the pigment gap of the larva was removed using a microsurgical knife (World Precision Instruments). The fin clip biopsy sample was transferred to a 96-well plate containing 15 µL of 50 mM NaOH and heated at 98 °C for 2 min, cooled down to RT and neutralized with 1/10 volume of 1 M Tris-HCl pH 8.0. The supernatant was used for genotyping by PCR. Each larva was recovered in a single well of a 24-well plate with water E3 until 5 dpf.

2.9.6 Caudal fin resection of transgenic zebrafish

Transgenic zebrafish at different months of age were transiently anaesthetized with MESAB 1X for 2 min at RT. Fish were moved onto a plastic lid of a 10-cm dish and the distal portion of the caudal fin (2 mm) was cut with a sterile micro-scalpel. The tissue biopsy sample was placed in a 1.5 mL Eppendorf tube and stored at - 80 ˚C. The fish was recovered for 5 min in a small tank with water E3 prior to returning to its tank of origin.

2.9.7 Whole mount fish and haematoxylin and eosin staining

Adult zebrafish were terminally anaesthetized with MESAB 1X for 5 min at RT and then placed on ice. Fish were fixed in 4 % paraformaldehyde for 24 h on a shaker at RT before proceeding with the deoxification process. After 5 min of washes with tap water, deoxification was initiated by incubation of the fish with I.E.D (Biocare Medical) for 4 h upon agitation at RT. Processing for paraffin embedding and preparation of H&E 10 μm sections were performed by the Cogentech S.c.a.r.l Tissue Processing Unit using standard protocols. Briefly, for H&E staining, sections were deparaffinized two times in xylene for 10 min and re-hydrated in descending scale of EtOH (100 % to 70 %) for 5 min each. Sections were stained with Papanicolau haematoxylin solution for 30 sec, washed in tap H_2O for 5 min and counterstained in eosin Y-solution for 8 sec. After washing in tap H_2O for 5 min, sections were dehydrated through ascending scale of EtOH (95 % to 100 %) for 5 min each. Stained sections were cleared by two steps in xylene for 5 min and mounted with xylene based mounting medium.

2.10 BIOINFORMATIC ANALYSIS

The Ion Torrent™ NGS data were analyzed with the software developed by Dr. Silvestro Conticello (Istituto Toscano Tumori, Italy). Briefly, the BAM files generated by the Ion PGM™ and Ion Torrent™ Systems were converted to FASTQ format using SAMtools (Li et al., 2009) and an *ad hoc* Perl script. The reads were aligned to the reference sequences (mouse *p53* and *Gapdh* mRNA; zebrafish *p53* and *β-actin* mRNA) with MOSAIK (Lee et al., 2014) and the resulting files were sorted by start position. The analysis of the resulting SAM files was performed using a custom Perl script that trims the low quality ends of the reads and calculates for each SAM file the following information: 1) coverage for each position of the reference sequence; 2) correct coverage for each position of the reference sequence, considering a Phred quality score higher than a set threshold of 25. This

threshold represented the best trade-off to obtain a suitable number of reads while considering the quality of the sequences. 3) Count of the bases at each position of the reference sequence, taking into account the threshold Phred score.

The scripts provided additional information, giving a quality overview of the run, such as 1) distribution of the lengths of the reads; 2) distribution of the nucleotide changes from the reference sequence on the reads; and 4) statistical information on the frequency of the changes across the reads. This last point was measured to assess any technical bias. The frequency of mutations in each sample was calculated using two different scripts that set an *a priori* threshold for the mutation frequency for each position of the segment set as reference sequence. The overall mutation frequency is calculated taking into account the positions of the reference sequence in which there might have been the presence of polymorphisms in the original sample.

The raw frequencies at each base were then processed as follows: only mutation frequencies that were above the minimum of 0.0005 were considered. Since the frequencies from each sample were derived from absolute number of reads, we adjusted the reads according to the highest SNPs frequency. This was based on the assumption that naturally occurring SNPs, which are part of the cell population, are independent of AID induced mutations and hence occurs at equal frequencies in all samples. To assess the significance of each mutation, any observed mutation frequency had to be larger than twice the combined frequencies of the other two samples:

$$\text{Frequency}_{\text{sample1}} = \text{frequency}_{\text{sample1}} - 2*(\text{frequency}_{\text{sample2}} + \text{frequency}_{\text{sample3}})$$

The expected frequency of mutations at dC/dG sites ($E_{C:G}$) was calculated as follow:

$$E_{C:G} = \frac{\#\text{ of mutations}}{(\%\text{ of dC/dG in mRNA sequence analyzed})}$$

The expected frequency of transitions at dC/dG sites ($E_{C:G\ Ts}$) was calculated as follow:

$$E_{C:G\ Ts} = E_{C:G} * \text{Ts frequency (0.333)}$$

The % of WRC present in the analyzed sequence was calculated prior to assess the expected frequency of mutations at WRC motifs (E_{WRC}).

$$\text{WRC } \% = \frac{\text{\# WRC bases in mRNA analyzed}}{\text{(total mRNA bases analyzed)}} \%$$

$$E_{WRC} = (\text{\# of mutations}) * (\text{WRC } \%)$$

The mutation rate for each run was calculated as follow:

$$\text{Mutation rate} = \frac{\sum(\text{absolute frequency}_{\text{significant mut}})}{\text{(total mRNA bases analyzed)}}$$

The list of TP53 human mutations was generated using TP53 database MUT-2.0 (Soussi et al., 2010), and TP53 Mutation Assessor database (TP53 Mut Assessor) (Leroy et al., 2013).

2.11 STATISTICAL ANALYSIS

2.11.1 Student's t-test

Statistical analysis of normal distributed values was performed by two-tailed unpaired student's t test. Differences were considered statistically significant if p value < 0.05.

2.11.2 χ^2 test

Statistical analysis of table of contingency was measured using the χ^2 test. Bonferroni correction was applied and differences were considered statistically significant if p value < 0.016.

CHAPTER 3 – RESULTS

3.1 TALE-AID INDUCES LOW FREQUENCY MUTAGENESIS *IN VITRO*

To study the process of low-frequency mutagenesis during cancer induction at the *trp53 (p53) locus*, we have generated new *in vitro* models by fusing the mouse DNA deaminase Aid to the sequence specific DNA binding protein TALE.

3.1.1 Generation of mouse TALE-Aid fusion protein

Three pairs (fused to a wild-type or mutant Aid protein) of TALEs (TALEa, TALEc, TALEe) were generated, each recognizing a unique sequence of 18 – 19 nucleotides within intron 4 of *p53* (Figure 8A). The choice of target was made based on the following: 1) the majority of *TP53* somatic mutations found in human tumors occur in the DBD; 2) targeting intron 4 allows us to observe mutations mostly in the DBD; 3) TALE is a large protein (120 kD), and by choosing the intron we limit the possibility of blocking the transcription machinery at the promoter or alter the coding sequence, and 4) previous work in our laboratory have shown that *in vivo* targeting of Aid to a specific *locus* induces DNA alterations of ~ 500 bp upstream and downstream of the DNA binding site (Franchini et al., 2014).

Mouse Aid cDNA was fused to the C-terminus of the TALE DNA binding domain followed by three Flag epitopes, with the expression driven by a CMV promoter (TALE-Aid-wt-FL). A control construct was generated, where the same TALE was fused to a catalytic inactive form of Aid (E58Q) (TALE-Aid-mt-FL) (Figure 8B).

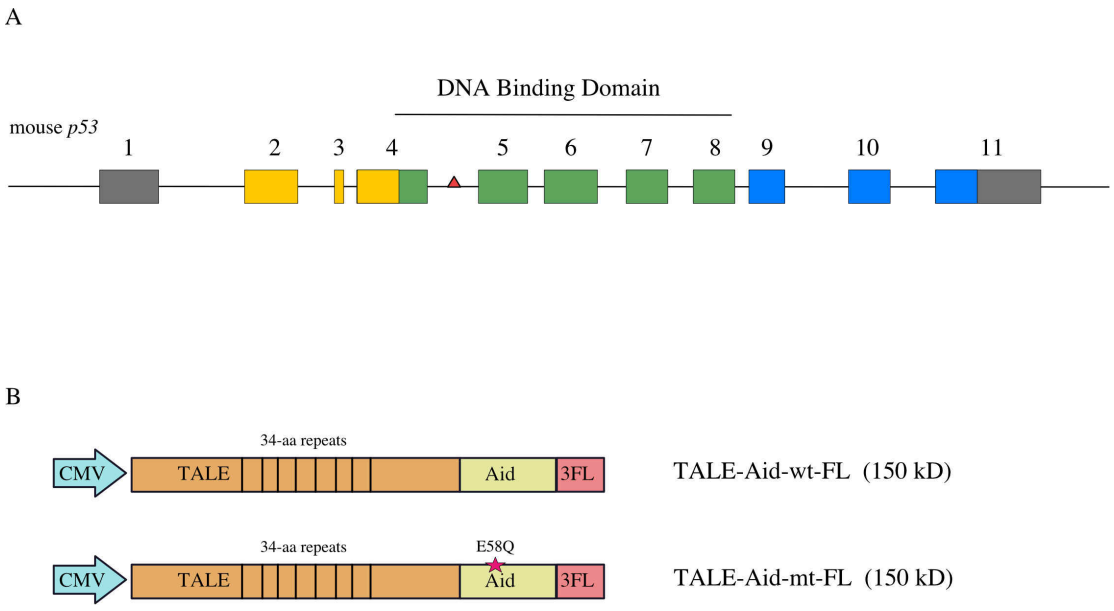


Figure 8: Targeting of *p53* using TALE-Aid fusion proteins.

A) Scheme of mouse *p53* gene structure and TALE binding site (red triangle). *p53* is composed of 11 exons, of which 10 are coding and colored according to the functional domains. Yellow: N-terminal domain; green: DNA binding domain; blue: C-terminal domain; grey: non coding region. The DNA binding domain is encoded beginning in exon 4 through exon 8, and TALE DNA binding protein targets 18-19 nt in the 4th intron. B) TALE-Aid-wt-FL (T-Aid-wt) fusion protein is composed of *p53* sequence specific TALE (orange) fused to the mouse Aid cDNA (yellow) at the C-terminus, followed by three Flag (3FL) peptide as a tag (red). TALE-Aid-mt-FL (T-Aid-mt) fusion protein differs from Aid-wt in one amino acid (E58Q) in the deaminase domain. A CMV promoter drives protein expression.

Given that the TALE itself is a very large protein (1200 aa), we wanted to ensure its full-length mammalian expression. TALE-Aid-FL variants a, c, and e (Ta-Aid, Tc-Aid, Te-Aid, respectively) were transiently transfected into mouse NIH3T3 cells, and analyzed for protein expression after 24 hours by Western blot. Full-length protein was produced for all TALE constructs (Figure 9A). Importantly, there was no significant difference in expression between T-Aid-wt and T-Aid-mt constructs (Figure 9B). Based on this analysis, we decided to test Ta-Aid constructs for targeting and mutational activity.

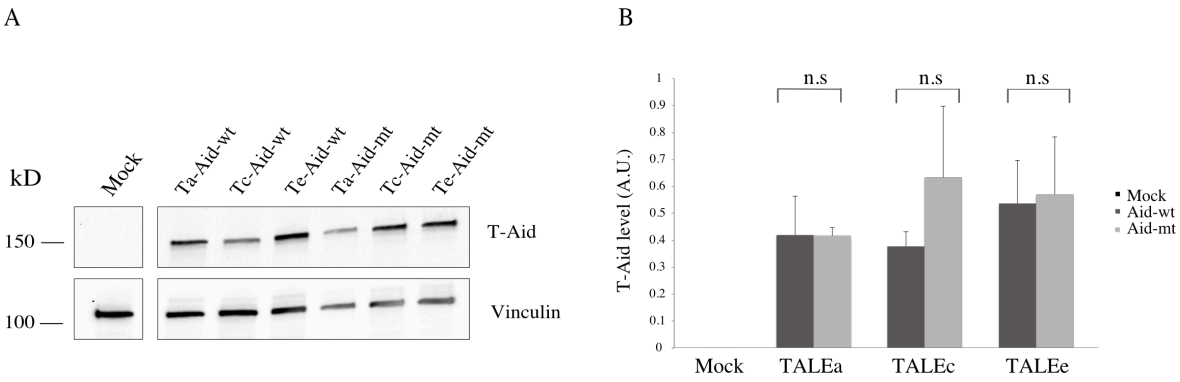


Figure 9: T-Aid protein levels in NIH3T3 transfected cells.

A) Immunoblot of protein lysates from NIH3T3 cells after 24 hours of transfection with T-Aid constructs (10 μg). T-Aid-wt and T-Aid-mt protein levels were determined using anti-Flag antibody. B) Protein levels were quantified in arbitrary units (A.U.) with ImageLab™ software (Bio-Rad) and normalized according to Vinculin protein level. The graph represents three independent experiments with mean ± SEM. Statistical significance was calculated using Student's t-test (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ and n.s = not significant).

3.1.2 TALE-Aid mutates the endogenous *p53 locus* in NIH3T3 cells

In order to determine Aid's ability to mutate the endogenous *locus* of *p53*, NIH3T3 cells were transiently co-transfected with Ta-Aid and EGFP reporter (pCDNA3-EGFP) plasmids. Western blot after 24 h confirmed Ta-Aid and EGFP protein expression, with Aid-wt and Aid-mt at similar levels (Figure 10).

After 48 h we sorted GFP^+ transfected cells and the efficiency was 36 % for Ta-Aid-wt and 32 % for (Figure 11).

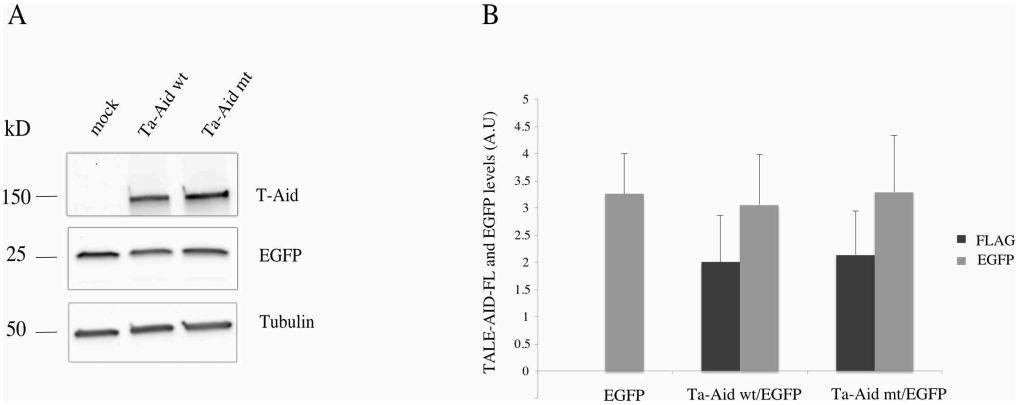


Figure 10: Ta-Aid and EGFP protein levels in NIH3T3 transfected cells.

A) Immunoblot of protein lysates from NIH3T3 cells collected at 24 hours after co-transfection of Ta-Aid (10 μg) and EGFP (1 μg) vectors. Ta-Aid and EGFP protein levels were determined using anti-Flag and anti-GFP antibodies, respectively. B) Protein levels were quantified in arbitrary units (A.U.) with ImageLab™ software (Bio-Rad) and normalized according to Tubulin protein level. The graph represents three independent experiments with mean ± SEM.

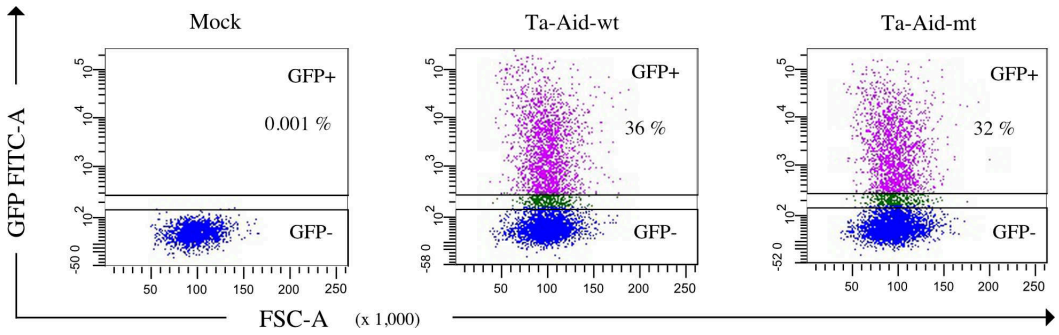


Figure 11: FACS sorting of transfected NIH3T3 cells

FACS sorting of transiently transfected NIH3T3 cells after 48 hours on a BD FACSAria™ III sorter for GFP^+ cells. Numbers within the blot represents the frequency of GFP^+ cells (purple dots) relative to GFP^- cells (blue dots) in mock, Ta-Aid-wt and Ta-Aid-mt samples.

To test the possibility that TALE proteins could block transcription at the target site through steric hindrance (Li et al., 2015), *p53* mRNA levels were assessed by qRT-PCR after sorting of GFP^+ cells. Similar *p53* expression levels were observed between mock-

transfected cells and Ta-Aid transfected cells (Figure 12), suggesting that Ta-Aid proteins do not interfere with gene expression in our system.

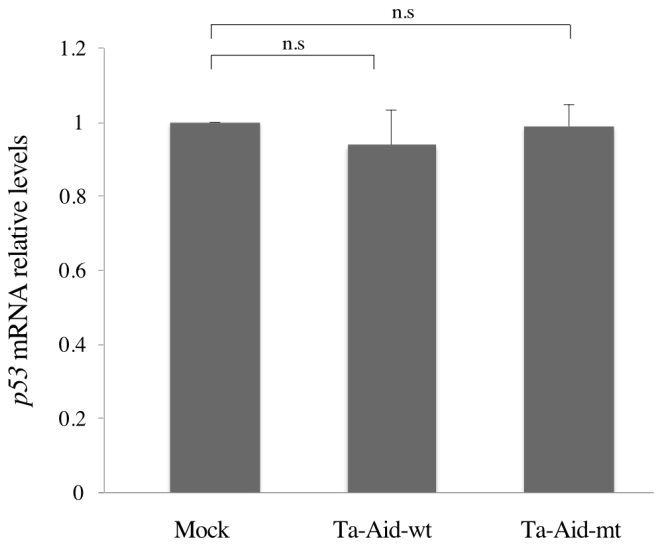


Figure 12: *p53* mRNA levels in NIH3T3 transfected cells.

A) Quantification by qRT-PCR of *p53* transcripts in NIH3T3 cells after 48 hours of transfection with Ta-Aid constructs (10 μg) and sorting of GFP^+ cells. Transcript Ct values were normalized to *Gapdh* mRNA abundance (ΔCt). The ΔΔCt values were calculated subtracting the ΔCt values of Ta-Aid transfected cells to the ΔCt values of the mock. The graph represents three independent experiments with mean ± SEM. Statistical significance was calculated using Student's t-test (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ and n.s = not significant).

The mRNA was isolated and the *p53* cDNA was amplified by high fidelity PCR and subjected to Ion Torrent™ NGS (Figure 13).

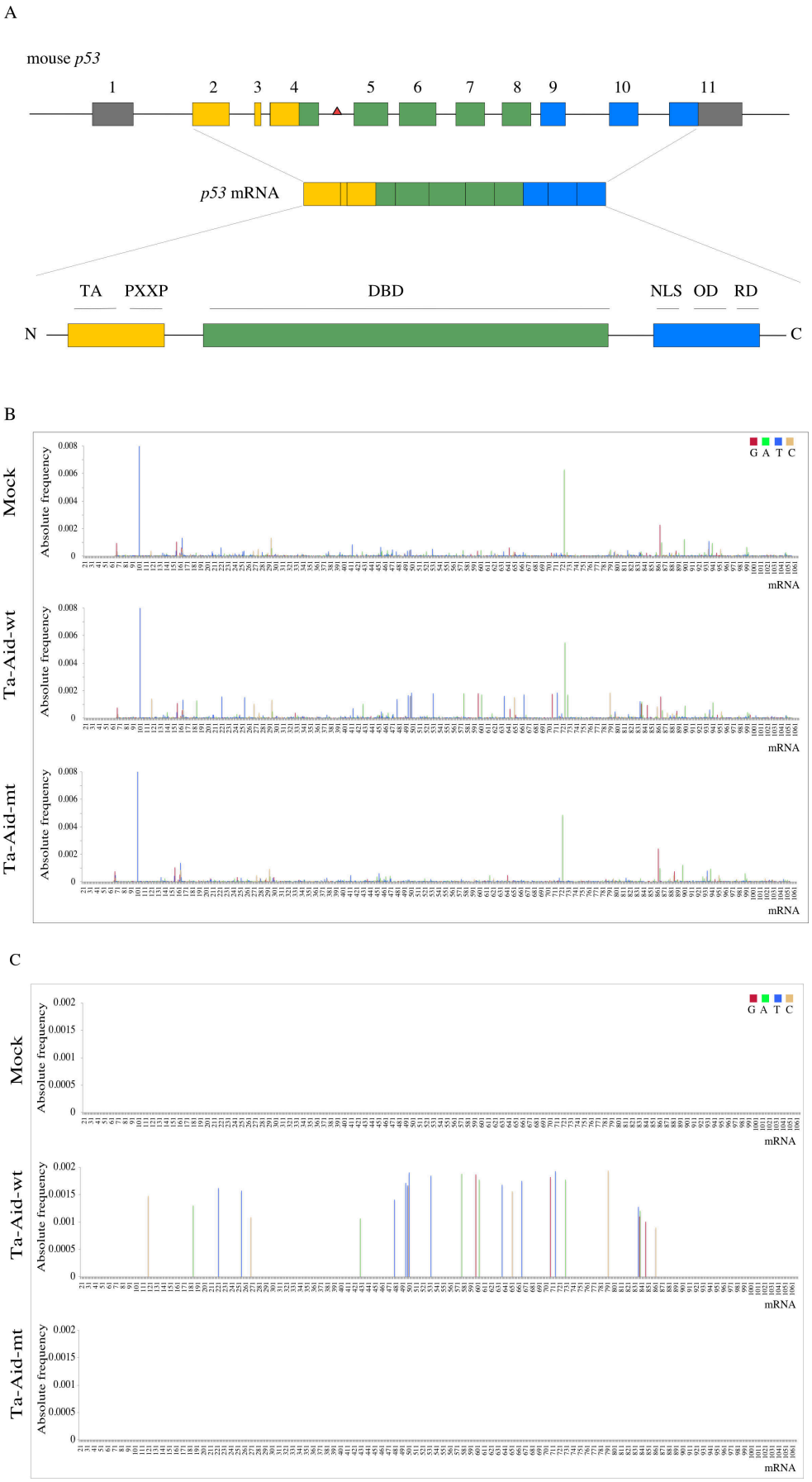


Figure 13: Mutational analysis of *p53* in Ta-Aid transfected NIH3T3 cells.

A) Schematic representation of *p53* gene with TALE target, mRNA product and protein structure (not in scale). Exons are colored according to the functional domains as described in Figure 8. B) *p53* mutation frequency per base of NIH3T3 transfected cells assessed with Ion PGM™ System. The x axis represents the position of 1047/1146 nucleotides analyzed of *p53* mRNA, and the y axis represent the absolute frequency of the mutations observed. C) Significant mutations (based on the formulae described in section 2.10) are indicated with colored bars, according to the mutated nucleotide. (TA: transactivation domain; PXXP: proline rich domain; DBD: DNA binding domain; NLS: nuclear localization signal; OD: oligomerization domain; RD: regulatory domain).

Two different Ion Torrent™ Systems were used for this analysis, Ion PGM™ and Ion Proton™, which differed in both the sensitivity and the DNA polymerase used for the sequencing reaction. Ion PGM™ System could read up to 10^8 bases per run and used a high-fidelity polymerase, while Ion Proton™ System had a higher throughput for each run (10^{10} bases), but used a lower fidelity polymerase.

A total of ~ 6.2 – 7.5x10^5 reads/sample were generated with Ion PGM™ and ~ 1 – 1.5x10^7 reads/sample with Ion Proton™, with an average read-length of ~ 200 bases. After the trimming of low-quality ends and the mapping to the reference sequence, the analysis was performed on ~ 6.1 – 7.3x10^5 reads/sample for Ion PGM™ and 0.8 – 1.2x10^7 reads/sample for Ion Proton™, with a mapping rate of 87 – 91 %.

From this raw analysis possible PCR or sequencing errors needed to be excluded. Based on the formulae described in section 2.10, significant mutations for all three samples were calculated. In the Ta-Aid-wt sample, 27 mutated bases were identified with Ion PGM™ with a mutation rate of 3.9x10^{-5} per bp analyzed, and 31 with Ion Proton™ with a mutation rate of 6.5x10^{-5}, with 26 in common. Neither mock nor Ta-Aid-mt samples showed any significant mutations (Figure 13C). To ensure accuracy of the Ion Torrent™ sequencing protocol and our analysis, *p53* cDNA from the Ta-Aid-wt sample was re-amplified, subjected to a new release of Ion Torrent™ library prep and the Ion Proton™ run, and 28

mutations were observed (with a mutation rate $3.9x10^{-5}$), with 19 in common among the three runs.

Of the 27 mutations observed in the Ion PGM™ run, 21 targeted dC and dG, with 13 being transitions and 7 occurring at WRC hotspot (A/T, A/G, C). The expected frequencies of dC/dG sites, transitions, and WRC were calculated and compared to the observed frequencies (Table 11). dC/dGs were targeted more often (56 % expected vs 77.7 % observed), leading to more transitions at dC/dG (33 % expected vs 86 % observed) and a different frequency of WRCs (11 % vs 26 %). The increase in transitions, dC targeting and WRCs was very suggestive of a DNA deaminase activity, with Ta-Aid recognizing and mutating the endogenous *p53 locus*.

Table 11: *p53* T-Aid-wt induced mutation analysis of NIH3T3 transfected cells.

Total mutations: 27	Expected	Observed
dC/dG targets	15.1	21
Ts at dC:dG	5	13
WRC	3	7

The table represents the expected frequencies of dC/dG targets, transitions (Ts) and WRC compared to the observed in Ion PGM™ System run. Calculations were made according to section 2.10.

In Table 12, the individual mutations in *p53* from Ta-Aid-wt transfected cells are listed. The majority of the mutations (23/37) were located in the DBD (amino acids 99 – 289), while the remaining were located at the N-terminus (8/37) and the oligomerization domain (6/37). In addition, 24 were missense mutations, 1 non-sense and 12 silent mutations. Some of these orthologous mutations were also conserved in human cancers, as analyzed in the TP53 mutation database (Leroy et al., 2013; Soussi et al., 2010). This included the C>T transition R279C (human R282W) in exon 8 that resides in a hot spot highly frequently mutated in human cancers (777 tumors), and leads to the complete protein

inactivation. In the same hot spot region two more mutations were conserved in human cancers, both leading to the inactivation of p53 protein: M243I (human M246I) was reported in 52 tumors and R264P (human R267P) was found in 25 tumors. Moreover, the R178C amino acid change in exon 5 (human R181C), arising from a C>T transition in a WRC motif, was observed in 38 tumors, with p53 activity being reduced, while W143* in exon 5 (human W146*), led to the translation of a truncated protein, with no functional activity left and was observed in 131 different cancers. These findings suggested that AID is targeting hotspot residues in human cancers, as it was already hypothesized (Lindley, 2013).

Table 12: *p53* mutations classification.

Mutation ID	PGM 1	Proton 1	Proton 2	Exon	Codon	amino acid	dC/dG transition	WRC	Distance to target	human amino acid	Frequency in human tumors	p53 activity	Comments
1	yes	yes	yes	4	119G>C	C39S	no	yes	-632	not conserved	-	-	-
2	no	no	yes	4	179C>T	A60V	yes	yes	-572	A63V	2	83.9	Very rare and does not display a significant loss of activity
3	yes	yes	yes	4	185G>A	R62P	yes	no	-556	R65K	0	65.8	Never been described and does not display a significant loss of activity
4	yes	yes	yes	4	222C>T	T74I	yes	no	-530	not conserved			
5	no	no	yes	4	242C>T	A81V	yes	no	-509	A84V	7	72.8	Not frequent and does not display a significant loss of activity
6	yes	yes	yes	4	255C>T	A85A	yes	no	-496	A88I	0	ND	Never been described and loss of activity is unknown
7	yes	no	yes	4	269T>C	L90P	no	no	-482	L93P	0	50	Never been described and does not display a significant loss of activity
8	no	yes	no	4	286T>C	S96P	no	no	-465	S99P	1	11	Very rare and inactive
9	no	no	yes	4	330C>G	F110L	no	yes	-421	F113L	6	16.3	Very rare and inactive
10	yes	yes	yes	5	429G>A	W143*	yes	no	390	W146*	131	ND	Frequent and the activity of the truncated is assumed to be null
11	yes	yes	no	5	479A>T	Y160F	no	yes	440	Y163F	2	59.1	Very rare and does not display a significant loss of activity
12	yes	yes	no	5	495C>T	H165H	yes	yes	457	H168H	10	x	Not freequent and the mutation can change splicing, translation of RNA stability. IT is a passenger mutation
13	yes	yes	no	5	498G>T	M166I	no	no	459	M169I	17	40.1	Partial activity
14	yes	yes	no	5	499A>C	T167P	no	no	460	T170P	2	36.7	Partial activity
15	yes	yes	no	5	500C>T	T167M	yes	no	461	T170M	15	46	Not frequent and parial activity
16	yes	yes	no	5	532C>T	R178C	yes	yes	493	R181C	38	26.1	Frequent with partial activity
17	yes	yes	yes	6	577C>A	R193R	no	no	616	R196R	8	x	Very rare and it can change splicing, translation or RNA stability. Passenger mutation
18	yes	yes	no	6	598C>G	P200A	no	no	637	not conserved	-	-	-
19	yes	yes	yes	6	603G>A	E201E	yes	yes	642	E204E	6	x	Very rare and it can change splicing, translation or RNA stability. Passenger mutation
20	yes	yes	yes	6	636C>T	S212S	yes	yes	675	S215S	2	-	nt change in human with no aa change
21	yes	yes	yes	6	651T>C	Y217Y	no	no	690	Y220Y	0	-	Not freequent and the mutation can change splicng, translation of RNA stability. IT is a passenger mutation
22	yes	yes	yes	7	665C>T	A22V	yes	no	1106	not conserved			-
23	yes	yes	yes	7	707A>G	N236S	no	no	1148	N239T	14	1.9	Not frequent and the p53 is inactive
24	yes	yes	yes	7	714C>T	S238S	yes	no	1155	S241S	5	x	Very rare and it can change splicing, translation or RNA stability. Passenger mutation
25	yes	yes	yes	7	729G>A	M243I	yes	no	1170	M246I	52	0.2	Frequent with inactivating mutaiton
26	yes	yes	yes	8	791G>C	R264P	no	no	1553	R267P	25	0	Frequent with inactivating mutaiton
27	yes	yes	yes	8	835C>T	R279C	yes	no	1598	R282W	777	0.55	HOT SPOT with inactivating mutation
28	yes	yes	yes	8	837C>G	R279R	no	no	1600	R282R	22	-	nt change in human with no aa change
29	yes	yes	yes	8	838C>A	R280S	no	yes	1601	R283S	1	12.4	Very rare and mutant is inactive
30	yes	yes	yes	8	846A>G	E282E	no	no	1609	E285E	7	x	Not frequent and it can change splicing, translation or RNA stability. Passenger mutation
31	yes	yes	yes	8	860G>C	R287P	no	no	1623	R290P	2	24.67	Very rare and mutant is inactive
32	no	yes	no	8	875T>C	L292P	no	no	1639	not conserved	-	-	-
33	no	yes	yes	8	885A>T	E295D	no	yes	1648	not conserved	-	-	-
34	no	no	yes	9	955C>A	P319T	no	yes	1798	P322T	1	105	Very rare and does not display a significant loss of activity
35	no	no	yes	10	993G>A	G331G	yes	no	2629	G334G	0	x	Never been described and it can change splicing, translation or RNA stability. Passenger mutation
36	no	yes	yes	10	997A>G	K333E	no	yes	2632	not conserved	-	-	-
37	no	yes	yes	10	1005C>T	F335F	yes	no	2640	F338F	6	x	Very rare and it can change splicing, translation or RNA stability. Passenger mutation

List of all significant mutations from Figure 13. Exon, codon, amino acid substitution, transitions at dC/dG, WRC context and distance to TALE target are represented for each mutation. The list of orthologous human mutations found in cancer was generated with TP53 mutation database (Leroy et al., 2013; Soussi et al., 2010), with the frequency, the predicted p53 activity, and the relevance to human cancers, shown.

Aside from mutations commonly found in human cancers, less frequent amino acidic changes were observed in the DBD. For example, N236S and R280S (human N239T and R283S respectively) were found at the hot spot region and are rarely reported in human cancers up to now, but lead to the p53 protein inactivation, affecting the p53 ability to bind the DNA. We did not score insertions/deletions, since the alignment algorithm of the raw Ion Torrent™ reads eliminated all insertion and deletion reads. This was done in part due to the inherit inaccuracy of the Ion Torrent™ System in misreading insertions and deletions.

Analysis of the distance of the observed mutations from the TALE binding site identified mutations ~ 600 bp upstream and ~ 2 kb downstream of the genomic target site (Table 12). These findings suggested that, although Ta-Aid binds to a specific sequence of *p53* more towards the N-terminus, the whole *locus* seemed accessible to Aid mutator activity.

To evaluate the specificity of Ta-Aid, *Gapdh* mRNA was amplified and sequenced (Figure 14). Applying the same criteria as for the *p53* analysis, only one significant mutation was detected from the T-Aid-mt sample, and none in the other samples – leading to a mutation rate of $3.3x10^{-7}$ per bp (now defined as the technical limitation in this assay). In the past it has been speculated that Ion Torrent™ sequence analysis was not sufficient for mutation detection, but in our assay we could distinguish a mutation rate from Ta-Aid that was 100 fold above technical error background, and 100 fold below that of SHM (Sale and Neuberger, 1998).

Together these findings suggested that our *in vitro* approach led us to mutagenize the *p53 locus* and identify, not only hotspot residues mutated by Aid in cancers, but also novel mutations that could be relevant for cancer progression.

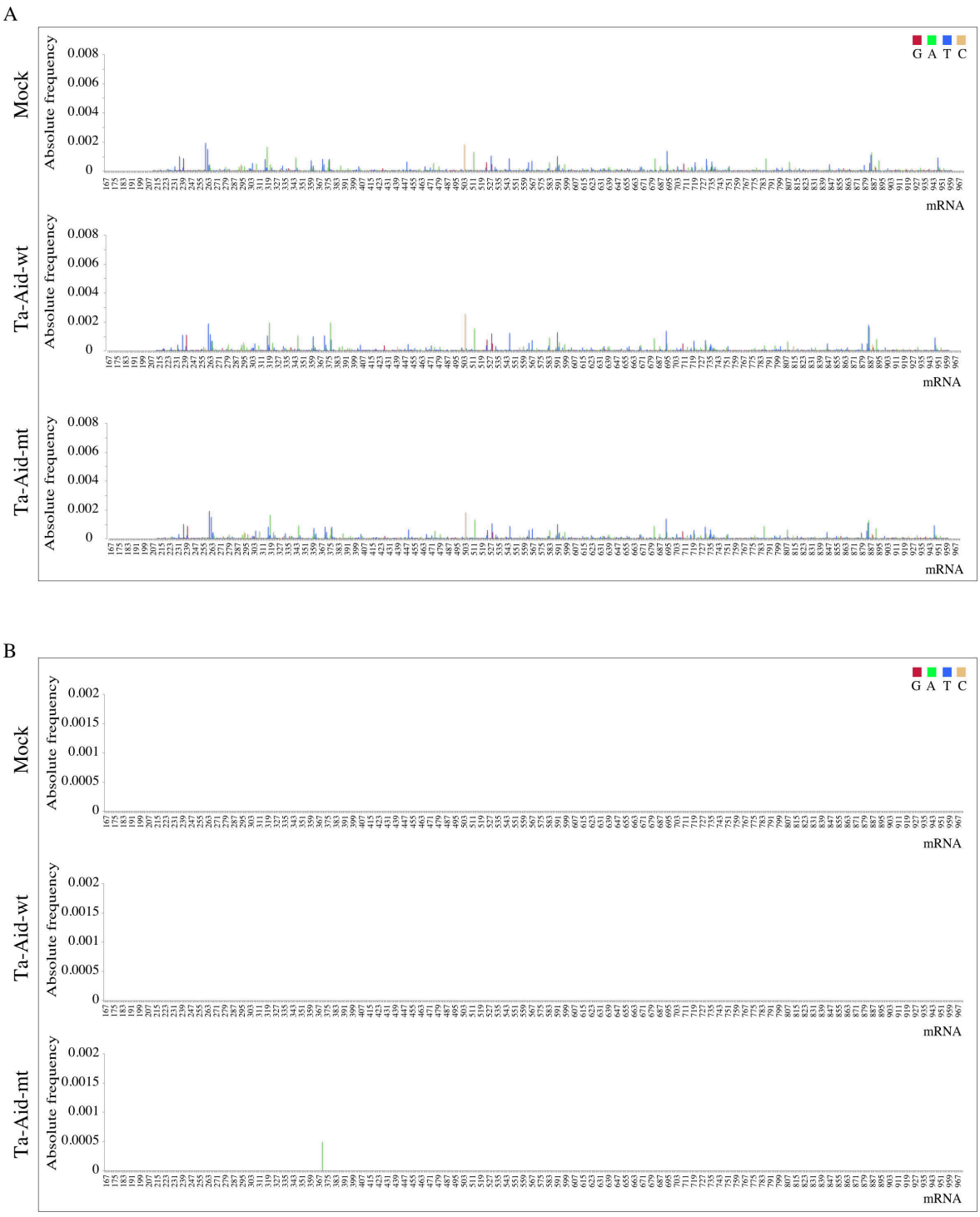


Figure 14: Mutational analysis of *Gapdh*.

A) *Gapdh* mutation frequency per base of NIH3T3 transfected cells is indicated with color bars according to the mutated nucleotide. The x axis represents the position of 804/1002 nucleotides analyzed of *Gapdh* mRNA and the y axis represents the absolute frequency of the mutations observed. B) Significant mutations (based on the formulae described in secion 2.10) are indicated with colored bars.

3.1.3 TALE-Aid soft-agar assay

To evaluate if our mutations in *p53* could have a positive effect on cellular transformation, an anchorage-independent growth soft agar assay was performed. NIH3T3 cells were transfected with all T-Aid pair constructs and 24 h later re-plated at the density of 10^3/well into agar medium. Cluster/colony formation was monitored for 4 weeks, and 8-cell clusters for each transfection were scored, with a slight increase in Tc-Aid-wt transfected cells compared to Tc-Aid-mt, but no significant increase compared to mock cells (Figure 15). Although there was an indication that Aid-wt induced a phenotypic change in the transfected cells, this was not significant. Given our mutation analysis result, it is likely that this assay would not have provided us with the proper frequency (both *p53* alleles needed to be targeted) to observe any phenotypic changes; hence we also attempted other assays.

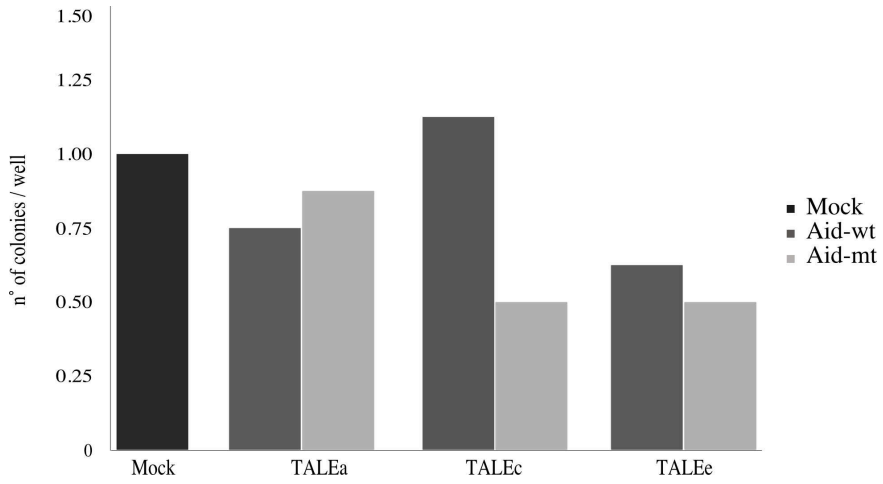


Figure 15: T-Aid does not affect anchorage-independent growth of NIH3T3 cells.

NIH3T3 cells were transfected with T-Aid constructs (10 μg) and at 24 h plated for anchorage independent growth. Colonies were counted after 4 weeks. Only 8-cell clusters were observed in the samples and are shown in the graph as number of 8-cell clusters/well of each sample, normalized to mock. The graph represents one experiment.

3.1.4 TALE-Aid induces p53-dependent loss of senescence in pMEFs

To overcome the limitations of the NIH3T3 immortalized fibroblasts, we took advantage of a more suitable experimental system, consisting of primary mouse embryonic fibroblasts (pMEFs), carrying one *p53* null allele. It is known that primary MEFs undergo stress-associated senescence after a limited number of passages in standard *in vitro* conditions (Parrinello et al., 2003). Mutations and inactivation of p53 can overcome the replication block, leading to immortalization (Harvey and Levine, 1991).

To evaluate if the T-Aid low-frequency mutagenesis could affect cellular senescence, pMEFs were derived from $p53^{+/-}$ embryos (Jacks et al., 1994) at E12.5 – E13.5. Early passage (p 4) $p53^{+/-}$ pMEFs were transfected with Ta-Aid constructs and EGFP reporter plasmid, and full-length protein expression was analyzed after 24 h by immunoblotting (Figure 16). Ta-Aid constructs were less expressed when transiently transfected in pMEFs as compared to NIH3T3 cells, probably due to the difficulties we encountered in transfecting mouse primary cells.

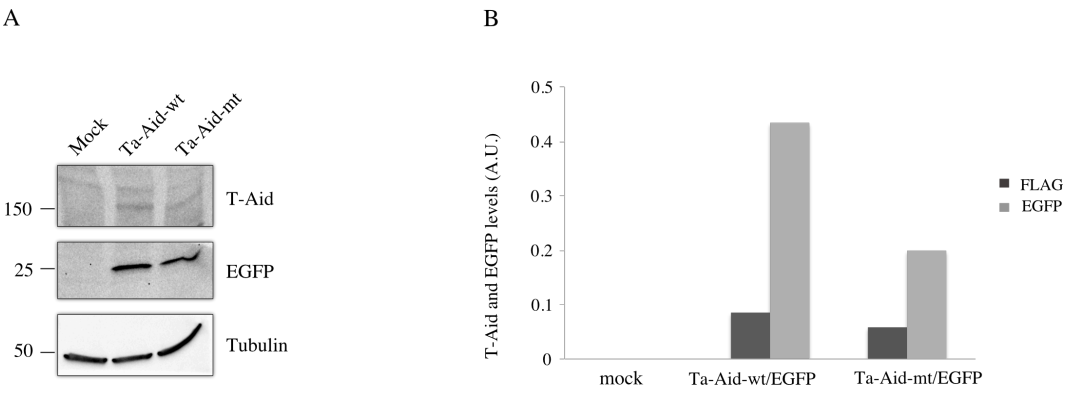


Figure 16: T-Aid and EGFP expression in transfected $p53^{+/-}$ pMEFs.

A) Immunoblot of protein lysates from $p53^{+/-}$ pMEFs collected at 24 hours after co-transfection with Ta-Aid (2 μg) and EGFP (200 ng) vectors. Ta-Aid and EGFP protein levels were determined using anti-Flag and anti-GFP antibodies, respectively. B) Protein levels were quantified in arbitrary units (A.U.) with ImageLab™ software (Bio-Rad) and normalized according to Tubulin protein level. The graph represents one experiment.

After 72 h, GFP$^+$ cells were sorted from Ta-Aid-wt (1.8 %) and Ta-Aid-mt (1.3 %) transfections (Figure 17), and re-plated at 300 cells/well. After 3 weeks 11 clones (9 % - from total well numbers) arising from Ta-Aid-wt transfected cells and 2 clones (2.4 %) from Ta-Aid-mt cells were scored, together with 21 clones (35 %) in the mock sample.

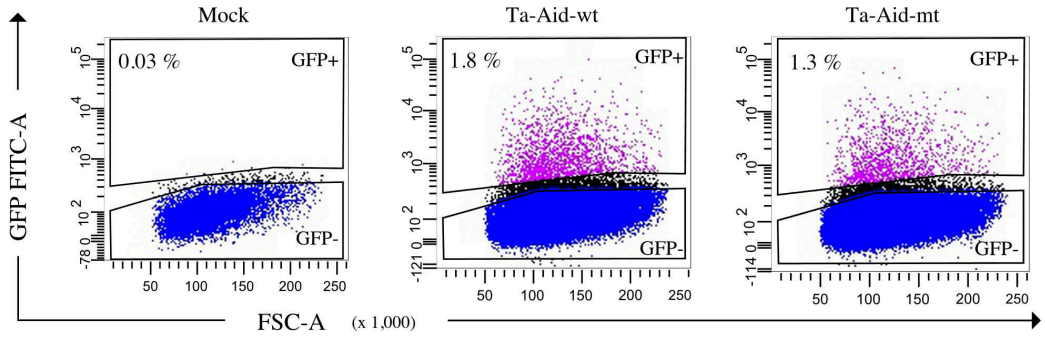


Figure 17: FACS sorting of transfected *p53*$^{+/-}$ pMEFs.

p53$^{+/-}$ pMEFs at passage 4 were transiently transfected with Ta-Aid and EGFP constructs and FACS sorted after 72 h on BD FACSAria™ III sorter for GFP$^+$ cells. Numbers within the blot represents the frequency of GFP$^+$ cells (purple dots) relative to GFP$^-$ cells (blue dots) in mock, Ta-Aid-wt and Ta-Aid-mt samples.

Total RNA was extracted from each clone and *p53* expression evaluated by RT-PCR. Only two clones from Ta-Aid-wt (#4 and #5) and one clone from the mock (#14) transfection expressed full-length *p53* mRNA (Figure 18).

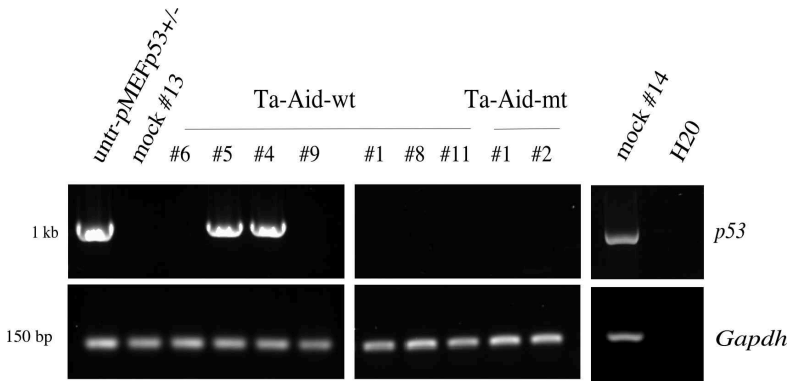


Figure 18: *p53* expression of *p53*$^{+/-}$ pMEF transfected clones.

Representative *p53* expression of transiently transfected Ta-Aid-wt, Ta-Aid-mt *p53*$^{+/-}$ pMEFs and mock clones was monitored by RT-PCR of full-length *p53* mRNA. Untreated pMEFs were used as positive control. *Gapdh* was amplified as loading control.

From those clones that produced *p53* mRNA, we sequenced the *p53* cDNA. Ta-Aid-wt clones #4 and #5 contained a 10-nucleotide deletion at position 85, which was absent in the mock clones (Figure 19A). *In silico* analysis predicted the translation of a truncated protein of 46 amino acids as a result of the frame-shift deletion. This includes the very N-terminus of p53 protein, with only one transactivation domain expressed. Importantly, at the 5' end of the deletion there was a WRC motif followed by a palindromic sequence forming a secondary structure (Figure 19B). These results suggested that Ta-Aid-wt deaminated the WRC sequence on ssDNA and the fragile structure downstream of the targeted dC formed a stem-loop resulting in a deletion, possibly via DNA polymerase slippage.

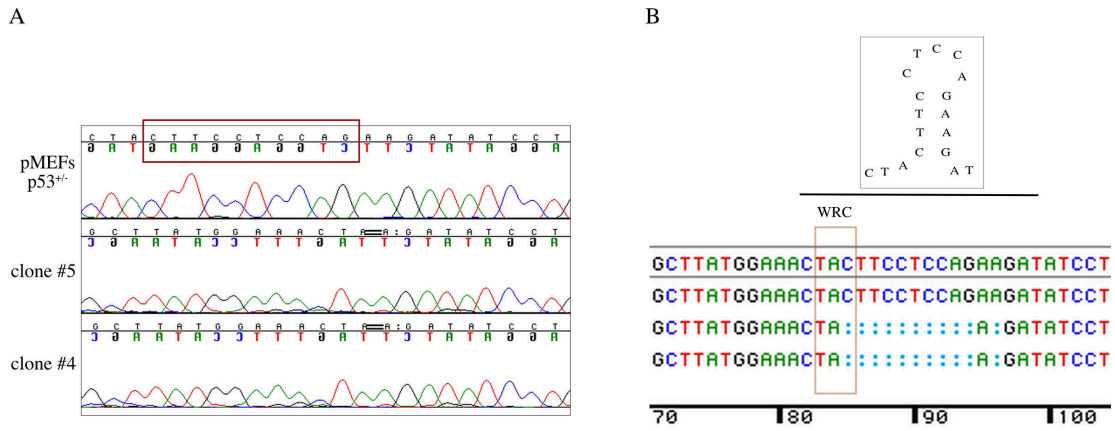


Figure 19: *p53* cDNA sequencing results from transfected *p53*$^{+/-}$ pMEF clones.

A) Sequencing chromatogram of untreated *p53*$^{+/-}$ pMEFs, Ta-Aid-wt clones #4 and #5. The deletion is highlighted with the red box and is present only in clones #4 and #5. B) Sequence of the deletion with schematic WRC motif highlighted at the 5' end (orange box). The deletion is a palindromic sequence and can form a stem loop as shown in the picture above. Nucleotides position of *p53* mRNA are shown at the bottom.

In the remaining Ta-Aid-wt, Ta-Aid-mt and mock clones, genomic *p53* was analyzed through a walking PCR approach (Figure 20A). The region spanning from exon 2 to exon 9 of *p53* wt allele was amplified and exon 2 – 7 were found deleted in all the T-Aid clones analyzed, with clones #4 and #5 serving as controls (Figure 20B). The PCR analysis showed a genomic band corresponding to exon 8 – 9, resulting from the amplification of the *p53*$^{+/-}$ targeted allele. These results suggested that there was the loss of the *p53* second allele. The loss of the second allele of a tumor suppressor is a common event during cancer progression, called loss of heterozygosity (LOH), and could explain why clones escaped senescence.

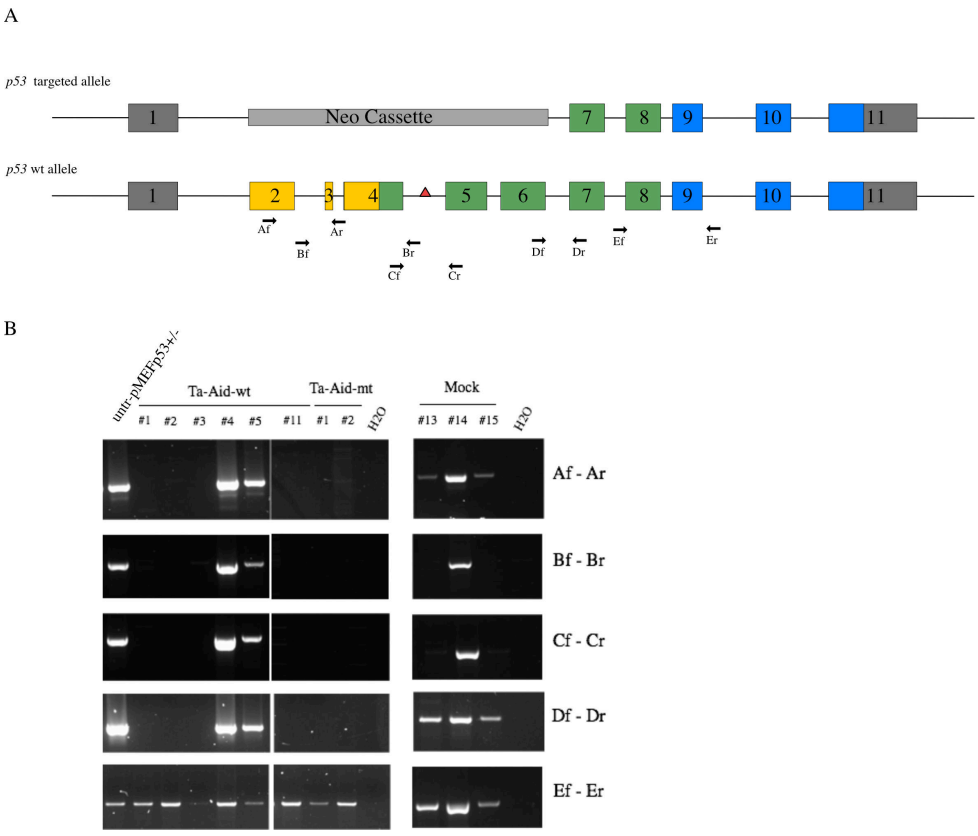


Figure 20: *p53* gDNA analysis of transiently transfected *p53*$^{+/-}$ pMEF clones.

A) Schematic representation of *p53* alleles in *p53*$^{+/-}$ pMEFs. In the recombined allele, a neomycin cassette replaces the region spanning from exon 2 to exon 6 (Jacks et al., 1994). Primer pairs (Af - Ar, Bf - Br, Cf - Cr, Df - Dr, Ef - Er) used for genomic amplification are depicted. B) Representative *p53* genomic amplification of transiently transfected Ta-Aid-wt, Ta-Aid-mt and mock clones. Untreated pMEFs were used as positive control.

PCR amplification of mock samples gDNA resulted in the deletion of intron 4 and possibly exon 5 (Figure 20B). Given that spontaneous *p53* inactivation can occur in pMEFs leading to an immortalized phenotype, it is possible that mock clones underwent such an Aid-independent event. Further experiments are needed to confirm our hypothesis.

3.2 STUDY OF LOW FREQUENCY MUTAGENESIS INDUCED BY TALE-AID *IN VIVO*

Aside from targeting *p53* in cell lines, we aimed to generate a new *in vivo* cancer model for low frequency mutagenesis, either by transiently expressing TALE-AID in zebrafish embryos or by creating a transgenic zebrafish expressing TALE-AID.

3.2.1 Generation of EGFP-TALE-AID fusion proteins

Analogous to the mouse approach, the tp53 (p53) DNA binding domain of zebrafish wt strain (AB) was targeted. Zebrafish p53 shares 46 % homology with the mouse counterpart with the DBD more conserved (Cheng et al., 1997). Two TALEs (TALEg, TALEh) were designed to target a set of 19 nucleotides in intron 6 (Figure 21A). The choice of the target was based on: 1) the above listed assumptions for mouse targeting, 2) our data on the mouse cells that show how TALE-Aid can target large section of the *p53* genome, 3) and the description of three *p53* zebrafish mutants, $tp53^{N168K}$ and $tp53^{I166T}$ in exon 6 and $tp53^{M214K}$ in exon 7, leading to neoplasia (Berghmans et al., 2005; Parant et al., 2010).

To visualize its expression during zebrafish embryo development, the N-terminus of the engineered TALE was fused to EGFP reporter protein, while the C-terminus was fused to human AID either wild type or the catalytic inactive form (E58Q). Expression of the EGFP-TALE-AID (EGFP-Tg-AID, EGFP-Th-AID) protein was driven by an EF1α promoter (Figure 21B).

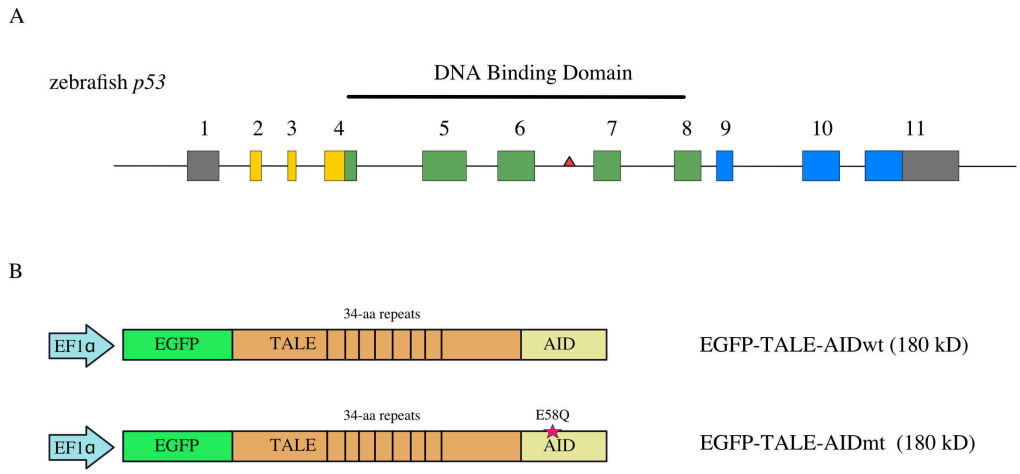


Figure 21: Targeting of zebrafish *p53* using EGFP-TALE-AID fusion proteins.

A) Schematic of the genomic *p53 locus* in zebrafish. TALE binding site is indicated by the red triangle. Exons are colored according to the functional domains. Yellow: N-terminal domain; green: DNA binding domain; blue: C-terminal domain; grey: non coding region. The DNA binding domain (green) is encoded beginning in exon 4 through exon 8, and TALE DNA binding proteins target 19 nt in the 6th intron. B) EGFP-TALE-AIDwt (EGFP-T-AIDwt) fusion protein is composed of EGFP at the N-terminus fused to *p53* sequence-specific TALE (orange) followed by human AID cDNA (yellow) at the C-terminus. EGFP-TALE-AIDmt (EGFP-T-AIDmt) fusion protein differs in one amino acid (E58Q) in the deaminase domain.

An EF1 α promoter drives protein expression of both constructs.

Prior to injecting the plasmid vector containing EGFP-T-AID constructs into zebrafish embryos, full-length expression was assessed in HEK 293T cell lines (Figure 22). EGFP expression was analyzed 24 h after transfection by fluorescence microscopy. EGFP-T-AID and EGFP-AID fusion proteins had reduced expression compared to EGFP alone (Figure 22A). To ensure that the C-terminal AID was also expressed, immunoblot analysis on cell extracts confirmed equivalent protein expression across all TALE-AID fusion proteins (Figure 22B).

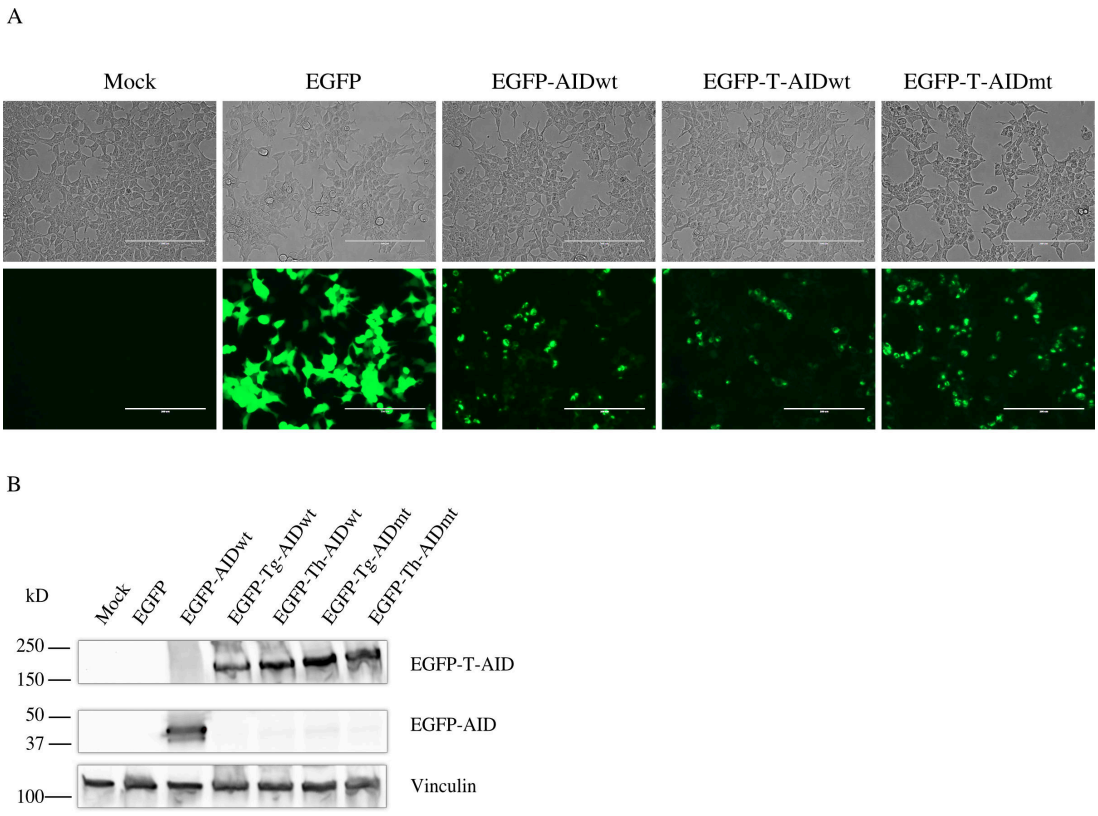


Figure 22: EGFP-T-AID expression in transfected HEK 293T cells.

A) HEK 293T cell were transfected with EGFP expressing vector (pT2AL200R150), EGFP-AIDwt, and EGFP-T-AID constructs (2 μg), and analyzed after 24 h for EGFP expression by fluorescence microscopy. Representative images are shown. Scale bar 200 μm. B) Immunoblot of protein lysates from transfected HEK 293T cells (24 h). EGFP-T-AID protein levels were determined using anti-AID antibody, and Vinculin was used as loading control.

3.2.2 EGFP-TALE-AID is expressed in zebrafish embryos

Fertilized eggs spawn by AB zebrafish strain were injected with EGFP-Tg-AID constructs together with Rhddext, and EGFP alone (pT2A200R150G) as positive control (Figure 23). Rhddext is a vital dye used to follow the injection's outcome without interfering with embryo's development. As shown in Figure 23, Rhddext expression was inconsistent in embryos analyzed at 24 hpf within the same injection (panel b', c' and d'), reflecting the intrinsic variability of the injection itself, which was also observed in EGFP expression (panel b", c", d"). Similar to the *in vitro* transfections, a decrease in EGFP expression was

observed in EGFP-Tg-AID compared to EGFP-only injected embryos. Once injected into fertilized one-cell stage embryo, expression in the developing embryo can be random, depending on the location of the insertion. EGFP injected embryos expressed EGFP in various tissues, including somites, eye and brain (panel b"'), while in EGFP-Tg-AIDwt injected embryos EGFP was detectable in the brain and in the yolk region (panel c"'). EGFP-Tg-AIDmt embryos expressed EGFP mainly in the eyes and the head (panel d"').

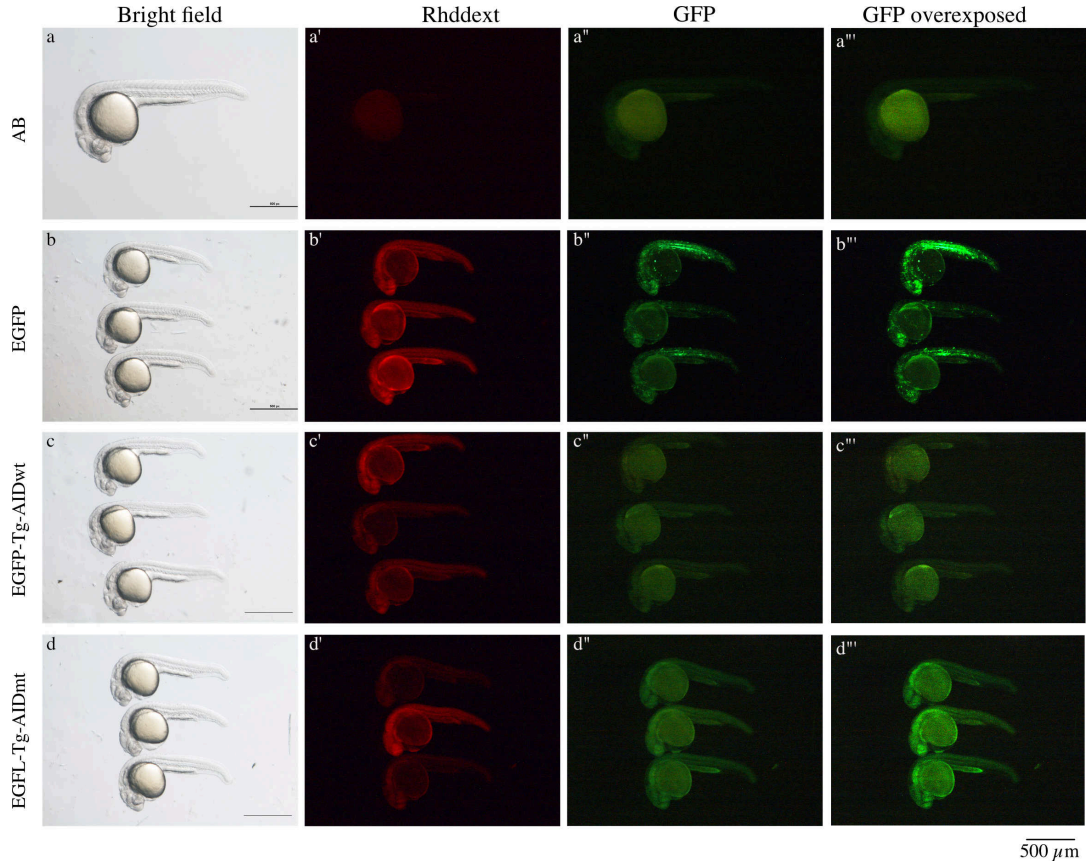


Figure 23: EGFP-T-AID expression in injected zebrafish embryos.

Representative bright field and fluorescence images of embryos at 24 hpf, previously anaesthetized with Tricaine and mounted on methylcellulose 3 %. 200 – 300 embryos were co-injected with Rhddext (5 pg) and EGFP (pT2AL200R150G) (b, b', b", b"'), EGFP-Tg-AIDwt (c, c', c", c"'), and EGFP-Tg-AIDmt (d, d', d", d"') (40 pg per plasmid). The control (AB) represents a non-injected embryo (a, a', a", a"'). Magnification 22.5X. Scale bar 500 μm.

3.2.3 Embryos injected with EGFP-TALE-AIDwt exhibit developmental defects

To test if TALE-AID fusion protein could alter embryonic development through *p53* mutagenesis, we monitored TALE-AID activity during embryogenesis. To this end, 24 hpf embryos injected with EGFP-Tg-AID and controls, EGFP-AIDwt or EGFP only, underwent morphological analysis. We used the 24 h end point because: 1) zygotic transcription starts only 3.5 hpf - at the mid-blastula transition when cell division becomes asynchronous (Kane and Kimmel, 1993); 2) there is an intrinsic mortality until 6 hpf due to the fertilization process which could bias the readout; 3) primary organogenesis and body axis start at 10 hpf, while at 24 hpf the circulation and fins appear, allowing the embryo to move; and 4) pigmentation, which complicates the morphological observations under the transmitted light, starts after 24 hpf (Kimmel et al., 1995).

After removing dead embryos at 6 hpf, at 24 hpf we categorized the embryos into 4 groups according to their phenotype: normal, mild, severe and dead. Normal phenotype was assigned to embryos that resembled a non-injected AB embryo (Figure 24A panels a, b, c, d and Figure 23 panel a). Mild phenotype represented embryos with a defined anterior-posterior axis, but short bodies, curved trunks and minor defects (panels a', b', c', d'), while severe phenotype included embryos with problems since the gastrulation stage (6 hpf), with a functional heart, but an indistinguishable body (panels a", b", c", d").

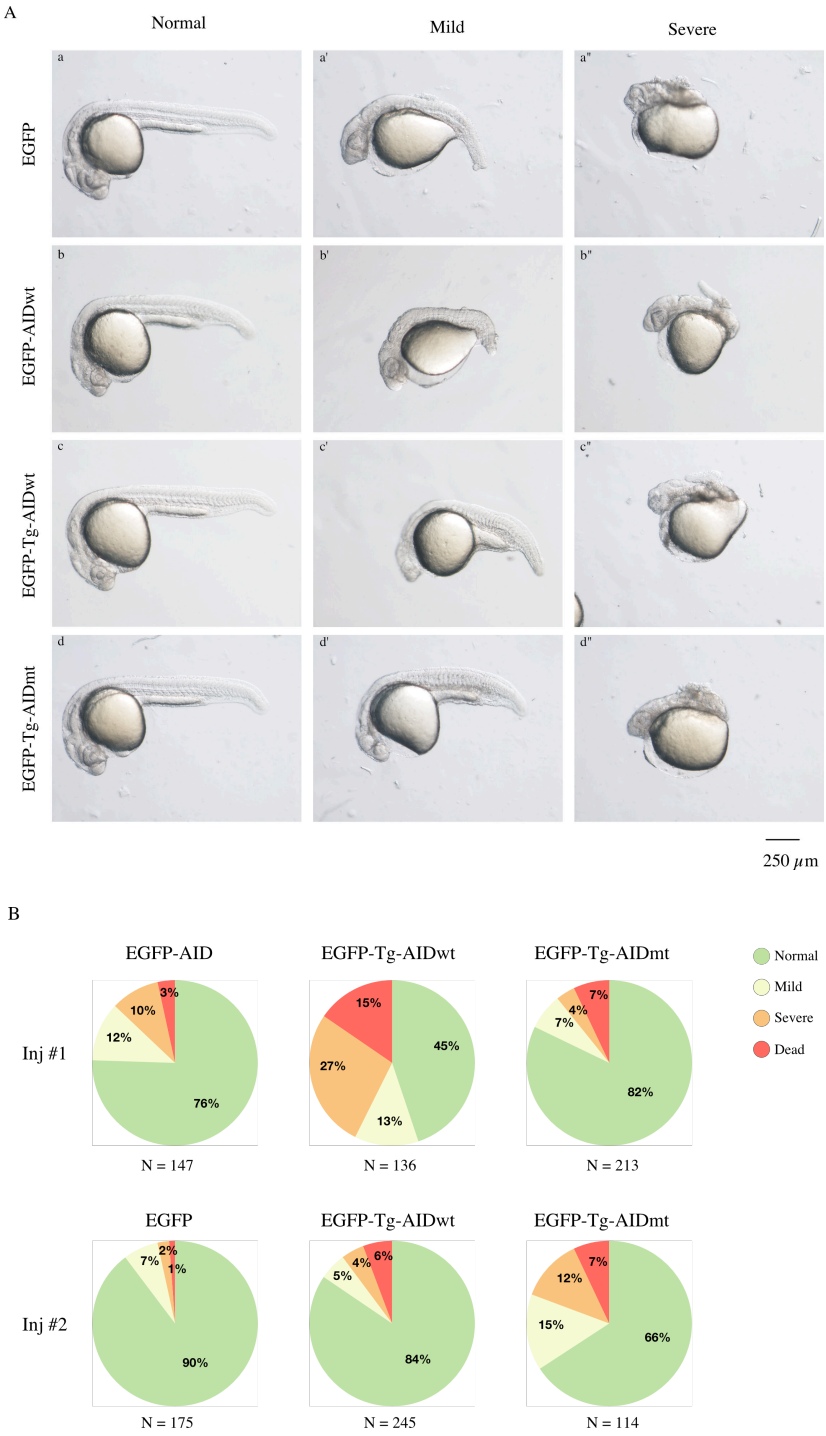


Figure 24: Post-injection developmental phenotypes.

A) Representative bright field images of 24 hpf embryos injected with EGFP (a, a', a"), EGFP-AIDwt (b, b', b"), EGFP-Tg-AIDwt (c, c', c") and EGFP-Tg-AIDmt (d, d', d") (40 pg per plasmid) previously anaesthetized in Tricaine and mounted on methylcellulose 3 %. The developing embryos were separated into four groups according to their phenotype: normal (a, b, c, d), mild (a', b', c', d'), severe (a", b", c", d"), and dead (not shown). Magnification 40X. Scale bar 250 μm. B) Representation of the observed frequencies of the phenotypes reported in Table 13.

Two independent injections were performed, and results are shown in Table 13. In one injection (#1), there was a significant increase in mild, severe, and dead embryos due to the presence by EGFP-Tg-AIDwt, compare to EGFP-Tg-AIDmt and EGFP-AID (Table 14). Furthermore, the difference between EGFP-Tg-AIDwt and EGFP-Tg-AIDmt injected embryos points towards the need of AID catalytic activity (Figure 24B). Un-injected AB embryos displayed normal developmental rates and none of the phenotypes described above was observed. These results suggested that TALE-AIDwt expression has an impact on zebrafish development, possibly via *p53* mutations. However, we cannot exclude that TALE protein could act as roadblock and influence embryo's development by altering *p53* mRNA production.

Table 13: Frequency of phenotypes in 24 hpf injected embryos.

Injection	Construct	Normal (%)	Mild (%)	Severe (%)	Death (%)	N
#1	EGFP-AIDwt	75.51	11.56	9.52	3.40	147
	EGFP-Tg-AIDwt	44.85	12.50	27.21	15.44	136
	EGFP-Tg-AIDmt	82.16	7.04	3.76	7.04	213
#2	EGFP	89.71	6.86	2.29	1.14	175
	EGFP-Tg-AIDwt	84.49	5.31	4.49	5.71	245
	EGFP-Tg-AIDmt	65.79	14.91	12.28	7.02	114

Frequency of the phenotypes observed in 2 independent injections (24 hpf). Injection #1 was performed with EGFP-AID as injection control (N = 147), EGFP-Tg-AIDwt (N = 136) and EGFP-Tg-AIDmt (N = 213). Injection #2 was performed with EGFP as injection control (N = 175), EGFP-Tg-AIDwt (N = 245) and EGFP-Tg-AIDmt (N = 114).

Table 14: Statistical analysis of phenotypes in 24 hpf injected embryos.

Injection	Pairwise comparison	p-value	χ^2 value	d.f
#1	EGFP-AID vs EGFP-Tg-AIDwt	$1.6x10^{-7}$	34.37	3
	EGFP-AID vs EGFP-Tg-AIDmt	0.026	9.295	3
	EGFP-Tg-AIDwt vs EGFP-Tg-AIDmt	$3.8x10^{-13}$	60.85	3
#2	EGFP vs EGFP-Tg-AIDwt	0.052	7.723	3
	EGFP vs EGFP-Tg-AIDmt	$1.7x10^{-5}$	27.34	3
	EGFP-Tg-AIDwt vs EGFP-Tg-AIDmt	$2.6x10^{-4}$	19.05	3

Statistical significance of observed phenotypes was calculated using χ^2 test. Bonferroni correction was applied and significance was considered only if p-value < 0.016. (d.f = degree of freedom).

3.2.4 TALE-AID mutagenesis in zebrafish embryos

Using NIH3T3 cells we were able to demonstrate *in vitro* that targeting Aid to an endogenous *locus* induced mutations in *p53*. To determine the frequency of targeted mutations in an organism, a mutational analysis of the *p53 locus* on 24 hpf injected embryos was performed. 20 embryos with mild phenotype from EGFP-Tg-AIDwt and EGFP-Tg-AIDmt injections, as well as normal AB embryos, were subjected to RNA isolation, cDNA synthesis, PCR amplification and Ion Proton™ NGS. Raw data analysis was carried out as for the NIH3T3 data and 10 unique mutations in EGFP-Tg-AIDwt sample were found that were absent in EGFP-Tg-AIDmt and mock samples (Figure 25).

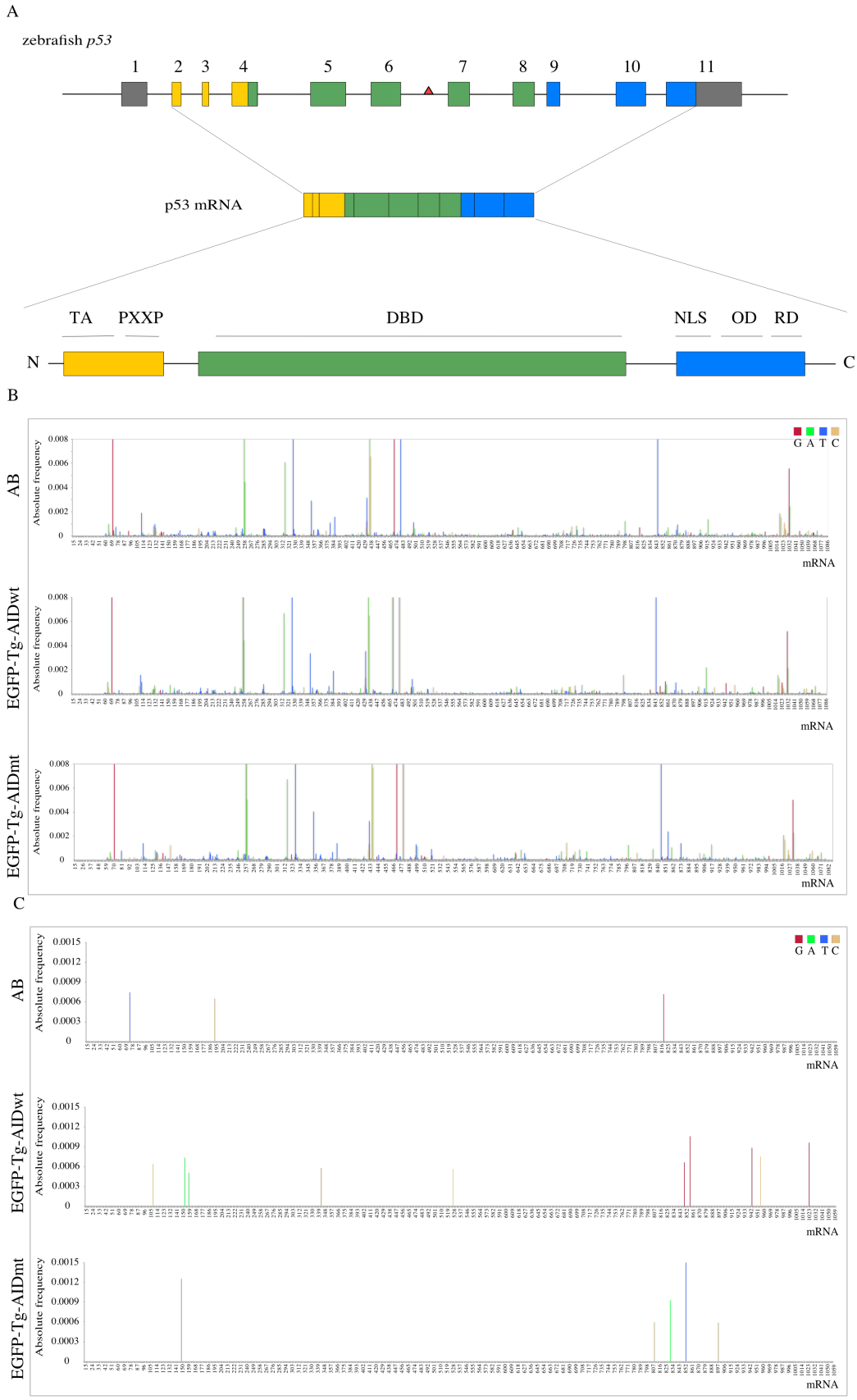


Figure 25: Mutational analysis of *p53* in EGFP-T-AID injected embryos.

A) Schematic representation of zebrafish *p53* gene with TALE target, mRNA product and protein structure. Exons are colored according to the functional domains as described in Figure 21 (not in scale). B) *p53* mutations frequency per base of 24 hpf embryos injected with EGFP-Tg-AIDwt, EGFP-Tg-AIDmt constructs and AB controls. The x axis represents the position of 1074/1110 nucleotides analyzed of *p53* mRNA and the y axis represents the absolute frequency of the mutations observed. C) Significant mutations are indicated with colored bars, according to the nucleotide. (TA: transactivation domain; PXXP: proline rich domain; DBD: DNA binding domain; NLS: nuclear localization signal; OD: oligomerization domain; RD: regulatory domain).

Of the 10 mutations, 3 were found at WRC motifs and 1 was a transition at dC-dG. Mutations were equally distributed upstream and downstream of the TALE target, spanning from exon 4 to exon 11, with the distance from target to mutation being up to 4 kb. Three orthologous mutations were found in human *TP53* (Table 15).

Table 15: zebrafish *p53* mutations classification.

Mutation ID	Exon	Codon	amino acid	dC/dG transition	WRC	Distance to target	Human amino acid	Frequency in human tumors	p53 activity	comments
1	4	114T>C	L38S	no	yes	- 4355	not conserved	-	-	
2	4	158G>A	E53K	yes	no	- 4311	not conserved	-	-	
3	4	164C>A	P55T	no	yes	- 4305	P87T	0	60	no significant loss of activity
4	5	348T>C	V116A	no	no	- 1429	V147A	7	11	Inactive mutant
5	6	532T>C	D117D	no	no	- 532	D208D	1	x	passenger mutation. It can change splicing, translation or RNA stability
6	9	867C>G	P289R	no	no	1956	not conserved	-	-	
7	9	875A>G	S292G	no	no	1964	not conserved	-	-	next to WRC
8	10	961A>G	K320K	no	no	4728	not conserved	-	-	
9	10	973T>C	S324S	no	no	4740	not conserved	-	-	
10	11	1041A>G	N347S	no	yes	4944	not conserved	-	-	

List of all significant mutations from Figure 25. Exons, codon, amino acid substitution, transitions at dC/dG, WRC context and distance to TALE target are represented for each mutation. The list of orthologous human mutations found in human cancers was generated with TP53 mutation database (Leroy et al., 2013; Soussi et al., 2010), with the frequency, the predicted p53 activity, and the relevance to human cancers, shown.

To ensure specificity of the targeting and the mutation approach, *β-actin* mRNA was amplified and sequenced. We were not able to score any significant mutation from three samples analyzed (not shown).

These results suggested that EGFP-Tg-AIDwt induced low-frequency mutagenesis at the *p53 locus* of mild embryos. However, we could not establish a direct correlation between phenotype and mutation frequency. Although the mutation frequency in the NIH3T3 targeting was significantly higher, the preliminary data from the zebrafish targeting is very encouraging and we are currently pursuing this approach more intently.

3.3 EGFP-TALE-AID TRANSGENIC ZEBRAFISH

3.3.1 Generation of transgenic lines

Aside from transiently mutating the endogenous *p53 locus* of zebrafish embryos, we generated transgenic lines expressing stably EGFP-T-AIDwt and the AIDmt control, to study the tumorigenic potential of AID-dependent *p53* loss.

To this end, we took advantage of the Tol2-mediated transposon host approach for plasmid integration into the genome (Urasaki et al., 2006). Here, a synthetic mRNA encoding the transposase and the transposon-donor containing plasmid are co-injected into fertilized eggs. The transposase protein catalyzes the excision of the transposon-like construct from the donor plasmid and induces its random integrations into the genome (Figure 26). The mRNA and the transposase protein gradually degrade and the Tol2 insertions become stable.

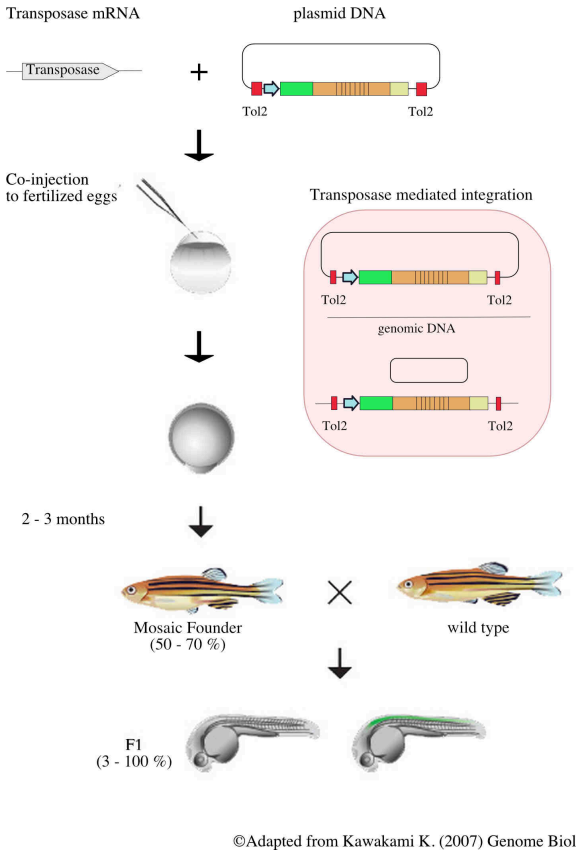


Figure 26: Schematic representation of transgenesis in zebrafish.

Generation of a transgenic zebrafish using Tol2 transposon mediated approach. The synthetic transposase mRNA and a transposon donor plasmid (EGFP-T-AID vectors) containing two minimal Tol2 fragments are co-injected into zebrafish fertilized eggs. The Tol2 containing fragment is excised from the donor plasmid and integrated into the genome. The integration happens in the germ cells at the frequency of 50 – 70 %, and the resulting germ cells of the transposed fish are mosaic (founder). Crossing the founder with a wild type fish resulted in non-transgenic and transgenic heterozygous embryos with a variable frequency (3 to 100 %) (Urasaki et al., 2006).

Zebrafish embryos are capable of independent feeding by 5 dpf, as the yolk supplies are largely depleted by the first week. EGFP-T-AID (wt and mt) constructs were microinjected together with transposase mRNA into AB fertilized eggs. From two independent microinjections with each construct (EGFP-Tg-AID and EGFP-Th-AID), non-fertilized eggs and dead embryos were removed and ~ 100 remaining larvae (6 dpf) per injection were raised to adulthood in the IFOM Zebrafish Unit aquarium. Only 10 to 30 fish per injection reached the adult stage (3 mpf) (Table 16), while in the non-injected AB embryos an intrinsic mortality of 30 – 40 % in the first three months was routinely observed.

Table 16: Germline transmission frequency of EGFP-TALE-AID injected zebrafish.

Injection	Construct	(pg)	N eggs 6 hpf	N alive larvae 6 dpf	N adult fish	N bred fish	N germline	Germline %
3.1	EGFP-Tg-AIDwt	40	130	100	15	8	2	17.6
3.2	EGFP-Tg-AIDwt	30	150	100	22	9	1	
5.1	EGFP-Tg-AIDmt	40	100	80	30	5	1	10
5.2	EGFP-Tg-AIDmt	30	110	80	25	5	0	
4.1	EGFP-Th-AIDwt	40	170	100	25	5	0	18
7.1	EGFP-Th-AIDwt	30	110	70	15	6	2	
4.2	EGFP-Th-AIDmt	40	150	70	30	3	1	33.3
7.2	EGFP-Th-AIDmt	30	110	60	15	3	2	
Total						44	9	

Two independent injections for each construct were performed. For each injection, the construct, the amount of injected plasmid (pg), the total number of eggs at 6 hpf and the total number of alive larvae at 6 dpf are indicated. The number of fish raised to adulthood, of fish outbred to AB fish and fish giving germline transmission are also reported, with the germline transmission frequency for each construct.

A breeding program was initiated to select founder animals. Injected adult animals were crossed to AB fish and resulting F_1 embryos were analyzed for EGFP expression under a fluorescence microscope and/or genotyped. Putative tg(EF1α-EGFP-T-AID) embryos were analyzed with fluorescence microscopy at 24 hpf, but expression was barely detectable. Confocal microscopy analysis was performed on single embryos and EGFP expression detected in different regions of the embryos, such as the eye, notochord and otic vesicle (Figure 27).

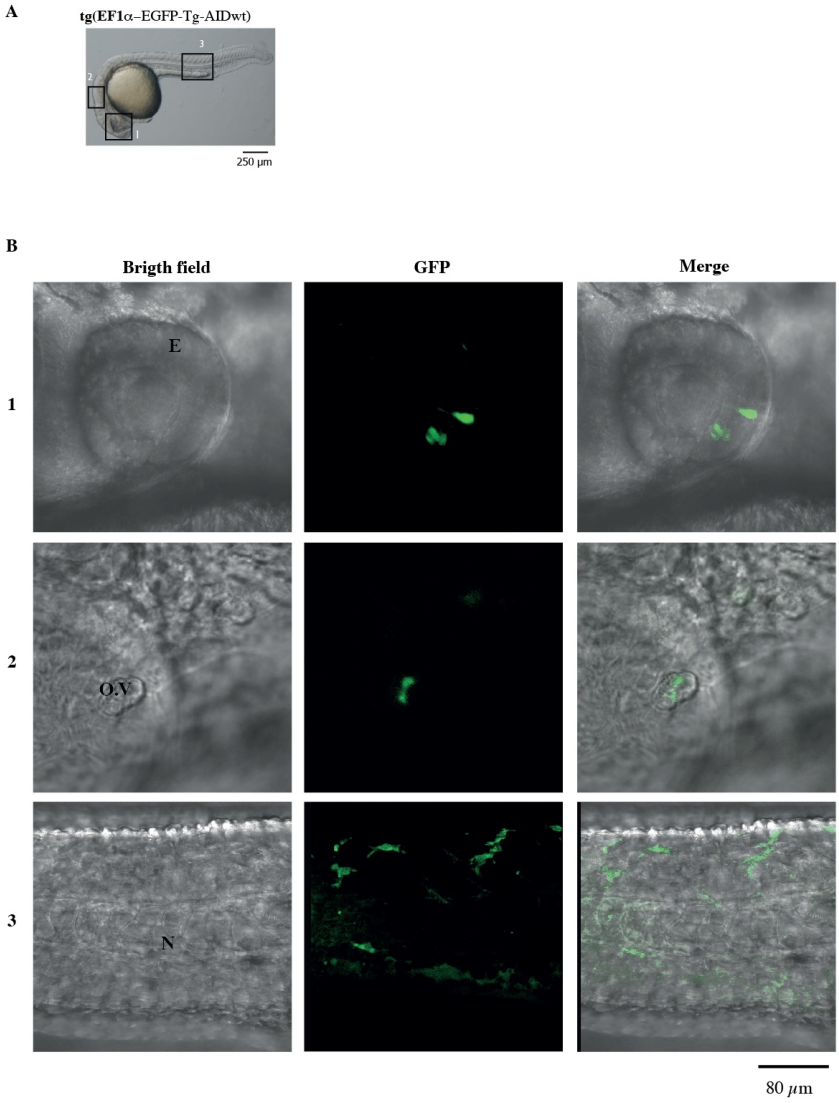


Figure 27: Tg(EF1α-EGFP-T-AID) embryos express EGFP-TALE-AID protein.

A) 24 hpf tg(EF1α-EGFP-T-AID) embryo was anaesthetized and mounted in 1.2 % low-melting agarose in E3 water and EGFP expression was analyzed by confocal microscopy. Magnification 40X. Scale bar 250 μm. B) Confocal images of different anatomical part of a representative embryo: 1) eye (E), 2) otic vesicle (O.V) and 3) notochord (N). Magnification 40X. Scale bar 80 μm.

Although the expression of the transgene could be verified, the weak EGFP signal limited the possibility of transgenic embryos screening based on EGFP expression.

To overcome this limitation, we established a recently published fin-clip approach on alive larvae at 72 hpf for genotyping (Wilkinson et al., 2013). The larvae were temporally anaesthetized and the embryonic tail cleaved, followed by rapid genomic DNA extraction

and genotyping by PCR (Figure 28A). Transgene expression was confirmed on RNA extracted from total larvae at 5 dpf that were fin-clip positive (Figure 28B).

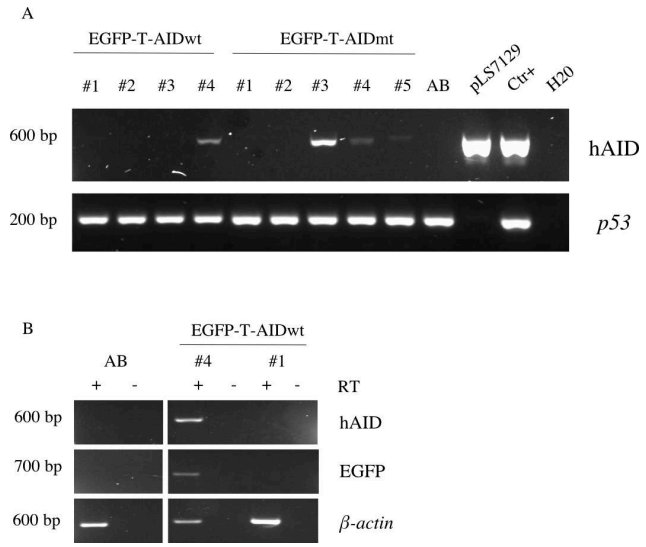


Figure 28: EGFP-T-AID genotyping and mRNA expression of transgenic embryos.

A) PCR amplification of AID transgene was performed on extracted DNA from fin-clip of 72 hpf larvae. Genomic *p53* was amplified as loading control. Larva #4 has integrated EGFP-Tg-AIDwt transgene, while larvae #3 and #4 are positive for EGFP-Tg-AIDmt. AB larvae are used as negative controls, while pLS7129 is amplified as positive control for the transgene presence. B) EGFP-T-AID mRNA expression was determined by RT-PCR on mRNA extracted from representative fin-clipped larvae at 5 dpf. *β-actin* was amplified as loading control.

As expected, the integration of the construct in the germ cell was very heterogeneous, ranging from 10 – 33 %, as summarized in Table 16. Moreover, germ cells of the founder fish are mosaic because of multiple insertions of the construct in the genome. Crossing founders with AB fish, the frequency of F_1 transgenic embryos is not mendelian, but can vary from 3 % to 100 % (Urasaki et al., 2006).

At least one founder for each transgenic line was identified, the F_1 generation established and crossed to AB fish, where a mendelian segregation of the integrated transgene was observed.

3.3.2 Characterization of EGFP-TALE-AID transgenic animals

Although the aim of my work is to generate a new animal model for neoplastic transformation, we also monitored the lifespan and fertility of our lines.

3.3.2.1 Lifespan

Under standard laboratory conditions, the zebrafish average lifespan is 24 – 36 months (Gerhard et al., 2002), yet some of our founder females from the EGFP-T-AIDwt line died after 9 to 13 months. Gross anatomical changes were not detected, and we are in the process of setting up a proper lifespan analysis.

3.3.2.2 Fertility

To determine the capacity of founder animals to induce spawning and produce fertilized eggs, transgenic females were mated to multiple AB males of similar age (to avoid the possibility of male-dependent fertility impairment). After 6 h of spawning period we scored three different types of eggs: fertilized, non-fertilized and dead (Figure 29).

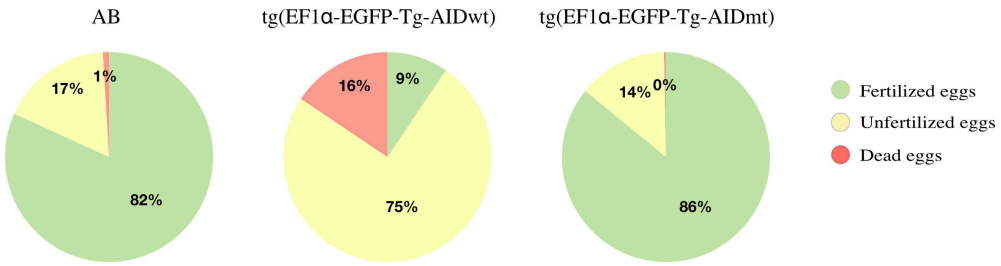


Figure 29: Fertilization success of tg(EF1α-EGFP-Tg-AID) females.

Representation of the fertilization success of tg(EF1α-EGFP-Tg-AIDwt) compared tg(EF1α-EGFP-Tg-AIDmt) founder females and AB controls from Table 15. Two independent breeding were set up for each female analyzed.

Fertilized eggs were characterized by the presence of the germ ring, a typical feature of 6 hpf stage. Non-fertilized eggs remained at one-cell stage, while dead eggs were made of apoptotic cells. Three tg(EF1α-EGFP-T-AIDwt) females, two tg(EF1α-EGFP-T-AIDmt) females and two AB females were induced to spawn in two independent breeding. Tg(EF1α-EGFP-T-AIDwt) animals showed a substantial increase in non-fertilized and dead eggs: females #1 and #3 produced 10 – 15 % of fertilized eggs compared to 80 – 100 % of tg(EF1α-EGFP-Tg-AIDmt) and AB females. Interestingly, tg(EF1α-EGFP-Tg-AIDwt) female #2 did not spawn any eggs in either experiment, even upon induction after manual squeezing (Table 17).

Table 17: Frequency of fertilization success of tg(EF1α-EGFP-Tg-AID) females.

	Fertilized %		Non fertilized %		Dead %		N of eggs	
Breeding:	A	B	A	B	A	B	A	B
AB #1	86.7	89.2	13.3	6.9	n.sp	3.8	98	130
AB #2	92.1	59.0	7.9	41	n.sp	n.sp	114	39
EF1α-EGFP-Tg-AIDwt #1	n.sp	14.7	n.sp	42.7	n.sp	42.7	0	75
EF1α-EGFP-Tg-AIDwt #2	n.sp	n.sp	n.sp	n.sp	n.sp	n.sp	0	0
EF1α-EGFP-Tg-AIDwt #3	2.9	10.8	91.7	84.9	n.sp	4.3	34	397
EF1α-EGFP-Tg-AIDmt #1	100	82.4	n.sp	17.6	n.sp	n.sp	48	17
EF1α-EGFP-Tg-AIDmt #2	62.8	98.3	36	1.7	1.2	n.sp	325	291

Frequency of fertilization success of tg(EF1α-EGFP-Tg-AID) females assessed in two independent breeding A and B (n.sp = non spawned eggs). Total number of eggs counted per breeding is reported.

Although a number of factors could lead to such a phenotype, the significant difference with the AIDwt animals, led us to analyze this phenomena more precisely to possibly identify a problem in the reproductive system due to AID-induced damages.

To begin the analysis of the female reproductive system, we initiated a time-course study at different months of age in the F_1 generation (Figure 30A). Histological analysis performed on paraffin sections of tg(EF1α-EGFP-Tg-AIDwt) animals showed regions of

highly disorganized cells in the stroma surrounding the developing ovarian follicles (Figure 30B panels b, b', b") that were not detected in tg(EF1α-EGFP-Tg-AIDmt) (panels c, c', c") or AB controls (panels a, a', a").

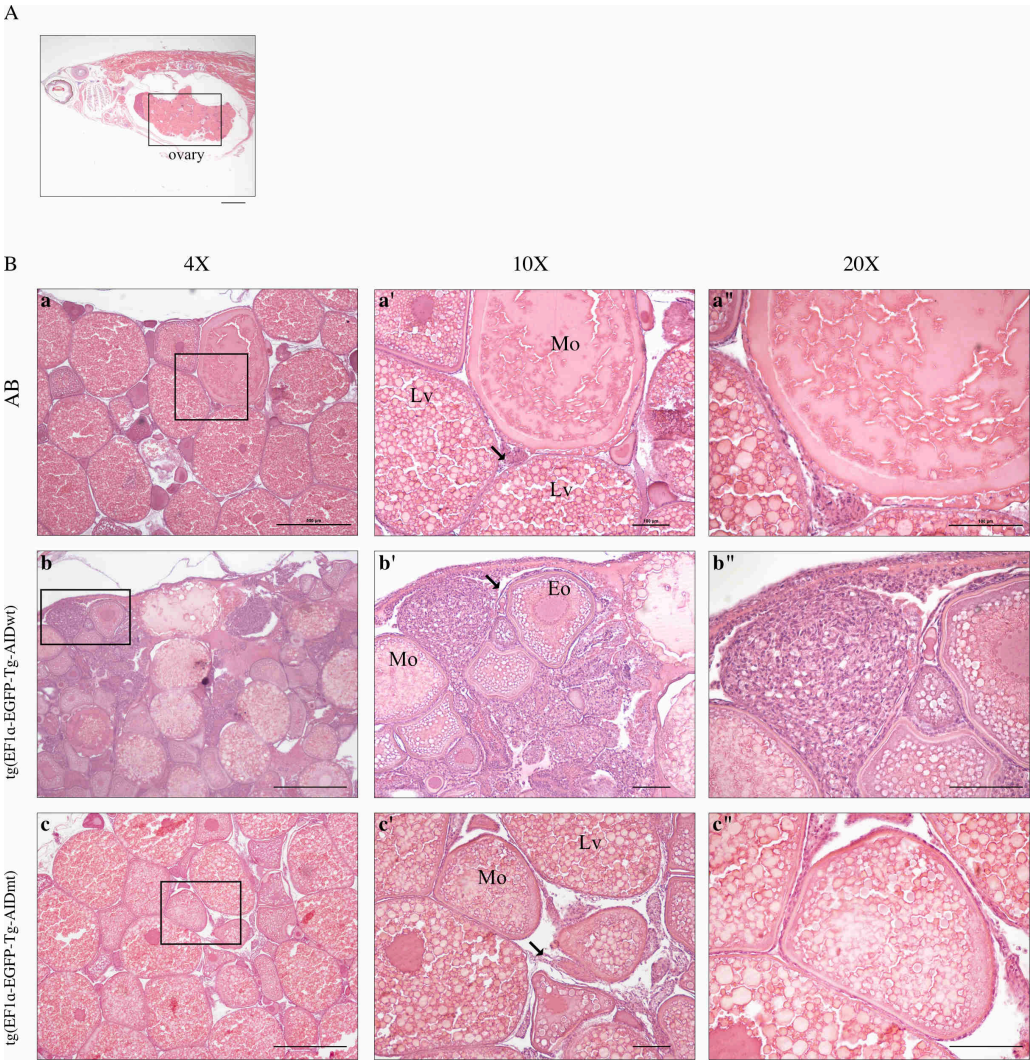


Figure 30: Histological analysis of transgenic females ovaries.

Tg(EF1α-EGFP-Tg-AIDwt) and Tg(EF1α-EGFP-Tg-AIDmt) females were analyzed at 4 and 6 mpf for gross abnormalities. A) Sagittal section of H&E staining of an adult female (6 mpf), with highlighted in the black box the position of the ovaries. Magnification 7X. Scale bar 1.5 mm. B) Representative images of H&E staining of ovaries from AB (a, a', a"), tg(EF1α-EGFP-Tg-AIDwt) (b, b', b") and tg(EF1α-EGFP-Tg-AIDmt) (c, c', c") females at 6 mpf. Oocytes at different stages of development are represented: Eo = Early vitellogenic oocyte, Lv = Late vitellogenic oocyte, Mo = mature/spawning oocyte. The stroma is indicated by arrows. Magnification 4X, 10X and 20X. Scale bar 500 μm, 100 μm and 100 μm respectively.

The stroma is composed of a vascular connective tissue and forms a continuous layer over the neighboring ovarian follicles, called theca layer, while single follicles are surrounded by granulosa cells (Patiño, 2002). These fibrotic-like regions were observed with a frequency of 66.6 % at 4 mpf and 62.5 % at 6 mpf, while they were absent in animals at 12 mpf both transgenic and controls (Table 18).

Table 18: Pathological phenotypes in zebrafish females at different mpf.

Genotype	Sex	Age (months)	Phenotype
AB	Female	4	0/2
tg(EF1α-EGFP-Tg-AIDwt)	Female	4	2/3 (66.6 %)
tg(EF1α-EGFP-Tg-AIDmt)	Female	4	0/3
AB	Female	6	0/2
tg(EF1a-EGFP-Tg-AIDwt)	Female	6	5/8 (62.5 %)
tg(EF1α-EGFP-Tg-AIDmt)	Female	6	0/1
AB	Female	12	0/3
tg(EF1α-EGFP-Tg-AIDwt)	Female	12	0/3
tg(EF1α-EGFP-Tg-AIDmt)	Female	12	0/2

Tg(EF1α-EGFP-Tg-AID) females at different mpf were analyzed for abnormalities in the reproductive system. Frequency of fibrotic-like tissue is reported for experimental animals and related AB controls.

3.4 GAL4/UAS MOUSE MODELS: A COMPLEMENTARY APPROACH FOR LOW FREQUENCY MUTAGENESIS

So far, we have demonstrated that targeting TALE-Aid fusion protein to an endogenous *locus in vitro*, we could induce mutations at low frequency. Moreover, in the *in vivo* model, preliminary results suggested that TALE-AID could induce mutations at the *p53 locus*, even if at significant lower frequency compared to the *in vitro* results. Our previous work demonstrated that Aid when fused to the Gal4 DNA binding domain retained the catalytic activity *in vivo* (Franchini et al., 2014).

To study the process of low-frequency mutagenesis with Aid *in vivo*, we created another animal model using the Gal4/UAS approach to target Aid to our *locus* of interest in a transgenic mouse. We decided to generate two UAS knock-in mice to target *p53* and *Apc*. Breeding Gal4-Aid mouse to the UAS knock-in mouse, Gal4-Aid could bind the UAS binding site and target our gene of interest, inducing low-frequency mutations.

Gal4-Aid deamination activity was tested *in vitro* before creating a Gal4-Aid transgenic mouse. The Gal4 DNA binding domain was fused to wt mouse Aid cDNA - to generate Gal4-Aid fusion protein (Gal4-Aid-wt) and the deaminase activity of the fusion protein was determined using the Rif-mutagenesis assay in *E. coli* (Franchini et al., 2014).

Two different transgenic mouse lines expressing either Gal4-Aid-wt or Gal4-Aid catalytic mutant (E58G) fusion proteins under the control of the CMV-promoter were generated (CMV-Gal4-Aid) (Franchini et al., 2014), and Gal4-Aid was identified to be ubiquitously expressed in various tissues in the adult by RT-PCR (Figure 31), confirming our previous results (Franchini et al., 2014).

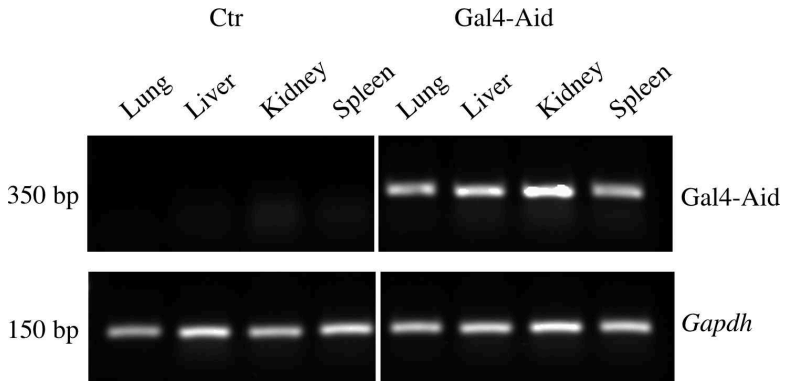


Figure 31: Gal4-Aid expression in transgenic CMV-Gal4-Aid mice.

Gal4-Aid expression in a representative transgenic CMV-Gal4-Aid-wt mouse and control (Ctr) at 3 months of age was monitored by RT-PCR. RNA samples were from lung, liver, kidney, and spleen. *Gapdh* was amplified as loading control.

Gal4-Aid activity was demonstrated *in vivo* by using a demethylation assay: female CMV-Gal4-Aid mice were bred to male mice harboring Gal4 binding sites in a methylated genomic *locus* (H19-DMR-UAS). Analysis of fetal liver samples and embryos showed that Gal4-Aid-wt could induced local-DNA demethylation (500 bp upstream / downstream of its binding site) near its targeted region, demonstrating that Gal4-Aid fusion can be used as a viable means for targeted DNA damage (Franchini et al., 2014).

3.4.1 Generation of $p53^{UAS/+}$ knock-in mouse

To generate the $p53^{UAS/+}$ knock-in mouse, we followed the above described strategy for TALE targeting, and three tandem UAS sites were inserted into intron 4 of *p53*. A classical targeting construct (pLS7070) was generated, where a neomycin (Neo) selection cassette was flanked by two *loxP* sites, imbedded within the homologous targeting sequence (Figure 32).

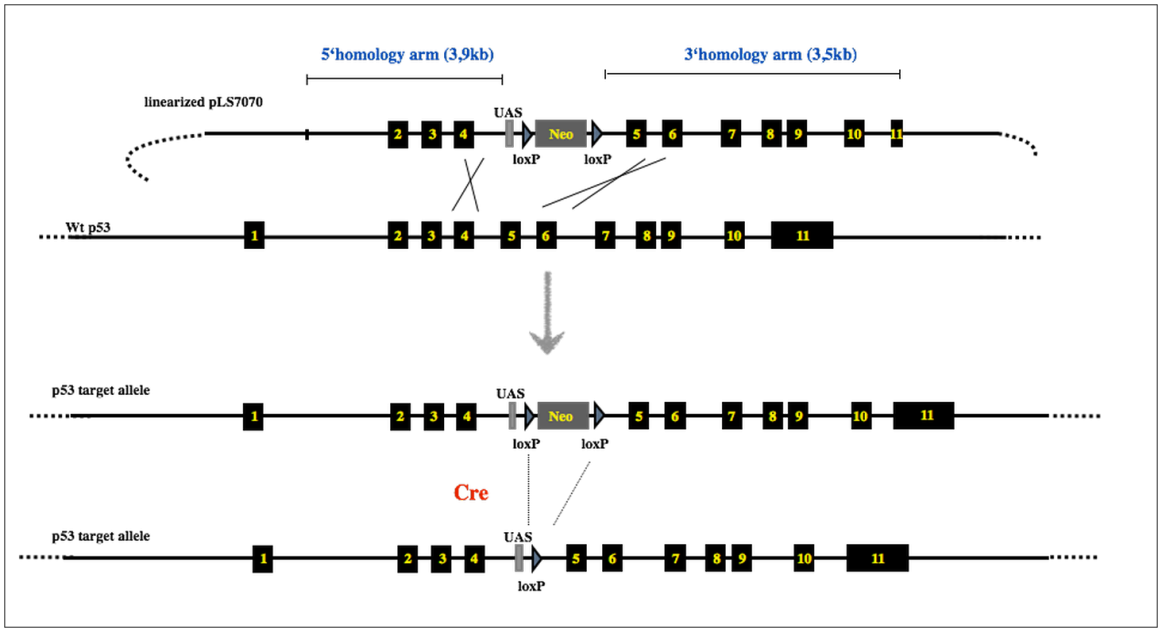
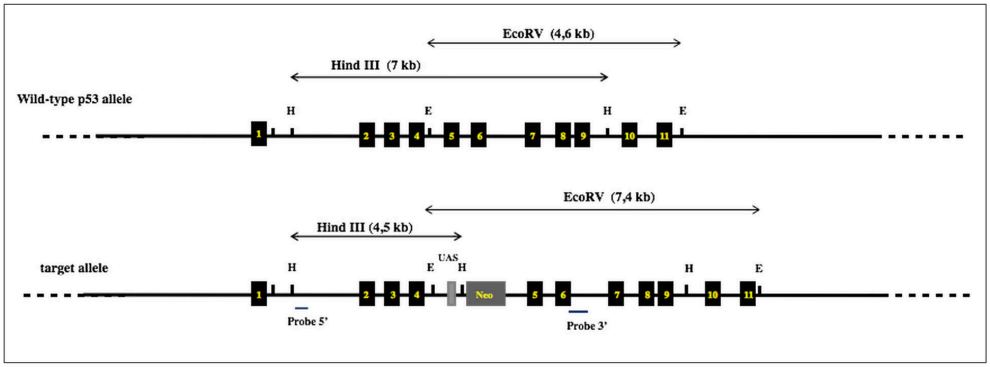


Figure 32: Strategy for the generation of *p53*^{UAS} allele via homologous recombination.

The targeting vector pLS7070 and the genomic organization of *p53* is shown at the top. pLS7070 contains three tandem UAS sites fused to intron 4 of *p53* and a Neo cassette flanked by two *loxP* sites, imbedded within the homologous targeting sequences. Theoretical crossover between endogenous *p53* and the linearized vector is depicted by the crossed lines. After homologous recombination, the UAS sequence with the Neo cassette have been integrated in intron 4. The Neo cassette is removed by crossing the founder *p53*^{UASNeo} mouse with a Deleter-Cre mouse.

E14tg2α ES cells were electroporated with the linearized vector and resulting clones were analyzed by southern blot for homologous recombination events (Figure 33A). Of 570 clones screened, only one had homologous recombined the targeting construct correctly (clone F41) (Figure 33B).

A



B

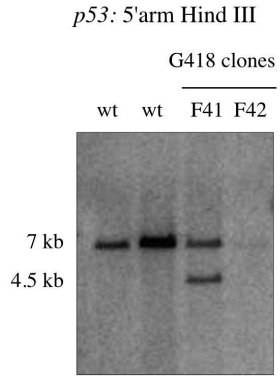


Figure 33: Southern blot analysis of electroporated ES cells with pLS7070.

A) Diagram showing probe design for *p53* targeting: digestion with Hind III and hybridization with 5' genomic probe resulted in a fragment of 7 kb for endogenous wt allele and 4.5 kb for the knock-in allele. Hybridization with 3' probe after EcoRV digestion results in a wt fragment of 4.6 kb and a recombined allele of 7.4 kb. B) ES cell DNA (3 µg) extracted from G418 resistant clones, was digested with Hind III, run on a 0.8% agarose gel, blotted onto a nylon membrane and hybridized with ^{32}P labeled 5' probe. Clone F41 has correctly homologous recombined the vector into *p53* gene.

After karyotype analysis with 76 % of euploid metaphases, F41 clone was injected into C57BL/6J blastocysts that were then implanted into pseudo-pregnant CD1 females. The first injection (#1) resulted in no born pups, while the second injection (#2) gave birth to 13 pups. Of these, one male chimera was found, with 45 % of agouti coat color. It was raised to adulthood and mated to Deleter-Cre (N = 4) or C57BL/6J (N = 2) females. All resulting pups (N = 146) had black coat color, with no germline transmission observed

(Table 19). For this reason, new injections will be done into C57BL/6J blastocysts to find new chimeras and to generate a Gal4-Aid/*p53*$^{UAS/+}$ transgenic mouse.

Table 19: *p53*$^{UAS/+}$ knock-in ES cells injections and pups analyzed.

Clone	Karyotype	# injection	N of pups	N of chimera	Sex	Chimera mated		
F41	76 %	1	0			**Female ID**	**N of pups**	**Germline transmission**
		2	13	1	M	Del-Cre #1	24	0
						Del-Cre #2	25	0
						Del-Cre #3	28	0
						Del-Cre #4	25	0
						C57BL/6J #1	34	0
						C57BL/6J #2	10	0
Total			13	1		6	146	0

The table summarizes two injections performed with the *p53*$^{UAS/+}$ recombined clone F41 into C57BL/6J blastocysts. The number of pups, number of chimeras and sex for each injection are represented. One chimera was found and was mated to different females. Breeding, number of pups and germline transmission are also summarized here.

3.4.2 Generation of *Apc*$^{UAS/+}$ knock-in mouse

To generate *Apc*$^{UAS/+}$ knock-in mouse, we decided to target intron 15 of *Apc* (Figure 34), based on the fact that the majority of *APC* mutations found in human colon cancers are situated in exon 16 (Mutation Cluster Region) (Cottrell et al., 1992).

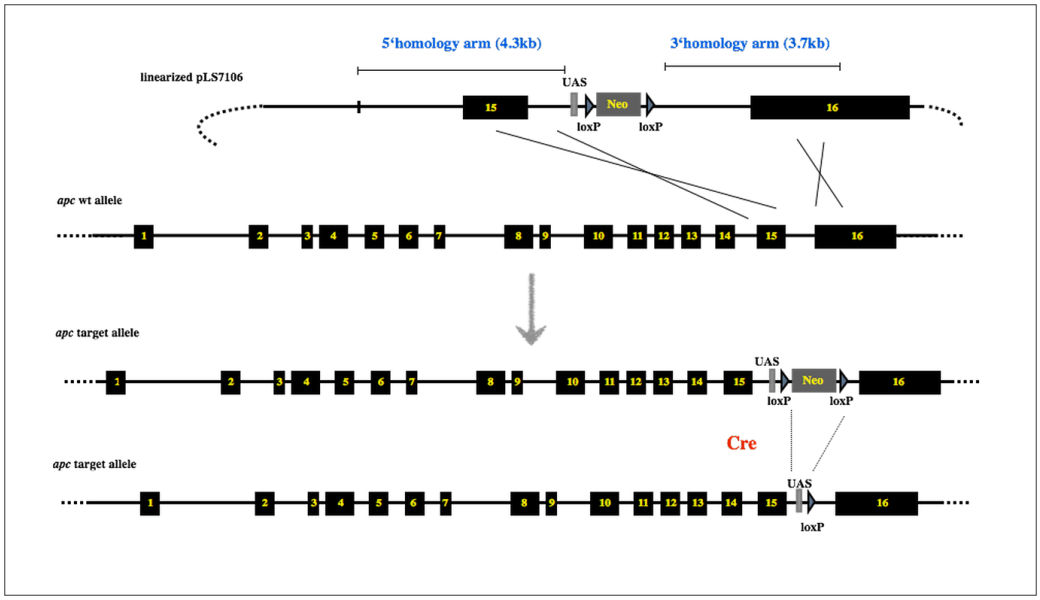
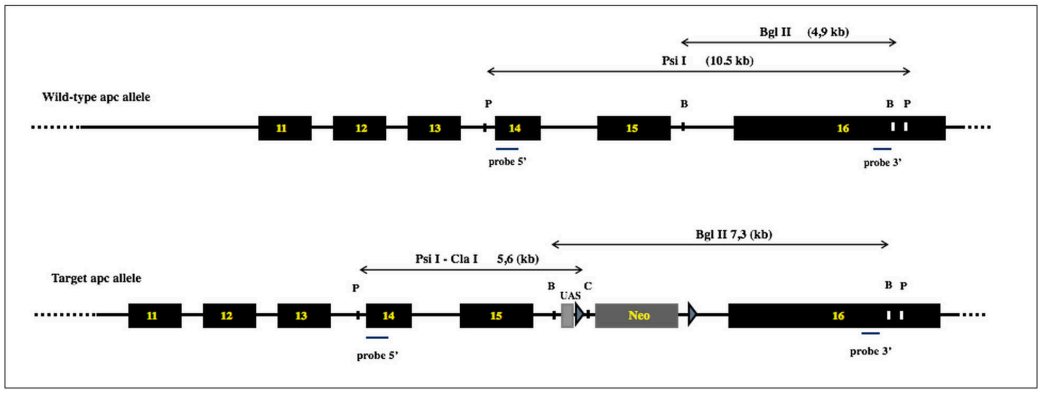


Figure 34: Strategy for the generation of *Apc*^{UAS} allele via homologous recombination.

The targeting vector pLS7106 and the genomic organization of *Apc* are shown at the top. pLS7106 contains three tandem UAS sites fused to intron 15 of *Apc* and a Neo cassette flanked by two *loxP* sites, imbedded within the homologous targeting sequences. Theoretical crossover between endogenous *Apc* and the linearized vector is depicted by the crossed lines. After homologous recombination, the UAS sequence with the Neo Cassette have been integrated in intron 15. The Neo cassette is removed by crossing the founder *Apc*^{UASNeo} mouse with a Deleter-Cre mouse.

As already described for *p53*$^{UAS/+}$ knock-in mouse generation, a classical targeting construct was made (pLS7106) and electroporated into E14tg2α ES cells. Of 480 clones screened by southern blot analysis (Figure 35A), only one had recombined correctly the vector into *Apc* gene (Figure 35B). The positive ES clone will be subjected to karyotype analysis and injected into C57BL/6J blastocyst to generate chimeras.

A



B

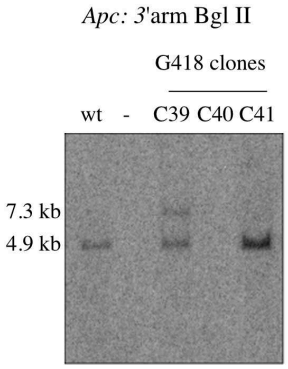


Figure 35: Southern blot analysis of electroporated ES cells with pLS7106.

A) Diagram showing probe design for *Apc* targeting: digestion with Psi I and Cla I and hybridization with 5' genomic probe resulted in a fragment of 10.5 kb for endogenous wt allele and 5.6 kb for the knock-in allele. Hybridization with 3' probe after Bgl II digestion results in a wt fragment of 4.9 kb and a recombined allele of 7.3 kb. B) ES cell DNA (3 µg) extracted from G418 resistant clones was digested with Bgl II, run on a 0.8 % agarose gel, blotted onto a nylon membrane and hybridized with ^{32}P labeled 5' probe. Clone C39 has correctly homologous recombined the vector into *Apc* gene.

CHAPTER 4 - DISCUSSION

Cancer is the second most common cause of death after cardiovascular diseases and accounts for 1 in 4 deaths in Europe (Ferlay et al., 2013). Cancer is an evolutionary process, characterized by the accumulation of mutations in individual cells over time, resulting in uncontrolled growth and spread of abnormal cells. 78 % of all cancers diagnosed are in patients of 55 years old or older. Given the long time (more that 20 years) until a tumor is clinically detected, it is reasonable to think that thousands of mutations have accumulated in tumor cells from the time an individual is exposed to a DNA damaging source to the clinical detection of the tumor. Importantly, animal models have contributed to the understanding of the molecular and cellular mechanisms underlying tumor initiation and progression, but the majority of these approaches have not lead to full disease recapitulation.

In the present study, I developed a new model that mimics the process of mutation accumulation during human cancer development and progression. Fusing the natural mutator AID to sequence-specific binding proteins, I could demonstrate AID-induced low frequency mutagenesis at the *p53 locus* of mouse cell lines and full organisms, thereby affecting cellular integrity and moreover inducing cellular transformation.

4.1 ESTABLISHMENT OF A NEW TOOL FOR LOW-FREQUENCY MUTAGENESIS *IN VITRO*

To study the process of low-frequency mutagenesis *in vitro*, the mouse DNA deaminase Aid was fused to a sequence specific TALE, customized to target intron 4 of the *p53 locus*. *TP53* is the most frequent mutated gene in human cancers, with the majority of the

mutations affecting the DBD. The choice of intron 4 ensured the T-Aid to target mostly the p53 DBD, which is encoded by exon 4 – 8, a genomic region of 2.4 kb.

The ability of Aid to retain the catalytic activity when fused to a DNA binding protein was demonstrated in a previous work of our laboratory, where Gal4-Aid mediated the demethylation of ~ 1.2 kb region of the mouse *H19 locus* (Franchini et al., 2014).

The choice of the TALE as the tool to target Aid at the *p53 locus* was made based on the following considerations: 1) in contrast to the Gal4 binding protein that needs the specific UAS sequence to bind DNA (Ornitz et al., 1991), TALE can recognize any sequence of interest with high specificity and high affinity, customizing its central amino acid repeats responsible for DNA binding (Boch et al., 2009; Meckler et al., 2013; Moscou and Bogdanove, 2009); 2) this tool facilitates the targeting of specific endogenous *loci* both in cell lines (Hockemeyer et al., 2011; Maeder et al., 2013b) and organisms (Sung et al., 2013; Zhang et al., 2014); and 3) the engineering of TALE proteins is cost and time-effective as compared to ZFP, where imperfections of the modularity with particular domains lead to the lack of high-affinity binding to all DNA triplets.

Aid was chosen to induce local DNA lesions at the *p53 locus* because, aside from its central role in the immune system, is capable of mutating DNA in almost any living cells. If deregulated, it was shown to induce mutations in key tumor suppressor genes, including *TP53* (Beale et al., 2004; Lindley, 2013), leading to cancer progression (Matsumoto et al., 2007; Shimizu et al., 2014), and the C>T DNA deaminase mutational signature was identified in many cancers (Burns et al., 2013b; Nik-Zainal et al., 2012). Furthermore, the deamination of cytosine to uracil is a natural occurring event reverted via physiological pathways, such as BER and MMR. Importantly, previous work demonstrated that the error-prone MMR is activated in non-B cells upon AID-damage (Pena-Diaz et al., 2012) and (Incorvaia E., and Petersen-Mahrt S.K. unpublished), possibly explaining the detrimental effect of AID expression outside the immune system.

Thus, the creation of TALE-Aid fusion proteins provides a novel tool in the field of genome editing, that enables the targeting of a natural mutagen to an endogenous *locus,* inducing local DNA lesions and cellular transformation, leading to the possibility of uncovering novel mutations important for human cancers.

TALE proteins, if targeted to a transcribed region, can block transcription at the target site through steric hindrance (Li et al., 2015). Similar levels of *p53* transcript were observed in mock and TALE-Aid transfected NIH3T3 cells (Figure 12), suggesting that transcriptional elongation of RNAPII was not affected by TALE proteins in our *in vitro* system.

Transient transfections of NIH3T3 cells with T-Aid and deep sequencing of the *p53* mRNA after 48 h, revealed the presence of 37 unique point mutations in T-Aid sample (Figure 13C). Importantly, this analysis was carried out with two recently developed NGS complementary approaches, Ion PGM™ and Ion Proton™ Systems. Although they differed in the throughput per run and the fidelity of the DNA polymerase used, both were able to detect 19 common mutations in the transfected NIH3T3 cells (Table 12), with a frequency of ~ 3.9 – 6.5x10^{-5} mutations per bp. Our results are in agreement with previous work that estimated a mutation rate of ~ 1– 90x10^{-5} mutations per bp in non-*Ig* genes, with the mutation rate in the oncogene *BCL-6* of 9x10^{-4} (Liu et al., 2008), much lower compared to the mutation rate at the *Ig locus* during SHM (~10^{-3} mutations per bp) (Sale and Neuberger, 1998). This difference is explained by the fact that non-*Ig* genes are protected from SHM by high fidelity repair of AID-dependent damage.

The types of mutations that AID induces on ssDNA have been extensively studied, which include transitions at dC in the sequence context of WRC (W= dA or dT; R= dA or dG) (Beale et al., 2004; Morgan et al., 2004; Pham et al., 2003). Importantly, in our analysis mutagenesis at dC/dG sites was observed (77.7 %) more often than expected (56 %), as were transitions at dC/dGs (86 % observed vs 33 % expected) and WRC (26 % observed vs 11 % expected) (Table 11), indicating that T-Aid functions at the *p53 locus*.

The majority of the mutations found in our mutational analysis of *p53* mRNA were located in the DBD (23/37) (Table 12), with the remaining distributed in the N-terminus and the oligomerization domain. The spectrum of p53 mutations can be explained by the previous findings that the DBD resides in the fragile central region of p53, where each residue can be found mutated at least once in each human cancer, while the N- and C-termini are less sensitive to amino acids substitution (Kato et al., 2003; Soussi et al., 2010). Importantly, some of the mutations discovered in our analysis were conserved across human cancers as analyzed in the TP53 mutation database (Edlund et al., 2012; Soussi et al., 2010) (e.g. R279C, M243I), suggesting that AID can target and mutate hotspot residues in human cancers. Our hypothesis is supported by several studies showing a C>T transition pattern of *p53* mutations in human cancers (Beale et al., 2004; Lindley, 2013; Shimizu et al., 2014). The mutations observed in our study were not situated in PTMs sites affecting p53 binding to its partners (Meek and Anderson, 2009), but could impair the binding to DNA, or p53 tetramerization, thus affecting the overall function. Aside from these frequently observed mutations, other altered amino acids were observed in our analysis and were rarely or never found in human cancers (e.g. N236S and R280S), underling the importance of our approach in defining new mutations possibly relevant for cancer progression.

In our previous work, Aid fused to the DNA binding domain of Gal4, was shown to mediate demethylation ~ 500 bp upstream / downstream of its original target site in the *H19 locus* of mice (Franchini et al., 2014). Importantly, our positional analysis identified mutations ~ 600 bp upstream and ~ 2 kb downstream of the genomic target site, highlighting the ability of Aid to deaminate DNA at long distance from its initial target site, which in our case spans the majority of the *locus*.

Although our analysis indicated that T-Aid is able to induce mutations in the *p53 locus* at relevant sites affecting p53 activity, when we tested the effect of T-Aid on cellular transformation by soft-agar assay, we could not observe any significant increase in

transformed clones (Figure 15). A possible explanation could be that the low frequency rate of mutations observed in our experiment was not sufficient to cause cellular transformation, suggesting that only one allele was targeted. To circumvent this possible problem, we determined the ability of *p53*$^{+/-}$ pMEFs that have been transiently transfected with T-Aid to undergo cellular senescence. Explanted MEFs initially continue to replicate *in vitro*, but then rapidly undergo stress-associated senescence due to *in vitro* conditions, especially oxidative stress caused by supraphysiological levels of O_2 supplied during standard culturing conditions (Parrinello et al., 2003). However, if the p53 pathway is damaged, MEFs can overcome this replication block, leading to gradual outgrowth of an immortalized population (Harvey and Levine, 1991). Our preliminary results showed a total of 11 clones (9 %) arising from *p53*$^{+/-}$ pMEFs transfected with T-Aid-wt and 2 clones (2.4 %) from *p53*$^{+/-}$ pMEFs transfected with T-Aid-mt. Of the wt clones, only two were expressing *p53* mRNA (Figure 18). Importantly, these two clones contained a 10 bp-deletion in exon 3 that was predicted to form a truncated protein of 46 amino acids, with only the transactivation domain included. At the 5' end of this deletion a WRC motif was followed by a palindromic sequence forming a secondary structure (Figure 19). We hypothesize that T-Aid in *p53*$^{+/-}$ pMEFs deaminated the cytosine at the WRC on ssDNA and the fragile structure downstream of the lesion formed a stem-loop. The deletion occurred probably as the result of DNA polymerase slippage during replication. Future experiments will be important to confirm our findings.

Overall, we have established a novel system to study the low-frequency mutagenesis in cancer genes *in vitro*, with the possibility of uncovering novel important mutations.

4.2 TALE-AID INDUCED LOW FREQUENCY MUTAGENESIS IN ZEBRAFISH EMBRYOS

To study the process of low-frequency mutagenesis in a living organism, we expressed EGFP-T-AID fusion proteins in zebrafish embryos, aimed at targeting intron 6 of *p53*. In recent years, zebrafish have become an established animal model for human diseases, including cancer. To test if EGFP-T-AID could alter *p53* expression possibly leading to developmental defects, EGFP-T-AID was monitored in the developing embryos, that were categorized 24 h after injection according to their phenotype: normal, mild, severe and dead (Figure 24). A significant increase of mild, severe and dead embryos due to EGFP-T-AIDwt expression was observed in one injection, but another injection did not show any differences (Table 14). These opposing results might be explained by two different hypothesis, 1) TALE-AID mutates *p53*, or 2) TALE physically acts as a roadblock for *p53* transcription (Li et al., 2015), both leading to p53-dependent developmental defects. *P53* is expressed early during zebrafish development, with maternal *p53* highly and ubiquitously expressed at 1 hpf, while at the pharyngula stage (24 – 48 hpf) *p53* is predominantly expressed in the head region (Cheng et al., 1997; Thisse et al., 2000). Several studies conducted in animal models reported that p53 has other functions above its fundamental role in the induction of apoptosis and growth arrest. For instance, in zebrafish embryos, the expression of a p53 isoform Δ113p53, but not wild type p53, was shown to arrest the development of the digestive tract at 3 dpf (Chen et al., 2005). In mice, $p53^{-/-}$ female embryos exhibit significant developmental abnormalities, including profound neural-tube defects associated with overgrowth of neural tissues and craniofacial malformations, leading to frequent (23 %) embryonic lethality (Armstrong et al., 1995; Sah et al., 1995). In addition, p53-deficient *Xenopus laevis* embryos show inhibition of mesodermal differentiation and severe gastrulation defects (Cordenonsi et al., 2003). Importantly, increased p53 levels due to the disruption of Mdm2, lead to embryonic developmental

defects both in mouse and zebrafish causing lethality (Jones et al., 1995; Montes de Oca Luna et al., 1995; Thisse et al., 2000).

To test our first hypothesis, a mutational analysis on injected embryos exhibiting the mild phenotype at 24 hpf was performed. Ion Proton™ NGS analysis revealed 10 unique mutations in embryos transiently injected with EGFP-T-AID embryos (Figure 25). Of these mutations, 3 were identified at WRC motif and 1 was a transition at dC/dG. Importantly, mutations spanned from exon 4 to exon 11, with a distance of up to 4 kb from the TALE target site. Based on these results it is possible that EGFP-T-AID induced low-frequency mutagenesis at the *p53 locus* in mild embryos, but we cannot establish a direct correlation between phenotype and mutation frequency. It will be interesting to test our second hypothesis, if the TALE on its own can physically alter *p53* expression, leading to embryonic developmental impairment.

4.3 PHENOTYPIC ANALYSIS OF TALE-AID ZEBRAFISH LINES

To study the tumorigenic potential of AID-induced mutagenesis, we generated a zebrafish line stably expressing EGFP-T-AID using the Tol2-mediated transposon host approach. The lack of efficient offspring generation (Table 16) could be due to p53-dependent developmental defects, as already hypothesized above. However, we cannot exclude the involvement of other factors, such as food up-take, slight changes in the environmental water (pH), and other environmental stress.

Genomic analysis of F_1 putative positive embryos revealed that the integration of the construct into the germ cell was heterogeneous (10 – 33 %) - as expected (Urasaki et al., 2006) (Table 16), and at least one founder fish for each aimed transgenic line was identified. When we initiated a breeding program with female founders to establish the transgenic F_1 generation, we observed a strong increase of non-fertilized and dead eggs in tg(EF1α-EGFP-Tg-AIDwt) animals compared to tg(EF1α-EGFP-Tg-AIDmt) and AB

females (Table 17). It is known from previous work that p53 is important for maternal fertility in mammals, by regulating the expression of the leukemia inhibitory factor (LIF), a cytokine crucial for blastocyst implantation. Loss of p53 in female mice leads to a decrease in fertility, due to the inefficient implantation (Hu et al., 2007), while in humans a common coding SNP at codon 72 results in either arginine (R72) or proline (P72), with the P72 allele associated with implantation defects in women (Kang et al., 2009; Kay et al., 2006). In three *p53* zebrafish mutants described so far (*tp53*^{N168K}, *tp53*^{M214K} and *tp53*^{I166T}) normal fertility was observed (Berghmans et al., 2005; Parant et al., 2010). Although zebrafish fertilization takes place *ex utero*, and many external factors could eventually influence this process (e.g. environmental water temperature and pH), we hypothesized that EGFP-T-AID induced alterations in *p53*, leading to impaired spawning of mature oocytes. In agreement with these findings, the histological analysis of the F_1 reproductive system of tg(EF1α-EGFP-Tg-AIDwt) females revealed an atypical stromal structure, with disorganized and fibrotic-like tissue surrounding the developing ovarian follicles (Figure 30). These results let us speculate that the impairment in spawning fertilized eggs is correlated to the pathological features observed in the reproductive organ. In support of this hypothesis, several studies showed that apoptosis plays an important role in the maturation of testes and ovaries in mammals and fish. During gonadal development in juvenile zebrafish (20 dpf), all individuals harbor undifferentiated ovary-like gonads (juvenile hermaphroditism) (Takahashi, 1977). In male zebrafish, oocytes disappear from the gonads by p53-mediated apoptosis between 21 – 25 dpf and spermatocytes develop (Rodriguez-Mari et al., 2010; Uchida et al., 2002), while in females, a programmed loss of oocytes occurs during ovarian development, where apoptotic cells are observed in the stromal tissue neighboring the developing oocytes by 40 dpf (Uchida et al., 2002). Moreover, during atresia of the pre-ovulatory follicles, hypertrophy and hyperplasia of the granulosa and theca cells within the stromal tissue are observed, followed by invasion and

absorption of the degenerated oocyte (Thome et al., 2009; Üçüncü and Çakıcı, 2009). Mature oocytes that escape atresia are extruded from the surrounding follicle cells (ovulation) and ready to spawn, with the stromal tissue undergoing structural changes to allow for the oocyte release (Patiño, 2002). Although many factors are involved in this process, including hormones (Chiang et al., 2001; Piferrer, 2001) and environmental stress (elevated water temperature) (Abozaid et al., 2012), I hypothesize that EGFP-T-AID dependent deregulated expression of p53 could alter a) the apoptotic process during ovarian maturation, leading to the production of dead and aberrant eggs, b) hyperplasia of the connective tissue. Indeed, these structural changes could cause the retention of the mature eggs during spawning, a process observed several times in our fertility analysis (Table 17). Based on the previous *p53* zebrafish models that develop sarcomas in the abdomen, the eye and the flank with variable penetrance (Berghmans et al., 2005; Parant et al., 2010), I can also hypothesize that the hyperplastic connective tissue observed in tg(EF1α-EGFP-Tg-AIDwt) females are derived from a pre-neoplastic lesion. However, given that none of the females analyzed at 12 mpf (0/3) developed the transformed-like phenotype in the ovaries (Table 18), further analysis on a larger cohort of animals should be performed to confirm our hypothesis.

4.4 FUTURE PLANS

4.4.1 Low frequency mutagenesis *in vitro*

In this study, I was able to develop a new tool based on TALE-Aid ability to target and mutate with low frequency the *p53 locus* in cell lines. Aside from TALE-Aid fusion protein, we have created dCas9-Aid-wt and mt fusion proteins and validated Aid activity with the Rif-mutagenesis assay in *E. coli*. We designed three different gRNAs to target intron 4 of mouse *p53*, and to further validate our results obtained with TALE-Aid, I will transfect the NIH3T3 cells with dCas9-Aid and perform a mutational analysis with Ion Torrent™ NGS. To assess the specific binding of our fusion proteins to the *p53 locus*, I will perform a Chromatin Immunoproecipitation (ChIP) on NIH3T3 cells transfected with TALE-Aid and dCas9-Aid. Considering the simplicity of the CRISPR/Cas9 system as compared to the TALE system, I will extend this study to other tumor suppressor genes, creating a library of gRNA targeted to multiple *loci* and to human cell lines.

4.4.2 Low frequency mutagenesis in zebrafish embryos

I have shown that EGFP-T-AID is able to target the *p53 locus* of zebrafish embryos causing developmental defects in the first 24 hpf. However, the mutation frequency observed in the embryos was significantly lowered compared to targeting the *p53 locus* in the murine NIH3T3 cells. To better understand this process, I will inject AB fertilized eggs with EGFP-T-AID and I will perform a mutational analysis on mild and severe phenotype embryos after 24 hpf. Moreover, to understand if the TALE protein can induce a transcriptional alteration of *p53,* I will inject fertilized eggs with EGFP-TALE construct and *p53* expression of both EGFP-T-AID and EGFP-TALE expressing embryos will be evaluated by RT-qPCR. To further analyze p53-mediated activity, I will treat EGFP-T-AID injected embryos with γ radiation at 24 hpf and evaluate their apoptotic response by

TUNEL assay at 30 hpf. If p53 retains its wild type activity, irradiation will result in massive apoptosis. On the contrary, if EGFP-T-AID embryos express a mutant p53, a block in the apoptotic response will be observed.

4.4.3 Low frequency mutagenesis in trangenic zebrafish lines

In this study I have established a new zebrafish transgenic line tg(EF1α-EGFP-Tg-AIDwt), with ovarian maturation defects. To make our histological analysis more significant, I will increase the number of analyzed animals and include other time points. In particular, considering that ovarian maturation starts at ~ 1 mpf, female will be sacrificed to analyze the phenotypic changes during early development. To further understand if the phenotype I observed is gender-specific, I will extend the histologic analysis to adult males. Moreover, because we observed a shorter lifespan in some of our founder animals, I will assess a lifespan analysis increasing the number of animals.

To better understand the role of p53 in our transgenic animals, I will evaluate *p53* mRNA and protein expression in the reproductive system of tg(EF1α-EGFP-Tg-AID) females, and perform a mutation analysis on the pathological tissue biopsy. In addition, pathological tissues will be analyzed for proliferation markers, such as PCNA or KI67. Moreover, to assess if EGFP-T-AID is able to accelerate the process of neoplastic transformation, I will cross tg(EF1α-EGFP-Tg-AIDwt) animals with *p53* mutant animals ($p53^{M214K/+}$ and $p53^{I166T/+}$), and monitor tumor formation and tumor onset.

4.4.4 A mouse model of Aid-induced low frequency mutagenesis

I have previously demonstrated that T-Aid is capable of mutating an endogenous *locus* both in cell lines and in a living organism. Moreover, our previous findings of Gal4-Aid/$H19^{UAS}$ animal model confirm that Aid fused to a DNA binding protein retains the

catalytic activity in mice. Due to unknown reasons, we have previously failed in the establishment of a *p53* $^{UAS/+}$ transgenic line and subsequently of the Gal4-Aid/*p53* $^{UAS/+}$ transgenic animals.

Considering our mutational analysis *in vitro*, I would like to generate T-Aid transgenic mice, with T-Aid targeted to the *p53 locus*. To overcome the limitation of only one allele being targeted, this mouse will be crossed to $p53^{+/-}$ mice. I will start to monitor for tumor formation at different time points. Based on the tumor spectra present in $p53^{-/-}$ mice, I expect to observe a variety of tumors in T-Aid mice, such as lymphomas and sarcomas that will be characterized by immunohistochemistry and FACS analysis, and sequenced to find new/known mutations in *p53*. In addition, blood samples will be collected at different time points and NGS analysis will be performed to find *p53* mutations during neoplastic transformation prior to full tumor formation. It was recently shown in humans, that circulating tumor DNA is found in patient's blood stream prior to the clinical detection of the tumor (Forshew et al., 2012). The analysis on circulating tumor DNA in our models will allow us to follow mutation accumulation due to Aid activity.

In parallel, I would like to generate dCas9-AID transgenic animals, which will allow us to target many *loci*, using different gRNAs.

CHAPTER 5 – REFERENCES

Abozaid, H., Wessels, S., and Horstgen-Schwark, G. (**2012**). Elevated temperature applied during gonadal transformation leads to male bias in zebrafish (Danio rerio). Sex Dev *6*, 201-209.

Aguilera, A., and Gomez-Gonzalez, B. (**2008**). Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet *9*, 204-217.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., *et al*. (**2013**). Signatures of mutational processes in human cancer. Nature *500*, 415-421.

Arakawa, H., Hauschild, J., and Buerstedde, J.M. (**2002**). Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science *295*, 1301-1306.

Armstrong, J.F., Kaufman, M.H., Harrison, D.J., and Clarke, A.R. (**1995**). High-frequency developmental abnormalities in p53-deficient mice. Curr Biol *5*, 931-936.

Bak, S.T., Sakellariou, D., and Pena-Diaz, J. (**2014**). The dual nature of mismatch repair as antimutator and mutator: for better or for worse. Front Genet *5*, 287.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (**2007**). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Barreto, V., Reina-San-Martin, B., Ramiro, A.R., McBride, K.M., and Nussenzweig, M.C. (**2003**). C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. Mol Cell *12*, 501-508.

Basu, U., Franklin, A., and Alt, F.W. (**2009**). Post-translational regulation of activation-induced cytidine deaminase. Philos Trans R Soc Lond B Biol Sci *364*, 667-673.

Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., *et al*. (**2011**). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. Cell *144*, 353-363.

Beale, R.C., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C., and Neuberger, M.S. (**2004**). Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J Mol Biol *337*, 585-596.

Beckman, K.B., and Ames, B.N. (**1997**). Oxidative decay of DNA. J Biol Chem *272*, 19633-19636.

Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F., 3rd (**1998**). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl

zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci U S A *95*, 14628-14633.

Begum, N.A., Kinoshita, K., Kakazu, N., Muramatsu, M., Nagaoka, H., Shinkura, R., Biniszkiewicz, D., Boyer, L.A., Jaenisch, R., and Honjo, T. (**2004**). Uracil DNA glycosylase activity is dispensable for immunoglobulin class switch. Science *305*, 1160-1163.

Berens, C., and Hillen, W. (**2003**). Gene regulation by tetracyclines. Constraints of resistance regulation in bacteria shape TetR for application in eukaryotes. Eur J Biochem *270*, 3109-3121.

Berghmans, S., Murphey, R.D., Wienholds, E., Neuberg, D., Kutok, J.L., Fletcher, C.D., Morris, J.P., Liu, T.X., Schulte-Merker, S., Kanki, J.P., *et al*. (**2005**). tp53 mutant zebrafish develop malignant peripheral nerve sheath tumors. Proc Natl Acad Sci U S A *102*, 407-412.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (**2010**). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (**2013**). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res *41*, 7429-7437.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (**2009**). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Bodmer, W.F., Bailey, C.J., Bodmer, J., Bussey, H.J., Ellis, A., Gorman, P., Lucibello, F.C., Murday, V.A., Rider, S.H., Scambler, P., *et al*. (**1987**). Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature *328*, 614-616.

Branzei, D., and Foiani, M. (**2005**). The DNA damage response during DNA replication. Curr Opin Cell Biol *17*, 568-575.

Brar, S.S., Watson, M., and Diaz, M. (**2004**). Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. J Biol Chem *279*, 26395-26401.

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., *et al*. (**2013a**). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature *494*, 366-370.

Burns, M.B., Temiz, N.A., and Harris, R.S. (**2013b**). Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet *45*, 977-983.

Cascalho, M., Wong, J., Steinberg, C., and Wabl, M. (**1998**). Mismatch repair co-opted by hypermutation. Science *279*, 1207-1210.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (**2011**). Efficient design and assembly

of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res *39*, e82.

Chahwan, R., Edelmann, W., Scharff, M.D., and Roa, S. (**2012**). AIDing antibody diversity by error-prone mismatch repair. Semin Immunol *24*, 293-300.

Chehab, N.H., Malikzay, A., Appel, M., and Halazonetis, T.D. (**2000**). Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53. Genes Dev *14*, 278-288.

Chen, J., Ruan, H., Ng, S.M., Gao, C., Soo, H.M., Wu, W., Zhang, Z., Wen, Z., Lane, D.P., and Peng, J. (**2005**). Loss of function of def selectively up-regulates Delta113p53 expression to arrest expansion growth of digestive organs in zebrafish. Genes Dev *19*, 2900-2911.

Chen, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R., *et al*. (**2015**). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell *160*, 1246-1260.

Chen, Y., Williams, V., Filippova, M., Filippov, V., and Duerksen-Hughes, P. (**2014**). Viral carcinogenesis: factors inducing DNA damage and virus integration. Cancers *6*, 2155-2186.

Cheng, R., Ford, B.L., O'Neal, P.E., Mathews, C.Z., Bradford, C.S., Thongtan, T., Barnes, D.W., Hendricks, J.D., and Bailey, G.S. (**1997**). Zebrafish (Danio rerio) p53 tumor suppressor gene: cDNA sequence and expression during embryogenesis. Mol Mar Biol Biotechnol *6*, 88-97.

Cheon, D.J., and Orsulic, S. (**2011**). Mouse models of cancer. Annu Rev Pathol *6*, 95-119.

Chiang, E.F., Yan, Y.L., Guiguen, Y., Postlethwait, J., and Chung, B. (**2001**). Two Cyp19 (P450 aromatase) genes on duplicated zebrafish chromosomes are expressed in ovary or brain. Mol Biol Evol *18*, 542-550.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (**2010**). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics *186*, 757-761.

Ciccia, A., and Elledge, S.J. (**2010**). The DNA damage response: making it safe to play with knives. Molecular cell *40*, 179-204.

Cleaver, J.E. (**2005**). Cancer in xeroderma pigmentosum and related disorders of DNA repair. Nat Rev Cancer *5*, 564-573.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al*. (**2013**). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M., and Zhang, F. (**2012**). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun *3*, 968.

Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K., and Neuberger, M.S. (**2005**). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol *22*, 367-377.

Cordenonsi, M., Dupont, S., Maretto, S., Insinga, A., Imbriano, C., and Piccolo, S. (**2003**). Links between tumor suppressors: p53 is required for TGF-beta gene responses by cooperating with Smads. Cell *113*, 301-314.

Cottrell, S., Bicknell, D., Kaklamanis, L., and Bodmer, W.F. (**1992**). Molecular analysis of APC mutations in familial adenomatous polyposis and sporadic colon carcinomas. Lancet *340*, 626-630.

Dasgupta, D., Majumder, P., and Banerjee, A. (**2012**). A revisit of the mode of interaction of small transcription inhibitors with genomic DNA. J Biosci *37*, 475-481.

de Yebenes, V.G., Belver, L., Pisano, D.G., Gonzalez, S., Villasante, A., Croce, C., He, L., and Ramiro, A.R. (**2008**). miR-181b negatively regulates activation-induced cytidine deaminase in B cells. J Exp Med *205*, 2199-2206.

Di Noia, J.M., and Neuberger, M.S. (**2007**). Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem *76*, 1-22.

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (**2013**). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res *41*, 4336-4343.

Dickerson, S.K., Market, E., Besmer, E., and Papavasiliou, F.N. (**2003**). AID mediates hypermutation by deaminating single stranded DNA. J Exp Med *197*, 1291-1296.

Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgomery, C.A., Jr., Butel, J.S., and Bradley, A. (**1992**). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature *356*, 215-221.

Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., Vandyk, J.K., and Bogdanove, A.J. (**2012**). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res *40*, W117-122.

Duffy, J.B. (**2002**). GAL4 system in Drosophila: a fly geneticist's Swiss army knife. Genesis *34*, 1-15.

Durkin, S.G., and Glover, T.W. (**2007**). Chromosome fragile sites. Annu Rev Genet *41*, 169-192.

Edlund, K., Larsson, O., Ameur, A., Bunikis, I., Gyllensten, U., Leroy, B., Sundstrom, M., Micke, P., Botling, J., and Soussi, T. (**2012**). Data-driven unbiased curation of the TP53 tumor suppressor gene mutation database and validation by ultradeep sequencing of human tumors. Proc Natl Acad Sci U S A *109*, 9551-9556.

Endo, Y., Marusawa, H., Kinoshita, K., Morisawa, T., Sakurai, T., Okazaki, I.M., Watashi, K., Shimotohno, K., Honjo, T., and Chiba, T. (**2007**). Expression of activation-induced cytidine deaminase in human hepatocytes via NF-kappaB signaling. Oncogene *26*, 5587-5595.

Endo, Y., Marusawa, H., Kou, T., Nakase, H., Fujii, S., Fujimori, T., Kinoshita, K., Honjo, T., and Chiba, T. (**2008**). Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. Gastroenterology *135*, 889-898, 898 e881-883.

Epeldegui, M., Hung, Y.P., McQuay, A., Ambinder, R.F., and Martinez-Maza, O. (**2007**). Infection of human B cells with Epstein-Barr virus results in the expression of somatic hypermutation-inducing molecules and in the accrual of oncogene mutations. Mol Immunol *44*, 934-942.

Fearnhead, N.S., Britton, M.P., and Bodmer, W.F. (**2001**). The ABC of APC. Hum Mol Genet *10*, 721-733.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J.W., Comber, H., Forman, D., and Bray, F. (**2013**). Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer *49*, 1347-1403.

Fields, S., and Jang, S.K. (**1990**). Presence of a potent transcription activating sequence in the p53 protein. Science *249*, 1046-1049.

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S.*, et al*. (**2015**). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res *43*, D805-811.

Forshew, T., Murtaza, M., Parkinson, C., Gale, D., Tsui, D.W., Kaper, F., Dawson, S.J., Piskorz, A.M., Jimenez-Linan, M., Bentley, D.*, et al*. (**2012**). Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med *4*, 136ra168.

Fousteri, M., and Mullenders, L.H. (**2008**). Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. Cell Res *18*, 73-84.

Franchini, D.M., Chan, C.F., Morgan, H., Incorvaia, E., Rangam, G., Dean, W., Santos, F., Reik, W., and Petersen-Mahrt, S.K. (**2014**). Processive DNA demethylation via DNA deaminase-induced lesion resolution. PLoS One *9*, e97754.

Franchini, D.M., Schmitz, K.M., and Petersen-Mahrt, S.K. (**2012**). 5-Methylcytosine DNA demethylation: more than losing a methyl group. Annu Rev Genet *46*, 419-441.

Frederico, L.A., Kunkel, T.A., and Shaw, B.R. (**1990**). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry *29*, 2532-2537.

Friedberg, E.C. (**2003**). DNA damage and repair. Nature *421*, 436-440.

Frosina, G., Fortini, P., Rossi, O., Carrozzino, F., Raspaglio, G., Cox, L.S., Lane, D.P., Abbondandolo, A., and Dogliotti, E. (**1996**). Two pathways for base excision repair in mammalian cells. J Biol Chem *271*, 9573-9578.

Furth, P.A., St Onge, L., Boger, H., Gruss, P., Gossen, M., Kistner, A., Bujard, H., and Hennighausen, L. (**1994**). Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. Proc Natl Acad Sci U S A *91*, 9302-9306.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (**2013**). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol *31*, 397-405.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (**2010**). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.

Gerhard, G.S., Kauffman, E.J., Wang, X., Stewart, R., Moore, J.L., Kasales, C.J., Demidenko, E., and Cheng, K.C. (**2002**). Life spans and senescent phenotypes in two strains of Zebrafish (Danio rerio). Exp Gerontol *37*, 1055-1068.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., *et al*. (**2013**). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451.

Gillet, L.C., and Scharer, O.D. (**2006**). Molecular mechanisms of mammalian global genome nucleotide excision repair. Chem Rev *106*, 253-276.

Gu, L., Hong, Y., McCulloch, S., Watanabe, H., and Li, G.M. (**1998**). ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. Nucleic Acids Res *26*, 1173-1178.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (**2005**). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol *1*, e60.

Hanahan, D., and Weinberg, R.A. (**2011**). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Harris, R.S., Sale, J.E., Petersen-Mahrt, S.K., and Neuberger, M.S. (**2002**). AID is essential for immunoglobulin V gene conversion in a cultured B cell line. Curr Biol *12*, 435-438.

Harvey, D.M., and Levine, A.J. (**1991**). p53 alteration is a common event in the spontaneous immortalization of primary BALB/c murine embryo fibroblasts. Genes Dev *5*, 2375-2385.

He, B., Raab-Traub, N., Casali, P., and Cerutti, A. (**2003**). EBV-encoded latent membrane protein 1 cooperates with BAFF/BLyS and APRIL to induce T cell-independent Ig heavy chain class switching. J Immunol *171*, 5215-5224.

Helleday, T. (**2010**). Homologous recombination in cancer development, treatment and development of drug resistance. Carcinogenesis *31*, 955-960.

Helleday, T., Eshtad, S., and Nik-Zainal, S. (**2014**). Mechanisms underlying mutational signatures in human cancers. Nature reviews. Genetics *15*, 585-598.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B., *et al*. (**2009**). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol *27*, 851-857.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., *et al*. (**2011**). Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol *29*, 731-734.

Hoeijmakers, J.H. (**2009**). DNA damage, aging, and cancer. The New England journal of medicine *361*, 1475-1485.

Horii, A., Nakatsuru, S., Ichii, S., Nagase, H., and Nakamura, Y. (**1993**). Multiple forms of the APC gene transcripts and their tissue-specific expression. Hum Mol Genet *2*, 283-287.

Hu, W., Feng, Z., Teresky, A.K., and Levine, A.J. (**2007**). p53 regulates maternal reproduction through LIF. Nature *450*, 721-724.

Hu, Y., Ericsson, I., Torseth, K., Methot, S.P., Sundheim, O., Liabakk, N.B., Slupphaug, G., Di Noia, J.M., Krokan, H.E., and Kavli, B. (**2013**). A combined nuclear and nucleolar localization motif in activation-induced cytidine deaminase (AID) controls immunoglobulin class switching. J Mol Biol *425*, 424-443.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., and Zhang, B. (**2011**). Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol *29*, 699-700.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (**2013**). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol *31*, 227-229.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (**1987**). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol *169*, 5429-5433.

Jacks, T., Remington, L., Williams, B.O., Schmitt, E.M., Halachmi, S., Bronson, R.T., and Weinberg, R.A. (**1994**). Tumor spectrum analysis in p53-mutant mice. Curr Biol *4*, 1-7.

Jackson, S.P., and Bartek, J. (**2009**). The DNA-damage response in human biology and disease. Nature *461*, 1071-1078.

Jacobs, A.L., and Schar, P. (**2012**). DNA glycosylases: in DNA repair and beyond. Chromosoma *121*, 1-20.

Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (**2002**). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol *43*, 1565-1575.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (**2012**). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jones, S.N., Roe, A.E., Donehower, L.A., and Bradley, A. (**1995**). Rescue of embryonic lethality in Mdm2-deficient mice by absence of p53. Nature *378*, 206-208.

Jung, D.W., Oh, E.S., Park, S.H., Chang, Y.T., Kim, C.H., Choi, S.Y., and Williams, D.R. (**2012**). A novel zebrafish human tumor xenograft model validated for anti-cancer drug screening. Mol Biosyst *8*, 1930-1939.

Kane, D.A., and Kimmel, C.B. (**1993**). The zebrafish midblastula transition. Development *119*, 447-456.

Kang, H.J., Feng, Z., Sun, Y., Atwal, G., Murphy, M.E., Rebbeck, T.R., Rosenwaks, Z., Levine, A.J., and Hu, W. (**2009**). Single-nucleotide polymorphisms in the p53 pathway regulate fertility in humans. Proc Natl Acad Sci U S A *106*, 9761-9766.

Kato, S., Han, S.Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R., and Ishioka, C. (**2003**). Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proc Natl Acad Sci U S A *100*, 8424-8429.

Kawakami, K. (**2007**). Tol2: a versatile gene transfer vector in vertebrates. Genome Biol *8 Suppl 1*, S7.

Kay, C., Jeyendran, R.S., and Coulam, C.B. (**2006**). p53 tumour suppressor gene polymorphism is associated with recurrent implantation failure. Reprod Biomed Online *13*, 492-496.

Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (**2007**). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science *318*, 648-651.

Kern, S.E., Kinzler, K.W., Bruskin, A., Jarosz, D., Friedman, P., Prives, C., and Vogelstein, B. (**1991**). Identification of p53 as a sequence-specific DNA-binding protein. Science *252*, 1708-1711.

Kim, H., Kim, K., Choi, J., Heo, K., Baek, H.J., Roeder, R.G., and An, W. (**2012**). p53 requires an intact C-terminal domain for DNA binding and transactivation. J Mol Biol *415*, 843-854.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W., and Kim, J.S. (**2009**). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res *19*, 1279-1288.

Kim, S.K., Marusawa, H., Eso, Y., Chiba, T., and Kudo, M. (**2013**). Novel mouse models of hepatocarcinogenesis with stepwise accumulation of genetic alterations. Dig Dis *31*, 454-458.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (**1996**). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A *93*, 1156-1160.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (**1995**). Stages of embryonic development of the zebrafish. Dev Dyn *203*, 253-310.

Kinzler, K.W., Nilbert, M.C., Su, L.K., Vogelstein, B., Bryan, T.M., Levy, D.B., Smith, K.J., Preisinger, A.C., Hedge, P., McKechnie, D., *et al*. (**1991**). Identification of FAP locus genes from chromosome 5q21. Science *253*, 661-665.

Knudson, A.G., Jr. (**1971**). Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A *68*, 820-823.

Komori, J., Marusawa, H., Machimoto, T., Endo, Y., Kinoshita, K., Kou, T., Haga, H., Ikai, I., Uemoto, S., and Chiba, T. (**2008**). Activation-induced cytidine deaminase links bile duct inflammation to human cholangiocarcinoma. Hepatology *47*, 888-896.

Kopetz, S., Lemos, R., and Powis, G. (**2012**). The promise of patient-derived xenografts: the best laid plans of mice and men. Clin Cancer Res *18*, 5160-5162.

Koster, R.W., and Fraser, S.E. (**2001**). Tracing transgene expression in living zebrafish embryos. Dev Biol *233*, 329-346.

Krokan, H.E., Drablos, F., and Slupphaug, G. (**2002**). Uracil in DNA--occurrence, consequences and repair. Oncogene *21*, 8935-8948.

Kubota, Y., Nash, R.A., Klungland, A., Schar, P., Barnes, D.E., and Lindahl, T. (**1996**). Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. The EMBO journal *15*, 6662-6670.

Kunkel, T.A., and Bebenek, K. (**2000**). DNA replication fidelity. Annu Rev Biochem *69*, 497-529.

Kunkel, T.A., and Erie, D.A. (**2005**). DNA mismatch repair. Annu Rev Biochem *74*, 681-710.

Lane, D.P. (**1992**). Cancer. p53, guardian of the genome. Nature *358*, 15-16.

Langerak, P., Nygren, A.O., Krijger, P.H., van den Berk, P.C., and Jacobs, H. (**2007**). A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. J Exp Med *204*, 1989-1998.

Lasken, R.S., and Goodman, M.F. (**1984**). The biochemical basis of 5-bromouracil-induced mutagenesis. Heteroduplex base mispairs involving bromouracil in G x C----A x T and A x T----G x C mutational pathways. J Biol Chem *259*, 11491-11495.

Lee, L.M., Seftor, E.A., Bonde, G., Cornell, R.A., and Hendrix, M.J. (**2005**). The fate of human malignant melanoma cells transplanted into zebrafish embryos: assessment of migration and cell division in the absence of tumor formation. Dev Dyn *233*, 1560-1570.

Lee, W.P., Stromberg, M.P., Ward, A., Stewart, C., Garrison, E.P., and Marth, G.T. (**2014**). MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. PLoS One *9*, e90581.

Lei, Y., Guo, X., Liu, Y., Cao, Y., Deng, Y., Chen, X., Cheng, C.H., Dawid, I.B., Chen, Y., and Zhao, H. (**2012**). Efficient targeted gene disruption in Xenopus embryos using engineered transcription activator-like effector nucleases (TALENs). Proc Natl Acad Sci U S A *109*, 17484-17489.

Leroy, B., Fournier, J.L., Ishioka, C., Monti, P., Inga, A., Fronza, G., and Soussi, T. (**2013**). The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. Nucleic Acids Res *41*, D962-969.

Li, F.P., Fraumeni, J.F., Jr., Mulvihill, J.J., Blattner, W.A., Dreyfus, M.G., Tucker, M.A., and Miller, R.W. (**1988**). A cancer family syndrome in twenty-four kindreds. Cancer Res *48*, 5358-5362.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (**2009**). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (**2011**). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res *39*, 359-372.

Li, X., and Heyer, W.D. (**2008**). Homologous recombination in DNA repair and DNA damage tolerance. Cell research *18*, 99-113.

Li, Y., Jiang, Y., Chen, H., Liao, W., Li, Z., Weiss, R., and Xie, Z. (**2015**). Modular construction of mammalian gene circuits using TALE transcriptional repressors. Nat Chem Biol *11*, 207-213.

Lieber, M.R. (**2010**). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem *79*, 181-211.

Lieschke, G.J., and Currie, P.D. (**2007**). Animal models of human disease: zebrafish swim into view. Nat Rev Genet *8*, 353-367.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K., *et al*. (**2009**). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069-1083.

Lin, J., Chen, J., Elenbaas, B., and Levine, A.J. (**1994**). Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to mdm-2 and the adenovirus 5 E1B 55-kD protein. Genes Dev *8*, 1235-1246.

Lindahl, T. (**1974**). An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. Proc Natl Acad Sci U S A *71*, 3649-3653.

Lindahl, T., and Barnes, D.E. (**2000**). Repair of endogenous DNA damage. Cold Spring Harbor symposia on quantitative biology *65*, 127-133.

Lindahl, T., and Nyberg, B. (**1974**). Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry *13*, 3405-3410.

Lindahl, T.N., B. (**1972**). Rate of depurination of native deoxyribonucleic acid. Biochemistry *11*, 3610-3618.

Lindley, R.A. (**2013**). The importance of codon context for understanding the Ig-like somatic hypermutation strand-biased patterns in TP53 mutations in breast cancer. Cancer Genet *206*, 222-226.

Liu, J., Li, C., Yu, Z., Huang, P., Wu, H., Wei, C., Zhu, N., Shen, Y., Chen, Y., Zhang, B., *et al*. (**2012**). Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. J Genet Genomics *39*, 209-215.

Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (**2008**). Two levels of protection for the B cell genome during somatic hypermutation. Nature *451*, 841-845.

Liu, Q., Guntuku, S., Cui, X.S., Matsuoka, S., Cortez, D., Tamai, K., Luo, G., Carattini-Rivera, S., DeMayo, F., Bradley, A., *et al*. (**2000**). Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. Genes Dev *14*, 1448-1459.

Livak, K.J., and Schmittgen, T.D. (**2001**). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Loeb, L.A., Springgate, C.F., and Battula, N. (**1974**). Errors in DNA replication as a basis of malignant changes. Cancer Res *34*, 2311-2321.

Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D., *et al*. (**2007**). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol *25*, 1298-1306.

Lozano, G. (**2010**). Mouse models of p53 functions. Cold Spring Harb Perspect Biol *2*, a001115.

Maeda, K., Singh, S.K., Eda, K., Kitabatake, M., Pham, P., Goodman, M.F., and Sakaguchi, N. (**2010**). GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. J Biol Chem *285*, 23945-23953.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (**2013a**). CRISPR RNA-guided activation of endogenous human genes. Nat Methods *10*, 977-979.

Maeder, M.L., Linder, S.J., Reyon, D., Angstman, J.F., Fu, Y., Sander, J.D., and Joung, J.K. (**2013b**). Robust, synergistic regulation of human gene expression using TALE activators. Nat Methods *10*, 243-245.

Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A., *et al*. (**2008**). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell *31*, 294-301.

Mahfouz, M.M., Li, L., Piatek, M., Fang, X., Mansour, H., Bangarusamy, D.K., and Zhu, J.K. (**2012**). Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. Plant Mol Biol *78*, 311-321.

Mailand, N., Falck, J., Lukas, C., Syljuasen, R.G., Welcker, M., Bartek, J., and Lukas, J. (**2000**). Rapid destruction of human Cdc25A in response to DNA damage. Science *288*, 1425-1429.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., *et al*. (**2015**). An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol *13*, 722-736.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (**2013**). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Mandell, J.G., and Barbas, C.F., 3rd (**2006**). Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. Nucleic Acids Res *34*, W516-523.

Marraffini, L.A., and Sontheimer, E.J. (**2010**). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet *11*, 181-190.

Matsumoto, Y., Marusawa, H., Kinoshita, K., Endo, Y., Kou, T., Morisawa, T., Azuma, T., Okazaki, I.M., Honjo, T., and Chiba, T. (**2007**). Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med *13*, 470-476.

Maynard, S., Schurman, S.H., Harboe, C., de Souza-Pinto, N.C., and Bohr, V.A. (**2009**). Base excision repair of oxidative DNA damage and association with cancer and aging. Carcinogenesis *30*, 2-10.

McCart, A.E., Vickaryous, N.K., and Silver, A. (**2008**). Apc mice: models, modifiers and mutants. Pathol Res Pract *204*, 479-490.

McLure, K.G., and Lee, P.W. (**1998**). How p53 binds DNA as a tetramer. EMBO J *17*, 3342-3350.

Meckler, J.F., Bhakta, M.S., Kim, M.S., Ovadia, R., Habrian, C.H., Zykovich, A., Yu, A., Lockwood, S.H., Morbitzer, R., Elsaesser, J., *et al*. (**2013**). Quantitative analysis of TALE-DNA interactions suggests polarity effects. Nucleic Acids Res *41*, 4118-4128.

Meek, D.W., and Anderson, C.W. (**2009**). Posttranslational modification of p53: cooperative integrators of function. Cold Spring Harb Perspect Biol *1*, a000950.

Meister, G.E., Chandrasegaran, S., and Ostermeier, M. (**2010**). Heterodimeric DNA methyltransferases as a platform for creating designer zinc finger methyltransferases for targeted DNA methylation in cells. Nucleic Acids Res *38*, 1749-1759.

Miyoshi, Y., Nagase, H., Ando, H., Horii, A., Ichii, S., Nakatsuru, S., Aoki, T., Miki, Y., Mori, T., and Nakamura, Y. (**1992**). Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. Hum Mol Genet *1*, 229-233.

Montes de Oca Luna, R., Wagner, D.S., and Lozano, G. (**1995**). Rescue of early embryonic lethality in mdm2-deficient mice by deletion of p53. Nature *378*, 203-206.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (**2004**). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem *279*, 52353-52360.

Moscou, M.J., and Bogdanove, A.J. (**2009**). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Moser, A.R., Pitot, H.C., and Dove, W.F. (**1990**). A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. Science *247*, 322-324.

Moser, A.R., Shoemaker, A.R., Connelly, C.S., Clipson, L., Gould, K.A., Luongo, C., Dove, W.F., Siggers, P.H., and Gardner, R.L. (**1995**). Homozygosity for the Min allele of Apc results in disruption of mouse development prior to gastrulation. Dev Dyn *203*, 422-433.

Msiska, Z., Pacurari, M., Mishra, A., Leonard, S.S., Castranova, V., and Vallyathan, V. (**2010**). DNA double-strand breaks by asbestos, silica, and titanium dioxide: possible biomarker of carcinogenic potential? Am J Respir Cell Mol Biol *43*, 210-219.

Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and Polakis, P. (**1995**). Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. Proc Natl Acad Sci U S A *92*, 3046-3050.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (**2000**). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553-563.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (**1999**). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem *274*, 18470-18476.

Nagase, H., and Nakamura, Y. (**1993**). Mutations of the APC (adenomatous polyposis coli) gene. Hum Mutat *2*, 425-434.

Nakamura, J., and Swenberg, J.A. (**1999**). Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. Cancer Res *59*, 2522-2526.

Nakamura, Y., Nishisho, I., Kinzler, K.W., Vogelstein, B., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., and Nagase, H. (**1991**). Mutations of the adenomatous polyposis coli gene in familial polyposis coli patients and sporadic colorectal tumors. Princess Takamatsu Symp *22*, 285-292.

Neuberger, M.S., Harris, R.S., Di Noia, J., and Petersen-Mahrt, S.K. (**2003**). Immunity through DNA deamination. Trends Biochem Sci *28*, 305-312.

Nichols, K.E., Malkin, D., Garber, J.E., Fraumeni, J.F., Jr., and Li, F.P. (**2001**). Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers. Cancer Epidemiol Biomarkers Prev *10*, 83-87.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., *et al*. (**2012**). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979-993.

Ornitz, D.M., Moreadith, R.W., and Leder, P. (**1991**). Binary system for regulating transgene expression in mice: targeting int-2 gene expression with yeast GAL4/UAS control elements. Proc Natl Acad Sci U S A *88*, 698-702.

Pabo, C.O., Peisach, E., and Grant, R.A. (**2001**). Design and selection of novel Cys2His2 zinc finger proteins. Annu Rev Biochem *70*, 313-340.

Parant, J.M., George, S.A., Holden, J.A., and Yost, H.J. (**2010**). Genetic modeling of Li-Fraumeni syndrome in zebrafish. Dis Model Mech *3*, 45-56.

Parrinello, S., Samper, E., Krtolica, A., Goldstein, J., Melov, S., and Campisi, J. (**2003**). Oxygen sensitivity severely limits the replicative lifespan of murine fibroblasts. Nat Cell Biol *5*, 741-747.

Pasqualucci, L., Guglielmino, R., Houldsworth, J., Mohr, J., Aoufouchi, S., Polakiewicz, R., Chaganti, R.S., and Dalla-Favera, R. (**2004**). Expression of the AID protein in normal and neoplastic B cells. Blood *104*, 3318-3325.

Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R.S., Klein, U., Kuppers, R., Rajewsky, K., *et al*. (**1998**). BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. Proc Natl Acad Sci U S A *95*, 11816-11821.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Kuppers, R., and Dalla-Favera, R. (**2001**). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature *412*, 341-346.

Patiño, R.S., C. V. (**2002**). Ovarian follicle growth, maturation, and ovulation in teleost fish. Fish Physiology and Biochemistry *26*, 57 - 70.

Pauklin, S., and Petersen-Mahrt, S.K. (**2009**). Progesterone inhibits activation-induced deaminase by binding to the promoter. J Immunol *183*, 1238-1244.

Pauklin, S., Sernandez, I.V., Bachmann, G., Ramiro, A.R., and Petersen-Mahrt, S.K. (**2009**). Estrogen directly activates AID transcription and function. J Exp Med *206*, 99-111.

Pavletich, N.P., and Pabo, C.O. (**1991**). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science *252*, 809-817.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., *et al*. (**2010**). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell *143*, 122-133.

Peltomaki, P. (**1997**). DNA mismatch repair gene mutations in human cancer. Environ Health Perspect *105 Suppl 4*, 775-780.

Peltomaki, P. (**2003**). Role of DNA mismatch repair defects in the pathogenesis of human cancer. J Clin Oncol *21*, 1174-1179.

Pena-Diaz, J., Bregenhorn, S., Ghodgaonkar, M., Follonier, C., Artola-Boran, M., Castor, D., Lopes, M., Sartori, A.A., and Jiricny, J. (**2012**). Noncanonical mismatch repair as a source of genomic instability in human cells. Mol Cell *47*, 669-680.

Petersen-Mahrt, S. (**2005**). DNA deamination in immunity. Immunol Rev *203*, 80-97.

Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (**2002**). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature *418*, 99-103.

Petitjean, A., Achatz, M.I., Borresen-Dale, A.L., Hainaut, P., and Olivier, M. (**2007a**). TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene *26*, 2157-2165.

Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S.V., Hainaut, P., and Olivier, M. (**2007b**). Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. Hum Mutat *28*, 622-629.

Pfeifer, G.P., You, Y.H., and Besaratinia, A. (**2005**). Mutations induced by ultraviolet light. Mutat Res *571*, 19-31.

Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (**2003**). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature *424*, 103-107.

Phillips, D.H., Hewer, A., Martin, C.N., Garner, R.C., and King, M.M. (**1988**). Correlation of DNA adduct levels in human lung with cigarette smoking. Nature *336*, 790-792.

Piferrer, F. (**2001**). Endocrine sex control strategies for the feminization of teleost fish. Aquaculture *197*, 229 - 281.

Politi, K., and Pao, W. (**2011**). How genetically engineered mouse tumor models provide insights into human cancers. J Clin Oncol *29*, 2273-2281.

Pope, C.A., 3rd, Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., and Thurston, G.D. (**2002**). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama *287*, 1132-1141.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (**2010**). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Prochnow, C., Bransteitter, R., Klein, M.G., Goodman, M.F., and Chen, X.S. (**2007**). The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature *445*, 447-451.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (**2013**). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Rada, C., Di Noia, J.M., and Neuberger, M.S. (**2004**). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. Mol Cell *16*, 163-171.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (**2008**). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201-1212.

Rajewsky, K. (**1996**). Clonal selection and learning in the antibody system. Nature *381*, 751-758.

Rajski, S.R., and Williams, R.M. (**1998**). DNA Cross-Linking Agents as Antitumor Drugs. Chem Rev *98*, 2723-2796.

Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A., and Nussenzweig, M.C. (**2004**). AID is required for c-myc/IgH chromosome translocations in vivo. Cell *118*, 431-438.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A.*, et al*. (**2008**). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell *135*, 1028-1038.

Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G.*, et al*. (**2013**). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet *45*, 970-976.

Rodriguez-Mari, A., Canestro, C., Bremiller, R.A., Nguyen-Johnson, A., Asakawa, K., Kawakami, K., and Postlethwait, J.H. (**2010**). Sex reversal in zebrafish fancl mutants is caused by Tp53-mediated germ cell apoptosis. PLoS Genet *6*, e1001034.

Romer, P., Hahn, S., Jordan, T., Strauss, T., Bonas, U., and Lahaye, T. (**2007**). Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. Science *318*, 645-648.

Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S.H., Masiarz, F.R., Munemitsu, S., and Polakis, P. (**1993**). Association of the APC gene product with beta-catenin. Science *262*, 1731-1734.

Ruggeri, B.A., Camp, F., and Miknyoczki, S. (**2014**). Animal models of disease: pre-clinical animal models of cancer and their applications and utility in drug discovery. Biochem Pharmacol *87*, 150-161.

Sablina, A.A., Budanov, A.V., Ilyinskaya, G.V., Agapova, L.S., Kravchenko, J.E., and Chumakov, P.M. (**2005**). The antioxidant function of the p53 tumor suppressor. Nature medicine *11*, 1306-1313.

Sah, V.P., Attardi, L.D., Mulligan, G.J., Williams, B.O., Bronson, R.T., and Jacks, T. (**1995**). A subset of p53-deficient embryos exhibit exencephaly. Nat Genet *10*, 175-180.

Sakamuro, D., Sabbatini, P., White, E., and Prendergast, G.C. (**1997**). The polyproline region of p53 is required to activate apoptosis but not growth arrest. Oncogene *15*, 887-898.

Sale, J.E., and Neuberger, M.S. (**1998**). TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. Immunity *9*, 859-869.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (**2011**). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res *39*, 9275-9282.

Sayegh, C.E., Quong, M.W., Agata, Y., and Murre, C. (**2003**). E-proteins directly regulate expression of activation-induced deaminase in mature B cells. Nat Immunol *4*, 586-593.

Scheer, N., and Campos-Ortega, J.A. (**1999**). Use of the Gal4-UAS technique for targeted gene expression in the zebrafish. Mech Dev *80*, 153-158.

Schreck, S., Buettner, M., Kremmer, E., Bogdan, M., Herbst, H., and Niedobitek, G. (**2006**). Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. J Pathol *210*, 26-31.

Schroeder, H.W., Jr., and Cavacini, L. (**2010**). Structure and function of immunoglobulins. J Allergy Clin Immunol *125*, S41-52.

Schwenk, F., Baron, U., and Rajewsky, K. (**1995**). A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. Nucleic Acids Res *23*, 5080-5081.

Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (**1998**). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science *280*, 1750-1752.

Shimizu, T., Marusawa, H., Matsumoto, Y., Inuzuka, T., Ikeda, A., Fujii, Y., Minamiguchi, S., Miyamoto, S., Kou, T., Sakai, Y., *et al.* (**2014**). Accumulation of somatic mutations in TP53 in gastric epithelium with Helicobacter pylori infection. Gastroenterology *147*, 407-417 e403.

Shinkura, R., Ito, S., Begum, N.A., Nagaoka, H., Muramatsu, M., Kinoshita, K., Sakakibara, Y., Hijikata, H., and Honjo, T. (**2004**). Separate domains of AID are required for somatic hypermutation and class-switch recombination. Nat Immunol *5*, 707-712.

Sorek, R., Kunin, V., and Hugenholtz, P. (**2008**). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol *6*, 181-186.

Soussi, T., Hamroun, D., Hjortsberg, L., Rubio-Nevado, J.M., Fournier, J.L., and Beroud, C. (**2010**). MUT-TP53 2.0: a novel versatile matrix for statistical analysis of TP53 mutations in human cancer. Hum Mutat *31*, 1020-1025.

Stracker, T.H., Usui, T., and Petrini, J.H. (**2009**). Taking the time to make important decisions: the checkpoint effector kinases Chk1 and Chk2 and the DNA damage response. DNA repair *8*, 1047-1054.

Strasser, A., Harris, A.W., Jacks, T., and Cory, S. (**1994**). DNA damage can induce apoptosis in proliferating lymphoid cells via p53-independent mechanisms inhibitable by Bcl-2. Cell *79*, 329-339.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (**2009**). The cancer genome. Nature *458*, 719-724.

Su, L.K., Johnson, K.A., Smith, K.J., Hill, D.E., Vogelstein, B., and Kinzler, K.W. (**1993**). Association between wild type and mutant APC gene products. Cancer Res *53*, 2728-2731.

Sung, Y.H., Baek, I.J., Kim, D.H., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J.S., and Lee, H.W. (**2013**). Knockout mice created by TALEN-mediated gene targeting. Nat Biotechnol *31*, 23-24.

Ta, V.T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., *et al*. (**2003**). AID mutant analyses indicate requirement for class-switch-specific cofactors. Nat Immunol *4*, 843-848.

Takahashi, H. (**1977**). Juvenile hermaphroditism in the Zebrafish, *Brachydanio rerio*. Bulletin of the Faculty of Fisheries Hokkaido University *28*, 57 - 65.

Takai, A., Marusawa, H., Minaki, Y., Watanabe, T., Nakase, H., Kinoshita, K., Tsujimoto, G., and Chiba, T. (**2012**). Targeting activation-induced cytidine deaminase prevents colon cancer development despite persistent colonic inflammation. Oncogene *31*, 1733-1742.

Tan, S., Guschin, D., Davalos, A., Lee, Y.L., Snowden, A.W., Jouvenot, Y., Zhang, H.S., Howes, K., McNamara, A.R., Lai, A., *et al*. (**2003**). Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. Proc Natl Acad Sci U S A *100*, 11997-12002.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (**2008**). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity *28*, 621-629.

Teodoro, J.G., Evans, S.K., and Green, M.R. (**2007**). Inhibition of tumor angiogenesis by p53: a new role for the guardian of the genome. J Mol Med (Berl) *85*, 1175-1186.

Thisse, C., Neel, H., Thisse, B., Daujat, S., and Piette, J. (**2000**). The Mdm2 gene of zebrafish (Danio rerio): preferential expression during development of neural and muscular tissues, and absence of tumor formation after overexpression of its cDNA during early embryogenesis. Differentiation *66*, 61-70.

Thomas, K.R., and Capecchi, M.R. (**1987**). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.

Thome, R.G., Santos, H.B., Arantes, F.P., Domingos, F.F., Bazzoli, N., and Rizzo, E. (**2009**). Dual roles for autophagy during follicular atresia in fish ovary. Autophagy *5*, 117-119.

Tran, P.T., Erdeniz, N., Symington, L.S., and Liskay, R.M. (**2004**). EXO1-A multi-tasking eukaryotic nuclease. DNA repair *3*, 1549-1559.

Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (**2010**). B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. Nat Immunol *11*, 148-154.

Uchida, D., Yamashita, M., Kitano, T., and Iguchi, T. (**2002**). Oocyte apoptosis during the transition from ovary-like tissue to testes during sex differentiation of juvenile zebrafish. J Exp Biol *205*, 711-718.

Üçüncü, S.I., and Çakıcı, O. (**2009**). Atresia and Apoptosis in Preovulatory Follicles in the Ovary of Danio rerio (Zebrafish). Turkish Journal of Fisheries and Aquatic Sciences *9*, 215-221.

Urasaki, A., Morvan, G., and Kawakami, K. (**2006**). Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. Genetics *174*, 639-649.

Vaidyanathan, B., Yen, W.F., Pucella, J.N., and Chaudhuri, J. (**2014**). AIDing Chromatin and Transcription-Coupled Orchestration of Immunoglobulin Class-Switch Recombination. Front Immunol *5*, 120.

Veinotte, C.J., Dellaire, G., and Berman, J.N. (**2014**). Hooking the big one: the potential of zebrafish xenotransplantation to reform cancer drug screening in the genomic era. Dis Model Mech *7*, 745-754.

Vogelstein, B., and Kinzler, K.W. (**2004**). Cancer genes and the pathways they control. Nat Med *10*, 789-799.

Vousden, K.H., and Ryan, K.M. (**2009**). p53 and metabolism. Nat Rev Cancer *9*, 691-700.

Wallace, S.S. (**1998**). Enzymatic processing of radiation-induced free radical damage in DNA. Radiat Res *150*, S60-79.

Wang, C.L., Harper, R.A., and Wabl, M. (**2004**). Genome-wide somatic hypermutation. Proc Natl Acad Sci U S A *101*, 7352-7356.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (**2014**). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80-84.

Weinberg, R.A. (**1995**). The retinoblastoma protein and cell cycle control. Cell *81*, 323-330.

White, R.M., Sessa, A., Burke, C., Bowman, T., LeBlanc, J., Ceol, C., Bourque, C., Dovey, M., Goessling, W., Burns, C.E., *et al*. (**2008**). Transparent adult zebrafish as a tool for in vivo transplantation analysis. Cell Stem Cell *2*, 183-189.

Wienholds, E., van Eeden, F., Kosters, M., Mudde, J., Plasterk, R.H., and Cuppen, E. (**2003**). Efficient target-selected mutagenesis in zebrafish. Genome Res *13*, 2700-2707.

Wilkinson, R.N., Elworthy, S., Ingham, P.W., and van Eeden, F.J. (**2013**). A method for high-throughput PCR-based genotyping of larval zebrafish tail biopsies. Biotechniques *55*, 314-316.

Willmann, K.L., Milosevic, S., Pauklin, S., Schmitz, K.M., Rangam, G., Simon, M.T., Maslen, S., Skehel, M., Robert, I., Heyer, V., *et al*. (**2012**). A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. J Exp Med *209*, 2099-2111.

Xu, Z., Zan, H., Pone, E.J., Mai, T., and Casali, P. (**2012**). Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. Nat Rev Immunol *12*, 517-531.

Zan, H., and Casali, P. (**2008**). AID- and Ung-dependent generation of staggered double-strand DNA breaks in immunoglobulin class switch DNA recombination: a post-cleavage role for AID. Mol Immunol *46*, 45-61.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (**2011**). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol *29*, 149-153.

Zhang, L., Zhou, Y., Cheng, C., Cui, H., Cheng, L., Kong, P., Wang, J., Li, Y., Chen, W., Song, B.*, et al*. (**2015**). Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. Am J Hum Genet *96*, 597-611.

Zhang, Z., Wu, E., Qian, Z., and Wu, W.S. (**2014**). A multicolor panel of TALE-KRAB based transcriptional repressor vectors enabling knockdown of multiple gene targets. Sci Rep *4*, 7338.

Zharkov, D.O. (**2008**). Base excision DNA repair. Cellular and molecular life sciences : CMLS *65*, 1544-1565.

Zhou, B.B., and Elledge, S.J. (**2000**). The DNA damage response: putting checkpoints in perspective. Nature *408*, 433-439.

AKNOLEDGMENTS

I would like to show my deepest gratitude to those people who made this thesis possible.

First I would like to thank my supervisor and mentor Svend Petersen-Mahrt for his valuable advices and his enthusiasm in guiding me in my scientific work.

I am grateful to Gianluca Deflorian for introducing me to the zebrafish world, his supervision during fish work and reagents sharing; I would like to thank Riccardo Pecori and Silvo Conticello for the successful collaboration on the TALE work and the great scientific discussions. I thank Elisa Allievi, Ambra Belpietro and Eleonora Verga for generating the knock-in mouse and for their help with the MEFs generation. I am thankful to Mirko Riboni and Simone Minardi for generating the NGS data; I thank Federica Pisati for troubleshooting with whole fish mounting and histological sections. I am thankful to Stefano Casola and Almudena Ramiro for helpful scientific advices and reagents sharing. I want to thank Bernardo Reina-San-Martin for generously giving the STR9 antibody, Pier Giuseppe Pelicci for generously giving the TRP53/C57 mouse strain, Koichi Kawakami for providing me with the Tol2 containing plasmid and Andrew Bassett for the collaboration on the dCas9-Aid work. I thank IFOM facilities for their great work to support our research.

Thanks to the current and past lab members, Kerstin Schmitz, Elisabetta Incorvaia, Ilaria Spadafora, Federica La Mastra and Don-Marc Franchini for fruitful discussions and help in everyday work. A big thank also for taking the time to proof read my thesis, your help was crucial.

I am grateful to Valentina Buttiglione, Valentina Petrocelli and Marialaura Mastrovito for always being there for scientific advice and personal support.

To Piero: thank you for your love, patience, support and encouragement during these years.

To my family: thank you for your endless support and believing in me no matter what.