Technische Universität München

Fakultät für Medizin

Transposon- and CRISPR-based tools for tumour suppressor gene screening *in vivo*

Julia Maria Weber

Vollständiger Abdruck der von der promotionsführenden Einrichtung Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender: Prof. Dr. Dieter Saur

Prüfende/-r der Dissertation:

1. Prof. Dr. Radu Roland Rad

2. Prof. Angelika Schnieke, Ph.D.

3. Prof. Dr. Heinrich Leonhardt

Die Dissertation wurde am 16.07.2019 bei der Technischen Universität München eingereicht und durch die promotionsführende Einrichtung Medizinische Fakultät am 07.04.2020 angenommen.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

°C	Degree Celsius
A	Adenine
ABC	Activated B-cell-like
aCGH	Array comparative genomic hybridisation
ATP	Activating transposon
B	Base
Bp	Base pairs
C	Cytosine
CAG	CMV enhancer/chicken β-actin
Cas	CRISPR associated system
CCDS	Consensus coding sequence
CCl_4	*Carbon tetrachloride*
cDNA	Complementary deoxyribonucleic acid
CI	Confidence Interval
CIS	Common insertion site
CMV	Cytomegalovirus
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CSR	Class switch recombination
DEN	*N-nitrosodiethylamine*
DLBCL	Diffuse large B-cell lymphoma
DNA	Deoxyribonucleic acid
DSB	Double-strand break
e.g.	*Exempli gratia*
ENU	*N-ethyl-N-nitrosourea*
et al.	*Et alii*
FBS	Foetal bovine serum
FFPE	Formalin-fixed, paraffin-embedded
FISH	Fluorescence *in situ* hybridisation
G	Guanine
g	Gram

GC	Germinal centre
GCB	GC B-cell-like
GEMM	Genetically engineered mouse model
Gy	Gray
h	Hour
H&E	Haematoxylin and eosin
H_2O	Water
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HDR	Homology-directed repair
HL	Hodgkin lymphoma
HR	Hazard Ratio
HTVI	Hydrodynamic tail vein injection
ICC	Intrahepatic cholangiocarcinoma
i.e.	*Id est*
indel	Insertion/deletion
ITP	Inactivating transposon
ITR	Inverted terminal repeat
L	Litre
LFD	Liver fluke disease
LINE	Long interspersed nuclear elements
LOH	Loss of heterozygosity
LPD	Lymphoproliferative disease
LTR	Long terminal repeat
M	Molar
m	Metre
MHC	Major histocompatibility complex
MRF	Mutant read frequency
MRI	Magnetic resonance imaging
MSCV	Murine stem cell virus
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next-generation sequencing

NHEJ	Non-homologous end joining
NHL	Non-Hodgkin lymphoma
NK cells	Natural killer cells
OR	Odds Ratio
OS	Overall survival
PAM	Protospacer adjacent motif
PB	*PiggyBac*
PCR	Polymerase chain reaction
PFS	Progression-free survival
PGK	Phosphoglycerate kinase
PLC	Primary liver cancer
PMBL	Primary mediastinal B-cell lymphoma
poly(A)	Polyadenylation
QiSeq	Quantitative transposon insertion site sequencing
RNA	Ribonucleic acid
RT	Room temperature
s	Second
SB	*Sleeping Beauty*
SCE	Sister chromatid exchange
sgRNA	Single guide RNA
SHM	Somatic hypermutation
shRNA	Short hairpin ribonucleic acid
SINE	Short interspersed nuclear elements
SNP	Single nucleotide polymorphism
SPF	Specific-pathogen-free
T	Thymine
TR	Terminal repeat
tracrRNA	Trans-activating crRNA
TSG	Tumour suppressor gene
UK	United Kingdom
USA	United States of America
V	Volt

Unit prefixes

T	10^{12}	**m**	10^{-3}
G	10^{9}	**µ**	10^{-6}
M	10^{6}	**n**	10^{-9}
k	10^{3}	**p**	10^{-12}

Gene Names/Symbols

Murine Gene Symbol	**Human Gene Symbol**	**Full Name**
Abcb4	*ABCB4*	ATP binding cassette subfamily B member 4
Alk	*ALK*	ALK receptor tyrosine kinase
Ankrd11	*ANKRD11*	Ankyrin repeat domain 11
Apc	*APC*	Adenomatous polyposis coli
Apob	*APOB*	Apolipoprotein B
Arid1a	*ARID1A*	AT-rich interaction domain 1a
Arid1b	*ARID1B*	AT-Rich interaction domain 1b
Arid2	*ARID2*	AT-Rich interaction domain 2
Arid5b	*ARID5B*	AT-Rich interaction domain 5b
Atm	*ATM*	ATM serine/threonine kinase
B2m	*B2M*	Beta-2-microglobulin
Bap1	*BAP1*	BRCA1 associated protein 1
Bcl2	*BCL2*	B-cell lymphoma 2
Bcl6	*BCL6*	B-cell lymphoma 6
Blm	*BLM*	Bloom syndrome, RecQ helicase-like
Brca1	*BRCA1*	Breast cancer 1
Brca2	*BRCA2*	Breast cancer 2
Card11	*CARD11*	Caspase recruitment domain family member 11
Cd58	*CD58*	CD58 molecule
Cd79a/b	*CD79A*	CD79a/b molecule
Cdkn2a	*CDKN2A*	Cyclin-dependent kinase inhibitor 2a
Cdkn2b	*CDKN2B*	Cyclin-dependent kinase inhibitor 2b
Chl1	*CHL1*	Cell adhesion molecule L1 like
Crebbp	*CREBBP*	CREB binding protein
Ctnnb1	*CTNNB1*	Catenin beta 1
Dido1	*DIDO1*	Death inducer-obliterator 1
Egf	*EGF*	Epidermal growth factor
Eml4	*EML4*	Echinoderm microtubule associated protein like 4
En2	*EN2*	Engrailed (Homeobox) 2
Ep300	*EP300*	E1a binding protein p300
Erp44	*ERP44*	Endoplasmic reticulum protein 44

Murine Gene Symbol	Human Gene Symbol	Full Name
Errfi1	*ERRFI1*	ERBB receptor feedback inhibitor 1
Ezh2	*EZH2*	Enhancer of zeste 2 polycomb repressive complex 2 subunit
Fas	*FAS*	Fas cell surface death receptor
Fgf19	*FGF19*	Fibroblast growth factor 19
Gna13	*GNA13*	G Protein subunit alpha 13
Gnas	*GNAS*	Nucleotide binding protein, alpha stimulating
Hnf4a	*HNF4A*	Hepatocyte nuclear factor 4 alpha
Hnrnpa2b1	*HNRNPA2B1*	Heterogeneous nuclear ribonucleoprotein a2/b1
Idh1	*IDH1*	Isocitrate dehydrogenase 1
Idh2	*IDH2*	Isocitrate dehydrogenase 2
Igsf10	*IGSF10*	Immunoglobulin superfamily member 10
Irf2	*IRF2*	Interferon regulatory factor 2
Kmt2b	*KMT2B*	Lysine methyltransferase 2B
Kmt2c	*KMT2C*	Lysine methyltransferase 2C
Kmt2d	*KMT2D*	Lysine methyltransferase 2D
Kras	*KRAS*	V-Ki-Ras2 Kirsten rat sarcoma viral oncogene homolog
Lmbrd1	*LMBRD1*	LMBR1 domain containing 1
Lnpep	*LNPEP*	Leucyl and cystinyl aminopeptidase
Malt1	*MALT1*	MALT1 paracaspase
Myc	*MYC*	MYC proto-oncogene, bHLH transcription factor
Myd88	*MYD88*	Myeloid differentiation primary response 88
Naa15	*NAA15*	N(alpha)-acetyltransferase 15, NatA auxiliary subunit
Nkx2-1	*NKX2-1*	NK2 homeobox 1
Notch 1	*NOTCH 1*	Notch 1
Pax5	*PAX5*	Paired box 5
Pbrm1	*PBRM1*	Polybromo 1
Prdm1	*PRDM1*	PR/SET domain 1
Psme4	*PSME4*	Proteasome activator subunit 4
Pten	*PTEN*	Phosphatase and tensin homolog
Ptprc	*PTPRC*	Protein tyrosine phosphatase, receptor type C
Rb1	*RB1*	RB transcriptional corepressor 1
Smad4	*SMAD4*	SMAD family member 4
Tert	*TERT*	Telomerase reverse transcriptase
Tet2	*TET2*	Tet methylcytosine dioxygenase 2
Tgfa	*TGFA*	Transforming growth factor alpha
Tnfaip3	*TNFAIP3*	TNF alpha induced protein 3
Tnfrsf14	*TNFRSF14*	TNF receptor superfamily member 14
Tox	*TOX*	Thymocyte selection associated high mobility group box
Trp53	*TP53*	Transformation related protein 53
Wac	*WAC*	WW domain containing adaptor with coiled-coil

1 ABSTRACT

Cancer is a complex disease with countless underlying genetic alterations. Progress in next-generation sequencing and other high-throughput technologies has led to the generation of large catalogues of putative cancer genes. However, the exact role of many of these genes in tumourigenesis is still far from being understood. Furthermore, the identification of non-mutated but transcriptionally or epigenetically dysregulated cancer genes by sequencing- or array-based approaches remains challenging. Forward genetic screening technologies can overcome some of these limitations of classic approaches to cancer genome analysis, especially with regard to the analysis of the non-mutated cancer genome.

In recent years, transposon-based insertional mutagenesis and CRISPR/Cas have emerged as powerful tools for genetic screening. *PiggyBac* and *Sleeping Beauty* transposon screens led to the discovery of a multitude of novel cancer genes in numerous tumour entities.

However, *in vivo* screening in a recessive context for the identification of new tumour suppressor genes has not been achieved so far. Likewise, while CRISPR/Cas has proven to be a powerful tool for the generation of gene knockouts and *in vitro* screening, its utilisation for forward genetic screening *in vivo* has not been demonstrated yet. In this work, novel transposon- and CRISPR/Cas-based tools for recessive screening *in vivo* have been generated and successfully utilised for tumour suppressor gene discovery in mice.

The first part of this thesis focused on the development of new “inactivating” transposon tools for tumour suppressor gene screening and demonstrated their applicability for genome-wide whole-body mutagenesis *in vivo*. Using a Bloom-mutated background prone to loss of heterozygosity, this has led to the identification of numerous known and novel tumour suppressor genes in diffuse large B-cell lymphoma. As transposon mobilisation took place in all somatic cells of the animals, occurrence of other tumour entities, such as liver, lung and intestinal cancers, was observed, thus demonstrating the high versatility of the model. Additionally, for the first time, an unbiased comparative analysis of the *in vivo* characteristics of *PiggyBac* and *Sleeping Beauty* was performed, pinpointing numerous biological differences between both transposition systems.

The second part of this thesis concentrated – in form of a proof-of-concept study – on the analysis of the suitability of CRISPR/Cas for tumour suppressor gene screening *in vivo*. The work demonstrated for the first time that CRISPR/Cas can be adapted for recessive genetic screening in adult mice. Analysis of the CRISPR/Cas-induced mutational pattern in mouse liver cancers revealed a striking positive selection for tumour-relevant mutations in already known but also in novel tumour suppressor genes such as *Tet2*. Moreover, the work points out that

CRISPR/Cas multiplexing can be exploited for chromosome engineering *in vivo*, showing modelling of large intrachromosomal deletions encompassing several megabases.

Taken together, this work is the first successful demonstration of transposon- and CRISPR/Cas-based *in vivo* tumour suppressor gene screens, highlighting the power of both tools for unravelling genetic landscapes of cancers and assigning biological function to genes.

2 ZUSAMMENFASSUNG

Krebs ist eine komplexe Erkrankung mit unzähligen zugrundeliegenden genetischen Veränderungen. Fortschritte bei neuen Sequenziertechnologien und anderen Hochdurchsatzmethoden haben zur Generierung umfangreicher Listen von putativen Krebsgenen geführt. Allerdings ist die genaue Rolle vieler dieser Gene in der Krebsentstehung immer noch unklar. Des Weiteren stellt die Identifizierung nicht-mutierter aber transkriptionell oder epigenetisch deregulierter Krebsgene immer noch eine große Herausforderung dar. Genetische Screens können manche dieser Limitationen der klassischen Herangehensweisen an die Krebsgenomanalyse überwinden, insbesondere im Hinblick auf die Analyse des nicht-mutierten Krebsgenoms. In den letzten Jahren haben sich Transposon-basierte Insertionsmutagenese und CRISPR/Cas als leistungsfähige Werkzeuge für das genetische Screening erwiesen. So konnte mit Hilfe von *PiggyBac* und *Sleeping Beauty* Transposonscreens eine Vielzahl neuer Krebsgene in zahlreichen Tumorentitäten identifiziert werden. Jedoch konnte bis jetzt noch nicht rezessives Screening für die Entdeckung neuer Tumorsuppressorgene erfolgreich *in vivo* umgesetzt werden. In ähnlicher Weise konnte zwar das große Potenzial des CRISPR/Cas-Systems für die Generierung von Gen-Knockouts und *in vitro* Screening gezeigt werden, nicht aber die Anwendbarkeit des Systems für genetisches Screening *in vivo*.

Im Rahmen dieser Arbeit wurden neue Transposon- und CRISPR/Cas-basierte Werkzeuge für das rezessive Screening in Mäusen generiert und zum ersten Mal erfolgreich für die Identifizierung von Tumorsuppressorgenen *in vivo* eingesetzt.

Der erste Teil dieser Arbeit fokussiert sich auf die Entwicklung neuer inaktivierender Transposon-Werkzeuge für das Screening nach Tumorsuppressorgenen und zeigt deren Anwendbarkeit für die genomweite Ganzkörper-Mutagenese *in vivo*. Durch die Verwendung eines Bloom-mutierten Hintergrunds, der hohe Heterozygotie-Verlustraten aufweist, konnten eine Vielzahl bekannter und neuer Tumorsuppressorgene assoziiert mit dem diffus großzelligen B-Zell-Lymphom identifiziert werden. Da die Transposon-Mobilisierung in allen somatischen Zellen der Mäuse stattgefunden hat, konnte darüber hinaus die Entwicklung weiterer Tumorentitäten, wie z.B. von Leber-, Lungen- und Darmtumoren, induziert werden, was die große Vielfältigkeit des Modells unterstreicht. Des Weiteren wurde im Rahmen dieser Arbeit die erste vergleichende Analyse der *in vivo* Eigenschaften von *PiggyBac* und *Sleeping Beauty* vorgenommen und zahlreiche biologische Unterschiede der beiden Transpositionssysteme herausgearbeitet.

Der zweite Teil dieser Arbeit konzentriert sich in Form einer „Proof-of-Concept"-Studie auf die Analyse der Eignung des CRISPR/Cas-Systems für das Tumorsuppressorgen-Screening *in vivo*. Im Rahmen dieser Arbeit konnte erstmalig gezeigt werden, dass das CRISPR/Cas-System erfolgreich für das rezessive Screening in adulten Mäusen adaptiert werden kann. So zeigte die Analyse des CRISPR/Cas-induzierten Indel-Musters in Lebertumoren eine bemerkenswerte positive Selektion für krebsrelevante Mutationen in bereits bekannten, aber auch in neuen Tumorsuppressorgenen, wie z.B. *Tet2*. Außerdem konnte in dieser Arbeit gezeigt werden, dass CRISPR/Cas-Multiplexing für „Chromosome Engineering" *in vivo* verwendet werden kann. So konnten beispielsweise große intrachromosomale Deletionen, die mehrere Megabasen umfassen, erfolgreich modelliert werden.

Zusammengefasst stellt diese Arbeit die erste erfolgreiche Demonstration von Transposon- und CRISPR/Cas-basierten *in vivo* Tumorsuppressorgen-Screens dar und hebt insbesondere die Leistungsfähigkeit der beiden Werkezeuge hervor, um die genetischen Landschaften von Tumoren zu entschlüsseln und um Genen ihre biologische Funktion zuzuweisen.

3 INTRODUCTION

Cancer, a group of more than 100 distinct neoplastic diseases characterised by abnormal cell growth, is among the leading causes of death worldwide, with about 8.2 million cancer-related deaths in 2012 (Torre *et al.*, 2015). While irregular cell proliferation can also cause benign tumour masses, cancer cells (and their surrounding microenvironment) are characterised by the so-called "hallmarks of cancer" proposed by Hanahan and Weinberg in 2000 (Hanahan and Weinberg, 2000) and updated in 2011 (Hanahan and Weinberg, 2011). These include, among others, self-sufficiency in growth signals, insensitivity to anti-growth signals, tissue invasion and metastasis, limitless replicative potential, sustained angiogenesis and evading apoptosis.

An estimate of 5-10% of all cancers is of hereditary origin (caused by germline mutations in high-penetrance cancer susceptibility genes) (Nagy *et al.*, 2004) while the vast majority is considered to be sporadic. The contribution of rare disease-causing variants and common low-penetrance alleles to the development of these "sporadic" cancers is the current focus of a multitude of genome-wide association studies (Fletcher and Houlston, 2010).

In general, genetic abnormalities play a pivotal role in tumour development. Therefore, the identification of oncogenes, tumour suppressor genes (TSG) and pathways with aberrant activities in cancer provides opportunities for the development of novel prognostic/diagnostic markers and targeted therapeutics. However, although enormous progress has been made over the past few decades, the complexity of genetic alterations driving tumourigenesis is still far from being understood. This underlines the critical need for more enhanced technologies for cancer gene screening and analysis/validation.

3.1 Tools for cancer gene discovery

The most profound impact on cancer genome analysis has emerged in recent years from progress in next-generation sequencing (NGS) technologies, which has tremendously facilitated the large-scale discovery of single nucleotide variations, insertions and deletions (indels), copy number variations, complex rearrangements and epigenetic modifications (Campbell *et al.*, 2008; Forbes *et al.*, 2008; Stratton *et al.*, 2009; Bignell *et al.*, 2010; Pleasance *et al.*, 2010; Alexandrov *et al.*, 2013; Kandoth *et al.*, 2013; Lawrence *et al.*, 2014). Complementary to NGS, genome-wide forward genetic screening in model organisms, such as mice, has contributed eminently to the identification of oncogenes and TSGs. Cancer gene

discovery approaches include (i) library-based screens using, for example, complementary deoxyribonucleic acid (cDNA), clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated system (Cas) or short hairpin ribonucleic acid (shRNA) libraries and (ii) mutagenesis screens utilising, for instance, irradiation, chemicals (e.g. *N-ethyl-N-nitrosourea*; ENU) or insertional mutagenesis tools (e.g. retroviruses and transposons). This thesis focuses on the applicability of transposon and CRISPR/Cas tools for TSG identification in mice.

3.2 Genetic screening using *Sleeping Beauty* and *PiggyBac* transposons

3.2.1 Transposons as mobile genetic elements within genomes

Transposons (or transposable elements) are mobile genetic elements, i.e. sequences of deoxyribonucleic acid (DNA) that are capable of switching their position within a genome (a mechanism called transposition). They were first discovered as “jumping genes” in maize by Barbara McClintock more than 60 years ago (McClintock, 1950), for which she was awarded the Nobel Prize in 1983. Transposons in form of retrotransposons with long terminal repeats (LTR), retrotransposons without LTRs (long and short interspersed nuclear elements; LINE and SINE) and DNA transposons constitute up to 45% of the human genome (Gregory, 2005) although the vast majority of them has been rendered inactive millions of years ago.

In general, there a two classes of transposons (Brookfield, 2005): Class I transposons (or retrotransposons) function in a self-amplifying manner. At first, DNA coding for the retrotransposon is transcribed to ribonucleic acid (RNA). Then, after transcription of RNA back to DNA, the latter integrates into a new genomic locus in a “copy-and-paste” manner. This second step is mediated by a reverse transcriptase, which is encoded by the transposon. In contrast, class II transposons (or DNA transposons) employ a “cut-and-paste” mechanism. A transposase protein (which the transposon encodes) transfers the transposon from one genomic locus to another. Transposon recognition and binding by the transposase occurs at specific sites within inverted terminal repeats (ITR), which flank the transposase sequence. Focus of this thesis is the application of DNA transposons for TSG discovery in mice.

3.2.2 Adaption of *Sleeping Beauty* and *PiggyBac* transposons for use in mammals

Transposon-based forward genetic screens have been performed in lower organisms (e.g. *Caenorhabditis elegans* and *Drosophila melanogaster*) for decades (Jorgensen and Mango, 2002; St Johnston, 2002; Thibault *et al.*, 2004). However, until recently, the lack of efficient insertional mutagenesis tools has hampered the implementation of these screening approaches in mammals/vertebrates. Hence, reconstruction/engineering of *Sleeping Beauty (SB) and PiggyBac (PB)* transposons for use in mammalian cells revolutionised forward genetic screening in higher organisms.

SB, belonging to the most abundant class of DNA transposons – Tc1/mariner – was reconstructed from the genome of fish in 1997 (Ivics *et al.*, 1997). Further improvements of this transposon system led to its successful application for insertional mutagenesis in mice (Collier *et al.*, 2005; Dupuy *et al.*, 2005). *PB*, another DNA transposon, originating from the cabbage looper moth *Trichoplusia ni*, was also shown to be active in mammalian cells (Ding *et al.*, 2005). Thereafter, the group of Allan Bradley (Wellcome Trust Sanger Institute, Hinxton/Cambridge, UK) pioneered *PB*-based cancer gene discovery in mice (Rad *et al.*, 2010).

In general, all frequently employed *SB* and *PB* transposon systems provide the transposase activity *in trans*, i.e. the transposase is not encoded by the transposon itself but in another locus. Instead, the transposon usually harbours several gene regulatory elements (such as promoters, splice donors/acceptors and polyadenylation (poly(A)) signals) between its ITRs.

3.2.3 Utilisation of *Sleeping Beauty* and *PiggyBac* transposons for cancer gene discovery in mice

The first two research papers describing the utilisation of the *SB* system for cancer gene discovery in mice were published back-to-back by the Largaespada (Collier *et al.*, 2005) and Copeland/Jenkins laboratories (Dupuy *et al.*, 2005) in 2005. Two novel *SB* transposons, named *T2/Onc* and *T2/Onc2* (which is an optimised variant of the former), were designed to work bi-functionally. While a promoter (the murine stem cell virus (MSCV) LTR) followed by a splice donor can induce gain-of-function mutations, bidirectional gene trapping elements (splice acceptors bordering poly(A) signals) can lead to disruption of gene function. For the general mechanism of oncogene activation and TSG inactivation by transposons, see Figure 3-1. In both studies, analysis of common insertions sites (CIS; sites within the genome which are more frequently hit by transposons than expected by chance) in tumours revealed

several already known cancer genes but also genes for which a role in tumourigenesis has not been described at that time.

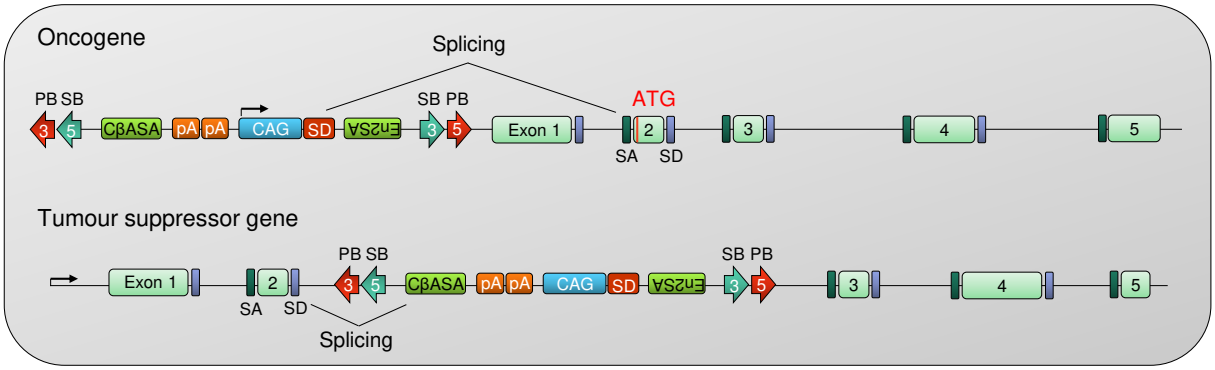

FIGURE 3-1. Mechanisms of transposon-mediated gene activation and inactivation.

Bi-functional (or activating) transposons can induce (over-)expression of genes (e.g. oncogenes) by inserting upstream of the transcriptional start site if promoter and splice donor (SD) are in sense orientation (upper scheme). Bi-functional (or inactivating) transposons can also disrupt expression of genes (e.g. tumour suppressor genes, TSG) by inserting in introns of a gene. Here, polyadenylation signals and splice acceptors function as bi-directional gene trapping cassettes (lower scheme). In rarer cases (since introns are usually much larger than exons and therefore the likelihood of insertions into introns is higher), TSG disruption can also occur by insertions in exons which lead to a frameshift and premature termination of expression (not shown). Figure shows the bi-functional *ATP1* transposon. Figure adapted from Friedrich *et al.*, 2017. PB, *PiggyBac* ITRs; SB, *Sleeping Beauty* ITRs; CβASA, carp β-actin splice acceptor; pA, SV40 polyadenylation signal; CAG, cytomegalovirus enhancer/chicken β-actin promoter; En2SA, *Engrailed 2* exon-2 splice acceptor; SA, splice acceptor.

Further advances in the *SB* technology were made by designing a transposon *(T2/Onc3)* (Figure 3-2) harbouring a cytomegalovirus (CMV) enhancer/chicken β-actin (CAG) promoter instead of the MSCV LTR and thereby shifting the transposon-induced tumour spectrum from haematopoietic to solid cancers (Dupuy *et al.*, 2009). In addition, conditional *SB* transposase knockin mouse lines were generated enabling tissue-specific expression of the transposase and thus overcoming limitations regarding embryonic lethality and induction of undesirable tumour entities (Dupuy *et al.*, 2009; Starr *et al.*, 2009). Various studies demonstrated the successful application of the *SB* system for cancer gene discovery in mice for a multitude of different cancers. Examples include blood cancer (van der Weyden *et al.*, 2011; Tang *et al.*, 2013; van der Weyden *et al.*, 2013), hepatocellular carcinoma (Keng *et al.*, 2009; Bard-Chapeau *et al.*, 2014), intestinal cancer (Starr *et al.*, 2009; Starr *et al.*, 2011; Takeda *et al.*, 2015), melanoma (Mann *et al.*, 2015), nervous system cancer (Genovesi *et al.*, 2013;

Rahrmann *et al.*, 2013), osteosarcoma (Moriarity *et al.*, 2015) and pancreatic cancer (Mann *et al.*, 2012; Perez-Mancera *et al.*, 2012).

In 2010, the first demonstration of the *PB* system as a tool for cancer gene discovery in mice was published by Allan Bradley's group (Rad *et al.*, 2010). They developed a *PB* transposon tool box consisting of two main components: (i) a *PB* transposase knockin mouse which constitutively expresses the *PB* transposase under control of the *Rosa26* promoter and (ii) 19 activating transposon (ATP) mouse lines varying in transposon type (*ATP1*, *ATP2* or *ATP3*) (Figure 3-2) as well as in locus and size of the transposon concatemer. ATP transposons are bi-functional (activating/inactivating) and harbour *SB* as well as *PB* ITRs for maximum flexibility (meaning either transposase can mobilise the transposon). Between the ITRs, promoters and gene trapping cassettes (similar to the T2/Onc transposon family) were included. Depending on the promoter type (CAG, MSCV or phosphoglycerate kinase (PGK)), mice predominantly developed solid cancers (CAG), haematopoietic tumours (MSCV) or both (PGK). More than 40% of all *PB*-CISs in haematopoietic tumours had not been identified before in haematologic *SB* and retroviral screens underlying *PB*'s unique insertion properties.

In 2015, Allan Bradley's group further expanded the *PB* transposon technology by generating a conditional *PB* transposase mouse line, which allows spatial restriction of *PB*-driven insertional mutagenesis to specific organs (Rad *et al.*, 2015). Published studies include *PB*-based cancer gene discovery in pancreatic cancer (Rad *et al.*, 2015) and in melanoma (Ni *et al.*, 2013). Figure 3-2 shows T2/Onc and ATP design, and differences between *SB* and *PB* transposon systems are discussed in Chapter 3.2.4.

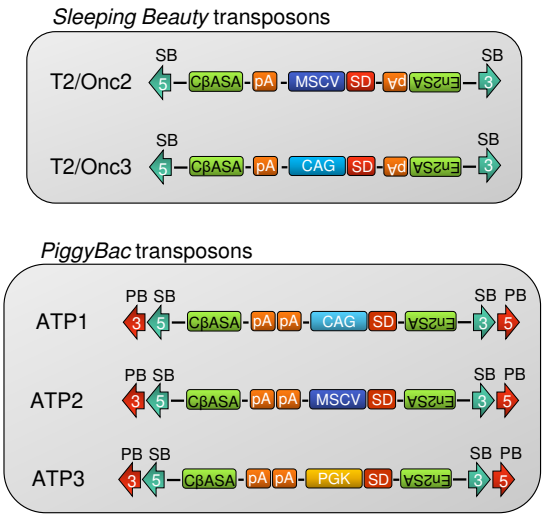

FIGURE 3-2. Structure of *Sleeping Beauty* and *PiggyBac* transposons.

Examples of bi-functional *Sleeping Beauty* (*T2/Onc2* and *T2/Onc3*; above) and *PiggyBac* transposons (*ATP1*, *ATP2* and *ATP3*; below) are shown. While *T2/Onc2* and *T2/Onc3* can be mobilised by the *Sleeping Beauty* transposase, mobilisation of *ATP1*, *ATP2* and *ATP3* can be mediated by the *PiggyBac* as well as the *Sleeping Beauty* transposase. SB, *Sleeping Beauty* ITRs; CβASA, carp β-actin splice acceptor; pA, SV40 polyadenylation signal; MSCV, murine stem cell virus promoter; CAG, cytomegalovirus enhancer/ chicken β-actin promoter; PGK, phosphoglycerate kinase promoter; SD, splice donor; En2SA, *Engrailed 2* exon-2 splice acceptor; PB, *PiggyBac* ITRs.

3.2.4 Comparison of *Sleeping Beauty* and *PiggyBac* characteristics

SB and *PB* have a number of different biological attributes. Not only do they have diverse insertion sites (TA for *SB* and TTAA for *PB*, although the latter might have non-canonical insertion sequences, too (Li *et al.*, 2013)), they also vary in their cargo capacity. Whereas for *SB* a 30% drop in transposition frequency has to be expected for every additional kilo base (kb) cargo (from its original 1.7 kb size) (Izsvak *et al.*, 2000), *PB* exhibits no decrease in transposition activity for up to 9 kb cargo size, allowing complex regulatory elements to be incorporated (Ding *et al.*, 2005). Moreover, *PB* has a much weaker tendency for local hopping, defined as re-integration of the transposon close to its original locus. While a large proportion (50-80%) of all *SB* transposons insert near (< 6 mega base (Mb)) their donor locus (Copeland and Jenkins, 2010), the local hopping activity of *PB* is much less pronounced (Rad *et al.*, 2010; Friedel *et al.*, 2013). Consequently, *SB*-induced tumours harbour a great number of insertions near the original transposon concatemer locus, making it difficult to identify “true” CISs on the donor chromosome. To overcome this problem and allowing coverage of the whole genome, often two transgenic *SB* mouse lines with distinct donor chromosomes are combined in *SB* screens (Bard-Chapeau *et al.*, 2014; Mann *et al.*, 2015). In contrast, there is no need to exclude donor chromosome insertions (apart from insertions within 1-3 Mb distance of the concatemer) from CIS analysis in *PB*-induced tumours (Rad *et al.*, 2010). Furthermore, *PB* preferentially integrates into open chromatin, expressed genes and transcription start sites while *SB* insertion sites are more randomly distributed and can also be found in genomic deserts (de Jong *et al.*, 2014). Another difference between the two transposon systems is that *SB* excision leaves a 5-bp-footprint (CAGTA or CTGTA) at its original site that might have mutagenic potential but also enables tracking of SB insertions (Liu *et al.*, 2004). In contrast, *PB* excision functions seamlessly (Fraser *et al.*, 1996). Additionally, the most commonly used *PB* transposase, *iPBase*, is nearly 250-fold more active in mammalian cells than the frequently used *SB* variant *SB11* (Liang *et al.*, 2009).

As observations in *Drosophila melanogaster* have shown that saturation mutagenesis can only be achieved if multiple transposon systems are deployed (Thibault *et al.*, 2004), *SB* and *PB* may be conclusively considered as complementary systems with many non-redundant properties.

3.3 The CRISPR/Cas system for genome engineering

3.3.1 Discovery of CRISPR/Cas as a prokaryotic immune system

The CRISPR/Cas system functions as an adaptive immune defence mechanism in about 50% of bacteria and more than 90% of archaea (Mojica *et al.*, 2000) and provides resistance to foreign viral and plasmid DNA (Mojica *et al.*, 2005). In prokaryotes, these CRISPR loci consist of a clustered set of *Cas* genes and the so-called CRISPR array, which contains short identical repeats interspaced by variable DNA sequences (spacers). The CRISPR/Cas immune defence involves three stages: (i) adaptation, (ii) expression and (iii) interference (van der Oost *et al.*, 2009). In the adaptation stage, a short fragment of foreign DNA (protospacer) is inserted as spacer into the CRISPR array. The CRISPR array is then transcribed as a precursor RNA (pre-crRNA), which is processed to individual mature CRISPR RNAs (crRNAs) (expression stage). In the interference stage, crRNAs together with Cas proteins target and cleave the foreign DNA (or in rarer cases RNA (O'Connell *et al.*, 2014)). While the adaptation stage is relatively conserved, expression and interference stages vary concerning involved enzymes and other auxiliary molecules between different CRISPR/Cas systems (Wright *et al.*, 2016).
A 2015 updated classification of CRISPR/Cas systems comprises two classes, five types and 16 subtypes based on genomic sequence analyses (Makarova *et al.*, 2015). Class 2 type II-A CRISPR/Cas systems are the most well-known, with the components from *Streptococcus pyogenes* and *Streptococcus thermophilus* first being adapted for use in mammalian cells (Cong *et al.*, 2013; Mali *et al.*, 2013b). In the past ten years, many research groups independently worked on CRISPR/Cas systems, often with a particular focus on the CRISPR/Cas9 system derived from *Streptococcus pyogenes* (*hSpCas9* system).

3.3.2 Adaptation of the CRISPR/Cas9 system for use in mammals

In the *hSpCas9* system, the processing of the pre-crRNA into multiple crRNAs is dependent on another RNA known as trans-activating crRNA (tracrRNA). The tracrRNA is able to hybridise to the repeat sequence within the CRISPR array due to sequence complementary (Deltcheva *et al.*, 2011), which allows processing of the pre-crRNAs by an RNA-specific ribonuclease (RNase III). Cas9, crRNA and tracrRNA form a ribonucleotide complex, in which the former, upon binding of the two RNAs, changes into an active DNA-binding conformation. The complex then can bind to the target DNA, which is complementary to the crRNA. A prerequisite for DNA binding is the presence of a so-called protospacer adjacent motif (PAM) next to the complementary sequence of the target DNA (for the *hSpCas9* system 5'-NGG-3'

directly downstream of the target sequence) (Hsu *et al.*, 2014). Upon binding, the Cas9 ribonucleotide complex undergoes a second conformational change, and is now able to separate the two strands of the double-stranded target DNA. Cleavage of both DNA strands by the Cas9 nuclease domains *RuvC* and *HNH* (Jinek *et al.*, 2014; Nishimasu *et al.*, 2014) induces a DNA double-strand break (DSB). For DSB repair, two different cellular mechanisms exist: (i) non-homologous end joining (NHEJ) and (ii) homology-directed repair (HDR) (Chu, 1997). In absence of a repair template (which is a prerequisite for HDR), NHEJ is the main repair pathway. NHEJ is error-prone and often leaves small indels at the site of repair, which can lead to a frameshift and disruption of gene expression. This CRISPR/Cas9-mediated inactivation of genes is now often exploited for "knockout" of candidate genes.

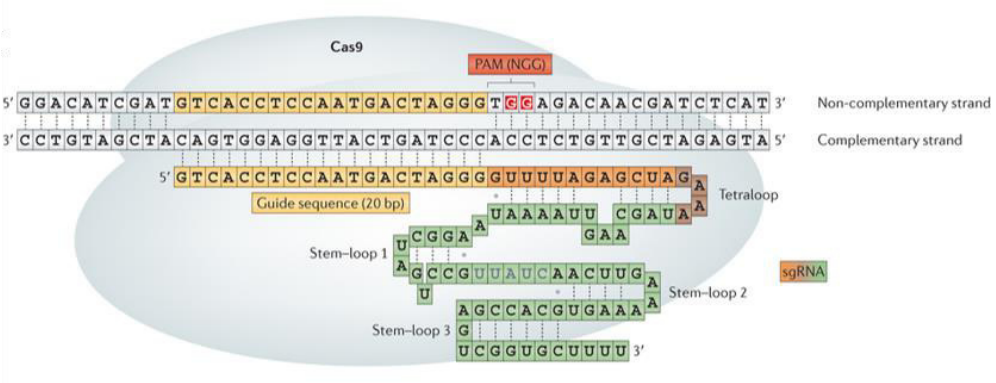

FIGURE 3-3. The CRISPR/Cas9 system for genome engineering.

The *hSpCas9* endonuclease is directed to a genomic locus of choice by a 20-bp guide sequence (yellow) within the single guide RNA (sgRNA). Different parts of the engineered sgRNA are shown in orange ("crRNA"), brown (tetraloop) and green ("tracrRNA"). Note that the protospacer adjacent motif (PAM; 5'-NGG-3') sequence is not part of the sgRNA. For targeting, the guide sequence pairs with the genomic target sequence (dotted lines) and a DNA double-strand break can be induced 3 bp upstream of the PAM sequence. Image reprinted by permission from the Nature Publishing Group (Kim and Kim, 2014).

In 2012, it was first shown that crRNA, tracrRNA, RNase III and Cas9 jointly induce DSBs *in vitro* (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012) and Jinek *et al.* constructed a single guide RNA (sgRNA) fusing crRNA and tracrRNA together (Jinek *et al.*, 2012) (Figure 3-3). A year later, Cong *et al.* and Mali *et al.* demonstrated the first application of the CRISPR/Cas9 system for genome engineering in mammalian cells *in vitro* (Cong *et al.*, 2013; Mali *et al.*, 2013b), and showed that RNase III is dispensable in this setting (suggesting the presence of endogenous RNase molecules with similar functions).

The CRISPR/Cas9 system is now used in countless applications/approaches in agricultural, biological and biomedical research with ongoing developments being made. For instance, catalytically inactive (nuclease-deficient) Cas9 variants (dCas9) are fused with repressor (e.g. Krueppel associated box) or activator (e.g. VP64) domains for gene repression (CRISPRi) or gene activation (CRISPRa) approaches (Dominguez *et al.*, 2016). Moreover, Cas9 variants with only one functional nuclease domain (Cas9 nickases; which induce single-strand breaks) can be used to foster HDR or to reduce off-target effects via double-nicking strategies (Shen *et al.*, 2014). In addition, applications of engineered Cas9 variants for base editing (introduction of point mutations) (Komor *et al.*, 2016) and epigenetic modifications (e.g. DNA methylation (Liu *et al.*, 2016)) were recently shown. Lastly, other CRISPR systems like CRISPR/Cpf1, which have features distinct from the CRISPR/Cas9 system, are under further development (Zetsche *et al.*, 2015a).

3.3.3 CRISPR/Cas9 as a tool in cancer research

As mentioned above (Chapter 3.3.2), the CRISPR/Cas9 system can be employed for a broad variety of biological applications. Focus of this chapter is the utilisation of the CRISPR/Cas9 system for cancer research.

CRISPR/Cas9-based tools have been used to model/study a broad variety of tumours *in vivo*, including lung cancer (Blasco *et al.*, 2014; Maddalo *et al.*, 2014; Platt *et al.*, 2014; Sanchez-Rivera *et al.*, 2014), liver cancer (Xue *et al.*, 2014; Weber *et al.*, 2015), pancreatic cancer (Maresch *et al.*, 2016), glioblastoma and medulloblastoma (Zuckermann *et al.*, 2015) and breast cancer (Annunziato *et al.*, 2016). Generation of constitutive and conditional Cas9 knockin mice (Platt *et al.*, 2014; Weber *et al.*, 2019) has further facilitated these approaches as Cas9 delivery has become dispensable.

Most models rely on transfection- or viral-based delivery methods. For instance, Xue *et al.* targeted the mouse liver by applying hydrodynamic tail vein injection (HTVI), a non-invasive transfection-mediated delivery method specific for the liver, which is based on high hydrodynamic pressure (Zhang *et al.*, 1999). Wild type mice treated with *carbon tetrachloride (CCl_4)* and targeted with Cas9 and sgRNAs directed against *Pten* and *Trp53* developed liver cancers with biliary differentiation (Xue *et al.*, 2014). In contrast, all lung cancer modelling approaches so far exploited viral-based methods, which are based on intranasal/-tracheal delivery of lentiviral, adenoviral or adeno-associated viral CRISPR/Cas9 components. In this organ system, not only knockout of known lung TSGs (e.g. *Apc*, *Lkb1*, *Nkx2-1* and *Pten*) was demonstrated (Platt *et al.*, 2014; Sanchez-Rivera *et al.*, 2014), but also modelling of

translocation events, which are common in human lung tumours, like *Eml4-Alk* (Blasco *et al.*, 2014; Maddalo *et al.*, 2014), was achieved.
Lastly, screening approaches using, for example, CRISPR/Cas9-based genome-wide libraries have been exploited in transplantation/xenograft models to discover novel genes regulating key processes in tumourigenesis, such as metastasis (Chen *et al.*, 2015).

3.4 Non-Hodgkin lymphoma

Lymphomas or lymphoid neoplasms are tumours arising from lymphocytes (B cells, T cells and natural killer (NK) cells). Traditionally, lymphomas are classified into two main classes, namely Hodgkin lymphomas (HL) and non-Hodgkin lymphomas (NHL). HL, a rather homogenous disease, is characterised by the presence of Reed-Sternberg cells (multinuclear giant cells derived from B cells), and is most often associated with a very good prognosis (>90% five-year survival rate) (Townsend and Linch, 2012). In contrast, NHL includes a very heterogeneous class of lymphoid neoplasms comprising three sub-classes: B-cell NHL, T-cell NHL and NHL of unknown lineage. For B-cell NHL, which encompasses about 85-90% of all NHL cases (Shankland *et al.*, 2012), two main sub-classes are known: precursor B-cell NHL (precursor B-lymphoblastic leukaemia/lymphoma) and mature B-cell NHL. While precursor B-cell NHL is mainly a childhood cancer, mature B-cell NHL predominately arises in adults, and is an extremely heterogeneous group of cancers. It includes among others, follicular lymphoma, mantle cell lymphoma, marginal zone lymphoma and diffuse large B cell lymphoma (DLBCL) (Swerdlow *et al.*, 2016).

3.4.1 Epidemiology, pathophysiology and treatment of diffuse large B-cell lymphoma

NHL is one of the most common tumour entities in Western countries (e.g. sixth most frequent cancer type among men and women in the US) (Siegel *et al.*, 2016) and its incidence is increasing in developed countries (Shankland *et al.*, 2012). DLBCL, the most prevalent form of NHL, accounts for about 30-40% of all adult NHL cases and is common worldwide without a geographical prevalence (Siegel *et al.*, 2013). Characterised by a diffuse growth pattern and abnormally large B cells (Martelli *et al.*, 2013), DLBCL is a very heterogeneous cancer type. While in principle, patients of all age groups can be affected by DLBCL, it primarily occurs in people of 60 years or older, with a median age of diagnosis of about 70 years (Smith *et al.*,

2011). DLBCL can arise *de novo* or, in rarer instances, from transformation of indolent B-cell lymphomas like follicular lymphoma (Lossos *et al.*, 2002). Being an aggressive, fast-growing type of lymphoma, it can occur inside and outside of the lymphatic system. First symptoms in patients often include weight loss, night sweats and fever (B symptoms) as well as lymphadenopathy. Standard treatment is the R-CHOP regime, which includes therapy with the monoclonal B-cell specific CD20 antibody rituximab in combination with chemotherapy (cyclophosphamide, hydroxydaunorubicin and vincristine) and the glucocorticoids prednisone/prednisolone (Roschewski *et al.*, 2014). Glucocorticoids efficiently induce apoptosis in lymphoid cells and are therefore included in all standard treatment regimens for lymphoid neoplasms (Schmidt *et al.*, 2004). Several studies showed that the addition of rituximab achieved a significant improvement of overall survival (OS) rates in DLBCL patients compared to therapy with chemotherapeutics/glucocorticoids (CHOP) alone (Coiffier *et al.*, 2010; Pfreundschuh *et al.*, 2011). While treatment with R-CHOP leads to durable remissions in a considerable proportion of patients, more than 30% of all DLBCL patients are either refractory to the therapy or will relapse with resistant tumours (Friedberg, 2011). Consequently, novel therapeutic combinations and molecularly targeted therapies are currently explored and tested in clinical trials (Roschewski *et al.*, 2014).

3.4.2 Classification and genetic landscape of diffuse large B-cell lymphoma

Whole exome/genome sequencing studies revealed that the mutational profile of DLBCL is extremely heterogeneous (Pasqualucci *et al.*, 2011b; Lohr *et al.*, 2012; Morin *et al.*, 2013; Zhang *et al.*, 2013; de Miranda *et al.*, 2014; Novak *et al.*, 2015). However, gene expression profiling can be utilised to sub-classify DLBCL into three subtypes: germinal centre (GC) B-cell-like (GCB)-DLBCL, activated B-cell-like (ABC)-DLBCL and primary mediastinal B-cell lymphoma (PMBL) (Alizadeh *et al.*, 2000). While PMBL cases are rare and have a very favourable prognosis (Dunleavy *et al.*, 2013), GCB- and ABC-DLBCL are the two major DLBCL subtypes.

ABC-DLBCL is the most aggressive form of DLBCL, and patients have a worse prognosis regarding progression-free survival (PFS) and OS compared to the GCB-subtype (three-year PFS rates: 40% vs 74%; three-year OS rates: 45% vs 80%) (Lenz *et al.*, 2008b). Post-GC activated B-cells (that have already undergone somatic hypermutation (SHM; process which introduces mismatches into the variable regions of the B-cell receptor that can enhance affinities between antigens and antibodies)) are most likely the cell of origin for ABC-DLBCL

whereas GCB-DLBCL resembles the transcriptional profile of GC B-cells with ongoing SHM (Lenz *et al.*, 2008b).

For both subtypes, inactivating mutations in the epigenetic modifiers *CREBBP*, *EP300*, *KMT2C (MLL3)* and *KMT2D (MLL2)* are frequent (Pasqualucci *et al.*, 2011a) while gain-of-function mutations in the histone-lysine N-methyltransferase *EZH2* are exclusive for GCB-subtype and found in about 20% of all cases (Morin *et al.*, 2010). Other alterations, which are only occurring in GCB-DLBCL, are loss-of-function mutations in *TNFRSF14* (which is a member of the tumour necrosis factor-receptor superfamily) and in *GNA13* (a G protein involved in Rho GTPase signalling) (Morin *et al.*, 2011; Morin *et al.*, 2013). Moreover, *PTEN* deletions are characteristic for GCB-DLBCL and are only infrequently present in the ABC-subtype (Pfeifer *et al.*, 2013). In contrast, the hallmark of ABC-DLBCL is constitutive active B-cell receptor/NF-κB signalling reflected by inactivating mutations in *TNFAIP3* and activating mutations in *CARD11*, *CD79A/B* and *MYD88* (Lenz *et al.*, 2008a; Compagno *et al.*, 2009; Davis *et al.*, 2010). In addition, up to 30% of all ABC-DLBCL cases show inactivating mutations in *PRDM1/BLIMP1*, a gene, which is indispensable for the terminal differentiation of B cells into antibody-producing plasma cells (Pasqualucci *et al.*, 2006). Common for both subtypes are genetic lesions associated with immune escape like loss-of-function mutations in *B2M* (about 30% of all cases) and *CD58* (about 20% of all cases), which affect major histocompatibility complex (MHC) I expression and activation of T cells and NK cells, respectively (Challa-Malladi *et al.*, 2011).

DLBCL cases not only show alterations on a mutational level, chromosomal rearrangements, like copy number variations and translocations, are also common. While rearrangements involving *MYC* are detectable in both subtypes (Savage *et al.*, 2009), t(14;18) translocations leading to overexpression of *BCL2* by the *immunoglobulin heavy chain (IgH)* gene enhancer are restricted to GCB-DLBCL and are identified in about one-third of all cases (Iqbal *et al.*, 2004). In contrast, t(3;14) translocations involving the *IgH* enhancer and *BCL6* are more often detected in the ABC-subtype (Iqbal *et al.*, 2007).

Lastly, aberrant SHM and class switch recombination (CSR), leading to hypermutation and translocation of genes outside the immunoglobulin gene clusters, are thought to be major contributing factors to DLBCL pathogenesis. Targets of illegitimate SHM and CSR include, for example, the well-known DLBCL oncogenes *BCL6*, *MYC* and *PAX5* (Pasqualucci *et al.*, 2001; Lenz *et al.*, 2007).

3.4.3 Mouse models of diffuse large B-cell lymphoma

While development of mature B-cell lymphomas is quite common in aged mice of certain inbred strains, only a few genetically engineered mouse models (GEMM) of DLBCL exist. Furthermore, in most GEMMs the penetrance for DLBCL is quite low. Exemplarily, three DLBCL mouse models are shortly described below.

Cattoretti *et al.* developed a mouse model in which the *Bcl6* gene is under control of the *IgH* enhancer mimicking the rearrangement identified in human DLBCL (Chapter 3.4.2) (Cattoretti *et al.*, 2005). Between 25% and 50% of all mice, starting at approximately 13 months of age, developed DLBCL though some mice displayed other mature B-cell lymphoma phenotypes.

In 2010, two mouse models for ABC-DLBCL were published, both based on B-cell specific inactivation of the TSG *Prdm1/Blimp1*, which is an important transcription factor regulating plasma cell differentiation (Minnich *et al.*, 2016) (Chapter 3.4.2) (Calado *et al.*, 2010; Mandelbaum *et al.*, 2010). Aged mice (>12 months) showed lymphoproliferative disease (LPD) and, occasionally, DLBCL.

Lastly, to generate another model for ABC-DLBCL, conditional $Myd88^{L252P}$ (Chapter 3.4.2) knockin mice were crossed with various B-cell specific Cre lines (*Cd19*, *Aid* and *Cd21*), leading to constitutive active NF-κB signalling in B-lymphocytes (Knittel *et al.*, 2016). Mice developed LPD and, in rare cases, ABC-DLBCL. Since *BCL2* amplifications and *MYD88* mutations often co-occur in human ABC-DLBCL cases (Wang *et al.*, 2014a), mice conditionally overexpressing *Bcl2* were crossed with $Cd19^{Cre}$;*LSL-*$Myd88^{L252P/+}$ mice. These triple-transgenic animals died significantly earlier than $Cd19^{Cre}$;*LSL-*$Myd88^{L252P/+}$ mice alone and showed a higher prevalence of ABC-DLBCL, demonstrating the co-operative effect of *Bcl2* and *Myd88* activation *in vivo*.

3.5 Liver tumourigenesis

3.5.1 Epidemiology, classification and treatment of primary liver cancer

Primary liver cancer (PLC) is the fifth and ninth most frequent cancer in men and women, respectively, and the second most common cause of cancer related death worldwide (Ferlay *et al.*, 2015). The most frequent types of PLC are hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC). While HCC accounts for 85-90% of all PLCs (Llovet *et al.*, 2016), ICC is less common and comprises up to 10% of all PLC cases (Shaib and El-Serag, 2004). Both, HCC and ICC, are much more frequent in men than in women and have their highest incidence rates in less developed countries (with “hotspots” being (South-)Eastern Asian countries for HCC and ICC and Sub-Saharan African countries for HCC) (Ferlay *et al.*,

2015; Torre *et al.*, 2015). The main reason for this geographic imbalance is the high prevalence of hepatitis B (HBV) infections (HCC and ICC) and liver fluke disease (LFD) (ICC) in (South-)Eastern Asia and of aflatoxin B1 exposure (HCC) in Sub-Saharan Africa, all being major risk factors for PLC development (Sia *et al.*, 2013; Llovet *et al.*, 2016). Other HCC/ICC risk factors include cirrhosis, hepatitis C (HCV) infections, alcohol abuse and metabolic syndrome (Llovet *et al.*, 2016). In addition, adeno-associated virus 2 infections (Nault *et al.*, 2015), non-alcoholic steatohepatitis (NASH) and non-alcoholic fatty liver disease (NAFLD) (Llovet *et al.*, 2016) are associated with risk of HCC development while hepatolithiasis and primary sclerosing cholangitis are ICC risk factors (Razumilava and Gores, 2014).

The five-year survival rate of PLC is less than 20%, as HCC and ICC are often diagnosed in advanced disease stages (Siegel *et al.*, 2016). Treatment options for HCC include surgical resection, liver transplantation, radiofrequency ablation (early stages; potentially curative treatments), chemoembolisation (intermediate stage; palliative treatment) and treatment with the tyrosine kinase inhibitor sorafenib (advanced stage; palliative treatment) (Llovet *et al.*, 2016). For ICC, the only current curative treatment is surgical resection while patients with advanced disease might have a survival benefit when receiving local or systemic chemotherapy (Razumilava and Gores, 2014). For both, HCC and ICC, numerous clinical trials of molecularly targeted therapy are ongoing (Forner *et al.*, 2012; Sia *et al.*, 2013; Llovet *et al.*, 2016).

3.5.2 Genetic landscape of primary liver cancer

The genomic landscape of HCC varies widely between the different aetiologies (e.g. alcohol-, HBV-, HCV-induced HCCs). In general, the most frequently mutated genes in HCC are *CTNNB1*, *TERT* and *TP53*. While *CTNNB1* and *TERT* mutations are more often identified in HCCs associated with alcohol abuse, mutations in *TP53* are particularly common in HBV-related tumours (Schulze *et al.*, 2015). Moreover, inactivating mutations in genes encoding components of chromatin remodelling complexes like *ARID1A*, *ARID1B* and *ARID2* are – in tendency – more frequent in alcohol-related HCCs, but also common in other HCC aetiologies (Li *et al.*, 2011; Huang *et al.*, 2012). Other recurrently altered genes in HCC include the cell cycle related genes *RB1*, *CCND1* and *CDKN2A*, the negative regulator of *WNT* signalling *AXIN1* and epigenetic regulators of the *MLL* family (e.g. *KMT2D (MLL2)*, *KMT2C (MLL3)* and *KMT2B (MLL4)*) (Totoki *et al.*, 2011; Guichard *et al.*, 2012; Sung *et al.*, 2012).

As for HCC, the mutational profile of ICC is quite diverse, i.e. differing between ICCs of distinct aetiologies. For instance, mutation in epigenetic regulators (e.g. *ARID1A*, *BAP1* and *PBRM1*)

are less often detected in LFD-associated ICCs than in HBC/HCV-associated cancers (Chan-On *et al.*, 2013; Jiao *et al.*, 2013). In contrast, LFD-related tumours more often show mutations in the TSGs *TP53*, *SMAD4* and *KMT2C (MLL3)*. The most prominent oncogenes over all ICC aetiologies are *KRAS*, *IDH1*, *IDH2* and *GNAS* (Ong *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Ross *et al.*, 2014).

3.5.3 Mouse models of primary liver cancer

To study HCC pathogenesis *in vivo*, (i) chemically induced models, (ii) diet-based models, (iii) transplantation-based approaches, (iv) viral and transfection-based delivery of transgenes (e.g. by HTVI) and (v) GEMMs are utilised (Heindryckx *et al.*, 2009).

Often applied chemicals for HCC induction are the carcinogen *N-nitrosodiethylamine (DEN)* and the hepatotoxin CCl_4. The pathomechanism of the latter is based on radical production causing lipid peroxidation and membrane damage in liver cells and an inflammatory response by Kupffer cells ultimately leading to tissue injury (Boll *et al.*, 2001; Luckey and Petersen, 2001). An example for diet-based models are mice receiving choline-deficient food, which develop HCCs after one year of age (Knight *et al.*, 2000).

In contrast, most conditional GEMMs rely on the liver-specific *Albumin-Cre (Albumin^{Cre})* mouse line (Postic *et al.*, 1999). Since HBV and HCV infections are major risk factors for HCC development, numerous GEMMs expressing viral HBV/HCV genes exist. For instance, mice which express the HBV antigen *HBx* in their livers usually display HCCs within one to two years of age (Takada *et al.*, 1995). Moreover, mice overexpressing oncogenes or with disruption (knockout) of TSGs are frequently used. Examples include mouse models with liver-specific overexpression of growth factors like *Tgf alpha* (Jhappan *et al.*, 1990), *Egf* (Borlak *et al.*, 2005) and *Fgf19* (Nicholes *et al.*, 2002), and with hepato-specific knockout of *Pten* (Watanabe *et al.*, 2007), *Abcb4* (an ABC transporter) (Katzenellenbogen *et al.*, 2007) and *Tak1* (a NF-κB signalling component) (Bettermann *et al.*, 2010).

As for HCC, several ICC GEMMs using the *Albumin^{Cre}* mouse line for driving liver-specific expression/depletion of genes exist. Models include combinatorial inactivation of *Pten* and *Smad4* (Xu *et al.*, 2006) and expression of oncogenic *Kras^{G12D}* together with *Trp53* inactivation (O'Dell *et al.*, 2012). Mice of the latter model developed ICCs with widespread metastases between a mean age of 19 weeks (homozygous *Trp53* knockout) and 56 weeks (heterozygous *Trp53* knockout), respectively. Additionally, liver-specific activation of the *Notch 1* intracellular domain (Zender *et al.*, 2013) or the *Idh2* oncogene (Saha *et al.*, 2014) also leads to ICC development in mice.

3.6 Bloom deficiency

Bloom syndrome (also called Bloom-Torre-Machacek syndrome) is a rare autosomal-recessive disorder that is caused by mutations in the RecQ DNA helicase *BLM* (Ellis *et al.*, 1995), and has its highest prevalence among Ashkenazi Jews (Li et al., 1998). Patients have a short stature, distinct morphologic features (e.g. a long, narrow face and prominent ears), exhibit sunlight sensitivity, and have a high risk to develop cancer of all cell types. Bloom syndrome's hallmark phenotype (and underlying disease mechanism) is genomic instability characterised by elevated rates of sister chromatid exchange (SCE) in somatic cells (about ten-fold higher than in cells of unaffected individuals) (Cunniff *et al.*, 2017).

In 2000, Bloom-mutated *($Blm^{m3/m3}$)* mice mimicking this disease were generated in the laboratory of Allan Bradley (then Baylor College of Medicine, Houston, Texas, USA) (Luo *et al.*, 2000). *$Blm^{m3/m3}$* mice, harbouring two hypomorphic *Blm* alleles, are viable, fertile and show no signs of growth retardation or other developmental abnormalities. However, comparable to human Bloom syndrome patients, they exhibit increased SCE rates (Figure 3-4) and about one-third of *$Blm^{m3/m3}$* mice develop tumours by the age of 20 months. Observed cancer types predominantly include lymphomas but also sarcomas and adenocarcinomas (e.g. of the liver, lung and intestine). Since elevated SCE rates result in an approximately 18-fold increased loss of heterozygosity (LOH) (Luo *et al.*, 2000), Blm-mutated embryonic stem cells and *$Blm^{m3/m3}$* mice are excellent models for TSG studies (Yusa *et al.*, 2004; Suzuki *et al.*, 2006; Wang *et al.*, 2009) as classical tumour suppressors require biallelic inactivation (Sherr, 2004).

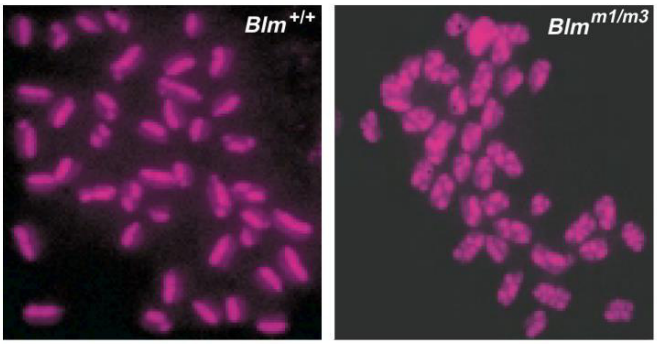

FIGURE 3-4. Elevated rates of sister chromatid exchange in Bloom-mutated mice.

Metaphase spreads of embryonic stem cells derived from wild type (left) and Bloom-mutated mice (in this case *$Blm^{m1/m3}$* animals) (right). Variations in sister chromatid exchange rates are visualised by differentially labelled sister chromatids (with either high or low intensity). Image reprinted by permission from the Nature Publishing Group (Luo *et al.,* 2000).

4 AIMS

As mentioned above, transposon-based insertional mutagenesis and the CRISPR/Cas system have proven to be invaluable tools for cancer gene screening and validation, respectively. Aim of this thesis was the further adaptation of both technologies for recessive cancer gene screening *in vivo*.

While transposons have been used for cancer gene discovery *in vivo* for the last decade, specific transposon-based recessive screening systems have been lacking so far. Therefore, this thesis wanted to investigate if:

- whole-body recessive transposition systems using the specifically designed inactivating *ITP1* and *ITP2* transposons are capable of inducing a broad tumour spectrum in mice;
- novel TSGs can be identified by analysing tumours derived from the recessive screen;
- a *Blm*-mutated background fosters LOH of TSGs *in vivo*;
- transposon insertion data obtained from the recessive screen can be exploited to systematically compare the *in vivo* characteristics of *Sleeping Beauty* and *PiggyBac*.

For the CRISPR/Cas system, it was shown that it could be applied for candidate gene validation *in vitro* and *in vivo* as well as for cancer gene screening *in vitro* and *ex vivo*. However, *in vivo* screening approaches for cancer gene discovery have been missing as of yet. Hence, this thesis aimed to analyse if:

- *in vivo* cancer modelling in the adult mouse liver is feasible using two different mouse models:
  (i) mice with liver-specific expression of oncogenic $Kras^{G12D}$ and
  (ii) mice treated with CCl_4 as part of a chemical-mediated liver injury model;
- CRISPR/Cas multiplexing can be exploited to perform a recessive genetic screen *in vivo* for TSG discovery;
- CRISPR/Cas *in vivo* multiplexing leads to unwanted off-target effects on either a mutational level (occurrence of indels at sites of the genome with sequence similarity to the target sequence) or regarding intra- and interchromosomal rearrangements.

5 MATERIALS

For detailed manufacturer information, see Table 5-10. All oligonucleotides were synthesised by Eurofins Genomics.

5.1 Equipment

Equipment	Source
Agilent Bioanalyzer 2100	Agilent Technologies
Centrifuge 5424	Eppendorf
Centrifuge 5810 R	Eppendorf
Class II Biological Safety Cabinet	Thermo Fisher Scientific
DynaMag™-96 Side Skirted Magnet	Thermo Fisher Scientific
Homogenisator Precellys® 24	Bertin Instruments
Incubator NCU-Line® IL 23	VWR International
M220 Focused Ultrasonicator	Covaris
Magnetic stirrer D-6010	neoLab
MiSeq System	Illumina
NU-5500 Incubator	NuAire
Primovert Microscope	Carl Zeiss
Qubit® 2.0 Fluorometer	Thermo Fisher Scientific
ROCKER 2D digital	IKA-Werke
StepOne Plus Real-Time PCR System	Applied Biosystems
Thermocycler Tpersonal 48	Biometra
Thermocycler TProfessional Basic 96	Biometra
ThermocyclerTProfessional Basic Gradient 96	Biometra
Thermomixer MHR 23	Ditabis
ThermoMixer® C	Eppendorf
ThermoMixer® comfort 5355	Eppendorf
Ultra Low-Temperature Freezer Innova® U725	Eppendorf
UVsolo 2 Gel Documentation System	Analytik Jena
Vortex-Genie 2	Scientific Industries
Weighing Scale A120S	Sartorius

TABLE 5-1. Equipment.

5.2 Reagents and enzymes

Reagent/Enzyme	Source
1 kb DNA Ladder	New England Biolabs
100 bp DNA Ladder	New England Biolabs
Acetic acid	Sigma-Aldrich
AflIII (10,000 units/mL)	New England Biolabs
Agarose	Sigma-Aldrich
Ampicillin	Sigma-Aldrich
BbsI (10,000 units/mL)	New England Biolabs
Calcium chloride	Carl Roth
Carbon tetrachloride *(CCl_4)*	Sigma-Aldrich
Corn oil	Sigma-Aldrich
Deoxynucleotide Mix, 10 mM each	Sigma-Aldrich
Dimethyl sulfoxide *(DMSO)*	Carl Roth
DirectPCR Lysis Reagent (Cell)	Viagen Biotech
Eosine	Waldeck
Ethanol absolute	Carl Roth
Ethidium bromide	Sigma-Aldrich
Ethylenediaminetetraacetic acid *(EDTA)*	Sigma-Aldrich
Forene® isoflurane	Abbott
Formalin	Carl Roth
Gel Loading Dye, Purple (6x)	New England Biolabs
Glycerol	Sigma-Aldrich
Haematoxylin	Merck
Isopropanol absolute	Carl Roth
LB-Agar (Luria/Miller)	Carl Roth
LB-Medium (Luria/Miller)	Carl Roth
Lipofectamine® 2000	Thermo Fisher Scientific
Magnesium chloride	Carl Roth
Magnesium chloride *($MgCl_2$)*	Carl Roth
Magnesium sulfate *($MgSO_4$)*	Carl Roth
NEBuffer 2	New England Biolabs
NotI (10,000 units/mL)	New England Biolabs
Pertex mounting medium	Medite
Phosphate buffered saline	Sigma-Aldrich
Polyethylene glycol 4000	Sigma-Aldrich
Potassium chloride *(KCl)*	Carl Roth
Proteinase K	Sigma-Aldrich
Q5® High-Fidelity DNA Polymerase	New England Biolabs
RNAlater	Sigma-Aldrich
Roti®-Histofix 4 %	Carl Roth
Shrimp Alkaline Phosphatase *(rSAP)* (1,000 units/mL)	New England Biolabs
Sodium chloride *(NaCl)*	Carl Roth
Surveyor® Mutation Detection Kit	Integrated DNA Technologies
SYBR® Select Master Mix	Thermo Fisher Scientific
T4 DNA Ligase	New England Biolabs
T4 Polynucleotide Kinase	New England Biolabs
TaKaRa Ex Taq DNA Polymerase	Takara Bio

Reagent/Enzyme	Source
Taq DNA Polymerase	New England Biolabs
Tris(hydroxymethyl)aminomethane *(TRIS)*	Sigma-Aldrich
Tris(hydroxymethyl)aminomethane hydrochloride *(TRIS-HCl)*	Sigma-Aldrich
Xylene	Carl Roth

TABLE 5-2. Reagents and enzymes.

5.3 Kits for nucleic acid isolation, purification and quantification

Kit	Source
DNeasy Blood & Tissue Kit	Qiagen
MinElute Reaction Cleanup Kit	Qiagen
NucleoBond® Xtra Midi EF	Macherey-Nagel
QIAprep Spin Miniprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification	Qiagen
Qubit® dsDNA BR Assay Kit	Thermo Fisher Scientific

TABLE 5-3. Reagents for nucleic acid isolation, purification and quantification.

5.4 Reagents for library preparation and sequencing

Reagent	Source
Agencourt AMPure XP magnetic beads	Genewiz
Agilent High Sensitivity DNA Kit	Agilent Technologies
EB buffer	Qiagen
KAPA DNA Standards and Primers for Illumina	Kapa Biosystems
KAPA HiFi HotStart ReadyMix (2x)	Kapa Biosystems
KAPA SYBR Fast qPCR ABI Mix (2x)	Kapa Biosystems
MiSeq Reagent Kit v2 (300 cycle)	Illumina
MiSeq Reagent Kit v3 (600 cycle)	Illumina
NEBNext DNA Sample Prep Reagent Set 1	New England Biolabs
NEBNext® Ultra DNA Library Prep Kit for Illumina®	New England Biolabs
Sodium hydroxide *(NaOH)*	Carl Roth

TABLE 5-4. Reagents for library preparation and sequencing.

5.5 Reagents for cell culture

Reagent	Source
Collagenase Type II	Worthington Biochemical
DMEM, high-glucose	Sigma-Aldrich
DPBS, no calcium, no magnesium	Thermo Fisher Scientific
FBS Superior	Biochrom
Gelatine	Sigma-Aldrich
L-Glutamine (200 mM)	Thermo Fisher Scientific
MEM Non-Essential Amino Acids Solution (100X)	Thermo Fisher Scientific
Penicillin-Streptomycin (5,000 U/ml)	Thermo Fisher Scientific
RPMI 1640 Medium	Thermo Fisher Scientific
Trypsin-EDTA (0.5%)	Thermo Fisher Scientific

TABLE 5-5. Reagents for cell culture.

5.6 Antibodies

Antibody	Source
Primary Antibodies	
A6	in-house production
AFP (AF5369)	R&D Systems
B220/CD45R (RA3-6B2)	R&D Systems
CD138 (281-2)	BD Biosciences
CD3 (A0452)	DAKO
Collagen-4 (CL50451AP)	Cedarlane
Cytokeratin 19 (TROMA-III)	Developmental Studies Hybridoma Bank
GP73 (sc-48011)	Santa Cruz Biotechnology
Ki67 (RM-9106)	Thermo Fisher Scientific
Myeloperoxidase (A0398)	DAKO
Secondary Antibodies	
AffiniPure Goat Anti-Rabbit IgG (H+L) (111-005-003)	Jackson ImmunoResearch
AffiniPure Rabbit Anti-Rat IgG (312-005-045)	Jackson ImmunoResearch
Polyclonal Rabbit Anti-Goat Immunoglobulins (P044901-2)	Agilent Technologies

TABLE 5-6. Antibodies.

5.7 Plasmids

Plasmid	Source
hSB5	Gift from Ursula Ehmer (Klinikum rechts der Isar, Technical University Munich, München, Germany)
pcDNA™6.2/EmGFP-Bsd/V5-DEST	Thermo Fisher Scientific

Plasmid	Source
pTnori	Gift from Ursula Ehmer (Klinikum rechts der Isar, Technical University Munich, München, Germany)
pX330	Addgene

TABLE 5-7. Plasmids.

5.8 Consumables

Consumable	Source
ABgene Storage Plate, 96-well, 2.2 mL, square well, conical	Thermo Fisher Scientific
Adhesive PCR Plate Foils	Thermo Fisher Scientific
Biopsy/tissue embedding cassettes	Simport
Cell culture dishes (100 mm)	Greiner Bio-One
Cell culture flasks (50 mL, 250 mL, 550 mL)	Greiner Bio-One
Cell culture plates (6-well, 12-well, 24-well, 96-well)	Corning
Cell scrapers	Sarstedt
Cell strainers (70 µm, 100 µm)	Corning
Combitips advanced® (0.2 mL, 0.5 mL, 1 mL, 5 mL, 10 ml)	Eppendorf
Conical tubes (15 mL, 50 mL)	Greiner Bio-One
Cover slips	Gerhard Menzel B.V.
Cryotubes (1.6 mL)	Sarstedt
Disposable blades	Swann-Morton
Disposable reservoirs	Integra Biosciences
Disposable scalpels	B. Braun Melsungen
Disposable spatulas	Carl Roth
DNA LoBind Tubes (1.5 mL)	Eppendorf
Glass slides SuperFrost™ Plus	Thermo Fisher Scientific
Hard-Shell® 96-Well PCR Plates, high profile, semi skirted	Bio-Rad Laboratories
Hard-Shell® Low-Profile Thin-Wall 96-Well Skirted PCR Plate	Bio-Rad Laboratories
MicroAmp® optical 96-well reaction plate	Thermo Fisher Scientific
MicroAmp® Optical Adhesive Film	Thermo Fisher Scientific
Microtome blades S35	Feather Safety Razor
microTUBE AFA Fiber Snap-Cap 6x16mm Case	Covaris
Needles 27 gauge	Seidel medipool
Pasteur pipettes	Brand
PCR stripes (8 tubes)	Sarstedt
Petri dishes (100 mm)	Greiner Bio-One
Pipette tips (10 µL, 200 µL)	Biozym
Pipette tips with filter (10 µL, 100 µL, 200 µL, 300 µL, 1250 µL)	Biozym
Reaction tubes safe-seal (0.5 mL, 1.5 mL, 2 mL)	Sarstedt
Reaction tubes safe-seal (5 mL)	Eppendorf
Serological pipettes (5 mL, 10 mL, 25 mL, 50 mL)	Greiner Bio-One
Syringes (1 mL, 30 mL)	B. Braun Melsungen

TABLE 5-8. Consumables.

5.9 Software

Software	Source
Genomic Workbench 7	Agilent Technologies
GraphPad Prism5	GraphPad Software
Office 2016	Microsoft Corporation
R Software Environment	The R Project, The R Foundation
Snapgene 3.1	GSL Biotech

TABLE 5-9. Software.

5.10 Manufacturers

Manufacturer	(Headquarter) Location
Abbott GmbH	Ludwigshafen, Germany
Addgene	Cambridge, Massachusetts, USA
Agilent Technologies, Inc.	Santa Clara, CA, USA
Analytik Jena AG	Jena, Germany
Applied Biosystems, Inc.	Carlsbad, CA, USA
B. Braun Melsungen AG	Melsungen, Germany
BD Biosciences, BD, Inc.	Franklin Lakes, NJ, USA
Bertin Instruments	Montigny-le-Bretonneux, France
Biochrom GmbH	Berlin, Germany
Biometra GmbH	Göttingen, Germany
Bio-Rad Laboratories, Inc.	Hercules, CA, USA
Biozym Scientific GmbH	Hessisch Oldendorf, Germany
Brand GmbH	Wertheim, Germany
Carl Roth	Karlsruhe, Germany
Carl Zeiss AG	Oberkochen, Germany
Cedarlane, Inc.	Burlington, ON, Canada
Corning, Inc.	Corning, NY, USA
Covaris, Inc.	Woburn, MA, USA
DAKO, Agilent Technologies, Inc.	Santa Clara, CA, USA
Developmental Studies Hybridoma Bank	Iowa City, IA, USA
Ditabis AG	Pforzheim, Germany
Eppendorf AG	Hamburg, Germany
Eurofins Genomics GmbH	Ebersberg, Germany
Feather Safety Razor Co., Ltd.	Osaka, Japan
Genewiz, Inc.	South Plainfield, NJ, USA
GraphPad Software, Inc.	San Diego, CA, USA
Greiner Bio-One GmbH	Kremsmünster, Austria
GSL Biotech LLC	Chicago, IL, USA
IKA-Werke GmbH	Staufen, Germany

Manufacturer	(Headquarter) Location
Illumina, Inc.	San Diego, CA, USA
Integra Biosciences AG	Biebertal, Germany
Integrated DNA Technologies, Inc.	Coralville, IA, USA
Jackson ImmunoResearch, Inc.	West Grove, PA, USA
Kapa Biosystems, Inc.	Wilmington, MA, USA
Macherey-Nagel GmbH	Düren, Germany
Medite GmbH	Burgdorf, Germany
Merck KGaA	Darmstadt, Germany
Microsoft Cooperation	Redmond, WA, USA
neoLab Migge GmbH	Heidelberg, Germany
New England Biolabs, Inc.	Ipswich, MA, USA
NuAire	Plymouth, MN, USA
Qiagen GmbH	Hilden, Germany
R&D Systems, Inc.	Minneapolis, MN, USA
Santa Cruz Biotechnology, Inc.	Dallas, TX, USA
Sarstedt AG	Nümbrecht, Germany
Sartorius AG	Göttingen, Germany
Scientific Industries, Inc.	Bohemia, NY, USA
Seidel medipool GmbH	Gauting-Buchendor, Germany
Sigma-Aldrich Corporation	St. Louis, MO, USA
Simport Scientific, Inc.	Beloeil, QC, Canada
Swann-Morton, Ltd.	Sheffield, United Kingdom
Takara Bio, Inc.	Kyoto, Japan
The R Project, The R Foundation	Vienna, Austria
Thermo Fisher Scientific Gerhard Menzel B.V.	Braunschweig, Germany
Thermo Fisher Scientific, Inc.	Waltham, MA, USA
Viagen Biotech, Inc.	Los Angeles, CA, USA
VWR International GmbH	Darmstadt, Germany
Waldeck GmbH	Münster, Germany
Worthington Biochemical Corporation	Lakewood, NJ, USA

TABLE 5-10. Manufacturers.

5.11 Buffers

Buffer	Composition
5x KCM Buffer	500 mM KCl 150 mM $CaCl_2$ 250 mM $MgCl_2$
1x Low TE Buffer	10 mM Tris-HCl, pH 8.0 0.1 mM EDTA

Buffer	Composition
50x Tris acetate EDTA (TAE) buffer, pH 8.5	2 M Tris 50 mM EDTA 5.71% Acetic acid
Tryptic soy broth (TSB) buffer	LB-Medium (Luria/Miller) pH 6.1 with 10% Polyethylene glycol 4000 5% DMSO 10 mM $MgCl_2$ 10 mM $MgSO_4$

TABLE 5-11. Buffers.

All buffers were prepared with bidestilled H_2O.

5.12 Oligonucleotides

Name	Forward Primer Sequence (5'→3')	Reverse Primer Sequence (5'→3')
Bloom-mut	TCATTTTGGCAGTCCACCTC	GTCGCTCTAATCCTTTCCATT
Bloom-wt	TCATTTTGGCAGTCCACCTC	TTAAGACCAGGGCTAGACAG
iPBase	GGCGGATCACAAGCAATAATAACCTGTAGTTT	CCAAAGTCGCTCTGAGTTGTTATCAG
ITP2-M	ACCTGGTTGTCATGGAGGAG	TGACGAGCTTGTTGGCTAGA
SB11-mut	TCCCTCGTGATCTGCAACTCCAGTCTT	GGAAATTGCTCCCAAGGATGAACCAGA
SB11-wt	CCAAAGTCGCTCTGAGTTGTTATCAG	GGCGGATCACAAGCAATAATAACCTGTAGTTT

TABLE 5-12. PCR primers for genotyping of transposon, transposase and Bloom-mutated mouse lines.

Name	Forward Primer Sequence (5'→3')	Reverse Primer Sequence (5'→3')
rs4137461	TGCATCCAAGTTAAATGTCAGGTAG	ACTGAATAGCCCTGCCTCTTT
rs220642642	AGCATGCTTCACTGGGAGAC	GGTCTGGAATTTGGCCATGC

TABLE 5-13. PCR primers for SNP analysis.

Name	Forward Oligonucleotide Sequence (5'→3')	Reverse Oligonucleotide Sequence (5'→3')
Apc_sgRNA	CACCGTCAGTTGTTAAAGCAAGTTG	AAACCAACTTGCTTTAACAACTGAC
Arid1a_sgRNA	CACCGTTAGTCCCACCATACGGCTG	AAACCAGCCGTATGGTGGGACTAAC
Brca1_sgRNA	CACCGAAATCTTAGAGTGTCCGATC	AAACGATCGGACACTCTAAGATTTC
Brca2_sgRNA	CACCGTAGGACCGATAAGCCTCAAT	AAACATTGAGGCTTATCGGTCCTAC
Cdkn2a-e1β_sgRNA	CACCGTGGTGAAGTTCGTGCGATCC	AAACGGATCGCACGAACTTCACCAC
Cdkn2a-e2_sgRNA	CACCGTGCGATATTTGCGTTCCGC	AAACGCGGAACGCAAATATCGCAC
Pten_sgRNA	CACCGCTAACGATCTCTTTGATGA	AAACTCATCAAAGAGATCGTTAGC
Smad4_sgRNA	CACCGACAACCCGCTCATAGTGATA	AAACTATCACTATGAGCGGGTTGTC
Tet2_sgRNA	CACCGAAAGTGCCAACAGATATCC	AAACGGATATCTGTTGGCACTTTC
Trp53_sgRNA	CACCGACACTCGGAGGGCTTCACT	AAACAGTGAAGCCCTCCGAGTGTC
Arid1b_sgRNA	CACCGTGTGCACCTGGGGGACCGT	AAACACGGTCCCCCAGGTGCACAC
Arid2_sgRNA	CACCGAGGCGCCTCCGGACGAGCG	AAACCCGCTCGTCCGGAGGCGCCTC
Arid5b_sgRNA	CACCGCTATGCAAATCGGATCCTT	AAACAAGGATCCGATTTGCATAGC
Atm_sgRNA	CACCGGCTGTCAACTTCCGAAAAC	AAACGTTTTCGGAAGTTGACAGCC
Cdkn2b_sgRNA	CACCGGCGCCTCCCGAAGCGGTTC	AAACGAACCGCTTCGGGAGGCGCC
Errfi1_sgRNA	CACCGAGCTCGGGACAGCGTGAAG	AAACCTTCACGCTGTCCCGAGCTC
Igsf10_sgRNA	CACCGGAGTCCGTAAAACGCCTCG	AAACCGAGGCGTTTTACGGACTCC
Irf2_sgRNA	CACCGTGCCGAGCCGCATGCATCC	AAACGGATGCATGCGGCTCGGCAC

TABLE 5-14. sgRNA oligonucleotides.

Name	Forward Primer Sequence (5'→3')	Reverse Primer Sequence (5'→3')
Apc_PCR	GCGAATAAGCACCACTCCTC	AAGAATGAACCAACACCAAGG
Arid1a_PCR	GTTCTGATTCCTGTGCTCGC	TCCATCACCTACCTGCTGTG
Brca1_PCR	AGCGTGAGAACTCCTCCAAA	CTGCCATGAGGAAGAACACA
Brca2_PCR	TCACGAGTTTCTCCGTGTCA	GCTCTGGCTGTCTCGAACTT
Cdkn2a-e1β_PCR	TCTCACCTCGCTTGTCACAG	AAGTACTCCATCTCCCGGGA
Cdkn2a-e2_PCR	TCAACTACGGTGCAGATTCG	CGGGTGGGTAAAATGGGAAC
Pten_PCR	TGCGAGGATTATCCGTCTTC	CATCCGTCTACTCCCACGTT
Smad4_PCR	TGCAGTGTCACAGATGCTCA	CTCAGGAACTGGAGGAAGCA
Tet2_PCR	CAGATGCTTAGGCCAATCAAG	AGAAGCAACACACATGAAGATG
Trp53_PCR	ACATAGCAAGTTGGAGGCCA	CCACTCACCGTGCACATAAC
Arid1b_PCR	AGTTCTGGGGTACTTGGAATCA	GGTACTGCAAGCCTCCCA
Arid2_PCR	ATGACTGAGCCCCGCCA	GAGCAGACTTTTCCGAGCAG
Arid5b_PCR	TGGCTTGCACGGACCTTATA	ATCAGCAGTTGGACGGTCTT
Atm_PCR	TCCTTTTCAACTGTTCCTGTTACA	GACAATGGAAAGGCGAGTCA
Cdkn2b_PCR	CCGAAGCTACTGGGTCTCC	CACTTGCCCAGCTTGTACG
Errfi1_PCR	GTGTTCCCCTACTCTGGCTC	TCTTCAGAGATGGGCAGTGG
Igsf10_PCR	CTGTCCACCTGAGTCCACTT	TGTCAGCCGGTTTCCTTCTA
Irf2_PCR	TGTCTGACAGTCGACTTCCC	ACTGGGAACTTCTGGGATGG

TABLE 5-15. PCR primers for sgRNA on-target region amplification.

Name	Forward Primer Sequence (5'→3')	Reverse Primer Sequence (5'→3')
hSpCas9_qPCR	GCCTATTCTGTGCTGGTGGT	ATCCCCAGCAGCTCTTTCAC
mouse_Apob_qPCR	CACGTGGGCTCCAGCATT	TCACCAGTCATTTCTGCCTTTG
CRISPR-SB_PCR	GAGGGCCTATTTCCCATGAT	CGACTCGGTGCCACTTTT
CRISPR-SB-qPCR	ACTATCATATGCTTACCGTAAC	
Cdkn2a-e1β_qPCR	CAAGAGAGGGTTTTCTTGGTGA	
Cdkn2a-e2_qPCR	ACAACATGTTCACGAAAGCCA	GGGACATCAAGACATCGTGC

TABLE 5-16. Quantitative real-time PCR primers.

Name	Forward Primer Sequence (5'→3')	Reverse Primer Sequence (5'→3')
Apc		
OT_Apc_1	CTGAGTGTGGTGCTATACTCAAG	ACTAGGATTAGGACCTAGGAAACA
OT_Apc_2	AGATCTGCAGTTCACCCCAA	GGGAGTCCAGGAAGCAGAAT
OT_Apc_3	AGTTACTGGTGGCTGTAAGACA	AGAGTGGCAGTTCAAGGTAGT
OT_Apc_4	ATCCAACGCTGATTCCTTGC	GGGAGGTGATTGAGAGGGAC
OT_Apc_5	CCTGGTTTTACGTTGCTGCT	CTATTTGCCTGCACCTCCAG
OT_Apc_6	CAATGCAAAAGGTGTTCTGACA	TCACCACCCTTGCTGTAACT
OT_Apc_7	CACTTGCTTCAGTCTGAGCC	CCTGCAGTCAACCTTGGTTC
OT_Apc_8	CGAACCTGTCAGTTGCAAGT	TGCGATGTTCTGGGCTATCT
Arid1a		
OT_Arid1a_1	TCCAGATGCCAACCCCTATC	GCCACAGACCCTATTCCTCA
OT_Arid1a_2	TGAGAGGGTCACGAGTTGG	CTATTGCCCCAGACCCAGAG
OT_Arid1a_3	TGTCTACGATCACAGTGCAGT	ACACAGGCTGTAACTCTGAAGA
OT_Arid1a_4	CAGAGGAAGTTGGGTGAGGA	TCATGCTCATCAGGGCTTCT
OT_Arid1a_5	GCCAACAGGTGAGTCTTCTAAC	CAGGCCCATGTTGTCTGAAG
OT_Arid1a_6	CGGCAAGTTCTGTTTGTGCT	GTCTGGGTCTCATCTCCTGG
OT_Arid1a_7	TCCTCGAAGTAGACATATCCACA	TGCAAAGGTTCTTCTGGAGC
Brca1		
OT_Brca1_1	GACTTCGTGGACAGAATGGC	TCCAGCCCTGTTTGATTCCT
OT_Brca1_2	GAGAACTGCAGAGCCCATTG	ACCGACATTTTCCCCTCCTT
OT_Brca1_3	TCCAAAGGCTGCTAGTGGAA	CCTCGACCCCTCCCAATTTT
OT_Brca1_4	CCCAACACAGCCCACTACA	ACCTGCAGAGTAAAGGGCTC
OT_Brca1_5	TGGATTCCAGCCTCTGTCAA	TGTCCCTAGCCAGTACCTCT
OT_Brca1_6	TAGCAGGGACCTCAAAGTGG	ATAGCAGCCCATGAAGCCAG
OT_Brca1_7	GCACTGTAAGCTCAACCCAG	CCTCTGCCACATGAGTACCA
OT_Brca1_8	ACATGACTGGAGTTAGAAAAGGA	TGTGCTTGCTATTCCTATGATGA
Brca2		
OT_Brca2_1	CACAGTAGGTTGGGTCTTCC	GACAGGGTTGGAGAGTGCC
OT_Brca2_2	GCGCTGTTATTTCCTCCGTT	AGCAAGGCCAGTGATCTCAT
OT_Brca2_3	TGAGCAAGTCACTTTGGAAAACA	AAGTGGGAACTTCAGGAGGG
OT_Brca2_4	CACTGAGTGTCATGCTTGGC	ACTAGTGAGCCCTGCCTTTC
OT_Brca2_5	GACACAGGAAGAGGGAGACA	ATCAAGCCACCAGAATCCCT
OT_Brca2_6	TGCATTTCCTTTGACACCAGT	ATCAGAGATCTCCGTGGCTG
OT_Brca2_7	AGAAGGAATTTGGGATTTTGGCA	TGGAGAGTGAGCTAGCCAAG

Name	Forward Primer Sequence (5'→3')	Reverse Primer Sequence (5'→3')
Cdkn2a-e1β		
OT_Cdkn2a-e1β_1	GCTTCCCTGAAACCTGCATC	CATCAAGGACTAGGAGCAATGA
OT_Cdkn2a-e1β_2	GTTGCCCTCATCTCAGACCT	TTCCAAGTGCAGCAAAGGTC
OT_Cdkn2a-e1β_3	GCGACTCACTCCAGGCTG	ACAAAAGGCATCTGGACAACT
OT_Cdkn2a-e1β_4	GGGGAGAGGGTCTAGAAGGA	TCCACAGATCATTGGCGAGA
OT_Cdkn2a-e1β_5	GGCATCTTTTCATTTGTCAGCC	ACACAGACACACAGATCCAAT
OT_Cdkn2a-e1β_6	ACTTCAGTGATCGCTAGGCC	CACACAGTGGGGCATAGAGA
OT_Cdkn2a-e1β_7	TGAGGACATGCACACAGACT	AATGCTTGGCTGGGTGATTG
OT_Cdkn2a-e1β_8	CTGCAGAGAGTTCCCAGGAA	CTCTTCATTGCTGATCCGCC
Cdkn2a-e2		
OT_Cdkn2a-e2_1	TGGGCTTGTTTTAAAGGGGC	CAATGTCTGCTGCTCACCTG
OT_Cdkn2a-e2_2	GTCTGTTTGGATGCCCTTGG	AGGCTACTCTTGCTGTCTCC
OT_Cdkn2a-e2_3	AAACTGAACTTGCTCGGCTC	TTGAGCATGAGAGGGAAGCA
OT_Cdkn2a-e2_4	TACCACTTCCTTCCCTGCAG	ATTGACTGTCCTACCCTGGG
OT_Cdkn2a-e2_5	TTACCTAACTCCTGGGGCAG	CAGGAAGCTAGACTGTGCCT
OT_Cdkn2a-e2_6	CCATCCTGTCAGTGGTTCCT	GCTACCTACCCACCACTACTC
OT_Cdkn2a-e2_7	ACTGGGGCATCTTCAGTCTC	AGTGAAAAGCCCCAATGATAAGT
Pten		
OT_Pten_1	CAAGAGAAAGACAAGGCATGGT	AGAAGGGAGGAGGGAAGGAA
OT_Pten_2	GGAGCAGCTTGGAGTCTGAT	CATTGCCAGCACAGTTCTCA
OT_Pten_3	GGAACATTAAGAGTGAAACAGCT	AAATAGGTGGCAGAACGGGT
OT_Pten_4	CATGCAACAACAGAGGACACA	TCCTTCTTCTGACCAAATGTGA
OT_Pten_5	AACAATGCTCAGAGGGTCCC	GATGGAATGTTGGGCCTCAA
OT_Pten_6	AAGGGTGGACTACAAAAGAGC	ACAGAAAGGTTTGTCTTGGCC
OT_Pten_7	GCTGTGGTATTTCAACTGGCT	TGACCTTCACGTTGCCAATG
OT_Pten_8	CCATAGCCATGTCCTCCCAT	GCTGCAAACATTAATGAAGAAGC
Smad4		
OT_Smad4_1	CATCATCTCCAAGGCCCTCA	GCCATTCCAGGGATCAAACC
OT_Smad4_2	CAGATATGGTGGTGCATGCC	TTGGAAAGCAGAGCAACAGG
OT_Smad4_3	GGGGTTCCTGGGAGTCTTTT	TACTGTGGCCTTGAGAAGCA
OT_Smad4_4	TAAGCAGCACTCACCACCAA	GCTCAGTCACCTAAGCTTGT
OT_Smad4_5	AAAGTGGGACTCATAGGGCC	TCCCGTCTCAGGTCACAAAA
OT_Smad4_6	TAATGCCTGCTGTCCCTTCA	TGAGATCATCTGACGGGCAA
Tet2		
OT_Tet2_1	AATTCAAGTGCAGAGCCAGG	GCCAGTCTGCAAATGAAATCT
OT_Tet2_2	CAACACACCTGCCTCCAAC	CTGAGTTCACTGTGCAAGCA
OT_Tet2_3	TCTAGGGAATGTGGCCTGAG	CCCTGCAGATCCCCTAAATGA
OT_Tet2_4	CCGCACCCATTTTCTGATAGG	CTTTCCGGTCCAGTTTCACC
OT_Tet2_5	GCTGTCCTGGAACTCACTCT	ACTGAGCCTAAGATTGTCCCA
OT_Tet2_6	TAATGCATCCTCCTTCACCCT	GGGGTTCAACATGGGGATCA
OT_Tet2_7	ACATGACCCAAGATTTCCCAA	GGCCTGAGAAGCGAAATGAG
OT_Tet2_8	CTATGAAGGCAAGGTGGGC	CATCCCCAGACTTACCCAGG

Name	Forward Primer Sequence (5'→3')	Reverse Primer Sequence (5'→3')
Trp53		
OT_Trp53_1	CCTAGCATTCAGGCCCTCAT	TGAGGGGAGGAGAGTACAGT
OT_Trp53_2	GGATTGTCCCTTGTACCACTTC	AACAAATGTGCGGGCAACTT
OT_Trp53_3	GCATGCACTGAACAGAAATTGG	TCAGAGGAGATTTGCTTGGGA
OT_Trp53_4	CCCTGGCTCTTCTGTGTGTA	GAACCCGCAGCATGTGATAG
OT_Trp53_5	CATGATGCCTGTTCACGAGG	CTGGTAAAAGGTGCTGGCTT
OT_Trp53_6	CATGCTGTTTGGGTGGAAGG	AGAAAAGAGGGGCTGGTTCC
OT_Trp53_7	CTACCCGGCAATGAACAGGT	CCAAGTGGCCAAGAAGCAAA
OT_Trp53_8	GGCTTGCCGTCTTTGTTGAT	AAGTGGACAGTTCTCCCAGC

TABLE 5-17. PCR primers for sgRNA off-target region amplification.

6 METHODS

Unless otherwise specified, all procedures were performed according to manufacturer's instructions.

6.1 General techniques

6.1.1 Generation and transformation of chemically competent bacteria

For all cloning experiments, homemade chemically competent bacteria (*DH10B*, *Stbl3* and *TOP10 E. coli* strains) were used.

For **generation of chemically competent bacteria**, bacteria were streaked and grown on agar plates containing 50 µg/mL streptomycin. A single bacterial colony was picked and 5 mL LB medium (with 50 µg/mL streptomycin) was inoculated. This culture was incubated at 37°C overnight and used the next morning for inoculation of 250 mL LB medium (containing 50 µg/mL streptomycin). Bacteria were grown at 37°C for approximately two hours until an optical density (OD_{600}) of 0.3-0.6 was reached. Then, bacteria were immediately placed on ice, centrifuged for 10 min at 1000 x g and 4°C and re-suspended in 25 mL ice-cold TSB buffer (Table 5-11). After 10 min incubation on ice, 500 µL bacteria aliquots were made and snap-frozen in liquid nitrogen. Bacteria were stored at -80°C until further usage.

For **transformation of chemically competent bacteria**, bacteria were thawed on ice for 20 min. 10-20 µL ligation reaction (or a few pg plasmid DNA for re-transformation) was mixed with 20 µL 5x KCM buffer (Table 5-11) and filled up with H_2O to 100 µL. 100 µL chemically competent bacteria were added, the mixture was first incubated for 20 min on ice and then for 10 min at room temperature (RT). After addition of 800 µL LB medium (without any antibiotic), bacteria were horizontally shaken (800 rpm) in a thermomixer at 28°C for 90 min (*DH10B*- and *Stbl3 E. coli*) or at 37°C for 60 min *(TOP10 E. coli)*. Bacteria were streaked and grown on agar plates containing the appropriate antibiotic (100 µg/mL ampicillin or 50 µg/mL kanamycin) at 28°C (*DH10B* and *Stbl3 E. coli*) or 37°C *(TOP10 E. coli)*.

6.1.2 Isolation and verification of plasmid DNA

For **isolation of plasmid DNA**, either plasmid mini- (QIAprep Spin Miniprep Kit) or midi-preparations (NucleoBond® Xtra Midi EF) were carried out. For both preparations, a single

bacterial colony was used to inoculate 5 mL (mini-preparations) or 150 mL (midi-preparations) LB medium containing the appropriate antibiotic (100 µg/mL ampicillin or 50 µg/mL kanamycin). Cultures were grown at 28°C for 24 h (*DH10B* and *Stbl3 E. coli*) or at 37°C for 16 h *(Top10 E. coli)* and DNA isolation procedures were performed according to manufacturer's instructions (for plasmid midi-preparations under endotoxin-free conditions). Plasmid DNA was quantified using the Qubit® fluorometer. In addition, diagnostic restriction digests or Sanger capillary sequencing were used for plasmid verification.

For **restriction digests**, usually 1 µg plasmid DNA was digested in a 20 µL reaction set-up with 10-50 units restriction enzyme (usually 0.5-1 µL) and the appropriate digestion buffer at the recommended temperature (37°C for most restriction enzymes) for 1 h. The result of the diagnostic digest was then visualised by gel electrophoresis (Chapter 6.1.4).

For **Sanger capillary sequencing**, 1-1,5 µg plasmid DNA was filled up with H_2O to 15 µL and 2 µL 10 µM sequencing primer (with a binding site within the plasmid) was added. Sanger capillary sequencing was performed at Eurofins Genomics. Sequences traces were visualised and analysed using SnapGene 3.1.

6.1.3 Isolation of DNA from cells and tissues

Isolation of DNA from cells and tissues was carried out using the Qiagen DNeasy Blood & Tissue Kit. Briefly, tissue was digested in ATL tissue lysis buffer supplied with 10% proteinase K (20 mg/mL) at 56°C overnight and the rest of the procedure was performed according to manufacturer's instructions. DNA concentrations were determined using the Qubit® fluorometer. For isolation of DNA from formalin-fixed, paraffin-embedded (FFPE) tissue, see Chapter 6.2.1.3.1.

6.1.4 Gel electrophoresis

For analytical gel electrophoresis, 1-2% agarose gels (depending on the size of the analysed DNA) were cast and run in 1x TAE buffer (Table 5-11). For visualisation of DNA bands, 1 mg/mL ethidium bromide was added in a 1:20,000 dilution to the gels. DNA was mixed with the appropriate volume of 6x DNA loading dye, a DNA ladder with a suitable fragment size (100-1,517 bp or 0.5-10 kb) was carried along and gels were generally run at 125 V for 75 min or until bands were separated adequately. Gel analysis was performed using an ultraviolet transillumination imaging system.

6.1.5 Histology

For histological analysis, tissue samples were fixed overnight in 4% formalin solution, dehydrated and embedded into paraffin. Paraffin blocks were cut in 3 µm and 10 µm thick sections for staining and DNA isolation (Chapter 6.2.1.3.1), respectively. For all samples, haematoxylin and eosin (H&E) staining was routinely performed according to standard protocols.

6.1.6 Primary cell culture

For establishment of primary cell cultures, mouse tumour tissues were washed in sterile DPBS and cut into small pieces (<1 mm) using a scalpel. Tissues were digested in RPMI 1640 supplemented with 10% foetal bovine serum (FBS), 1x penicillin-streptomycin (5,000 U/mL) and 200 U/mL collagenase type II at 37°C in a thermomixer (600 rpm) until complete disintegration was accomplished. Cells were then centrifuged and pellets were re-suspended in RPMI 1640 containing 10% FBS, 1x penicillin-streptomycin (5,000 U/mL), 1x L-glutamine (200 mM) and 1x non-essential amino acids and sown in six-well-plates coated with 1% gelatine. All cell culture experiments were performed in laminar flow hoods under sterile conditions. Cells were stored in cell culture incubators at 37°C and 5% carbon dioxide.

6.1.7 Animal experiments

All animal experiments were carried out in compliance with the requirements of the European guidelines for the care and use of laboratory animals and were approved by the local authorities. Mice were maintained in the animal facilities of the Wellcome Trust Sanger Institute, Hinxton/Cambridge, UK and Klinikum rechts der Isar, Technical University Munich, München, Germany under specific-pathogen-free (SPF) conditions on a 12 h light/dark cycle, receiving food and water *ad libitum.* For further information, see Chapters 6.2.1.2 and 6.2.2.5.

6.1.8 Statistical analyses

Diagrammatic representation and statistical analysis of the data was performed using GraphPad Prism5 and the R software environment. Methods for statistical hypothesis testing are directly stated in Chapter 7. If necessary, correction for multiple testing was applied using

the Bonferroni or Benjamini-Hochberg correction. In general, the significance level was set to 0.05.

6.2 Specific techniques

6.2.1 Transposon-based recessive genetic screening

6.2.1.1 Design of *ITP1* and *ITP2* transposons

In order to facilitate transposon-based TSG identification, inactivating transposons (ITP) were designed by Roland Rad (then Wellcome Trust Sanger Institute, Hinxton/Cambridge, UK). In contrast to ATP transposons, ITP transposons carry no promoters but only gene-trapping elements. Two different types of ITP transposons were generated: *ITP1* harbours, in addition to three poly(A) signals, a promoter-less β-galactosidase-neomycin fusion transcript *(bGEO)*, which can serve as a reporter to visualise gene trapping events. *ITP2* consists of a series of five bi-directional poly(A) signals. In both transposons, two splice acceptors in opposite orientation flank the reporter and/or gene-trapping elements. *SB* and *PB* ITRs allow transposition catalysed by the *SB* transposase as well as the *PB* transposase. The ITP trapping efficiency was tested by targeting ITP transposons to the X-chromosomal *Hprt* locus in embryonic stem cells derived from male mice. Efficient gene trapping at this locus confers resistance of the cells to *6-thioguanine*. The results showed that the gene-trapping efficiency is strongly influenced by the splice acceptor included in the gene-trapping cassette. Therefore, out of a panel of splice acceptors, the ones with the highest trapping efficiency (adenovirus splice acceptor and mouse *Engrailed 2 (En2)* exon-2 splice acceptor) were selected and used for the transposon design. Splice acceptor tests were performed by Roland Rad. Figure 6-1 shows design of *ITP1* and *ITP2*.

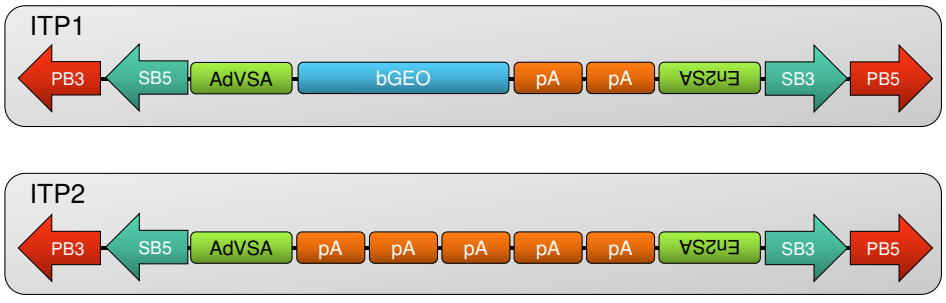

FIGURE 6-1. *ITP1* and *ITP2* design.

Two inactivating transposons, *ITP1* (top) and *ITP2* (bottom), were designed. Both transposons have *Sleeping Beauty* and *PiggyBac* inverted terminal repeats (ITRs) for mobilisation with either transposase, two splice acceptors and two *(ITP1)* or five *(ITP2)* bi-directional polyadenylation signals. Additionally, *ITP1* harbours a β-galactosidase-neomycin fusion transcript *(bGEO)* for visualisation of gene-trapping events. PB3, *PiggyBac* ITR 3'; SB5, *Sleeping Beauty* ITR 5'; AdVSA, adenovirus splice acceptor; pA, SV40 polyadenylation signal;

FIGURE 6-1 (continued)

En2SA, *Engrailed 2* exon-2 splice acceptor; SB3, *Sleeping Beauty* 3' ITR; PB5, *PiggyBac* 5' ITR.

6.2.1.2 Animal experiments

6.2.1.2.1 Generation of mouse strains

All mouse lines used for the ITP screen were already established at the Wellcome Trust Sanger Institute, Hinxton/Cambridge, UK.

Transgenic *ITP1* and *ITP2* mouse strains. For generation of transgenic *ITP1* and *ITP2* mouse strains, *ITP1* and *ITP2* transposons were cut out of the *pBluescript* donor vector by restriction digest. Linearised transposon constructs were then used for pronuclear injection to establish transgenic mouse lines following standard protocols. For each transposon, several mouse lines (*ITP1*: 2; *ITP2*: 3) were established. The transgenic mouse lines differ in type of transposon (*ITP1* or *ITP2*), transposon copy number and location of the transposon concatemer (donor locus) (Table 6-1). For the ITP screen, two medium-to-high copy number mouse lines (*ITP1-C* and *ITP2-M*) were selected. Roland Rad generated and characterised all ITP mouse lines.

Mouse Line Identifier	Transposon	Copy Number	Donor Locus
ITP1-C	ITP1	70 copies	Chr 14
ITP1-I	ITP1	8 copies	n.d.
ITP2-M	ITP2	35 copies	Chr 14
ITP2-N	ITP2	10 copies	Chr 14
ITP2-P	ITP2	2 copies	Chr X

TABLE 6-1. Characteristics of transgenic ITP mouse lines.

Listed are all transgenic mouse lines generated for *ITP1* and *ITP2* transposons. Two *ITP1* and three *ITP2* mouse lines differing in transposon copy number (ranging from 2 *[ITP2-P]* to 70 copies *[ITP1-C]*) and donor locus were established. The donor locus was not determined for the *ITP1-I* mouse line. n.d., not determined.

***PB* and *SB* transposase knockin mice.** *Rosa26*$^{PB/+}$ and *Rosa26*$^{SB/+}$ mice have been generated by Pentao Liu (Wellcome Trust Sanger Institute, Hinxton/Cambridge, UK) as published in Rad *et al.*, 2010.

Bloom-mutated mice. Bloom-mutated *Blm*$^{m3/m3}$ mice, which show increased LOH rates in their cells, were generated and characterised by Allan Bradley's group (then Baylor College of Medicine, Houston, Texas, USA) (Luo *et al.*, 2000). For more information on *Blm*$^{m3/m3}$ mice, see Chapter 3.6.

6.2.1.2.2 Generation of mouse cohorts

All mice analysed in the recessive ITP screen study have two copies of the hypomorphic *Blm*$^{m3/m3}$ allele. To generate experimental (triple-transgenic) and control (double-transgenic) animal cohorts, *Rosa26*$^{PB/+}$*;Blm*$^{m3/m3}$ and *Rosa26*$^{SB/+}$*;Blm*$^{m3/m3}$ mice were crossed with either *Tg(ITP1)C;Blm*$^{m3/m3}$ mice or *Tg(ITP2)M;Blm*$^{m3/m3}$ mice. All triple-transgenic experimental mice harbour one copy of each, transposase and transposon, while double-transgenic control mice lack either transposase activity or the transposon concatemer. In the following, triple-transgenic *Tg(ITP2)M;Rosa*$^{PB/+}$*;Blm*$^{m3/m3}$ animals are abbreviated as IPB mice, triple-transgenic *Tg(ITP2)M;Rosa*$^{SB/+}$*;Blm*$^{m3/m3}$ animals as ISB mice and triple-transgenic *Tg(ITP1)C;Rosa*$^{PB/+}$*;Blm*$^{m3/m3}$ animals as IcPB mice. For nomenclature of ITP mouse cohorts, see also Table 7-1. Mouse cohorts were generated by Roland Rad. Genotyping of mouse lines was performed with primers listed in Table 5-12 according to Friedrich *et al.*, 2017.

6.2.1.2.3 Tumour watch and sample collection

All tumour-watch animals were monitored regularly for signs of sickness (e.g. inactivity, palpable/visible masses and poor grooming). Mice were sacrificed and necropsied when displaying symptoms of severe distress/illness. At necropsy, a gross inspection of all internal organs (including brain and lymph nodes) was carried out. In addition to tumour/abnormal tissue samples, a broad range of organs (usually brain, femur, gastrointestinal tract, heart, kidneys, liver, lungs, lymph nodes, pancreas, reproductive organs, spleen and thymus) – even without macroscopically visible signs of tumourigenesis – was routinely prepared for paraffin embedding by overnight fixation in 4% formalin solution. Organs/samples were paraffin-embedded and H&E stained at the Addenbrooke's Hospital Tissue Bank, Cambridge, UK. Moreover, tumour tissue and ear/tail samples (the latter for eventual re-genotyping) were stored in RNAlater for later DNA/RNA isolation. Mouse necropsies were performed by Roland Rad.

All H&E sections were analysed/diagnosed by Gary Hoffmann (School of Pathology and Laboratory Medicine, University of Western Australia, Australia), an experienced pathologist. Moreover, for all haematopoietic tumour samples, immunohistochemistry (B220/CD45R, CD3, CD138 and myeloperoxidase) (Table 6-2; for detailed information, see Table 5-6) was conducted at the Addenbrooke's Hospital Tissue Bank, Cambridge, UK. Mouse expert haematologic pathologists Leticia Quintanilla-Martinez de Fend and Ursula Kohlhofer (University of Tübingen, Tübingen, Germany) carried out the histopathological analysis of the lymphoma samples.

Primary Antibody	Host	Pre-Treatment	Dilution
B220/CD45R	rat	EDTA; 37°C; 20 min	1:40
CD138	rat	EDTA; 37°C; 20 min	1:50
CD3	rabbit	EDTA; 37°C; 20 min	1:100
Myeloperoxidase	rat	EDTA; 37°C; 20 min	1:100

TABLE 6-2. Antibodies for immunohistochemistry.

6.2.1.3 Sequencing of transposon insertion sites

A detailed protocol for sequencing of transposon insertion sites (quantitative transposon insertion site sequencing; QiSeq) was published in Nature Protocols in 2017 (Friedrich *et al.*, 2017). A short description of the main experimental steps is given below and Figure 6-2 displays a schematic overview. DNA was isolated from RNAlater-stored tissue (2-3 mm tissue slice/sample) using the Qiagen DNeasy Blood & Tissue Kit according to manufacturer's instructions. For tissue samples only available in form of FFPE tissue, a modified DNA isolation protocol was developed (Chapter 6.2.1.3.1).

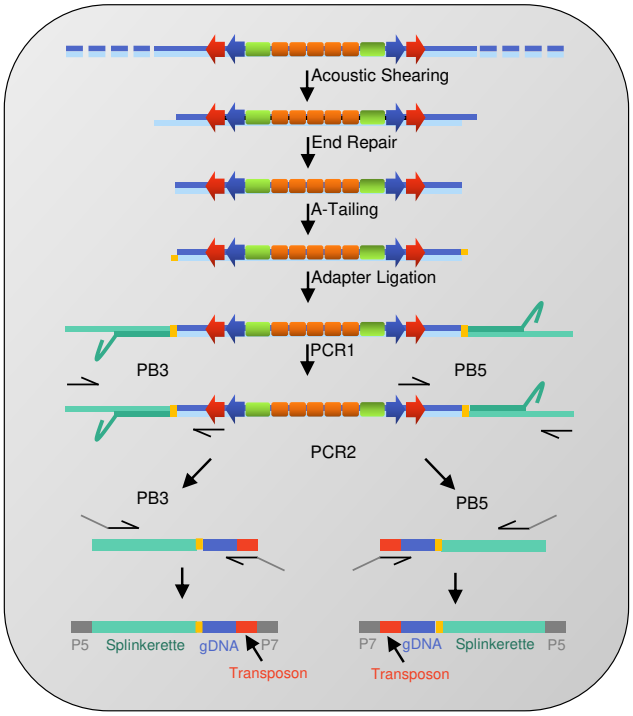

FIGURE 6-2. QiSeq library preparation overview.

For sequencing of transposon insertion sites, the QiSeq (quantitative transposon insertion site sequencing) protocol was carried out. Genomic DNA is sheared, end-repaired, A-tailed and ligated to a splinkerette adapter. Two PCRs (PCR1 and PCR2) with transposon- and splinkerette-specific primers that specifically amplify transposon-containing DNA fragments are performed. The 5' and 3' transposon end is amplified separately. PCR2 adds the Illumina flow cell binding sites (P5 and P7) to the fragments. Figure adapted from Friedrich *et al.*, 2017. PCR, polymerase chain reaction; PB, *PiggyBac*; gDNA, genomic DNA.

6.2.1.3.1 Isolation of DNA from formalin-fixed, paraffin-embedded tissue

For isolation of DNA from FFPE tissue, a new protocol was developed. For each sample, three 10 µm sections were cut and deparaffinised in xylene for 2 x 5 min. Xylene was removed by submerging the slides in absolute ethanol for 2 min and the sections were air-dried for 5 min. 20-30 µL ATL tissue lysis buffer (from the Qiagen DNeasy Blood & Tissue Kit) was then pipetted on top of the tissue area of interest and the tissue was carefully scraped off using a pipette tip. The tissue/lysis buffer mix was transferred into a 1.5 mL safe-lock tube. Per sample, a total of 180 µL ATL tissue lysis buffer was used. After addition of 20 µL proteinase K (20 mg/mL; as supplied in the DNeasy Blood & Tissue Kit), the samples were incubated in a thermomixer (56°C; 650 rpm) overnight. The next day, 10-20 µL proteinase K was added and the samples were incubated for another 24 h. This step was repeated on day three. After an approximately 60-h-incubation, DNA isolation was continued according to manufacturer's instructions of the DNeasy Blood & Tissue Kit with following modifications: (i) instead of the DNeasy mini spin columns provided in the kit, MinElute spin columns (which allow a smaller elution volume) were used and (ii) the elution step was performed with 50 µL pre-heated (56°C) AE buffer, which was incubated on the column for 10 min.

6.2.1.3.2 Library preparation

QiSeq is a method to identify and quantify transposon insertion sites in (tumour) tissues/cells. Key element is acoustic DNA shearing (fragmentation), leading to a random fragment cluster (with a mean fragment size of 250 bp) for each genomic position. With this approach, a major issue of enzymatic-based fragmentation can be circumvented: using frequently cutting restriction enzymes, the distances from the transposon insertion site to the next restriction sites are position-specific, meaning that one specific transposon insertion site is always represented by a fragment of a distinct size. Since the subsequent library preparation steps have a bias for certain fragment lengths (short fragments are "easier" amplified; long fragments might exceed the NGS size limit), an accurate quantification of the transposon insertion sites is not feasible. With QiSeq, such an insertion site-specific bias is much less likely to occur because a pool of differently sized fragments represents each transposon insertion site (and this size distribution is similar at every transposon insertion/genomic position).
After shearing, the fragmented DNA is end-repaired (DNA ends are blunted), A-tailed (an adenosine is added to 3' DNA ends) and a splinkerette adapter is ligated to each DNA end. For the 5' and 3' end, subsequent steps (amplification and sequencing of transposon-genome junctions) are separately performed. The specific structure of the splinkerette adapter (Y-shaped design with a template and a hairpin strand) ensures that only transposon-genome junction fragments (and not genomic fragments without transposon insert) can be amplified in the following first PCR step (which is conducted with transposon- and splinkerette-specific primers). Afterwards, a second nested PCR step is performed for further amplification, barcoding of samples and extension with Illumina flow cell binding sites (P5 and P7). Each sample is then quantified with quantitative real-time PCR (using P5- and P7-specific primers). Subsequently, samples are equimolarly mixed and the library pool is again quantified.

6.2.1.3.3 Sequencing

Sequencing was conducted on the Illumina MiSeq desktop sequencer (MiSeq Reagent Kit v2 (300 cycle)) in the DNA Sequencing Facility of the Wellcome Trust Sanger Institute, Hinxton/Cambridge, UK. A modified Illumina sequencing recipe including 12 initial chemistry-only/dark cycles was used. Since sequencing starts from the transposon end, the first 12 bases of each cluster are identical (sequence of the transposon ITR). Nucleotide diversity is especially important during these first cycles; therefore, the sequencing success would be impaired. To overcome this problem, read-1 consists of 12 dark cycles (transposon-specific

sequence; no imaging) and 75 regular imaging cycles (genomic sequence). Subsequently, the first 12 skipped bases are sequenced regularly (12 cycles, transposon-read) followed by sequencing of the barcode (8 cycles, index-read). After template turnaround, sequencing from the splinkerette adapter end is performed (75 cycles, read-2).

6.2.1.3.4 Downstream analysis and methods for determination of common insertion sites

For downstream analysis, the Wellcome Trust Sanger Institute Bioinformatics Pipeline was used (with scripts generated by Hannes Ponstingl) (for detailed information, see also Friedrich *et al.*, 2017). Roughly summarised, read-1 and transposon-read are re-assembled and then the transposon sequence is trimmed only leaving the *PB* (TTAA) and *SB* (TA) insertion sequence. Reads are mapped against the *Mus musculus* C57BL/6J GRCm38 assembly (*mm10* database) and several filtering steps based on read quality are performed. Insertion coordinates are determined and reads for each insertion site from the 5' and 3' end are correlated. In the end, for each sample a file containing all relevant information about the transposon insertion sites is generated.

For CIS analysis, which was performed by Maxim Barenboim (Klinikum rechts der Isar, Technical University Munich, München, Germany), TAPDANCE (Transposon Annotation Poisson Distribution Association Network Connectivity Environment) (Sarver *et al.*, 2012) and CIMPL (Common Insertion Site Mapping Platform; based on a Gaussian kernel convolution (using 30 k as scale parameter)) (de Ridder *et al.*, 2006) were carried out.

Pathway enrichment analysis was conducted using Ingenuity® Pathway Analysis (IPA®; Qiagen).

6.2.1.4 Loss of heterozygosity analysis

To determine LOH occurrence in tumours from IPB and ISB mice, single nucleotide polymorphism (SNP) analysis was applied. For this, DNA from a cell line derived from a small intestine tumour of an ISB mouse (IMSB_1.5C) and corresponding tail tissue was extracted (Qiagen DNeasy Blood & Tissue Kit). For amplification of ~650 bp regions flanking SNPs *rs4137461* and *rs220642642* within the *Apc* gene, 10 ng DNA was amplified with the Q5 High-Fidelity DNA Polymerase using following conditions (Table 6-3 and Table 6-4; primers listed in Table 5-13).

Component	Volume (µL)
5x Q5 Reaction Buffer	10
dNTPs (10 mM each)	1
10 µM Forward Primer	2.5
10 µM Reverse Primer	2.5
Q5 High-Fidelity DNA Polymerase	0.5
10 ng Template DNA	variable
H_2O	to 33.5

TABLE 6-3. PCR set-up for amplification of genomic regions within the *Apc* gene flanking SNPs *rs4137461* and *rs220642642*.

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	98	30	1
Denaturation	98	10	35
Annealing	67	20	
Extension	72	20	
Final Extension	72	120	1
Hold	10	pause	1

TABLE 6-4. Thermocycler programme for amplification of genomic regions within the *Apc* gene flanking SNPs *rs4137461* and *rs220642642*.

Sanger capillary sequencing (Eurofins Genomics) was performed using the respective forward primer as sequencing primer.

6.2.2 CRISPR/Cas-based recessive genetic screening

Part of this chapter can be also found in the original publication (Weber *et al.*, 2015), which appeared in *Proceedings of the National Academy of Sciences* in 2015.

6.2.2.1 Selection of single guide RNAs for tumour suppressor gene targeting

For selection of sgRNAs, consensus coding sequences (CCDS) were downloaded from the Ensembl database (http://www.ensembl.org; EMBL-EBI and Wellcome Trust Sanger Institute, Hinxton/Cambridge, UK) based on the *Mus musculus* C57BL/6J GRCm38 assembly (*mm10* database). Using the CRISPR design tool (http://crispr.mit.edu; Massachusetts Institute of Technology, Cambridge, MA, USA), for each gene, a sgRNA targeting exonic sequence near the transcription start site with minimal predicted off-targets was selected (Table 6-5).

Gene	Ensembl Identifier	CCDS Id.	sgRNA Target Sequence (PAM)	T.E.
Apc	ENSMUST00000079362	CCDS29125	TCAGTTGTTAAAGCAAGTTG (AGG)	2
Arid1a	ENSMUST00000105897	CCDS38908	TTAGTCCCACCATACGGCTG (AGG)	2
Brca1	ENSMUST00000017290	CCDS25474	AAATCTTAGAGTGTCCGATC (TGG)	2
Brca2	ENSMUST00000044620	CCDS39411	TAGGACCGATAAGCCTCAAT (TGG)	3
Cdkn2a-e1β	ENSMUST00000107131	CCDS18350	TGGTGAAGTTCGTGCGATCC (CGG)	1
Cdkn2a-e2	ENSMUST00000060501	CCDS38812	GTGCGATATTTGCGTTCCGC (TGG)	2
	ENSMUST00000107131	CCDS18350		
Pten	ENSMUST00000013807	CCDS29753	GCTAACGATCTCTTTGATGA (TGG)	1
Smad4	ENSMUST00000025393	CCDS29337	ACAACCCGCTCATAGTGATA (TGG)	2
Tet2	ENSMUST00000098603	CCDS51071	GAAAGTGCCAACAGATATCC (AGG)	3
Trp53	ENSMUST00000171247	CCDS48826	GACACTCGGAGGGCTTCACT (TGG)	4
Arid1b	ENSMUST00000115797	CCDS49929	CTGTGCACCTGGGGGACCGT (AGG)	2
Arid2	ENSMUST00000096250	CCDS37185	AGGCGCCTCCGGACGAGCGG (AGG)	1
Arid5b	ENSMUST00000020106	CCDS35929	GCTATGCAAATCGGATCCTT (TGG)	2
Atm	ENSMUST00000118282	CCDS40636	GGCTGTCAACTTCCGAAAAC (GGG)	7
Cdkn2b	ENSMUST00000097981	CCDS18351	GGCGCCTCCCGAAGCGGTTC (AGG)	1
Errfl1	ENSMUST00000030811	CCDS18974	AAGCTCGGGACAGCGTGAAG (AGG)	4
Igsf10	ENSMUST00000039419	CCDS50915	TGAGTCCGTAAAACGCCTCG (GGG)	4
Irf2	ENSMUST00000034041	CCDS22295	GTGCCGAGCCGCATGCATCC (AGG)	3

TABLE 6-5. Sequences of sgRNAs.

For each gene, the selected transcript (Ensembl and CCDS identifier), sgRNA target sequence with protospacer adjacent motif (PAM) and the targeted exon is listed. CCDS, consensus coding sequence; Id., Identifier; T.E., targeted exon.

6.2.2.2 *CRISPR-SB* vector cloning

For cloning of *Sleeping Beauty* terminal repeats (*SB* TRs) into *pX330* (Addgene; 42230), *SB* TRs were amplified from *pTnori* (Yant *et al.*, 2000) with *AflIII* and *NotI* restriction site overhangs and cloned into *pX330* that was sequentially opened with *AflIII* and *NotI*.

6.2.2.3 Cloning of single guide RNAs into the *CRISPR-SB* vector

To clone sgRNAs into the *CRISPR-SB* vector, the single-stranded forward and reverse sgRNA oligonucleotides (sequences listed in Table 5-14) were annealed and treated with T4 Polynucleotide Kinase to generate double-stranded and phosphorylated sgRNA oligonucleotides (Table 6-6 and Table 6-7).

Component	**Volume (µL)**
100 µM Forward Oligonucleotide	1
100 µM Reverse Oligonucleotide	1
10x T4 DNA Ligase Reaction Buffer	1
T4 Polynucleotide Kinase	0.5
H_2O	6.5

TABLE 6-6. Reaction set-up for sgRNA oligonucleotide annealing.

Temperature (°C)	**Time (min)**
37	30
95	5
then ramp down to 25°C at 0.1°C/sec	
25	pause

TABLE 6-7. Thermocycler programme for sgRNA oligonucleotide annealing.

The annealed oligonucleotides were then diluted 1:50 in H_2O.1 µg *CRISPR-SB* plasmid was digested with *BbsI* at 37°C for 30 min (Table 6-8).

Component	Volume (µL)
1 µg *CRISPR-SB*	variable
10x NEBuffer 2	2
10,000 units/mL *BbsI*	1
H2O	to 20

TABLE 6-8. Reaction set-up for *BbsI* digest of *CRISPR-SB*.

Subsequently, 1 µL Shrimp Alkaline Phosphatase was added to de-phosphorylate the vector and the whole reaction was incubated for another 30 min. Afterwards, heat-inactivation was performed at 65°C for 20 min. The ligation reaction was set-up and ligation was performed at RT for 10 min (Table 6-9).

Component	Volume (µL)
50 ng *BbsI*-digested Vector	1
Annealed and diluted Oligonucleotide	1
2x Quick Ligation Reaction Buffer	10
Quick T4 DNA Ligase	0.5
H_2O	7.5

TABLE 6-9. Reaction set-up for ligation of sgRNA oligonucleotide and *CRISPR-SB* vector.

6.2.2.4 Surveyor assays

To determine the on-targeting efficiency of the selected sgRNAs *in vitro*, the mouse pancreatic cancer cell lines *PPT-53631* and *PPT-4072*, which were cultured in DMEM supplemented with 10% FBS and 1x Penicillin-Streptomycin (5,000 U/mL), were used. 80,000 cells per well were sown in a 24-well-plate and transfected 24 h later with 450 ng *CRISPR-SB* and 50 ng *pcDNA™6.2/EmGFP-Bsd/V5-DEST* (which contains a blasticidin resistance cassette) using Lipofectamine® 2000 according to manufacturer's instructions. 24 h post transfection, selection of transfected cells was carried out using 5 µg/mL blasticidin for 48 h. Afterwards, cells were lysed with DirectPCR lysis reagent (Cell) according to manufacturer's instructions. Target region amplification was performed with TaKaRa Ex Taq DNA Polymerase (primers listed in Table 5-15). PCR products were denatured and reannealed in NEBuffer 2 in a thermocycler. Surveyor nuclease reaction was performed according to manufacturer's instructions and indel frequency was calculated as stated in Ran *et al.*, 2013.

6.2.2.5 Animal experiments

For **hydrodynamic tail vein injection** (HTVI), eight weeks old wild type or *Albumin$^{Cre/+}$;LSL-Kras$^{G12D/+}$* mice were used (Postic *et al.*, 1999; Hingorani *et al.*, 2003). To deliver the plasmid DNA into the mouse liver, 10 µg/mL *hSB5* transposase (Yant *et al.*, 2007) and ten (eighteen) *CRISPR-SB* sgRNA vectors (10 µg/mL in total) were dissolved in 2 mL (female mice) or 2.5 mL (male mice) 0.9% saline, respectively and injected into the lateral tail vein over six to ten seconds. Control mice were injected with a Cas9-only expressing plasmid and *hSB5* transposase in the same manner as described above.

For some cohorts, liver tumourigenesis was chemically accelerated using ***CCl_4* treatment**. For this, mice were intra-peritoneally injected once a week with 1 µL/g body weight 10% *CCl_4* for a total of nine weeks beginning two weeks post HTVI.

All mice were regularly watched for signs of illness and tumour development.

Starting 20 weeks after HTVI, mice were repeatedly (in two- to four-week intervals) monitored using **magnetic resonance imaging (MRI)** in collaboration with the Department of Radiology, Klinikum rechts der Isar, Technical University Munich, München, Germany (Rickmer Braren and Irina Heid). MRI was conducted using a 3 Tesla clinical MRI system (Ingenia 3T, Philips Healthcare) with a human 8-channel wrist coil (SENSE Wrist coil 8 elements) following a previously described protocol (Braren *et al.*, 2011), which was adapted to the 3 Tesla scanner. To this end, longitudinal T2-weighted turbo spin-echo imaging (slice thickness: 0.7 mm, in-plane resolution: 0.3 x 0.38 mm^2, TR/TE: 4352 ms/101 ms, TF: 21, NSA: 9, total scan duration: 5.22 minutes) was performed for tumour detection and volumetric analysis.

In general, mice were sacrificed when displaying signs of illness or MRI-diagnosed tumours reached a size greater or equal 3 mm in diameter. For all mice, tumour and healthy liver tissue was stored in RNAlater, and for tumours with a size equal or larger 1 mm histological analysis (Chapter 6.1.5) was performed.

Subcutaneous implantation of cell lines from mouse primary liver tumours was conducted using *NOD.Cg-PrkdcscidIl2rgtm1Wjl/SzJ* (NOD Scid gamma) mice. For this, trypsinised tumour cells were washed twice with DBPS and 5 x 10^5 cells in 150 µL were injected into the right and left flank of the mice using a 1-mL-syringe with a 27-gauge-needle. Mice were monitored regularly for signs of sickness and tumour growth and were sacrificed once tumours reached a size of 1 cm in diameter.

6.2.2.6 Quantitative Cas9 analysis

To detect *hSpCas9* presence in liver samples of mice two weeks post HTVI, 7.5 ng genomic DNA was used for real-time quantitative PCR (SYBR® Select Master Mix). *HSpCas9* copy numbers were normalised to mouse *Apolipoprotein B* (*Apob*) copy numbers. Sequences for *hSpCas9_qPCR* and *mouse_Apob_qPCR* primers are listed in Table 5-16.

6.2.2.7 Quantitative guide distribution analysis

To analyse the distribution of sgRNAs in liver samples of mice two weeks after HTVI, 10 ng DNA per 20 µL sample was amplified with Taq Polymerase using *CRISPR-SB_PCR* primers listed in Table 5-16. PCR products were cleaned-up (QIAquick PCR Purification Kit) and 10 pg purified PCR product was used for sgRNA-specific real-time quantitative PCR (SYBR® Select Master Mix). For this, a universal forward primer (*CRISPR-SB-qPCR*; sequence listed in Table 5-16) and sgRNA-specific reverse primers (reverse sgRNA oligonucleotide; sequences listed in Table 5-14) were used.

6.2.2.8 Characterisation of liver tumours

Tissue processing and H&E staining was conducted according to standard protocols (Chapter 6.1.5). For hepatic tumours, immunohistochemistry was performed with markers listed in Table 6-10 by the group of Mathias Heikenwälder (Institute of Virology, Technical University Munich, München, Germany). For more information on primary and secondary antibodies, see also Table 5-6.

Primary Antibody	Host	Pre-Treatment	Dilution	Secondary Antibody	Dilution
A6	rat	Proteinase; 37°C; 10 min	1:100	rabbit-anti-rat	1:1000
AFP	goat	Citrate; 100°C; 30 min	1:100	rabbit-anti-goat	1:300
Collagen-4	rabbit	Proteinase; 37°C; 10 min	1:50	goat-anti-rabbit	1:500
Cytokeratin 19	rat	EDTA; 100°C; 20 min	1:500	rabbit-anti-rat	1:1000
GP73	goat	EDTA; 100°C; 30 min	1:100	rabbit-anti-goat	1:500
Ki67	rabbit	EDTA; 95°C; 30 min	1:200	goat-anti-rabbit	1:500

TABLE 6-10. Antibodies used for immunohistochemistry of liver tumours.

6.2.2.9 Microdissection and DNA isolation from microdissected tissue

For heterogeneity analysis, tumour sections were microdissected under a microscope using 20-gauche-needles. DNA was isolated from the microdissected regions as described in Chapter 6.2.1.3.1.

6.2.2.10 Sequencing of single guide RNA target regions

For sequencing of sgRNA target regions, 5 ng DNA were amplified with Q5 High-Fidelity DNA Polymerase using following conditions (Table 6-11, Table 6-12 and Table 6-13):

Component	Volume (µL)
5x Q5 Reaction Buffer	6
dNTPs (10 mM each)	0.6
10 µM Forward Primer	1.5
10 µM Reverse Primer	1.5
Q5 High-Fidelity DNA Polymerase	0.6
5 ng Template DNA	variable
H_2O	to 30

TABLE 6-11. PCR set-up for amplification of sgRNA target regions.

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	98	30	1
Denaturation	98	10	35
Annealing	variable	20	
Extension	72	15	
Final Extension	72	120	1
Hold	10	pause	1

TABLE 6-12. Thermocycler programme for amplification of sgRNA target regions.

sgRNA Target Region	Annealing Temperature (°C)
Apc	60
Arid1a	60
Arid5b	60
Atm	60
Brca1	60
Brca2	60
Cdkn2a-e1β	60
Cdkn2a-e2	60
Igsf10	60
Pten	60
Smad4	60
Tet2	60
Trp53	60
Arid1b	62
Cdkn2b	62
Errfi1	62
Irf2	62
Arid2	65

TABLE 6-13. Annealing temperatures for amplification of sgRNA target regions.

For amplification of the GC-rich *Arid2* target locus, the reaction set-up was slightly modified and a 5x GC Enhancer included (Table 6-14).

Component	Volume (µL)
5x Q5 Reaction Buffer	6
5x GC Enhancer	6
dNTPs (10 mM each)	0.6
10 µM Forward Primer	1.5
10 µM Reverse Primer	1.5
Q5 High-Fidelity DNA Polymerase	0.6
5 ng Template DNA	variable
H_2O	to 30

TABLE 6-14. PCR set-up for amplification of *Arid2* target region.

For NGS, all PCR products from one tumour were pooled and purified using the QIAquick PCR purification kit. Library preparation was performed as described in Quail *et al.*, 2009. Briefly, concentration of pooled PCR products was measured using the Qubit® fluorometer. End-repair, A-tailing and adapter ligation were performed using the NEBNext® Ultra DNA Library Prep Kit for Illumina®. Adapter-ligated PCR products were cleaned up (AMPure XP Beads)

and PCR enrichment (12 cycles) was conducted using the KAPA HiFi HotStart ReadyMix. After bead clean-up, all samples were quantified (Qubit® fluorometer) and equimolarly pooled. Quantification of the library was carried out using qPCR with primers specific for the Illumina P5 and P7 flow cell binding sites. The libraries were sequenced on the Illumina MiSeq desktop sequencer using the MiSeq Reagent Kit v3 (600 cycle) with 20% PhiX spiking.

6.2.2.11 Sequencing of single guide RNA off-target regions

Coordinates of potential off-target sites for the sgRNAs were downloaded from the CRISPR design tool (http://crispr.mit.edu). For the top five off-targets (exonic, intronic and intergenic) and (if not already included in the top five list) top three exonic off-targets of each sgRNA, flanking PCR primers (sequences listed in Table 5-17) were designed. PCRs and amplicon-based next generation sequencing were performed in the same manner as described above for the sgRNA on-target regions (Chapter 6.2.2.10).

6.2.2.12 Bioinformatics analyses

Maxim Barenboim (Klinikum rechts der Isar, Technical University Munich, München, Germany) performed bioinformatics analyses. MiSeq Illumina paired 300 nucleotide reads were mapped to *mm10* with *BBMAP short read aligner* (http://bbmap.sourceforge.net) using default settings. BAM files were sorted and indexed with *samtools (v0.1.19)* (Li *et al.*, 2009). After mapping, only paired reads (about 3% were unpaired) were extracted based on *bitwise flag 0x2*. This resulted in BAM files containing only correctly paired reads. In order to obtain data in pileup format with the number of reads covering sites *samtools (v0.1.6) pileup command with option (-i)*, which only displays lines containing indels, was employed. Pileup files were processed with *VarScan (v2.3.6) pileup2indel command* (Koboldt *et al.*, 2009).

6.2.2.13 Fusion analysis for detection of large chromosomal deletions

To test for possible intra-chromosomal large deletions/fusion products caused by combinatorial sgRNA targeting, PCRs spanning the potential fusion location, as predicted by the sgRNA target sites, were performed. For this, 10 ng genomic DNA in a 30 µL reaction were amplified using TaKaRa Ex Taq DNA Polymerase and the respective forward and reverse primers of the

target sites. Resulting PCR products were cleaned-up (QIAquick PCR Purification Kit) and Sanger capillary sequencing was performed. To quantify the *Cdkn2a* fusion product in the different regions of Tu1 (R1, R2 and R3), 10 ng DNA of the respective samples was used for quantitative real-time PCR (SYBR® Select Master Mix) with primers displayed in Table 5-16. The primer combination *Cdkn2a-e1β-qPCR-F/Cdkn2a-e2-qPCR-R* was used for quantification of the fusion product. Quantification of other alleles at that position (wild type and with small indels) was carried out with *Cdkn2a-e2-qPCR-F* and *Cdkn2a-e2-qPCR-R* primers.

6.2.2.14 Multicolour fluorescence *in situ* hybridisation

Multicolour fluorescence in situ hybridisation (M-FISH) was performed by Beiyuan Fu and Fengtang Yang (Wellcome Trust Sanger Institute, Hinxton/Cambridge, UK) as described in Jentsch *et al.*, 2001.

6.2.2.15 Array comparative genomic hybridisation

Array comparative genomic hybridisation (aCGH) was carried out by the group of Kristian Unger (Helmholtz Zentrum München, Neuherberg/München, Germany) and analysis was performed by Thomas Engleitner (Klinikum rechts der Isar, Technical University Munich, München, Germany). For this, Agilent 60k mouse CGH arrays with custom design (AMADID 041078) were used as stated in Wolf *et al.*, 2014. CGH data was pre-processed with the Agilent Genomic Workbench software. Raw log ratios were re-centred by adding or subtracting a constant value to insure that the zero point is reflecting the most common ploidy state. Segmentation and aberration calling were performed with the implemented ADM-2 algorithm. Normalised data was imported into *R version 3.1*. For each detected aberration, the closest off-targets surrounding the aberration borders up- and downstream were investigated. The distance and the number of probes between the aberration border and the predicted off-target site were calculated. An aberration was called potentially induced by a CRISPR/Cas off-target if 20 probes or less were located between the aberration and the off-target site and the distance between them was lower than 500,000 nucleotides.

7 RESULTS

7.1 Study of lymphomagenesis using transposon-based recessive genetic screening

Some results from this chapter can be also found in Weber *et al.*, 2019, which was published in *Nature Communications* in 2019.

To adapt the transposon-based insertional mutagenesis technology for recessive genetic screening in mice, ITP transposons that only harbour gene-trapping but no gene-activating elements were generated. In addition, the screen was conducted in a *Blm*-mutated background fostering LOH in tumour cells and thereby identification of classical TSGs.

For the generation of experimental and control cohorts, (i) constitutive *Rosa26*$^{PB/PB}$ and *Rosa26*$^{SB/SB}$ transposase knockin mice, (ii) transgenic transposon mice harbouring the *ITP1 (Tg(ITP1)C)* or *ITP2 (Tg(ITP2)M)* transposon concatemers and (iii) Bloom-mutated *Blm*$^{m3/m3}$ mice were crossed (Figure 7-1a). Resulting triple-transgenic mouse cohorts are further abbreviated as IPB *(Tg(ITP2)M;Rosa26*$^{PB/+}$*;Blm*$^{m3/m3}$*)*, ISB *(Tg(ITP2)M;Rosa26*$^{SB/+}$*;Blm*$^{m3/m3}$*)* and IcPB *(Tg(ITP1)C;Rosa26*$^{PB/+}$*;Blm*$^{m3/m3}$*)* mice.

For control cohorts (double-transgenic mice lacking either the transposon concatemer or the transposase activity), following abbreviations are used: PB *(Rosa26*$^{PB/+}$*;Blm*$^{m3/m3}$*)*, SB *(Rosa26*$^{SB/+}$*;Blm*$^{m3/m3}$*)* and IB (*Tg(ITP2)M;Blm*$^{m3/m3}$*)*.

7.1.1 Genetic screening using ITP transposons leads to solid and haematopoietic tumour development

Since *in vivo* transposition might be associated with embryonic lethality in mice (particularly when using mice with constitutive transposase activity), numbers of viable triple-transgenic offspring were determined for IPB, ISB and IcPB breedings and compared to the expected numbers in accordance with Mendel's law. While ISB mice were born close to Mendelian frequency (91.7% of the expected), moderate and extensive embryonic lethality was observed for IPB (59.2% of the expected) and IcPB mice (13.8% of the expected), respectively (Figure 7-1b).

Experimental and control mice were aged and sacrificed when showing signs of tumour development (Chapter 6.2.1.2.3). For all 123 IPB, 81 ISB, 34 PB and 57 SB mice, survival data as well as detailed tumour histology data is available; for IB mice (30 animals) only survival data exists (Table 7-1).

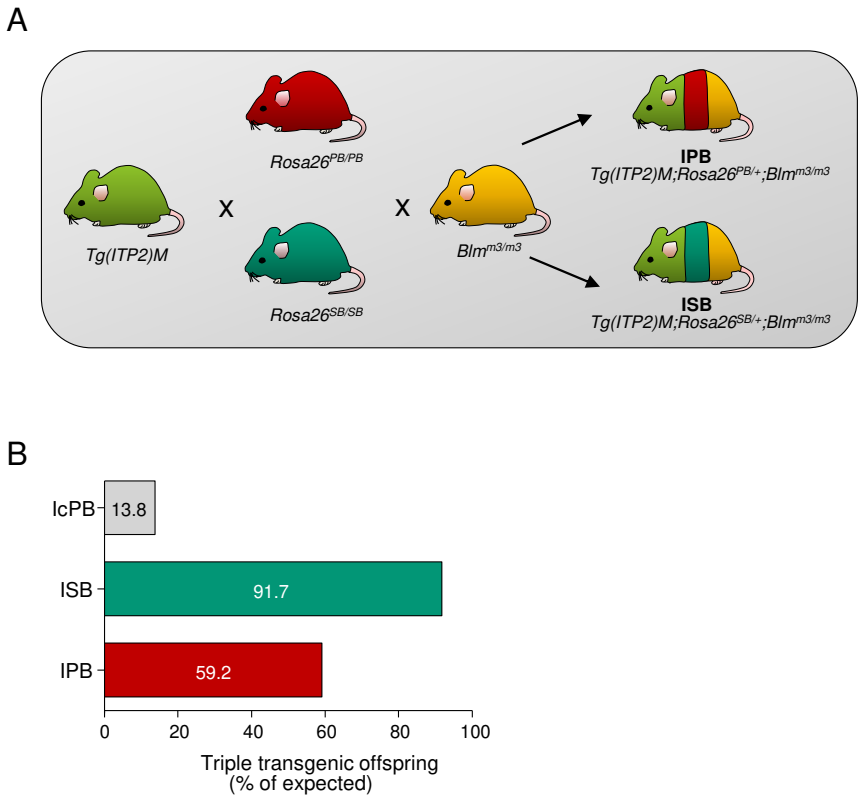

FIGURE 7-1. Generation of ITP experimental and control mouse cohorts.

(**A**) Crossing of *Rosa26*$^{PB/PB}$/*Rosa26*$^{SB/SB}$, *Tg(ITP2)M* and *Blm*$^{m3/m3}$ mice was performed to establish triple-transgenic IPB *(Tg(ITP2)M;Rosa26*$^{PB/+}$*;Blm*$^{m3/m3}$*)* and ISB *(Tg(ITP2)M;Rosa26*$^{SB/+}$*;Blm*$^{m3/m3}$*)* mice. Control cohorts include double-transgenic mice without transposon concatemer (PB: *Rosa26*$^{PB/+}$*;Blm*$^{m3/m3}$ and SB: *Rosa26*$^{SB/+}$*;Blm*$^{m3/m3}$) or transposase activity (IB: *Tg(ITP2)M;Blm*$^{m3/m3}$). (**B**) Embryonic lethality was analysed for IPB, ISB and IcPB mice by determining the numbers of triple-transgenic offspring for all three cohorts and then comparing with the expected numbers in accordance with Mendel's Law.

Line	Abbreviation	Cohort Type	Cohort Size	Tumour Histology Available
Tg(ITP2)M;Rosa26$^{PB/+}$*;Blm*$^{m3/m3}$	IPB	experimental	123	yes
Tg(ITP2)M;Rosa26$^{SB/+}$*;Blm*$^{m3/m3}$	ISB	experimental	81	yes
Tg(ITP1)C;Rosa26$^{PB/+}$*;Blm*$^{m3/m3}$	IcPB	experimental	7	no
Rosa26$^{PB/+}$*;Blm*$^{m3/m3}$	PB	control	34	yes
Rosa26$^{SB/+}$*;Blm*$^{m3/m3}$	SB	control	57	yes
Tg(ITP2)M;Blm$^{m3/m3}$	IB	control	30	no

TABLE 7-1. ITP mouse cohorts.

All ITP mouse cohorts (experimental and control) with respective abbreviation and number of mice are listed. For each cohort, availability of tumour histology data is stated.

Survival analysis showed that IPB mice lived significantly shorter than PB ($p<0.001$ (all p-values are corrected for multiple testing with the Benjamini-Hochberg correction); Hazard Ratio (HR) = 2.8; 95% confidence interval (CI) = 2.1 – 3.8) and IB mice ($p<0.001$; HR = 2.0; 95% CI = 0.4 – 1.7) while PB and IB mice exhibited no significant differences in survival time (Figure 7-2a). Similarly, ISB mice had a significantly shorter survival time than SB ($p<0.01$; HR = 1.9; 95% CI = 1.3 – 2.8) and – in tendency – IB mice (p-value not significant) whereas SB and IB mice showed no significant survival time variations (Figure 7-2b).

Since survival analysis resulted in no significant differences between all three control mouse lines, PB, SB and IB cohorts were merged and referred to as “control mice” in the following. When comparing survival times of both triple-transgenic mouse lines (IPB and ISB; which only differ in the transposase used for transposon mobilisation), IPB mice died significantly earlier than ISB mice ($p<0.001$; HR = 1.7; 95% CI = 1.3 – 2.3) while both IPB and ISB had a significantly shorter survival time compared to control mice (IPB vs controls: $p<0.001$; HR = 2.0; 95% CI = 1.4 – 2.7 and ISB vs controls: $p<0.001$; HR = 3.0; 95% CI = 2.2 – 4.2) (Figure 7-2c).

As partly reflected in the survival time differences between IPB, ISB and control mice, the tumour spectrum is also quite diverse (Figure 7-2d). Mice of all three cohorts developed solid and haematopoietic tumours (lymphomas), in some cases both in the same animal. However, the distribution of solid and haematopoietic cancers differed tremendously between the cohorts. While 66% of all IPB mice (with tumours) solely showed lymphomas, this was only the case for 25% of all ISB tumour mice. In contrast, solid tumour development was more prominent in the ISB tumour mouse cohort (solid tumour only: 33%; solid and haematopoietic tumour: 42%) than in the IPB tumour mouse cohort (solid tumour only: 18%; solid and haematopoietic tumour: 16%). Control mice (with tumours) showed a similar tumour spectrum as IPB mice (haematopoietic tumour only: 59%; solid tumour only: 17%; solid and haematopoietic tumour: 23%).

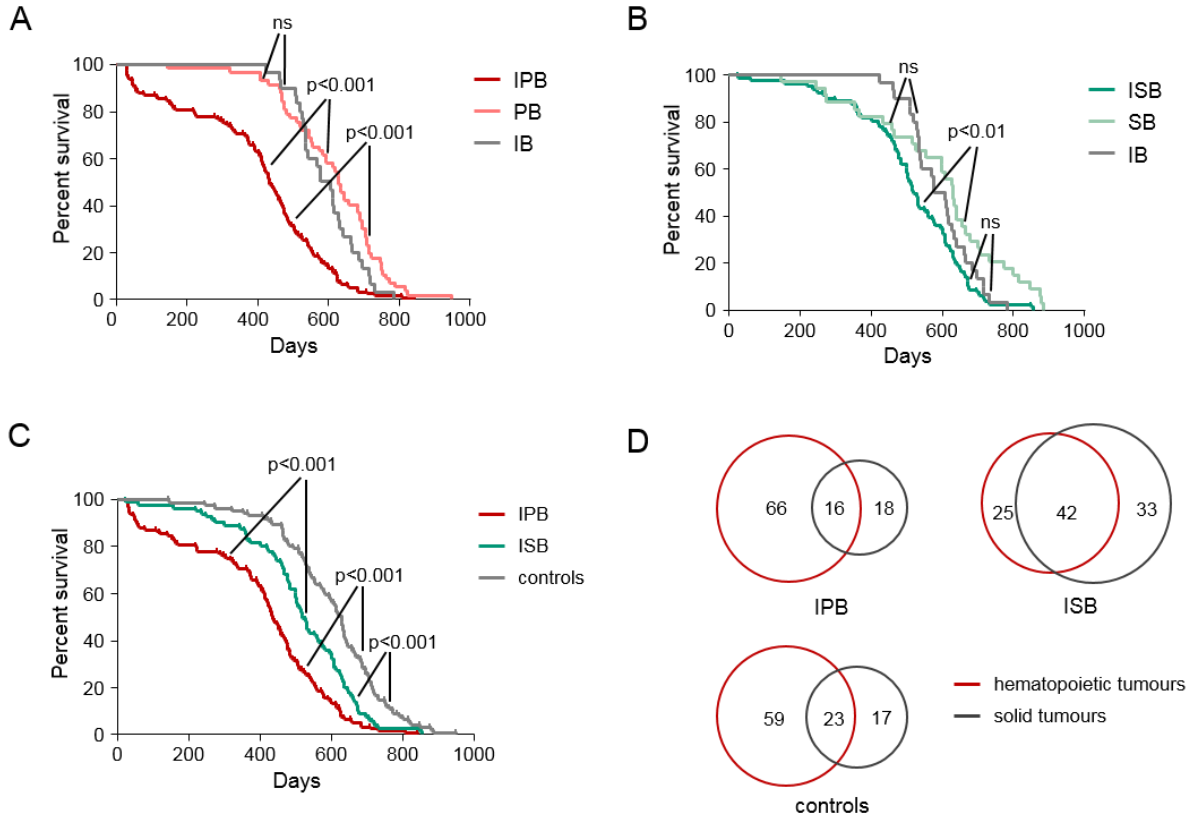

FIGURE 7-2. Survival of and tumour spectrum in ITP mice.

(**A**) Kaplan-Meier survival curves for IPB *(Tg(ITP2)M;Rosa26$^{PB/+}$;Blm$^{m3/m3}$)*, PB *(Rosa26$^{PB/+}$;Blm$^{m3/m3}$)* and IB *(Tg(ITP2)M;Blm$^{m3/m3}$)* mice. (**B**) Kaplan-Meier survival curves for ISB *(Tg(ITP2)M;Rosa26$^{SB/+}$;Blm$^{m3/m3}$)*, SB *(Rosa26$^{SB/+}$;Blm$^{m3/m3}$)* and IB mice. (**C**) Kaplan-Meier survival curves for IPB, ISB and control (*Rosa26$^{PB/+}$;Blm$^{m3/m3}$*, *Rosa26$^{SB/+}$;Blm$^{m3/m3}$* and *Tg(ITP2)M;Blm$^{m3/m3}$*) mice. (**D**) Venn diagrams display tumour spectra of IPB, ISB and control mice. Percentage of tumour-bearing mice with haematopoietic tumours, solid tumours or both is indicated. Red circles represent haematopoietic tumours, grey circles solid tumours.

7.1.2 IPB and ISB mice show a broad variety of solid tumours

Of all IPB, ISB and control tumour mice, 34%, 75% and 40%, respectively, developed at least one solid tumour. Solid tumours included intestinal tumours (mainly dysplastic polyps and minor adenomas located in the small intestine), papillary adenomas and adenocarcinomas of the lung, liver tumours (mainly HCCs but also Kupffer cell sarcomas), tumours of the reproductive organs (predominantly invasive adenocarcinomas of the uterus, but also tumours located in the ovaries and prostate), squamous cell carcinomas and stomach tumours (mostly adenocarcinomas) (Figure 7-3a). In addition, more than 40% of ISB and 20% of IPB mice with solid tumours developed pituitary adenomas (adrenocorticotropic hormone-producing tumours

of the pituitary intermediate lobe). Another intracranial tumour entity, namely astrocytoma, was only (rarely) identified in IPB mice (6% of all IPB tumour mice). Control mice showed a rather similar tumour spectrum as IPB and ISB mice; however, no pituitary adenomas, astrocytomas and stomach tumours were observed.

While only about 20% of all IPB solid tumour mice had two solid cancers, 37% of all ISB mice with solid tumours had two or more cancers (10% even had three individual solid tumours) (Figure 7-3b).

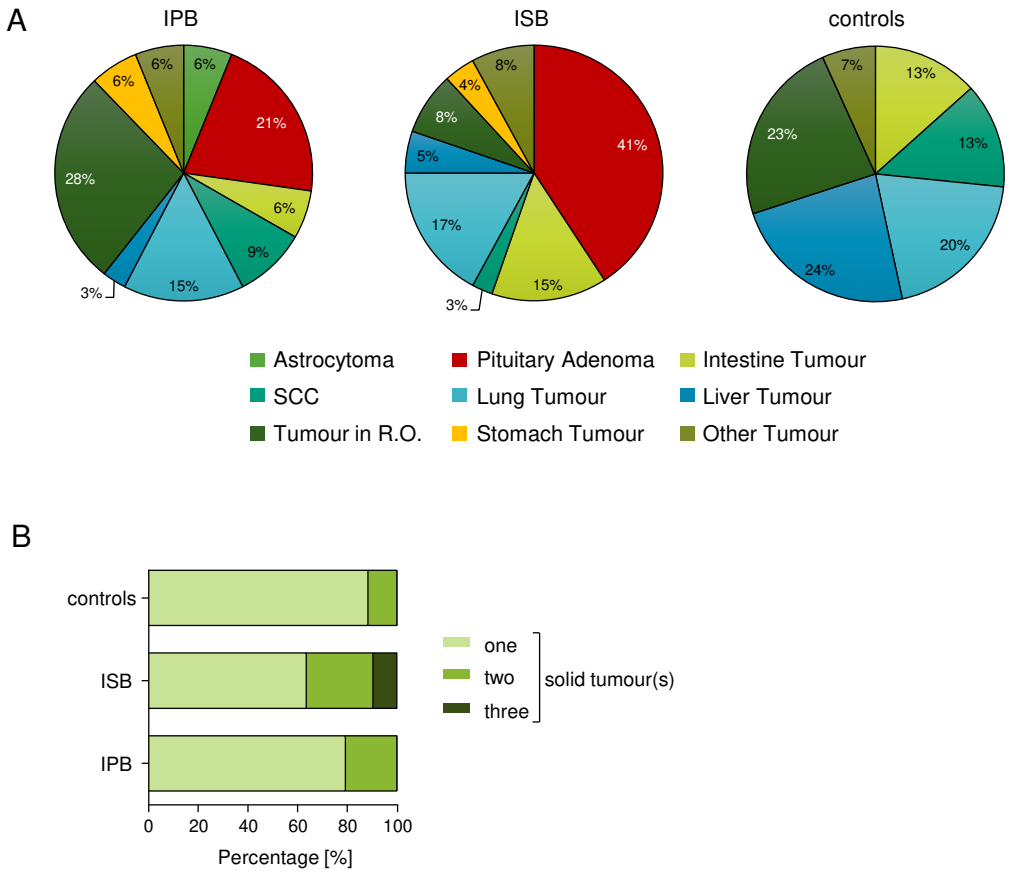

FIGURE 7-3. Solid tumour spectrum in IPB, ISB and control mice.

(**A**) Pie charts show solid tumour spectrum in IPB *(Tg(ITP2)M;Rosa26$^{PB/+}$;Blm$^{m3/m3}$)*, ISB *(Tg(ITP2)M;Rosa26$^{SB/+}$;Blm$^{m3/m3}$)* as well as in control (*Rosa26$^{PB/+}$;Blm$^{m3/m3}$*, *Rosa26$^{SB/+}$;Blm$^{m3/m3}$* and *Tg(ITP2)M;Blm$^{m3/m3}$*) mice. (**B**) Bar graph shows number of solid cancers in the IPB, ISB and control solid tumour mouse cohorts. SCC, squamous cell carcinoma; R.O., reproductive organs.

7.1.3 IPB and ISB mice frequently develop B-cell lymphomas

All haematopoietic tumours were classified as lymphoma, defining it the most prominent tumour entity of the ITP screen: 82%, 67% and 82% of all IPB, ISB and control tumour mice developed lymphomas. IPB and ISB lymphoma mice had a significantly shorter survival time compared to control mice with lymphoma (IPB vs controls: $p<0.001$; HR = 3.2; 95% CI = 2.1 – 4.7 and ISB vs controls: $p<0.05$; HR = 1.9; 95% CI = 1.2 – 3.1). IPB lymphoma mice had a tendency to die earlier than ISB lymphoma mice (p-value not significant; mean survival time IPB: 476 days; mean survival time ISB: 529 days) (Figure 7-4a).

To further characterise the lymphomas, immunohistochemical analyses were performed and expression of B220/CD45R (B cell marker), CD3 (T cell marker), myeloperoxidase (myeloid marker) and CD138 (plasma cell marker) was analysed. While lymphomas showed either strong B220 or CD3 expression depending on their origin in the B-cell or T-cell lineage, respectively, no haematopoietic tumours from the myeloid lineage were present in this study. Likewise, no CD138-positive plasmacytomas/multiple myelomas were detected, though some DLBCLs exhibited CD138 positivity and were classified as DLBCLs with plasmacytic differentiation.

For all cohorts, the most prevalent lymphoma subtype was DLBCL (IPB: 39/46 diagnosed lymphoma cases (84.8%); ISB: 13/26 diagnosed lymphoma cases (50.0%); controls: 25/26 diagnosed lymphoma cases (96.2%)). In 12.8%, 7.7% and 16.0% of all IPB, ISB and control DLBCL cases, mice had an additional tumour within the same haematopoietic organ, which was in most cases a histiocytic sarcoma (7/10 cases; 70%). Other identified lymphoma entities included T-cell lymphoblastic lymphoma (IPB: 6/46 diagnosed lymphoma cases (13.0%); ISB and controls: in each cohort, 1/26 diagnosed lymphoma cases (3.8%)) and different B-cell lymphoma types, which were mainly identified in the ISB cohort: Marginal zone lymphoma of the spleen (IPB: 1/46 diagnosed cases (2.2%), ISB: 2/26 diagnosed cases (7.7%)), follicular lymphoma (ISB: 1/26 diagnosed cases (3.8%) and B-cell lymphoblastic lymphoma (ISB: 2/26 diagnosed cases (7.7%)) (Figure 7-4b).

Figure 7-4c exemplarily shows histology and surface marker expression of several B-cell lymphoma subtypes.

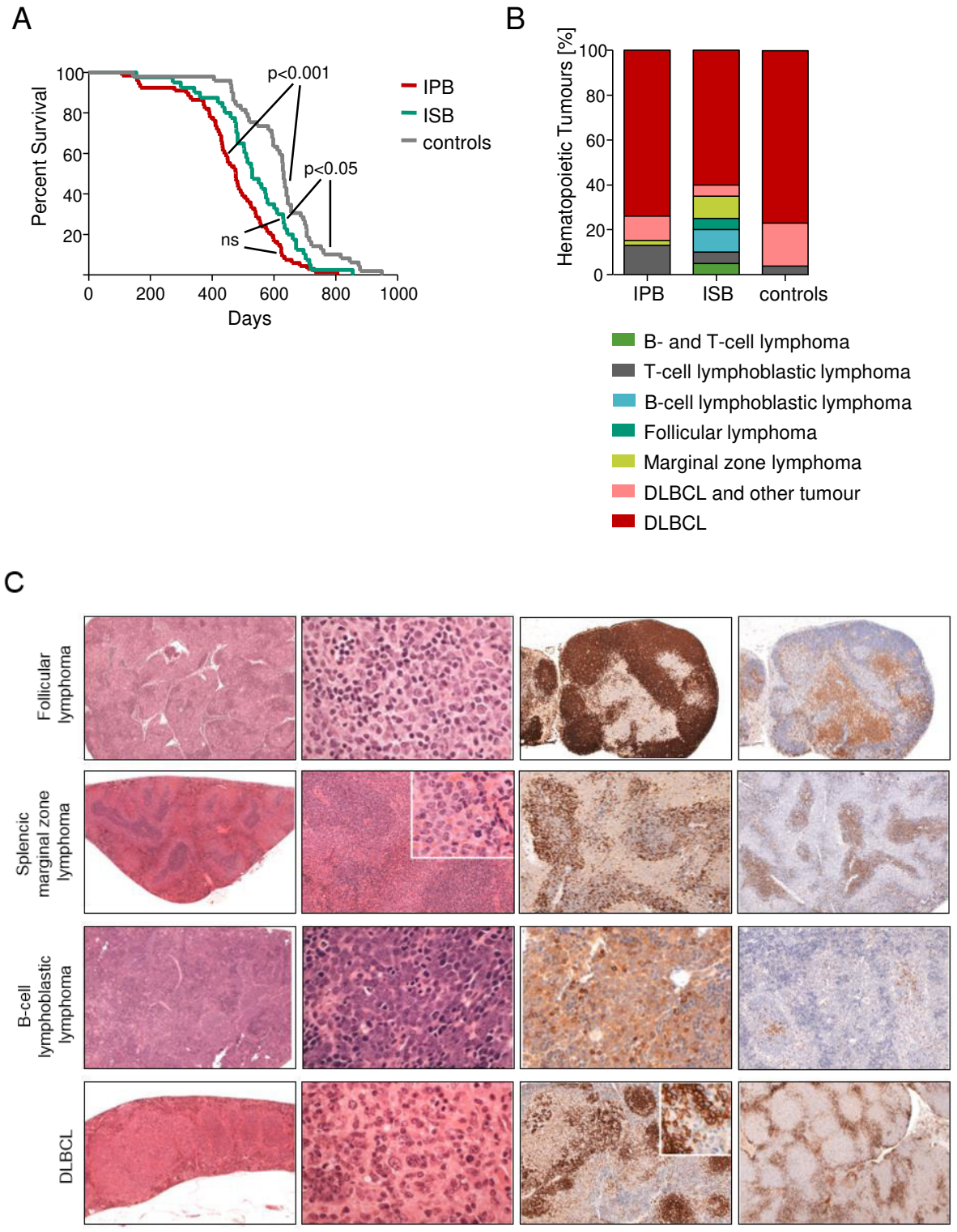

FIGURE 7-4. Lymphomas in IPB, ISB and control mice.

(**A**) Kaplan-Meier survival curves for IPB mice *(Tg(ITP2)M;Rosa26$^{PB/+}$;Blm$^{m3/m3}$)*, ISB mice *(Tg(ITP2)M;Rosa26$^{SB/+}$;Blm$^{m3/m3}$)* and control mice (*Rosa26$^{PB/+}$;Blm$^{m3/m3}$*, *Rosa26$^{SB/+}$;Blm$^{m3/m3}$* and *Tg(ITP2)M;Blm$^{m3/m3}$*) with lymphomas. (**B**) Bar graph shows spectrum of lymphomas in IPB, ISB and control mice. (**C**) Overview of B-cell lymphoma histology and surface marker expression. Shown are representative

FIGURE 7-4 (continued)

images of a follicular lymphoma (row 1; lymph node), splenic marginal zone lymphoma (row 2; spleen), B-cell lymphoblastic lymphoma (row 3; spleen) and diffuse large B-cell lymphoma (DLBCL; row 4; spleen). Columns 1 and 2 show haematoxylin and eosin stainings, column 3 B220 immunohistochemistry (IHC) and column 4 (with the exception of row 4 that displays CD138 IHC) CD3 IHC. All B-cell lymphomas strongly expressed B220 and were negative for CD3 (for DLBCL not shown). This DLBCL case also exhibited CD138 expression and was therefore classified as DLBCL with plasmacytoid differentiation. Magnifications: column 1: row 1: 50x, rows 2-4: 25x; column 2: rows 1, 3, 4: 630x, row 2: 100x, insert: 630x; column 3: rows 1 and 4: 50x, row 2: 100x, row 3: 400x; column 4: rows 1 and 2: 50x; row 3: 100x; row 4: 25x.

7.1.4 IPB and ISB mice show insertions in lymphoma relevant genes

All tumours (and tails from tumour-bearing mice) for which RNAlater stored tissue was available, were sequenced and transposon insertion sites were determined using the QiSeq pipeline. For DLBCLs derived from the IPB cohort (n = 34), CIMPL and TAPDANCE analyses were conducted. For both CIS analyses, the identical cut-off (sum of read coverage of 5' and 3' read ≥ 20) was used to exclude very low coverage insertions (which most likely play only minor roles in transposon-driven tumourigenesis). While CIMPL analysis led to a total of 261 CISs located within unique genes, only 94 intragenic CISs were reported by TAPDANCE. Of these 94 CISs, the vast majority, namely 86, were also detected using CIMPL. Four out of the eight remaining "TAPDANCE-unique" CISs were located on the donor chromosome, for which accurate CIS determination is in general more challenging due to the high number of transposon insertions. For the remaining four genes (*Dido1*, *Lmbrd1*, *Psme4* and *Selt*) no reason could be determined why they were not identified by CIMPL analysis.

Table 7-2 shows the Top 50 CISs (ranked by p-value library; CISs on the donor chromosome are not shown) and corresponding candidate genes from TAPDANCE analysis. For some genes (e.g. *Prprc* and *Pten*) more than one CIS was found, indicating that transposon insertions are distributed over larger regions of the gene. While many genes derived from the CIS list are already known for their involvement in lymphomagenesis, some genes have not been implicated with lymphoma/cancer so far. The most frequently hit genes were, among others, *Fas*, *Hnrnpa2b1, Pten* and *Ptprc,* all of which have known roles in tumourigenesis (Kojima *et al.*, 2006; Porcu *et al.*, 2012; Pfeifer *et al.*, 2013; Kataoka *et al.*, 2015). Moreover, *Gna13*, which is a recently discovered TSG in B-cell lymphoma (Muppidi *et al.*, 2014), was also commonly hit. Top-ranked genes not yet implicated in lymphomagenesis included for example *Naa1*5 and *Erdr1*. Figure 7-5 exemplarily shows ITP insertion patterns in *Pten*, *Ptprc*

and *Gna13*. Transposon insertions are distributed all over the genes, indicative of a TSG (inactivating) insertion pattern.

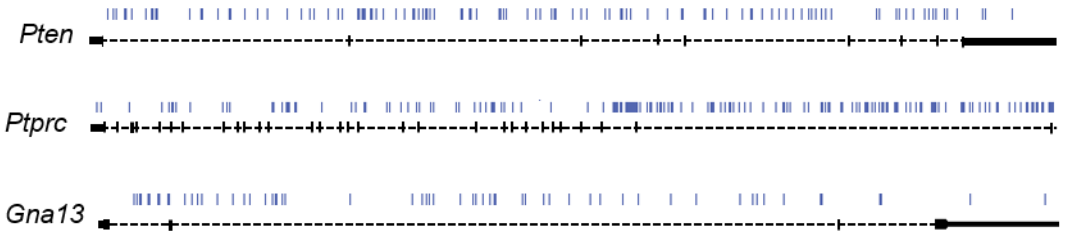

FIGURE 7-5. ITP insertion pattern in three selected tumour suppressor genes.

All ITP transposon insertions (indicated by blue lines) derived from the ITP-DLBCL study (IPB cohort) are shown for *Pten*, *Ptprc* and *Gna13*. For each gene, gene structure of the CCDS is shown (*Pten*: CCDS29753, *Ptprc*: CCDS48383 and *Gna13*: CCDS25577).

For pathway enrichment analysis, CIMPL-CISs were used as input. In general, CIMPL analysis is less stringent than TAPDANCE, therefore leading to identification of larger numbers of CISs (Rad *et al.*, 2015). Table 7-3 shows results from Ingenuity® Pathway Analysis (IPA®; Qiagen). Among the top enrichment pathways were pathways related to immune signalling like B-cell receptor signalling ($p = 1.6 \times 10^{-6}$), phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K) signalling in B lymphocytes ($p = 1.11 \times 10^{-4}$), T cell receptor signalling ($p = 1.11 \times 10^{-4}$) and CD28 signalling in T helper cells ($p = 5.07 \times 10^{-4}$). Moreover, pathways that are typically enriched in cancer, like protein kinase A signalling ($p = 7.96 \times 10^{-4}$), axonal guidance signalling ($p = 1.39 \times 10^{-3}$) and p53 signalling ($p = 2.64 \times 10^{-3}$), were also identified.

#	Significant Window	# of Libraries	P-Value Library	Gene
1	chr19:34286718-34333718	27	8,23894E-28	*Fas*
2	chr19:32759415-32823415	28	9,59682E-26	*Pten*
3	chr11:109364104-109384104	18	5,71324E-21	*Gna13*
4	chr1:138108882-138172882	24	4,2338E-20	*Ptprc*
5	chr11:23737011-23748011	15	1,24518E-19	*Rel* (1)
6	chr18:65442122-65453122	13	8,38704E-16	*Malt1* (2)
7	chr18:60803820-60814820	13	8,38704E-16	*Cd74*
8	chr1:171918101-171929101	12	6,17492E-14	*Slamf6*
9	chr6:51458297-51469297	12	6,17492E-14	*Hnrnpa2b1*
10	chrX:18168210-18271210	22	4,16709E-13	*Kdm6a*
11	chr15:25355868-25419868	17	6,11519E-11	*Basp1*

#	Significant Window	# of Libraries	P-Value Library	Gene
12	chr13:83565247-83585247	12	6,97489E-11	*Mef2c*
13	chr17:29035929-29046929	10	2,61435E-10	*Srsf3*
14	chr4:154916864-154927864	10	2,61435E-10	*Tnfrsf14*
15	chr11:44651614-44777614	21	2,64089E-10	*Ebf1*
16	chr3:51415665-51447665	13	6,39819E-10	*Naa15*
17	chr2:103785774-103796774	9	1,48062E-08	*Caprin1*
18	chrY:90799220-90810220	9	1,48062E-08	*Erdr1*
19	chr4:48216146-48280146	15	1,57602E-08	*Erp44*
20	chr18:50011771-50094771	16	4,77603E-08	*Tnfaip8*
21	chr16:33380024-33463024	16	4,77603E-08	*Zfp148*
22	chr18:7865497-7929497	14	2,30116E-07	*Wac*
23	chr8:80705031-80737031	11	3,7829E-07	*Smarca5*
24	chr4:6868705-6951705	15	5,73562E-07	*Tox*
25	chr7:25120281-25131281	8	7,54683E-07	*Pou2f2*
26	chr18:39418569-39521569	16	1,09594E-06	*Nr3c1*
27	chr3:60507279-60610279	16	1,09594E-06	*Mbnl1*
28	chr1:178316262-178336262	9	2,78251E-06	*Hnrnpu*
29	chr10:95517075-95537075	9	2,78251E-06	*Ube2n*
30	chr11:11690813-11754813	13	3,13593E-06	*Ikzf1*
31	chr9:88448855-88480855	10	8,10102E-06	*Syncrip*
32	chr1:87623240-87670240	11	2,04025E-05	*Inpp5d*
33	chr1:165928337-165975337	11	2,04025E-05	intergenic
34	chr11:79419153-79568153	17	2,70994E-05	*Nf1*
35	chr4:6977685-6988685	7	3,41927E-05	*Tox*
36	chr4:44770862-44781862	7	3,41927E-05	*Zcchc7*
37	chr9:51213664-51224664	7	3,41927E-05	*Pou2af1*
38	chr1:24679721-24690721	7	3,41927E-05	*Lmbrd1*
39	chr3:83992515-84003515	7	3,41927E-05	*D930015E06Rik*
40	chr3:123505203-123516203	7	3,41927E-05	intergenic
41	chr8:122934415-122998415	12	3,96827E-05	*Ankrd11*
42	chr11:3142151-3206151	12	3,96827E-05	*Sfi1* (3)
43	chr11:44914844-44978844	12	3,96827E-05	*Ebf1*
44	chr17:17520605-17603605	13	6,78559E-05	*Lnpep*
45	chr11:115620278-115703278	13	6,78559E-05	*Grb2*
46	chr13:83590432-83610432	8	7,80048E-05	*Mef2c*
47	chr11:98457171-98477171	8	7,80048E-05	*Ikzf3*
48	chr12:32954710-32974710	8	7,80048E-05	*Sypl*
49	chr2:6705649-6725649	8	7,80048E-05	*Celf2*
50	chr16:52053585-52156585	14	9,62206E-05	*Cblb*

TABLE 7-2. CISs identified in diffuse large B-cell lymphoma cohort derived from IPB mice.

Top 50 candidate genes (ranked by p-value library) derived from TAPDANCE analysis of all diffuse large B-cell lymphomas (n = 34) from the IPB cohort. Only insertions with a summarised read coverage (5' and 3' read) ≥ 20 were included. Notes: (1) Most insertions are located downstream of *Rel*; (2) Truncating insertion pattern; (3) CISs detected in *Sfi1* are known artefacts since multiple not-annotated copies of this gene exist in the mouse genome (Quinlan *et al.*, 2010). #, number.

Pathway	P-Value	# of Molecules
B Cell Receptor Signalling	1,63E-06	17
PI3K Signalling in B Lymphocytes	1,11E-04	12
T Cell Receptor Signalling	1,11E-04	11
CD28 Signalling in T Helper Cells	5,07E-04	11
Protein Kinase A Signalling	7,96E-04	19
Axonal Guidance Signalling	1,39E-03	20
p53 Signalling	2,64E-03	9
Cardiac Hypertrophy Signalling	3,12E-03	13
Ephrin Receptor Signalling	3,12E-03	11
Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes	3,12E-03	8
fMLP Signalling in Neutrophils	3,20E-03	9
iCOS-iCOSL Signalling in T Helper Cells	3,20E-03	9
PI3K/AKT Signalling	3,34E-03	9
Role of NFAT in Regulation of the Immune Response	3,49E-03	11
Actin Nucleation by ARP-WASP Complex	4,48E-03	6
Superpathway of Inositol Phosphate Compounds	4,48E-03	12
ERK/MAPK Signalling	5,40E-03	11
Telomerase Signalling	5,69E-03	8
Phospholipase C Signalling	5,69E-03	12
Insulin Receptor Signalling	5,69E-03	9
3-phosphoinositide Degradation	7,72E-03	9
Integrin Signalling	9,25E-03	11
3-phosphoinositide Biosynthesis	9,25E-03	10

TABLE 7-3. Pathway enrichment analysis of CISs derived from diffuse large B-cell lymphomas of the IPB cohort.

Results from Ingenuity® Pathway Analysis using CIMPL-CISs as input. All pathways with a p-value <0.01 (corrected for multiple testing with the Benjamini-Hochberg method) are shown. For each pathway, numbers of involved genes, which harbour CIMPL-CISs, are listed. #, number.

Lastly, CISs derived from ISB-DLBCL samples (n = 11) were compared to IPB-DLBCL-CISs (Table 7-4). For ISB-DLBCL samples, 40 CISs positioned within genes were found, only four of which (*Chl1*, *Malt1*, *Pten* and *Ptprc*) were not located on the donor chromosome. Of these four genes, *Malt1*, *Pten* and *Ptprc* were also detected in the IPB-cohort while no CISs within *Chl1* were identified (neither using CIMPL nor TAPDANCE). For a more detailed comparison of IPB- and ISB-insertions in DLBCL and tail samples, see Chapter 7.1.6.

Cohort	CIS Analysis	Number of CISs		Number of CISs within Unique Genes	
		Total	Chr 14 excl.	Total	Chr 14 excl.
IPB-DLBCL	TAPDANCE	136	94	94	88
IPB-DLBCL	CIMPL	280	263	261	254
ISB-DLBCL	TAPDANCE	46	6	16	4

TABLE 7-4. Overview of numbers of CISs identified in the IPB-DLBCL and ISB-DLBCL cohort.

Number of CISs (total and chromosome 14 excluded) and number of CISs within unique genes (total and chromosome 14 excluded) are listed. CISs, common insertion sites; Chr, chromosome; excl., excluded.

7.1.5 The Bloom-mutated background can induce loss of heterozygosity

As mentioned in Chapter 3.6, *Blm*$^{m3/m3}$ mice show increased rates of mitotic recombination in their cells, which fosters LOH. To demonstrate that by crossing *Rosa26*$^{PB/PB}$*;ITP2-M* and *Rosa26*$^{SB/SB}$*;ITP2-M* mice with *Blm*$^{m3/m3}$ mice, biallelic inactivation of TSGs can be achieved, SNP analysis was performed. This method exploits the fact that all mouse lines were generated, inter-crossed and maintained on a mixed *C57BL/6 x Sv/129* background. To show that LOH is indeed occurring in ITP tumours, DNA from a cell line established from a small intestine tumour of an ISB mouse and DNA from tail tissue of the same animal was used for SNP analysis. Study of the transposon insertion data showed that the tumour had a high-coverage insertion within *Apc*, a classical TSG, for which it is known that only bi-allelic inactivation promotes tumourigenesis (Sparks *et al.*, 1998). Regions containing two potential SNPs (*rs4137461* and *rs220642642*) were amplified and sequenced (Figure 7-6a). For both SNPs, two alleles (C/T) were present in tail DNA while only one SNP allele was identified in the tumour DNA (*rs4137461*: T and *rs220642642*: C, respectively) (Figure 7-6b), demonstrating loss of the second allele in the tumour.

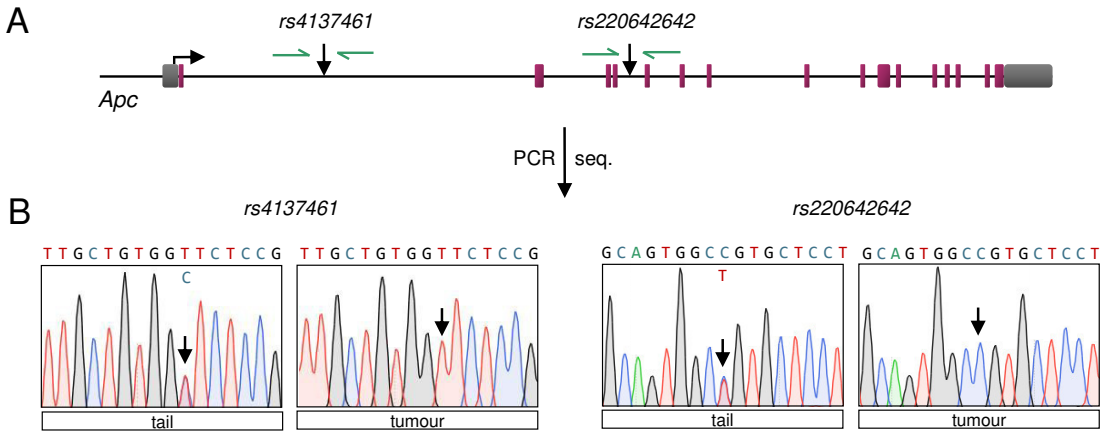

FIGURE 7-6. Detection of loss of heterozygosity in ITP tumours.

(**A**) Structure of the mouse *Apc* gene with indicated positions of single nucleotide polymorphisms (SNPs) *rs4137461* and *rs220642642* and of PCR primers (green arrows). (**B**) Sanger capillary sequencing results from tumour and tail DNA for each SNP (left: *rs4137461*; right: *rs220642642*) show that loss of heterozygosity takes place in the tumour (only one SNP allele identified) but not in the tail (two SNP alleles present). Figure adapted from Weber *et al.,* 2019. Seq., sequencing.

7.1.6 *PiggyBac* and *Sleeping Beauty* show different insertion properties in tumours and in tails

Since the identical transposon, namely *ITP2*, was mobilised by both transposases, the ITP screen allows conducting a side-by-side comparison of the *in vivo* characteristics/behaviours of *PB* and *SB*. For this, number of insertions per tumour, insertion pattern and local hopping behaviour of *PB* and *SB* were analysed. By comparing the number of insertions per DLBCL sample from the IPB and ISB cohort, it can be shown that – in tendency – IPB tumour samples harbour more transposon insertions than ISB tumours (average number of insertions: 7164 (IPB) vs 5400 (ISB)) (Figure 7-7a). To analyse the extent of local hopping in the ITP screen, percentages of insertions located on the donor chromosome (chromosome 14) were analysed for IPB and ISB DLBCL and tail DNA samples, respectively (Figure 7-7b). In DLBCL samples, while only 7.9% of all insertions in the IPB cohort were located on chromosome 14, 23.0% of all ISB insertions were. In the tail DNA, the bias for the donor chromosome was even more pronounced. Only 13.0% of all IPB-CISs were detected on chromosome 14, however, in the ISB cohort 39.9% of all insertions were identified on the donor chromosome. Lastly, the proportion of insertions in intra- and intergenic regions was compared in DLBCL and tail samples of IPB and ISB mice (Figure 7-7c). While 50.8% of all IPB insertions in the tail DNA were found to be intragenic, there were significantly less insertions within genes in the ISB

cohort (37.5%; p<0.01; Odds Ratio (OR) = 1.62; CI = 1.59 – 1.65). Even more intragenic insertions were identified in tumour samples: 55.8% of all insertions in IPB-DLBCLs were intragenic whereas only 45.1% of all insertions in ISB-DLBCL samples were located within genes (p<0.01; OR = 1.54; CI = 1.51 – 1.56). These differences were less pronounced when specifically analysing the insertion distribution on the donor chromosome (Figure 7-7d). In tail DNA, for both, the IPB and the ISB cohort, more than two-third of all insertions on chromosome 14 (IPB: 72.7%; ISB: 73.0%) were found to be intergenic (p-value not significant; OR = 1.54; CI = 1.51 – 1.56). In contrast, in DLBCL samples, there were more intergenic insertions in the ISB cohort (68.7%) than in the IPB cohort (63.1%) (p<0.01; OR = 1.29; CI = 1.23 – 1.35).

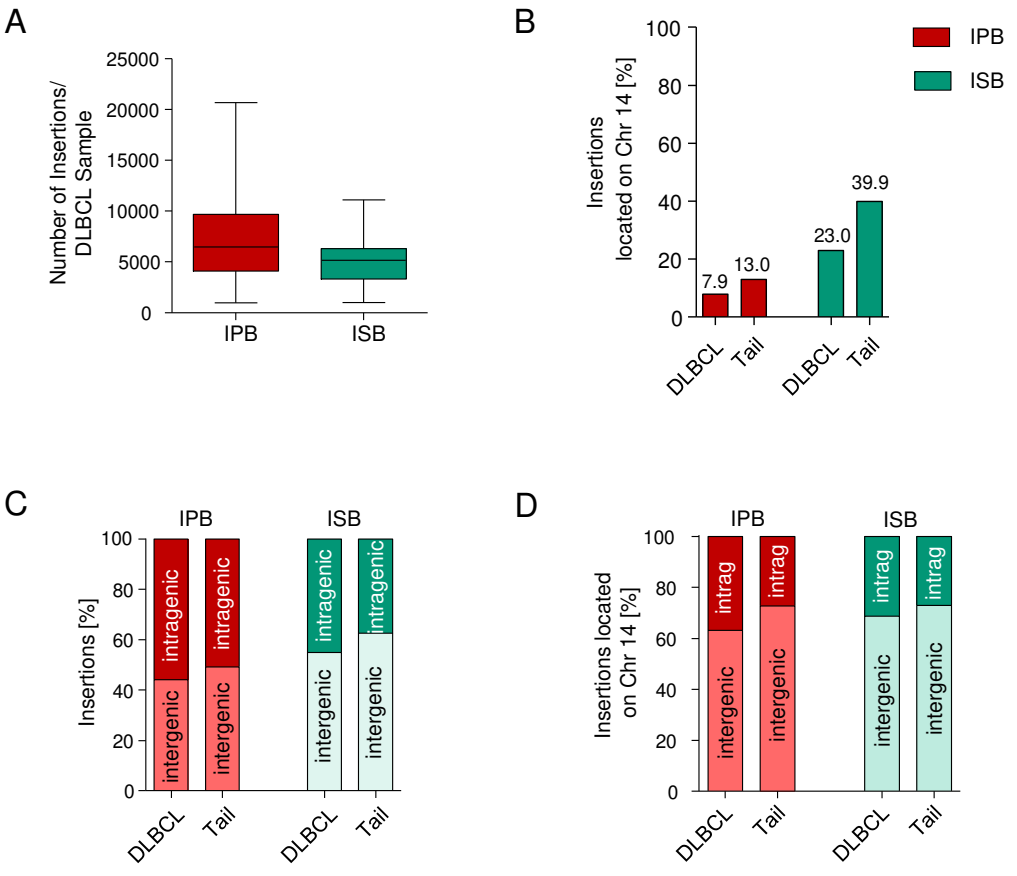

FIGURE 7-7. Comparison of *PiggyBac* and *Sleeping Beauty* insertion characteristics.

(**A**) Box plots show numbers of insertions per tumour (DLBCL) derived from IPB (n = 34; *Tg(ITP2)M;Rosa26*$^{PB/+}$*;Blm*$^{m3/m3}$) and ISB (n = 11; *Tg(ITP2)M;Rosa26*$^{SB/+}$*;Blm*$^{m3/m3}$) mice. Horizontal lines indicate mean numbers of insertions. (**B**) Bar graph shows percentage of insertions located on the donor chromosome (Chr 14) in DLBCLs as well as in tails from IPB (DLBCL, n = 34; tail, n = 12) and ISB

FIGURE 7-7 (continued)

(DLBCL, n = 11; tail, n = 12) mice. (**C – D**) Bar graphs show percentages of insertions located within genes (intragenic) and in intergenic regions in DLBCLs and tails from IPB and ISB mice. (**C**) shows distribution of all insertions whereas in (**D**) only insertions identified on chromosome 14 are included. DLBCL, diffuse large B-cell lymphoma; Chr, chromosome; intrag, intragenic.

7.2 CRISPR/Cas-based recessive genetic screening in the mouse liver

Most results from this chapter can be also found in Weber *et al.*, 2015, which was published in *Proceedings of the National Academy of Sciences* in 2015. Rupert Öllinger (Klinikum rechts der Isar, Technical University Munich, München, Germany) and I share first authorship of this publication.

7.2.1 Selection of tumour suppressor genes and single guide RNA validation

To study (i) the feasibility of somatic multiplex-mutagenesis using the CRISPR/Cas system and (ii) the utility of the CRISPR/Cas system for recessive genetic screening in the adult mouse liver, initially a panel of ten tumour suppressors, for which ten sgRNAs were designed, was selected. Six of the chosen genes had an already known role in ICC/HCC pathogenesis; among these were

- *Apc* since Wnt pathway activation is detected in many ICC and HCC cases;
- the chromatin remodelling gene *Arid1a*, for which mutations are found in up to one-third of all ICCs;
- *Pten*, for which promoter hyper-methylation and loss of expression is found in many ICC/HCC cases;
- the *TGF-β* pathway gene *Smad4*, for which mutations and loss of expression were particularly reported in ICCs;
- and the master regulatory transcription factor *Trp53*, which is frequently mutated in ICC and HCC (for references see Table 7-5).

Moreover, for *Cdkn2a*, which – by alternative splicing – encodes for two tumour suppressor proteins, $p16^{Ink4A}$ and $p19^{Arf}$, two sgRNAs were included: one *(sg_Cdkn2a-e1β)* targeting exon-1β, which solely encodes for $p19^{Arf}$, and the other one *(sg_Cdkn2a-e2)* targeting exon-2, which is "shared" by both tumour suppressors (Figure 7-8a). Epigenetic dysregulation and loss of expression of *CDKN2A* is reported in a majority of human ICCs and HCCs.

Additionally, the TSG *Tet2* was selected for targeting; although nearly no genetic alterations of this gene have been reported in ICC/HCC so far, it is a putative negatively regulated downstream target of *IDH1/2, which* are mutated in up to 10% of all human ICCs. Lastly, sgRNAs directed against *Brca1* and *Brca2*, which seem to play no crucial roles in liver tumourigenesis, were included as negative controls. Table 7-5 gives an overview of genetic alterations detected for the targeted TSGs in ICC/HCC based on an extensive literature search.

Gene	ICC Range [%]	Sources	HCC Range [%]	Sources
APC	**PM:** 26.6 – 47.2 **WNT** - *A:* 82 - *B:* 15	Sugimachi *et al.*, 2001; Lee *et al.*, 2002; Yang *et al.*, 2005	**MUT:** 0 – 3.0 **DEL:** 0 – 0.5 **LOSS:** 53.0	Cerami *et al.*, 2012; Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Huang *et al.*, 2012; Cleary *et al.*, 2013; Gao *et al.*, 2013; Kan *et al.*, 2013; Schulze *et al.*, 2015
ARID1A	**MUT:** 9 – 35.5	Ong *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Goeppert *et al.*, 2014	**MUT:** 2.0 – 16.0 **DEL:** 0 – 1.4	Cerami *et al.*, 2012; Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Huang *et al.*, 2012; Cleary *et al.*, 2013; Gao *et al.*, 2013; Schulze *et al.*, 2015
BRCA1	**MUT:** 0 – 3.6	Ong *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Goeppert *et al.*, 2014	**MUT:** 0 – 2.0 **DEL:** 0 – 0.3	Cerami *et al.*, 2012; Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Huang *et al.*, 2012; Cleary *et al.*, 2013; Gao *et al.*, 2013; Schulze *et al.*, 2015
BRCA2			**MUT:** 0 – 5.7 **DEL:** 0 – 0.8	Cerami *et al.*, 2012; Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Huang *et al.*, 2012; Cleary *et al.*, 2013; Gao *et al.*, 2013; Schulze *et al.*, 2015
CDKN2A	**MUT:** 0 – 5.6 **PM:** 15.7 – 83.0 **LOSS:** 35.7 **DEL:** 18.0	Tannapfel *et al.*, 2000; Kang *et al.*, 2002; Lee *et al.*, 2002; Yang *et al.*, 2005; Sriraksa *et al.*, 2011; Ong *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Goeppert *et al.*, 2014; Ross *et al.*, 2014	**MUT:** 0 – 2.9 **PM:** 17.6 **LOSS:** 72.2 **DEL:** 4.0 – 6.4	Yang *et al.*, 2003; Cerami *et al.*, 2012; Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Huang *et al.*, 2012; Cleary *et al.*, 2013; Gao *et al.*, 2013; Schulze *et al.*, 2015
PTEN	**MUT:** 0 – 10.7 **PM:** 35.3	Ong *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Goeppert *et al.*, 2014; Ross *et al.*, 2014	**MUT:** 0 – 4.0 PM: 16.1 **DEL:** 4.0 **LOSS:** 40.9 – 57.1	Fukai *et al.*, 2005; Wang *et al.*, 2007; Totoki *et al.*, 2011; Cerami *et al.*, 2012; Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Huang *et al.*, 2012; Cleary *et al.*, 2013; Gao *et al.*, 2013; Schulze *et al.*, 2015
SMAD4	**MUT:** 0 – 16.7 **LOSS:** 45.2	Tannapfel *et al.*, 2000; Ong *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Goeppert *et al.*, 2014	**MUT:** 0 - 0.9 **DEL:** 0 - 0.8	Cerami *et al.*, 2012; Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Huang *et al.*, 2012; Cleary *et al.*, 2013; Gao *et al.*, 2013; Schulze *et al.*, 2015
TET2			**MUT:** 0 – 2.0 **DEL:** 0 - 0.8	Cerami *et al.*, 2012; Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Huang *et al.*, 2012; Cleary *et al.*, 2013; Gao *et al.*, 2013; Schulze *et al.*, 2015
TP53	**MUT:** 6 – 44.4 PM: 61.1	Ong *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Sia *et al.*, 2013; Goeppert *et al.*, 2014	**MUT:** 18 – 51.8 **DEL:** 0 - 3.0	Cerami *et al.*, 2012; Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Huang *et al.*, 2012; Cleary *et al.*, 2013; Gao *et al.*, 2013; Schulze *et al.*, 2015

TABLE 7-5. Literature-based study of tumour suppressor gene alterations in human liver cancers.

To select tumour suppressor genes for targeting by CRISPR/Cas, an extensive literature search was conducted. Mutations, epigenetic dysregulations (in form of promoter hypermethylation), loss of (protein) expression and deletions in human intrahepatic cholangiocarcinomas (ICCs) and hepatocellular carcinomas (HCCs) identified in whole genome/exome sequencing and other genetic/genomic studies are listed. Frequently altered genes (per gene at least one alteration found in more than >25% of all cases) are shaded in dark blue. Aberrant WNT pathway activity was linked to reduced membranous expression of β-catenin (A) and aberrant nuclear expression of

TABLE 7-5 (continued)

β-catenin (B). A modified version of this table can be found in Weber *et al.*, 2015. PM, promoter hypermethylation; MUT, mutation; DEL, gene located in a deleted region; LOSS, loss of expression; WNT, WNT pathway activation.

For each TSG, one sgRNA (with exception of *Cdkn2a,* for which two sgRNAs were generated), targeting one of the first exons near the transcription start site, was designed (Table 6-5) and cloned into the *CRISPR-SB* vector that was specifically generated for this research project (Figure 7-8b). In this vector, *SB* TRs flank sgRNA and Cas9 expression cassettes. In combination with a *SB* transposase *(hSB5)* (Figure 7-8c), this might enable integration of the CRISPR/Cas components into the genome of (liver) cells thereby allowing ongoing Cas9 and sgRNA expression.

All ten sgRNAs were validated in two mouse cancer cell lines, *PPT-53631* (which harbours a homozygous *Cdkn2a* deletion and was therefore not suited for *sg_Cdkn2a-e1β* and *sg_Cdkn2a-e2* testing) and *PPT-4072*. On-target editing efficiencies were determined by mismatch cleavage assays showing that all sgRNAs efficiently cleave their target DNA, with indel frequencies ranging from 31-59% in *PPT-53631* (Figure 7-8d) and 12-15% in *PPT-4072* cells (Figure 7-8e).

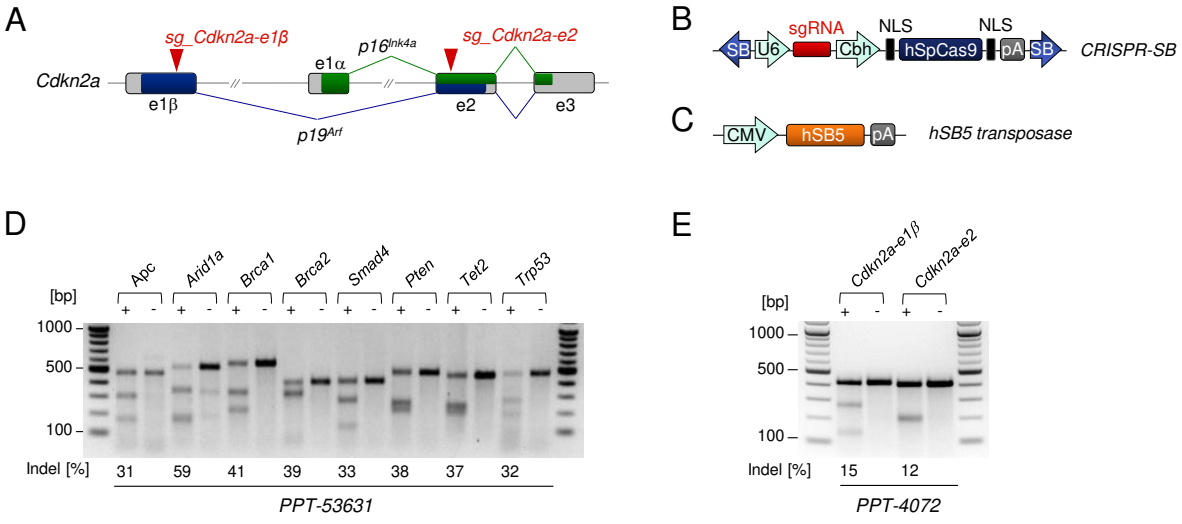

FIGURE 7-8. Vector design and singe guide RNA validation.

(**A**) For *Cdkn2a*, two single guide RNAs (sgRNAs; *sg_Cdkn2a-e1β* and *sg_Cdkn2a-e2*) targeting two different exons, exon-1β and exon-2, were designed. (**B**) A vector *(CRISPR-SB)* harbouring a sgRNA as well as a Cas9 expression cassette flanked by *SB* terminal repeats was co-injected with (**C**) a *SB* transposase vector *(hSB5)*.

FIGURE 7-8 (continued)

(**D-E**) For determination of on-target editing efficiencies, mismatch cleavage assays were conducted in *PPT-53631* and *PPT-4072* cell lines. Figure adapted from Weber *et al.*, 2015. SB, *Sleeping Beauty*; U6, U6 RNA polymerase III promoter; sgRNA, single guide RNA; Cbh, chicken β-actin hybrid intron promoter; *hSpCas9*, *Streptococcus pyogenes* Cas9; NLS, nuclear localisation signal; pA, polyadenylation signal; CMV, cytomegalovirus; *hSB5*, hyperactive *Sleeping Beauty* transposase 5.

7.2.2 CRISPR/Cas components are efficiently delivered into the mouse liver

To test if the CRISPR/Cas components are efficiently delivered into the mouse liver by HTVI, five wild type mice, which received the Cas9/10-sgRNA/*hSB5* cocktail (Figure 7-9a), were sacrificed two weeks post injection. Isolated liver DNA was analysed by quantitative real-time PCR for (i) presence of Cas9 DNA and (ii) sgRNA distribution. All five mice showed detectable Cas9 DNA in varying levels (Figure 7-9b), and a random distribution of the 10 sgRNAs was observed in most mouse livers (with the exception of mouse *28.1*, for which only four sgRNAs were detectable) (Figure 7-9c), demonstrating efficient and mostly unbiased CRISPR/Cas delivery.

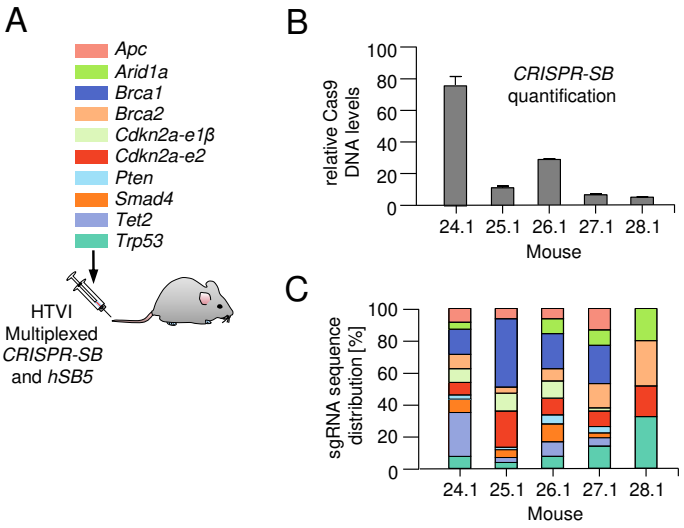

FIGURE 7-9. Cas9 and single guide RNA detection in the mouse liver two weeks post hydrodynamic tail vein injection.

(**A**) Ten different single guide RNAs (sgRNAs) genes were delivered into the mouse liver by hydrodynamic tail vein injection (HTVI). (**B**) Cas9 DNA copies were quantified in liver samples from mice two weeks post HTVI by quantitative real-time PCR and normalised to the *Apolipoprotein B (Apob)* copy number. (**C**) Distribution of sgRNA vectors detected in liver samples from mice two weeks post HTVI by quantitative real-time PCR. Colour code from A. Figure adapted from Weber *et al.*, 2015.

7.2.3 Hepatic delivery of Cas9 and ten single guide RNAs induces liver tumourigenesis in a predisposing context

Next, the ten *CRISPR-SB* vectors (and *hSB5* transposase) were delivered into the livers of *Albumin*$^{Cre/+}$*;LSL-Kras*$^{G12D/+}$ mice by HTVI. Mice were aged until they showed signs of tumour development or liver tumours were detected via MRI scans. A total of 21 liver tumours from eight animals was collected 20 to 30 weeks post HTVI; no metastases were observed in the Cas9/10-sgRNA/hSB5 cohort.

Mice developed ICCs (6 out of 10 diagnosed liver tumours; 60%) as well as HCCs (4 out of 10 diagnosed liver tumours; 40%). Eleven tumours were very small (<1 mm) and therefore no histopathological analysis was possible (unclassified tumours). ICCs showed diverse differentiation stages (well, moderately and poorly differentiated; Figure 7-10a) and tumour cells were positive for the ductal marker cytokeratin 19 (CK19) and the oval cell marker A6. Collagen IV was strongly expressed in the tumour-associated stroma (Figure 7-10a). HCCs were moderately to poorly differentiated (Figure 7-10b). The tumour cells exhibited expression of the HCC marker Golgi phosphoprotein 2/Golgi membrane protein GP73 (GOLM1/GP73) and slight to moderate expression of α-fetoprotein (AFP), which is another marker for HCC. All HCCs displayed high proliferative activity as indicated by intensive Ki67 staining (Figure 7-10b).

No liver tumours (ICCs/HCCs) were observed in *Albumin*$^{Cre/+}$*;LSL-Kras*$^{G12D/+}$ mice, which were injected with *hSB5* transposase and a Cas9-only expression plasmid (n = 8) and aged up to 30 weeks post HTVI (control cohort).

While no indels were identified in livers derived from mice of the control cohort, NGS of all target regions in the 21 liver tumours derived from the Cas9/10-sgRNA/*hSB5* cohort revealed a total of 167 indels (with a mutant read frequency (MRF; defined as fraction of mutant-reads/all-reads at individual target sites) >1%) (Figure 7-11a). Of these, about three-fourth were deletions and approximately one-fourth were insertions. About 75% of all indels ranged within 1 bp and 3 bp (1-bp-deletion: 34.1% of all indels; 2-bp-deletion: 8.4%; 3-bp-deletion: 8.4%; 1-bp-insertion: 21.6%). Whereas deletions as large as 201 bp were detected, the largest identified insertion was only 15 bp in size. In contrast to the predominate occurrence of deletions, some target sites seemed to be prone for induction of insertions (Figure 7-11b). For instance, 64.1% of all indels detected within the *Apc* target region were insertions, with deletions only accounting for 35.9%. Figure 7-11c exemplarily shows indels detected in three different target regions in three individual tumours (*Arid1a* in Tu1; *Tet2* in Tu4 and *Pten* in Tu2).

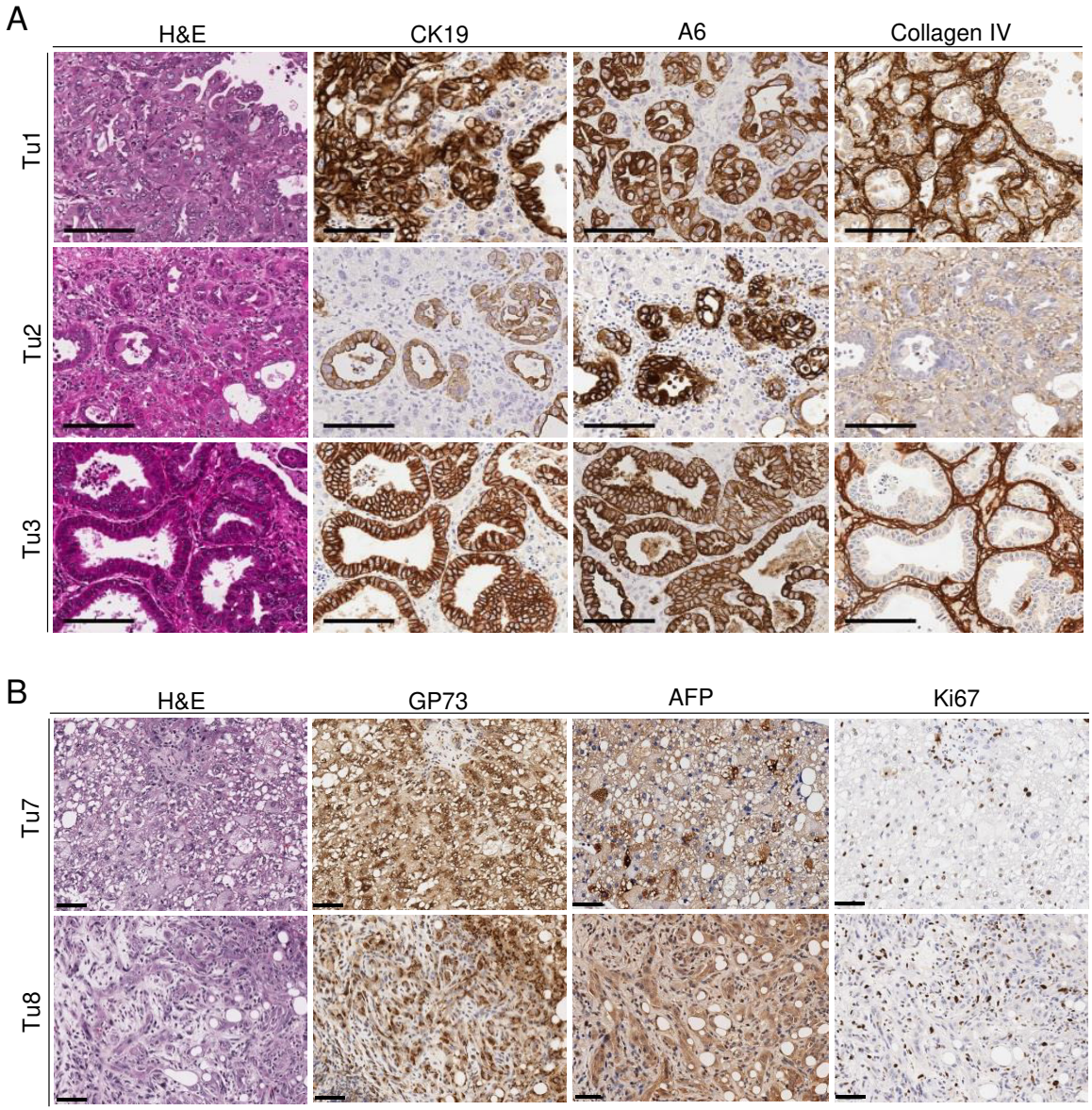

FIGURE 7-10. Liver tumour histology.

(**A**) Intrahepatic cholangiocarcinomas showed different stages of differentiation (Tu1: moderately to poorly differentiated (poorly differentiated parts not shown); Tu2: well to moderately differentiated; Tu3: well differentiated) and highly expressed cytokeratin 19 (CK19) and A6. Tumour stroma expressed collagen IV. Bars, 100 µm. (**B**) Hepatocellular carcinomas exhibited moderate (Tu7) to poor differentiation (Tu8) and expressed GP73 (strongly) and AFP (slightly). Proliferative activity was high (Ki67). Scale bars, 50 µm. Figure adapted from Weber *et al.*, 2015. H&E, haematoxylin and eosin; GP73, Golgi membrane protein GP73; AFP, α-fetoprotein.

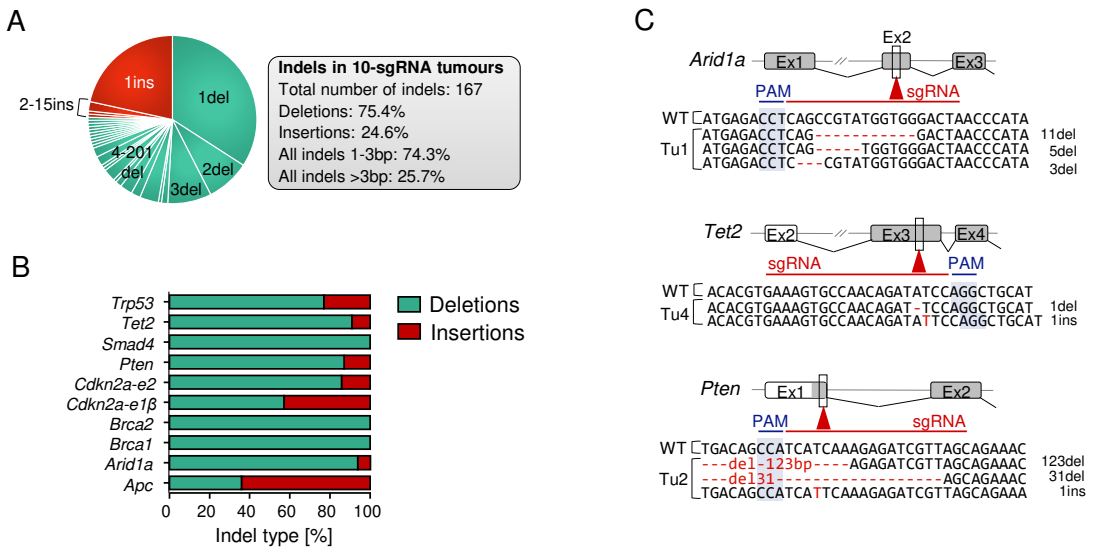

FIGURE 7-11. CRISPR/Cas-induced indel types at genomic target sites.

(**A**) Analysis of tumours from the Cas9/10-sgRNA/*hSB5* cohort showed that the majority of induced indels at the target sites were deletions, with insertions only accounting for about 25%. About 75% of all indels were small and ranged between 1 bp and 3 bp. All indels with a mutant read frequency >1% are included. (**B**) The ratio between deletions and insertions mostly depended on the sequence of the target region and was therefore quite variable. All indels with a mutant read frequency >1% are included. (**C**) Exemplarily, alignments of selected reads harbouring different indels in three target regions (*Arid1a*, *Tet2* and *Pten*) found in three tumours (Tu1, Tu4 and Tu2) are shown. Figure adapted from Weber *et al.*, 2015. Indel, insertion/deletion; ins, insertion; del, deletion; WT, wild type; PAM, protospacer adjacent motif; sgRNA, single guide RNA.

7.2.4 Cancer-relevant indels undergo clonal selection during tumourigenesis

Whereas healthy liver samples from tumour-bearing mice showed no or only very few indels with low MRFs (healthy-liver-sample-1_*Pten*: 1.3% and healthy-liver-sample-4_*Tet2*: 1.2%), all liver tumours exhibited indels over the MRF threshold set to 4% (which was used to exclude their origin in normal tissue) (Figure 7-12a). In individual tumours, MRFs reached up to 40-60% (Tu1_*Pten*: 62%; Tu7_*Pten*: 41%; Tu10_*Pten*: 41%), reflecting clonal indel expansion. While some tumours revealed indels with high MRFs (e.g. Tu1, Tu5 and Tu10), others had indels with considerable lower MRFs (e.g. Tu2, Tu3 and Tu4). This can be at least partly explained with varying tumour-cell/non-tumour-cell content, which is exemplary shown in Figure 7-12b. In this example, the non-tumour-cell content (e.g. stromal cells) in Tu2 was much higher than in Tu1, which is mirrored by lower MRFs observed at all target loci.

CRISPR/Cas-induced indels showed a non-random distribution over all ten different target sites ($p = 2.2 \times 10^{-15}$; x^2 test), meaning that some target regions exhibited significantly more indels than others. For instance, *Brca1* and *Brca2* mutations were largely non-existent (only one *Brca1* indel detected in 1/21 tumours). In comparison to *Brca1*/*Brca2*, several target genes were significantly more frequently hit: *Pten* ($p = 6.4 \times 10^{-15}$), *Apc* ($p = 9.3 \times 10^{-7}$), *Tet2* ($p = 6.6 \times 10^{-5}$), *Cdkn2a-e2* ($p = 0.0007$), *Trp53* ($p = 0.007$) and *Arid1a* ($p = 0.02$; Fisher's Exact Test). The *Pten* target locus was found to be mutated in 21/21 tumours (100%), *Apc* in 13/21 tumours (62%), *Tet2* in 10/21 tumours (48%), *Cdkn2a-e2* and *Trp53* in 6/21 tumours (29%), and *Arid1a* in 5/21 tumours (24%). In the ICC cohort, the most commonly disrupted genes/target loci were *Pten* and *Tet2* (6/6 ICCs; 100%), *Apc*, *Cdkn2a-e2* and *Trp53* (4/6 ICCs; 67%) and *Arid1a* (3/6 ICCs; 50%). In contrast, in the HCC cohort, *Pten* was mutated in every tumour (4/4 HCCs; 100%) whereas *Tet2*, *Apc* and *Cdkn2a-e2* were mutated in only one tumour each (1/4 HCCs; 25%). Additionally, for the two sgRNAs (*sg_Cdkn2a-e1β* and *sg_Cdkn2a-e2)* targeting the *Cdkn2a* locus, significant differences regarding indel occurrence were observed: whereas no indels (above a MRF cut-off of 0.2% which was set to account for technical sequencing errors) were found in the $p19^{Arf}$-encoding *Cdkn2a-e1β* target locus, the *Cdkn2a-e2* locus (which codes for $p16^{Ink4a}$ as well as $p19^{Arf}$) was targeted in one-third (7/21) of all tumours ($p = 0.009$; Fisher's Exact Test).

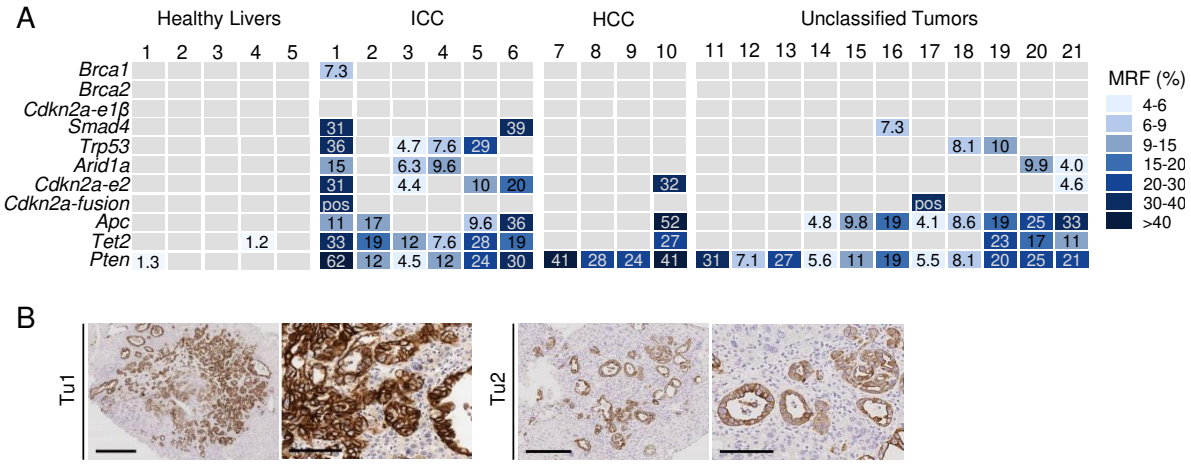

FIGURE 7-12. Analysis of mutant read frequencies in liver tumours derived from the Cas9/10-sgRNA/*hSB5* cohort.

(**A**) Mutant read frequencies (MRFs) determined by amplicon-based next-generation sequencing are shown for healthy liver samples of tumour-bearing mice (healthy livers 1-5) and for liver tumours (Tu1-Tu21). MRFs are defined as fraction of mutant-reads/all-reads at individual target sites. All frame-shift causing indels with MRFs >1% and 4% are shown for healthy livers and tumours, respectively.

FIGURE 7-12 (continued)

(**B**) Comparison of tumour cell/non-tumour-cell ratio in two different intrahepatic cholangiocarcinomas (ICCs; Tu1 and Tu2). Cytokeratin 19 was used as a marker for tumour cells. Scale bars, right images Tu1 and Tu2, 400 µm and left images Tu1 and Tu2, 100 µm. Figure adapted from Weber *et al.*, 2015. HCC, hepatocellular carcinoma.

7.2.5 Some liver cancers show signs of CRISPR-related genetic heterogeneity

In some liver tumours (e.g. Tu1, Tu4, Tu5 and Tu21), MRFs differed widely between individual target sites (e.g. Tu1_*Pten*: 62% and Tu1_*Arid1a*: 15%), indicating potential intra-tumour heterogeneity. To examine this phenomenon further, Tu1 was analysed in more detail. In addition to the already examined sample (R1) (Figure 7-13a), two further areas of Tu1 were microdissected and analysed (Figure 7-13a and b): R2, which showed a well-differentiated tubular growth pattern, and R3 displaying a poorly differentiated solid growth pattern.

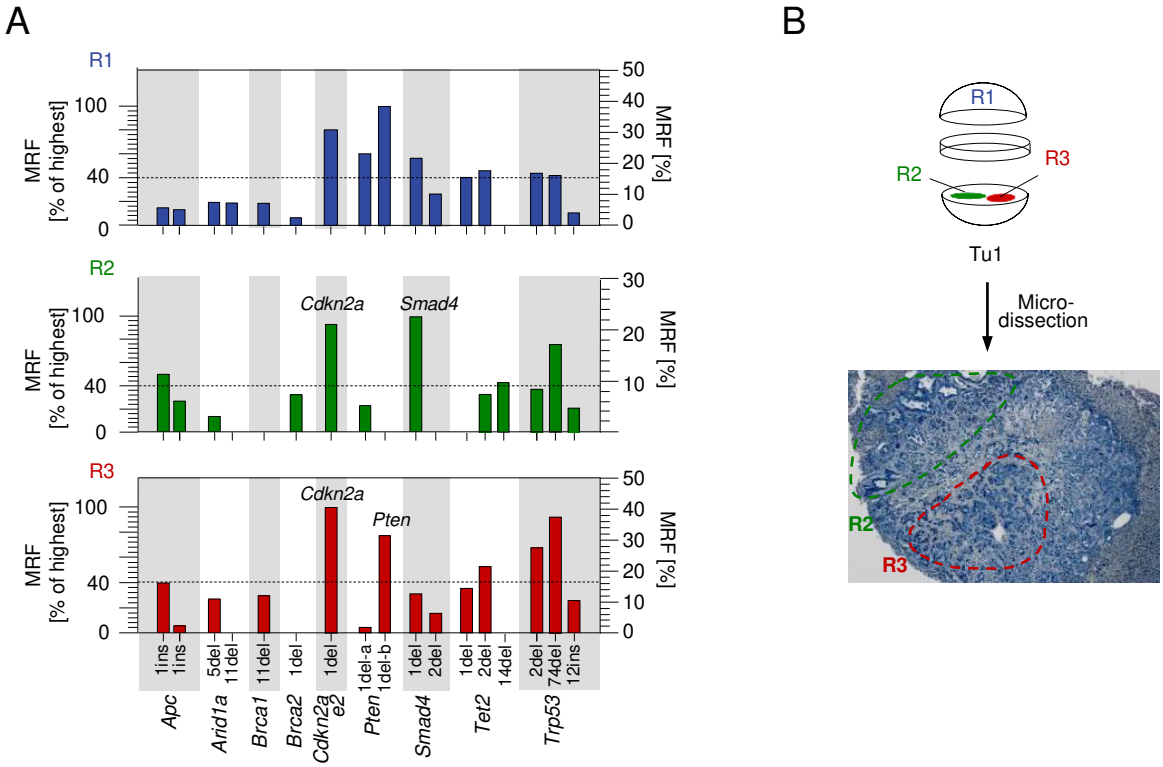

FIGURE 7-13. Analysis of intra-tumour heterogeneity.

(**A**) Analysis of mutant read frequencies (MRFs) in region 1 (R1), region 2 (R2) and region 3 (R3) of Tu1. (**B**) Tu1 was microdissected and DNA from two morphologically different regions (R2 and R3) was isolated. Figure adapted from Weber *et al.*, 2015.

All three regions harboured a specific deletion of 1 bp and an additional 18-kb-deletion (which is described in Chapter 7.2.6) within the *Cdkn2a-e2* target region. This indicates that these deletions might by an early event during tumour development. R2 showed a predominant 1-bp-deletion in the *Smad4* target locus (R2_*Smad4*_1-del: 21.6%), which had a much lower MRF in R3 (R3_*Smad4*_1-del: 12.3%; note that the MRFs in R3 are generally higher than in R2). In contrast, R3 had a high-frequent 1-bp-deletion in the *Pten* target region (R3_*Pten*_1-del-b: 30.9%) that was not detectable in R2. These results suggest that genetic heterogeneity might underlie phenotypic intra-tumour diversity (i.e. different differentiation stages of tumours).

7.2.6 CRISPR/Cas somatic multiplex-mutagenesis can induce chromosomal rearrangements in liver tumours

Since with the multiplexed CRISPR/Cas approach, some chromosomes have multiple target regions (chromosome 4: three; chromosome 11: two and chromosome 18: two) (Figure 7-14a), the induction of large deletions by CRISPR/Cas might be a possibility. To screen for these intra-chromosomal rearrangements, a PCR-based approach was used: out of 105 possible deletions in 21 tumours (five possible deletions/tumour), evidence for an approximately 18 kb large *Cdkn2a-e1β*/*Cdkn2a-e2* fusion product was found in 2/21 tumours (Tu1 and Tu17) (Figure 7-14b and c). Quantitative real-time PCR was performed to determine the relative copy number of the *Cdkn2a-e1β*/*Cdkn2a-e2* fusion allele in Tu1 (R1) and its regions R2 and R3. While no fusion allele was detected in negative control DNA samples (28.1 and wild type DNA), the amount of the fusion allele (in relation to the *Cdkn2a* allelic locus) was calculated to be about 20% in Tu1 (R1), R2 and R3 (Figure 7-14d). To rule out additional large deletions within the *Cdkn2a* locus, long-range PCRs (~3 kb product size) were conducted (Figure 7-14e). No large deletions were found in the analysed liver tumour (Tu1-Tu5) and normal liver samples (24.1, 25.1, 26.1, 27.1 and 28.1).

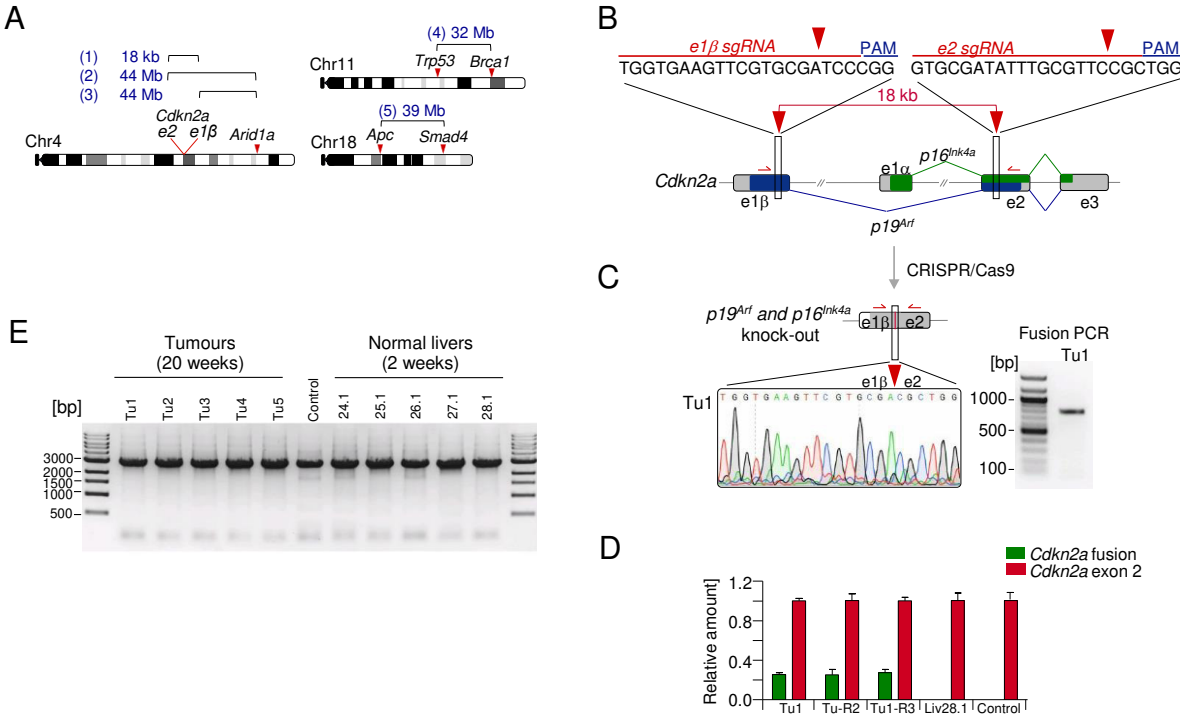

FIGURE 7-14. Induction of intra-chromosomal large deletions in the Cas9/10-sgRNA/*hSB5* cohort.

(**A**) Possible intra-chromosomal rearrangements in the Cas9/10-sgRNA/*hSB5* cohort. (**B**) Overview of the *Cdkn2a* locus, which was targeted by two different sgRNAs (*sg_Cdkn2a-e1β* and *sg_Cdkn2a-e2*). (**C**) Induction of an about 18-kb-deletion is shown by Sanger capillary sequencing chromatogram and PCR for Tu1. (**D**) Quantitative real-time PCR was performed to detect the allelic frequency of the *Cdkn2a* fusion allele. (**E**) Long-range PCRs were conducted to detect potential large deletions in the *Cdkn2a* locus. Figure adapted from Weber *et al.*, 2015.

7.2.7 Further expansion of the multiplexed-mutagenesis approach leads to results comparable to the 10-sgRNA cohort

To study if a further expansion of the multiplex-mutagenesis approach is feasible, in an additional set of experiments, a Cas9/18-sgRNA/*hSB5* cocktail was administered by HTVI. The 18-sgRNA-mixture included the ten original sgRNAs and eight novel sgRNAs targeting TSGs, which are involved in liver cancer (as recently published in several NGS studies (Fujimoto *et al.*, 2012; Guichard *et al.*, 2012; Cleary *et al.*, 2013; Kan *et al.*, 2013)): *Arid1b*, *Arid2*, *Arid5b*, *Atm*, *Cdkn2b*, *Errfi1*, *Igsf10* and *Irf2*.

In separate experiments, these 18 sgRNAs were injected into (i) *Albumin*$^{Cre/+}$*;LSL-Kras*$^{G12D/+}$ mice (Kras experimental cohort; n = 3) and (ii) wild type mice that were treated with nine weekly doses of CCl_4 (liver injury model; CCl_4 experimental cohort; n = 7) (Figure 7-15a). While all mice from both cohorts developed liver tumours, mock-treated mice did show no signs of

hepatic tumourigenesis. For mock treatment, *Albumin*$^{Cre/+}$*;LSL-Kras*$^{G12D/+}$ mice (Kras control cohort; n = 8) and *CCl*$_4$-treated wild type mice (*CCl*$_4$ control cohort; n = 4) were injected with a Cas9-only expression plasmid in combination with *hSB5* transposase.

Indel analysis of tumours from the Kras (n = 6; Tu22-Tu27) and *CCl*$_4$ experimental cohort (n = 35; for eight tumours NGS results are available: Tu28-Tu35) revealed a similar indel distribution as observed in the Cas9/10-sgRNA/*hSB5* cohort (Figure 7-15b). For instance, the most frequently targeted region was *Pten*, which was altered in 13/14 tumours (92.9%). In addition, only 2 out of 14 tumours had no indels in at least one of the genes of the ARID family (*Arid1a*, *Arid1b*, *Arid2* and *Arid5b*), which are involved in chromatin remodelling. For example, while *Arid1a* was targeted in 4/14 cancers (28.6%), the *Arid1b* target region was altered in 11/14 tumours (78.6%). In contrast, *Brca1* and *Brca2* were rarely targeted, with *Brca1* being mutated in 1/14 tumours (7.1%) and *Brca2* in 2/14 tumours (14.3%).

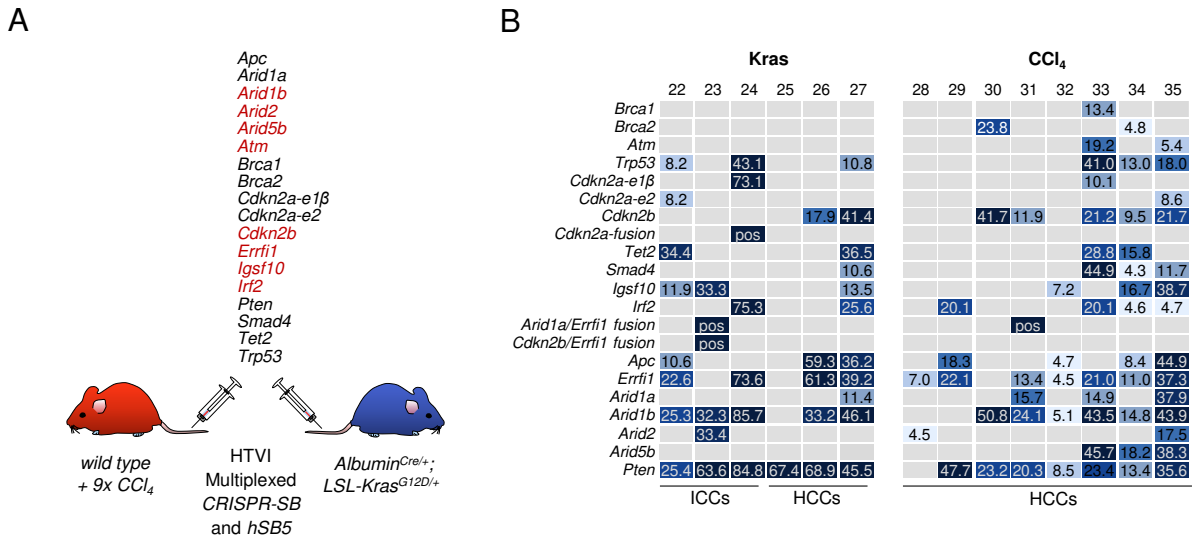

FIGURE 7-15. 18-sgRNA CRISPR/Cas multiplex-mutagenesis approach for liver mutagenesis.

(**A**) 18 sgRNAs were simultaneously injected into wild type mice (treated with nine weekly doses of *carbon tetrachloride (CCl*$_4$*)* after injection) or *Albumin*$^{Cre/+}$*;LSL-Kras*$^{G12D/+}$ mice. Genes not targeted in the 10-sgRNA approach are highlighted in red. (**B**) Mutant read frequencies determined by amplicon-based next-generation sequencing are shown for tumours of the two 18-sgRNA cohorts. All frame-shift causing indels with MRFs >4% are shown. For explanation of the colour code, see Figure 7-12. HTVI, hydrodynamic tail vein injection; ICC, intrahepatic cholangiocarcinoma; HCC, hepatocellular carcinoma.

For large deletion analysis, PCR-based deletion testing was performed. Out of 533 possible intrachromosomal deletions (13 possible deletions/tumour; 41 tested tumours in total) (Figure

7-16a), four large deletions (Tu24: 18-kb-deletion in the *Cdkn2a* locus; Tu23: 62-Mb-deletion between *Cdkn2b* and *Errfi1* and Tu23 and Tu31: 17-Mb-deletion between *Arid1a* and *Errfi1*) were detected (Figure 7-16b and c).

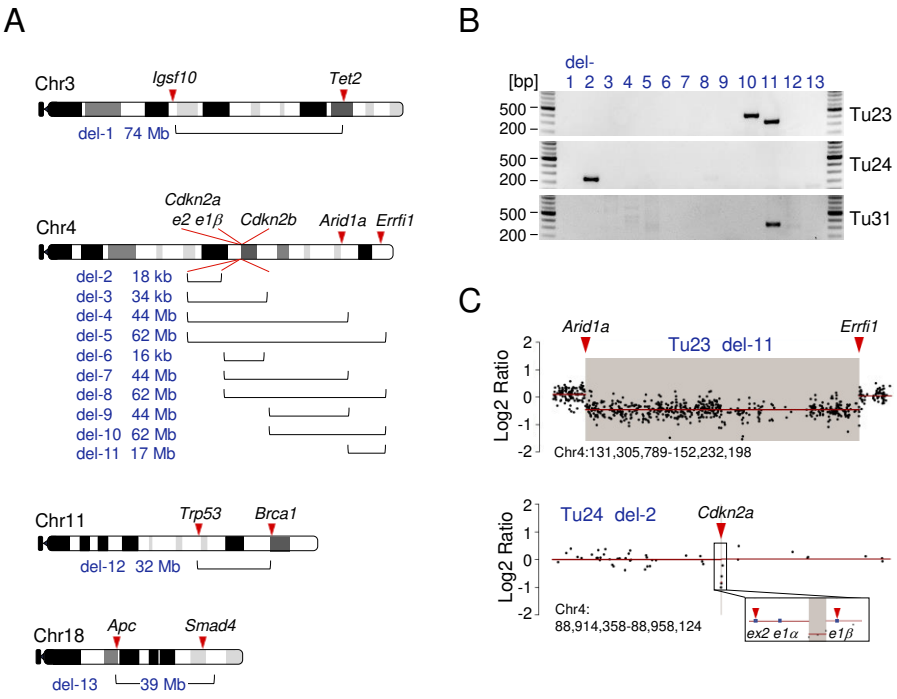

FIGURE 7-16. Induction of intra-chromosomal large deletions in the Cas9/18-sgRNA/*hSB5* cohort.

(**A**) Possible intra-chromosomal rearrangements in the 18-sgRNA cohort. (**B**) PCRs for detection of large deletions. (**C**) Large deletions detected by array comparative genomic hybridisation. Figure adapted from Weber *et al.*, 2015. Chr, chromosome; del, deletion.

Cell lines were generated from Tu23 and Tu24 and M-FISH was performed. M-FISH analysis showed a tetraploid stable chromosome set for both analysed cell lines: the composite karyotype of the cell line derived from Tu23 (Tu23-CL) was 77,XXXX,Del(4),-15,-17,-19 (Figure 7-17a) and Tu24-CL had an 80,XXXX karyotype (Figure 7-17b). For Tu23-CL, the CRISPR/Cas9 induced 62 Mb large deletion on chromosome 4 was clearly visible in one out of four chromosomes. Further analysis of additional metaphases of Tu23-CL confirmed three different states of chromosome 4 as already identified by PCR: (i) without any visible alterations (wild type chromosome), (ii) with the CRISPR/Cas induced 17-Mb-deletion between *Arid1a* and *Errfi1* and (iii) with the CRISPR/Cas induced 62-Mb-deletion between *Cdkn2b* and *Errfi1* (Figure 7-17c).

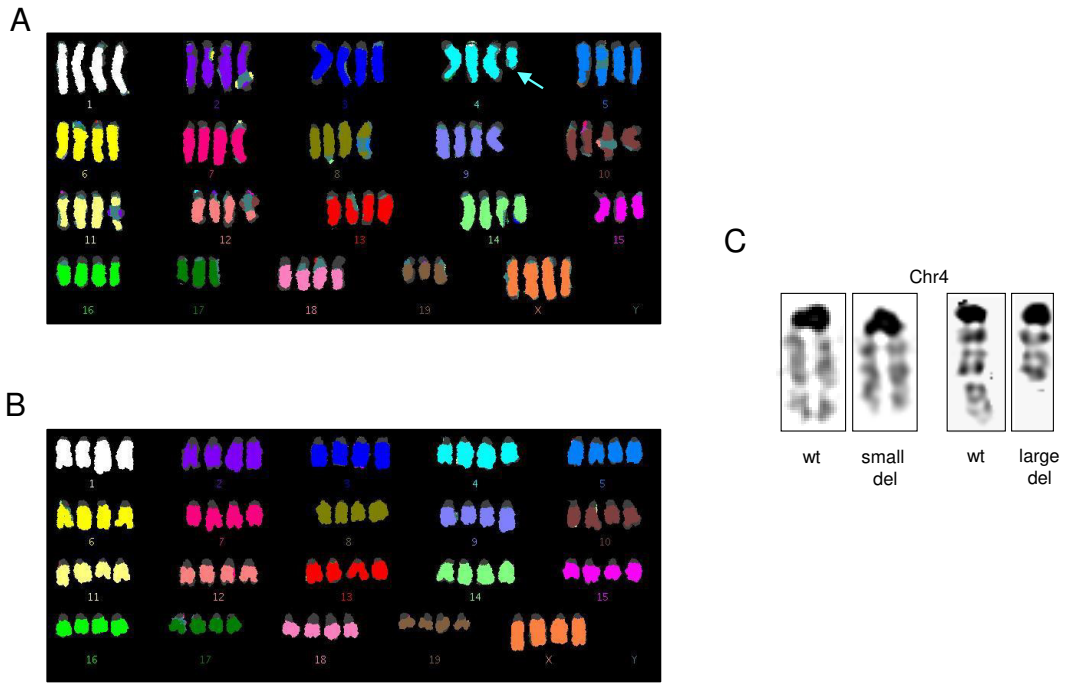

FIGURE 7-17. Multi-colour fluorescence *in situ* hybridisation of two liver tumour cell lines.

(**A**) Multi-colour fluorescence *in situ* hybridisation (M-FISH) analysis of Tu23-CL. (**B**) M-FISH analysis of Tu24-CL. (**C**) Analysis of different metaphases of chromosome 4 of Tu23-CL. Figure adapted from Weber *et al.*, 2015. Chr, chromosome; wt, wild type; del, deletion.

In addition, aCGH was performed for a sub-set of tumours to analyse further copy number alterations (e.g. large deletions). While the 62-Mb-deletion was not identified in the aCGH analysis (indicating its subclonal occurrence), the 18-kb-deletion in the *Cdkn2a* locus in Tu24 was detected as well as the 17-Mb-deletion between *Arid1a* and *Errfi1* (Figure 7-16c).

7.2.8 Three-fourth of all CRISPR/Cas induced indels are biallelic

To determine the incidence of mono- versus bi-allelic indel induction by CRISPR/Cas, two cancer cell lines (Tu23-CL and Tu24-CL), established from a mouse with an early onset ICC with multiple metastases, were analysed (Figure 7-18). For this, extensive geographical sampling of the primary tumour mass (n = 10) and of lymph node, lung and peritoneal metastases (n = 9) was performed (Figure 7-19a).

NGS revealed three different primary tumours (Tu22, Tu23 and Tu24) with Tu24 being the predominant clone (found in 8/10 samples) (Figure 7-19b). Analysis of the CRISPR/Cas

induced indel pattern allowed phylogenetic tracking of metastatic clones and showed that all metastases originate from the predominant tumour Tu24.

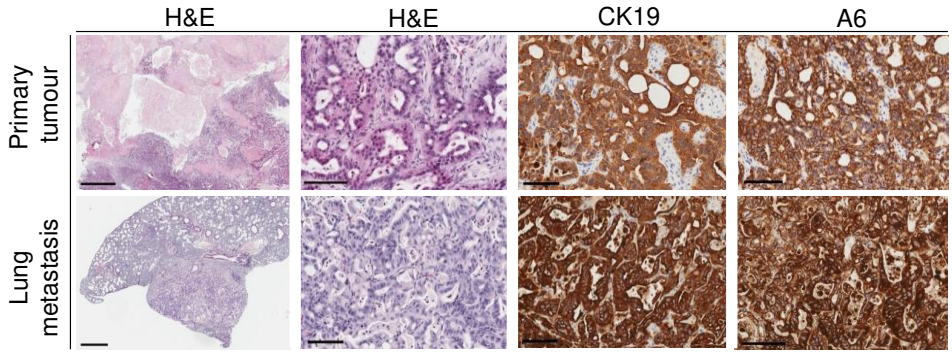

FIGURE 7-18. Histology of Tu24.

Moderately to poorly differentiated intrahepatic cholangiocarcinoma with extensive central necrosis (upper panels). Tumour cells strongly express cytokeratin 19 (CK19) and A6. Lower panels show sub-pleural lung metastasis with tubular growth pattern and intensive CK19 and A6 staining. Scale bars, 100 µm except upper left: 500 µm and lower left: 1 mm. Figure adapted from Weber *et al.*, 2015. H&E, haematoxylin and eosin.

While accurate MRF estimation in the tumour tissue might be difficult due to varying tumour cell/non-tumour cell content, cell line-based MRF analysis clearly demonstrated if indels in the respective target loci were hetero- or homozygous. An integrated quantitative analysis of (i) wild type read frequencies, (ii) indel frequencies and (iii) presence or absence of large deletions at the respective mutated target regions demonstrated that 79% of all mutated target loci were bi-allelically inactivated (Figure 7-19c and Table 7-6).

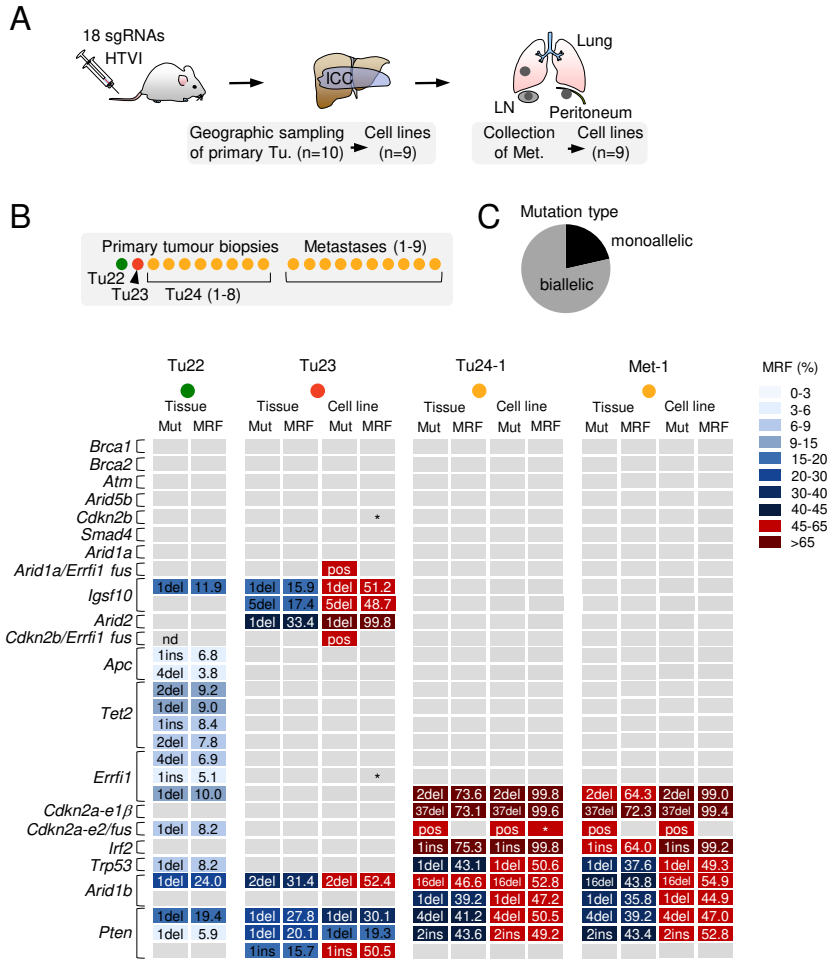

FIGURE 7-19. Mutant read frequency comparison in cell culture and tissue samples.

(**A**) Geographical sampling of an early onset intrahepatic cholangiocarcinoma (ICC) and of multiple metastases was performed. (**B**) Mutant read frequencies of three primary tumours (Tu22, Tu23 and Tu24-1) and one metastasis (Met-1) is shown. Tumours with indicated fusion products are marked as positive (pos). Asterisks indicate a lack of wild type sequence. Figure adapted from Weber *et al.*, 2015. Tu., tumour; Met., metastases; LN, lymph node; Mut, mutation; MRF, mutant read frequency.

Tumour	Target	Results of NGS and Deletion PCRs	Mono-allelic	Bi-allelic
	Arid1b	Indel 50% | wild type 50%	1	
	Arid2	Indel 100% | No wild type		1
	Igsf10	Indel-1 50% | Indel-2 50%		1
Tu23	*Pten*	Indel-1 28% | Indel-2 20% | Indel-3		1
	Arid1a	Large Deletion | Wild type	1	
	Cdkn2b	Large Deletion | No wild type		1
	Errfi1	Large Deletion | Large Deletion		1
	Trp53	Indel 50% | wild type 50%	1	
	Irf2	Indel 100% | No wild type		1
	Errfi1	Indel 100% | No wild type		1
Tu24	*Pten*	Indel-1 50% | Indel-2 50%		1
	Arid1b	Indel-1 50% | Indel-2 50%		1
	Cdkn2a-e1β	Large Deletion | Indel		1
	Cdkn2a-e2	Large Deletion | No wild type		1
		Sum (Percent)	3 (21%)	11 (79%)

TABLE 7-6. Mono- and bi-allelic indel induction.

TABLE 7-6 (continued)

Shown are the combined results of amplicon-based next-generation sequencing and deletion PCRs and mono- or bi-allelic allele status is noted. Table adapted from Weber *et al.*, 2015. NGS, next-generation sequencing.

7.2.9 CRISPR/Cas-induced liver tumours are transplantable

To show that CRISPR/Cas-induced liver tumours are transplantable, Tu23-CL and Tu24-CL were subcutaneously transplanted into NOD scid gamma mice. All mice developed tumours (n = 4 per cell line) up to 1 cm in diameter within two weeks after implantation (Figure 7-20a). For both cell lines, tumours display a solid growth pattern with multifocal necroses and infiltrate the adjacent adipose tissue (Figure 7-20b).

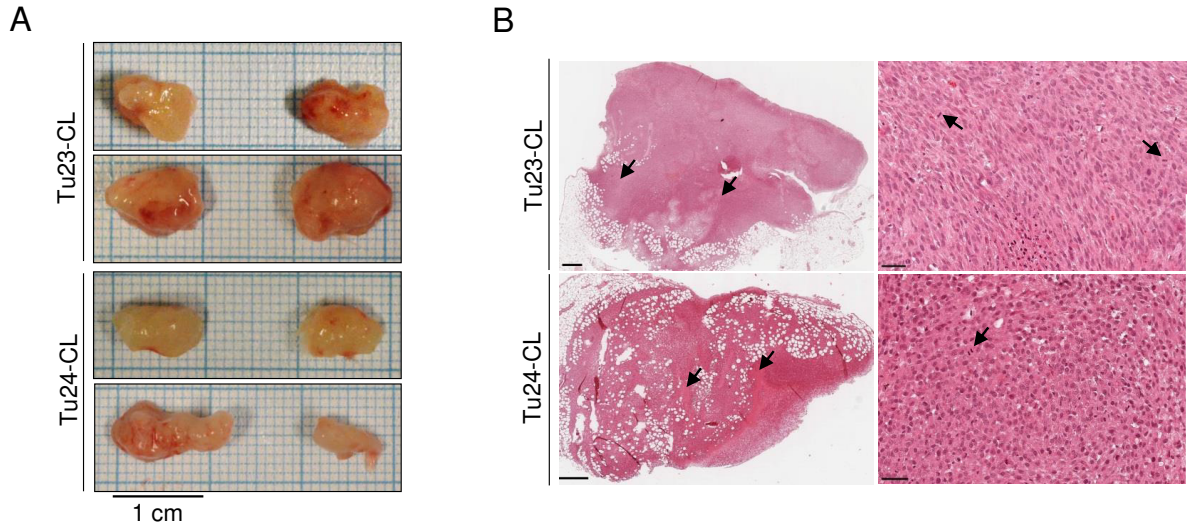

FIGURE 7-20. Subcutaneous implantation of tumour cell lines.

(**A**) Mice developed tumours up to 1 cm within 2 weeks post implantation. (**B**) Microscopic H&E images of the allograft tumours. Figure adapted from Weber *et al.*, 2015. Scale bars, 500 µm left panels, 50 µm right panels.

7.2.10 CRISPR/Cas somatic multiplex-mutagenesis does not induce detectable off-target effects

To screen for undesirable CRISPR/Cas-induced off-target effects, eight liver tumours (Tu6, Tu9, Tu12, Tu21, Tu23, Tu24, Tu32 and Tu35) were selected and analysed. For each sgRNA, the top five off-target regions (and if not already included in the top-five off-target list, at least three exonic off-targets) were sequenced. There were no indels at off-target sites with a mutant read frequency of 0.2% or higher (the same cut-off was used to exclude sequencing errors for on-target site analyses). Additionally, aCGH data of six tumours (Tu6, Tu9, Tu13, Tu23, Tu24 and Tu30) was screened for 266,778 potential intrachromosomal deletions, which result from combinations of potential off-target cleavage events (1,010 and 1,550 possible off-target sites for tumours of the Cas9/10-sgRNA/*hSB5* and Cas9/18-sgRNA/*hSB5* cohort, respectively). Off-target sites were defined to be potentially causative if they were within a distance of 500,000 bp (and 20 probes or fewer) to an aberration detected by aCGH. No chromosomal deletions, which were attributable to off-target effects, were identified.

8 DISCUSSION

Focus of this thesis was the implementation of two systems for TSG screening in mice: (i) transposon-based insertional mutagenesis (using *PB* and *SB* transposition tools) and (ii) CRISPR/Cas (exploiting the CRISPR/Cas9 system from *Streptococcus pyogenes*).

The transposon-based study demonstrated that utilization of a specifically designed "inactivating-only" transposon *(ITP2)* in a LOH-prone background (*Blm*-mutated mice) leads to induction of a broad spectrum of solid as well as haematopoietic tumours in mice. Furthermore, analysis of a tumour cohort (DLBCLs derived from IPB mice) revealed CISs in many known lymphoma TSGs but also in genes, for which a tumour suppressive function in haematologic cancers has not been described yet. Hence, this study provides the first description of a successful transposon-based recessive screen in mice. While several recessive *in vitro* screens have been conducted in *Blm*-mutated embryonic stem cells (Yusa *et al.*, 2004; Wang *et al.*, 2009), recessive *in vivo* screening using *Blm*-mutated mice has not been fruitful so far. For instance, implementation of a recessive retroviral screen in *Blm*-mutated mice mainly led to identification of oncogenes due to strong enhancers and promoters located within the retroviral LTRs that can act as gene-activating elements (Suzuki *et al.*, 2006). In contrast, ITP transposons only carry gene-trapping elements in form of polyadenylation signals and splice acceptors but harbour no elements that can enhance gene expression (although gene activation via truncation is possible). In addition, mobilisation of the *ITP2* transposon by both, the *PB* as well as the *SB* transposase, made it possible to conduct the first unbiased *in vivo* comparison of these transposon systems regarding activity, integration preferences and local hopping behaviour.

CRISPR/Cas is a novel system for genome engineering that is suitable for forward and reverse genetic screening approaches. While many research groups apply CRISPR/Cas tools to validate genes of interest (e.g. for cancer gene knockout in cell cultures systems or for generation of knockout animals), the vast majority of CRISPR/Cas-based screens has been conducted *in vitro* so far. This work provides one of the first studies that shows that CRISPR/Cas can be exploited to induce tumours somatically in adult mice. Moreover, the first proof-of-principle CRISPR/Cas-based genetic *in vivo* screen is presented here.

8.1 Comparison of the applicability of transposon- and CRISPR/Cas-based tools for forward genetic screening *in vivo*

Transposons and CRISPR/Cas are both exceptionally suitable tools for forward genetic screening *in vivo* although numerous differences apply for (i) library/screen design, (ii) identification of gene perturbations, (iii) delivery of components, (iv) suitability for tumour suppressor gene screening and (v) perturbation preferences/biases.

(I) Library/screen design

One major difference between transposon and CRISPR/Cas tools is that CRISPR/Cas screens rely on *a prior* knowledge for sgRNA design whereas transposons allow hypothesis-free genome-wide screening. This means that sgRNA libraries usually only target already known/annotated genes and/or regulatory elements (Sanjana *et al.*, 2014). Moreover, no genome-wide CRISPR/Cas screens targeting the whole coding and non-coding genome have been achieved *in vivo* so far. In contrast to this, transposon screens typically cover the whole genome, thus also allowing identification of novel protein-coding genes and regulatory elements, such as enhancers, silencers and non-coding RNAs. However, biases for intergration into loci harbouring certain genomic and epigenetic features are described (de Jong *et al.*, 2014).

CRISPR/Cas libraries are usually mono-functional, meaning that most often screens rely on either CRISPR/Cas activating or inactivating libraries though some designs for multiplexed CRISPR/Cas activation and repression libraries have been generated (Zalatan *et al.*, 2015). In contrast, the most commonly used transposon types (*SB*: *T2/Onc2* and *T2/Onc3*; *PB*: *ATP1* and *ATP2*) are bi-functional (Collier *et al.*, 2005; Dupuy *et al.*, 2005; Rad *et al.*, 2010) and thereby allow simultaneous gene activation and inactivation screens. However, for transposons, there are some biases regarding oncogene activation/TSG inactivation described, such as that in solid cancers *SB* transposons more likely identify TSGs than oncogenes whereas in haematopoietic tumours activating insertion patterns prevail (Mann *et al.*, 2015).

A potential advantage of the CRISPR/Cas system might be that design of sub-libraries targeting genes of interest (e.g. genes involved in certain cellular processes such as apoptosis (Horlbeck *et al.*, 2016) or with related functions like kinases (Wang *et al.*, 2014b)) is easily feasible. Moreover, CRISPR/Cas regional screening approaches to identify regulatory elements (Canver *et al.*, 2015) or relevant cancer genes within commonly altered regions in tumours are achievable although the possible induction of unwanted intrachromosomal rearrangements has to be carefully considered (Chapter 7.2.6). Notably, *SB* transposons,

which show a pronounced local hopping behaviour, are also suitable for regional mutagenesis, especially in the context of cancer gene discovery (unpublished data; Roland Rad, Technical University Munich, München, Germany).

(II) Identification of gene perturbations

"Tracking" of induced gene perturbations is also different between the two systems. While transposon insertion sites can be easily and rather inexpensively identified by specifically adapted sequencing protocols (Friedrich *et al.*, 2017), complex CRISPR/Cas libraries require genomic integration of the sgRNA or other barcode sequences to assign phenotypic effects to gene perturbations. Consequently, transfection-based approaches are ineligible for large-scale CRISPR/Cas libraries because the sgRNA sequence is usually not mobilised into the targeted genome, thus not allowing "tagging" of cells. Therefore, for large-scale or genome-wide CRISPR/Cas libraries, viral- or transposon-based delivery methods have to be considered.

(III) Delivery of components

To allow transposon screening *in vivo*, usually transgenic transposon mice are crossed with transposase knockin mice although recently some viral-based transposon screens have been performed (Molyneux *et al.*, 2014). In contrast, *in vivo* delivery of the CRISPR/Cas components poses a major challenge. While single sgRNAs or small scale sgRNA subpool libraries were delivered by transfection into the mouse liver (this study (Weber *et al.*, 2015) and Xue *et al.*, 2014) and pancreas (Maresch *et al.*, 2016) or via viral vectors into the lungs of mice (Maddalo *et al.*, 2014; Platt *et al.*, 2014; Sanchez-Rivera *et al.*, 2014), *in vivo* delivery of genome-wide sgRNA libraries has not been achieved so far. Furthermore, large-scale combinatorial screening, i.e. analysis of synergistic or antagonistic gene effects requiring the delivery of multiple sgRNAs into a single cell will be even more challening. Due to cargo size limitations, separation of Cas endonuclease and sgRNA expression cassettes is indispensable for many virus-based applications. For this reason, Cas9 knockin mouse lines (Platt *et al.*, 2014; Annunziato *et al.*, 2016; Weber *et al.*, 2019) are frequently used for *in vivo* experiments and only sgRNA cassettes are somatically delivered.

(IV) Suitability for tumour suppressor gene screening

Since most CRISPR/Cas-induced indels are biallelic (Chapter 7.2.8), the CRISPR/Cas system is particularly suited to screen for classical TSGs, which require biallelic inactivation. However, it is assumed that the vast majority of TSGs is haploinsufficient and the CRISPR/Cas

technology might be less powerful to identify those type of TSGs as their "full knockout" can have adverse effects. In contrast to this, it is statistically unlikely that two (or more) transposons integrate into the (at least) two alleles of a gene. Therefore, identification of classical TSGs, which requires biallelic inactivation, is hampered. However, it might be possible (at least partly) to overcome this problem by performing transposon screens in a LOH-prone background (Chapter 7.1.5).

(V) Perturbation preferences/biases

Both, transposons and CRISPR/Cas, exhibit perturbation biases to at least some extent. This means that perturbations do not only depend on the resulting selective pressure (e.g. influence on cellular fitness or proliferation) but also on other tool-intrinsic features. For CRISPR/Cas, the cleavage efficiency of different sgRNAs can be variable (Doench *et al.*, 2014) and, in particular, activating CRISPR libraries show severe efficiency biases, depending on the structures of the endogenous gene promoters (Konermann *et al.*, 2015). In contrast, transposons show pseudo-random insertion patterns and have integration biases for certain genomic and epigenetic features (open chromatin, transcription start sites, etc.) (de Jong *et al.*, 2014).

In conclusion, it is dependent on design and aim of a study if transposon mutagenesis or CRISPR/Cas is the preferable system as both tools have various context-specific advantageous and adverse characteristics. However, while transposons have proven to be a valuable screening tool and transposon-based screens have successfully been performed for decades, the validity of the CRISPR/Cas system for high-throughput genome-wide *in vivo* screening has yet to be shown.

8.2 Study of lymphomagenesis using transposon-based recessive genetic screening

8.2.1 ITP screen characteristics are comparable to bi-functional whole-body transposon screens

To compare the characteristics of the ITP screen with already published *PB* and *SB* whole-body transposon studies, embryonic lethality as well as tumour spectrum of triple-transgenic mouse lines was analysed, and marked differences between IcPB, IPB and ISB mice were revealed. The pronounced occurrence of embryonic lethality in the IcPB cohort can be at least

partly explained with the presence of a high transposon copy number (70 copies) in *ITP1-C* mice compared to medium-copy *ITP2-M* transposon animals (which harbour 35 copies). Transgenic *ITP1* mice were not crossed with *SB* transposase knockin animals as the *ITP1* transposon would have exceeded the cargo capacity of *SB* (the *bGEO* cassette alone has a size of 3.9 kb) and a tremendous drop in the transposition efficiency would have been expected (Copeland and Jenkins, 2010). There were also some differences between the IPB and ISB cohorts, with more triple-transgenic offspring being born in the latter group, which most likely reflects the distinct activity levels of the *PB* and *SB* transposase in the mouse genome (Liang *et al.*, 2009). When comparing the frequencies of embryonic lethality observed in the ITP screen with available data from *PB* and *SB* whole-body screens, similar embryonic lethality rates were identified in mice from the IPB cohort and in ATP mouse lines with a comparable copy number (e.g. *ATP1-H12* and *ATP2-H33*) (Rad *et al.*, 2010). Interestingly, frequently used *SB* mouse lines harbour up to 358 *T2/Onc2* transposon copies (Dupuy *et al.*, 2005) and drops in the numbers of viable offspring have been observed for high-copy number transposon lines (Dupuy *et al.*, 2005). However, embryonic lethality in *SB* transposon mice is less pronounced than in *PB* transposon mouse lines harbouring a similar transposon copy number (Rad *et al.*, 2010), which again might be attributed to the different transposition levels of *PB* and *SB* in the mouse genome.

Triple-transgenic IPB and ISB mice developed solid as well as haematopoietic tumours. The solid tumour spectrum was similar to those described in previously published *PB* and *SB* whole-body transposon studies (Collier *et al.*, 2005; Dupuy *et al.*, 2005; Rad *et al.*, 2010). In contrast, the observed range of haematopoietic cancers was quite distinct. Whereas earlier whole-body screens reported the development of either predominantly T-cell lymphomas (Dupuy *et al.*, 2005) or a very mixed spectrum with large numbers of myeloid cancers (Rad *et al.*, 2010), mice in the ITP screen mainly displayed B-cell lymphomas (Chapter 8.2.2).

8.2.2 Development of diffuse large B-cell lymphomas might be attributed to the Bloom-mutated background

Nearly all B-cell lymphomas derived from the ITP screen were diagnosed as DLBCLs. So far, all published B-cell-centric haematopoietic transposon screens focused on B-cell precursor acute lymphoblastic leukaemia (van der Weyden *et al.*, 2011; van der Weyden *et al.*, 2015). Additionally, no other large B-cell lymphoma/leukaemia cohort was systematically analysed in *PB/SB* whole-body screens until now. This study presents for the first time a comprehensive DLBCL cohort with detailed analysis of transposon insertion sites and CISs. Since *Blm*-

mutated control mice (without active transposition) also frequently develop DLBCL, it can be hypothesised that the *Blm*-mutated background is the major contributing factor to DLBCL occurrence. RecQ DNA helicase *Blm* is a helicase involved in DNA repair mechanisms of the cell. Defects in this enzyme lead to increased LOH rates and therefore mice harbouring two hypomorphic *Blm* alleles ($Blm^{m3/m3}$ mice) are tumour-prone (Luo *et al.*, 2000). The observed tumour spectrum included some solid cancers but predominantly lymphomas. However, an exact characterisation of the latter is mostly lacking so far. In a 2010 study, $Blm^{m3/m3}$ and $Blm^{m3/+}$ mice were irradiated (gamma irradiation; 4 Gy) and resulting tumours were analysed in detail (Warren *et al.*, 2010). Irradiated *Blm* mice, which frequently developed lymphomas, mainly showed T-cell lymphoblastic lymphomas (29%), DLBCLs (20%), B-cell lymphoblastic lymphomas (8.2%), follicular lymphomas (6.6%) and plasmacytomas (5.5%). More than 95% of all diagnosed haematopoietic tumours in the ITP control cohort were DLBCLs, with only one T-cell lymphoblastic lymphoma being identified. While some other B-cell lymphoma entities (B-cell lymphoblastic lymphoma, follicular lymphoma, marginal zone lymphoma of the spleen) were observed in the transgenic cohorts (especially in ISB mice), DLBCL was also the main lymphoma type in these animals, accounting for about 75% and 65% of all haematopoietic tumours in the IPB and ISB cohort, respectively. This indicates that the *Blm*-mutated background of IPB, ISB and control mice is most likely the main causative factor for DLBCL development and that transposition effects only play secondary roles in shifting the tumour spectrum towards DLBCL.

Similar to the tumour spectrum in mice, human patients with Bloom syndrome have a high risk to develop haematopoietic cancers, including B-cell (e.g. Burkitt lymphoma) and T-cell non-Hodgkin lymphoma (but not Hodgkin lymphoma) and acute myeloid leukaemia (Cunniff *et al.*, 2017), further demonstrating the similarity of phenotypes in human and murine Bloom deficiency.

8.2.3 DLBCLs show integrations in many known lymphoma genes and are enriched for lymphoma- and cancer-specific pathways

DLBCL samples derived from the IPB and ISB cohorts were sequenced and CIS as well as pathway enrichment analyses (the latter for IPB-DLBCL) were carried out. As expected and shown before (Rad *et al.*, 2010), CIMPL, compared to TAPDANCE, led to the identification of a larger number of CISs. The vast majority of CIS-containing genes detected by TAPDANCE analysis (86/94; 91.5%) was also identified when applying CIMPL analysis. Both methods, in addition to gCIS – a gene-centric CIS analysis procedure (Brett *et al.*, 2011) –, are commonly

and side-by-side used for CIS identification in *SB*- and *PB*-based transposon screens (van der Weyden *et al.*, 2013; Moriarity *et al.*, 2015; Takeda *et al.*, 2015).

Among the list of candidate genes were many TSGs with already implicated roles in DLBCL, such as *Gna13*, *Fas*, *Pten* and *Tnfrsf14*. *Gna13*, for example, was recently discovered to be involved in G protein-coupled receptor signalling, which mediates GC B-cell persistence (Muppidi *et al.*, 2014; Healy *et al.*, 2016), and is mutated in about 10% of all DLBCL cases. Furthermore, inactivating mutations in *Tnfrsf14* were identified in 20% of all DLBCLs (Lohr *et al.*, 2012), even though its exact function in lymphomagenesis is still unknown. Besides, also genes not yet directly associated with DLBCL development/maintenance but with well-studied roles in lymphomagenesis, like *Ankrd11* and *Tox* (Gonzalez-Aguilar *et al.*, 2012; Vasmatzis *et al.*, 2012), were frequently hit. Moreover, the ITP screen not only identified well-studied cancer/lymphoma genes, but also genes with completely unknown roles in tumourigenesis, such as *Erp44*, *Lnpep*, *Naa15* and *Wac*. The potential function of these genes in DLBCL could be validated in further experiments (i.e. functional *in vivo* and *in vitro* characterisation and cross-species analysis exploiting human DLBCL datasets) (Chapter 8.2.5).

Signalling pathways related to B-cell biology and cancer were among the most significantly enriched results derived from the Ingenuity® Pathway Analysis. The top enriched pathway was B-cell receptor signalling, a finding, which correlates well with human DLBCL genetics, for which constitutive activation of this pathway is a hallmark, particularly in the ABC-DLBCL subtype (Davis *et al.*, 2010). PI3K signalling in B-lymphocytes was the second most enriched signalling pathway playing a well-studied role in DLBCL (Kloo *et al.*, 2011; Pfeifer and Lenz, 2013). Moreover, rituximab, which is an integral component of the standard treatment regime for DLBCL – R-CHOP –, downregulates PI3K signalling (Suzuki *et al.*, 2007). Notably, T-cell receptor signalling and CD28 signalling in T helper cells were most likely present among the top enriched pathways due to the large overlap of genes involved in B-cell receptor signalling/PI3K signalling in B-lymphocytes as well as in T-cell receptor signalling/CD28 signalling in T helper cells.

8.2.4 *PiggyBac* and *Sleeping Beauty* are complementary tools because of their different characteristics

Several *in vitro* studies, mainly performed in mouse embryonic stem cells, revealed that *PB* and *SB* display dissimilar features (Liang *et al.*, 2009; Li *et al.*, 2013). However, *in vivo* data directly comparing the behaviour of *SB* and *PB* has been missing so far. This study provided the unique opportunity to perform an unbiased comparison of *PB* and *SB's* characteristics

because the identical transposon *(ITP2)* with the same copy number (35 copies) was mobilised from the one distinct genomic locus (chr14 E1-2.2). Furthermore, the identical library preparation and sequencing method (QiSeq) as well as analyses for identification of CISs (CIMPL and TAPDANCE) were carried out.

Clear differences in activity, insertion pattern and local hopping behaviour were observed. Unsurprisingly, there were – in tendency – more insertions in tumours derived from the IPB cohort than from the ISB mice since it is known that *PB* transposases exhibit in general a higher catalytic activity in the mouse genome compared to *SB* transposases. In this study, *iPBase* and *SB11* transposases catalysed transposon mobilisation, with the former being 250-fold more active in murine cells as determined by *in vitro* studies (Liang *et al.*, 2009). For both transposon systems, transposase versions with higher transposition rates exist: a mouse-codon optimised *(mPBase)* (Cadinanos and Bradley, 2007) and hyperactive variant *(hypPBase)* (Yusa *et al.*, 2011) for *PB* and the highly active *SB100* transposase (Mates *et al.*, 2009) for *SB*.

While it is known that *PB* exhibits a bias for actively transcribed genes, transcription start sites and open chromatin, most publications claim that *SB* insertions are more random and that there is no bias for intergenic or intragenic regions of the genome (Vigdal *et al.*, 2002; Liang *et al.*, 2009). This study showed that the distribution of transposon insertion sites is markedly different between *PB* and *SB* and that there are significantly more intergenic insertions in *SB*-derived tumours and tails compared to *PB*-derived samples. Based on the structure of the mouse genome, an unbiased distribution of transposon insertion sites would lead to about one-third of all insertions to be found within genes (Liang *et al.*, 2009). However, both transposon systems displayed significantly more intragenic insertions in tumour and tail DNA than expected by chance.

As predicated, due to the high selective pressure in cancer cells, percentages of intragenic *PB* and *SB* insertions were higher in tumour than in tail samples. Similarly, on the donor chromosome, which naturally shows a high transposon insertion density due to the local hopping behaviour of *SB* and, attenuated, *PB*, but less selective pressure exists, percentages of intragenic transposon insertions were smaller compared to the overall genome.

When comparing the local hopping behaviour of both transposons, the bias for donor chromosome insertions was, as expected, much more pronounced (three-fold higher) in the ISB than in the IPB cohort. For both transposon types, significantly more local hopping occurred in tail samples compared to tumour samples as there is less selective pressure being present in the former.

8.2.5 Future directions: further characterisation of murine diffuse large B-cell lymphomas and validation of candidate genes

Future steps of the transposon-based study will include a detailed characterisation of DLBCL subtypes as well as selection and further analysis of potential DLBCL candidate TSGs. For validation of the murine DLBCL samples, immunohistochemistry (e.g. *Bcl6* and *Irf4*), gene expression profiling and analysis of the B-cell receptor repertoire will be conducted. For the latter, immunoglobulin heavy chain rearrangements are determined by sequencing enabling clonality analysis of tumours. Examination of SHM occurrence within the immunoglobulin heavy chain region will further contribute to DLBCL subtype characterisation as ongoing SHM is distinctive for GCB-DLBCL (Lenz *et al.*, 2008b). To study the role of candidate genes in lymphomagenesis, several *in vitro* and *in vivo* experiments are being planned. For instance, *in vitro* studies exploiting the CRISPR/Cas system for knockout of TSGs in DLBCL cell lines (e.g. SuDHL4, SuDHL6, RIVA and HT) followed by gene expression profiling will help to elucidate gene functions. To add a second level of *in vivo* gene validation (the first being the ITP screen itself), transplantation-based approaches using the well-established Eμ-myc model (Bouchard *et al.*, 2007; Mills *et al.*, 2013) are currently being performed. Lastly, human DLBCL datasets (mutation, copy number variation, expression and clinical data) will be analysed for all potential candidate TSGs.

8.3 CRISPR/Cas-based recessive genetic screening in the mouse liver

8.3.1 Selection of tumour suppressor genes for forward genetic screening in the mouse liver

In the initial 10-sgRNA panel, seven TSGs with well-studied functions in ICC/HCC development (*Apc*, *Arid1a, Cdkn2a* encoding for $p16^{Ink4a}$ and $p19^{Arf}$, *Pten*, *Smad4* and *Trp53*), one gene unknown for its role in liver tumourigenesis *(Tet2)* and two TSGs, which are not involved in ICC/HCC tumourigenesis ("negative controls"; *Brca1* and *Brca2*), were selected. For the 18-sgRNA approach, eight additional TSGs with recently discovered roles in ICC/HCC were included (*Arid1b*, *Arid2*, *Arid5b*, *Atm*, *Cdkn2b*, *Errfi1*, *Igsf10* and *Irf2*). Among these, many genes are involved in chromatin remodelling (e.g. *Arid1b*, *Arid2* and *Arid5b*), a commonly altered epigenetic mechanism in liver cancer (Li *et al.*, 2011; Fujimoto *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013).

BRCA1 and *BRCA2* are part of a complex that is involved in DNA DSB repair by homologous recombination. Germline mutations in *BRCA1* and *BRCA2* are associated with the hereditary breast-ovarian cancer syndrome in which (female) patients have tremendously increased risks to develop breast and ovarian cancer but also other tumours (e.g. pancreatic cancer) (Welcsh and King, 2001). However, an elevated liver cancer risk for *BRCA1/BRCA2* mutation carriers has not been proven so far. Additionally, whole-genome/whole-exome sequencing studies only rarely identified mutations in *BRCA1* and *BRCA2* in ICC/HCC patients (Totoki *et al.*, 2011; Fujimoto *et al.*, 2012; Ong *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Schulze *et al.*, 2015).

TET2 is a putative downstream target of *IDH1* and *IDH2*, the latter being mutated in up to 10% of all ICC cases (Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Wang *et al.*, 2013). IDH1/IDH2's oncogenic function is linked to an abnormal enzymatic activity, which allows them to convert *α-ketoglutarate* to *2-hydroxyglutarate* leading to dioxygenase inhibition (Dang *et al.*, 2009; Ward *et al.*, 2010). Moreover, it was recently shown that mutant IDH blocks liver differentiation through suppression of HNF-4α, a key protein in hepatocyte identity and quiescence (Saha *et al.*, 2014). TET2 is a 2OH-dependent dioxygenase and a putative downstream target of IDH1/IDH2. *TET2* and *IDH1/2* mutations induce comparable hypermethylation phenotypes and are mutually exclusive in acute myeloid leukaemia (Guilhamon *et al.*, 2013; Wang *et al.*, 2013), which might suggest similar effects on cellular transformation (Figueroa *et al.*, 2010).

8.3.2 CRISPR/Cas somatic multiplex-mutagenesis induces liver tumours *in vivo*

While parallel work to this study showed that delivery of Cas9 and two sgRNAs directed against *Pten* and *Trp53* can induce liver tumours with bile duct differentiation features in CCl_4-treated wild type mice (Xue *et al.*, 2014), CRISPR/Cas multiplex-mutagenesis targeting a broad selection of TSGs has not been reported so far. In this study, the CRISPR/Cas system was utilised in the context of forward genetic screening. Therefore, as aforementioned (Chapter 8.3.1), genes with unknown function in liver cancer *(Tet2)*, putative TSGs, which were not yet validated *in vivo* (e.g. *Arid1a*, *Arid1b*, *Arid2*), well-known liver TSGs (e.g. *Pten*, *Smad4* and *Trp53*) and genes, which play no role in liver tumourigenesis (*Brca1* and *Brca2*), were selected. In addition, potential roles of these genes were studied in the context of liver-specific expression of oncogenic $Kras^{G12D}$ and in a CCl_4-mediated liver injury model. Mutations in *KRAS* are found in more than 10% of all ICC cases (Ong *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013) and *RAS/MAPK/ERK* pathway activation plays a crucial role in human HCC (Delire and Starkel, 2015). CCl_4 is a potent hepatotoxic agent, which can lead to lipid peroxidation and membrane damage in liver cells caused by formation of trichloromethyl radicals after metabolisation (Boll *et al.*, 2001). Furthermore, an inflammatory response triggered by Kupffer cells resulting in secretion of cytokines, chemokines and other factors and attraction of different immune cells (e.g. monocytes, neutrophils and lymphocytes), can also lead to liver tissue damage (Luckey and Petersen, 2001; Liedtke *et al.*, 2013).

While in the $Kras^{G12D}$ background ICCs and HCCs were induced, only HCCs were observed in the CCl_4-treated cohort. In the Kras group, the development of both, ICC and HCC, was expected as activation of oncogenic $Kras^{G12D}$ and *RAS/MAPK/ERK* signalling are central events in hepatic tumourigenesis of both lineages. For CCl_4, it is known that treatment can induce – in a dose-dependent manner – liver damage (lower dosages) and HCC (higher dosages) (Liedtke *et al.*, 2013). Development of ICC following CCl_4-treatment was not described so far in the literature, which is consistent with the induction of only HCCs (but not ICCs) in this study.

8.3.3 CRISPR/Cas multiplex-mutagenesis is suitable for forward genetic *in vivo* screening in the mouse liver

Analysis of liver tumours derived from the Cas9/10-sgRNA/*hSB5* cohort showed that the global indel distribution pattern is not random, meaning that some target regions are significantly more often hit than others. All sgRNAs were tested before their *in vivo* application *in vitro* using Surveyor nuclease assays, which are considered to be the gold standard for indel efficiency testing (Ran *et al.*, 2013), and showed comparable cleavage efficiencies. Therefore, the uneven indel distribution in the liver tumours suggests that indels/mutations in genes, which are biologically relevant for liver tumourigenesis, undergo positive selection *in vivo*.

Strikingly, *Pten* was mutated in all 21 analysed hepatic tumours underlying its crucial function in ICC/HCC development and the importance of PI3K signalling in hepatic tumourigenesis (Chen *et al.*, 2011). *Tet2* was mutated in all six analysed ICCs but only in one out of four HCCs. This indicates that while alterations in *TET2* are significant in ICC development, its role in HCC might be less pivotal. This is consistent with the high prevalence of mutations in *TET2*'s upstream players *IDH1/IDH2* in ICC but not in HCC (Chan-On *et al.*, 2013; Jiao *et al.*, 2013; Wang *et al.*, 2013).

Positive selection for cancer-relevant mutations was likewise observed when comparing the indel efficiencies of the *Cdkn2a-e1β* and *Cdkn2a-e2* sgRNAs in ICCs/HCCs. Both gene products from the *Cdkn2a* locus, $p16^{Ink4a}$ and $p19^{Arf}$, have important tumour suppressive functions in the liver. While the *Cdkn2a-e1β* sgRNA can only inactivate $p19^{Arf}$, the *Cdkn2a-e2* sgRNA disrupts both, $p16^{Ink4a}$ as well as $p19^{Arf}$. $p16^{INK4A}$ mainly promotes senescence in interaction with *RB1* (Ohtani *et al.*, 2004) whereas $p19^{ARF}$ induces cell cycle arrest in a p53-dependent manner (Ozenne *et al.*, 2010) (though it also exhibits some functions independent of *TP53* (Weber *et al.*, 2000)). As expected, there was a positive selection for indels in the *Cdkn2a-e2* target locus to occur leading to inactivation of both TSGs. Consequently, no selective pressure was present for induction of mutations in the *Cdkn2a-e1β* target region, and, consistent with this, indels at this target site were only rarely identified.

Additionally – as also observed in the 18-sgRNA cohort – mutations in chromatin remodelling genes/proteins were abundant, as indicated by frequent indel occurrence in the *Arid1a*, *Arid1b, Arid2* and *Arid5b* target regions. This demonstrates, as stated earlier, the substantial involvement of this class of genes in ICC/HCC tumourigenesis (Li *et al.*, 2011; Fujimoto *et al.*, 2012; Chan-On *et al.*, 2013; Jiao *et al.*, 2013). In marked contrast to this, *Brca1* and *Brca2* were only very infrequently hit, reflecting their minor roles in hepatic tumourigenesis.

8.3.4 CRISPR/Cas somatic multiplex-mutagenesis can induce intra-tumour heterogeneity and intrachromosomal rearrangements

Some CRISPR/Cas-induced liver tumours show signs of intra-tumour heterogeneity as indicated by varying indel frequencies detected at different target regions in individual cancers. Moreover, frequently more than two mutations existed at distinct target sites within a tumour. One possible explanation for this might be that while some indels already occur in the initially transfected tumour cell clone (founder cell), other indels are induced after division into subsequent daughter cells. To examine this further, one of the tumours (Tu1) was microdissected (regions R2 and R3) and indel sequencing was performed. However, even within the microdissected regions, indel frequencies were unexpectedly low and the subclonal structure of the tumour could only be partially resolved, which might indicate the presence of further intraregional minority clones.

In general, MRFs were determined from mouse tissue, which can be inaccurate due to contamination of tumour tissue with stromal cells. Therefore, for some cancers, primary cell cultures were established allowing precise indel frequency determination. However, generation of primary cell cultures was exclusively possible from mouse ICCs (and here only from larger tumours) but not from murine HCCs as cells derived from the latter do not grow well under regular cell culture conditions.

Several studies showed that CRISPR/Cas could induce intra- and interchromosomal rearrangements in cells *in vitro* and *in vivo*. In one of the first studies, He *et al*. demonstrated that deletions reaching up to 1 Mb in size can be generated in cell culture systems (He *et al.*, 2015). Later publications also showed induction of large-scale deletions *in vivo* (Li *et al*., 2015). While in this study only intrachromosomal rearrangements were observed, a similar CRISPR/Cas-based project, aiming to model and analyse pancreatic cancer (Maresch *et al.*, 2016), also revealed the (albeit infrequent) occurrence of CRISPR/Cas-induced interchromosomal rearrangements (translocations). Maresch *et al.* utilised, comparable to this work, a transfection-based sgRNA delivery approach targeting the mouse pancreas with twelve sgRNAs (which are – apart from two “neutral” sgRNAs directed against the *Rosa26* locus – a sub-set derived from the 18-sgRNA cohort of this study). Although in both studies the number of mutated target sites per single cell was similar, incidence of CRISPR/Cas-triggered structural rearrangements was higher in the pancreas-centric study. This might reflect tissue-specific grades of susceptibility for chromosomal aberrations. Overall, both studies illustrate that CRISPR/Cas multiplexing can lead to induction of chromosomal rearrangements *in vivo*, which on one side can be exploited to model cancer-associated

structural aberrations (e.g. Blasco *et al.*, 2014; Maddalo *et al.*, 2014), but on the other side has to be carefully considered in order to avoid undesirable off-target effects.

8.3.5 The CRISPR/Cas system only has no off-target effects *in vivo* at predicted off-target sites

No off-target effects (indels as well as large intra- and interchromosomal deletions) were observed in eight tested hepatic tumours. It is known that particularly *in vitro* systems, which are based on transduction-based approaches (with ongoing Cas9 expression), often exhibit indels at predicted off-target sites (Zetsche *et al.*, 2015b) and up to five mismatches in the sgRNA sequence are tolerated (Hsu *et al.*, 2013; Mali *et al.*, 2013a; Pattanayak *et al.*, 2013). However, other studies which performed *in vivo* approaches comparable to this study also identified very few to no off-target mutations (Sanchez-Rivera *et al.*, 2014; Xue *et al.*, 2014), even so some of these experiments were based on stable Cas9 expression (Sanchez-Rivera *et al.*, 2014). This indicates that there might be different off-target susceptibilities in *in vitro* and *in vivo* settings although the underlying mechanism remains unknown.

8.3.6 Future directions: Library expansion and implementation of additional mouse models

This work demonstrated – in form of a proof-of-principle study – that CRISPR/Cas somatic multiplex-mutagenesis could be exploited for recessive genetic screening *in vivo* since selection of cancer-relevant indels occurs. Further expansions of this approach might include larger TSG libraries but also specific sub-libraries (for example, based on gene functions, gene domains or genomic locations). Moreover, liver-specific CRISPR/Cas multiplex-mutagenesis studies using other experimental settings (e.g. viral-based delivery methods) and additional murine backgrounds (i.e. predisposing genetic mutations and/or chemical induction models) could be carried out. For the latter, already known ICC/HCC models (Chapter 3.5.3) might be harnessed, which could facilitate in-depth analysis of yet unidentified gene-cooperative effects.

9 REFERENCES

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., Boyault, S., Burkhardt, B., Butler, A.P., Caldas, C., Davies, H.R., Desmedt, C., Eils, R., Eyfjord, J.E., Foekens, J.A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinski, M., Jager, N., Jones, D.T., Jones, D., Knappskog, S., Kool, M., Lakhani, S.R., Lopez-Otin, C., Martin, S., Munshi, N.C., Nakamura, H., Northcott, P.A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J.V., Puente, X.S., Raine, K., Ramakrishna, M., Richardson, A.L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T.N., Span, P.N., Teague, J.W., Totoki, Y., Tutt, A.N., Valdes-Mas, R., van Buuren, M.M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L.R., Australian Pancreatic Cancer Genome, I., Consortium, I.B.C., Consortium, I.M.-S., PedBrain, I., Zucman-Rossi, J., Futreal, P.A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S.M., Siebert, R., Campo, E., Shibata, T., Pfister, S.M., Campbell, P.J., and Stratton, M.R. (2013). Signatures of mutational processes in human cancer. *Nature* **500**, 415-421.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., and Staudt, L.M. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511.

Annunziato, S., Kas, S.M., Nethe, M., Yucel, H., Del Bravo, J., Pritchard, C., Bin Ali, R., van Gerwen, B., Siteur, B., Drenth, A.P., Schut, E., van de Ven, M., Boelens, M.C., Klarenbeek, S., Huijbers, I.J., van Miltenburg, M.H., and Jonkers, J. (2016). Modeling invasive lobular breast carcinoma by CRISPR/Cas9-mediated somatic genome editing of the mammary gland. *Genes Dev* **30**, 1470-1480.

Bard-Chapeau, E.A., Nguyen, A.T., Rust, A.G., Sayadi, A., Lee, P., Chua, B.Q., New, L.S., de Jong, J., Ward, J.M., Chin, C.K., Chew, V., Toh, H.C., Abastado, J.P., Benoukraf, T., Soong, R., Bard, F.A., Dupuy, A.J., Johnson, R.L., Radda, G.K., Chan, E.C., Wessels, L.F., Adams, D.J., Jenkins, N.A., and Copeland, N.G. (2014). Transposon mutagenesis identifies genes driving hepatocellular carcinoma in a chronic hepatitis B mouse model. *Nat Genet* **46**, 24-32.

Bettermann, K., Vucur, M., Haybaeck, J., Koppe, C., Janssen, J., Heymann, F., Weber, A., Weiskirchen, R., Liedtke, C., Gassler, N., Muller, M., de Vos, R., Wolf, M.J., Boege, Y., Seleznik, G.M., Zeller, N., Erny, D., Fuchs, T., Zoller, S., Cairo, S., Buendia, M.A., Prinz, M., Akira, S., Tacke, F., Heikenwalder, M., Trautwein, C., and Luedde, T. (2010). TAK1 suppresses a NEMO-dependent but NF-kappaB-independent pathway to liver cancer. *Cancer Cell* **17**, 481-496.

Bignell, G.R., Greenman, C.D., Davies, H., Butler, A.P., Edkins, S., Andrews, J.M., Buck, G., Chen, L., Beare, D., Latimer, C., Widaa, S., Hinton, J., Fahey, C., Fu, B., Swamy, S., Dalgliesh, G.L., Teh, B.T., Deloukas, P., Yang, F., Campbell, P.J., Futreal, P.A., and Stratton, M.R. (2010). Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893-898.

Blasco, R.B., Karaca, E., Ambrogio, C., Cheong, T.C., Karayol, E., Minero, V.G., Voena, C., and Chiarle, R. (2014). Simple and rapid in vivo generation of chromosomal rearrangements using CRISPR/Cas9 technology. *Cell Rep* **9**, 1219-1227.

Boll, M., Weber, L.W., Becker, E., and Stampfl, A. (2001). Mechanism of carbon tetrachloride-induced hepatotoxicity. Hepatocellular damage by reactive carbon tetrachloride metabolites. *Z Naturforsch C* **56**, 649-659.

Borlak, J., Meier, T., Halter, R., Spanel, R., and Spanel-Borowski, K. (2005). Epidermal growth factor-induced hepatocellular carcinoma: gene expression profiles in precursor lesions, early stage and solitary tumours. *Oncogene* **24**, 1809-1819.

Bouchard, C., Lee, S., Paulus-Hock, V., Loddenkemper, C., Eilers, M., and Schmitt, C.A. (2007). FoxO transcription factors suppress Myc-driven lymphomagenesis via direct activation of Arf. *Genes Dev* **21**, 2775-2787.

Braren, R., Altomonte, J., Settles, M., Neff, F., Esposito, I., Ebert, O., Schwaiger, M., Rummeny, E., and Steingoetter, A. (2011). Validation of preclinical multiparametric imaging for prediction of necrosis in hepatocellular carcinoma after embolization. *J Hepatol* **55**, 1034-1040.

Brett, B.T., Berquam-Vrieze, K.E., Nannapaneni, K., Huang, J., Scheetz, T.E., and Dupuy, A.J. (2011). Novel molecular and computational methods improve the accuracy of insertion site analysis in Sleeping Beauty-induced tumors. *PLoS One* **6**, e24668.

Brookfield, J.F. (2005). The ecology of the genome - mobile DNA elements and their hosts. *Nat Rev Genet* **6**, 128-136.

Cadinanos, J., and Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res* **35**, e87.

Calado, D.P., Zhang, B., Srinivasan, L., Sasaki, Y., Seagal, J., Unitt, C., Rodig, S., Kutok, J., Tarakhovsky, A., Schmidt-Supprian, M., and Rajewsky, K. (2010). Constitutive canonical NF-kappaB activation cooperates with disruption of BLIMP1 in the pathogenesis of activated B cell-like diffuse large cell lymphoma. *Cancer Cell* **18**, 580-589.

Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C., Teague, J.W., Menzies, A., Goodhead, I., Turner, D.J., Clee, C.M., Quail, M.A., Cox, A., Brown, C., Durbin, R., Hurles, M.E., Edwards, P.A., Bignell, G.R., Stratton, M.R., and Futreal, P.A. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* **40**, 722-729.

Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., Luc, S., Kurita, R., Nakamura, Y., Fujiwara, Y., Maeda, T., Yuan, G.C.,

Zhang, F., Orkin, S.H., and Bauer, D.E. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-197.

Cattoretti, G., Pasqualucci, L., Ballon, G., Tam, W., Nandula, S.V., Shen, Q., Mo, T., Murty, V.V., and Dalla-Favera, R. (2005). Deregulated BCL6 expression recapitulates the pathogenesis of human diffuse large B cell lymphomas in mice. *Cancer Cell* **7**, 445-455.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., and Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404.

Challa-Malladi, M., Lieu, Y.K., Califano, O., Holmes, A.B., Bhagat, G., Murty, V.V., Dominguez-Sola, D., Pasqualucci, L., and Dalla-Favera, R. (2011). Combined genetic inactivation of beta2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. *Cancer Cell* **20**, 728-740.

Chan-On, W., Nairismagi, M.L., Ong, C.K., Lim, W.K., Dima, S., Pairojkul, C., Lim, K.H., McPherson, J.R., Cutcutache, I., Heng, H.L., Ooi, L., Chung, A., Chow, P., Cheow, P.C., Lee, S.Y., Choo, S.P., Tan, I.B., Duda, D., Nastase, A., Myint, S.S., Wong, B.H., Gan, A., Rajasegaran, V., Ng, C.C., Nagarajan, S., Jusakul, A., Zhang, S., Vohra, P., Yu, W., Huang, D., Sithithaworn, P., Yongvanit, P., Wongkham, S., Khuntikeo, N., Bhudhisawasdi, V., Popescu, I., Rozen, S.G., Tan, P., and Teh, B.T. (2013). Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. *Nat Genet* **45**, 1474-1478.

Chen, K.F., Chen, H.L., Tai, W.T., Feng, W.C., Hsu, C.H., Chen, P.J., and Cheng, A.L. (2011). Activation of phosphatidylinositol 3-kinase/Akt signaling pathway mediates acquired resistance to sorafenib in hepatocellular carcinoma cells. *J Pharmacol Exp Ther* **337**, 155-161.

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

Chu, G. (1997). Double strand break repair. *J Biol Chem* **272**, 24097-24100.

Cleary, S.P., Jeck, W.R., Zhao, X., Chen, K., Selitsky, S.R., Savich, G.L., Tan, T.X., Wu, M.C., Getz, G., Lawrence, M.S., Parker, J.S., Li, J., Powers, S., Kim, H., Fischer, S., Guindi, M., Ghanekar, A., and Chiang, D.Y. (2013). Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology* **58**, 1693-1702.

Coiffier, B., Thieblemont, C., Van Den Neste, E., Lepeu, G., Plantier, I., Castaigne, S., Lefort, S., Marit, G., Macro, M., Sebban, C., Belhadj, K., Bordessoule, D., Ferme, C., and Tilly, H. (2010). Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood* **116**, 2040-2045.

Collier, L.S., Carlson, C.M., Ravimohan, S., Dupuy, A.J., and Largaespada, D.A. (2005). Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**, 272-276.

Compagno, M., Lim, W.K., Grunn, A., Nandula, S.V., Brahmachary, M., Shen, Q., Bertoni, F., Ponzoni, M., Scandurra, M., Califano, A., Bhagat, G., Chadburn, A., Dalla-Favera, R., and Pasqualucci, L. (2009). Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**, 717-721.

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

Copeland, N.G., and Jenkins, N.A. (2010). Harnessing transposons for cancer gene discovery. *Nat Rev Cancer* **10**, 696-706.

Cunniff, C., Bassetti, J.A., and Ellis, N.A. (2017). Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. *Mol Syndromol* **8**, 4-23.

Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., Marks, K.M., Prins, R.M., Ward, P.S., Yen, K.E., Liau, L.M., Rabinowitz, J.D., Cantley, L.C., Thompson, C.B., Vander Heiden, M.G., and Su, S.M. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744.

Davis, R.E., Ngo, V.N., Lenz, G., Tolar, P., Young, R.M., Romesser, P.B., Kohlhammer, H., Lamy, L., Zhao, H., Yang, Y., Xu, W., Shaffer, A.L., Wright, G., Xiao, W., Powell, J., Jiang, J.K., Thomas, C.J., Rosenwald, A., Ott, G., Muller-Hermelink, H.K., Gascoyne, R.D., Connors, J.M., Johnson, N.A., Rimsza, L.M., Campo, E., Jaffe, E.S., Wilson, W.H., Delabie, J., Smeland, E.B., Fisher, R.I., Braziel, R.M., Tubbs, R.R., Cook, J.R., Weisenburger, D.D., Chan, W.C., Pierce, S.K., and Staudt, L.M. (2010). Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88-92.

de Jong, J., Akhtar, W., Badhai, J., Rust, A.G., Rad, R., Hilkens, J., Berns, A., van Lohuizen, M., Wessels, L.F., and de Ridder, J. (2014). Chromatin landscapes of retroviral and transposon integration profiles. *PLoS Genet* **10**, e1004250.

de Miranda, N.F., Georgiou, K., Chen, L., Wu, C., Gao, Z., Zaravinos, A., Lisboa, S., Enblad, G., Teixeira, M.R., Zeng, Y., Peng, R., and Pan-Hammarstrom, Q. (2014). Exome sequencing reveals novel mutation targets in diffuse large B-cell lymphomas derived from Chinese patients. *Blood* **124**, 2544-2553.

de Ridder, J., Uren, A., Kool, J., Reinders, M., and Wessels, L. (2006). Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. *PLoS Comput Biol* **2**, e166.

Delire, B., and Starkel, P. (2015). The Ras/MAPK pathway and hepatocarcinoma: pathogenesis and therapeutic implications. *Eur J Clin Invest* **45**, 609-623.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473-483.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32**, 1262-1267.

Dominguez, A.A., Lim, W.A., and Qi, L.S. (2016). Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol* **17**, 5-15.

Dunleavy, K., Pittaluga, S., Maeda, L.S., Advani, R., Chen, C.C., Hessler, J., Steinberg, S.M., Grant, C., Wright, G., Varma, G., Staudt, L.M., Jaffe, E.S., and Wilson, W.H. (2013). Dose-adjusted EPOCH-rituximab therapy in primary mediastinal B-cell lymphoma. *N Engl J Med* **368**, 1408-1416.

Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., and Jenkins, N.A. (2005). Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**, 221-226.

Dupuy, A.J., Rogers, L.M., Kim, J., Nannapaneni, K., Starr, T.K., Liu, P., Largaespada, D.A., Scheetz, T.E., Jenkins, N.A., and Copeland, N.G. (2009). A modified sleeping beauty transposon system that can be used to model a wide variety of human cancers in mice. *Cancer Res* **69**, 8150-8156.

Ellis, N.A., Lennon, D.J., Proytcheva, M., Alhadeff, B., Henderson, E.E., and German, J. (1995). Somatic intragenic recombination within the mutated locus BLM can correct the high sister-chromatid exchange phenotype of Bloom syndrome cells. *Am J Hum Genet* **57**, 1019-1027.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**, E359-386.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., Tallman, M.S., Sun, Z., Wolniak, K., Peeters, J.K., Liu, W., Choe, S.E., Fantin, V.R., Paietta, E., Lowenberg, B., Licht, J.D., Godley, L.A., Delwel, R., Valk, P.J., Thompson, C.B., Levine, R.L., and Melnick, A. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair haematopoietic differentiation. *Cancer Cell* **18**, 553-567.

Fletcher, O., and Houlston, R.S. (2010). Architecture of inherited susceptibility to common cancer. *Nat Rev Cancer* **10**, 353-361.

Forbes, S.A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J.W., Futreal, P.A., and Stratton, M.R. (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* **Chapter 10**, Unit 10 11.

Forner, A., Llovet, J.M., and Bruix, J. (2012). Hepatocellular carcinoma. *Lancet* **379**, 1245-1255.

Fraser, M.J., Ciszczon, T., Elick, T., and Bauser, C. (1996). Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol Biol* **5**, 141-151.

Friedberg, J.W. (2011). Relapsed/refractory diffuse large B-cell lymphoma. *Hematology Am Soc Hematol Educ Program* **2011**, 498-505.

Friedel, R.H., Friedel, C.C., Bonfert, T., Shi, R., Rad, R., and Soriano, P. (2013). Clonal expansion analysis of transposon insertions by high-throughput sequencing identifies candidate cancer genes in a PiggyBac mutagenesis screen. *PLoS One* **8**, e72338.

Friedrich, M.J., Rad, L., Bronner, I.F., Strong, A., Wang, W., Weber, J., Mayho, M., Ponstingl, H., Engleitner, T., Grove, C., Pfaus, A., Saur, D., Cadinanos, J., Quail, M.A., Vassiliou, G.S., Liu, P., Bradley, A., and Rad, R. (2017). Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nat Protoc* **12**, 289-309.

Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K.A., Hosoda, F., Nguyen, H.H., Aoki, M., Hosono, N., Kubo, M., Miya, F., Arai, Y., Takahashi, H., Shirakihara, T., Nagasaki, M., Shibuya, T., Nakano, K., Watanabe-Makino, K., Tanaka, H., Nakamura, H., Kusuda, J., Ojima, H., Shimada, K., Okusaka, T., Ueno, M., Shigekawa, Y., Kawakami, Y., Arihiro, K., Ohdan, H., Gotoh, K., Ishikawa, O., Ariizumi, S., Yamamoto, M., Yamada, T., Chayama, K., Kosuge, T., Yamaue, H., Kamatani, N., Miyano, S., Nakagama, H., Nakamura, Y., Tsunoda, T., Shibata, T., and Nakagawa, H. (2012). Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* **44**, 760-764.

Fukai, K., Yokosuka, O., Imazeki, F., Tada, M., Mikata, R., Miyazaki, M., Ochiai, T., and Saisho, H. (2005). Methylation status of p14ARF, p15INK4b, and p16INK4a genes in human hepatocellular carcinoma. *Liver Int* **25**, 1209-1216.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., and Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-2586.

Genovesi, L.A., Ng, C.G., Davis, M.J., Remke, M., Taylor, M.D., Adams, D.J., Rust, A.G., Ward, J.M., Ban, K.H., Jenkins, N.A., Copeland, N.G., and Wainwright, B.J. (2013). Sleeping Beauty mutagenesis in a mouse medulloblastoma model defines networks that discriminate between human molecular subgroups. *Proc Natl Acad Sci U S A* **110**, E4325-4334.

Goeppert, B., Konermann, C., Schmidt, C.R., Bogatyrova, O., Geiselhart, L., Ernst, C., Gu, L., Becker, N., Zucknick, M., Mehrabi, A., Hafezi, M., Klauschen, F., Stenzinger, A., Warth, A., Breuhahn, K., Renner, M., Weichert, W., Schirmacher, P., Plass, C., and Weichenhan, D. (2014). Global alterations of DNA methylation in cholangiocarcinoma target the Wnt signaling pathway. *Hepatology* **59**, 544-554.

Gonzalez-Aguilar, A., Idbaih, A., Boisselier, B., Habbita, N., Rossetto, M., Laurenge, A., Bruno, A., Jouvet, A., Polivka, M., Adam, C., Figarella-Branger, D., Miquel, C., Vital, A., Ghesquieres, H., Gressin, R., Delwail, V., Taillandier, L., Chinot, O., Soubeyran, P., Gyan, E., Choquet, S., Houillier, C., Soussain, C., Tanguy, M.L., Marie, Y., Mokhtari, K., and Hoang-Xuan, K. (2012). Recurrent mutations of MYD88 and TBL1XR1 in primary central nervous system lymphomas. *Clin Cancer Res* **18**, 5203-5211.

Gregory, T.R. (2005). Synergy between sequence and size in large-scale genomics. *Nat Rev Genet* **6**, 699-708.

Guichard, C., Amaddeo, G., Imbeaud, S., Ladeiro, Y., Pelletier, L., Maad, I.B., Calderaro, J., Bioulac-Sage, P., Letexier, M., Degos, F., Clement, B., Balabaud, C., Chevet, E., Laurent, A., Couchy, G., Letouze, E., Calvo, F., and Zucman-Rossi, J. (2012). Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet* **44**, 694-698.

Guilhamon, P., Eskandarpour, M., Halai, D., Wilson, G.A., Feber, A., Teschendorff, A.E., Gomez, V., Hergovich, A., Tirabosco, R., Fernanda Amary, M., Baumhoer, D., Jundt, G., Ross, M.T., Flanagan, A.M., and Beck, S. (2013). Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. *Nat Commun* **4**, 2166.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* **100**, 57-70.

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

He, Z., Proudfoot, C., Mileham, A.J., McLaren, D.G., Whitelaw, C.B., and Lillico, S.G. (2015). Highly efficient targeted chromosome deletions using CRISPR/Cas9. *Biotechnol Bioeng* **112**, 1060-1064.

Healy, J.A., Nugent, A., Rempel, R.E., Moffitt, A.B., Davis, N.S., Jiang, X., Shingleton, J.R., Zhang, J., Love, C., Datta, J., McKinney, M.E., Tzeng, T.J., Wettschureck, N., Offermanns, S., Walzer, K.A., Chi, J.T., Rasheed, S.A., Casey, P.J., Lossos, I.S., and Dave, S.S. (2016). GNA13 loss in germinal center B cells leads to impaired apoptosis and promotes lymphoma in vivo. *Blood* **127**, 2723-2731.

Heindryckx, F., Colle, I., and Van Vlierberghe, H. (2009). Experimental mouse models for hepatocellular carcinoma research. *Int J Exp Pathol* **90**, 367-386.

Hingorani, S.R., Petricoin, E.F., Maitra, A., Rajapakse, V., King, C., Jacobetz, M.A., Ross, S., Conrads, T.P., Veenstra, T.D., Hitt, B.A., Kawaguchi, Y., Johann, D., Liotta, L.A., Crawford, H.C., Putt, M.E., Jacks, T., Wright, C.V., Hruban, R.H., Lowy, A.M., and Tuveson, D.A. (2003). Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437-450.

Horlbeck, M.A., Gilbert, L.A., Villalta, J.E., Adamson, B., Pak, R.A., Chen, Y., Fields, A.P., Park, C.Y., Corn, J.E., Kampmann, M., and Weissman, J.S. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5**.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., and Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832.

Huang, J., Deng, Q., Wang, Q., Li, K.Y., Dai, J.H., Li, N., Zhu, Z.D., Zhou, B., Liu, X.Y., Liu, R.F., Fei, Q.L., Chen, H., Cai, B., Zhou, B., Xiao, H.S., Qin, L.X., and Han, Z.G. (2012). Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma. *Nat Genet* **44**, 1117-1121.

Iqbal, J., Greiner, T.C., Patel, K., Dave, B.J., Smith, L., Ji, J., Wright, G., Sanger, W.G., Pickering, D.L., Jain, S., Horsman, D.E., Shen, Y., Fu, K., Weisenburger, D.D., Hans, C.P., Campo, E., Gascoyne, R.D., Rosenwald, A., Jaffe, E.S., Delabie, J., Rimsza, L., Ott, G., Muller-Hermelink, H.K., Connors, J.M., Vose, J.M., McKeithan, T., Staudt, L.M., Chan, W.C., and Leukemia/Lymphoma Molecular Profiling, P. (2007). Distinctive patterns of BCL6 molecular alterations and their functional consequences in different subgroups of diffuse large B-cell lymphoma. *Leukemia* **21**, 2332-2343.

Iqbal, J., Sanger, W.G., Horsman, D.E., Rosenwald, A., Pickering, D.L., Dave, B., Dave, S., Xiao, L., Cao, K., Zhu, Q., Sherman, S., Hans, C.P., Weisenburger, D.D., Greiner, T.C., Gascoyne, R.D., Ott, G., Muller-Hermelink, H.K., Delabie, J., Braziel, R.M., Jaffe, E.S., Campo, E., Lynch, J.C., Connors, J.M., Vose, J.M., Armitage, J.O., Grogan, T.M., Staudt, L.M., and Chan, W.C. (2004). BCL2 translocation defines a unique tumor subset within the germinal center B-cell-like diffuse large B-cell lymphoma. *Am J Pathol* **165**, 159-166.

Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501-510.

Izsvak, Z., Ivics, Z., and Plasterk, R.H. (2000). Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J Mol Biol* **302**, 93-102.

Jentsch, I., Adler, I.D., Carter, N.P., and Speicher, M.R. (2001). Karyotyping mouse chromosomes by multiplex-FISH (M-FISH). *Chromosome Res* **9**, 211-214.

Jhappan, C., Stahle, C., Harkins, R.N., Fausto, N., Smith, G.H., and Merlino, G.T. (1990). TGF alpha overexpression in transgenic mice induces liver neoplasia and abnormal development of the mammary gland and pancreas. *Cell* **61**, 1137-1146.

Jiao, Y., Pawlik, T.M., Anders, R.A., Selaru, F.M., Streppel, M.M., Lucas, D.J., Niknafs, N., Guthrie, V.B., Maitra, A., Argani, P., Offerhaus, G.J., Roa, J.C., Roberts, L.R., Gores, G.J., Popescu, I., Alexandrescu, S.T., Dima, S., Fassan, M., Simbolo, M., Mafficini, A., Capelli, P., Lawlor, R.T., Ruzzenente, A., Guglielmi, A., Tortora, G., de Braud, F., Scarpa, A., Jarnagin, W., Klimstra, D., Karchin, R., Velculescu, V.E., Hruban, R.H., Vogelstein, B., Kinzler, K.W., Papadopoulos, N., and Wood, L.D. (2013). Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nat Genet* **45**, 1470-1473.

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.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A.T., Charpentier, E., Nogales, E., and Doudna, J.A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997.

Jorgensen, E.M., and Mango, S.E. (2002). The art and design of genetic screens: caenorhabditis elegans. *Nat Rev Genet* **3**, 356-369.

Kan, Z., Zheng, H., Liu, X., Li, S., Barber, T.D., Gong, Z., Gao, H., Hao, K., Willard, M.D., Xu, J., Hauptschein, R., Rejto, P.A., Fernandez, J., Wang, G., Zhang, Q., Wang, B., Chen, R., Wang, J., Lee, N.P., Zhou, W., Lin, Z., Peng, Z., Yi, K., Chen, S., Li, L., Fan, X., Yang, J., Ye, R., Ju, J., Wang, K., Estrella, H., Deng, S., Wei, P., Qiu, M., Wulur, I.H., Liu, J., Ehsani, M.E., Zhang, C., Loboda, A., Sung, W.K., Aggarwal, A., Poon, R.T., Fan, S.T., Wang, J., Hardwick, J., Reinhard, C., Dai, H., Li, Y., Luk, J.M., and Mao, M. (2013). Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res* **23**, 1422-1433.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., Leiserson, M.D., Miller, C.A., Welch, J.S., Walter, M.J., Wendl, M.C., Ley, T.J., Wilson, R.K., Raphael, B.J., and Ding, L. (2013). Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-339.

Kang, Y.K., Kim, W.H., and Jang, J.J. (2002). Expression of G1-S modulators (p53, p16, p27, cyclin D1, Rb) and Smad4/Dpc4 in intrahepatic cholangiocarcinoma. *Hum Pathol* **33**, 877-883.

Kataoka, K., Nagata, Y., Kitanaka, A., Shiraishi, Y., Shimamura, T., Yasunaga, J., Totoki, Y., Chiba, K., Sato-Otsubo, A., Nagae, G., Ishii, R., Muto, S., Kotani, S., Watatani, Y., Takeda, J., Sanada, M., Tanaka, H., Suzuki, H., Sato, Y., Shiozawa, Y., Yoshizato, T., Yoshida, K., Makishima, H., Iwanaga, M., Ma, G., Nosaka, K., Hishizawa, M., Itonaga, H., Imaizumi, Y., Munakata, W., Ogasawara, H., Sato, T., Sasai, K., Muramoto, K., Penova, M., Kawaguchi, T., Nakamura, H., Hama, N., Shide, K., Kubuki, Y., Hidaka, T., Kameda, T., Nakamaki, T., Ishiyama, K., Miyawaki, S., Yoon, S.S., Tobinai, K., Miyazaki, Y., Takaori-Kondo, A., Matsuda, F., Takeuchi, K., Nureki, O., Aburatani, H., Watanabe, T., Shibata, T., Matsuoka, M., Miyano, S., Shimoda, K., and Ogawa, S. (2015). Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat Genet* **47**, 1304-1315.

Katzenellenbogen, M., Mizrahi, L., Pappo, O., Klopstock, N., Olam, D., Jacob-Hirsch, J., Amariglio, N., Rechavi, G., Domany, E., Galun, E., and Goldenberg, D. (2007). Molecular mechanisms of liver carcinogenesis in the mdr2-knockout mice. *Mol Cancer Res* **5**, 1159-1170.

Keng, V.W., Villanueva, A., Chiang, D.Y., Dupuy, A.J., Ryan, B.J., Matise, I., Silverstein, K.A., Sarver, A., Starr, T.K., Akagi, K., Tessarollo, L., Collier, L.S., Powers, S., Lowe, S.W., Jenkins, N.A., Copeland, N.G., Llovet, J.M., and Largaespada, D.A. (2009). A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nat Biotechnol* **27**, 264-274.

Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. *Nat Rev Genet* **15**, 321-334.

Kloo, B., Nagel, D., Pfeifer, M., Grau, M., Duwel, M., Vincendeau, M., Dorken, B., Lenz, P., Lenz, G., and Krappmann, D. (2011). Critical role of PI3K signaling for NF-kappaB-dependent survival in a subset of activated B-cell-like diffuse large B-cell lymphoma cells. *Proc Natl Acad Sci U S A* **108**, 272-277.

Knight, B., Yeoh, G.C., Husk, K.L., Ly, T., Abraham, L.J., Yu, C., Rhim, J.A., and Fausto, N. (2000). Impaired preneoplastic changes and liver tumor formation in tumor necrosis factor receptor type 1 knockout mice. *J Exp Med* **192**, 1809-1818.

Knittel, G., Liedgens, P., Korovkina, D., Seeger, J.M., Al-Baldawi, Y., Al-Maarri, M., Fritz, C., Vlantis, K., Bezhanova, S., Scheel, A.H., Wolz, O.O., Reimann, M., Moller, P., Lopez, C., Schlesner, M., Lohneis, P., Weber, A.N., Trumper, L., German International Cancer Genome Consortium Molecular Mechanisms in Malignant Lymphoma by Sequencing Project, C., Staudt, L.M., Ortmann, M., Pasparakis, M., Siebert, R., Schmitt, C.A., Klatt, A.R., Wunderlich, F.T., Schafer, S.C., Persigehl, T., Montesinos-Rongen, M., Odenthal, M., Buttner, R., Frenzel, L.P., Kashkar, H., and Reinhardt, H.C. (2016). B-cell-specific conditional expression of Myd88p.L252P leads to the development of diffuse large B-cell lymphoma in mice. *Blood* **127**, 2732-2741.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283-2285.

Kojima, Y., Tsurumi, H., Goto, N., Shimizu, M., Kasahara, S., Yamada, T., Kanemura, N., Hara, T., Sawada, M., Saio, M., Yamada, T., Takahashi, T., Tomita, E., Takami, T., and Moriwaki, H. (2006). Fas and Fas ligand expression on germinal center type-diffuse large B-cell lymphoma is associated with the clinical outcome. *Eur J Haematol* **76**, 465-472.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., Nureki, O., and Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495-501.

Lee, S., Kim, W.H., Jung, H.Y., Yang, M.H., and Kang, G.H. (2002). Aberrant CpG island methylation of multiple genes in intrahepatic cholangiocarcinoma. *Am J Pathol* **161**, 1015-1022.

Lenz, G., Davis, R.E., Ngo, V.N., Lam, L., George, T.C., Wright, G.W., Dave, S.S., Zhao, H., Xu, W., Rosenwald, A., Ott, G., Muller-Hermelink, H.K., Gascoyne, R.D., Connors, J.M., Rimsza, L.M., Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Fisher, R.I., Chan, W.C., and Staudt, L.M. (2008a). Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676-1679.

Lenz, G., Nagel, I., Siebert, R., Roschke, A.V., Sanger, W., Wright, G.W., Dave, S.S., Tan, B., Zhao, H., Rosenwald, A., Muller-Hermelink, H.K., Gascoyne, R.D., Campo, E., Jaffe, E.S., Smeland, E.B., Fisher, R.I., Kuehl, W.M., Chan, W.C., and Staudt, L.M. (2007). Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell-like diffuse large B cell lymphoma. *J Exp Med* **204**, 633-643.

Lenz, G., Wright, G.W., Emre, N.C., Kohlhammer, H., Dave, S.S., Davis, R.E., Carty, S., Lam, L.T., Shaffer, A.L., Xiao, W., Powell, J., Rosenwald, A., Ott, G., Muller-Hermelink, H.K., Gascoyne, R.D., Connors, J.M., Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Rimsza, L.M., Fisher, R.I., Weisenburger, D.D., Chan, W.C., and Staudt, L.M. (2008b). Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci U S A* **105**, 13520-13525.

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, L., Eng, C., Desnick, R.J., German, J., and Ellis, N.A. (1998). Carrier frequency of the Bloom syndrome blmAsh mutation in the Ashkenazi Jewish population. *Mol Genet Metab* **64**, 286-290.

Li, M., Zhao, H., Zhang, X., Wood, L.D., Anders, R.A., Choti, M.A., Pawlik, T.M., Daniel, H.D., Kannangai, R., Offerhaus, G.J., Velculescu, V.E., Wang, L., Zhou, S., Vogelstein, B., Hruban, R.H., Papadopoulos, N., Cai, J., Torbenson, M.S., and Kinzler, K.W. (2011). Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat Genet* **43**, 828-829.

Li, M.A., Pettitt, S.J., Eckert, S., Ning, Z., Rice, S., Cadinanos, J., Yusa, K., Conte, N., and Bradley, A. (2013). The piggyBac transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites. *Mol Cell Biol* **33**, 1317-1330.

Li, Y., Park, A.I., Mou, H., Colpan, C., Bizhanova, A., Akama-Garren, E., Joshi, N., Hendrickson, E.A., Feldser, D., Yin, H., Anderson, D.G., Jacks, T., Weng, Z., and Xue, W. (2015). A versatile reporter system for CRISPR-mediated chromosomal rearrangements. *Genome Biol* **16**, 111.

Liang, Q., Kong, J., Stalker, J., and Bradley, A. (2009). Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. *Genesis* **47**, 404-408.

Liedtke, C., Luedde, T., Sauerbruch, T., Scholten, D., Streetz, K., Tacke, F., Tolba, R., Trautwein, C., Trebicka, J., and Weiskirchen, R. (2013). Experimental liver fibrosis research: update on animal models, legal issues and translational aspects. *Fibrogenesis Tissue Repair* **6**, 19.

Liu, G., Aronovich, E.L., Cui, Z., Whitley, C.B., and Hackett, P.B. (2004). Excision of Sleeping Beauty transposons: parameters and applications to gene therapy. *J Gene Med* **6**, 574-583.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233-247 e217.

Llovet, J.M., Zucman-Rossi, J., Pikarsky, E., Sangro, B., Schwartz, M., Sherman, M., and Gores, G. (2016). Hepatocellular carcinoma. *Nat Rev Dis Primers* **2**, 16018.

Lohr, J.G., Stojanov, P., Lawrence, M.S., Auclair, D., Chapuy, B., Sougnez, C., Cruz-Gordillo, P., Knoechel, B., Asmann, Y.W., Slager, S.L., Novak, A.J., Dogan, A., Ansell, S.M., Link, B.K., Zou, L., Gould, J., Saksena, G., Stransky, N., Rangel-Escareno, C., Fernandez-Lopez, J.C., Hidalgo-Miranda, A., Melendez-Zajgla, J., Hernandez-Lemus, E., Schwarz-Cruz y Celis, A., Imaz-Rosshandler, I., Ojesina, A.I., Jung, J., Pedamallu, C.S., Lander, E.S., Habermann, T.M., Cerhan, J.R., Shipp, M.A., Getz, G., and Golub, T.R. (2012). Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci U S A* **109**, 3879-3884.

Lossos, I.S., Alizadeh, A.A., Diehn, M., Warnke, R., Thorstenson, Y., Oefner, P.J., Brown, P.O., Botstein, D., and Levy, R. (2002). Transformation of follicular lymphoma to diffuse large-cell lymphoma: alternative patterns with increased or decreased expression of c-myc and its regulated genes. *Proc Natl Acad Sci U S A* **99**, 8886-8891.

Luckey, S.W., and Petersen, D.R. (2001). Activation of Kupffer cells during the course of carbon tetrachloride-induced liver injury and fibrosis in rats. *Exp Mol Pathol* **71**, 226-240.

Luo, G., Santoro, I.M., McDaniel, L.D., Nishijima, I., Mills, M., Youssoufian, H., Vogel, H., Schultz, R.A., and Bradley, A. (2000). Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nat Genet* **26**, 424-429.

Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E., Lowe, S.W., and Ventura, A. (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516**, 423-427.

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., Horvath, P., Moineau, S., Mojica, F.J., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R., and Koonin, E.V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722-736.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-838.

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

Mandelbaum, J., Bhagat, G., Tang, H., Mo, T., Brahmachary, M., Shen, Q., Chadburn, A., Rajewsky, K., Tarakhovsky, A., Pasqualucci, L., and Dalla-Favera, R. (2010). BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. *Cancer Cell* **18**, 568-579.

Mann, K.M., Ward, J.M., Yew, C.C., Kovochich, A., Dawson, D.W., Black, M.A., Brett, B.T., Sheetz, T.E., Dupuy, A.J., Australian Pancreatic Cancer Genome, I., Chang, D.K., Biankin, A.V., Waddell, N., Kassahn, K.S., Grimmond, S.M., Rust, A.G., Adams, D.J., Jenkins, N.A., and Copeland, N.G. (2012). Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. *Proc Natl Acad Sci U S A* **109**, 5934-5941.

Mann, M.B., Black, M.A., Jones, D.J., Ward, J.M., Yew, C.C., Newberg, J.Y., Dupuy, A.J., Rust, A.G., Bosenberg, M.W., McMahon, M., Print, C.G., Copeland, N.G., and Jenkins, N.A. (2015). Transposon mutagenesis identifies genetic drivers of Braf(V600E) melanoma. *Nat Genet* **47**, 486-495.

Maresch, R., Mueller, S., Veltkamp, C., Ollinger, R., Friedrich, M., Heid, I., Steiger, K., Weber, J., Engleitner, T., Barenboim, M., Klein, S., Louzada, S., Banerjee, R., Strong, A., Stauber, T., Gross, N., Geumann, U., Lange, S., Ringelhan, M., Varela, I., Unger, K., Yang, F., Schmid, R.M., Vassiliou, G.S., Braren, R., Schneider, G., Heikenwalder, M., Bradley, A., Saur, D., and Rad, R. (2016). Multiplexed pancreatic genome engineering and cancer induction by transfection-based CRISPR/Cas9 delivery in mice. *Nat Commun* **7**, 10770.

Martelli, M., Ferreri, A.J., Agostinelli, C., Di Rocco, A., Pfreundschuh, M., and Pileri, S.A. (2013). Diffuse large B-cell lymphoma. *Crit Rev Oncol Hematol* **87**, 146-171.

Mates, L., Chuah, M.K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., VandenDriessche, T., Ivics, Z., and Izsvak, Z. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet* **41**, 753-761.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* **36**, 344-355.

Mills, J.R., Malina, A., Lee, T., Di Paola, D., Larsson, O., Miething, C., Grosse, F., Tang, H., Zannis-Hadjopoulos, M., Lowe, S.W., and Pelletier, J. (2013). RNAi screening uncovers Dhx9 as a modifier of ABT-737 resistance in an Emu-myc/Bcl-2 mouse model. *Blood* **121**, 3402-3412.

Minnich, M., Tagoh, H., Bonelt, P., Axelsson, E., Fischer, M., Cebolla, B., Tarakhovsky, A., Nutt, S.L., Jaritz, M., and Busslinger, M. (2016). Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation. *Nat Immunol* **17**, 331-343.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174-182.

Mojica, F.J., Diez-Villasenor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36**, 244-246.

Molyneux, S.D., Waterhouse, P.D., Shelton, D., Shao, Y.W., Watling, C.M., Tang, Q.L., Harris, I.S., Dickson, B.C., Tharmapalan, P., Sandve, G.K., Zhang, X., Bailey, S.D., Berman, H., Wunder, J.S., Izsvak, Z., Lupien, M., Mak, T.W., and Khokha, R. (2014). Human somatic cell mutagenesis creates genetically tractable sarcomas. *Nat Genet* **46**, 964-972.

Moriarity, B.S., Otto, G.M., Rahrmann, E.P., Rathe, S.K., Wolf, N.K., Weg, M.T., Manlove, L.A., LaRue, R.S., Temiz, N.A., Molyneux, S.D., Choi, K., Holly, K.J., Sarver, A.L., Scott, M.C., Forster, C.L., Modiano, J.F., Khanna, C., Hewitt, S.M., Khokha, R., Yang, Y., Gorlick, R., Dyer, M.A., and Largaespada, D.A. (2015). A Sleeping Beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. *Nat Genet* **47**, 615-624.

Morin, R.D., Johnson, N.A., Severson, T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., Kuchenbauer, F., Yap, D., Humphries, R.K., Griffith, O.L., Shah, S., Zhu, H., Kimbara, M., Shashkin, P., Charlot, J.F., Tcherpakov, M., Corbett, R., Tam, A., Varhol, R., Smailus, D., Moksa, M., Zhao, Y., Delaney, A., Qian, H., Birol, I., Schein, J., Moore, R., Holt, R., Horsman, D.E., Connors, J.M., Jones, S., Aparicio, S., Hirst, M., Gascoyne, R.D., and Marra, M.A. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* **42**, 181-185.

Morin, R.D., Mendez-Lago, M., Mungall, A.J., Goya, R., Mungall, K.L., Corbett, R.D., Johnson, N.A., Severson, T.M., Chiu, R., Field, M., Jackman, S., Krzywinski, M., Scott, D.W., Trinh, D.L., Tamura-Wells, J., Li, S., Firme, M.R., Rogic, S., Griffith, M., Chan, S., Yakovenko, O., Meyer, I.M., Zhao, E.Y., Smailus, D., Moksa, M., Chittaranjan, S., Rimsza, L., Brooks-Wilson, A., Spinelli, J.J., Ben-Neriah, S., Meissner, B., Woolcock, B., Boyle, M., McDonald, H., Tam, A., Zhao, Y., Delaney, A., Zeng, T., Tse, K., Butterfield, Y., Birol, I., Holt, R., Schein, J., Horsman, D.E., Moore, R., Jones, S.J., Connors, J.M., Hirst, M., Gascoyne, R.D., and Marra, M.A. (2011). Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298-303.

Morin, R.D., Mungall, K., Pleasance, E., Mungall, A.J., Goya, R., Huff, R.D., Scott, D.W., Ding, J., Roth, A., Chiu, R., Corbett, R.D., Chan, F.C., Mendez-Lago, M., Trinh, D.L., Bolger-Munro, M., Taylor, G., Hadj Khodabakhshi, A., Ben-Neriah, S., Pon, J., Meissner, B., Woolcock, B., Farnoud, N., Rogic, S., Lim, E.L., Johnson, N.A., Shah, S., Jones, S., Steidl, C., Holt, R., Birol, I., Moore, R., Connors, J.M., Gascoyne, R.D., and Marra, M.A. (2013). Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* **122**, 1256-1265.

Muppidi, J.R., Schmitz, R., Green, J.A., Xiao, W., Larsen, A.B., Braun, S.E., An, J., Xu, Y., Rosenwald, A., Ott, G., Gascoyne, R.D., Rimsza, L.M., Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Braziel, R.M., Tubbs, R.R., Cook, J.R., Weisenburger, D.D., Chan, W.C., Vaidehi, N., Staudt, L.M., and Cyster, J.G. (2014). Loss of signalling via Galpha13 in germinal centre B-cell-derived lymphoma. *Nature* **516**, 254-258.

Nagy, R., Sweet, K., and Eng, C. (2004). Highly penetrant hereditary cancer syndromes. *Oncogene* **23**, 6445-6470.

Nault, J.C., Datta, S., Imbeaud, S., Franconi, A., Mallet, M., Couchy, G., Letouze, E., Pilati, C., Verret, B., Blanc, J.F., Balabaud, C., Calderaro, J., Laurent, A., Letexier, M., Bioulac-Sage, P., Calvo, F., and Zucman-Rossi, J. (2015). Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat Genet* **47**, 1187-1193.

Ni, T.K., Landrette, S.F., Bjornson, R.D., Bosenberg, M.W., and Xu, T. (2013). Low-copy piggyBac transposon mutagenesis in mice identifies genes driving melanoma. *Proc Natl Acad Sci U S A* **110**, E3640-3649.

Nicholes, K., Guillet, S., Tomlinson, E., Hillan, K., Wright, B., Frantz, G.D., Pham, T.A., Dillard-Telm, L., Tsai, S.P., Stephan, J.P., Stinson, J., Stewart, T., and French, D.M. (2002). A mouse model of hepatocellular carcinoma: ectopic expression of fibroblast growth factor 19 in skeletal muscle of transgenic mice. *Am J Pathol* **160**, 2295-2307.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949.

Novak, A.J., Asmann, Y.W., Maurer, M.J., Wang, C., Slager, S.L., Hodge, L.S., Manske, M., Price-Troska, T., Yang, Z.Z., Zimmermann, M.T., Nowakowski, G.S., Ansell, S.M., Witzig, T.E., McPhail, E., Ketterling, R., Feldman, A.L., Dogan, A., Link, B.K., Habermann, T.M., and Cerhan, J.R. (2015). Whole-exome analysis reveals novel somatic genomic alterations associated with outcome in immunochemotherapy-treated diffuse large B-cell lymphoma. *Blood Cancer J* **5**, e346.

O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M., and Doudna, J.A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**, 263-266.

O'Dell, M.R., Huang, J.L., Whitney-Miller, C.L., Deshpande, V., Rothberg, P., Grose, V., Rossi, R.M., Zhu, A.X., Land, H., Bardeesy, N., and Hezel, A.F. (2012). Kras(G12D) and p53 mutation cause primary intrahepatic cholangiocarcinoma. *Cancer Res* **72**, 1557-1567.

Ohtani, N., Yamakoshi, K., Takahashi, A., and Hara, E. (2004). The p16INK4a-RB pathway: molecular link between cellular senescence and tumor suppression. *J Med Invest* **51**, 146-153.

Ong, C.K., Subimerb, C., Pairojkul, C., Wongkham, S., Cutcutache, I., Yu, W., McPherson, J.R., Allen, G.E., Ng, C.C., Wong, B.H., Myint, S.S., Rajasegaran, V., Heng, H.L., Gan, A., Zang, Z.J., Wu, Y., Wu, J., Lee, M.H., Huang, D., Ong, P., Chan-on, W., Cao, Y., Qian, C.N., Lim, K.H., Ooi, A., Dykema, K., Furge, K., Kukongviriyapan, V., Sripa, B., Wongkham, C., Yongvanit, P., Futreal, P.A., Bhudhisawasdi, V., Rozen, S., Tan, P., and Teh, B.T. (2012). Exome sequencing of liver fluke-associated cholangiocarcinoma. *Nat Genet* **44**, 690-693.

Ozenne, P., Eymin, B., Brambilla, E., and Gazzeri, S. (2010). The ARF tumor suppressor: structure, functions and status in cancer. *Int J Cancer* **127**, 2239-2247.

Pasqualucci, L., Compagno, M., Houldsworth, J., Monti, S., Grunn, A., Nandula, S.V., Aster, J.C., Murty, V.V., Shipp, M.A., and Dalla-Favera, R. (2006). Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J Exp Med* **203**, 311-317.

Pasqualucci, L., Dominguez-Sola, D., Chiarenza, A., Fabbri, G., Grunn, A., Trifonov, V., Kasper, L.H., Lerach, S., Tang, H., Ma, J., Rossi, D., Chadburn, A., Murty, V.V., Mullighan, C.G., Gaidano, G., Rabadan, R., Brindle, P.K., and Dalla-Favera, R. (2011a). Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189-195.

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.

Pasqualucci, L., Trifonov, V., Fabbri, G., Ma, J., Rossi, D., Chiarenza, A., Wells, V.A., Grunn, A., Messina, M., Elliot, O., Chan, J., Bhagat, G., Chadburn, A., Gaidano, G., Mullighan, C.G., Rabadan, R., and Dalla-Favera, R. (2011b). Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet* **43**, 830-837.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839-843.

Perez-Mancera, P.A., Rust, A.G., van der Weyden, L., Kristiansen, G., Li, A., Sarver, A.L., Silverstein, K.A., Grutzmann, R., Aust, D., Rummele, P., Knosel, T., Herd, C., Stemple, D.L., Kettleborough, R., Brosnan, J.A., Li, A., Morgan, R., Knight, S., Yu, J., Stegeman, S., Collier, L.S., ten Hoeve, J.J., de Ridder, J., Klein, A.P., Goggins, M., Hruban, R.H., Chang, D.K., Biankin, A.V., Grimmond, S.M., Australian Pancreatic Cancer Genome, I., Wessels, L.F., Wood, S.A., Iacobuzio-Donahue, C.A., Pilarsky, C., Largaespada, D.A., Adams, D.J., and Tuveson, D.A. (2012). The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. *Nature* **486**, 266-270.

Pfeifer, M., Grau, M., Lenze, D., Wenzel, S.S., Wolf, A., Wollert-Wulf, B., Dietze, K., Nogai, H., Storek, B., Madle, H., Dorken, B., Janz, M., Dirnhofer, S., Lenz, P., Hummel, M., Tzankov, A., and Lenz, G. (2013). PTEN loss defines a PI3K/AKT pathway-dependent germinal center subtype of diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A* **110**, 12420-12425.

Pfeifer, M., and Lenz, G. (2013). PI3K/AKT addiction in subsets of diffuse large B-cell lymphoma. *Cell Cycle* **12**, 3347-3348.

Pfreundschuh, M., Kuhnt, E., Trumper, L., Osterborg, A., Trneny, M., Shepherd, L., Gill, D.S., Walewski, J., Pettengell, R., Jaeger, U., Zinzani, P.L., Shpilberg, O., Kvaloy, S., de Nully Brown, P., Stahel, R., Milpied, N., Lopez-Guillermo, A., Poeschel, V., Grass, S., Loeffler, M., Murawski, N., and MabThera International Trial, G. (2011). CHOP-like chemotherapy with or without rituximab in young patients with good-prognosis diffuse large-B-cell lymphoma: 6-year results of an open-label randomised study of the MabThera International Trial (MInT) Group. *Lancet Oncol* **12**, 1013-1022.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., Graham, D.B., Jhunjhunwala, S., Heidenreich, M., Xavier, R.J., Langer, R., Anderson, D.G., Hacohen, N., Regev, A., Feng, G., Sharp, P.A., and Zhang, F. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440-455.

Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.L., Ordonez, G.R., Bignell, G.R., Ye, K., Alipaz, J., Bauer, M.J., Beare, D., Butler, A., Carter, R.J., Chen, L., Cox, A.J., Edkins, S., Kokko-Gonzales, P.I., Gormley, N.A., Grocock, R.J., Haudenschild, C.D., Hims, M.M., James, T., Jia, M., Kingsbury, Z., Leroy, C., Marshall, J., Menzies, A., Mudie, L.J., Ning, Z., Royce, T., Schulz-Trieglaff, O.B., Spiridou, A., Stebbings, L.A., Szajkowski, L., Teague, J., Williamson, D., Chin, L., Ross, M.T., Campbell, P.J., Bentley, D.R., Futreal, P.A., and Stratton, M.R. (2010). A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191-196.

Porcu, M., Kleppe, M., Gianfelici, V., Geerdens, E., De Keersmaecker, K., Tartaglia, M., Foa, R., Soulier, J., Cauwelier, B., Uyttebroeck, A., Macintyre, E., Vandenberghe, P., Asnafi, V., and Cools, J. (2012). Mutation of the receptor tyrosine phosphatase PTPRC (CD45) in T-cell acute lymphoblastic leukemia. *Blood* **119**, 4476-4479.

Postic, C., Shiota, M., Niswender, K.D., Jetton, T.L., Chen, Y., Moates, J.M., Shelton, K.D., Lindner, J., Cherrington, A.D., and Magnuson, M.A. (1999). Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. *J Biol Chem* **274**, 305-315.

Quail, M.A., Swerdlow, H., and Turner, D.J. (2009). Improved protocols for the illumina genome analyzer sequencing system. *Curr Protoc Hum Genet* **Chapter 18**, Unit 18 12.

Quinlan, A.R., Clark, R.A., Sokolova, S., Leibowitz, M.L., Zhang, Y., Hurles, M.E., Mell, J.C., and Hall, I.M. (2010). Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. *Genome Res* **20**, 623-635.

Rad, R., Rad, L., Wang, W., Cadinanos, J., Vassiliou, G., Rice, S., Campos, L.S., Yusa, K., Banerjee, R., Li, M.A., de la Rosa, J., Strong, A., Lu, D., Ellis, P., Conte, N., Yang, F.T., Liu, P., and Bradley, A. (2010). PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. *Science* **330**, 1104-1107.

Rad, R., Rad, L., Wang, W., Strong, A., Ponstingl, H., Bronner, I.F., Mayho, M., Steiger, K., Weber, J., Hieber, M., Veltkamp, C., Eser, S., Geumann, U., Ollinger, R., Zukowska, M., Barenboim, M., Maresch, R., Cadinanos, J., Friedrich, M., Varela, I., Constantino-Casas, F., Sarver, A., Ten Hoeve, J., Prosser, H., Seidler, B., Bauer, J., Heikenwalder, M., Metzakopian, E., Krug, A., Ehmer, U., Schneider, G., Knosel, T., Rummele, P., Aust, D., Grutzmann, R., Pilarsky, C., Ning, Z., Wessels, L., Schmid, R.M., Quail, M.A., Vassiliou, G., Esposito, I., Liu, P., Saur, D., and Bradley, A. (2015). A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nat Genet* **47**, 47-56.

Rahrmann, E.P., Watson, A.L., Keng, V.W., Choi, K., Moriarity, B.S., Beckmann, D.A., Wolf, N.K., Sarver, A., Collins, M.H., Moertel, C.L., Wallace, M.R., Gel, B., Serra, E., Ratner, N., and Largaespada, D.A. (2013). Forward genetic screen for malignant peripheral nerve sheath tumor formation identifies new genes and pathways driving tumorigenesis. *Nat Genet* **45**, 756-766.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308.

Razumilava, N., and Gores, G.J. (2014). Cholangiocarcinoma. *Lancet* **383**, 2168-2179.

Roschewski, M., Staudt, L.M., and Wilson, W.H. (2014). Diffuse large B-cell lymphoma-treatment approaches in the molecular era. *Nat Rev Clin Oncol* **11**, 12-23.

Ross, J.S., Wang, K., Gay, L., Al-Rohil, R., Rand, J.V., Jones, D.M., Lee, H.J., Sheehan, C.E., Otto, G.A., Palmer, G., Yelensky, R., Lipson, D., Morosini, D., Hawryluk, M., Catenacci, D.V., Miller, V.A., Churi, C., Ali, S., and Stephens, P.J. (2014). New routes to targeted therapy of intrahepatic cholangiocarcinomas revealed by next-generation sequencing. *Oncologist* **19**, 235-242.

Saha, S.K., Parachoniak, C.A., Ghanta, K.S., Fitamant, J., Ross, K.N., Najem, M.S., Gurumurthy, S., Akbay, E.A., Sia, D., Cornella, H., Miltiadous, O., Walesky, C., Deshpande, V., Zhu, A.X., Hezel, A.F., Yen, K.E., Straley, K.S., Travins, J., Popovici-Muller, J., Gliser, C., Ferrone, C.R., Apte, U., Llovet, J.M., Wong, K.K., Ramaswamy, S., and Bardeesy, N. (2014). Mutant IDH inhibits HNF-4alpha to block hepatocyte differentiation and promote biliary cancer. *Nature* **513**, 110-114.

Sanchez-Rivera, F.J., Papagiannakopoulos, T., Romero, R., Tammela, T., Bauer, M.R., Bhutkar, A., Joshi, N.S., Subbaraj, L., Bronson, R.T., Xue, W., and Jacks, T. (2014). Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* **516**, 428-431.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784.

Sarver, A.L., Erdman, J., Starr, T., Largaespada, D.A., and Silverstein, K.A. (2012). TAPDANCE: an automated tool to identify and annotate transposon insertion CISs and associations between CISs from next generation sequence data. *BMC Bioinformatics* **13**, 154.

Savage, K.J., Johnson, N.A., Ben-Neriah, S., Connors, J.M., Sehn, L.H., Farinha, P., Horsman, D.E., and Gascoyne, R.D. (2009). MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. *Blood* **114**, 3533-3537.

Schmidt, S., Rainer, J., Ploner, C., Presul, E., Riml, S., and Kofler, R. (2004). Glucocorticoid-induced apoptosis and glucocorticoid resistance: molecular mechanisms and clinical relevance. *Cell Death Differ* **11 Suppl 1**, S45-55.

Schulze, K., Imbeaud, S., Letouze, E., Alexandrov, L.B., Calderaro, J., Rebouissou, S., Couchy, G., Meiller, C., Shinde, J., Soysouvanh, F., Calatayud, A.L., Pinyol, R., Pelletier, L., Balabaud, C., Laurent, A., Blanc, J.F., Mazzaferro, V., Calvo, F., Villanueva, A., Nault, J.C., Bioulac-Sage, P., Stratton, M.R., Llovet, J.M., and Zucman-Rossi, J. (2015). Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* **47**, 505-511.

Shaib, Y., and El-Serag, H.B. (2004). The epidemiology of cholangiocarcinoma. *Semin Liver Dis* **24**, 115-125.

Shankland, K.R., Armitage, J.O., and Hancock, B.W. (2012). Non-Hodgkin lymphoma. *Lancet* **380**, 848-857.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., and Skarnes, W.C. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* **11**, 399-402.

Sherr, C.J. (2004). Principles of tumor suppression. *Cell* **116**, 235-246.

Sia, D., Hoshida, Y., Villanueva, A., Roayaie, S., Ferrer, J., Tabak, B., Peix, J., Sole, M., Tovar, V., Alsinet, C., Cornella, H., Klotzle, B., Fan, J.B., Cotsoglou, C., Thung, S.N., Fuster, J., Waxman, S., Garcia-Valdecasas, J.C., Bruix, J., Schwartz, M.E., Beroukhim, R., Mazzaferro, V., and Llovet, J.M. (2013). Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. *Gastroenterology* **144**, 829-840.

Siegel, R., Naishadham, D., and Jemal, A. (2013). Cancer statistics, 2013. *CA Cancer J Clin* **63**, 11-30.

Siegel, R.L., Miller, K.D., and Jemal, A. (2016). Cancer statistics, 2016. *CA Cancer J Clin* **66**, 7-30.

Smith, A., Howell, D., Patmore, R., Jack, A., and Roman, E. (2011). Incidence of haematological malignancy by sub-type: a report from the Haematological Malignancy Research Network. *Br J Cancer* **105**, 1684-1692.

Sparks, A.B., Morin, P.J., Vogelstein, B., and Kinzler, K.W. (1998). Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res* **58**, 1130-1134.

Sriraksa, R., Zeller, C., El-Bahrawy, M.A., Dai, W., Daduang, J., Jearanaikoon, P., Chau-In, S., Brown, R., and Limpaiboon, T. (2011). CpG-island methylation study of liver fluke-related cholangiocarcinoma. *Br J Cancer* **104**, 1313-1318.

St Johnston, D. (2002). The art and design of genetic screens: Drosophila melanogaster. *Nat Rev Genet* **3**, 176-188.

Starr, T.K., Allaei, R., Silverstein, K.A., Staggs, R.A., Sarver, A.L., Bergemann, T.L., Gupta, M., O'Sullivan, M.G., Matise, I., Dupuy, A.J., Collier, L.S., Powers, S., Oberg, A.L., Asmann, Y.W., Thibodeau, S.N., Tessarollo, L., Copeland, N.G., Jenkins, N.A., Cormier, R.T., and Largaespada, D.A. (2009). A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323**, 1747-1750.

Starr, T.K., Scott, P.M., Marsh, B.M., Zhao, L., Than, B.L., O'Sullivan, M.G., Sarver, A.L., Dupuy, A.J., Largaespada, D.A., and Cormier, R.T. (2011). A Sleeping Beauty transposon-mediated screen identifies murine susceptibility genes for adenomatous polyposis coli (Apc)-dependent intestinal tumorigenesis. *Proc Natl Acad Sci U S A* **108**, 5765-5770.

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

Sugimachi, K., Taguchi, K., Aishima, S., Tanaka, S., Shimada, M., Kajiyama, K., Sugimachi, K., and Tsuneyoshi, M. (2001). Altered expression of beta-catenin without genetic mutation in intrahepatic cholangiocarcinoma. *Mod Pathol* **14**, 900-905.

Sung, W.K., Zheng, H., Li, S., Chen, R., Liu, X., Li, Y., Lee, N.P., Lee, W.H., Ariyaratne, P.N., Tennakoon, C., Mulawadi, F.H., Wong, K.F., Liu, A.M., Poon, R.T., Fan, S.T., Chan, K.L., Gong, Z., Hu, Y., Lin, Z.,

Wang, G., Zhang, Q., Barber, T.D., Chou, W.C., Aggarwal, A., Hao, K., Zhou, W., Zhang, C., Hardwick, J., Buser, C., Xu, J., Kan, Z., Dai, H., Mao, M., Reinhard, C., Wang, J., and Luk, J.M. (2012). Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* **44**, 765-769.

Suzuki, E., Umezawa, K., and Bonavida, B. (2007). Rituximab inhibits the constitutively activated PI3K-Akt pathway in B-NHL cell lines: involvement in chemosensitization to drug-induced apoptosis. *Oncogene* **26**, 6184-6193.

Suzuki, T., Minehata, K., Akagi, K., Jenkins, N.A., and Copeland, N.G. (2006). Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice. *EMBO J* **25**, 3422-3431.

Swerdlow, S.H., Campo, E., Pileri, S.A., Harris, N.L., Stein, H., Siebert, R., Advani, R., Ghielmini, M., Salles, G.A., Zelenetz, A.D., and Jaffe, E.S. (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**, 2375-2390.

Takada, S., Tsuchida, N., Kobayashi, M., and Koike, K. (1995). Disruption of the function of tumor-suppressor gene p53 by the hepatitis B virus X protein and hepatocarcinogenesis. *J Cancer Res Clin Oncol* **121**, 593-601.

Takeda, H., Wei, Z., Koso, H., Rust, A.G., Yew, C.C., Mann, M.B., Ward, J.M., Adams, D.J., Copeland, N.G., and Jenkins, N.A. (2015). Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor progression. *Nat Genet* **47**, 142-150.

Tang, J.Z., Carmichael, C.L., Shi, W., Metcalf, D., Ng, A.P., Hyland, C.D., Jenkins, N.A., Copeland, N.G., Howell, V.M., Zhao, Z.J., Smyth, G.K., Kile, B.T., and Alexander, W.S. (2013). Transposon mutagenesis reveals cooperation of ETS family transcription factors with signaling pathways in erythro-megakaryocytic leukemia. *Proc Natl Acad Sci U S A* **110**, 6091-6096.

Tannapfel, A., Benicke, M., Katalinic, A., Uhlmann, D., Kockerling, F., Hauss, J., and Wittekind, C. (2000). Frequency of p16(INK4A) alterations and K-ras mutations in intrahepatic cholangiocarcinoma of the liver. *Gut* **47**, 721-727.

Thibault, S.T., Singer, M.A., Miyazaki, W.Y., Milash, B., Dompe, N.A., Singh, C.M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H.L., Ryner, L., Cheung, L.M., Chong, A., Erickson, C., Fisher, W.W., Greer, K., Hartouni, S.R., Howie, E., Jakkula, L., Joo, D., Killpack, K., Laufer, A., Mazzotta, J., Smith, R.D., Stevens, L.M., Stuber, C., Tan, L.R., Ventura, R., Woo, A., Zakrajsek, I., Zhao, L., Chen, F., Swimmer, C., Kopczynski, C., Duyk, G., Winberg, M.L., and Margolis, J. (2004). A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. *Nat Genet* **36**, 283-287.

Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J Clin* **65**, 87-108.

Totoki, Y., Tatsuno, K., Yamamoto, S., Arai, Y., Hosoda, F., Ishikawa, S., Tsutsumi, S., Sonoda, K., Totsuka, H., Shirakihara, T., Sakamoto, H., Wang, L., Ojima, H., Shimada, K., Kosuge, T., Okusaka, T., Kato, K., Kusuda, J., Yoshida, T., Aburatani, H., and Shibata, T. (2011). High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* **43**, 464-469.

Townsend, W., and Linch, D. (2012). Hodgkin's lymphoma in adults. *Lancet* **380**, 836-847.

van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., and Brouns, S.J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* **34**, 401-407.

van der Weyden, L., Giotopoulos, G., Rust, A.G., Matheson, L.S., van Delft, F.W., Kong, J., Corcoran, A.E., Greaves, M.F., Mullighan, C.G., Huntly, B.J., and Adams, D.J. (2011). Modeling the evolution of ETV6-RUNX1-induced B-cell precursor acute lymphoblastic leukemia in mice. *Blood* **118**, 1041-1051.

van der Weyden, L., Giotopoulos, G., Wong, K., Rust, A.G., Robles-Espinoza, C.D., Osaki, H., Huntly, B.J., and Adams, D.J. (2015). Somatic drivers of B-ALL in a model of ETV6-RUNX1; Pax5(+/-) leukemia. *BMC Cancer* **15**, 585.

van der Weyden, L., Rust, A.G., McIntyre, R.E., Robles-Espinoza, C.D., del Castillo Velasco-Herrera, M., Strogantsev, R., Ferguson-Smith, A.C., McCarthy, S., Keane, T.M., Arends, M.J., and Adams, D.J. (2013). Jdp2 downregulates Trp53 transcription to promote leukaemogenesis in the context of Trp53 heterozygosity. *Oncogene* **32**, 397-402.

Vasmatzis, G., Johnson, S.H., Knudson, R.A., Ketterling, R.P., Braggio, E., Fonseca, R., Viswanatha, D.S., Law, M.E., Kip, N.S., Ozsan, N., Grebe, S.K., Frederick, L.A., Eckloff, B.W., Thompson, E.A., Kadin, M.E., Milosevic, D., Porcher, J.C., Asmann, Y.W., Smith, D.I., Kovtun, I.V., Ansell, S.M., Dogan, A., and Feldman, A.L. (2012). Genome-wide analysis reveals recurrent structural abnormalities of TP63 and other p53-related genes in peripheral T-cell lymphomas. *Blood* **120**, 2280-2289.

Vigdal, T.J., Kaufman, C.D., Izsvak, Z., Voytas, D.F., and Ivics, Z. (2002). Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. *J Mol Biol* **323**, 441-452.

Wang, J.Q., Jeelall, Y.S., Beutler, B., Horikawa, K., and Goodnow, C.C. (2014a). Consequences of the recurrent MYD88(L265P) somatic mutation for B cell tolerance. *J Exp Med* **211**, 413-426.

Wang, L., Wang, W.L., Zhang, Y., Guo, S.P., Zhang, J., and Li, Q.L. (2007). Epigenetic and genetic alterations of PTEN in hepatocellular carcinoma. *Hepatol Res* **37**, 389-396.

Wang, P., Dong, Q., Zhang, C., Kuan, P.F., Liu, Y., Jeck, W.R., Andersen, J.B., Jiang, W., Savich, G.L., Tan, T.X., Auman, J.T., Hoskins, J.M., Misher, A.D., Moser, C.D., Yourstone, S.M., Kim, J.W., Cibulskis, K., Getz, G., Hunt, H.V., Thorgeirsson, S.S., Roberts, L.R., Ye, D., Guan, K.L., Xiong, Y., Qin, L.X., and Chiang, D.Y. (2013). Mutations in isocitrate dehydrogenase 1 and 2 occur frequently in intrahepatic cholangiocarcinomas and share hypermethylation targets with glioblastomas. *Oncogene* **32**, 3091-3100.

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

Wang, W., Bradley, A., and Huang, Y. (2009). A piggyBac transposon-based genome-wide library of insertionally mutated Blm-deficient murine ES cells. *Genome Res* **19**, 667-673.

Ward, P.S., Patel, J., Wise, D.R., Abdel-Wahab, O., Bennett, B.D., Coller, H.A., Cross, J.R., Fantin, V.R., Hedvat, C.V., Perl, A.E., Rabinowitz, J.D., Carroll, M., Su, S.M., Sharp, K.A., Levine, R.L., and Thompson, C.B. (2010). The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225-234.

Warren, M., Chung, Y.J., Howat, W.J., Harrison, H., McGinnis, R., Hao, X., McCafferty, J., Fredrickson, T.N., Bradley, A., and Morse, H.C., 3rd (2010). Irradiated Blm-deficient mice are a highly tumor prone model for analysis of a broad spectrum of hematologic malignancies. *Leuk Res* **34**, 210-220.

Watanabe, S., Horie, Y., Kataoka, E., Sato, W., Dohmen, T., Ohshima, S., Goto, T., and Suzuki, A. (2007). Non-alcoholic steatohepatitis and hepatocellular carcinoma: lessons from hepatocyte-specific phosphatase and tensin homolog (PTEN)-deficient mice. *J Gastroenterol Hepatol* **22 Suppl 1**, S96-S100.

Weber J., de la Rosa J., Grove C.S., Schick M., Rad L., Baranov O., Strong A., Pfaus A., Friedrich M.J., Engleitner T., Lersch R., Öllinger R., Grau M., Gonzalez Menendez I., Martella M., Kohlhofer U., Banerjee R., Turchaninova M.A., Scherger A., Hoffman G.J., Hess J., Kuhn L.B., Ammon T., Kim J., Schneider G., Unger K., Zimber-Strobl U., Heikenwälder M., Schmidt-Supprian M., Yang F., Saur D., Liu P., Steiger K., Chudakov D.M., Lenz G., Quintanilla-Martinez L., Keller U., Vassiliou G.S., Cadiñanos J., Bradley A., Rad R. (2019). PiggyBac transposon tools for recessive screening identify B-cell lymphoma drivers in mice. *Nat Commun* **10**, 1415.

Weber, J., Ollinger, R., Friedrich, M., Ehmer, U., Barenboim, M., Steiger, K., Heid, I., Mueller, S., Maresch, R., Engleitner, T., Gross, N., Geumann, U., Fu, B., Segler, A., Yuan, D., Lange, S., Strong, A., de la Rosa, J., Esposito, I., Liu, P., Cadinanos, J., Vassiliou, G.S., Schmid, R.M., Schneider, G., Unger, K., Yang, F., Braren, R., Heikenwalder, M., Varela, I., Saur, D., Bradley, A., and Rad, R. (2015). CRISPR/Cas9 somatic multiplex-mutagenesis for high-throughput functional cancer genomics in mice. *Proc Natl Acad Sci U S A* **112**, 13982-13987.

Weber, J.D., Jeffers, J.R., Rehg, J.E., Randle, D.H., Lozano, G., Roussel, M.F., Sherr, C.J., and Zambetti, G.P. (2000). p53-independent functions of the p19(ARF) tumor suppressor. *Genes Dev* **14**, 2358-2365.

Welcsh, P.L., and King, M.C. (2001). BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet* **10**, 705-713.

Wolf, M.J., Adili, A., Piotrowitz, K., Abdullah, Z., Boege, Y., Stemmer, K., Ringelhan, M., Simonavicius, N., Egger, M., Wohlleber, D., Lorentzen, A., Einer, C., Schulz, S., Clavel, T., Protzer, U., Thiele, C., Zischka, H., Moch, H., Tschop, M., Tumanov, A.V., Haller, D., Unger, K., Karin, M., Kopf, M., Knolle, P., Weber, A., and Heikenwalder, M. (2014). Metabolic activation of intrahepatic CD8+ T cells and NKT cells causes nonalcoholic steatohepatitis and liver cancer via cross-talk with hepatocytes. *Cancer Cell* **26**, 549-564.

Wright, A.V., Nunez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29-44.

Xu, X., Kobayashi, S., Qiao, W., Li, C., Xiao, C., Radaeva, S., Stiles, B., Wang, R.H., Ohara, N., Yoshino, T., LeRoith, D., Torbenson, M.S., Gores, G.J., Wu, H., Gao, B., and Deng, C.X. (2006). Induction of intrahepatic cholangiocellular carcinoma by liver-specific disruption of Smad4 and Pten in mice. *J Clin Invest* **116**, 1843-1852.

Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N.S., Cai, W., Yang, G., Bronson, R., Crowley, D.G., Zhang, F., Anderson, D.G., Sharp, P.A., and Jacks, T. (2014). CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* **514**, 380-384.

Yang, B., Guo, M., Herman, J.G., and Clark, D.P. (2003). Aberrant promoter methylation profiles of tumor suppressor genes in hepatocellular carcinoma. *Am J Pathol* **163**, 1101-1107.

Yang, B., House, M.G., Guo, M., Herman, J.G., and Clark, D.P. (2005). Promoter methylation profiles of tumor suppressor genes in intrahepatic and extrahepatic cholangiocarcinoma. *Mod Pathol* **18**, 412-420.

Yant, S.R., Huang, Y., Akache, B., and Kay, M.A. (2007). Site-directed transposon integration in human cells. *Nucleic Acids Res* **35**, e50.

Yant, S.R., Meuse, L., Chiu, W., Ivics, Z., Izsvak, Z., and Kay, M.A. (2000). Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nat Genet* **25**, 35-41.

Yusa, K., Horie, K., Kondoh, G., Kouno, M., Maeda, Y., Kinoshita, T., and Takeda, J. (2004). Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. *Nature* **429**, 896-899.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A* **108**, 1531-1536.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., and Lim, W.A. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-350.

Zender, S., Nickeleit, I., Wuestefeld, T., Sorensen, I., Dauch, D., Bozko, P., El-Khatib, M., Geffers, R., Bektas, H., Manns, M.P., Gossler, A., Wilkens, L., Plentz, R., Zender, L., and Malek, N.P. (2013). A critical role for notch signaling in the formation of cholangiocellular carcinomas. *Cancer Cell* **23**, 784-795.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., Koonin, E.V., and Zhang, F. (2015a). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771.

Zetsche, B., Volz, S.E., and Zhang, F. (2015b). A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **33**, 139-142.

Zhang, G., Budker, V., and Wolff, J.A. (1999). High levels of foreign gene expression in hepatocytes after tail vein injections of naked plasmid DNA. *Hum Gene Ther* **10**, 1735-1737.

Zhang, J., Grubor, V., Love, C.L., Banerjee, A., Richards, K.L., Mieczkowski, P.A., Dunphy, C., Choi, W., Au, W.Y., Srivastava, G., Lugar, P.L., Rizzieri, D.A., Lagoo, A.S., Bernal-Mizrachi, L., Mann, K.P., Flowers, C., Naresh, K., Evens, A., Gordon, L.I., Czader, M., Gill, J.I., Hsi, E.D., Liu, Q., Fan, A., Walsh, K., Jima, D., Smith, L.L., Johnson, A.J., Byrd, J.C., Luftig, M.A., Ni, T., Zhu, J., Chadburn, A., Levy, S., Dunson, D., and Dave, S.S. (2013). Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A* **110**, 1398-1403.

Zuckermann, M., Hovestadt, V., Knobbe-Thomsen, C.B., Zapatka, M., Northcott, P.A., Schramm, K., Belic, J., Jones, D.T., Tschida, B., Moriarity, B., Largaespada, D., Roussel, M.F., Korshunov, A., Reifenberger, G., Pfister, S.M., Lichter, P., Kawauchi, D., and Gronych, J. (2015). Somatic CRISPR/Cas9-mediated tumour suppressor disruption enables versatile brain tumour modelling. *Nat Commun* **6**, 7391.

10 PUBLICATIONS

Parts of this thesis have been published in *Nature Communications* and *Proceedings of the National Academy of Sciences*:

Weber J.*, de la Rosa J.*, Grove C.S.*, Schick M., Rad L., Baranov O., Strong A., Pfaus A., Friedrich M.J., Engleitner T., Lersch R., Öllinger R., Grau M., Gonzalez Menendez I., Martella M., Kohlhofer U., Banerjee R., Turchaninova M.A., Scherger A., Hoffman G.J., Hess J., Kuhn L.B., Ammon T., Kim J., Schneider G., Unger K., Zimber-Strobl U., Heikenwälder M., Schmidt-Supprian M., Yang F., Saur D., Liu P., Steiger K., Chudakov D.M., Lenz G., Quintanilla-Martinez L., Keller U., Vassiliou G.S., Cadiñanos J., Bradley A., Rad R. (2019). PiggyBac transposon tools for recessive screening identify B-cell lymphoma drivers in mice. *Nat Commun* **10**, 1415.
*shared first-authorship

Weber, J.*, Ollinger, R.*, Friedrich, M., Ehmer, U., Barenboim, M., Steiger, K., Heid, I., Mueller, S., Maresch, R., Engleitner, T., Gross, N., Geumann, U., Fu, B., Segler, A., Yuan, D., Lange, S., Strong, A., de la Rosa, J., Esposito, I., Liu, P., Cadinanos, J., Vassiliou, G.S., Schmid, R.M., Schneider, G., Unger, K., Yang, F., Braren, R., Heikenwalder, M., Varela, I., Saur, D., Bradley, A., and Rad, R. (2015). CRISPR/Cas9 somatic multiplex-mutagenesis for high-throughput functional cancer genomics in mice. *Proc Natl Acad Sci U S A* **112**, 13982-13987.
*shared first-authorship

Other thesis-related publications:

de la Rosa, J.*, **Weber, J.***, Friedrich, M. J., Li, Y., Rad, L., Ponstingl, H., Liang, Q., de Quiros, S. B., Noorani, I., Metzakopian, E., Strong, A., Li, M. A., Astudillo, A., Fernandez-Garcia, M. T., Fernandez-Garcia, M. S., Hoffman, G. J., Fuente, R., Vassiliou, G. S., Rad, R., Lopez-Otin, C., Bradley, A. and Cadinanos, J. (2017). A single-copy Sleeping Beauty transposon mutagenesis screen identifies new PTEN-cooperating tumor suppressor genes. *Nat Genet* **49**, 730-741.
*shared first-authorship

Rad, R., Rad, L., Wang, W., Strong, A., Ponstingl, H., Bronner, I. F., Mayho, M., Steiger, K., **Weber, J.**, Hieber, M., Veltkamp, C., Eser, S., Geumann, U., Ollinger, R., Zukowska, M., Barenboim, M., Maresch, R., Cadinanos, J., Friedrich, M., Varela, I., Constantino-Casas, F., Sarver, A., Ten Hoeve, J., Prosser, H., Seidler, B., Bauer, J., Heikenwalder, M., Metzakopian, E., Krug, A., Ehmer, U., Schneider, G., Knosel, T., Rummele, P., Aust, D., Grutzmann, R., Pilarsky, C., Ning, Z., Wessels, L., Schmid, R. M., Quail, M. A., Vassiliou, G., Esposito, I., Liu, P., Saur, D. and Bradley, A. (2015). A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nat Genet* **47**, 47-56.

Friedrich, M. J., Rad, L., Bronner, I. F., Strong, A., Wang, W., **Weber, J.**, Mayho, M., Ponstingl, H., Engleitner, T., Grove, C., Pfaus, A., Saur, D., Cadinanos, J., Quail, M. A., Vassiliou, G. S., Liu, P., Bradley, A. and Rad, R. (2017). Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nat Protoc* **12**, 289-309.

Weber, J. and Rad, R. (2019). Engineering CRISPR mouse models of cancer. *Curr Opin Genet Dev* **54**, 88-96.

Maresch, R., Mueller, S., Veltkamp, C., Ollinger, R., Friedrich, M., Heid, I., Steiger, K., **Weber, J.**, Engleitner, T., Barenboim, M., Klein, S., Louzada, S., Banerjee, R., Strong, A., Stauber, T., Gross, N., Geumann, U., Lange, S., Ringelhan, M., Varela, I., Unger, K., Yang, F., Schmid, R. M., Vassiliou, G. S., Braren, R., Schneider, G., Heikenwalder, M., Bradley, A., Saur, D. and Rad, R. (2016). Multiplexed pancreatic genome engineering and cancer induction by transfection-based CRISPR/Cas9 delivery in mice. *Nat Commun* **7**, 10770.

Yuan, D., Huang, S., Berger, E., Liu, L., Gross, N., Heinzmann, F., Ringelhan, M., Connor, T. O., Stadler, M., Meister, M., **Weber, J.**, Ollinger, R., Simonavicius, N., Reisinger, F., Hartmann, D., Meyer, R., Reich, M., Seehawer, M., Leone, V., Hochst, B., Wohlleber, D., Jors, S., Prinz, M., Spalding, D., Protzer, U., Luedde, T., Terracciano, L., Matter, M., Longerich, T., Knolle, P., Ried, T., Keitel, V., Geisler, F., Unger, K., Cinnamon, E., Pikarsky, E., Huser, N., Davis, R. J., Tschaharganeh, D. F., Rad, R., Weber, A., Zender, L., Haller, D. and Heikenwalder, M. (2017). Kupffer Cell-Derived Tnf Triggers Cholangiocellular Tumorigenesis through JNK due to Chronic Mitochondrial Dysfunction and ROS. *Cancer Cell* **31**, 771-789.

de la Rosa, J., **Weber, J.**, Rad, R., Bradley, A. and Cadinanos, J. (2017). Disentangling PTEN-cooperating tumor suppressor gene networks in cancer. *Mol Cell Oncol* **4**, e1325550.

Other publications:

Bassani-Sternberg, M., Braunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., **Weber, J.**, Slotta-Huspenina, J., Specht, K., Martignoni, M. E., Werner, A., Hein, R., D, H. B., Peschel, C., Rad, R., Cox, J., Mann, M. and Krackhardt, A. M. (2016). Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* **7**, 13404.

Frey, B., Ruckert, M., **Weber, J.**, Mayr, X., Derer, A., Lotter, M., Bert, C., Rodel, F., Fietkau, R. and Gaipl, U. S. (2017). Hypofractionated Irradiation Has Immune Stimulatory Potential and Induces a Timely Restricted Infiltration of Immune Cells in Colon Cancer Tumors. *Front Immunol* **8**, 231.

11 ACKNOWLEDGEMENTS

First, I am very thankful to Professor Roland Rad for giving me the fantastic opportunity to perform my PhD research in his lab. I would like to thank him very much for offering me such interesting and outstanding projects, for always being open to discussions and ideas and for supporting my work in any way possible. I definitely consider it a privilege to have been one of his first PhD students!

I am also very grateful to Professor Angelika Schnieke for being my second PhD advisor and to Professor Dieter Saur for being my mentor during my PhD research. I would like to thank both of them for their invaluable scientific input and great interest in my work.

I would also like to thank Professor Allan Bradley for hosting me in his lab during the first months of my PhD, for giving me the opportunity to gain profound insights into the transposon technology and for the long lasting productive collaboration. Many thanks also to all members of his lab for their help and support!

I am also thankful to Professor Roland Schmid for giving me the opportunity to work in his department and to all the coordinators of DKTK Munich.

Furthermore, I am also very grateful to all my collaboration partners. Thanks a lot to Katja Steiger and Mathias Heikenwälder (and their teams!) for all their support with pathology and immunohistochemistry. Thank you very much also to Ursula Ehmer, Leticia Quintanilla de Fend and her team, Jorge de la Rosa, Juan Cadiñanos, George Vassiliou and all the others who contributed to this work.

I am very thankful to all members of the Rad lab for the great working atmosphere, the helpful discussions and all the support during the years. Thank you very much to Rupert Öllinger for working together on the CRISPR project and sharing his scientific knowledge with me. I would also like to thank Thomas Engleitner for all his bioinformatics support during the years. Many thanks also to Anja Pfaus and Robert Lersch for working together on the lymphoma projects!

Finally, and most importantly, this dissertation would not have been possible without the constant help and tremendous support from my friends and family. Thank you so much!