Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

ALEXANDROS PANAGIOTIS DRAINAS

M.Sc. MOLECULAR BIOLOGY

born in: Ioannina, Greece

DATE OF ORAL EXAMINATION: 25^{th} November 2016

CRISPR/Cas9-Based Approaches for Investigating Mechanisms of Oncogene Activation and Tumor Suppression

Referees:

Dr. Martin Jechlinger

Prof. Dr. Frank Lyko

Abstract

Cancer is largely a disease of the genome. Cancer development is thought to involve the gradual acquisition of mutations that can activate oncogenes and/or inactivate tumor suppressor genes, resulting in a series of genetic changes that stimulate growth, attenuate cell death, destroy checkpoint controls, promote further genetic instability, and enable metastasis.

In the first part of my thesis, I focused on deciphering how activation of oncogenes by structural copy number alterations (SCNAs) that relocate enhancers in close proximity to oncogenes can be achieved, rather than activation by mutation or amplification. This mechanism was recently described as enhancer hijacking (EHJ). I contributed to the understanding of the EHJ mechanism at two genomic loci, *IGF2*, a known oncogenic locus in colorectal cancer and *IRS4*, a gene identified as a top pan-cancer EHJ candidate. To achieve this, I recapitulated the rearrangements associated with EHJ in colorectal and lung cell lines using the CRISPR/Cas9 genomic engineering system and tested for *IGF2* and *IRS4* overexpression, respectively. The rearrangements were successfully reconstructed; however an increase in gene expression was not achieved, suggesting a more complex mechanism of activation or context-dependency than initially anticipated. Investigation of the tumor promoting role of *IRS4* was supported using mouse xenografts, where constitutive overexpression of *IRS4* leads to formation of larger tumors in comparison to control tumors.

In the second part of this thesis I emphasized on the identification of genes, which, when disrupted, lead to sustained cell growth and can be potential tumor suppressors. To achieve this, I employed systematic screens on cells with different genetic backgrounds using a combination of CRISPR/Cas9-based whole genome knockout libraries and the powerful anchorage independent growth assay. I was able to verify known tumor suppressor genes, which include components of the Hippo and mTOR pathways, as well as to identify novel candidates including *FRYL* and *AHR*. Furthermore, a growth screen under non-selecting conditions was performed and identified numerous candidates found in the initial anchorage independent growth screen, which further supports the growth promoting roles of the candidate genes. In conclusion, in my study I identified potential tumor suppressors that lead untransformed cells to enhanced as well as anchorage independent growth.

Zusammenfassung

Krebs ist größtenteils eine Erkrankung des Genoms. Man geht davon aus, dass sich Krebs durch eine schrittweise Folge von Mutationen entwickelt, die über die Aktivierung von Onkogenen und/oder Inaktivierung von Tumorsuppressorgenen schlussendlich zur Wachstumsstimulierung, Apoptosehemmung, Aufhebung der Checkpoint-Kontrolle, Förderung der genetischen Instabilität und Metastasierung führt.

Im ersten Teil meiner Doktorarbeit beschäftigte ich mich mit der Frage, wie durch strukturelle Kopiezahlveränderungen (SCNA), durch die Enhancer in die Nähe von Onkogenen verlagert werden, Onkogene aktiviert werden können. Dieser Mechanismus wurde vor kurzem als Enhancer Hijacking (EHJ) beschrieben. Mit meiner Arbeit habe ich zum Verständnis des Mechanismus von EHJ am Beispielen zweier genomischer Loci beigetragen: *IGF2*, ein bekannter onkogener Lokus in kolorektalen Tumoren und *IRF4*, dessen Gen als ein Hauptkandidat für EHJ im Pan-Cancer-Kontext identifiziert wurde. Mein Ansatz dabei war, die mit EHJ assoziierten strukturellen Rearrangierungen zu erfassen, die in CRISPR/CAS9-editierten kolorektalen und Lungen-Krebszelllinien zur Überexprimierung von *IGF2* und *IRS4* führen. Die Rearrangierungen wurden erfolgreich rekonstruiert, es konnte jedoch kein Anstieg in der Genexpression festgestellt werden. Das alles deutet darauf hin, dass ein komplexerer Aktivierungsmechanismus als ursprünglich angenommen, verantwortlich ist oder die Ergebnisse vom Modellsystem abhängig sind. Im Anschluss wurde die tumorpromovierende Rolle von *IRS4* in einem Maus-Xenograft-Modell herausgestellt, wobei die konstitutive Überexprimierung von *IRS4* zur Entstehung größerer Tumore im Vergleich zur Kontrolle führt.

Im zweiten Teil meiner Arbeit lag der Fokus auf der Identifizierung von solchen Genen, die durch genetische Veränderung zu anhaltendem Zellwachstum führen und die deshalb potentielle Tumorssuppressoren sind. Für systematische Screens in Zellen mit unterschiedlichem genetischem Hintergrund wurde ein kombinierter Ansatz von CRISPR/Cas 9-basierten Whole-genome-knockout-Bibliotheken und dem verankerungsunabhängigen Wachstumsversuch genutzt. Dabei konnte ich ein bekanntes Tumorsuppressorgen, das Komponenten des Hippo und mTOR Pathways einschließt, bestätigen, aber auch neue Kandidaten einschließlich *FRYL* and*AHR* identifizieren. In einem Wachstumstest unter nichtselektiven Bedinungen wurden mehrere Kandidaten ermittelt, die im vorangegangenen verankerungsunabhängigen Wachstumsversuch bereits identifiziert worden sind.

Dies spricht dafür, dass die Kandidatengene wachstumsfördernde Auswirkungen haben. Zusammenfassend ergeben sich aus meinen Untersuchungen potentielle Tumorsuppressoren, die untransformierte Zellen zu erhöhtem wie auch verankerungsunabhängigem Wachstum anregen.

Acknowledgements

First and foremost I would like to thank my supervisor Dr. Jan Korbel who gave me the opportunity to conduct my PhD in his laboratory. With constant encouragement and excitement he showed me how to pursuit projects and outcoming results. He gave me the chance to familiarize myself with the growing field of bioinformatics and its interplay with experimental molecular biology, a strong combination in today's science. Jan mentored, supported and motivated me to explore and investigate various ideas during my research, which I greatly appreciated.

Besides my supervisor, I would like to acknowledge my Thesis Advisory Committee (TAC) members Dr. Furlong, Dr. Haering, Prof. Schiebel and Prof. Lyko who provided me with useful comments and suggestions in every TAC meeting. These meetings played a highly important role in the outcome of my thesis since all of them were important decision-making timepoints. Also, I am very grateful to Dr. Jechlinger who was able to become the first examiner for my thesis and defense in short notice.

Next, I would like to express my deep gratitude to Balca who supported me in daily bases throughout my PhD. Balca supervised me, helped me design and conduct experiments and introduced me to several experimental techniques. She taught me how to think critically about my research and draw conclusions carefully. Balca vitally influenced my development as a scientist.

Furthermore, I am grateful to Joachim, with whom I strongly collaborated in the enhancer hijacking project throughout most of my PhD. During numerous scientific discussions, even over a beer sometimes, we tried to figure out how to solve the mysteries of the results we were confronted with. Working with Joachim was exciting and fun and I gathered a lot of knowledge from his expertise.

Many thanks to Sebastian who helped me during several projects, especially when involving bioinformatics. He provided many creative ideas and advice important for the development of my work. Sebastian's passion for science inspired me moreover to see biological phenomenons from different angles and discuss their meaning in great detail.

Specifically I would like to thank Adrian and Benjamin who supported me throughout many experiments, especially towards the end of my PhD. Benjamin's thorough experimental abilities helped bringing projects forward in an optimized matter. Adrian's help,

support and advice cannot simply be put in a few words. He advised me how to unbiasedly interpret results and how to correctly evaluate them. In several occasions he pointed out when it was worth pursuing a result and when to concentrate on next steps. Even Friday results, or "turning green" had their meaning and aside of science we had several good laughs.

Also thanks to Charis, a friend with whom we studied together from the diploma studies through masters up to the PhD. His cheerful and positive mood made even bad days into a fun experience and I enjoyed the numerous serious scientific discussions about crazy ideas of his.

I would also like to thank certain laboratory members as well as EMBL colleagues that helped me throughout my PhD in diverse matters. Chris, Elli, George, Hernando, Markus, Mike, Nina, Sasha, Serap, Stephanie, Ruxandra, Thomas and Tobias all played an important role in the development of my thesis. Along with them I would like to thank the rest of the Korbel laboratory members for stimulating scientific discussions.

I greatly appreciate the help of the EMBL facilities, such as the FACS, the Gene Core, ALMF, IT and as well as the rest of the EMBL facilities for their great assistance when needed.

Additionally I would like to thank my friends Flo, Michi, Nik, Ola, Hernando, Gemma, Ana Rita and Simone for the flauben times as well as Manu, Paola, Moritz, Doro, George and Vaso for at the great times we had together. They made life in Heidelberg a great experience.

Furthermore I would like to express my gratitude to my family for their provided support at any time needed. My mother Virginia and my siblings Ari and Elaine supported me throughout my PhD. Also, I would like to thank my uncle Denis and our family friend Jim for the scientific discussions and help especially before big presentations. Especially I would like to thank my father who played an essential role for my development as a scientist and as a person. The critical thinking and life values he passed on to me, brought me where I am and will guide me to where I will continue. He will always be remembered.

And finally I would like to thank my wife Sarah who supported me throughout my whole PhD, who encouraged me in the difficult times and shared excitement in the

positive times. She always gave me feedback when I needed it from a different scientific perspective that stimulated my thoughts and made them more concrete. She supported my work but also made me realize when to set it aside and take a break to enjoy activities and trips we planned. Together now with our expected son Constantin we are looking forward to upcoming adventures.

To my family

Contents

Abbreviations

ANIG	Anchorage Independent Growth
bp	base pair
CAST	Complex Alterations after Selection and Transformation
cDNA	Complementary Deoxyribonucleic Acid
ChIP	Chromatin Immunoprecipitation
chr	Chromosome
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
DNA	Deoxyribonucleic Acid
FACS	Fluorescent Activated Cell Sorting
FDR	False Discovery Rate
FISH	Fluorescent *in situ* hybridization
GeCKO	Genome-wide CRIPSR/CAS9 Knock-Out
GFP	Green Fluorescent Protein
GI	Genetic Interaction
HR	Homologous Recombination
Kb	Kilobase(s)
Mb	Megabase(s)
MEM	Mixed Effect Model
MOI	Multiplicity of Infection
NHEJ	Non-Homologous End Joining
PCR	Polymerase Chain Reaction
PM	Point Mutation
SCNA	Structural Copy Number Alteration
TAD	Topological Associated Domain
TALEN	Transcription Activator-Like Effector Nuclease

TCGA	The Cancer Genome Atlas
ZNF	Zinc-Finger Nuclease

Chapter 1

Introduction

Throughout evolution, mammalian genomes have formed various regulatory programs to ensure the precise development and propagation of cells in an organism. In most multicellular organisms, including mammals, proper orchestration of these regulatory programs are necessary to help development during which they specify cell types and determine the body plan of the organism. Therefore, changes in the genome are a great threat to the integrity of cell growth. To ensure that the genome is not compromised, cells have evolved several repair mechanisms, in case errors arise. Although the regulatory networks are redundant and robust, they can still encounter errors that jeopardize their integrity and function. Possible layers of prevention in case of malfunction are self-destruction, senescence or recruitment of cells that will terminate dysfunctional cells, in order to preserve a healthy equilibrium. Even though there are many barriers preventing a cell from going awry, there is still the possibility of failure, leading the cells to a selfish path of uncontrolled growth. This path of cells gone astray is a disease of the genome known as cancer.

Cancer is thought to occur through a gradual accumulation of somatic mutations resulting in uncontrolled cell growth (Cavenee and White, 1995). Throughout the evolution of the cancer genome, cells find diverse ways to induce aggressive proliferation, resulting in a large population of transformed cells, some of which will inevitably invade multiple tissues (Martincorena and Campbell, 2015). Over a hundred different types of cancer exist, which may arise from nearly any cell type (http://www.cancer.gov/types). Although

continuous growth is the common characteristic of all cancers, mutational processes that affect growth are mostly different and specific for each cancer type.

Up to date there has not been any "magic bullet" to target all the different vulnerabilities of each cancer. To better understand these, investigation of multiple mutational processes is required. In this thesis, I investigated the mechanisms of oncogene activation by introducing perturbations in regulatory regions around certain oncogenes. Additionally, I sought to understand the initial steps of cancer development by identifying tumor suppressors using an unbiased, genome-wide approach.

1.1 Hallmarks of Cancer

For a cancer cell to evolve, it needs to overcome several barriers and adopt certain non-physiological traits. These include sustained proliferative signaling, evasion of growth suppressors, resistance of cell death, enabling replicative immortality, induction of angiogenesis and activation of invasion and metastasis (Hanahan and Weinberg, 2000). Furthermore, tumor cells will typically alter the microenvironment resulting in the promotion of inflammation, the deregulation of their cellular energetics and the avoidance of immune destruction (Hanahan and Weinberg, 2011). Genomic alterations including point mutations (PMs) or structural variations (SVs) in tumors underlie these events and generate the genetic diversity that enables cells to become malignant (Hanahan and Weinberg, 2011) (Fig. 1.1).

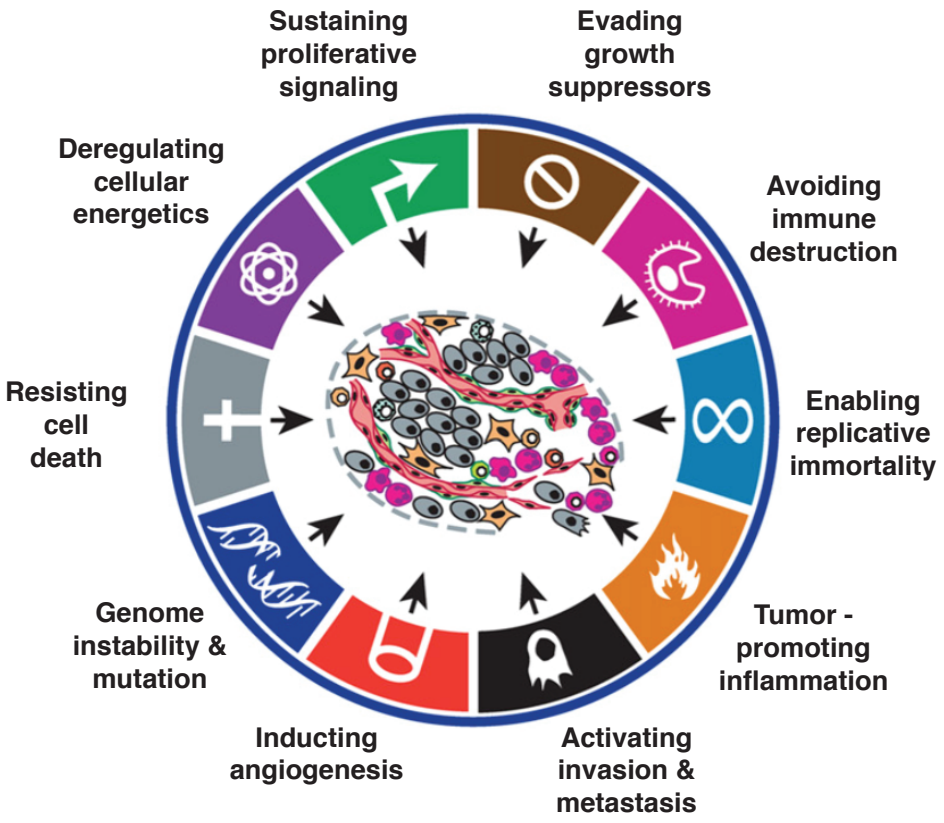


FIGURE 1.1: Cancer Hallmarks.

1.1.1 Hallmarks capabilities of sustainability and tumor growth

A simplistic model of tumor formation (tumorigenesis) is caused by defects in tumor suppressors or proto-oncogenes. Tumor suppressors are important to restrain uncontrolled growth whereas oncogenes (altered proto-oncogenes by mutation causing constitutive activation) perform the opposite: they promote growth. Both tumor suppressors and oncogenes act on various layers of cell regulation, which effect different cancer hallmarks. Strategies to investigate the cancer roles of genes include analysis of the mutational rate in cancer genomes, investigation of the growth promoting effect and transformation in cell lines, xenograft models with cells with the altered gene, modification of the gene in *in vivo* organisms, etc. This deregulation leads to uncontrolled growth and eventually metastasis – a state where cells detach from the main tumor and circulate to a new location (Fig. 1.1 and 1.2).

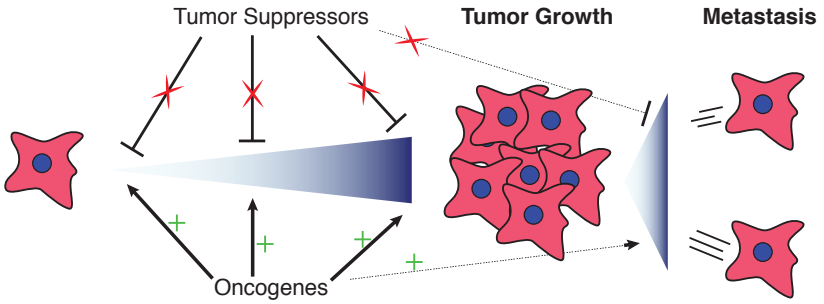


FIGURE 1.2: Tumor Suppressors and Oncogenes. Tumor Suppressors fail to inhibit growth and metastasis while oncogenes promote it.

Cells have evolved several mechanisms to preserve and tightly regulate the cell state in order to restrain uncontrolled growth. Therefore, initial attributes cancer cells acquire are alterations in pathways that sustain their living state as well as promoting growth. One of these attributes is the ability to resist cell death whereby the cell builds up the potential of constant proliferation without activating pathways that lead to self-destruction (i.e. apoptosis). One of the main players in activating apoptosis is a tumor suppressor gene known as the guardian of the genome, *TP53* (Lane, 1992; Vousden and Prives, 2009). *TP53* is a transcription factor and is known to activate apoptosis by several mechanisms, such as activation of Apaf-1 (a coactivator of caspase-9) (Kannan et al., 2001), upregulation of expression of caspase-6 (MacLachlan and El-Deiry, 2002), or activation through cytochrome C release (Schuler et al., 2000), all which lead to the activation of the caspase cascade. Caspases have cysteine protease activity that cleaves targeted proteins and plays a core role in the apoptosis pathway (Fridman and Lowe, 2003).

Furthermore, *TP53* in addition to its ability to lead the cells to apoptosis is responsible for cell cycle arrest in response to damage as well as senescence and therefore considered as a growth suppressor. For example, in DNA damaged cells, *TP53* gets activated in order to enable the cells to repair their genome before continuing the cell cycle. This allows mutations that would potentially activate oncogenes to get repaired, before passing the damaged genome onto daughter cells. If the DNA damage is severe, the cells might be driven to senescence, a state where the cell cycle arrests irreversibly.

Additional functions of *TP53* are its association with metabolism, autophagy and stem cell biology, which also play important roles in tumorigenesis (Bieging et al., 2014).

Previous studies have shown that complete dysfunction of *TP53* leads to 100% penetrance in tumor formation in mice (Donehower et al., 1992). However, *TP53* is not always inactivated, but some tumors harbor point mutations in domains that disrupt *TP53* partially which can then be advantageous for the tumor (Zhu et al., 2015a). Analysis of several tumor types have shown to have *TP53* mutated with a penetrance up to 80%, such as in breast basel carcinoma (Hoadley et al., 2014). Lastly, the whole TP53 pathway is found to be the most commonly affected pathway across cancers (Stracquadanio et al., 2016). In summary, *TP53* is involved in several pathways that sustain the cell state, control growth and cell death and therefore is considered to be a major tumor suppressor gene (Bieging et al., 2014; Cosme-Blanco et al., 2007; Liu et al., 2004).

Apart from tumor suppressors that are responsible for activating apoptosis or restraining growth, oncogenes also play an important role in tumorigenesis. Oncogenes can be classified in six groups: transcription factors, chromatin remodelers, growth factors, growth factor receptors, signal transducers and apoptosis regulators (Croce, 2008). The first example of an oncogene was the the constitutive activation of the *SRC* gene (Stehelin et al., 1977). *SRC* was discovered through the formation of tumors by the Rous sarcoma virus in chickens (Stehelin et al., 1977). *SRC* plays an important role for signal transduction in many pathways affecting proliferation, differentiation, cell adhesion and migration (Parsons and Parsons, 2004). Oncogenic *SRC* harbors mutations that result in a constitutively active SRC protein. Hence, the cell receives constant growth signaling through oncogenic SRC resulting in tumorigenesis.

Another important factor in sustaining growth is replicative immortality. Generally, cells have a limited number of divisions due to telomere shortening. Telomeres are repetitive sequences in the end of each chromosome to protect them from fusion and shortening. When telomeres become too short, cells are driven to senescence or to crisis (d'Adda di Fagagna et al., 2003; Hayashi et al., 2015; Maciejowski et al., 2015). In cancer though, cells achieve a state of immortalization with infinite divisions, which is linked to telomere restoration (Blasco, 2005). The main gene responsible for telomere lengthening is telomerase and it is overexpressed in approximately 90% of cancers (Hanahan and Weinberg, 2011; Kim et al., 1994). The remaining cancers have activated alternative lengthening pathways of telomeres termed ALT (Alternative Lengthening of Telomers) (Henson et al., 2002).

Besides its role in cancer evolution, telomerase immortalization has thoroughly been used in establishing cell lines in order to achieve unlimited proliferative potential. These immortalized cell lines can be used to investigate hypotheses of cancer development and are therefore of high importance.

1.1.2 Cancer epistasis

Tumor genomes typically evolve by combinations of genomic alterations rather than changes acting on single genes. *MYC* and mutant *RAS* both known oncogenes, were the first example to demonstrate such synergistic effect. Whereas overexpression of either *MYC* or *RAS* was unable to generate tumors, the combination of both oncogenes was able to generate fast growing tumors in mice, indicating their cooperative actions (Land et al., 1983). Since then, such interactions have been found to be quite common in cancer, contrary to initial expectations (Wang et al., 2014b).

This type of gene interaction (GI) is known as epistasis and describes the situation where a resulting phenotype of a gene is dependent on another gene (Park and Lehner, 2015). There, the observed phenotype of these GIs is not additive (i.e. adding the effect of both genes) but can be in some cases synergistic, for example in proliferation (Fig. 1.3). Genes that express proteins involved in same pathways are considered to have masking effects since the outcome of alterations of either or both genes results in the same phenotype (Wang et al., 2014b). Masking effects were found to be more common than epistatic effects when investigating mutated cancer genes in breast cancer, indicating that mutated genes are frequently involved in the same pathways (Wang et al., 2014b).

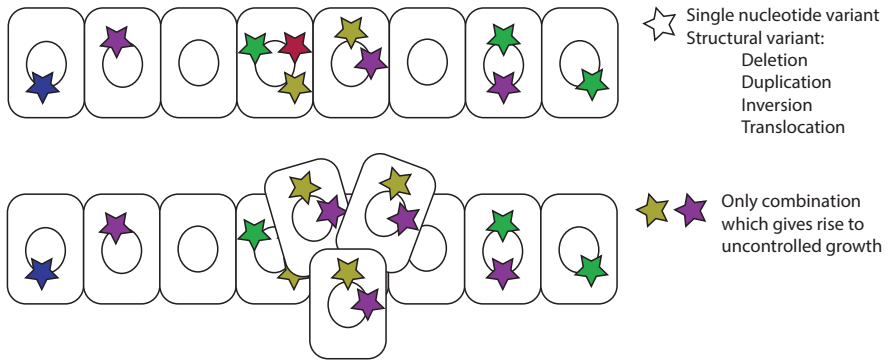


FIGURE 1.3: Epistasis resulting in proliferation.

As described, epistasis plays an important role in cell growth. Epistatic effects can also be exploited to inhibit growth. An analogous concept to epistasis in treating cancers is that of synthetic lethality. Synthetic lethality describes when the inhibition of two genes in combination lead to cell death while the individual inhibition of either does not, or has a weaker effect. Knockdown screens have shown that *RAS* oncogenic mutant tumors have a synthetic lethal effect in combination with inhibition of the APC/C (the anaphase promoting complex), among several other candidates (Luo et al., 2009). Therefore, synthetic lethality is an attractive approach for tumor treatment (Astsaturov et al., 2010; Kaelin, 2005; Luo et al., 2009).

Furthermore, data mining projects have attempted to investigate the co-occurrence of tumor suppressors and oncogenes in tumors. Several identified combinations of tumor suppressors and oncogenes were linked and simultaneously mutated in tumors (Zack et al., 2013; Zhu et al., 2015b).

A good example of a GI that highlights the importance of epistasis in cancer was identified between *Brac1* and *Trp53* (Xu et al., 2001). Mutant *Brca1* in mice (exon 11 deficient mutant) could not give rise to viable mice due to activation of apoptosis. By depleting *Trp53*, mice were viable rescuing the *Brca1* deleterious phenotype. This GI has importance for tumorigenesis since *Brca1* repairs double strand breaks in the chromosomes and *Trp53* activates apoptosis, resulting in a deficient combination that favors tumor cell evolution.

Epistasis therefore has an important effect in tumors and studying modification of cells in different genetic backgrounds can shed light in previously unidentified GIs. Understanding the forces that enable the cells to overcome the initial barriers for tumorigenic growth is therefore essential to understand cancer initiation and develop further therapeutic strategies. For this reason, one of my aims was to identify GIs of tumor suppressors of initial driver events that result in uncontrolled cell growth.

1.1.3 Genomic alterations

Genomic alterations in the form of PMs and SVs are important hallmarks that can enhance tumor evolution by effecting gene expression. PMs and SVs could be either acquired gradually, as the classical cancer model describes (Cavenee and White, 1995),

or in bursts, where multiple alterations could happen under one event (Nik-Zainal et al., 2012; Stephens et al., 2011). These two different routes may compliment each other, for instance with gradual mutation accumulation being punctuated by periodic mutational bursts (Yates and Campbell, 2012) (Fig. 1.4).

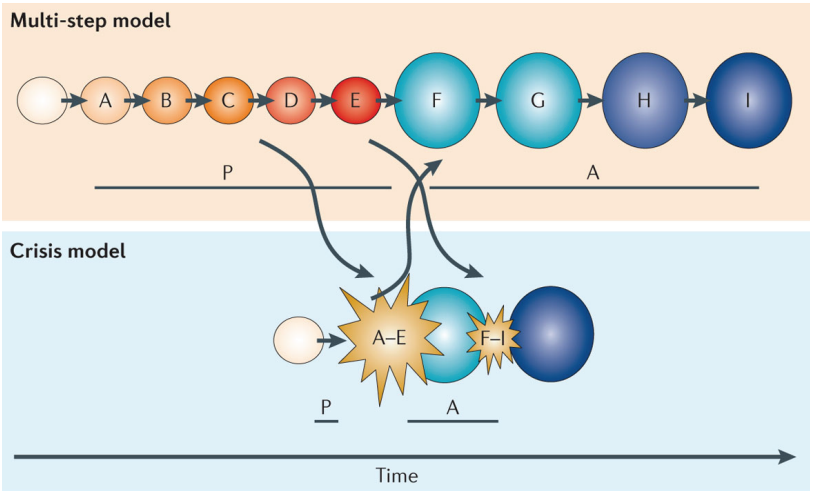


FIGURE 1.4: Gradual and crisis tumor evolution. Cancer genomes can evolve gradually or in bursts. Interplay between these two routs could also influence cancer evolution. Figure adapted from Yates and Campbell (2012).

PMs, alterations of a single bases, can lead to activation of proto-oncogenes or inactivation of tumor suppressors. Constitutive activation of oncogenes has been observed in many cancers such as a point mutation in *KRAS* (McCoy et al., 1984) that results to altered amino acid leading to its activation or point mutations in *EGFR* (Lynch et al., 2004) resulting in alterations in the ATP-binding pocket of the tyrosine kinase domain, both of which result in stimulation of growth in the same pathway. In contrary to oncogene activation, PMs can inactivate tumor suppressors by introducing a stop codon within the coding sequence, such as in *RB*, a major G1 checkpoint gene, leading to the production of truncated, non-functional protein (Horowitz et al., 1989). Furthermore, instead of PMs affecting genes, they may also influence activation of genes by altering promoter sequences, allowing new transcription factors to bind, for example seen in melanomas with the *TERT* promoter (Huang et al., 2013).

In contrary to PMs, larger rearrangements, structural variations (SVs) can result in alterations of several genes at once. SVs can occur as deletions duplications/amplifications, inversions and translocations (Weischenfeldt et al., 2013; Zhang and Pellman, 2015).

Deletions involve a loss in genetic material and may be as large as whole chromosome or focal and can contain tumor suppressors (Yin and Shen, 2008). Duplications and amplifications provide the cell with additional copies of a genomic locus or whole chromosomes, which could give rise to numerous copies of oncogenes (Albertson, 2006; Saliba et al., 2015). Inversions can switch the orientation of genes. Although they do not alter the dosage of genetic material they can modify the chromosome structure, leading to changes in regulatory landscapes (e.g. enhancer positioning) which may lead to changes in gene expression (Roberts et al., 2002). Translocations also result in shuffling material of different chromosomes and are linked to cancer by causing instability (Mitelman et al., 2007; Stankiewicz and Lupski, 2002). Complex karyotypes containing many SVs have been associated with poor prognosis in cancer (Zack et al., 2013). Additionally, several SVs have shown to be recurrent even across cancers indicating common driving forces for the evolution of some cancer types (Zack et al., 2013).

SVs can also result in gene fusions, whereby an oncogene may be overexpressed due to the stitching of its loci to a non-tumorigenic, but highly expressed, gene. These chimeric genes have been shown to be responsible for the outcome of several cancers making gene fusion a common mechanism of oncogene activation (Mertens et al., 2015; Yoshihara et al., 2015).

Shuffling of the genome has not only shown to cause direct alterations on the genes but also in the genomic neighborhood affecting enhancers. Enhancers are genomic sequences that employ transcription factors to activate the expression of a nearby gene. Clustered enhancers in a region are commonly referred as super-enhancers and are considered to drive strong gene expression (Hnisz et al., 2013). Therefore SVs can shuffle enhancers to new targets, which can activate proto-oncogenes (Gröschel et al., 2014). This has been thoroughly investigated for group III type medulloblastoma for the activation of the *GFI1* gene through different types of rearrangements (duplications, inversions and deletions) (Northcott et al., 2014). The investigation of SVs of the non-coding genome and deciphering their mechanism of action was one of the focuses of this thesis.

Independently of the path tumors take to evolve, each cell in the tumor population acquires its own set of mutations, which results in heterogeneity within the population. Beneficial mutations can lead to clonal expansions that may outgrow the rest of the cells under different conditions or stresses. Genomic instability enhances heterogeneity and

can serve as an indicator of the potency and aggressiveness of the tumor and also of its drug resistance. Hence, heterogeneity, a result of many mutational processes, is crucial for tumor growth and also a great burden for cancer treatment (Fisher et al., 2013).

1.2 Genetic Engineering

Due to advances in technology, especially in the last decade, genetic engineering has become a vital procedure in research. Genetic engineering is a process of precisely modifying the genome of an organism in order to alter properties of the cell or the organism. Main uses involve loss of function and gain of function experiments, which result in complete loss or modification of gene expression in cells. Another application is gene tagging, which can be utilized for protein localization and trafficking experiments, flow cytometry and cell sorting experiments as well as chromatin immunoprecipitation (Dominguez et al., 2016).

One of the first model organisms that genome engineering was thoroughly applied to was budding yeast, due to their efficient homologous recombination (HR) repair mechanism which allowed for efficient integration of foreign genetic material (Orr-Weaver et al., 1981). Since then yeast genetics have become a very powerful system to study gene function. Until recently, establishing an efficient system in other organisms had been very challenging. Tools that played an immense role in genetic engineering are restriction enzymes also called "molecular scissors", which introduce breaks in the DNA in precise locations. The cell will repair the breaks by mechanisms such as non-homologous end joining (NHEJ) and HR (Sancar et al., 2004). Repair by NHEJ can result in alteration of the DNA sequence and produce a frame-shift in a coding region. HR is a more precise repair mechanism and by using a repair template can lead to a specific genomic edit.

1.2.1 Programmable nucleases

The use of nucleases greatly increases the efficiency for genetic engineering. However, common restriction enzymes cut the genome too frequently, which makes them impractical for genetic editing. Therefore, nucleases that cut the genome with high specificity are the preferred enzymes to alter the genome. One of the first class of nucleases used for

this purpose were meganucleases, which recognize a large sequence in the genome and introduce a double strand break (Epinat et al., 2003). Later, by editing the sequence of the meganucleases, they could be programmed to target different sites (Epinat et al., 2003). Despite the potential of meganucleases precise editing was very challenging. The further investigation of new classes of engineered nucleases, zinc-finger nucleases (ZNFs) and transcription activator-like effector nucleases (TALENs), enhanced the field of genetic engineering.

ZNFs make use of fusing a zinc-finger DNA binding domain to a non-specific nuclease domain of the FOKI restriction enzyme (Kim et al., 1996). Zinc-finger proteins utilize zinc to fold the protein and have been identified to bind DNA, RNA, proteins, etc. (Gamsjaeger et al., 2007). ZNFs can be engineered and used in tandem to recognize and break specific DNA sequences (Urnov et al., 2010). Therefore ZNFs can be directed to specific sites in the genome (Fig. 1.5a).

TALENs are another class of DNA binding proteins, which similar to ZNFs, are fused with FOKI nuclease. Transcription activator-like effectors (TALEs) are proteins secreted by a bacteria species called *Xanthomonas*. TALEs contain highly variable sequences and are shown to bind strongly to specific nucleotide sequences. The variable sequence they carry is termed repeat variable di-residues (RVDs) which can bind specifically to certain nucleotides (Boch et al., 2009) and hence can be programmable to target specific sequences. Therefore TALENs can easily direct the nuclease domain to specific sequences in the genome and introduce a break (Fig. 1.5b).

A promising successor in the genetic engineering field after the programmable TALENs is a nuclease found in bacteria named Cas9, which is a core component of the bacterial adaptive immune system Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) (Fig. 1.5c).

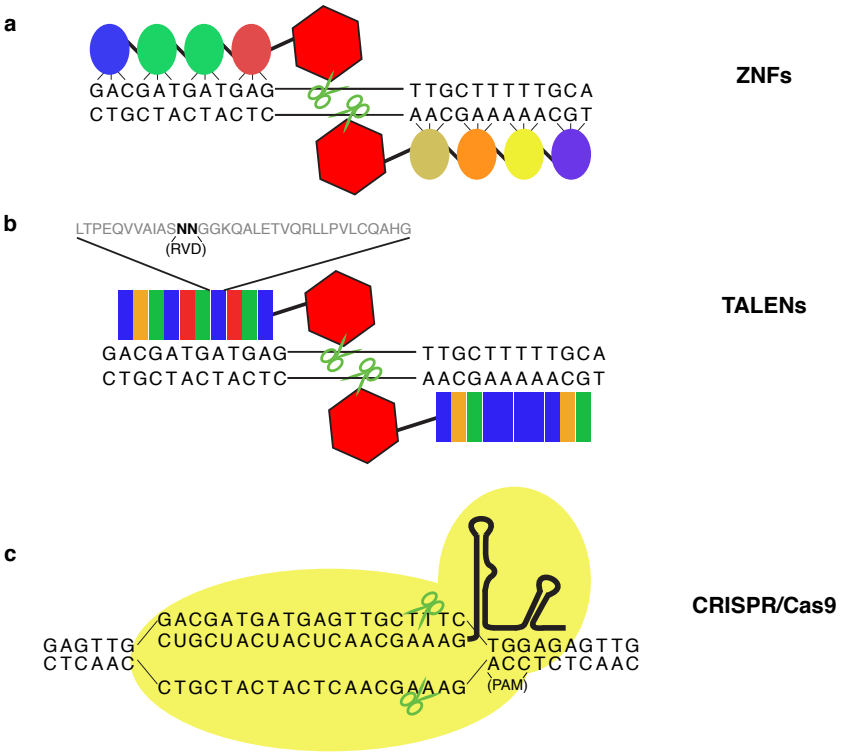


FIGURE 1.5: Programmable nucleases. (**a**) Genome targeting by zinc-finger nucleases (ZNFs). Zinc-finger domains recognize a set of three DNA bases. When the whole Zinc-finger array recognizes its target the attached FOKI nucleases can introduce breaks on the DNA. (**b**) Genome targeting by transcription activator-like effector nucleases (TALENs). Each domain in the TALEN recognizes a specific base according to the repeat variable di-residues (RVDs) domain. Similary when the whole array is attached to the DNA target the FOKI nucleases introduce breaks on the DNA. (**c**) Genome targeting by CRISPR/Cas9 nuclease. A gRNA guides the Cas9 nuclease to a targeted site. Firstly, the Cas9 recognizes a protospacer adjacent motif (PAM) with sequence of "NGG" and then if the gRNA is compliment to the adjacent DNA sequence, the Cas9 introduces a double strand break.

1.2.2 CRISPR/Cas9 based genetic engineering

CRISPR is a defense system found in multiple bacteria species (Doudna and Charpentier, 2014; Wright et al., 2016). The CRISPR locus is defined by palindrome repeats separated by short sequences called protospacers (Fig. 1.5c). The transcribed protospacer sequences coupled with CRISPR associated proteins (Cas) bind uniquely to a targeted DNA sequence, which is then cleaved by the nuclease domain of the Cas protein. The protospacer sequence is integrated by specialized Cas proteins, inserting pieces from the foreign DNA into the genome (Barrangou et al., 2007; Datsenko et al., 2012; Nuñez et al., 2015). The repeats and protospacer sequence are transcribed and matured and guide the nuclease Cas protein to cut foreign DNA. By this means, bacteria can acquires adaptive immunity to the foreign DNA, particularly that of invading bacteriophages. Although

the bacteria host transcribes the protospacer sequences, the Cas proteins do not target the bacteria's genome, due to the absence of the protospacer adjacent motif (PAM). PAM is a short sequence found adjacent to the target site of the protospacer motif. The Cas protein firstly recognizes the PAM sequence and then the protospacer sequence attempts to anneal at the targeted locus. A cut is only introduced if the protospacer anneals perfectly. Therefore PAMs are essential for the CRISPR system since they provide an elegant system to distinguish between the bacteria genome and the invading species.

The CRISPR bacteria immune system is classified into five types (Makarova et al., 2015) and all work under the same principle; a nuclease guided by a transcribed RNA (Barrangou, 2015). The most popular and well studied system is the type II CRISPR/Cas9 isolated from *Streptococcus pyogenes*. It consists of three basic components: the CRISPR associated protein 9 (Cas9), the CRISPR RNA (crRNA) and the trans-activating crRNA (trRNA). The Cas9 binds to a crRNA-trRNA complex and is guided to a targeted site. The Cas9 protein recognizes the PAM sequence of "NGG" and only in those sites it will attempt to anneal the crRNA (Fig. 1.5c). After the correct hybridization of the guide RNA duplex, the Cas9 protein creates a double strand break three bases upstream of the PAM sequence. The Cas9 nuclease domains are RuvC and HNH (Tsai and Joung, 2016) and each creates a nick in the opposite strand.

Advances in the CRISPR/Cas9 system involve the construction of a fused crRNA and trRNA which is called guide RNA (gRNA) (Jinek et al., 2012). This chimeric RNA molecule abolishes the need of a trRNA co-transcribed with the crRNA, constituting a highly efficient molecule to guide the Cas9 nuclease to a targeted site (Jinek et al., 2012). After the break occurs, the cell signals DNA repair pathways to stich back the cut DNA. In mammalian cells the DNA may then be repaired using the error prone NHEJ repair mechanism (Tsai and Joung, 2016). If the DNA is correctly repaired, the active Cas9 can bind again and introduce another cut. Therefore, Cas9 stops cleaving only when the repair pathways introduce errors at the break site, altering the targeted sequence. These errors can include deletions or insertions of a single or multiple base pairs. If the alteration is in a coding region and not in frame with the rest of the coding sequence, the transcript will be translated to a non-functional protein (Fig. 1.6).

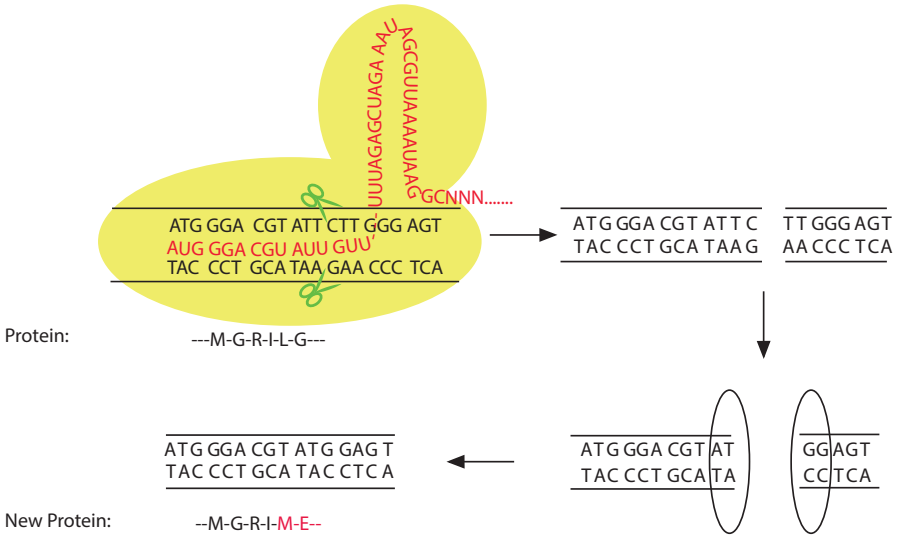


FIGURE 1.6: Erroneous repair. CRISPR/Cas9 targets a specific locus guided by the gRNA and introduces breaks to each DNA strand. Then the repair machinery will repair the break. If the repair is inaccurate, a frame-shift may occur that when translated it will produce a functioning protein.

1.2.3 Further applications in genetic engineering

With the advances in programmable nucleases, genetic engineering has become a routine task in laboratories. One of the main applications is the construction of gene knockout libraries as previously described. Using this approach, large sets of guides can be synthesized and cloned in parallel, creating a pool of guides that can target up to all genes in the human genome (Shalem et al., 2014; Wang et al., 2014a).

Another advance is fusion of genes to fluorophores resulting in tagged proteins. By cutting at the first or last codon of the gene, fluorescent markers can be inserted with a repair template. The repair template consists of homologous sequences to the adjacent locus. Mediated by HR, the cell can insert the repair template in a precise manner and create a gene fusion (Stewart-Ornstein and Lahav, 2016).

Furthermore, combinations of guides can be used to create structural rearrangements in the genome. They can be used to create deletions in exons of genes, or remove large genomic loci. It has also been shown that introducing two guides to cut two adjacent sequences can create a variety of rearrangements, including deletions, duplications, amplifications and inversions (Kraft et al., 2015).

Moreover, gene expression can be regulated by activation or repression according to the protein fused to DNA binding domain of the nuclease. Such DNA binding domain and regulatory domain fusions were initially created with zinc-finger proteins or TALEs (Gersbach and Perez-Pinera, 2014). Since Cas9 proved to be a simple straightforward system, by eliminating the nuclease domains with point mutations, the dead-Cas9 (dCas9) is converted to a DNA binding molecule (Gilbert et al., 2013, 2014; Perez-Pinera et al., 2013). The dCas9 can then be fused to activator proteins resulting in overexpression of genes, or repressor proteins, which can result in controlled reduction of gene expression without perturbing the gene sequence.

Lastly, dCas9 can be tagged with fluorophores and targeted loci can be visualized in living cells (Chen et al., 2013), showing an alternative method to fluorescent *in vitro* hybridization (FISH). In addition to this note, different Cas genes have been identified to target RNA and therefore live RNA-like FISH experiments.

In conclusion CRISPR systems have revolutionized the field of genetic engineering. Elegantly designed older strategies are combined with a simple programmable system that can readily, precisely and in high throughput manner edit almost any loci in the genome. It is expected that many more advances will occur in the next years utilizing CRISPR.

Chapter 2

Recapitulating Genomic Rearrangements of *IRS4* and *IGF2* Implicated in Enhancer Hijacking

This chapter describes the efforts of recapitulating enhancer hijacking related rearrangements of *IGF2* and *IRS4* identified in tumor genomes. Furthermore, the driver role of *IRS4* in tumorigenesis is investigated and supported using mouse xenografts. The results described in this chapter are partly included in the recently submitted manuscript entitled:

Pan-cancer analysis implicates *IRS4* and *IGF2* in enhancer hijacking

Joachim Weischenfeldt*, Taronish Dubash*, **Alexandros P. Drainas***, Balca R. Mardin, Yuanyuan Chen, Adrian M. Stütz, Sebastian M. Waszak, Graziella Bosco, Ann Rita Halvorsen, Benjamin Raeder, Theocharis Efthymiopoulos, Serap Erkek, Christine Siegl, Hermann Brenner, Odd Terje Brustugun, Sebastian M. Dieter, Paul A. Northcott, Iver Petersen, Stefan M. Pfister, Martin Schneider, Steinar K. Solberg, Erik Thunissen, Wilko Weichert, Thomas Zichner, Roman Thomas, Martin Peifer, Aslaug Helland, Claudia R. Ball, Martin Jechlinger, Rocio Sotillo, Hanno Glimm#, & Jan O. Korbel#

* Equally contributed to the project; # Joint senior authors

Contributions

Jan Korbel, Joachim Weischenfeldt and I conceptualized and designed the work presented in this chapter. I designed all used vectors with the support of Balca Mardin and Adrian Stütz. Furthermore, with the help of Joachim Weischenfeldt I designed the screening strategy, which I then conducted together with Adrian Stütz and Benjamin Raeder. I and Benjamin Raeder carried out the majority of the cell culture work with occasional help of Adrian Stütz and Theocharis Efthymiopoulos. The qPCR experiments were conducted by Adrian Stütz. Library preparation for sequencing was performed by Adrian Stütz and Benjamin Raeder. I contributed to the design of the mouse xenograft project in collaboration with Yuanyuan Chen, Adrian Stütz, Joachim Weischenfeldt, Martin Jechlinger, Rocio Sotillo and Jan Korbel. I mainly designed the vectors used in the mouse project with support of Adrian Stütz and Joachim Weischenfeldt. All mouse related work was performed by Yuanyuan Chen and I contributed by conducting downstream experiments on the mouse tumors, interpretation and analysis of results. For the related paper I played a key role in the development of figures. Jan Korbel guided and supervised the project providing valuable and insightful feedback.

2.1 Introduction and Motivation

Up to date, structural variations (SVs) and point mutations (PMs) of the genome have been extensively studied in protein-coding regions. SVs and PMs have shown to promote tumorigenesis in several ways. For example, PMs causing missense or frame-shift alterations can result in hyperactivation of oncogenes or inactivation of tumor suppressors. Larger SVs, such as deletions or amplifications, can remove tumor suppressors or amplify oncogenes respectively. However, only 1% of the genome is translated to proteins, whereas the rest is composed of cis-regulatory elements (Roadmap Epigenomics Consortium et al., 2015; Zhu et al., 2013). Cis-regulatory elements can act from a distance up to several megabases and affect gene expression.

Although the non-coding genome seems to be more tolerant to mutations and somatic variants, recent studies suggested that SVs and PMs could influence the non-coding genome and promote tumorigenesis (Horn et al., 2013; Northcott et al., 2014; Peifer et al., 2015). Such mechanisms may result in generating new binding sites for transcription

factors resulting in increased expression of oncogenes or disruption of transcription factor binding sites at tumor suppressors (Ludlow et al., 1996; Mansour et al., 2014). Recent examples in medulloblastoma, neuroblastoma and leukemia have shown SVs to juxtapose an enhancer to an oncogene, which can result in high upregulation of the respective gene (Hnisz et al., 2016; Northcott et al., 2014; Peifer et al., 2015). This phenomenon has been described as enhancer hijacking (EHJ) (Northcott et al., 2014).

Recently we described the extent of EHJ in a systematic pan-cancer study of 7,423 tumor genomes from 26 tumor types and have shown several genes to be affected by EHJ (Weischenfeldt, Dubash, **Drainas** et al. in revision), suggesting EHJ an important, alternative and frequent mechanism of tumorigenesis.

We employed an approach which applies the statistical concepts of expression quantitative trait locus (eQTLs) and somatic copy number alterations (SCNAs – non-balanced SVs, such as deletions, duplications and amplifications) to identify SCNAs that lead to deregulation of gene expression *in cis* (Fig. 2.1). The principal lies in partitioning the genome based on topological associated domains (TADs) and investigating within these TADs the association between SCNAs and gene expression. TADs are large (estimated average size 1Mb) chromatin interacting units that have high levels of chromatin associations in comparison to the rest of the genome and are thought to help mediate control of gene regulation (e.g. enhancer-promoter interactions) within the 3D structure of TADs (Dixon et al., 2012). TADs can be identified based on Hi-C experiments – a technique that can capture physically interacting chromatin in 3D space (Dixon et al., 2012; Pombo and Dillon, 2015). Comparison and association of SCNAs and TADs was applied for each cancer individually as well as across multiple cancers (pan-cancer). The SCNAs were defined based on SNP6 arrays, which are DNA microarrays that can identify single nucleotide polymorphisms (http://cancergenome.nih.gov/).

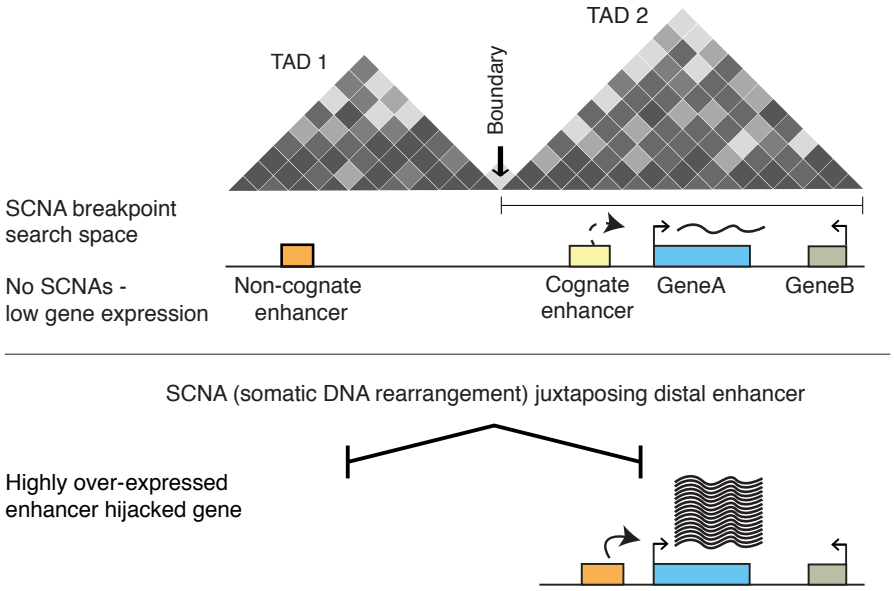


FIGURE 2.1: The principle behind the approach used for identification of SCNAs associated with overexpression of genes in close proximity is depicted. Topological associated domains (TADs) are shown from Hi-C contact maps (a technique that identifies long range interactions). The darker shading indicates stronger interaction. The search space of the method used is according to TAD size. After a rearrangement a non-cognate enhancer can be relocated next to a gene and drive its high expression. Figure was adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

2.1.1 Colorectal cancer analysis implicates IGF2 in enhancer hijacking

Among the different tumor types analyzed, we identified colorectal as the most affected tumor type of potential EHJ events. The top candidate was *IGF2* which was associated with a cluster of SCNAs that led to its upregulation. *IGF2* belongs to the insulin family of growth factors. *IGF2* is involved in development and growth but was also suggested to act as an oncogene in cancer when misregulated (Brouwer-Visser and Huang, 2015; Cui et al., 2003).

Specifically, we identified a cluster of duplications encompassing *IGF2* which was associated with high upregulation (up to 1000 fold) of *IGF2* (Fig. 2.2). The high levels of *IGF2* overexpression exceeded the amount that would be overexpressed by a dosage model. A dosage model would only result to an overexpression of 1.5 fold (since the genome will have three copies instead of two), which is much lower than the identified *IGF2* levels observed, suggesting an alternative model of activation. Apart from duplications, deletions were also identified in close proximity to *IGF2* that led to its high upregulation (deletions depicted in Fig. 2.7).

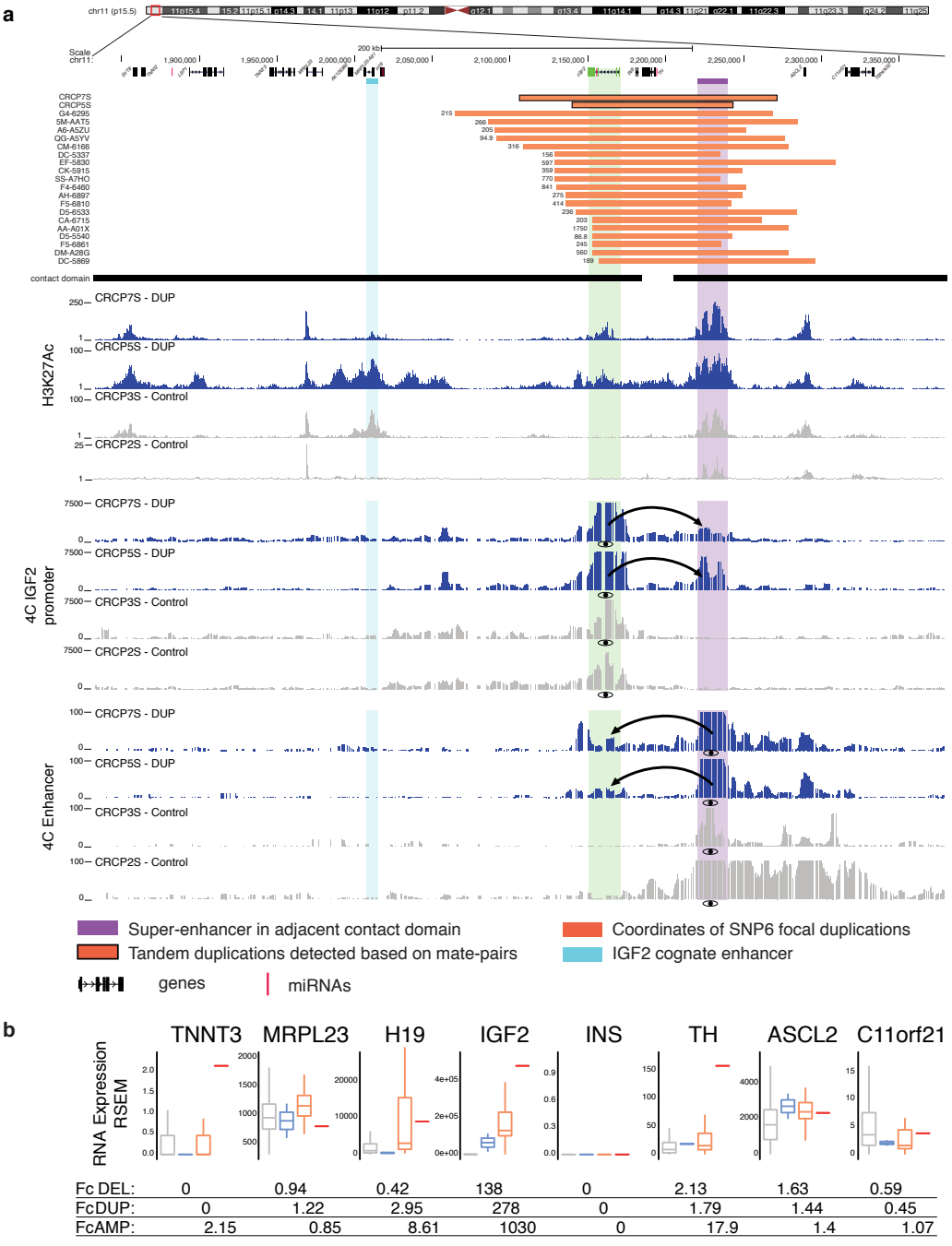


FIGURE 2.2: *IGF2* enhancer hijacking. (**a**) Active chromatin marks, H3K27Ac (Ernst et al., 2011), show signals at the IGF2 locus in the samples that carry the duplication but not in the controls. Next, 4C-Seq experiments indicate a physical interaction between *IGF2* and the super-enhancer in samples that carry the duplication but not in the controls. The opposite, using the super-enhancer as viewpoint verifies the specific physical interaction with *IGF2* in the samples that carry the duplication. (**b**) The boxplots show the relationship between expression and SCNAs for all genes within the respective TAD. *IGF2* has the highest expression in the samples with the recurrent SCNAs *in cis* but not in the control samples lacking SCNAs. The boxplots depict deletion (DEL), duplication (DUP) and amplification carriers (AMP; >4 copies) as well as controls. Figure is adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

In order to further study the mechanism of how the SCNAs result into *IGF2* overexpression, we screened colorectal spheroid cell lines to identify cells that carried high *IGF2* overexpression. Spheroids are cell lines kept in 3D cultures and in comparison to normal (2D) cultures are considered to be more similar to actual tumors, making them more appropriate models for studying tumorigenesis. DNA sequencing of two spheroid lines with high upregulation of *IGF2* revealed that they carried the same duplication found in the colorectal tumors. Then the H3K27Ac status was tested, which can be evaluated as a marker for active chromatin and is tested by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) (Ernst et al., 2011). H3K27Ac marks on the gene were present in the samples that carried the duplication but not in the non-carriers. Afterwards, 4C-seq (van de Werken et al., 2012), a method to identify long range genomic interactions in 3D, indicated a physical connection of *IGF2* to a nearby super-enhancer, which can explain the increased expression levels of *IGF2*. The experiment identifies physical connections of a targeted genomic region (called the viewpoint) to any other locations in the genome. This experiment was conducted from the viewpoints of the gene and of the super-enhancer and both indicated an interaction between the gene and the super-enhancer (Fig. 2.2a) but not to its known cognate enhancer (Leighton et al., 1995). None of the other genes in the vicinity of *IGF2* showed a high overexpression (Fig. 2.2b), indicating in this case the super-enhancer interacts presumably only with *IGF2* and drives its expression.

Hence, the duplication coupled with high overexpression suggests that *IGF2* is activated by enhancer hijacking rather than by gene dosage in which the latter would result in only 1.5 increase of expression (Fig. 2.3).

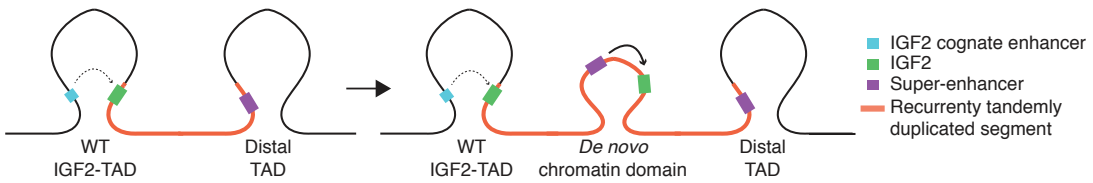


FIGURE 2.3: *IGF2* enhancer hijacking mechanism of activation. The proposed model depicts the formation of a new contact domain that brings in close proximity *IGF2* and the super-enhancer from the adjacent TAD. Figure is adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

2.1.2 Pan-cancer analysis implicates *IRS4* in enhancer hijacking

We further analyzed all available cancer genomes simultaneously (pan-cancer) in order to identify genes activated by enhancer hijacking across cancers. With this approach, we identified the insulin receptor substrate 4 (*IRS4*) as the major candidate. *IRS4* can be phosphorylated by the insulin tyrosine kinase receptor and when phosphorylated, *IRS4* plays a role in growth and proliferation as well as in insulin metabolism (Qu et al., 1999). Clustered deletions in proximity to *IRS4* associated them with high expression of *IRS4*. Even more pronounced were the clustered deletions found specifically to lung squamous carcinoma (Fig. 2.4). The clustered deletions cross a TAD and result in up to 2000 times upregulation of *IRS4* (Fig. 2.4). SCNAs across multiple cancers are shown as a heatmap, further supporting that these clustered deletions match the most common deleted region (Fig. 2.4).

Lung tumors were screened for *IRS4* upregulation and then sequenced for the presence of the deletions. Cells positive for the SCNA and overexpression of IRS4 were investigated further. Firstly, the H3K27Ac marks status were tested, and were present in the samples that carried the deletion but not in the non-carriers. Then we investigated whether *IRS4* is connected to a cis regulator element in close proximity. In both *IRS4* overexpressing samples and controls, *IRS4* was connected to transcription factor binding sites found in close proximity to *IRS4* (Fig. 2.4). We hypothesize that the TAD boundary or a loss of an insulator ensued to spreading of active chromatin. In the context of an already established promoter-enhancer interaction the active chromatin can result in activation of *IRS4* (Fig. 2.5, Weischenfeldt, Dubash, **Drainas** et al., in revision).

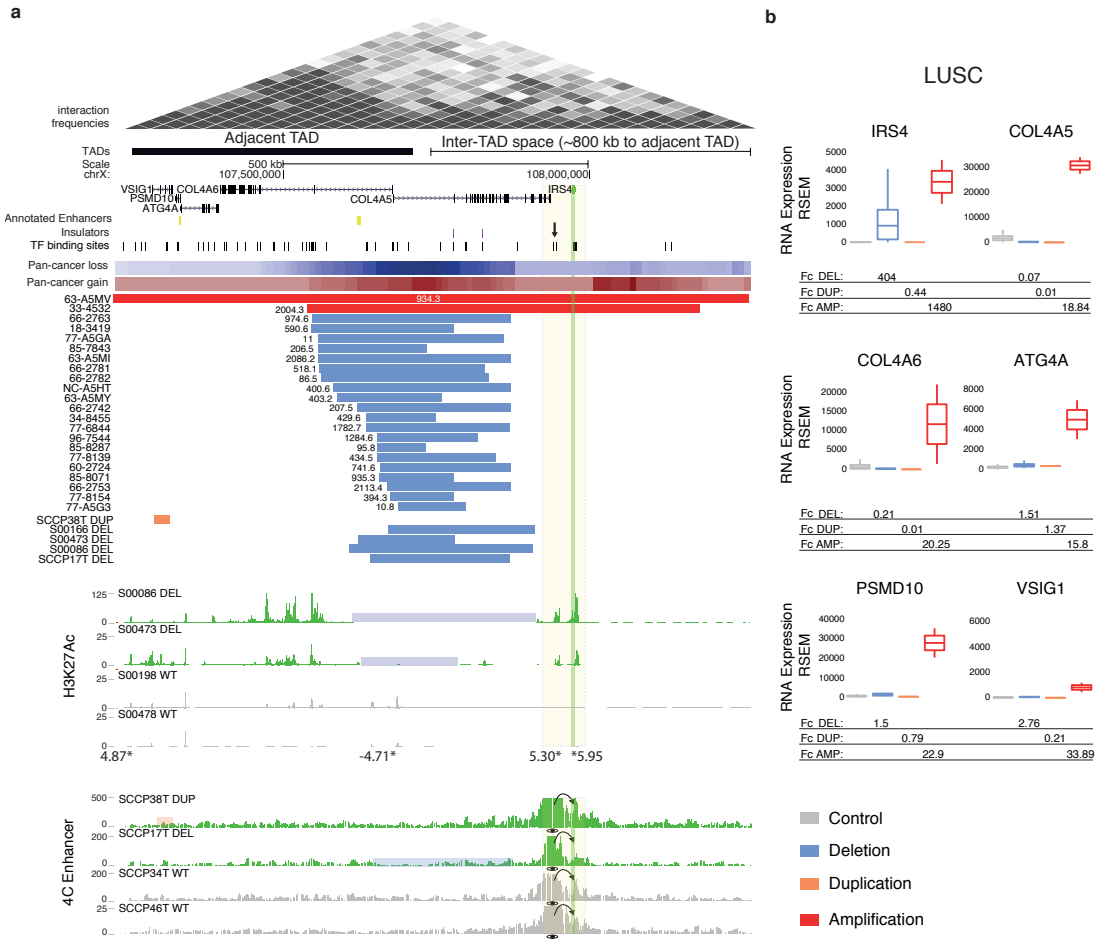


FIGURE 2.4: Recurrent SCNAs *in cis* are associated with *IRS4* expression increase. (**a**) Recurrent deletions at a TAD boundary near *IRS4* and amplifications in the locus, are associated with *IRS4* upregulation in lung squamous cell carcinoma. The figure depicts as heatmaps summarized SCNAs across cancer types (pan-cancer copy-number gains and losses). The deletion carriers have elevated H3K27Ac marks at a region near *IRS4*, which is absent in the controls. Transcription factor (TF) binding sites (candidate cis regulatory element) are highlighted with an arrow. Samples (S00086, S00473, SCCP38T, SCCP17T) that carried an SCNA exhibited higher expression using semi-quantitative RT-PCR and qPCR (Weischenfeldt, Dubash, **Drainas** et al., in revision). Lastly, 4C-Seq experiments using the candidate cis regulatory elements as a viewpoint in deletion carriers versus non-carrier control samples are depicted. DEL, deletion; DUP, duplication; WT, wild-type locus. (**b**) Expression measurements in LUSC (unadjusted RSEM gene expression values) for deletion carriers versus controls, reveal *IRS4* as the most overexpressed gene in the locus. Figure is adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

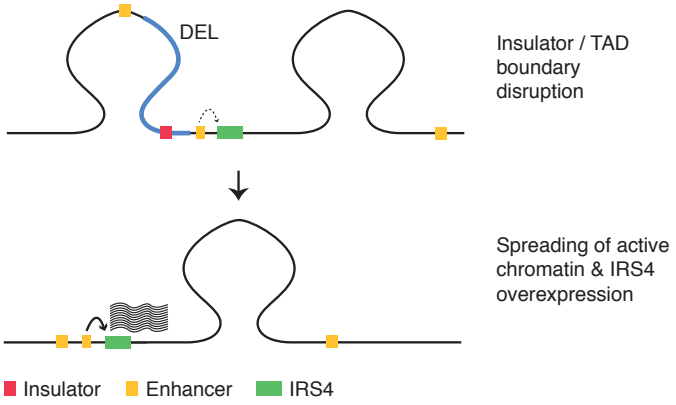


FIGURE 2.5: Proposed model for IRS4 overexpression. The clustered deletions result in loss of the TAD boundary *in cis* and allow active chromatin spreading and activation of *IRS4*. Figure is adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision)

IRS4 has not been characterized as an oncogene yet. The presented analysis suggests it as a potential new oncogene in cancer. This finding gives reason to screen for *IRS4* overexpression in tumors (particularly in lung tumors) and potential therapeutics could be explored for treatment.

Motivated from the above results, we sought to recapitulate the genomic alterations that resulted in enhancer hijacking in cell lines. Using CRISPR/Cas9 technology, we attempted to re-engineer alterations associated to *IGF2* expression in colorectal cancer cell lines, as well as *IRS4* in lung cancer cell lines. As described in the introduction (see section 1.2.3), a combination of gRNAs can be used to generate a variety of rearrangements. We therefore used gRNAs in combination with or without templates to achieve the specific rearrangements.

2.2 Results

The strategies chosen to perform the genetic engineering are shown in Figure 2.6. In order to perform the deletions, I designed gRNAs that introduce breaks at the genomic coordinates of rearrangements found in tumors. To select a specific rearrangement, a donor template construct was provided with a selection marker (Fig. 2.6a and Table 5.2). Cells were then selected and enriched using flow cytometry. In order to perform the duplication, a similar strategy was designed using a template in an orientation that if homologous recombination occurs between separate chromosomes, it can result in a duplication (Fig. 2.6b). The strategy for the duplication was based on a similar approach applied in yeast (Huber et al., 2014). In order to screen cells that obtained the vector expressing the gRNAs and Cas9 as well as the template vector, selection markers were used. After selection I tested with PCR whether the rearrangements were present in the population of cells (see Appendix Fig. A.1). Then, single cells were sorted with FACS in order to obtain individual cell clones. After the cells were confluent in the sorted plates, they were collected from all wells and subpools of approximately twelve clones were generated. From these subpools, DNA was extracted and each rearrangement was tested again by PCR. The individual clones in these subpools that gave the expected PCR product were then individually grown to larger wells and DNA was extracted from each of the clones. Clones with the desired rearrangement were further tested by qPCR in order to asses the overexpression levels of the gene of interest (*IGF2* or *IRS4*). Additionally, whole genome sequencing was performed to verify whether the rearrangement was indeed present and also to identify any other possible secondary rearrangements that might have occurred (Fig. 2.6c). The screen was applied to *IGF2* and *IRS4*, which were the most promising candidates that were identified from the tumor specific and pan-cancer analysis for enhancer hijacking (Weischenfeldt, Dubash, **Drainas** et al. in revision). An overview of the screen is depicted in Appendix (Fig. A.2).

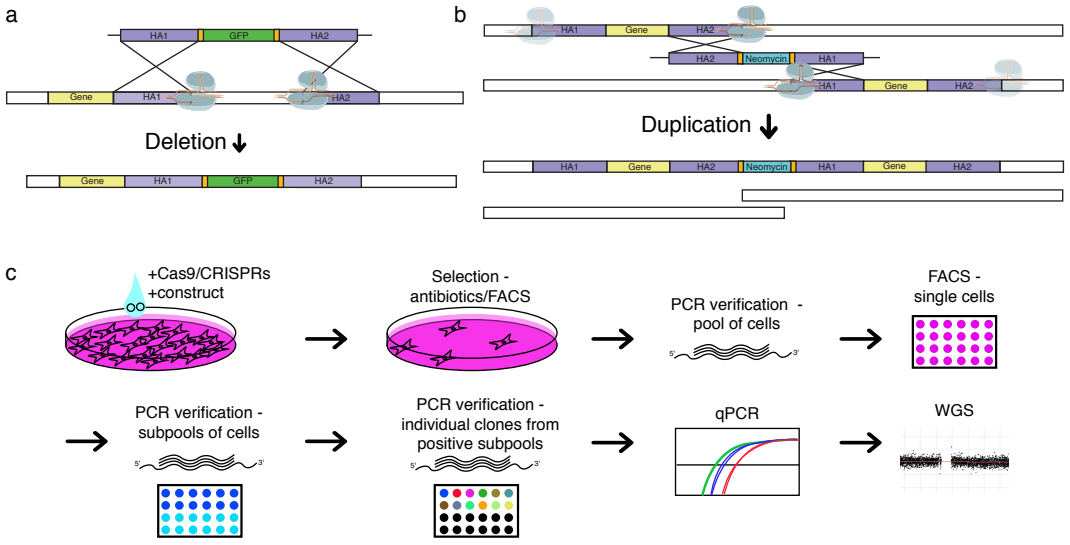


FIGURE 2.6: Screening strategy to generate rearrangements. (**a**) Targeted deletion generation strategy using a template to drive homologous recombination through the homologous arms (HA). Selection is based on GFP intensity and shorting with FACS. ((**b**) Targeted duplication generation strategy using a template to drive homologous recombination through the HAs. Selection is based on neomycin resistance. ((**c**) Screening strategy to identify cells that harbor the engineered rearrangements. First cells are transfected or transduced with the designed gRNS that target for the rearrangement. Then selection takes place to enriched for the cells that obtained the gRNAs. Next, a PCR reaction is performed to test whether the rearrangement is present in the population of cells. If successful, single clones are sorted. Subscreening with PCR of smaller pools accelerates the screening process. Finally single clones are screened for the presence of the rearrangement and qPCR as well as whole genome sequencing is performed.

2.2.1 Generation of *IGF2* enhancer hijacking associated rearrangements

First, we sought to recreate a duplication event for the cluster of duplications found in colorectal cancer (Fig. 2.2) as well as for the deletions (Fig. 2.7), which all led to *IGF2* overexpression. Although the deletions were not clustered, we considered recreating deletions a more achievable task in comparison to duplications, which structurally are more complicated. To achieve these rearrangements I used the HTC116 colorectal cancer cell line (ATCC) as the model cell line. HTC116 was chosen due to its simplicity in culture as well as to its effectiveness in transfection of foreign DNA.

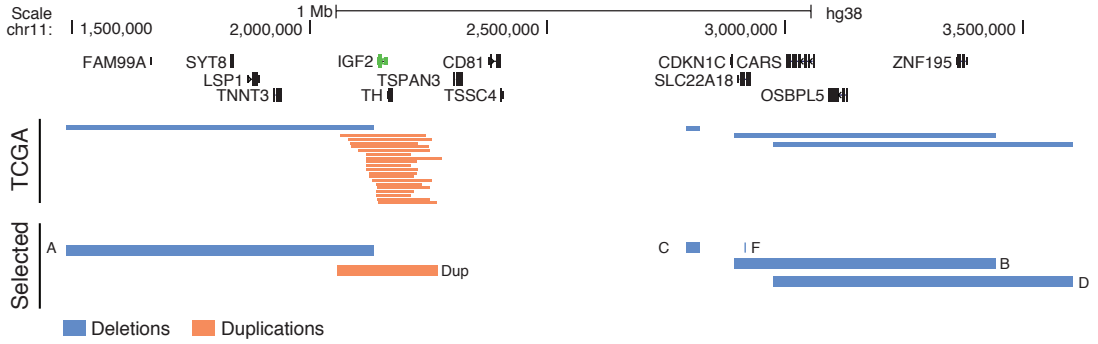
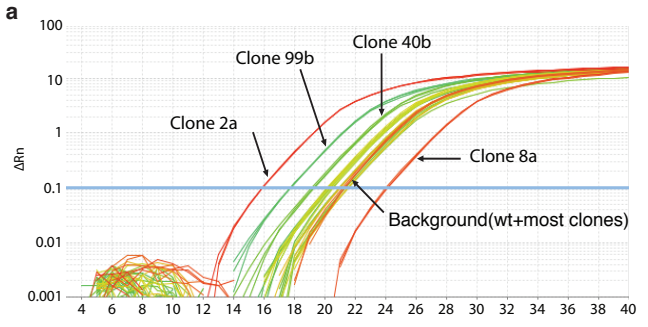


FIGURE 2.7: Selected rearrangements in the *IGF2* locus. A set of deletions and a duplication were selected according to the SCNAs identified in colorectal cancers (from the cancer genome atlas TCGA data) that led to high *IGF2* overexpression.

We applied the screening strategy depicted in Figure 2.6 to identify clones that resulted in the selected rearrangements (Fig. 2.7). A summary of the results is shown in Table 2.1. Performing the PCR-based assays as described in the previous section, I obtained clones for four out of the six rearrangements (Table 2.1). For these we observed an overexpression with qPCR in two clones (Fig. 2.8ab). Since the duplication event showed the most promising results, I screened more clones carrying potential duplicated regions in comparison to the deletions (Table 2.1).

TABLE 2.1: *IGF2* genetic engineering results summary. Plates from all conditions were screened. PCR verified clones were obtained from conditions DelC, DelD, DelF and Dup. From those, only two duplication cases showed higher expression levels by qPCR. MP-seq was performed on the qPCR positive clones but the expected duplication was not observed.

		HTC116				
Rearrangement	Wells	**Clones**	**Construct**	**PCR Verif.**	**qPCR Verif.**	**WGS Verif.**
DelA (1.6Mb)	192	7	yes	0	0	NA
DelB (510kb)	192	2	yes	0	0	NA
DelC (20kb)	384	24	yes & no	12	0	NA
DelD (7.2Mb)	384	24	yes	1	0	NA
DelF (CTCF)	192	12	yes & no	7	0	NA
Dup (220kb)	768	65	yes	10	2	0



b

Clone	PCR	Sanger seq	Q-PCR	WGS
2a	+	duplication sig.	55.2	duplication
8a	+	duplication sig.	0.1	deletion
10a	-	-	1.0	-
7b	+	one end only	1.0	-
8b	-	-	0.7	-
9b	-	-	2.2	-
14b	-	-	1.3	-
21b	+	duplication sig.	1.1	-
23b	-	-	1.1	-
40b	-	-	3.4	no SV
41b	+	duplication sig.	1.1	-
97b	+	duplication sig.	1.6	-
99b	not clear	-	7.9	telomere del.

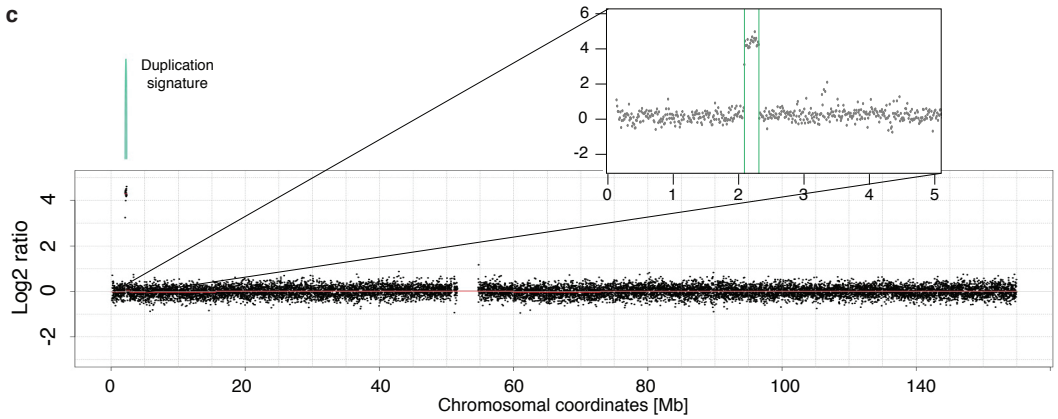


FIGURE 2.8: qPCR and sequencing results for clones that were verified by PCR to harbor a duplication event. (**a**) qPCR was performed on duplication positive as well on negative WT clones (qPCR Figure and experiment performed by Dr. Adrian Stütz). (**b**) Summary of data. The clones with higher overexpression are highlighted in yellow. The clone with the expected fold change is highlighted in blue. (**c**) Read depth plot of MP-seq of clone with highest *IGF2* overexpression. Chromosome 11 is depicted. *IGF2* locus carries a duplication signature and is also highly amplified. Zoom in of 0-5 Mb region is displayed (sequencing libraries prepared by Dr. Adrian Stütz and Benjamin Raeder).

We further examined the clones with the highest overexpression with Sanger sequencing (Materials and Methods 5.3.7) and verified that the clones indeed carried a duplication rearrangement. Additionally, we performed long-range paired end (mate pair) sequencing (MP-seq) to fully characterize the rearrangements we observed in the qPCR (Materials and Methods 5.1.6). Such an approach can give a detailed representation of structural rearrangements. Only one out of the two clones showed a duplication signature in which a sharp increase in the read depth was observed, indicative for a high level amplification of the affected locus (Fig. 2.8c). The degree of overexpression (52 times) can be explained by the degree of amplification, which is approximately 44 times higher. Therefore, this overexpression clone was due to high-level amplification rather than EHJ. From these results no EHJ event was detected. PCR verification of the rest of the clones (also by verified by Sanger sequencing) indicates that the rearrangement may have been achieved,

but only the qPCR positive clones (the ones with higher *IGF2* expression) were tested by MP-seq.

2.2.2 Generation of *IRS4* enhancer hijacking associated rearrangements

Considering the EHJ analysis across tumors and specifically in lung squamous carcinoma, I sought to recreate the deletions associated with *IRS4* overexpression. Since the deletion cluster was more prominent in lung squamous cancer, I used the following lung squamous cancer cell lines to model the deletions: HCC15, H520 and H2170 (Table 5.1). From these H520 and H2170 gave few clones and could not be scanned in all conditions. Recreation of six deletions (DELs) were tested as depicted in Figure 2.9. Genomic coordinates of Del-A and Del-C were selected according to inner and outer coordinates of the deletion cluster, whereas Del-B was selected as a separate deletion. The remaining three deletions were chosen according to actual deletion coordinates in the lung tumors that had the highest expression of *IRS4*. For the first three (Del-A, Del-B and Del-C) two gRNAs were used to introduce breaks in each coordinate to increase efficiency, whereas for the next three (Del-1, Del-2 and Del-3) only one gRNA per site was used based on the cutting efficiency (Materials and Methods 5.3.8).

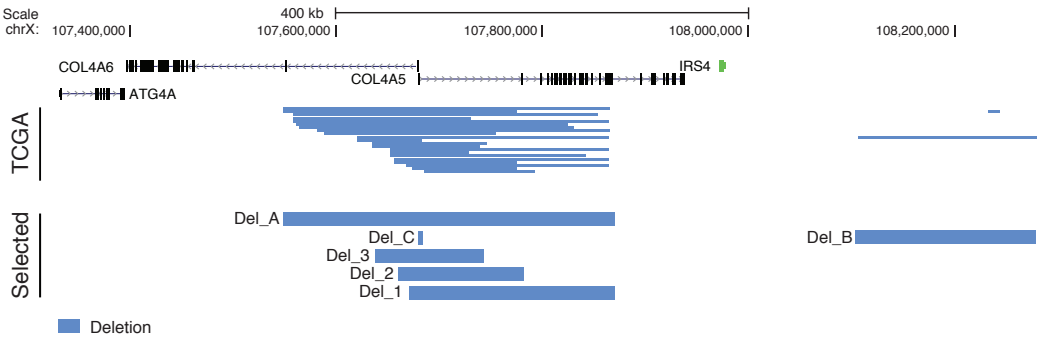


FIGURE 2.9: Selected rearrangements in the *IRS4* locus. A set of deletions were selected according to the clustered SCNAs identified in lung cancers (from the cancer genome atlas TCGA data) that were associated with high *IRS4* overexpression. Del-A was selected according the to the outer coordinates of the clustered deletions, Del-C was selected according to the inner coordinates of the clustered deletions, Del-B as a deletion outside the clustered deletions. Del-1,2,3 were selected according to actual deletion coordinates from the TCGA data with the highest expression levels.

The screening strategy was applied as described in Figure 2.6 with a modification. Instead of transfection, virus transduction was used (Materials and Methods 5.1.4). All deletions selected were initially verified in the population of cells by PCR. From the screening method single clones were identified and verified by PCR, which carried the deletion. A summary of screening results is depicted in Table 2.2.

TABLE 2.2: Summary of *IRS4* screening results. All clones obtained were tested by PCR for the presence of the rearrangement. PCR positive clones were further Sanger sequencing, whole genome sequenced and also tested for *IRS4* expression levels with qPCR. Positive results are indicated in the Table. NA: non-applicable

HCC15/H520						
Rearrangement	**Total wells**	**Clones**	**PCR**	**Sanger**	**WGS**	**qPCR**
Del-A (323kb)	**192/192**	16/19	0/1	NA	NA	0/0
Del-B (175kb)	**192/192**	18/29	2/0	2/NA	2/NA	0/0
Del-C (5kb)	**192/192**	21/19	1/0	1/NA	not clear	0/0
HCC15/H2170						
Del-1 (200kb)	**384/384***	192/192	5/17	5/NA	1/NA	0/0
Del-2 (121kb)	**384/192**	192/38	8/1	8/NA	0/NA	0/0
Del-3 (105kb)	**768/384**	384/76	1/NA	NA	NA	0/NA

*more than one cell sorted in wells in order to increase screening depth

None of the generated clones exhibited overexpression of *IRS4*. In order to verify that the deletions were indeed generated, two approaches were undertaken. The first approach was to verify the junction of the rearrangement with Sanger sequencing. For the second approach low-coverage whole genome sequencing was applied (LC-WGS) in order to identify the deletion based on the read depth information (Materials and Methods 5.1.7). From the first approach 36 positive clones were identified, indicating that such a rearrangement is present in the clones. Two samples of Del-B were verified by MP-seq and the rest were tested with LC-WGS. From the LC-WGS results we were identified one sample that carried the deletion (Fig. 2.10). Though, the absence of the deletion in the sequencing data but present in the Sanger sequencing data, may suggest that the rearrangement had occurred but the piece was not lost from the cell, but rather integrated somewhere else in the genome. Therefore we tested all Sanger sequenced positive cases with qPCR. From the 16 cases tested we were unable to identify any *IRS4* overexpressing

clone. Although I achieved to recreate the rearrangements in a lung squamous cancer cell line, no overexpression resulted from this rearrangement. This indicates that there might be additional factors involved that lead to overexpression of *IRS4* as seen in the tumors.

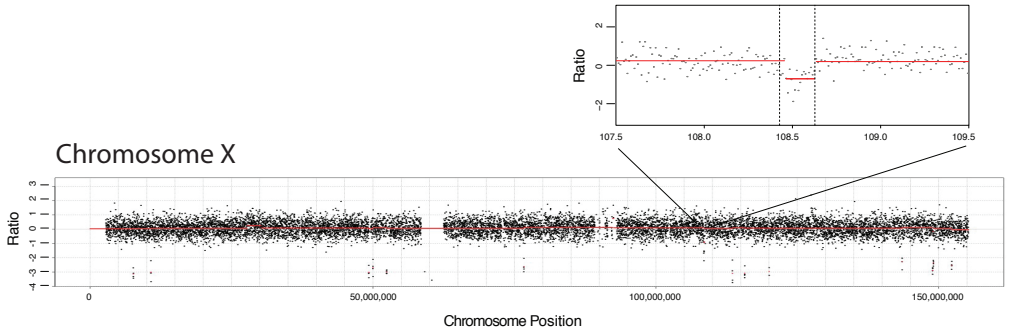


FIGURE 2.10: Read depth plot of LC-WGS of a clone positive for the deletion in PCR. Chromosome X is depicted. *IRS4* adjacent locus has a deletion at the expected coordinates. Zoom in of 107.5-109.5 Mb region is also displayed.

2.2.3 Identification of *IRS4* as a potential oncogene

IRS4 was found to be frequently overexpressed in the pan-cancer analysis and suggested as a potential oncogene (Weischenfeldt, Dubash, **Drainas** et al. (in revision)). In order to investigate its oncogenic properties *in vivo*, we designed mouse xenograft experiments and studied the growth rate of tumors with and without *IRS4* overexpression. To this end, we subcutaneously injected HCC15-IRS4 and HCC15-mock cells (lung squamous cancer cell lines) into athymic nude mice. This experiment was conducted with two independent replicates (N=8 for each group in the first experiment, N=9 for control and N=12 for *IRS4* overexpressing sample in the second experiment).

The outline of the experiment is depicted in Figure 2.11. To obtain a verified *IRS4* overexpressing transgene vector, *IRS4* transgene was ordered (Origene). Then *IRS4* was removed from the vector in order to have a control vector (Materials and Methods 5.2.2). Lentivirus was then prepared from each vector and the cells were infected. Next, I performed two rounds of fluorescent activated cell sorting (FACS), to sort cells according to their GFP intensity. By this, I obtained enriched cells that have the vector stably integrated in their genome. Then I tested whether the protein indeed is translated to IRS4 by immunoblotting (immunoblotting of control and IRS4 HCC15 cell lines can be seen in Fig. 2.13c).

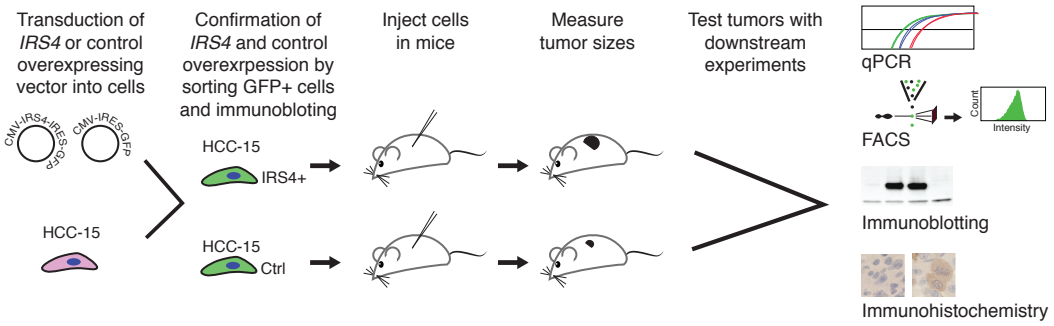


FIGURE 2.11: Scheme of xenograft experiments in mice. *IRS4* and control vector overexpressing HCC15 cell lines were generated by lentivirus integration. Then by measuring GFP intensity stable integration of the vectors was confirmed. The cloned HCC15 cell lines were then injected into mice and tumor sizes were measured after one week every four to five days. The tumors were harvested and qPCR, FACS, immunoblotting and immunohistochemistry were conducted.

After confirming the expression of *IRS4* in the protein level we proceeded with injection of the cells into mice (all mouse related work was kindly conducted by Dr. Yuanyuan Chen, Deutsches Krebsforschungszentrum, Heidelberg). We observed tumor formations in both *IRS4* overexpression and control cell injections. The tumors harboring the *IRS4* overexpressing plasmid were significantly larger indicating increased tumor growth (P=0.046 and P=0.03, respectively in last time point; two-tailed t-test; Fig. 2.12). These experiments support the tumor promoting role of *IRS4* in carcinogenesis.

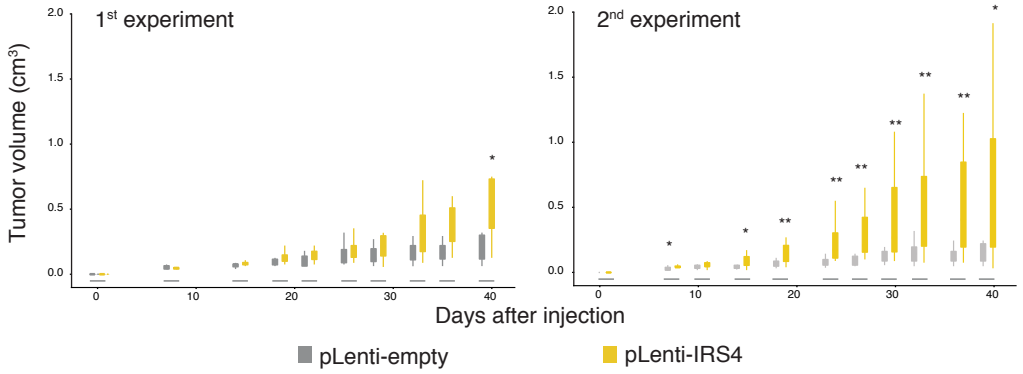


FIGURE 2.12: Growth experiments between *IRS4* overexpressing and control xenografts. Boxplots depicting mouse tumor progression curves of HCC15 cells containing IRS4-expressing lentiviral constructs (pLenti-IRS4) versus mock control (pLenti-empty HCC15); 1x106 cells injected respectively; last time point: 1st P=0.046, 2nd P=0.03; two-tailed t-test; two-tailed t-test computed at last measured time point (day 39); N=8 for each group in first experiment, N=9 for control and N=12 for IRS4 overexpressing sample in second experiment. Figure and caption text adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

We then tested whether resected tumors maintained *IRS4* overexpression using four independent approaches. Firstly, we measured RNA expression with qPCR and showed that in both experiments, *IRS4* overexpressing cell lines indeed had higher *IRS4* levels in comparison to the controls. Secondly, FACS also indicated that a large population of cells in each tumor was indeed GFP positive, which indicates the cells tested carried the vectors. Thirdly, with immunoblotting I confirmed that five out of eight *IRS4* overexpressing tumors indeed contained higher levels of IRS4 protein. Lastly, immunohistochemistry showed in both experiments that IRS4 protein was overexpressed in the expected tumors (Fig. 2.13).

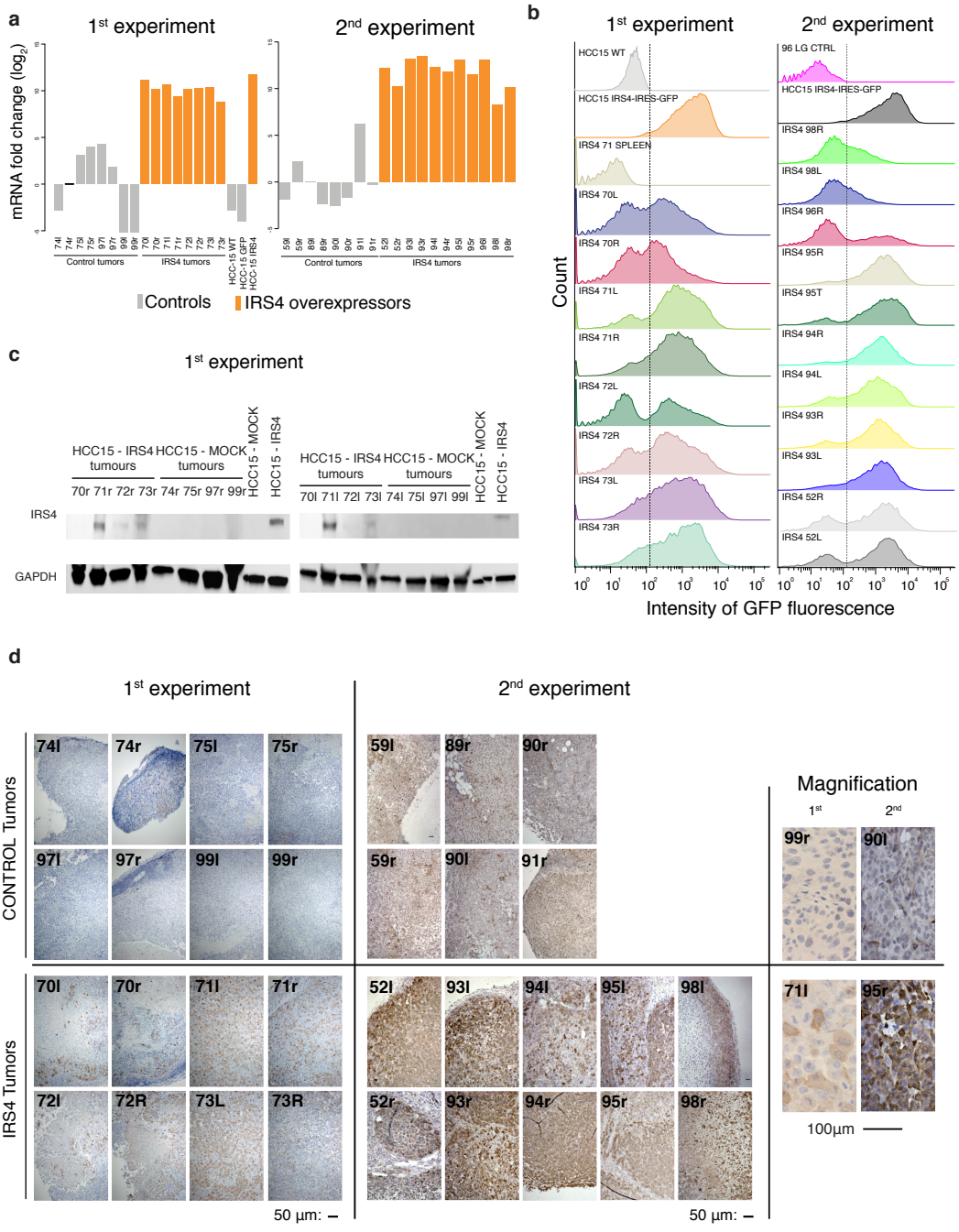


FIGURE 2.13: (IRS4 overexpression verification experiments. **a**) qPCR of the control and *IRS4* overexpressing HCC15 tumors confirming *IRS4* overexpression (experiment performed by Dr. Yuanyuan Chen). (**b**) Flow cytometry of tumors injected with HCC15-IRS4-GFP confirm the expression of the vector by measuring the GFP intensity. (**c**) Immunoblotting of protein extracts of the xenografts of the first experiment, indicating protein expression in at least seven out of the ten xenografts. (**d**) Immunohistochemistry represents IRS4 protein expression in the HCC15-IRS4 harvested tumors but not in the control tumors in both experimental replicates. Figure is adapted and modified from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

2.3 Discussion

In the presented study we estimated potential enhancer-hijacking events across cancers and identified potential candidates. The analysis was based on SNP6 microarrays and associated deletions and duplications with high expression levels of nearby genes (Weischenfeldt, Dubash, **Drainas** et al., in revision). From this analysis we investigated further the mechanism of activation of two genes: *IGF2* a top candidate in colorectal carcinoma and *IRS4* a top candidate across many different cancers, specifically in lung squamous adenocarcinoma.

For *IGF2* we identified a cluster of recurrent duplications resulting in a much higher expression of *IGF2* that cannot be explained by a gene dosage model. We could further verify the connection of *IGF2* with an enhancer from another topological associated domain (TAD). This was verified in spheroid cell lines that harbored the duplication and had high *IGF2* expression.

IRS4 locus revealed a set of recurring deletions within the COL4A5/A6 genes, but not in *IRS4*, which resulted in high *IRS4* overexpression. The deletion in the COL genes resulted in their under-expression as expected. This further implies that *IRS4* is the only candidate of the locus that is associated with high expression.

Based on these observations and results, we attempted to recreate these rearrangements that were associated with overexpression. The cell lines selected for the experiments were specific to each tumor type, therefore a colorectal cell line and lung cell lines were used for *IGF2* and *IRS4* respectively. During my PhD thesis work was published that recreated rearrangements in the genome using the CRISPR/Cas9 system (Kraft et al., 2015; Li et al., 2015) but they did not investigate enhancer hijacking (EHJ). They showed that genomic engineering had low efficiency, which is in line with my observations. However, these studies did not further characterize the rearrangements by whole genome sequencing which gives another layer of detailed validation, an important step we performed that confirms the presence of the rearrangements.

Initially, I attempted to re-engineer the rearrangements associated with *IGF2* overexpression and selected a set of deletions and a duplication (from the duplication cluster). Although the genomic coordinates of the deletions were not as similar to each other as the duplications, I initially attempted to generate those since deletions are structurally

simpler than duplications. I achieved to recreate four out of the six rearrangements designed for *IGF2* in the HCT116 cell lines (Fig. 2.7). Only for the duplication I was able to observe a difference in expression thus proceeded to analyze these events further. From all clones that I had tested, only one clone led to high expression of *IGF2*. PCR and Sanger sequencing verified this clone to carry the duplication. To investigate the rearrangement in detail, long range paired end sequencing (MP-seq) was used to examine the read depth and structural rearrangements. MP-seq revealed, despite the confirmation of a duplication, a high level amplification (Fig. 2.8). Therefore this clone was dismissed as a putative enhancer hijacking event since the level of overexpression could be explained by the amplification. Unfortunately the remaining clones that I had similarly analyzed did not exhibit overexpression of *IGF2*.

For *IRS4*, I attempted to engineer the deletions according to the cluster of deletion found in lung cell carcinoma (as well as in the pan-cancer screen). We designed different approaches for selecting the deletions and for introducing the breaks. The first three were designed according to inner and outer coordinates of the deletion cluster in order to capture the whole range of the deletion as well as the minimal region. With this we reasoned that we would have more insight on the region that results in EHJ. Also we designed guides that target a deletion outside the clustered rearrangements, which also resulted in overexpression of IRS4. For these rearrangements I applied two gRNAs per coordinate, to assure that a break will occur. For the next three, deletion coordinates were selected based on the events observed in the tumor, to assure that we recreate actual deletions observed. Furthermore, one gRNA per side was designed and tested with an assay that measures break efficiency. One gRNA was used in order to minimize the amount of off target effects that each gRNA might introduce to the genome, for example from having two from each side. We screened over 3500 wells and identified 36 clones according to PCR verification. We achieved to recreate all deletions according to the Sanger sequencing data. For further validation we confirmed with whole genome sequencing that only three cases have the deletion (two with MP-seq data and one with LP-WGS). The confirmation rate with the LP-WGS was very low in comparison to the PCR and Sanger sequencing verification. LP-WGS though is not able to identify the orientation and position of the sequenced data, which might explain the low confirmation rate. The reads are aligned to a reference therefore only read depth information is obtained but not an actual position. Therefore if the piece is not lost from the cell, we

will not see any drop in the read depth in the expected position. The PCR and Sanger sequencing data suggest that that the rearrangement did occur, however the piece was not lost from the cell but presumably integrated somewhere else in the genome. In order to verify this hypothesis, two approaches can be used, MP-seq on the clones tested only by low coverage sequencing or fluorescent *in vitro* hybridization (a technique that can visualize DNA with a targeted probe).

Unfortunately, by recreating the respective SV events in the cell lines, we did not observe any *IRS4* overexpression as hypothesized from the pan-cancer analysis data. Although we verified three clones (two harboring the deletions out of the cluster and one within) with all techniques, no *IRS4* overexpression was observed. Moreover no *IRS4* overexpression was observed in any of the other clones verified by PCR and Sanger sequencing.

Multiple parameters may influence why we did not observe gene overexpression from enhancer hijacking. All cell lines used in culture are grown in optimal conditions and overexpression of these genes may have no benefit for the cells in culture. Another reason may be that the targeted enhancers are not active in the cells in culture. Therefore the resulting rearrangement will then not give any increase in expression. Moreover the chromosomal state in the cell lines is different to tumors, as observed in previous studies for measuring epigenetic H3K27Ac marks and DNA methylation, which influence open and active chromatin (Hovestadt et al., 2014; Lin et al., 2016). The latter parameter is important because the structural conformation of the chromosomes, which is influenced by the epigenome of the DNA, might provide an explanation to why we did not to see enhancer hijacking in this context.

In conclusion, targeted rearrangements were obtained and verified by PCR and Sanger sequencing but no overexpression of *IGF2* nor *IRS4* was observed. Experiments *in vivo* may be more relevant to actual tumors, since the correct cell types can be targeted with the natural chromosomal state and may result in the observed enhancer hijacking events that lead to gene overexpression.

Next, we investigated the tumorigenic role of *IRS4*. To achieve this we investigated the tumor growth in mice xenografts. I generated cell lines that stably overexpressed *IRS4* or a control vector, by lentivirus infection. The cells were subcutaneously injected in mice and tumor growth was measured regularly. Although individual mice exhibited different rates of tumor growth, overall we concluded from two independent experiments that the

IRS4 overexpression significantly increased tumor growth. All tumors were also tested for *IRS4* overexpression with various techniques (Fig. 2.13). Therefore these results indicate that *IRS4* plays an oncogenic role in promoting tumor growth.

2.4 Conclusions

Enhancer hijacking was observed more commonly as previously thought in cancer as indicated from the pan-cancer analysis described in the introduction. To recapitulate such events is not trivial since it is likely an effect resulting from the combination of multiple parameters, including cellular context or possibly modifiers not yet understood. Although we were unable to detect EHJ in the cell lines tested, we were able to recreate the majority of the selected rearrangements found in tumors, indicating the functionality of genomic engineering in the cell types tested.

As observed from the pan-cancer analysis, some candidates were found across several cancers but others were cancer specific. This could be due to gene specificity in the different tumors. Furthermore the microenvironment and the chromosomal state of the cell might have a large influence in enhancer hijacking. Therefore, investigating in *in vivo* models may greatly enhance the possibility of targeting a cell with optimal chromosomal state that will lead to enhancer hijacking.

Lastly, we investigated the tumorigenic role of *IRS4* in *in vivo* models. We used lung squamous cell lines overexpressing *IRS4* and performed xenograft experiments in mice. The tumors observed from the overexpression *IRS4* group in comparison to the control, in two individual experiments, gave significantly larger tumors. This points to a tumorigenic role of *IRS4* and hence may be an important target to investigate for cancer therapy of *IRS4* overexpressing tumors.

Chapter 3

Identification of Drivers of Growth as Potential Tumor Suppressor Genes via Genome Wide CRISPR/Cas9 Knockout Screens

This chapter describes genome wide knockout screens for the identification of genes that drive cell growth and may be potential tumor suppressors. The screens led to the development of ScrispR, a tool to analyze the data. As a result, genes related to the Hippo and the mTOR pathways were mostly enriched. Novel candidates of promoting growth, *AHR* and *FRYL* were further investigated and combined with the outcome of additional analysis of the cancer genome atlas are suggested as potential tumor suppressors.

Contributions

This chapter contains mainly my own work with support of others along the development of the project. I conceptualized the work of this chapter together with Balca Mardin and Jan Korbel. I designed and led all experiments related to the screens. Furthermore, I conducted the initial library preparations, which were then optimized by Adrian Stütz and Benjamin Raeder. Sebastian Wazsak proposed the main concept of the ScrispR tool. I then led the development and application of the tool. The final package was written and optimized by Mike Smith. I conducted the analyses of all screens receiving valuable input and advice from Balca Mardin and Sebastian Wazsak. I supervised the analysis performed by Ruxandra Lambuta relating the ubiquitination pathway. Moreover, I established the crystal violet assay and its analysis with help from Benjamin Raeder and Ruxandra Lambuta. Also, I conducted the downstream analysis for the *AHR* and *FRYL* genes. I developed the method to identify knockout clones by interpreting complex Sanger sequencing information. Adrian Stütz and Benjamin Raeder prepared the libraries for sequencing. The TCGA analysis, which helped me further support the results of the screens, was kindly conducted by Christopher Buccitelli. Balca Mardin and Jan Korbel guided and supervised the project and provided valuable and insightful feedback. From the work regarding this chapter a manuscript will be written where I will be the first author.

Publication related to this chapter:

A cell-based model system links chromothripsis with hyperploidy

Mardin B. R., **Drainas A. P.**, Waszak S. M., Weischenfeldt J., Isokane M., Stütz A. M., Raeder B., Efthymiopoulos T., Buccitelli C., Segura-Wang M., Northcott P., Pfister S., Lichter P., Ellenberg J., Korbel J.O.
Molecular systems biology, 11(9) 828–828, 2015.

3.1 Introduction and Motivation

In order to model neoplastic growth *in vitro*, reconstructing the important steps in cancer genome evolution is essential. Previous studies aimed at identifying contributors to tumor growth in different cell types at different stages, using either shRNA based screens (Silva et al., 2008) (downregulation) or CRISPR/Cas9 screens (Shalem et al., 2014) (complete knockout). These screens not only identified known tumor suppressors such as *PTEN*, but also revealed novel candidates, such as *REST* (Westbrook et al., 2005), JNK pathway associated genes (Eskiocak et al., 2011) or genes important for tumor invasion and metastasis such as *XPO4* (Zender et al., 2008), *ADAMTS18* (Ly et al., 2012), *GAS1* (Gobeil et al., 2008), *SALL1* (Wolf et al., 2014), *NF2*, *PTEN* and *CDKN2A* (Chen et al., 2015).

Until recently, screens that identified novel tumor suppressors have only employed shRNA libraries. However with shRNA based libraries complete inactivation of a gene is not possible as opposed to knockouts e.g. achieved by CRISPR/Cas9 system. Furthermore, screens that studied essential genes showed that CRISPR/Cas9 based screens outperformed shRNA screens in effectiveness, false positives and off target effects (Evers et al., 2016). This can be due to various reasons, one of which is low expression of genes being sufficient to reduce or prevent a phenotype. In order to study the complete loss-of-function of a gene, CRISPR/Cas9 based libraries have been constructed and employed.

Moreover, up to date, a systematic whole genome knockout screen to identify genes involved in tumor initiation has not been performed. Previous screens did not target the whole genome, but a smaller set of genes. Lastly, as mentioned in the introduction (see section 1.1.2), genetic interactions (GIs) play an important role in tumorigenesis and several tumor associated genes have shown to interact genetically, indicating that epistasis in cancer is a frequent event (Wang et al., 2014b). A detailed investigation of the epistatic interactions and their importance to tumor initiation has not been performed yet.

In order to study epistatic interactions I made use of isogenic cell lines in my CRISPR/-Cas9 based screens, which already contain certain hallmarks of tumor evolution such as replicative immortality (e.g. hTERT activation), resistance to cell death (*TP53* inactivation) and hyperploidy (as a result of induced genomic instability) (Table 3.1).

Despite having these attributes, the cells are mostly unable to grow unattached to a surface; thus they are considered as non-transformed (and therefore can be considered non-tumorigenic), but have shown to be capable of transformation (Mardin et al., 2015). I used isogenic cells lines with *TP53* deficiency and hyperploidy since these backgrounds have shown to increase cancer risk substantially (Mardin et al., 2015). This encouraged us to test for genes with potential tumor suppressor activity in these genetic backgrounds.

To understand the mechanisms of transformation I used a powerful tumorigenic predictor: anchorage-independent growth (ANIG). ANIG is utilized in a soft agar assay, where transformed cells are able to grow in an agarose matrix (Fig. 3.1). Over the past few decades ANIG on soft agar has been used as a marker for *in vitro* transformation (Mori et al., 2009). It also has been reported to be associated with tumorigenicity and metastatic potential *in vivo* (Freedman and Shin, 1974).

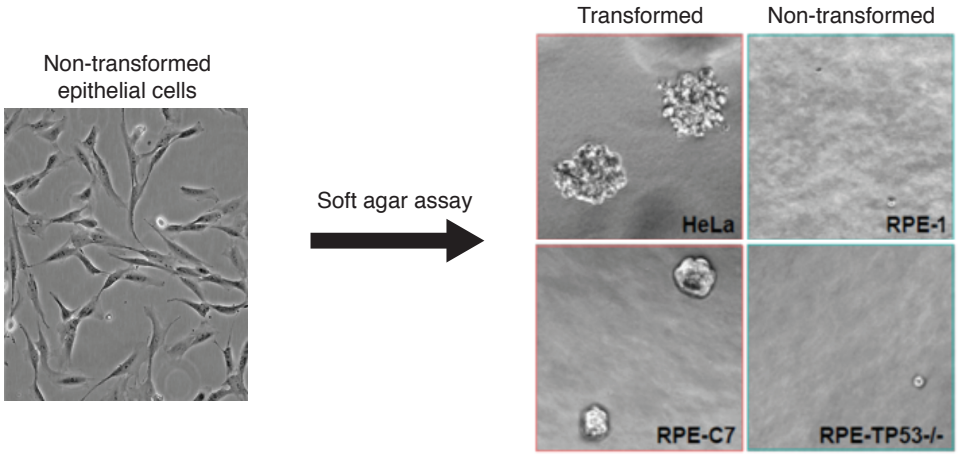


FIGURE 3.1: Soft agar assay. Transformed epithelial cells are able to grow in soft agar in comparison to non-transformed cells. HeLa cells and RPE-C7 are transformed cells whereas RPE-1^{WT} and RPE-$1^{TP53^{-/-}}$ are unable to grow and thus considered non-transformed.

ANIG was previously employed in our laboratory leading to the development of a method termed CAST (Complex Alterations after Selection and Transformation) (Mardin et al., 2015). CAST is based on applying perturbations on cell lines (chemical or genetic) and afterwards subjecting them on soft agar to test for ANIG. Cells that grow in soft agar, which are considered to be transformed, are then pooled in batches and subjected to low-pass massively parallel whole genome sequencing. Applying CAST we were able to identify several transformed clones some of which carried complex rearrangements consistent with patterns of chromothripsis (Mardin et al., 2015). In addition, we were able

to demonstrate that a chemical or genetic perturbation can be sufficient to induce transformation, showing the robustness and efficiency of the assay. We then reasoned that combining genome-wide knockout screens (Hart et al., 2015; Shalem et al., 2014; Wang et al., 2014a) (see Section 1.2.2) with a powerful tumorigenic assay, such as CAST, can have the potential to lead to the discovery of genes that contribute to cell transformation representing potentially tumor suppressors.

Table 3.1: RPE and MCF10A isogenic cell lines used in study.

Cell line	Information	Reference
RPE^{WT}	Retina pigment epithelial cells	ATCC
$RPE^{TP53^{-/-}}$	RPE with disfunctional TP53	Mardin et al. (2015)
$RPE^{TP53^{-/-},hyperploid}$	$RPE^{TP53^{-/-}}$ with whole genome duplicated	Mardin et al. (2015)
MCF10A	Mammary epithelial cells	ATCC
$MCF10A^{TP53^{-/-}}$	MCF10A with disfunctional TP53	Custom Generated

Previous screens have used similar methods to ANIG, combining them with shRNA libraries that target several genes (Westbrook et al., 2005). CRISPR/Cas9 screens have the advantage of complete destruction of the gene rather than diminishing their expression in comparison to RNAi based libraries (i.e. siRNA or shRNA). siRNA screens have to be performed in multi-well format and therefore make large scale screens very laborious. shRNA screens can be performed in bulk, however since down-regulation of the gene is not complete, minimal expression of the gene might be sufficient for the function of the gene, thus can mask the relevant phenotypes. Therefore CRISPR/Cas9 screens offer a superior alternative to the RNAi based screens and have shown to be quite effective (Hart et al., 2015; Shalem et al., 2014; Wang et al., 2014a).

For our experiments I employed the GeCKO (Genome-wide CRISPR/Cas9 Knockout) library screening approach (Shalem et al., 2014). The GeCKO library was designed to target the majority of genes in the human genome. Initially the authors obtained a list of genes with the most common transcripts expressed across several cell types. The strategy was to design gRNAs that target the first exons of each gene, and a set of six gRNAs per gene were designed. The gRNAs were planned to have minimal off target effects in the genome (Hsu et al., 2013). In addition, GeCKO library targets 1,864 miRNAs, introducing potentially interesting targets in non-gene elements. Furthermore GeCKO

also encompasses 1000 control gRNAs (gRNAs without any matches in the genome) and therefore can be used as a negative control for the phenotype of interest. In total 123,411 gRNAs were designed (Fig. 3.2). Since the set of gRNAs was considered to be too large for a single library, the library was divided to two subsets: Libraries A and B. Each library contains three guides per gene and also the same 1000 control guides. Library A contains additionally the gRNAs that are designed to target miRNAs.

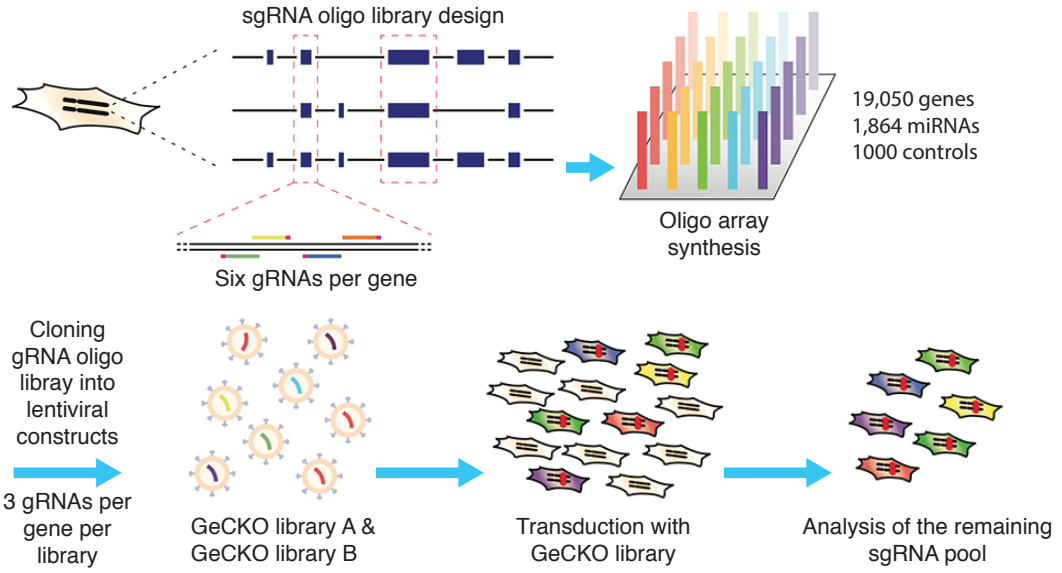


FIGURE 3.2: GeCKO library design. Initially most common transcripts were identified and gRNAs were designed to target the first exons, in total six gRNAs per gene. The gRNA sequences were synthesized using oligo array synthesis technology. The oligos were cloned into a lentiviral backbone and viral particles are constructed. The cells are transduced with the virus and are subsequently selected for antibiotic resistance. The surviving cells are processed for the screen of interest. Image and strategy adapted from Sanjana et al. (2014) and Shalem et al. (2014).

The designed library was acquired from Addgene as a pooled library. The strategy of the screen has three main steps: first library amplification, second virus packaging and last transduction and selection. In the end of the selection, analysis of the remaining gRNA representation indicates which gene knockouts resulted in the phenotype of interest (Sanjana et al., 2014; Shalem et al., 2014)(Fig. 3.2).

3.2 Results

3.2.1 Methodological overview

The strategy of the screen is depicted in Figure 3.3. Initially lentivirus was prepared to infect the non-transformed cell lines in all cell lines in replicate (Materials and Methods 5.3.1.1, Table 3.1). Lentiviruses have the ability to integrate their genome into the host genome even in non-replicating cells. This gives the system the ability for stable integration of genes of interest, which for the GeCKO library are Cas9, gRNA and a puromycin resistance gene used for selection. Therefore, puromycin selection enriches cells that incorporated the viral DNA. In order to assure transduction of ideally one gRNA per cell, I conducted a titration experiment to estimate the multiplicity of infection (MOI) and a low MOI was selected (Material and Methods 5.3.1.2).

Following antibiotic selection, a portion of the cells was collected. This serves as a baseline of the representation of gRNA sequences before soft agar selection (indicated as "I", Fig. 3.3). The remaining cells were embedded in soft agar for the duration of one month. During this time, only the cells that acquire a knockout that mediates their growth in soft agar divide and form colonies. After one month in culture the cells were collected. These post soft agar cells contain the final representation of gRNA sequences (indicated as "II", Fig. 3.3).

DNA was extracted from the bulk population of cells "I" and "II". As depicted in Figure 3.4 libraries were prepared separately for each sample. Then, a two-step nested PCR amplification was carried out. The first PCR amplification step is specific, in order to ensure accurate amplification of the integrated DNA in the cell's genome. The second PCR amplifies the gRNA sequence while adding the necessary barcode and adaptor sequences. After sequencing, the comparison of the representation of gRNAs before ("I") and after ("II") soft agar indicated the differential representation of gRNAs (and subsequently their respective genes) between the two populations (Fig. 3.4).

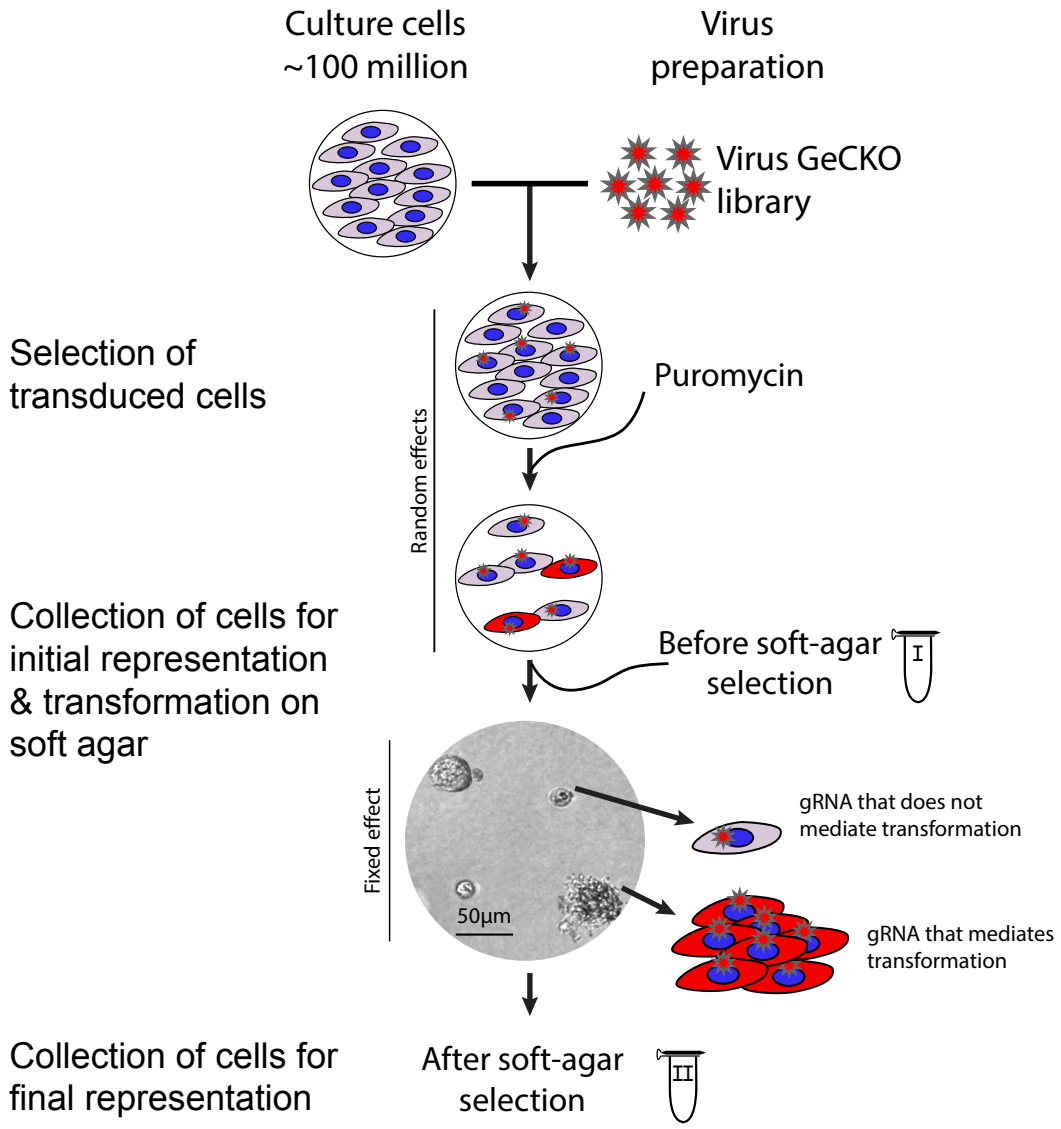


FIGURE 3.3: Scheme of the transformation screen. 100 million cells are transduced with the GeCKO virus library. The transduced cells are then selected for puromycin resistance. After selection, a sample of cells (approximately 1/3) is collected in order to have the initial gRNA representation in the cell population. The cells are then cultured in soft agar. Transformed cells grow in soft agar forming clumps in comparison to non-transformed cells. After one month the cells are collected in bulk as the final gRNA representation. For the analysis of the screen, the transformation assay is considered a fixed effect whereas the gRNAs and the infection of the cells are considered as random effects.

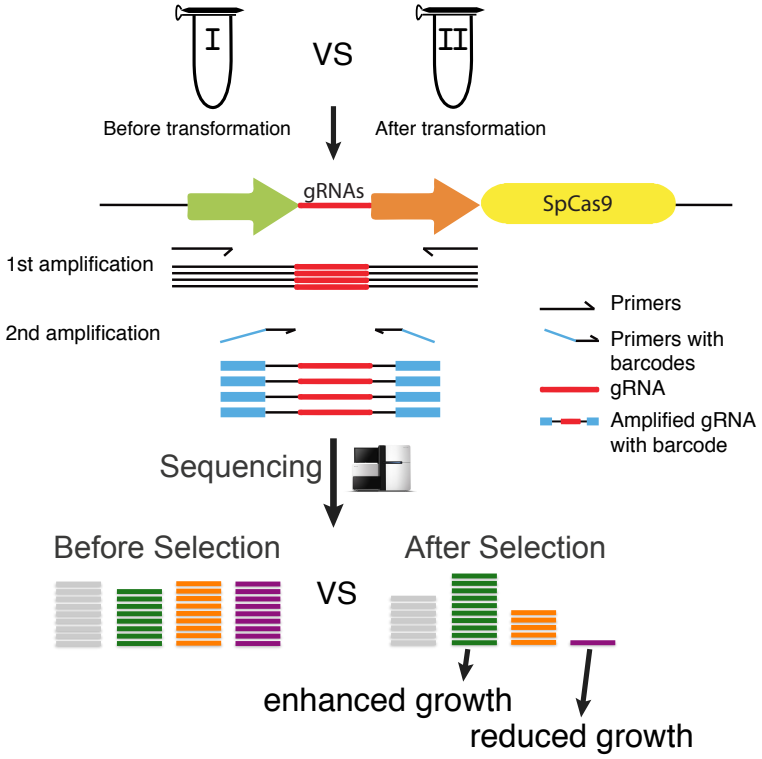


FIGURE 3.4: Library preparation. Nested PCR amplification strategy for CRISPR/-Cas9 vector (LentiCRISPRv2) and sequencing. First, the gRNA sequence is amplified with specific primers and then a more targeted amplification attaches the barcodes. The amplicons are then subjected to sequencing. Analysis of representation of gRNAs prior ("I") and post ("II") soft agar indicates the differentially represented genes.

3.2.2 Analysis

The data produced from each screen consist of two replicates. Each replicate contains GeCKO library "A" and "B". Each library encompasses three gRNAs per gene and therefore in total both libraries contain six guides per gene (Fig. 3.2). Multiple gRNAs are designed in order to account for gRNA efficiency as well as potential off target effects. The gRNAs have been designed to have minimal off targets, but nevertheless off target effects still occur (Tsai et al., 2015). Hence, having multiple guides for the same gene, serves as a biological control. Therefore, if the majority of gRNAs targeting the same gene give a similar phenotype, the gene that is targeted by these gRNAs is considered as a likely true candidate.

Statistical approaches have therefore been designed to take into account gRNA variability. These approaches were adapted from shRNA libraries, which are known to have technical and biological variability (Hu and Luo, 2012). In such libraries, each gene is

investigated separately and ranked in a list according to their differential representation of shRNAs in the population. Afterwards statistical approaches assess whether there is a bias of gRNA clustering in the ranked list and they calculate an enrichment and significance score based on a permutation test (König et al., 2007; Luo et al., 2008). Another more recent statistical approach named MAGeCK (Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout) is designed specifically for CRISPR libraries (Li et al., 2014b). MAGeCK uses a ranking approach and it assumes that if a gene knockout has no effect on the phenotype, all guides will be randomly distributed in the list. On the other hand, if there is an influence on the phenotype of interest they will be clustered. With this rule significance of the target gene being a true hit can be derived (Li et al., 2014b). Lastly, caRpools (Winter et al., 2015) combines the use of a variety of statistical approaches and provides the ability to intersect the results from all tools.

Although the previous statistical approaches take guide variability by indirect methods, they fail to take the random effect of each guide as well as the combination of different libraries and experiments containing guides that target the same gene into account. Therefore a statistical model that takes multiple parameters into consideration was lacking.

We thus decided to develop ScrispR, a novel statistical analysis package that is designed to take fixed effects as well as random effects of a CRISPR/Cas9 screening experiment into account. (Materials and Methods 5.3.2). Technical effects are hereby treated as 'random effects' in order to account for effects that are specific to each guide. For example, the knockout efficiency of each guide can be dependent on its "GC" content, favored bases in the seed region, point mutations in the seed region, chromosome compaction state, position within target gene region, etc. (Doench et al., 2014; Wang et al., 2014a). Moreover, we have treated the GeCKO library as an additional random effect as it is separated in two parts (A & B) and processed them independently (e.g. transduction efficiencies might vary across libraries as well as the division of gRNAs across libraries). In contrast, the screen is considered as a fixed effect since measures only one effect: growth or no growth in soft agar.

A scheme of the analysis is depicted in Figure 3.5. Initially, ScrispR aligns the gRNAs to its respective reference genome. After their alignment, a count table depicting each individual gRNA is obtained. Quality controls test whether the sequencing depth is

sufficient and whether the data is reproducible – by using principle component and correlation analyses. A mixed effect model (MEM) is then employed in order to account for both fixed (e.g. before/after transformation) and random effects (e.g. gRNA) of the screen, as described above. The MEM model assesses whether the log-transformed abundance profile of all gRNAs per gene is different between two conditions before ("I") and after ("II") transformation and accounts for quantification differences between gRNAs and GeCKO libraries.

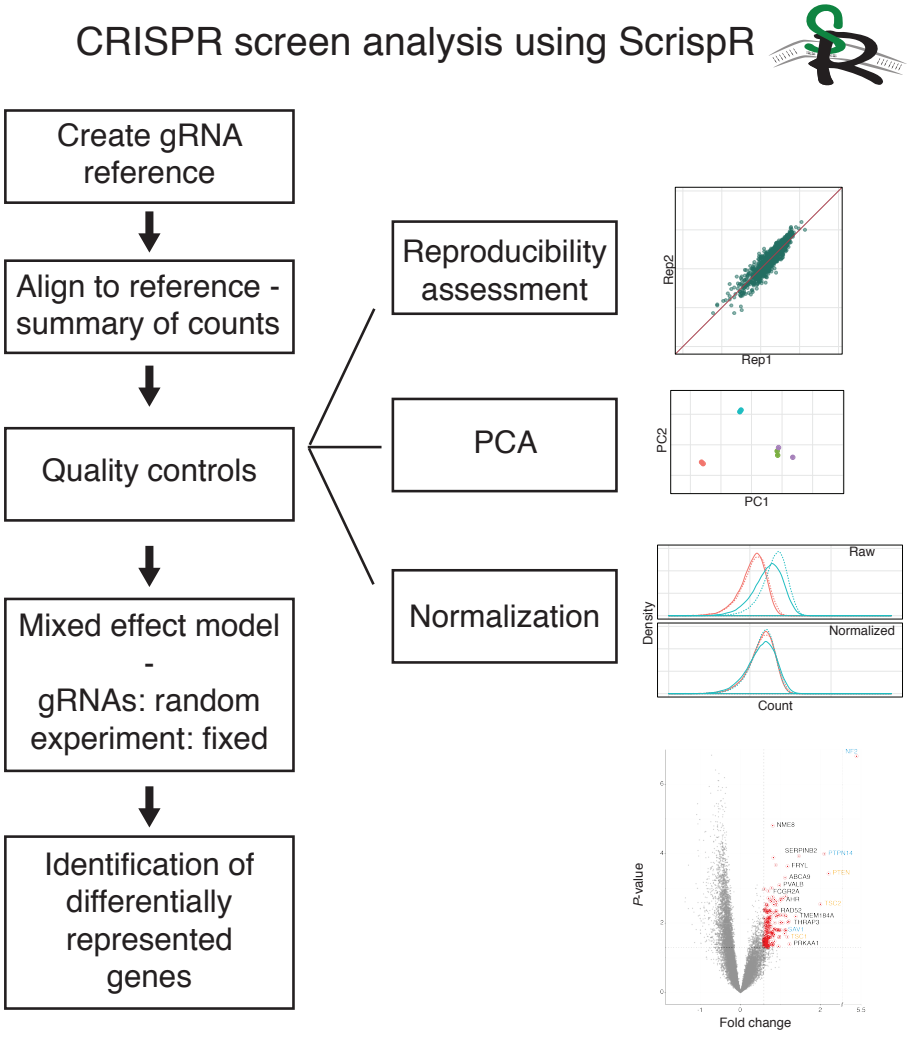


FIGURE 3.5: Scheme of ScrispR CRISPR/Cas9 screen analysis. Initially the raw sequenced data is aligned to the reference, which is defined by the user. Next, quality controls are performed in order to assess the reproducibility of the screen. Lastly, using a MEM, ScrispR assesses the significance of a gene being overrepresented or underrepresented in comparison to a control population.

3.2.3 Screen results

Selected hTERT-RPE-1 (hereafter RPE) and MCF10A cell lines were screened and analyzed by ScrispR (Table 3.1 and Fig. 3.5). Initially quality controls were conducted for each screen in order to evaluate the reproducibility and the sequencing depth (Figure 3.6 demonstrates the reproducibility of the screens in the $RPE^{TP53^{-/-}}$). Next, ScrispR tests the correlations between the two replicates. In $RPE^{TP53^{-/-}}$ the gRNA counts both before and after agar selection are well correlated demonstrating the reproducibility of the screen. The data is then normalized to sequencing depth. Principle component analyses also indicate the reproducibility of the screen after normalization. Finally, density plots of gRNA count further demonstrate the success in reproducibility and normalization. The remaining cell lines, RPE^{WT} and $RPE^{TP53^{-/-},\,hyperploid}$, had also successful quality controls, which allowed me to proceed with the analysis (Appendix Fig. B.1). Unfortunately, the libraries of MFC10A and $MCF10A^{TP53^{-/-}}$ cell lines did not exhibit reproducibility after agar selection, indicating technical problems during the screening process (Appendix Fig. B.2). Since these libraries did not have sufficient quality as assessed by the control guides, I could not use these libraries for further analyses.

ScrispR then runs a MEM on the normalized data and outputs FDR corrected P-values for each gene (Materials and Methods 5.3.2). Plots illustrating P-values and fold changes for each gene were generated ("Volcano plots", Fig. 3.7).

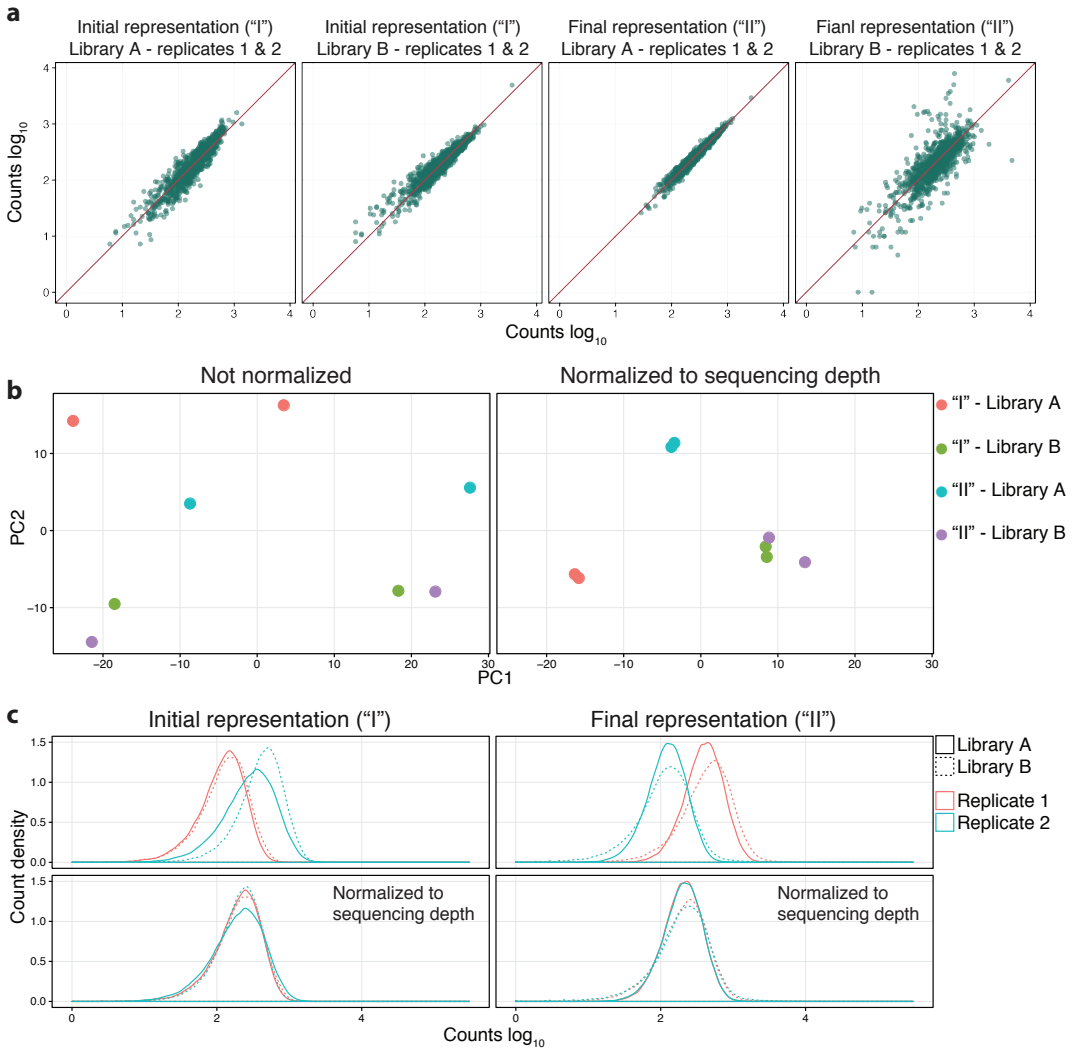


FIGURE 3.6: Quality controls of the screen. (**a**) The representation of each control guide replicate was plotted against each other in order to evaluate the reproducibility of the screen. (**b**) Principle component analysis plots indicate that after normalization the replicates are similar. (**c**) Density plots before and after normalization. Replicates overlap after normalization.

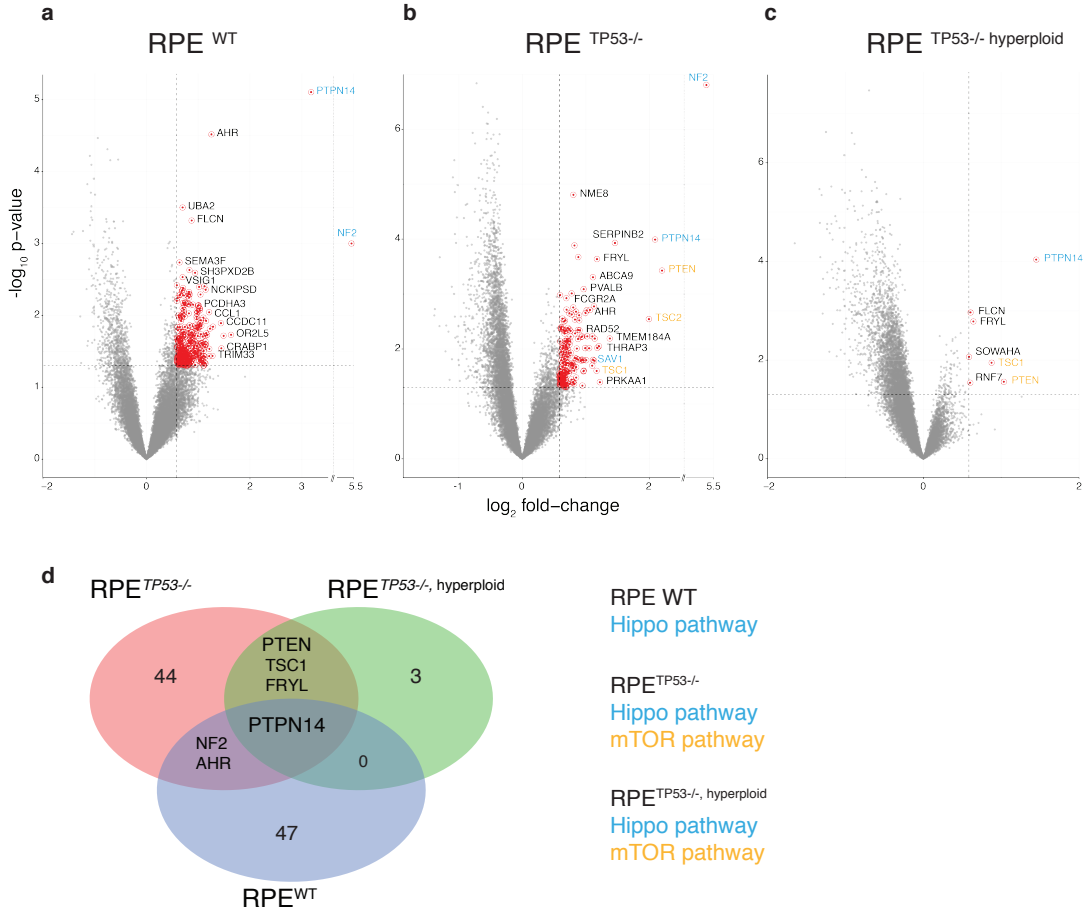


FIGURE 3.7: Volcano plots depicting overrepresented and underrepresented genes from each screen. Genes are highlighted according to P-value > 0.05 and fold change > 1.5. (**a**) RPE$^{\text{WT}}$ screen, (**b**) RPE$^{\text{TP53}^{-/-}}$ screen, (**c**) RPE$^{\text{TP53}^{-/-},\text{hyperploid}}$ screen. (**d**) Intersection of genes from the soft agar CRISPR/Cas9 screens. *PTPN14* is common hit in all screens. *NF2* and *AHR* are shared in the *TP53* knockout and wild type (WT) cells. *PTEN*, *TSC1* are genes related to the mTOR pathway and were only shared between the *TP53* knockout cell lines. *FRYL* was also only shared in the *TP53* knockout cell lines. There were no common significantly enriched genes between *TP53* knockout hyperploid and the WT cells.

The plots demonstrate two populations of cells: underrepresented and overrepresented genes. In all screens investigated, the underrepresented genes were enriched in essential pathways, such as ribosome biogenesis, transcription, cell cycle, etc (Appendix Table B.2). Knockouts in these pathways result to slower growth, senescence, apoptosis or necrosis, therefore it explains the underrepresentation of these gRNAs in the population of cells after transformation. Additionally, in the soft agar screen, genes that do not contribute to transformation are also expected to be underrepresented. In spite of the enrichment in essential pathways in the underrepresented gene list, essential genes cannot be reliably evaluated using soft agar screens. Such screens have multiple parameters

(e.g. selection), therefore assessment of essential genes can be biased by the soft agar enrichment step.

My interest lies in the overrepresented gene lists, since they encompass enrichment of gRNAs in response to their respective gene knockouts. In these experiments the over-represented gRNAs likely represent contributors to cell transformation thus potential tumor suppressors. Top overrepresented candidates were known tumor suppressor genes (such as *NF2* and *PTEN*), which stresses the potential of the method to identify genes important for the cancer development (Table 3.2).

TABLE 3.2: Top candidate genes that were enriched in the isogenic RPE cell lines. The gens are sorted according to fold change and the *P*-values are indicated.

RPE^{WT}			$RPE^{TP53^{-/-}}$			$RPE^{TP53^{-/-},hyperploid}$		
Gene	**Fold change**	***P*-value**	**Gene**	**Fold change**	***P*-value**	**Gene**	**Fold change**	***P*-value**
NF2	44.0	1.00E-03	*NF2*	41.9	1.56E-07	*PTPN14*	2.7	9.18E-05
PTPN14	9.1	7.84E-06	*PTEN*	4.6	3.73E-04	*PTEN*	2.0	2.74E-02
OR2L5	3.1	1.86E-02	*PTPN14*	4.3	1.02E-04	*TSC1*	1.8	1.13E-02
PSKH1	2.8	1.92E-02	*TSC2*	4.0	2.86E-03	*FRYL*	1.6	1.65E-03
CRABP1	2.7	2.86E-02	*SERPINB2*	2.8	1.17E-04	*FLCN*	1.5	1.08E-03
CCDC11	2.7	1.28E-02	*TMEM184A*	2.6	6.43E-03	*RNF7*	1.5	2.87E-02
SHISA6	2.4	1.43E-02	*PRKAA1*	2.3	4.02E-02	*SOWAHA*	1.5	8.56E-03
TRIM33	2.4	3.66E-02	*THRAP3*	2.3	8.89E-03			
AHR	2.4	3.03E-05	*TCEAL7*	2.3	9.56E-03			
CPNE8	2.3	1.53E-02	*FRYL*	2.3	2.31E-04			
PAX7	2.3	2.91E-02	*TSC1*	2.3	2.51E-02			
CCL1	2.3	9.03E-03	*GOLGA8B*	2.2	6.46E-03			
F8	2.3	3.69E-02	*KRTAP9-2*	2.2	1.63E-02			
OR10A4	2.3	1.86E-02	*NAPB*	2.2	1.67E-03			
CRISPLD1	2.2	1.77E-02	*NAA30*	2.2	1.58E-02			
OR4X1	2.2	2.54E-02	*ABCA9*	2.2	4.94E-04			
CFHR2	2.2	1.19E-02	*C6orf222*	2.1	5.89E-03			
NCKIPSD	2.2	4.40E-03	*SAV1*	2.1	2.01E-02			
MAP10	2.2	4.67E-02	*PRAMEF8*	2.1	1.94E-03			
FZD7	2.2	2.53E-02	*RSPH10B2*	2.1	9.71E-03			
EDC4	2.2	1.93E-02	*PDCD7*	2.0	5.89E-03			
NOX5	2.2	2.75E-02	*AHR*	2.0	1.98E-03			
KIRREL	2.2	3.77E-02	*IL6*	2.0	1.60E-02			
IFNL3	2.2	8.83E-02	*CAP1*	2.0	2.21E-03			
GRHL2	2.2	4.56E-02	*DDO*	2.0	9.65E-03			

Protein tyrosine phosphatase, non-receptor type 14 (*PTPN14*) was the only candidate which was significantly enriched in all screens performed in the RPE isogenic cell lines

(Fig. 3.7). *PTPN14* is a phosphatase involved in the PTP family. The PTP family genes play important roles in several cellular processes. PTPN14 protein was shown to assist LATS1 activation (pLATS1) that phosphorylates and inactivates YAP (pYAP – inactive), the final target of the Hippo pathway (Wang et al., 2012; Wilson et al., 2014).

The Hippo pathway is a kinase-cascade driven signaling pathway that is conserved across animal species. Physiologically it plays an important role in development and organ growth as well as in stem cell function. The core proteins of the Hippo pathway are the kinases MST1/2, which together with SAV1 activate LATS proteins. LATS1/2 phosphorylate YAP and TAZ that result in their nuclear export and therefore inactivate them. Active YAP and TAZ are transported to the nucleus and result in cell proliferation and inhibit cell death (Johnson and Halder, 2014). Tumor cells exploit this pathway by its inactivation, and consequently suppression of growth cannot be achieved via this pathway. Hence, in cancer it is considered as a tumor suppressor pathway and therefore its inactivation is evident in many malignancies (Johnson and Halder, 2014).

PTPN14 has been shown to support the activation of the Hippo tumor suppressor pathway but independently of MST1 and MST2 - core proteins of this pathway (Wilson et al., 2016). In summary, *PTPN14* acts as a tumor suppressor through activation of YAP. Furthermore, loss-of-function mutation of *PTPN14* has been identified in basal cell carcinoma, in support of its role as a tumor suppressor (Bonilla et al., 2016).

In addition to *PTPN14*, other genes involved in the Hippo pathway were also enriched such as *NF2* (Hamaratoglu et al., 2006; Li et al., 2014a). *NF2* was enriched in the RPE^{WT} as well as in the $RPE^{TP53^{-/-}}$ (Fig. 3.7d). *NF2* was not significantly enriched in the $RPE^{TP53^{-/-},\ hyperploid}$ cell line, although investigation of raw counts showed two guides indeed being enriched (Fig. B.5). Insignificance of *NF2* in the hyperploid cell line can be due to the inefficient modification of all the alleles resulting into knockouts. Hence, one allele that is still expressing a functional protein might be enough to buffer for the loss of the others.

Furthermore, *PTEN*, *TSC1* and *TSC2* were also enriched in the *TP53* knockout cell lines but not in the RPE^{WT} cells. *PTEN*, *TSC1* and *TSC2* are genes involved in the Mammalian Target of Rapamycin (mTOR). The mTOR pathway is well known to promote cell growth and cell cycle progression and therefore constitutively active mTOR pathway promotes tumorigenesis (Laplante and Sabatini, 2012). The enrichment of TSC1/2 in a

TP53 null genetic background suggests a genetic interaction between *TP53* clones and mTOR pathway genes. *TP53* has been shown to inhibit the mTOR pathway in cell lines as well as *in vivo*, confirming the previously suggested epistatic interactions between *TSC1/2* and *TP53* (Akeno et al., 2015; Hasty et al., 2013).

Encouraged by the results confirming the functions of known tumor suppressors as well as previously identified genetic interactions, I moved on to investigate the novel candidates that were enriched in these screens and studied their role as potential tumor suppressors.

3.2.4 Evaluating hits

My selection was based on the investigation of genes that are not known to be involved in transformation, uncharacterized genes and a set of known tumor suppressors. As a positive control I selected *NF2* since it was the top candidate and a known tumor suppressor as well as *PTPN14*, which recently was described as a tumor suppressor in basal cell carcinoma (Bonilla et al., 2016). For testing genetic interactions with *TP53* I selected *TSC1*, *TSC2* and *PTEN*, all involved in the mTOR pathway and shown to have epistatic effects with p53, protein of *TP53* (Akeno et al., 2015). Next I selected *AHR* which was shared between RPE^{WT} and $RPE^{TP53^{-/-}}$ as a gene that does not show a genetic interaction with *TP53*. I then selected *FRYL*, *SERPINB2*, *TMEM184A*, *THRAP3* and *C6orf222* all of which were enriched in the $RPE^{TP53^{-/-}}$ but not in the RPE^{WT} cells, indicating potential genetic interactions with p53. Lastly, I selected *RNF7* as a gene specific for $RPE^{TP53^{-/-},\ hyperploid}$ cell line (genes and functions are summarized in Table 3.3).

In order to assure that the genes lead to transformation, individual knockouts were generated for each gene. Two gRNAs from the screen were selected and a third gRNA was newly designed for each gene (Table B.4). The last gRNA was designed to assess whether an independently designed guide has the same effect on the phenotype as the gRNAs of the screen. After transducing each gRNA individually I subjected the cells to soft agar (Materials and Methods 5.3.3).

In order to measure the transformation potential of each gene I set up an assay based on crystal violet staining. Crystal violet is a dye that is able to bind to proteins and DNA, which allows easy visualization of colonies due to its violet color (Franken et al., 2006).

TABLE 3.3: Selected genes for verification and summary of their function.

Gene name	Summary of gene function
PTPN14	PTP family - negative regulator of the Hippo pathway (Wang et al., 2012)
NF2	Negative regulator of the Hippo pathway (Hamaratoglu et al., 2006)
TSC1	mTOR pathway (Laplante and Sabatini, 2012)
TSC2	mTOR pathway (Laplante and Sabatini, 2012)
PTEN	mTOR pathway (Laplante and Sabatini, 2012)
FRYL	Furry like homolog
AHR	Aryl hydrocarbon receptor – known for its biological responce to aromatic hydrocarbons (Murray et al., 2014)
SERPINB2	Plasminogen activator inhibitor-2 (Croucher et al., 2008)
TMEM184A	Transmembrane protein - May activate the MAPK kinase signaling pathway (Pugh et al., 2016)
THRAP3	Involved in pre-mRNA splicing, DNA-damage response protein (Beli et al., 2012)
C6orf222	Unknown
RNF7	Interacting with PCNA, component of a E3 ligase complex (Cooper et al., 2015)

After applying the cells to soft agar I collected the transformed clones in each well and transferred the cells back to plates so they can attach to the surface. After one week of culturing, the amount of colonies in the plate is a direct indication of the transformation potential of the respective knockout clone. In order to visualize the colonies I applied crystal violet and then the plates were photographed (Materials and Methods 5.3.3, Fig. 3.8).

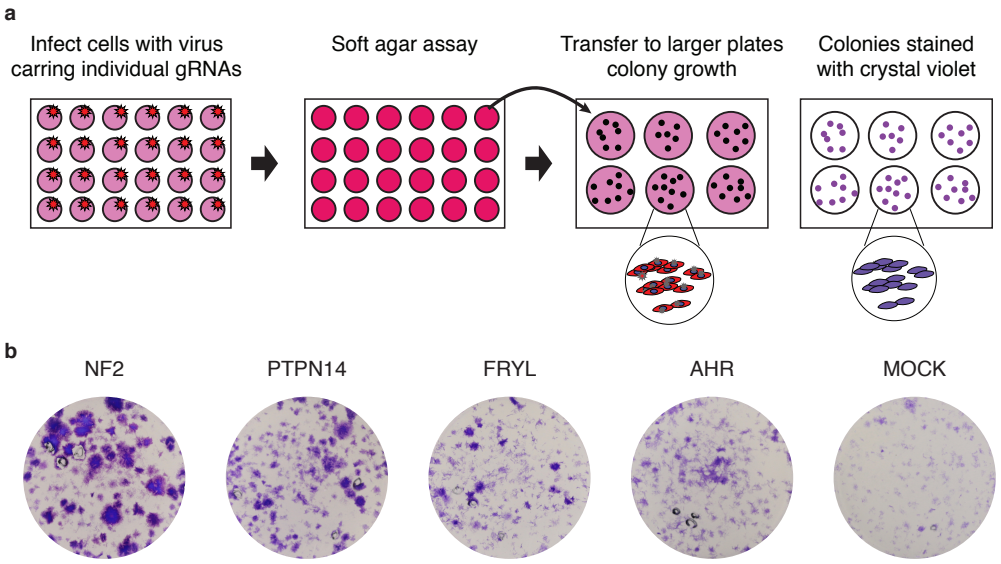


FIGURE 3.8: Scheme of crystal violet screen. (**a**) Cells are individually treated with the gRNAs to knockout the gene candidates. Then the cells are subjected to soft agar and grown for one month. Afterwards each well is transferred to a larger plate, to allow the cells to attach to the surface of the plate. Crystal violet was used to visualize the colonies. Then each plate was photographed and the intensity of crystal violet was measured by an in-house developed image analysis tool. (**b**) Raw photographs of $RPE^{TP53^{-/-}}$ cell colonies after applying crystal violet.

The intensity of crystal violet corresponds to the amount of colonies that grew on the plate and can be measured. The known tumor suppressors have expectedly high signals of crystal violet, indicating the feasibility of the method to detect cells that transformed and grew in soft agar. On the contrary, control samples (no virus, mock virus and virus targeting only GFP) cannot transform – and have low intensity levels from this assay (Fig. 3.9).

All RPE isogenic cell lines were tested. From the analysis, *NF2* and *PTPN14* had elevated growth indicating their high transformation potential. *TSC1*, *TSC2* and *PTEN* were mostly enriched in the $RPE^{TP53^{-/-}}$ cell lines but not in the RPE^{WT} cells suggesting the genetic interaction between mTOR proteins and p53. The remaining samples also showed more growth comparing to the controls, but a statistical test could not be applied since there were insufficient amounts of controls (Fig. 3.9a).

In order to examine whether the selected genes transform RPE cells specifically, I repeated the experiment in isogenic mammary epithelial cell lines (MCF10A). From the selected genes I continued with three positive controls (*NF2*, *PTPN14* and *PTEN*) and three new candidates (*AHR*, *FRYL*, *SERPINB2*). These candidates were selected due to their transformation efficiency in the RPE cells. Each plate was manually evaluated by the amount of colonies and colony sizes as well as their potential role in transformation. Both *AHR* and *SERPINB2* were suggested to increase proliferation and *FRYL* may have a link to the Hippo pathway (Couzens et al., 2013; Croucher et al., 2008; Murray et al., 2014). I performed the experiment in several replicates in order to have sufficient data for statistical testing. A t-test was applied for a pair-wise comparison in order to investigate the significance of each observation. All knockout conditions transformed the cells significantly (Fig. 3.9b) in comparison to mock knockouts (targeting GFP) as controls. This verified that the selected genes transformed two different epithelia cell lines and therefore can be considered as potential tumor suppressors.

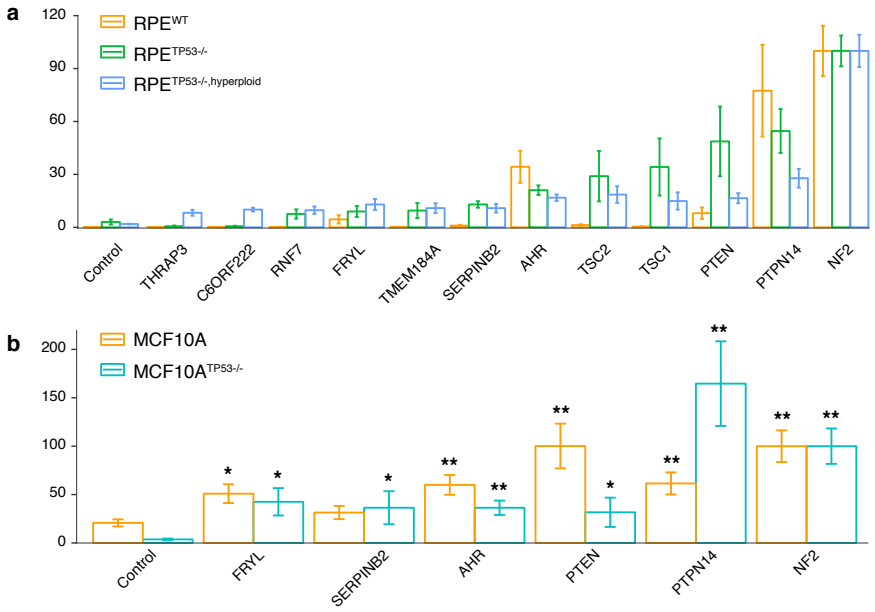


FIGURE 3.9: Results from the analysis of the crystal violet screen. Three guides per gene were grouped together. (**a**) Crystal violet intensities of the RPE cell lines normalized to the intensity of the NF2 knockout clones (n=3 replicates per gRNA). (**b**) Crystal violet intensities in the MCF10A isogenic cell lines normalized to the intensity of the NF2 knockout clones (n=6 replicates per gRNA). Significance was calculated by a t-test in comparison to the respective control group of each cell line (* < 0.05, ** $<$ 0.01).

Using the crystal violet assay, I was able to demonstrate the transformation potential of individual gRNAs in different cell lines. Known tumor suppressors indeed transformed the cells and were readily detected by the assay. Negative controls showed the opposite effect, indicating the functionality of the method. In summary, from the screen and from these experiments I identified three genes, *AHR*, *FRYL* and *SERPINB2*, that have not been previously characterized as tumor suppressors.

In the following section I focus on the investigation of the roles of *AHR* and *FRYL* in promoting cell growth and transformation.

3.2.5 Investigation of mechanism of transformation of *AHR*

Aryl Hydrocarbon Receptor (*AHR*) is a ligand activated transcription factor that induces transcription of several enzymes important for the metabolism of aromatic hydrocarbons (Murray et al., 2014). *AHR* is also known to be involved in various cellular processes, such as cell cycle, migration and immune function (Murray et al., 2014). Lack of *AHR* has been suggested to increase tumor formation after diethylnitrosamine treatment (a chemical

used in industry) in comparison to the wild type *AHR* (Fan et al., 2010). Additionally, in a prostate tumor mouse model TRAMP$^{\text{AHR}^{-/-}}$ has been reported to increase the incidence of tumorigenesis (Fritz et al., 2007). On the other hand, constitutive activation of *AHR* has also been linked to carcinogenesis (Andersson et al., 2002). In summary, there is evidence of involvement of *AHR* in carcinogenesis via two different mechanisms: loss of *AHR* or activation of *AHR*.

AHR loss-of-function might also mediate tumorigenesis through *RB* and β-catenin (*CTNNB1*). *AHR* was shown to reduce the phosphorylation of retinoblastoma protein *RB* (a known tumor suppressor), which when phosphorylated, results in the blockage of cell cycle progression (Puga et al., 2000). Additionally, *AHR* promotes degradation of *CTNNB1*, a protein involved in the WNT signaling pathway, whose constitutive activation by mutation is found in several cancers (Ohtake et al., 2009).

In order to investigate the exact function of *AHR* during transformation, I first generated cell lines with *AHR* loss-of-function alleles. Using individual gRNAs that target *AHR* specifically, I isolated several clones that had potentially *AHR* gene disruption. In order to verify the knockouts, I tested for loss of protein expression by immunoblotting (Materials and Methods 5.1.5). Unfortunately the assay was unable to detect AHR protein, although from mRNA sequencing data we had identified *AHR* expression (data not shown). Therefore I sought to investigate the genomic locus and determine whether the modification produces a potential knockout. I developed a method based on Sanger DNA sequencing to identify potential knockouts (see Materials and Methods 5.3.9). This method uses the information from the chromatogram produced by Sanger DNA sequencing, which indicates the presence of two bases instead of one clear base peak. Using this information I was able to decipher the sequence of both alleles and therefore draw conclusions if I have a frameshift in the coding sequence, which will result in a knockout (Materials and Methods Fig. 5.4).

In summary with this method I was able to identify six knockout cell lines out of 13 cases. I also identified two cell lines with no allele modifications whereas five cases were inconclusive (Table 3.4).

Then I investigated the localization of *AHR* by immunofluorescence (see Material and Methods 5.3.6). I investigated two antibodies with two different cell fixation conditions.

Both antibodies showed different subcellular localizations of AHR protein, requiring further investigation and optimization steps (Appendix Fig. B.4).

TABLE 3.4: Summary of knockouts of *AHR* and *FRYL* clones in RPE cell lines

	Knockouts	**WT**	**N.A.**	**Total clones**
AHR	6	2	5	13
FRYL	6	2	4	12

3.2.6 Investigation of the transformation mechanism of *FRYL*

Another interesting candidate the screen revealed is the Furry-homolog like (*FRYL*) gene. This protein shares 60% similarity to *FRY*, which plays a role in spindle pole integrity during mitosis (Ikeda et al., 2012). FRYL was also identified to have a physical interaction with SLMAP (Couzens et al., 2013), which belongs to the striatin-interacting phosphatase and kinase complex (STRIPAK) that regulates a variety of cellular processes, such as cell signaling, cell cycle, apoptosis and migration (Madsen et al., 2015). SLMAP was shown to interact with MST1/MST2 (Couzens et al., 2013), central components of the Hippo pathway (Harvey and Tapon, 2007). Interestingly, the phosphatase PP2A of the STRIPAK complex has already been shown to negatively regulate the Hippo pathway, which results in cell growth (Ribeiro et al., 2010).

Similarly to *AHR*, in order to study the mechanisms of transformation of *FRYL*, I performed experiments to identify knockouts. Immunoblotting experiments did not work for FRYL protein although *FRYL* is expressed (data not shown). Therefore, as described above, I used the method I developed based on Sanger DNA sequencing. With this method, I was able to detect six knockout clones, two wild type clones and four cases were inconclusive (Table 3.4).

Next, I sought to investigate the localization of *FRYL* (Materials and Methods 5.3.6). I performed immunofluorescence with an anti-FRYL antibody. I observed FRYL to localize in foci with a subcellular localization resembling P-bodies (Fig. 3.10). P-bodies are mRNA processing bodies containing aggregates of riboproteins, mRNA as well as proteins involved in RNAi. Thus they are involved in mRNA repression and decay (Eulalio et al., 2007; Parker and Sheth, 2007).

Next, to verify that FRYL localization is not an artifact, I tested whether *FRYL* knockout clones abolishes FRYL localization. I performed immunofluorescence on all these cell lines (Table 3.4). I observed that for four out of the six knockout cell lines there was indeed no foci localization of FRYL. The two functional FRYL cell lines showed the expected foci. As for the unclear cases, all of them showed a mixed signal of cells: some with foci and some without. The latter could be due to mixed populations of cells, which may be the reason of unsuccessful analysis with the Sanger sequencing data.

Future experiments will employ siRNAs to knock down *FRYL* and check the FRYL protein localization with immunofluorescence. This experiment will show whether FRYL indeed localizes to the P-body like foci. Additionally, in order to assure that the foci I observed are indeed P-bodies, I will perform immunofluorescence to identify AGO2 localization. AGO2 is an important protein involved in the RNAi machinery, which is known to localize at P-bodies (Eulalio et al., 2007; Parker and Sheth, 2007).

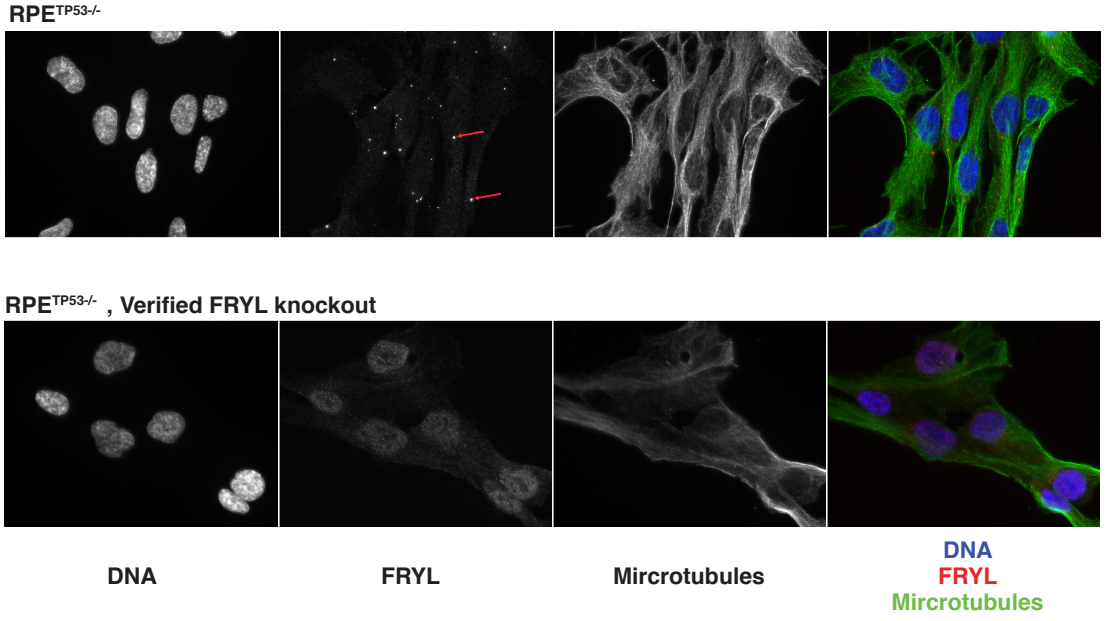


FIGURE 3.10: Immunofluorescence of RPE$^{TP53^{-/-}}$ stained for DNA, FRYL and microtubules (alpha-tubulin). RPE$^{TP53^{-/-}}$ shows the appearance of foci in the FRYL plain in comparison to the RPE$^{TP53^{-/-}}$ *FRYL* knockout cell line.

3.2.7 Structural alterations in transformed cell lines

In order to investigate the mechanisms of transformation, clones from the RPE cell lines were picked from every condition prior testing with crystal violet. Based on previous screens in the laboratory, a mechanism of transformation is by acquisition of somatic

copy number alterations (SCNAs) in the genome (Mardin et al., 2015). These alterations can alter the genome in various ways resulting in deregulation of the transcriptome and can result in activation of pathways leading to transformation. Hence, I investigated whether these clones acquired any SCNAs. We sequenced 69 cell lines with mass LC-WGS (Materials and Methods 5.1.7) and investigated gross SCNAs in these clones based on read depth analysis (Table B.1).

Analysis of all read depth plots investigated revealed that 28 out of 69 clones sequenced had at least one alteration (Fig. 3.11). More than one alteration was observed in 14 of the clones, of which seven of them were *NF2* clones and three *PTPN14*. This suggests that inactivation of the Hippo pathway may lead to increase number of SCNAs. This could be due to the increased proliferation rate of the cells, resulting in improper repair of potential DNA damage while continuing the cell cycle. To statistically evaluate this, I will sequence spontaneously transformed cells as controls as well as more clones from knockouts of tumor suppressor genes of the Hippo pathway.

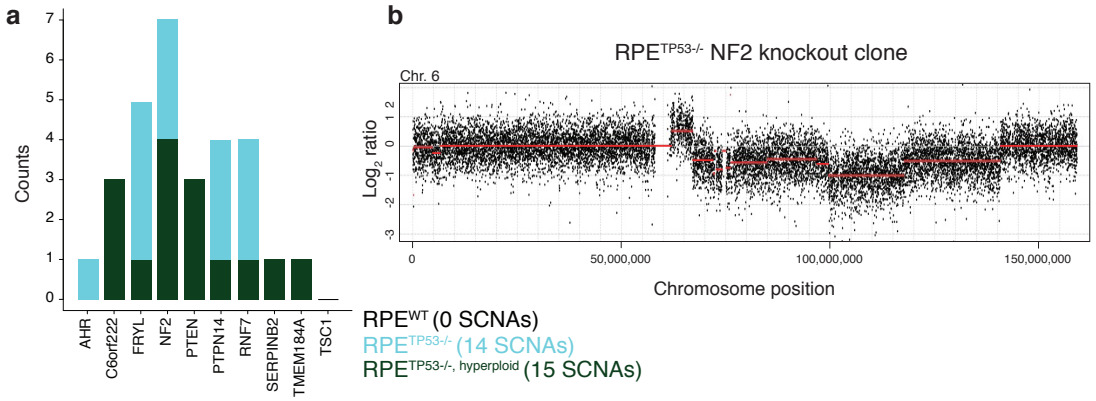


FIGURE 3.11: Summary of SCNAs in sequenced clones. (**a**) Amount of SCNAs observed in $RPE^{TP53^{-/-}}$ and $RPE^{TP53^{-/-},hyperploid}$. RPE^{WT} did not exhibit any SCNAs. (**b**) Example of sequencing read depth of $RPE^{TP53^{-/-}}$ with *NF2* knockout which carries extensive SCNAs on the q arm of chromosome 6.

3.2.8 Screen to identify potential growth suppressors

In order to understand whether the transformation is caused by growth advantage or mechanisms specialized to anchorage independent growth, I performed screens with no selection but growth as a constraint (Fig. 3.12). I subjected two isogenic cell lines to the growth screen, $RPE^{TP53^{-/-}}$ and $RPE^{TP53^{-/-},\,hyperploid}$. The cells were grown for two

weeks being split every three days. Cells were then collected and libraries were prepared and sequenced.

Analysis using ScrispR between day 0 and day 14 was performed according to the default settings of the program. The samples had successful quality controls, which allowed me to proceed with the analysis (Appendix Fig. B.3). From the analysis *P*-values and fold changes of each gene were calculated and then volcano plots were generated (Fig. 3.12b).

The intersection between overrepresented genes of both $RPE^{TP53^{-/-}}$ and $RPE^{TP53^{-/-},hyperploid}$ was further investigated. As depicted in Figure 3.12, 75 genes were commonly enriched for both screens. Gene ontology indicated significance of several pathways including Hippo, mTOR and ubiquitin related pathways (Appendix Table B.3). Genes involved in the Hippo pathway are *NF2*, *PTPN14*, *SAV1*, *LATS1*, *LATS2*, *AMOTL2* and *TAOK1*. Interestingly *TAOK1* knockout was highly enriched in the no-selection growth screen but not in the transformation screen. *TAOK1* is shown to activate the Hippo pathway by activating *MST1/2* (Boggiano et al., 2011; Poon et al., 2011). Furthermore *TAOK1* has shown to regulate chromosome congression and checkpoint signaling indicating multiple roles of *TAOK1* (Draviam et al., 2007). This suggests that inactivation of the Hippo pathway promotes growth, as expected. Next, genes in the mTOR pathway were enriched: *PTEN*, *TSC2*, *TSC1*, *CAB39*. Although, these genes point to the same genetic interaction with TP53, complete evaluation is not possible since the wild type no-selection growth screen has not been performed yet.

Interestingly, genes involved in ubiquitination and neddylation were enriched in the no-selection growth screen. Genes important for neddylation included *CUL3*, *CAND1*, *KEAP1*, as well as E2 ligases *UBE2M* and *UBE2F*. These proteins are involved in activation of cullins that eventually contribute to ubiquitin transfer to proteins and target degradation (Petroski and Deshaies, 2005). Cullins target genes for degradation which are involved in the cell cycle (Soucy et al., 2010; Tateishi et al., 2001). Apart from neddylation, genes involved in ubiquitination were enriched: *RNF7*, *KLHL21*, *ARIH1*, *USP47*, *UBE2L3* and *KCTD10*). *KCTD10* is an interesting candidate as it was also highly enriched in the no-selection growth screen. *KCTD10* was shown to promote the cell cycle presumably via interaction with proliferating cell nuclear antigen (PCNA), an important protein for replication and cell cycle progression (Wang et al., 2009). Furthermore *KCTD10* was shown to interact with CUL3 (also top candidate of the screen) and

NOTCH1 and it promotes NOTCH1 degradation (Ren et al., 2014). Notch1 pathway has shown to be upregulated in cancer (Rizzo et al., 2008). Hence, this data suggests that disruption of specific genes in the ubiquitination pathway lead to enhanced growth.

From the no-selection growth screen, the top candidates are indeed involved in pathways that were suggested to promote proliferation. Interestingly most of the genes were also enriched in the transformation screen, indicating that elevated growth itself is one of the main contributors for transformation.

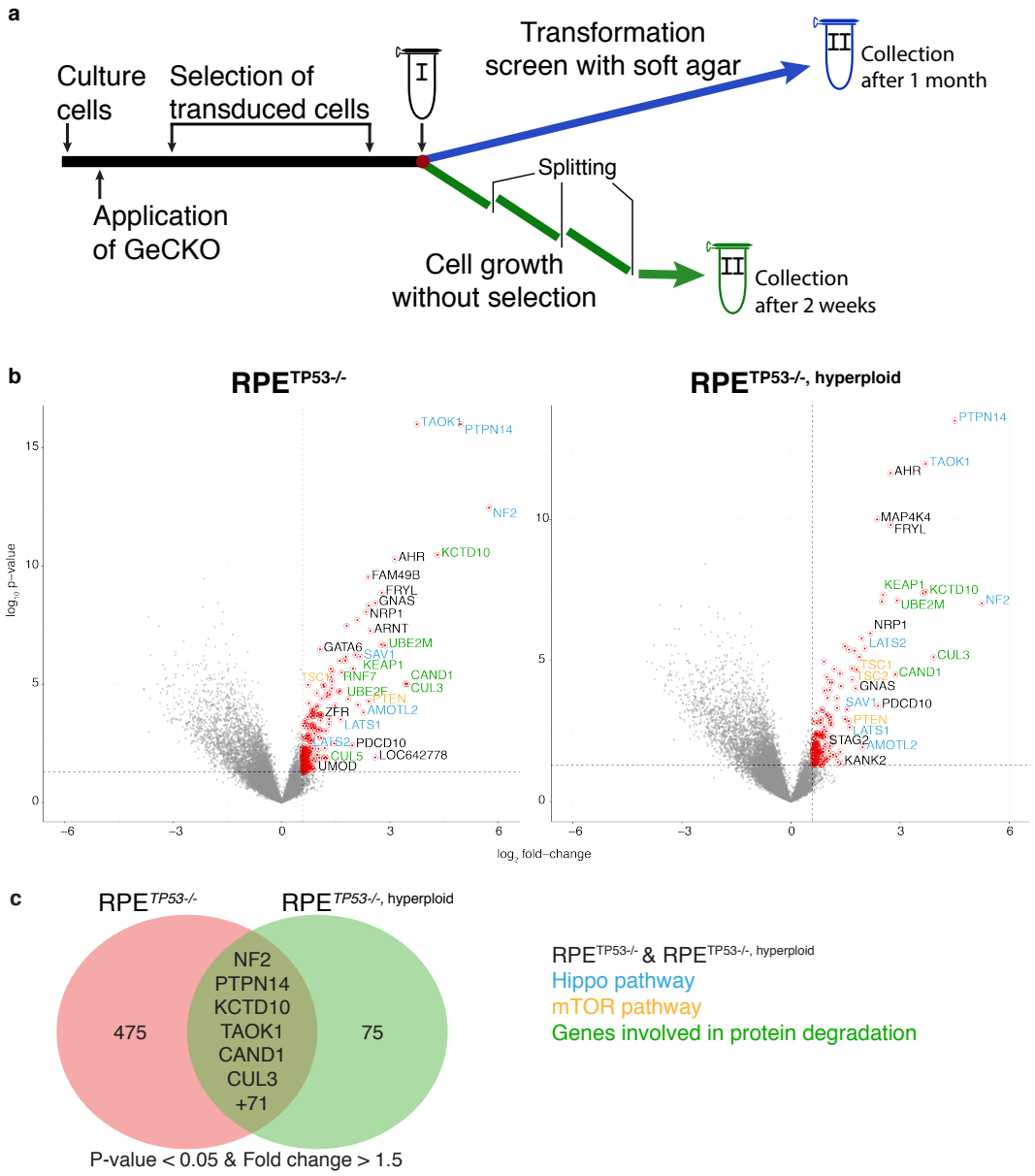


FIGURE 3.12: Scheme of GeCKO screens. (**a**) Approximately 50 million cells were cultured and transduced with the GeCKO virus library. Then cells were selected with puromycin resistance for virus uptake. A sample of cells was then collected as the initial representation. The cells were either embedded in soft agar (Fig. 3.3) or the cells were grown for two weeks having them split every three days. Lastly, the cells were collected for the final representation. (**b**) Volcano plots form no-selection growth screen. Many top candidates are shared with the transformation screen. (**c**) Intersection between $RPE^{TP53^{-/-}}$ and $RPE^{TP53^{-/-},\ hyperploid}$ cells.

3.2.9 Analysis of cancer genomes for mutational occurrence of the candidate genes

As an initial assessment as to whether candidates reflect true tumor suppressors in clinical tumors, frequencies of occurrence of point mutations in samples from The Cancer Genome Atlas (TCGA) were compared (analysis conducted by Christopher Buccitelli). We focused on point mutations, as their effects are gene-specific rather than large copy number variants, which involve many genes. The tumor types most enriched for our candidates included bladder carcinoma (BLCA), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC) and uterine corpus endometrial carcinoma (UCEC).

The most enriched gene across was the tumor suppressor *PTEN* as expected. Interestingly next was *FRYL* which is also frequently mutated to approximately 8% in BLCA, LUAD and UCEC. Combined with the growth promoting and transformation potential as found from the screens described in the previous sections, *FRYL* is suggested to function as a potential new tumor suppressor. Another interesting candidate is *KEAP1*, which is related to the ubiquitination pathway. *KEAP1* was commonly mutated in LUAD and LUSC with 19% and 12% respectively. KEAP1 protein has been suggested to play a role in cancer through regulation of NRF2 protein degradation (Hast et al., 2014; Jaramillo and Zhang, 2013). *NRF2* is important for oxidative stress and therefore cancer cells activate this response and acquire advantages in survival and growth. Only *CUL3* was mutated up to 6.7% in LUSC but none of the other genes of the ubiquitination pathway were as mutated as high as *KEAP1*.

Hippo pathway genes were under 3% mutational rate across cancers. Only *PTPN14* and *LATS1* had over 3% in five cancers tested with highest occurrence in UCEC (9.4% and 4.5% respectively). *NF2* though, the highest hit from the screen, was not as commonly mutated across cancers, a finding that came to surprise given its high growth and transformation potential from the screens. This could be also due to investigation of mutations and not larger SCNAs, which may lead to loss of the gene. mTOR pathway genes, besides *PTEN* that also signals through mTOR, *TSC1* and *TSC2* had higher mutation rates in the specific cancers tested but not across cancers.

Lastly, *AHR* and *ARNT* were mutated just over 3% in four cancers tested but none over 3% across all cancers. Further investigation as a complex as well as the mutational result

still needs to be conducted.

In summary, mutational frequency analysis of the candidate genes show that the candidates are mutated and potentially relevant across many different malignancies. Some candidates even show higher mutation frequencies than known tumor suppressors. This result further supports the role of the candidate genes in tumor suppression.

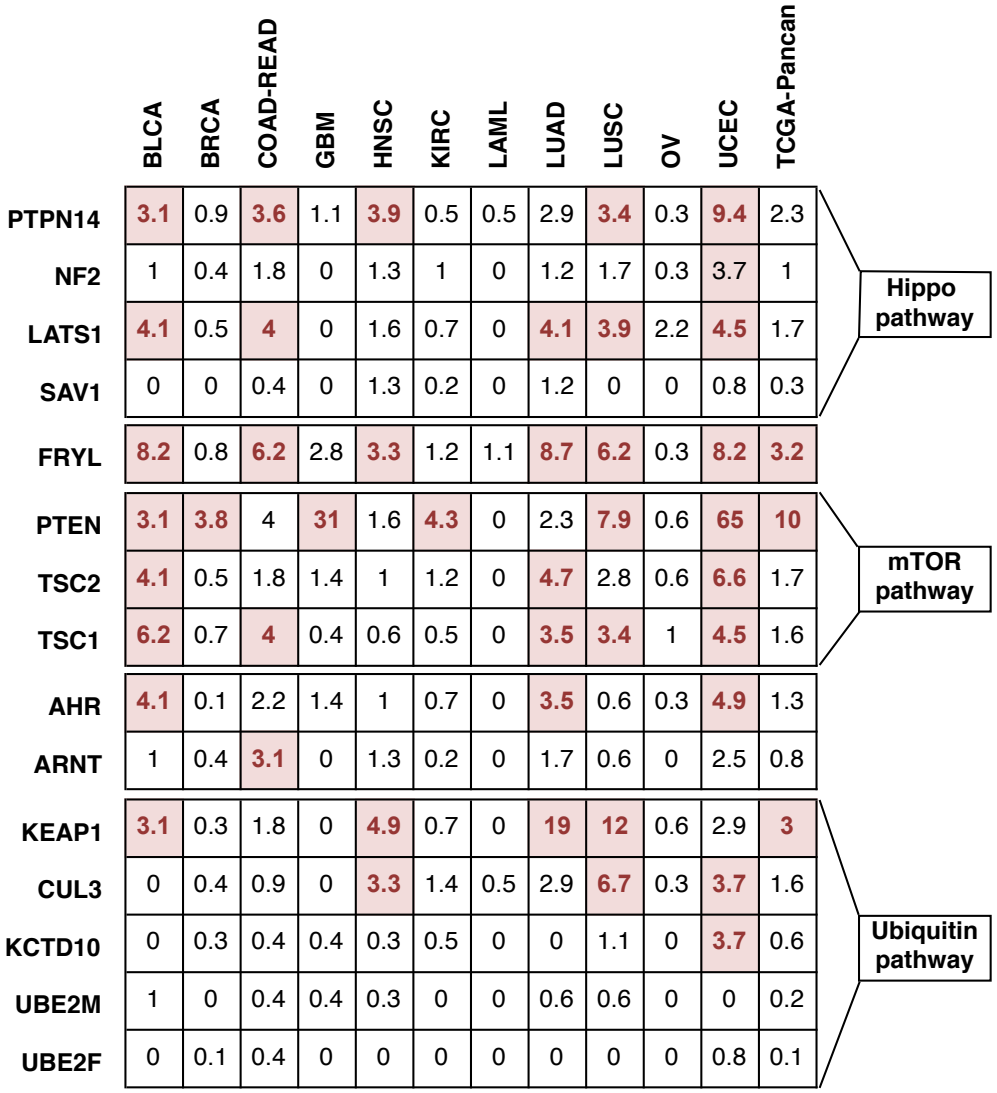


FIGURE 3.13: Mutational frequency of selected genes in most enriched cancers as well as across cancers (TCGA-Pancan). Mutational frequency over 3% is highlighted. BLCA, bladder carcinoma (N=95); BRCA, breast cancer (N=742); COAD-READ, Colorectal adenocarcinoma (N=85); GBM, Glioblastoma multiforme (N=138); HNSC, Head and Neck squamous cell carcinoma (N=299); KIRC, Kidney renal clear cell carcinoma (N=400); LAML, Acute Myeloid Leukemia (N=166); LUAD, lung adenocarcinoma (N=167); LUSC, lung squamous cell carcinoma (N=177); OV, Ovarian serous cystadenocarcinoma (N=154); UCEC, uterine corpus endometrial carcinoma (N=237); TCGA-Pancan, all cancers (N=2660).

3.3 Discussion

In this study, the transformation screen was designed to identify genes that can result in anchorage independent growth when knocked out (by measuring their ability to grow in soft agar). I hypothesized that known tumor suppressors will be among the top hits of the assay. *NF2*, *PTEN*, *TSC1* and *TSC2* - all known tumor suppressors - were among the most enriched genes, verifying the ability of the assay to identify tumor suppressors. The assay additionally revealed candidates of cell transformation, which are potential tumor suppressors. Two pathways were highlighted in our screen for their ability to induce cell growth and transformation, the Hippo and mTOR pathways (Hamaratoglu et al., 2006; Laplante and Sabatini, 2012; Li et al., 2014a).

The candidate that was shared in all screens was *PTPN14*. This phosphatase was shown to activate *LATS*, which subsequently leads to inactivation of *YAP*. PTPN14 protein activates the "executer" of the Hippo pathway, YAP, independently of the core Hippo proteins MST1/2, indicating that the Hippo signaling cascade could be activated downstream of the MST1/2 kinases. Furthermore, *PTPN14* inactivation has been identified in basal cell carcinoma (Bonilla et al., 2016), revealing its tumor suppressor role.

Other components of the Hippo pathway such as *NF2*, were enriched in the transformation screen in both RPE and MCF10A isogenic cell lines, indicating that inactivation of the Hippo pathway is able to transform cells. Therefore inactivation of the Hippo pathway leads to activation of YAP through phosphorylation (Harvey and Tapon, 2007), which results in proliferation. Despite Hippo pathway inactivation transforms cells independently of the *TP53* status, in fact it is also connected to *TP53*. It is evident that the Hippo pathway inactivation is enough to promote tetraploidization of cells by inevitably inactivating p53 (protein product of *TP53* gene) – demonstrating the connection of the Hippo pathway proteins and p53 (Ganem et al., 2014). Hence, in the RPE^{WT}, inactivation of the Hippo pathway might work in two layers, inactivation of *TP53* and promotion of growth through YAP activation. In contrary to the *TP53* knockout cell lines Hippo inactivation promotes only growth.

In addition to the Hippo pathway, *NF2*, the top candidate of most screens, was also shown to block anchorage independent growth (Bosco et al., 2010; Chiasson-MacKenzie et al., 2015; Curto et al., 2007; Okada et al., 2005). NF2 was shown to inhibit EGFR

in confluent cells, a receptor important for mitogenic signaling (Tomas et al., 2014). Knockout cell lines of *NF2* can result in constant signaling of EGFR which promotes proliferation. Additionally, NF2 is shown to suppress RAC, a gene important for cell cycle and contact inhibition (Bid et al., 2013), and sustain growth signaling in confluent cells. Thus, despite NF2's function in the Hippo pathway it functions in yet another layer, anchorage independent growth.

Moreover, cell detachment or anchorage independence has also been shown to activate the Hippo pathway (Zhao et al., 2012). Cells were found to activate the Hippo pathway when led to anoikis – a state where the cell detaches from the colony. The exact mechanism of Hippo activation has not been formally demonstrated, however it is tempting to speculate that it may function through *NF2*, suggesting *NF2* as a probable link between anchorage independence and Hippo pathway.

In addition to the transformation screen, *NF2* was found to be the most enriched gene in the no-selection growth screen. This result indicates that growth signaling is pronounced in *NF2* knockout cells and in combination with the contact inhibition functions of *NF2*, there is substantial growth also in the transformation screen.

Lastly, given that *NF2* drives considerable growth through multiple pathways, acquisition of structural rearrangements may be an expected outcome as described below. The cells signal for constant growth, thus reducing the time necessary to repair any damaged chromosomes. If the damage is not threatening for cell survival, it may get repaired improperly and altered chromosomes can propagate to daughter cells. Additionally, SCNA acquisition was observed only in *TP53* knockout clones ($RPE^{TP53^{-/-}}$ and $RPE^{TP53^{-/-,\,hyperploid}}$) but not in the RPE^{WT} cells (Fig. 3.11). This suggests that in the RPE^{WT} cells, p53 gets activated upon DNA damage, which is corrected by the repair machinery before continuing the cell cycle. Therefore, *NF2* results in SCNAs but as an indirect outcome of the cells' ample growth, rather than playing a direct role in DNA damage.

In summary, *NF2* is shown to act on multiple pathways that all converge into accelerated growth and consequently transformation.

Another pathway, which appeared only in the context of $TP53^{-/-}$ cell lines, is the mTOR pathway. Main regulators of mTOR, *TSC1* and *TSC2*, were both identified in the

TP53$^{-/-}$ cell lines, which suggest that active p53 protein prevents transformation through mTOR. Recent studies have shown that mTOR is highly activated upon p53 loss, resulting in proliferation (Akeno et al., 2015). Therefore the transformation screen was able to identify epistatic effects between *TP53* and known tumor suppressors. These genes were also enriched in the no-selection growth screen, but to confirm the epistatic effect in this screen, a no-selection growth screen should also be conducted in the RPE$^{\mathrm{WT}}$ cells.

Novel candidates, *AHR* and *FRYL* were enriched and verified in both RPE and MCF10A isogenic cell lines. Both have a potential influence on cell growth as described below.

Aryl Hydrocarbon Receptor (*AHR*) is an important transcription factor for the metabolism of aromatic hydrocarbons (Murray et al., 2014). Thus, *AHR* plays an important physiological role in protection of the cells of potentially toxic contaminants (Denison and Nagy, 2003). AHR binds with the nuclear protein ARNT (Aryl Hydrocarbon Receptor Nuclear Translocator) and the dimer binds highly to DNA and activates xenobiotic responsive elements, such as *CYP1A1* (Ikuta et al., 2000). Besides AHR's response to xenobiotic compounds (as described in the results section 3.2.5) it is also involved in various cellular processes (Murray et al., 2014; Puga et al., 2009). There is indication that *AHR* is involved in cell cycle progression and apoptosis although the exact mechanisms are unknown (Puga et al., 2009). Some may include reduction of Rb phosphorylation by AHR (Puga et al., 2000) or degradation of CTNNB1 (Ohtake et al., 2009), both will result in progression of the cell cycle.

Studies disrupting a ligand binding domain resulted in a constitutive active *AHR*, where it induced numerous stomach tumors in mice (Andersson et al., 2002). On the other hand, *AHR*$^{-/-}$ deficient mice also showed an increased incident of tumors in prostate model mouse lines (Fritz et al., 2007).

In the present study I further support the view that *AHR* functions as a tumor suppressor with two lines of evidence. First, *AHR* was enriched and verified in the soft agar transformation screen in the RPE$^{\mathrm{WT}}$ and in the RPE$^{\mathrm{TP53}^{-/-}}$ cell lines, indicating that *AHR* could act independently of *TP53*. Second, *AHR* as well as nuclear transporter protein *ARNT*, were both enriched in the no selection growth screen, providing additional evidence that destruction of either protein of the dimer complex has an effect in cell growth. Based on these findings I hypothesize that *AHR* acts on several pathways that

converge on enhanced cell proliferation. Experimental testing of simultaneous activation/inactivation of these pathways need to be performed in order to identify the tumor suppressor function of *AHR*.

Another interesting candidate that was enriched in both screens (transformation and no-selection growth screen) was the Furry-homolog-like (FRYL) gene. This gene has not been previously characterized and the protein shares approximately 60% homology to FRY, a protein important for spindle pole integrity in mitosis. Based on their similarity one can hypothesize that FRYL protein might also play a role in mitosis. This may explain the genomic losses $RPE^{TP53^{-/-}}$ and FRYL knockouts carried indicating defects in the genome of these cell lines. SCNAs were only present in $RPE^{TP53^{-/-}}$ cells but not in the $RPE^{TP53^{-/-},\ hyperploid}$ cell lines, although *FRYL* was also enriched in the $RPE^{TP53^{-/-},\ hyperploid}$ transformation screen. This could be due to the low efficiency of knockouts in the cell line with extra chromosomes, and I did not detect any actual knockouts in the hyperploid cells. Additionally in the wild type RPE cells, having functional p53 protein is a constraint for allowing rearrangements to occur, similarly to NF2 as discussed before.

Investigating the role of FRYL in mitosis, for example with live cell imaging, can reveal a possible connection of this protein to cell cycle machinery.

Apart from mitosis, FRYL immunofluorescence showed localization of FRYL protein potentially in P-bodies in the wild type RPE cells but not in the $RPE^{TP53^{-/-}}$ verified clones. P-bodies are important for mRNA decay and therefore might play a role in affecting cell growth. Further investigation of colocalization of FRYL to P-bodies will be carried out with co-immunofluorescence with AGO2, a protein involved in the RNAi machinery and known to localize in P-bodies (Eulalio et al., 2007). In addition, siRNA experiments will be performed to see whether downregulation of *FRYL* will result in the same phenotype as the knockouts, i.e. no localization in P-bodies. The knockdown and colocalization with AGO2 experiments will reveal whether FRYL protein indeed localizes to P-bodies.

Moreover, a study showed an indirect connection of FRYL to the Hippo pathway through immunoprecipitation of Hippo genes (Couzens et al., 2013). FRYL was co-immunoprecipitated with SLMAP, a component of the STRIPAK complex. SLMAP has been shown to be

connected to the central Hippo kinases MST1/2 (Couzens et al., 2013). STRIPAK complexes play roles in many biological processes. Interestingly, the STRIPAK complex with PP2A (Protein phosphatase 2) has been proposed to be a negative regulator of the Hippo pathway (Ribeiro et al., 2010). The above suggests that *FRYL* may indeed be connected with the Hippo pathway. Investigation of YAP activation will indicate if there is such a connection. If YAP gets activated (or stays activated) in *FRYL* knockout cell lines in conditions that would normally be inactive, then the phenotype of enhanced growth in the transformation screen as well as in the no-selection growth screen can be explained.

Independently of the mechanism of *FRYL* action in transformation, *FRYL* was also enriched in the no-selection screen. This indicates that *FRYL* knockout clones proliferate faster than the rest of the knockouts, pointing to an acceleration in the cell cycle. Similarly to the *NF2* deficient clones $RPE^{TP53^{-/-}}$, accelerated growth seems to result in increase of SCNAs. This was evident only in the $RPE^{TP53^{-/-}}$ but not in the RPE^{WT} and $RPE^{TP53^{-/-},\,hyperploid}$ cells. An explanation for the WT cells is similar to *NF2* deficient clones, where functional p53 protein is the likely explanation of no SCNA acquisition. As for the hyperploid cells, extra chromosomes might result in inefficient modification of all *FRYL* alleles and thus I could not detect any complete knockouts.

In summary, I envision two possible functions of FRYL: involvement in cell division – presumably with a role in P-bodies – and regulating the Hippo pathway. The no-selection growth screen indicates that *FRYL* knockouts indeed are more represented in the final population which points to *FRYL* influencing cell growth. I will further test these possibilities using a variety of cell biological experiments, comparing WT cells to the deletion or depletion of *FRYL*.

Collectively, the sequencing results (as depicted in Fig. 3.11) point to a general mechanism of SCNA acquisition: accelerated growth. As described already for *NF2* and *FRYL*, both had acquired at least one SCNA in the $RPE^{TP53^{-/-}}$ cell lines but not in the WT cells. Similarly PTPN14 as well as RNF7 (a component of the SCF E3 ubiquitin ligase) had SCNAs in several sequenced clones in the $RPE^{TP53^{-/-}}$ cell line. In total 29 cells out of 69 sequenced cell lines acquired SCNAs (none for RPE^{WT}, 14/20 for $RPE^{TP53^{-/-}}$ and 15/34 for $RPE^{TP53^{-/-},hyperploid}$). None were identified in the RPE^{WT} cells, indicating the importance of *TP53* inactivation for SCNA accumulation. Also, apart from *C6orf222* and *TMEM184A*, all remaining genes sequenced (*AHR*, *NF2*, *FRYL*, *PTEN*, *PTPN14*,

RNF7, *TSC1*) were enriched in the no-selection growth screen. From these genes *TSC1* showed no SCNA accumulation, and *AHR* only had one clone to carry an SCNA. Despite this, for the remaining genes at least three clones harboring a SCNA were detected in the *TP53* deficient cell lines. This suggests that these genes, which accelerate growth, may indirectly lead to SCNA acquisition. Hence, the mechanism of SCNA formation in the tested knockouts is an indirect consequence of accelerated progression through the cell cycle resulting to insufficient time to repair potential errors.

From the transformation screen, *NF2*, *PTPN14*, *PTEN*, *SERPINB2*, *FRYL* and *AHR* were all verified in two different cell lines, retina and mammary epithelial cells. These results indicate that the transformation potential of these genes are not specific to a cell line but are able to drive transformation in different cell lines of epithelial origin.

Moreover, the no-selection growth screen was conducted in order to evaluate whether the results from the transformation screen are specific to transformation or whether they are dependent on activation of proliferation pathways. Interestingly, similar to the transformation screen, several Hippo pathway genes were enriched. One candidate, *TAOK1*, was highly enriched, but not in the transformation screen, indicating it as a candidate that is growth specific but not transformation specific. This hypothesis needs to be tested with a soft agar assay to assess whether the cells can grow unattached to a surface. Next, mTOR pathway genes *TSC1*, *TSC2* and *PTEN* were all enriched in the no-selection growth screen just as in the transformation screen. Lastly, the enrichment of genes related in the ubiquitin pathway and specifically to the neddylation pathway was interesting. E2 ligases, UBE2F and UBE2M, were enriched in $RPE^{TP53^{-/-}}$ and $RPE^{TP53^{-/-},\,hyperploid}$ cell lines as well as CUL3, a cullin activated by these E2 ligases with NEDD8 (the neddylation protein). Active CUL3 promotes ubiquitin transfer and degradation of proteins. These proteins are important for the degradation of proteins involved in cell cycle progression, indicating a driver role in the cells cycle (Petroski and Deshaies, 2005; Soucy et al., 2010; Tateishi et al., 2001). Knockouts of these ubiquitin related genes point to accelerated growth and hence may be potential tumor suppressors.

Lastly, the mutational frequency of the top candidates from all the screens conducted in actual tumors was assessed through analysis of the cancer genome atlas (TCGA). Only *PTEN* was highly mutated across all cancers (10%) and up to 65% in UCEC, which is expected as a known tumor suppressor gene. Other tumor suppressors such

as *TCS1*, *TCS2*, *LATS1* and *NF2* had mutational frequencies up to 6.6% but only in specific cancers analyzed. Interestingly *FRYL* had higher mutation rates from the known tumor suppressors tested, even further supporting its involvement in cancer. Together with the tumor promoting role and transformation potential *FRYL* can be suggested as a potential novel tumor suppressor gene. *AHR* and *ARNT* together had higher mutational frequencies in some cancers but not across all cancers in general, therefore these genes may be cancer specific. Further analysis needs to be conducted to find out whether these mutations are inactivating the genes. Finally, from the ubiquitination pathway *KEAP1* was highly mutated, which also has been suggested to play a role in cancer through *NRF2* stabilization and handling oxidative stress. In summary, the selected genes have relatively high mutation frequencies in specific cancers suggesting their possible involvement in tumor progression.

3.4 Conclusions

The screens carried out in this study were designed to identify contributors to cell growth under different conditions, which may be potential tumor suppressors. Both the transformation screen as well as the no-selection growth screen identified known tumor suppressors as well as novel genes that regulate cell growth. Two pathways were highlighted from both screens: the Hippo pathway and the mTOR pathway. The latter was also found to have an epistatic interaction with *TP53*, since these genes were only identified in the *TP53* knockout background but not in the wild type cells. The mTOR *TP53* genetic interaction has already been described, indicating that the screens can identify epistatic interactions.

From the transformation screen, a set of genes was selected for in depth investigation. The selection was based on positive controls, potential epistatic interactions and novel candidates. The selected genes were individually verified for their transformation potential. To further support the transformation effect of the selected genes, six genes, three controls and three candidates, were able to transform additionally mammary epithelia cell lines, MCF10A and MCF10A $^{TP53^{-/-}}$. This result shows the transformation potential of these genes in different epithelial cell types, indicating that they are not cell specific.

Also many of the identified genes play important roles in several pathways. For example *NF2*, acts as an activator of the Hippo pathway and suppressor of anchorage dependent growth. Therefore, it was not a surprise that *NF2* knockouts were highly enriched in all screens, since it promotes growth in multiple ways. *PTPN14* was also enriched in all screens which is also connected to the Hippo pathway. Moreover, *PTPN14* was only recently described as a tumor suppressor in a basal cell carcinoma. Here I further support the role of *PTPN14* in promoting growth in two different cell types.

From the remaining candidates I focused on *AHR* and *FRYL*. *AHR* has shown to enhance tumorigenesis by constitutive activation as well as downregulation. The knockout screen highlighted the role of *AHR* disruption, where it enhanced growth and promoted transformation. The latter was also verified in two different epithelial cell lines showing that it is not cell specific. Apart of *AHR*'s enrichment in the no-selection growth screens, astonishingly ARNT, the protein that binds and transfers AHR protein to the nucleus, was also enriched. This even further suggests the importance of AHR and specifically

the complex AHR-ARNT for growth regulation. AHR was suggested to activate the tumor suppressor RB protein, that inhibits progression of the cell cycle and also AHR was shown to promote degradation of CTNNB1 protein, which promotes growth. These findings indicate multiple functions of *AHR* to control cell growth and proliferation and suggest *AHR* as a potential tumor suppressor.

Besides *AHR*, *FRYL* was also enriched in the transformation screen but was specific to *TP53* knockout cell lines, which may indicate an epistatic interaction between *FRYL* and *TP53*. Confirming its growth promoting function even further, *FRYL* was enriched in the no-selection growth screen. Since the wild type screen has not been performed yet, the epistatic interaction of *FRYL* and *TP53* cannot be confirmed in the no-selection growth screen. This study suggests FRYL protein to potentially localize in P-bodies and the literature suggests a connection of *FRYL* with the Hippo pathway (Couzens et al., 2013). How *FRYL* promotes growth through P-bodies is still under investigation. As for the connection to the Hippo pathway, I will test whether *FRYL* knockouts lead to the activation of the YAP protein, the executor of the Hippo pathway. *FRYL* has not been characterized yet, making it an interesting candidate to investigate its effect in promoting growth and transformation and classifying it as a potential tumor suppressor.

Furthermore, apart from the Hippo and mTOR pathways being enriched in the no-selection growth screen, ubiquitin mediated degradation genes were also enriched. Certain genes, also specific to the neddylation pathway (part of the ubiquitin pathway), were connected in the same cascade indicating that either one of them being disrupted may result in promoting growth. Whether these genes also promote transformation still needs to be investigated.

Both transformation screen and no-selection growth screen complement each other. The transformation screen revealed tumor suppressors as expected, but also identified potential new contributors to cell transformation. The no-selection growth screen also revealed known tumor suppressors and contributors to cell growth, some of which were also enriched in the transformation screen. Therefore transformation can be considered to be highly influenced by constant growth signaling. In addition, several candidates have more than one function, many of which result in accelerated growth. Therefore, dysfunctional genes that result in enhanced growth and override anchorage dependence will be more enriched than other genes. Hence, the transformation screen can be biased

to select transformed clones that cover these two functions. Nevertheless, it revealed new candidate genes, which have not been previously suggested as promoters of growth and as tumor suppressors. Furthermore, the tumor suppressor role of the novel candidates can be supported by the higher mutational frequencies observed in tumors from TCGA, which may result in inactivation of the gene. A summary of the results is depicted in Figure 3.14, which indicates the targets and interactions of the candidates.

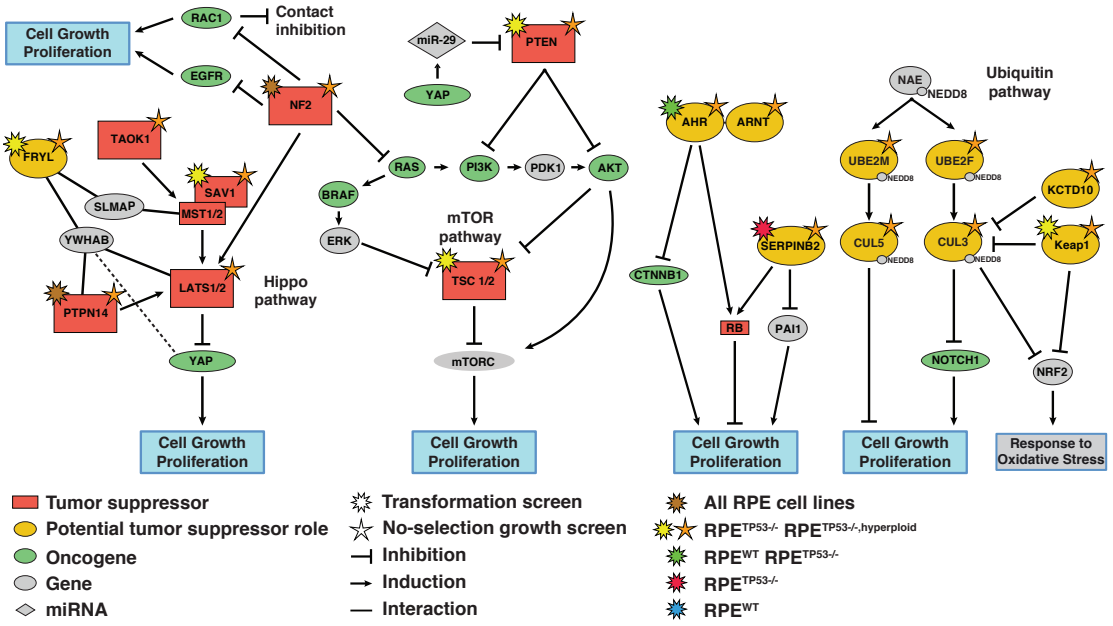


FIGURE 3.14: Summary of candidates and their role in pathways. All candidates result in promoting growth, indicating it as a large attribute to cell transformation and furthermore cancer progression. (Beli et al., 2012; Chen and Chen, 2016; Chien et al., 2005; Collins et al., 2013; Couzens et al., 2013; Croucher et al., 2008; Hariharan, 2015; Hasumi et al., 2015; Jaramillo and Zhang, 2013; Johnson and Halder, 2014; Mihaylova and Shaw, 2011; Murray et al., 2014; Ooi et al., 2013; Rattan et al., 2010; Ren et al., 2014; Rizzo et al., 2008; Samant et al., 2014; Soucy et al., 2010; Starheim et al., 2009; Wang et al., 2012, 2009; Wilson et al., 2014)

In conclusion, the screens identified known tumor suppressors as well as novel candidates. Furthermore, I found that anchorage independent growth, which is an indicator of cell transformation, is largely driven by accelerated cell growth. The fact that the same candidates in the no-selection growth screen were most enriched in the transformation screen indicates that fast growth itself can be sufficient to transform cells. How exactly the novel candidates promote cell growth, is still under investigation. Several follow-up experiments need to be conducted in order to understand the mechanisms of actions of genes and pathways and how their deregulation leads to uncontrolled growth.

Eventually the goal of this project is to transfer this knowledge from basic research to medical applications. By identifying novel tumor suppressors and their mechanisms of action, we can shed light into novel mechanisms of cancer initiation and progression.

Chapter 4

Final Conclusions

In this thesis I investigated mechanisms of cancer progression using CRISPR/Cas9-based approaches. I used targeted approaches to decipher the mechanism of oncogene activation through enhancer hijacking and a systematic unbiased approach to identify potential tumor suppressors.

In the first project (Chapter 2) I sought to investigate a mechanism involving structural changes in the genome relocating enhancers to nearby genes and leading to their increase in gene expression, described as enhancer hijacking (EHJ). From analysis of cancer genomes I was involved in (Weischenfeldt, Dubash, **Drainas** et al., in revision), we identified *IGF2* and *IRS4* to be implicated in EHJ and further investigated their mechanism of activation. To this end, I recreated rearrangements associated with *IGF2* and *IRS4* overexpression using CRISPR/Cas9-based technology in relevant cell lines.

We screened over 3000 clones and obtained 37 clones verified by PCR. The strategy used was similar to recent published work (Kraft et al., 2015; Li et al., 2015) with some differences that allowed us to screen clones more efficiently. From our clones that harbored the rearrangement according to PCR and Sanger sequencing validation, three cases were further verified also by whole genome sequencing. Although we successfully engineered the rearrangements, overexpression of the genes was not observed. This could be dependent on the cell context and chromosomal epigenetic state or in addition indicate a more complex mechanism of activation. Further experimental work in *in vivo*, for example in mouse, may overcome these limitations and the applied modification can target the correct cell type and chromosomal state.

Another aim I achieved in this chapter was to address the oncogenic role of *IRS4*. *IRS4* was identified as the major EHJ candidate from the analysis across several cancers, especially in lung squamous carcinoma. As *IRS4* has not been described as an oncogene, we showed the growth promoting role *IRS4* in tumor xenografts in mice, indicating its oncogenic properties. Hence, this result further supports *IRS4* as a potentially novel oncogene. This work is part of a manuscript we recently submitted (Weischenfeldt, Dubash, **Drainas** et al., in revision).

Apart from oncogene activation, I investigated tumor suppressors using genome wide screens (Chapter 3). The aim was to conduct CRISPR/Cas9-based screens and identify novel tumor suppressors. I used two approaches, with anchorage independent growth as a selection barrier (transformation screen), and a growth assay without any selection (no-selection growth screen) and used epithelial cells in these screens. Many of the enriched genes in the transformation screen were known tumor suppressors indicating the reliability of the screen. Surprisingly, similar genes were also the highly enriched in the growth screen. This result indicates that promoting growth is a major contributor for overcoming anchorage dependency. Furthermore it supports that the potential novel candidates may indeed play tumor promoting roles since they were identified by two independent assays along with known tumor suppressors.

Another aim in this chapter was to identify potential genetic interactions with *TP53*. From the candidates identified, a previously known genetic interaction between *TP53* and mTOR pathway was confirmed. I also identified two novel tumor suppressors; *AHR* that was enriched in all screens and *FRYL* that was enriched in the $TP53^{-/-}$ cells, indicating a possible genetic interaction. Moreover, the transformation ability of these candidates were verified, also in different cell lines, showing that the genes may promote anchorage independent growth across different cell lines. *AHR* and *FRYL* were also found to be mutated in cancer genomes and together suggesting these genes as potential tumor suppressors.

During my PhD I was also involved in the development of a tool to analyze the screen data. We developed ScrispR, which uses information of random and fixed effects from the assay, an approach that has not been used before to our knowledge, for the analysis of such screens. A future outlook is the comparison of this tool with other recently developed

tools, which are mostly based on indirect methods to account for gRNA variability (Li et al., 2014b; Winter et al., 2015).

In conclusion, during my PhD work, I generated targeted rearrangements in the genome and revealed that the mechanisms of EHJ are more complicated than initially anticipated and likely context specific. Furthermore, I contributed to the discovery of *IRS4* as a potential oncogene and I identified potential tumor suppressors from CRISPR/Cas9-based genome wide screens. These results will increase our understanding of tumor formation mechanisms.

Chapter 5

Materials and Methods

5.1 Experimental Procedures Related to All Chapters

5.1.1 Cell lines

The cell lines used for all projects and relevant information are indicated in Table 5.1. All different medias were supplemented with Antiotic-Antimycotic (Thermo Fischer Scientific). The media for the MCF10A cell lines contained additionally 5% cholera toxin, 1 ng/mL human insulin, 10 μg/mL epidermal growth factor, 10 ng/mL hydrocortisone. All cell lines were regularly checked for mycoplasma contamination.

5.1.2 CRISPR design

CRISPRs were designed according to the webtool (http://crispr.mit.edu/) and were further selected based on their sequence adjacent to the PAM region (Doench et al., 2014). The least favored sequence is C-**TGG**-G with in any case eliminating any PAMs that follow three guanines. These restrictions were based according to the cutting efficiency (Doench et al., 2014). Selected CRISPRs (or gRNAs) were cloned and their cutting efficiency was tested with the surveyor assay (Section 5.3.8,Integrated DNA Technologies).

TABLE 5.1: Cell lines used in study. FBS: Fetal Bovine Serum, HS: Horse Serum. ND: not determined.

Cell line	Puromycin	Media	Distributor
RPE-1 WT	5 μg/ml	DMEM/F12-FBS	ATCC
RPE-1 $^{TP53-/-}$	10 μg/ml	DMEM/F12-FBS	(Mardin et al., 2015)
RPE-1 $^{TP53-/-,hyperploid}$	10 μg/ml	DMEM/F12-FBS	(Mardin et al., 2015)
MCF10A WT	0.5 μg/ml	DMEM/F12-HS+	ATCC
MCF10A $^{TP53-/-}$	0.5 μg/ml	DMEM/F12-HS+	This study
HEK293FT	ND	DMEM-FBS	Thermo Fischer
HCC15	ND	RPMI-FBS	DSMZ
H520	ND	RPMI-FBS	ATCC
H2170	ND	RPMI-FBS	ATCC

5.1.3 CRISPR cloning

The CRISPR sequences were cloned according to a modified protocol from Shalem et al. (2014) to the vectors of choice (see Table 5.2). Each CRISPR pair was annealed by adding 1 μl of each CRISPR primer, 1 μl 10X T4 Ligation Buffer (NEB) up to 10 μl H_20. Then the CRISPRs were placed in the thermocycler and incubated for 30 minutes at 37 ^oC, heated to 100 ^oC and ramped down by 1 ^oC/sec to room temperature. In parallel, the vector was digested with the appropriate restriction enzyme to allow annealing with the gRNAs. The annealed guides are then diluted 1:100. The ligation reaction is set up according to Table 5.3 and incubated for ten minutes. Then we transform Stbl3 chemically competent bacteria (Thermo-Fischer Scientific). The bacteria are thawed on ice and 2.5 μl from the reaction are used with 25 μl of the bacteria. The mixture is incubated on ice for 30 minutes and then a heat-shock for 60 seconds is applied at 42 ^oC. After the heat-shock the bacteria is placed back on ice for five minutes. 250 μl of SOC media is then added and the cells are incubated at 37 ^oC for one hour. The bacteria are then pelleted at 4000rpm for three minutes and are plated on agar plates with a selection marker (Ampicillin for lentiCRISPRv2 and pX330 vectors). The next day individual clones are picked and cultured in 5 ml LB liquid cultures. The day after, 0.5 ml is mixed with 60% glycerol to prepare a bacteria stock. The remaining sample is

pelleted and plasmid DNA is isolated (Qiagen). DNA pellets are then sent for Sanger sequencing with the appropriate primers to verify the cloning of the guide sequence.

Table 5.2: Vectors used in study

Vector	**Distributor or Specification**
pX330	Addgene ref: 42230
pX330-P2G	pX330 with gRNA insertion sites
pLenti-IRS4-Myc-DDK	Origene, ref: RC218385L1
pIRES2-AcGFP1	Takara-Clonetech
pLenti-IRS4-Myc-DDK-IRES-GFP	pLenti-IRS4-Myc-DDK with IRES-GFP from pIRES2-AcGFP1
pLenti-IRES-GFP	Control without IRS4
pDup	Construct with homologous arms for IGF2 duplication
pDelA	Construct with homologous arms for IGF2 deletion A
pDelB	Construct with homologous arms for IGF2 deletion B
pDelC	Construct with homologous arms for IGF2 deletion C
pDelD	Construct with homologous arms for IGF2 deletion D
pDelF	Construct with homologous arms for IGF2 deletion F
psPAX2	Addgene, ref: 12260
pMD2.G	Addgene, ref: 12259
lentiCRISPRv2	Addgene, ref: 52961
lentiCRISPRv2-EGFP	lentiCRISPRv2 with GFP as a selection marker

Table 5.3: Ligation Reaction

X μl	Digested plasmid (50 ng)
1 μl	Annealed gRNAs
5 μl	2X Quick Ligase Buffer (NEB)
X μl	ddH_2O
10 μl	subtotal
1 μl	Quick Ligase (NEB M2200S) total
11 μl	total

5.1.4 Virus production and infection

HEK 293FT cells (Table 5.1) were cultured to 80% confluence in 6-well plates. 1 μg of vectors psPAX2, pMD2.G and the respective lentiviral vector (see Table 5.2) were mixed

1:1:1 in OptiMem (Thermo Fisher Scientific) and transfected with Lipofectamine 3000 to the 293FT cells according to the manufactures protocol.

Virus was added directly to the media of the cells with the addition of 8 μg/mL polybrene (Sigma-Aldrich) (Davis et al., 2002). The cells were then centrifuged at 2000rpm for hours hours (spinfection). After centrifugation the media was replaced with fresh media. According to the type of selection the virus vector carried, the cells were either treated the next day with puromycin, or were sorted according to their GFP intensity three to four days later.

5.1.5 Immunoblotting

Cells were grown in 6-well plates until confluence. The cells were then trypsionized, washed twice with PBS and could be stored at -80 °C with snap freezing. For every one million of cells (roughly one confluent 6-well) 50 μl of RIPA buffer was added (10 mM Tris-Cl (pH 8.0) 1 mM EDTA. 1% Triton X-100. 0.1% sodium deoxycholate. 0.1% SDS, 140 mM NaCl, 1 mM PMSF). Additionally a protease inhibitor cocktail was added (Thermo-Fischer Scientific). For whole lysate, the sample was sonicated for 15 minutes with one minute break intervals at 4 °C (protein extracts were kept always at 4 °C). Protein concentration was measured using Bradford assay (Bio-Rad) as described below. A 96-well plate was prepared with BSA concentrations ranging from 5 μg/μl to 40 μg/μl in triplicate (1 μl of RIPA buffer was added to each solution and then 200 of 1x Bradford solution was added to each well). 1 μl of protein solution in triplicate was added and then 200 μl of the Bradford solution was added. The plate was incubated for five minutes at RT and then was measured at 595nm. The BSA slope was used as a reference to estimate the protein amount of the samples. For each sample 4x sample buffer (Bio-Rad) was added and the samples were then boiled for five minutes at 99 °C. Then the samples were stored at -20 °C. Approximately 20 μg of protein were added to a 4–20% gradient gel (Bio-Rad) and ran at 180V for one hour. The gel was then transferred to a nitrocellulose membrane by a Bio-Rad transfer system and ran according to the manufacturer's pre-settings. After the transfer the membrane was incubated for 30 minutes in 10% low fat milk in TBS-T (50 mM Tris pH 7.5, 150 mM NaCL, 0,1% Tween). Then the membrane was washed three times with TBS-T. Antibodies were added to 5% milk in TBS-T and the membrane was then incubated overnight at 4 °C. The next day the membrane was washed three

times with TBS-T. The secondary antibody was added to 5% milk in TBS-T and the membrane was incubated for one hour. The membrane was washed three times with TBS-T and horseradish-peroxidase substrate (Bio-Rad) was added to the membrane. After five minutes the membranes were photographed for chemiluminescence (Bio-Rad).

5.1.6 Long-range paired-end sequencing (MP-seq)

Long-range (or 'Mate-pair') DNA library preparation was carried out using the Nextera Mate Pair Sample Preparation Kit (Illumina). In brief, 4 μg of high molecular weight genomic DNA were fragmented by the Tagmentation reaction in 400 μl, followed by the strand displacement and AMPure XP (Agencourt) cleanup reaction. Samples were size selected to 4-6 kb with a gel step following the Gel-Plus path of the protocol. 300-550ng of size-selected DNA were circularized in 400 μl for 16 hours at 30 °C. The library was then constructed after an exonuclease digestion step to get rid of remaining linear DNA, fragmentation to 300-700 bp with a Covaris S2 instrument (LGC Genomics), binding to streptavidin beads and Illumina Truseq adapter ligation. The final library was obtained after PCR for one minute at 98 °C, followed by nice cycles of 30 seconds at 98 °C, 30 seconds at 60 °C, one minute at 72 °C and a final five minutes at 72 °C step. Deep sequencing was carried out with the Illumina HiSeq2000 (2x101bp) instrument using v3 chemistry to reach an average physical coverage of 20-30x. After sequencing, the reads were aligned to the hg19 assembly of the human reference genome. Library preparation performed by Dr. Adrian Stütz and Benjamin Reader. Text adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

5.1.7 Mass-low coverage whole genome sequencing (LC-WGS)

Libraries were prepared using NEB Ultra kit (New England Biolabs) according to manufacturers specifications. Libraries were then sequenced on a Hiseq 2000 (Illumina). Library preparation performed by Dr. Adrian Stütz and Benjamin Reader.

5.2 Experimental Procedures Related to Chapter 2

5.2.1 Flow cytometry

Transduced HCC-15 cells were sorted for GFP expression on a MoFloXDP cell sorter (Beckman Coulter Inc) equipped with a Coherent Innova 90C Argon ion laser (Coherent Inc.), tuned to 488 nm at 200 mW. Cells were sorted using a 100 μm Nozzle while running BD FACSFlow as sheath at 20 psi/RT. Forward and side scatter height and area signals were used for gating of live cells and singlets. GFP fluorescence was detected using a 530/40 nm bandpass filter combined with a 488 notch filter. GFP positive cells were sorted in purity mode (one drop envelope) into 6-well or 96-well dishes with culture media respectively. In order to measure GFP intensity HCC-15 cells were run through LSR-Fortessa SORP instrument (BD Biosciences) with a 488 nm laser (530/30 BP). All post acquisition analysis was done with FlowJo 10.0.8 (Tree Star, Inc). Text adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

5.2.2 *IRS4* vectors and virus preparation

HCC15 cell line was purchased from DSMZ and cultured in RPMI 1640 medium (Thermo Fisher Scientific) supplemented with 10% FBS (Thermo Fisher Scientific) and Antibiotic-Antimycotic (Thermo Fisher Scientific). An IRS4 overexpressing vector, pLenti-IRS4-Myc-DDK, was purchased from OriGene. An IRES-eGFP sequence was cloned from a pIRES2-AcGFP1 vector (Takara-Clonetech) into the pLenti-IRS4-Myc-DDK vector using In-Fusion HD cloning kit (Takara-Clonetech) and it is referred as pLenti-IRS4. We used the following primers for this purpose: F-GGCCGCGGTCTGTACActtcgaattctgca-gtcgacg; and R-GAATCCTACTTGTACAtcacttgtacagctcatccatgcc. The control vector was created by removing IRS4-Myc-DDK by restriction enzyme digest with EcoRI and it is referred as pLenti-empty. Plasmids used for lentivirus production were pMD2.G (VSV-G envelope) and psPAX2 (2nd generation lentiviral packaging plasmid); both gifts from Didier Trono - Addgene plasmids ref: 12259 and 12260. Lentivirus production was conducted by transfection with Lipofectamine 3000 Reagent (Thermo Fisher Scientific) of equal amounts of pMD2.G, psPAX2 and pLenti-IRS4-Myc-DDK-IRES-GFP/pLenti-IRES-GFP, in 293FT cells (Thermo Fisher Scientific) according to the manufactures protocol. Cells were transduced with produced virus with the addition of 8 μg/mL

polybrene (Sigma-Aldrich) by spinfection (centrifuge 2000 rpm for two hours) with the produced virus and were enriched by sorting according to eGFP intensity (see Flow cytometry methods). All cell lines were regularly checked for mycoplasma contamination. Text adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

5.2.3 Mouse injections

One million transduced HCC15 cells were suspended in DMEM mixed 1:1, v/v with Matrigel (BD Biosciences) and subcutaneously implanted into both flanks of nude mice (Charles River Laboratories, NMRI-Foxn1^{nu} /Foxn1^{nu} (homozygous) male mice; eight weeks old at time of injection). The total number of tumors were N=8 for each cell line (i.e. using four mice for each line, whereby we performed experiments in both flanks in each mouse). While at this sample size effect sizes are not robustly estimated, differences in tumor growth became readily evident. Mice were randomly assigned into two groups and tumor sizes were measured twice weekly in two dimensions (length and width). Tumor volumes (V) were calculated as: V (cm^3) = 0.5 X (length X width^2). Mice were euthanized once the biggest tumor volume was 2 cm^3. Mice were housed and maintained according to animal use guidelines at EMBL Heidelberg. Both mouse grouping as well as tumor volume measurements were blinded. Text adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

5.2.4 qPCRs

Single-stranded cDNA was synthesized from 500 ng of total RNA using the SuperScript III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen) according to the manufacturers' protocol.

The qPCR primers were designed using the online Primer3 Plus program (Untergasser et al., 2007) with the qPCR settings activated. Primer sequences are IRS4_F: CCCACACATGAGCAGAGAGA, IRS4_R: CTGACTGTCTGGGTTCAGCA, Globulin_F: TACATGTCTCGATCCCACTTAACTAT, Globulin_R: AGCGTACTCCAAAGATTCAGGTT, IGF2_F: TGGCATCGTTGAGGAGTGCTGT and IGF2_R: ACGGGGTATCTGGGGAAGTTGT. All primers were tested by running a standard curve and requiring the primer efficiency to be between 90-100% and as close as possible to that of the house

keeping primer pair. The primer efficiency for globulin was 91.3%, 91.6% for *IGF2* and 95.6% for *IRS4*. In addition, a single and discrete peak was detected in the melt curve analysis for all primers tested. The qPCR experiments were performed on a StepOnePlus 96 Fast machine (Applied Biosystems) in 20 μl using a 96-well plate. The mastermix contained 10 μl 2xSYBR Green PCR Master Mix (Applied Biosystems), 0.4 μl of each primer (10 μM), 2.5-5 ng of sample cDNA in 5 μl, and 4.2 μl nuclease free H_2O. The reaction program was run in default ramping speed mode and cycling conditions were 10 min at 95 ^oC, 40 cycles of 95 ^oC for 15 seconds and 60 ^oC for one minute, followed by a melting curve stage. Non-template controls were included in all experiments, replacing cDNA with H_2O, and typically resulted in no detection at all. The results were analyzed using the StepOne analysis software v2.3 (Applied Biosystems). Relative expression levels for *IGF2* and *IRS4* were calculated after normalization to the house keeping gene globulin using the $\Delta\Delta$-Ct method in Microsoft Excel. Experiment performed by Dr. Adrian Stütz. Text adapted from Weischenfeldt, Dubash, **Drainas** et al. (in revision).

5.3 Experimental Procedures Related to Chapter 3

5.3.1 CRISPR soft-agar screen protocol

In order to perform the transformation screen based on CRISPR/Cas9 and soft agar the following steps take place. Firstly, the virus needs to be prepared in large amounts to be used in the screen. Next, the multiplicity of infection (MOI) needs to be calculated for optimal virus use. Afterwards the infection of the screen takes place with the estimated MOI. Then the clones that were selected for virus uptake are subjected to soft agar. Cells collected before and after agar selection have their DNA extracted and then libraries containing the gRNA information of each population are prepared and sequenced. The next section describe these steps in detail.

5.3.1.1 Virus preparation

The GeCKOv2 library was purchased from Addgene and was amplified according to Shalem *et al.* with small modifications (Shalem et al., 2014). For virus production two

5-layer flasks (Corning) were seeded with 25 million 293FT cells each in DMEM complete media (+10% FBS, NEAA, pyruvate, glutamine, Antibiotic-antimycotic - (Thermo Fischer Scientific)). The next day, the media was removed and 110 ml of fresh DMEM complete media was added. For each flask, 500 μl of P3000 reagent (Lipofectamine 3000 kit – Thermo Fischer Scientific) was diluted in 20 ml OptiMEM (Thermo Fischer Scientific) with 100 μg of GeCKOv2 library (Addgene - ref:1000000048), 50 μg of pMD2.G (Addgene – ref: 12259), and 75 μg of psPAX2 (Addgene - ref: 12260). 500 μl of Lipofectamine 3000 (Thermo Fischer Scientific) was added to 20 ml of OptiMEM and after five minutes was mixed with the DNA-P3000 solution. The complete mixture was incubated for ten minutes at room temperature and then was added to each flask. After 60 hours, the media was collected and centrifuged at 3,000 rpm at 4 ^{o}C for ten minutes to pellet cell debris. The supernatant was then filtered through a 0.45 μm low protein binding membrane (Millipore Steriflip HV/PVDF). To achieve 100-300X concentration of the pooled library, the virus was ultracentrifuged 24,000 rpm for two hours at 4 ^{o}C (Beckman Coulter). Media was removed and the pellet was resuspended overnight at 4 ^{o}C with agitation in 8 ml of DMEM supplemented with 1% BSA. Stocks of 50 μl were prepared, for multiplicity of infection testing and 500 μl aliquots for use. The aliquots were stored at –80 ^{o}C.

5.3.1.2 Multiplicity of infection (MOI)

To achieve this one million cells with 3 ml of media were plated in each well of a 6-well plate 16 hours before transduction. Then, the media was aspirated from each well and 2 ml of media supplemented with polybrene (8 μg/μl) was added. To each well 2 μl, 7 μl, 15 μl, 25 μl of virus was added. The remaining wells served as positive and negative controls. The plate was then centrifuged for two hours at 37 ^{o}C at 2000 rpm. After centrifugation, the virus-containing media was replaced with fresh media and the cells were allowed to recover for one day. The next day the cells from each well were split into 15 cm dishes with media supplemented with puromycin (Table 5.1). Puromycin was also added to the negative control but not to the positive control in order to check for cell viability. Then the cells were grown for three days and the number of cells was counted for each condition. The MOI was determined by dividing the amount of cells in each condition with the amount of the positive control cells (Fig. 5.1). A MOI of approximately 20% was achieved for each experiment. Over 30-40% MOI was not

observed, perhaps due to the toxicity of the virus to the cells, therefore a preferable range was close to 20%.

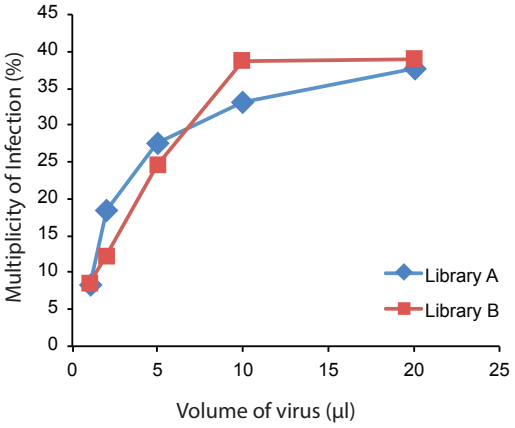


FIGURE 5.1: Example of multiplicity of infection (MOI) determination. $RPE^{TP53^{-/-},\ hyperploid}$ cells were infected with deferent amounts of the produced virus. After infection the cells were selected with puromycin. The percentage of cell survival in comparison to cells with no selection was calculated and considered as the MOI. A MOI of approximately 20% is selected for the virus screens.

5.3.1.3 Screen and infection

98 million cells were grown per library per replicate. RRE^{WT} and $RPE^{TP53^{-/-}}$ were aliquoted to one million cells per well into 6-well plates, as for $RPE^{TP53^{-/-},\ hyperploid}$ were 0.5 million cells per 6-well. The MCF10A and $MCF10A^{TP53^{-/-}}$ cells lines were plated as two million cells per 6-well. Lastly, two 6-well plates were used as positive (virus and no puromycin – for MOI estimation) and negatives controls (no virus, puromycin). Each well contained 2 ml of media. The next day, the media was removed and replaced with new media supplemented with polybrene and the appropriate virus amount (this can be achieved easily by making a master-mix and using a 15 ml pipet). Cells were spinfected for two hours at 37 °C at 2000 rpm. Then the media was replaced and the cells recovered for one day. The following day 12 wells were pulled together into one 5-layer flask with media with puromycin. Eight 5-layers were needed in total. Positive and negative controls were planed in 15 cm dishes. After three days, the cells were collected, filtered in a 70 μm cell strainer (Falcon) and counted to estimate the complete number of cells that survived. Up to 80 million cells were used for soft agarose selection. When less, two thirds, was used for agarose selection. The rest was frozen in order to have a representation of gRNAs before agar selection.

5.3.1.4 Agar selection

In order to prepare the agarose layer, 2x media (Thermo-Fischer Scientific) was prepared and filtered with a 0.22 μm filter (Millipore). Then it was mixed 1:1 (at 37 ^{o}C) with 0.7% agarose (autoclaved). 8 CellSTACK flasks (636 cm^2 growth area) with low attachment surface (Corning) were used for each screen. Approximately eight to ten million (through a strainer in order not to form clumps) cells per stack with 150 ml of final volume were used. The flasks were let in room temperature for 30 minutes to solidify. After the agarose solidified 50 ml of media was added. The cells were incubated then for at least five weeks. After incubation and colonies appeared in the soft agar, the cells were collected into a beaker and pipetted several times in order to break the agarose layer. 100 ml of media was added per flask and pipetted and then approximately 70 ml was added to 5x15 cm^2 dishes. Two day later the 5x15 cm^2 dishes were pipeted to 6x15 cm^2 dishes after the addition of 70 ml more media. Any floating colonies were collected and plated in an individual plate. Although there was a loss of cells that did not attach to the plate, an acceptable amount of cells attached to the plate. The cells were then incubated for five to seven days in order to allow them to attach. For collection the dishes were washed twice with PBS then trypsinized, pelleted and frozen at -80 ^{o}C.

5.3.1.5 PCR amplification of GeCKO libraries and sequencing

Genomic DNA was isolated with QIAamp DNA Blood Maxi Kit and quantified with Qubit. Then eight reactions of 0.5 μg DNA per reaction were prepared (Table 5.4) for a first PCR amplification step with the settings shown in Table 5.5. The samples were pooled together into two vials (four samples in each) and the PCR reaction was purified, eluted in 30 μl H_2O and quantified. Four reactions for the secondary PCR reaction were prepared with 6 μl of DNA per reaction (Table 5.4).

The primers used for the reactions are depicted in Table 5.6. After the second PCR all reactions were pooled together, purified using AMPure beads (Beckman Coulter) and eluted in 30 μl. All samples were then run in a 1% agarose gel for one hour. A band of approximately 390 bp was cut, purified and quantified with Qubit. All samples were pooled appropriately and submitted for sequencing.

TABLE 5.4: PCR mix used for GeCKO library amplification

Hercules mix (1x of 50 μl)	
Buffer	10 μl
dNTPs	0.5 μl
Primers	1.25 μl each
Enzyme	0.5 μl
H_2O	up to 50

TABLE 5.5: PCR run settings

Hercules PCR reactions		
cycle	temperature	time
1	98 °C	2:00
2	95 °C	0:20
3	62 °C ^{1st} (60^{o} ^{2nd})	0:20
4	72 °C	0:30
5	GoTo 2	17 times^{1st}, (7 times ^{2nd})
6	72 °C	3:00
7	12 °C	forever

5.3.2 Mapping raw representation of gRNAs and downstream analysis

Short reads were initially pre-processed to remove barcodes and afterwards aligned to the self-constructed guide reference database (n=123,411 contigs) using Subread, a short-read alignment program in R (Liao et al., 2013). The total number of aligned reads per guide sequence was then used as the quantification measure. ScrispR takes raw read counts per guide as input and generates initial quality control plots to evaluate the reproducibility of the screen. Raw read counts are adjusted for sequencing depth by scaling guide counts by the total number of sequenced reads that mapped onto the full guide sequence database. Differences between conditions, in this case before and after transformation, for each guide are evaluated using a mixed effect model (MEM) that is implemented in the lme4 R package (Bates et al., 2015) (see section 3.2.2). Note, log-transformation is applied to satisfy the normality assumption of residuals. Finally, normal distributions are used to approximate nominal model P-values for each gene. The False Discovery Rate (FDR) procedure by Benjamini-Hochberg is used to account

TABLE 5.6: Primers for 1st & 2nd PCR reactions

Primers	
Lenti_AF	atggactatcatatgcttaccgtaacttg
Lenti_AR	cgactactgcacttatatacggttctc
Universal	caagcagaagacggcatacgagatcggtctcggcattcctgctgaaccgctcttccga
	tctcaacttctcggggactgtgggcga
GeCKO_1	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct
	-atgcga-tcttgtggaaaggacgaaacaccg
GeCKO_2	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct
	-cataa-tcttgtggaaaggacgaaacaccg
GeCKO_3	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct
	-gccg-tcttgtggaaaggacgaaacaccg
GeCKO_4	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct
	-tga-tcttgtggaaaggacgaaacaccg

for the multiple testing. Genes are scored both by its fold change and FDR corrected *P*-values. Genes can be visualized by volcano plots that depict fold change in the x-axis and *P*-values on the y-axis.

5.3.3 Candidate verification assay

For each candidate selected from the screen, three gRNAs were selected for verification. Two of gRNAs were selected showing the highest efficiency in the screen and an additional one was designed. The gRNAs selected were cloned according to section 5.1.3 and virus was produced according to 5.1.4. RPE and MCF10A cell lines (Table 5.1) were plated in 96-well plates and were infected with virus when they reached a confluence of 80%. The cells then were spinfected for two hours at 37 °C at 2000 rpm. After spinfection the cells were incubated overnight and the next day were split in concentration 3/5 to new 96well plates. Puromycin selection was conducted according to the concentration depicted in Table 5.1. After three to five days of selection (according to the negative control), cells were split in half to low attachment plates (Corning), mixed 1:1 (total volume of 150 μl) with agarose (0.7%). The remaining half are put back into a normal 96-well plate and the next day were treated with MTT (see section 5.3.4). The MTT assay estimated the

amount of cells that were going to be embedded in soft agar. The agarose plate was then solidified in room temperature for 30 minutes. The solidified plates were then incubated at 37 °C. The next day a feeder layer of 100 μl of media was added to the wells. The plates were wrapped with airtight plastic wrap and incubated for one month at 37 °C.

After one month of culture, each well was collected separately to 6-well plates. To allow better solubilization of the agar, 10 μl solubilization solution (Cell Biolabs) was added to each well prior collection. The colonies were incubated for one to two weeks - until colonies visually appeared in the positive control wells (*i.e.* transformed cell lines). The 6-well plates were washed twice with PBS and crystal violet of concentration of 0.05% was added for ten minutes in each well. Crystal violet is a violet color chemical that binds proteins and DNA (Franken et al., 2006). The colonies were colored and each plate was photographed (S&P Robotics). The data was analyzed by Fiji (Schindelin et al., 2012) by a code kindly provided by Hernando Martinez Vergara. In brief the code converts each image to 8-bit and then sets a threshold in order to capture the colonies (set manually). Then it measures the amount of black and white pixels that correspond to growth and empty space. This is then output to a csv file, which is then further processed using "R". The data is normalized to the highest positive control values (in these experiments, according to *NF2*) in order to compare growth across experiments. This is then plotted with barplots with error bars of the standard deviation of the mean (Fig. 3.9).

5.3.4 MTT assay

MTT solution (Sigma) was prepared to 5mg/ml (10x stock) in PBS and passed through a 0.2 μm filter. MTT is stored in -20 °C. Cells are cultured in 96-well plates and 10 μ of MTT solution is added to each well. The cells are then incubated for three hours in 37 °C. Then, all media (with MTT) is aspirated and cells are then left to dry for one hour. Afterwards, the cells are dissolved in 100 μl isopropanol and the optical density is measured at 570 nm (and reference over 650 nm).

5.3.5 Creating *TP53*$^{-/-}$ cell lines

For each gene, gRNAs targeting within the first exons were designed according to section 5.1.2. Two gRNAs targeting *TP53* were then cloned in vector px330-P2G (Table 5.2). Cells in 10 cm dishes were transfected with Lipofectamine 3000 (Thermo-Fischer scientific) according to the manufacturers protocol. The next day the cells were split into a 15 cm dish and puromycin was added for selection. After two days the cells were single cell sorted on 96-well plates. Cells clones were then subsequently grown to 6-well plates. In order to test p53 protein expression, cells were treated one hour with doxorubicin 1.5 μM in order to damage the DNA and induced a DNA damage response. Finally, protein expression was tested by immunoblotting (Fig. 5.2).

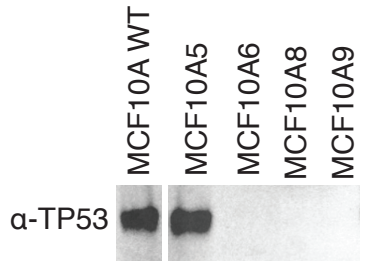


FIGURE 5.2: *TP53* knock outs in MCF10A cell lines. *TP53* in clones 6, 8 and 9 was successfully disrupted wereas the control and clone 5 exhibit TP53 protein expression.

5.3.6 Immunofluorescence

Cells are grown on glass slides (cover slips) until 95% confluence. Then the cells are fixed either with formaldehyde 4% in PBS for ten minutes in room temperature or with methanol for five minutes in -20 °C. After fixation the cover slips were washed once with PBS and stored in 4 °C. The fixation method depends on the antibody activity, therefore for initial tests both fixing methods were used. Afterwards, the cells were treated with 0.1% Triton-X 100 for ten minutes for membrane permeabilization. The cells were then washed once with PBS, incubated with 10% FBS (Fetal Bovine Serum) for 30 minutes and then washed with PBS again. Antibody was diluted in 3% BSA (Bovine serum albumin) in the suggested concentration. Then 10-15 μl (according to cover slip size) of antibody was added to each cover slip (added on parafilm and at the cover slip upside down on the solution). The antibody was incubated on the cells for one hour in a humidified chamber. The cover slips were then washed three times with PBS

and the secondary antibody (an antibody conjugated to a fluorophore and targets the species of the primary antibody) was then added, together with 1:10000 HOECHST dye (DNA binding dye) for 30 minutes. After incubation the cover slips were washed three times with PBS and the slides were fixed on glass slides with ProLong (Thermo Fischer Scientific). The cells could then be visualized on a fluorescence microscope.

5.3.7 Sanger sequencing verification

Sanger sequencing was performed by GATC biotec (https://www.gatc-biotech.com/). For verification of plasmids, DNA was extracted from bacteria (Qiagen) and appropriate primers were sent to GATC biotec. The data was analyzed using the Snapgene (http://www.snapgene.com/) software. For verification of PCR products, the PCR reaction was ran initially in a eletrophoresis gel. Then the amplified pieces were extracted from the gel and purified (Qiagen). The purified PCR product with the appropriate primers was sent to GATC biotec and analyzed by Snapgene.

5.3.8 Surveyor assay

Cells were transfected or infected with the gRNA of interest. After three days the cells were collected and DNA was extracted. The assay was performed according to the manufacturer's specifications (Integrated DNA Technologies). In brief the assay works as follows. First a reaction PCR with primers designed adjacent to the targeted site is performed. Next, the amplified piece is dissociated and reannealed by heating and cooling. The heteroduplex/homoduplex mixer is then treated with Surveyor nuclease. Lastly, an electrophoresis is performed. Acquisition of extra bands that add to the expected size of the PCR amplification and thier relative intensity indicates the functionality of the gRNAs (Fig. 5.3).

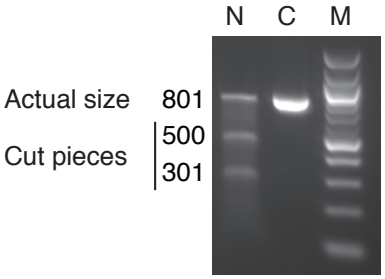


FIGURE 5.3: Surveyor assay. Surveyor nucleas indroduces a double strand break in the heteroduplexes but not in homoduplexes. A digestion of a piece is observed in the treated sample (N) in comparison to the control (C). M=1000 kb marker.

5.3.9 Identification of knockout clones via Sanger sequencing

In order to detect knockouts I developed a strategy exploiting Sanger sequencing information of double bases. Initially primers are designed adjacent to the targeted site of the gRNA. The piece is amplified by PCR and the reaction is then loaded in an electrophoresis gel. If there are larger deletions, two bands will appear in the gel. Purifying those bands will give the sequence information of each allele. In most of the cases, there is only a single band in the gel. The band is purified and sent for Sanger sequencing (Materials and Methods 5.3.7). From the data generated, firstly the chromatogram is read and the region where double bases appear are identified. Both bases are then documented. Then, sequenced from the reverse primer, the common region is noted. Next, the common allele is then identified in the double bases from the forward sequence. The sequences are disentangled and both alleles can be solved according to this information. In the depicted figure a deletion and an insertion are shown (Fig. 5.4).

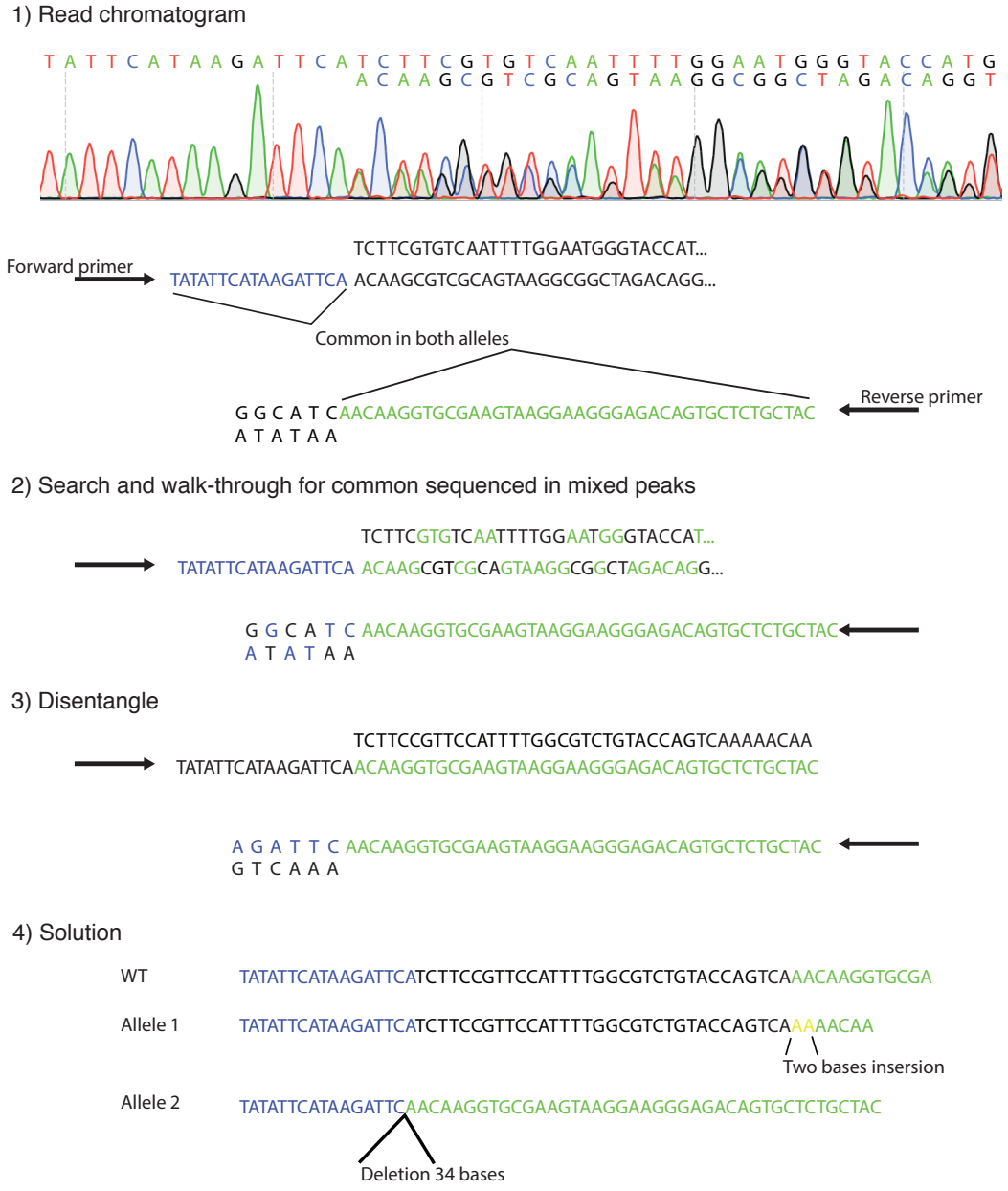


FIGURE 5.4: Strategy to identify knockouts using Sanger sequencing information. 1) The chromatogram from the Sanger sequenced data is read. The region with the double read bases is identified and the common sequence of both alleles is noted. 2) The common allele is then identified from the reverse sequenced and then it is checked in the double bases of the forward sequence. 3) The information is disentangled according to the identified sequence. 4) The two alleles are solved according to the generated information. In this case one allele harbors a deletion of 34 bases and the other allele harbors an insertion of two bases. Both alleles generate a frame shift in the sequence, which results into a knockout.

Appendix A

Appendix Related to Chapter 2

A.1 Additional Figures

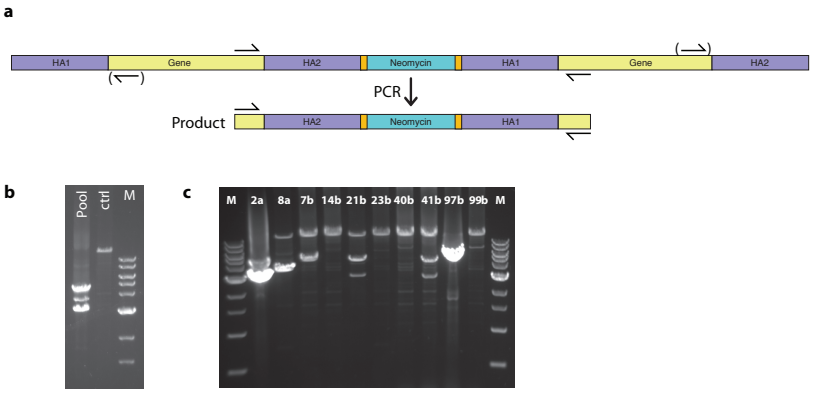


FIGURE A.1: Example of PCR validation in recreating the duplication. (**a**) PCR strategy to capture the duplication event. (**b**) PCR on the pool of cells, which were transfected with the construct versus control. Appearance of three bands are observed in the pool. The top band represents the expected size of the duplication event with the insertion of the selection marker as depicted in panel "a". The second band was sequenced to carry the duplication with the selection marker but with an aditional deletion. The third band represents the duplication event without the selection marker. (**c**) Individual clones were screened to identify duplication positive clones.

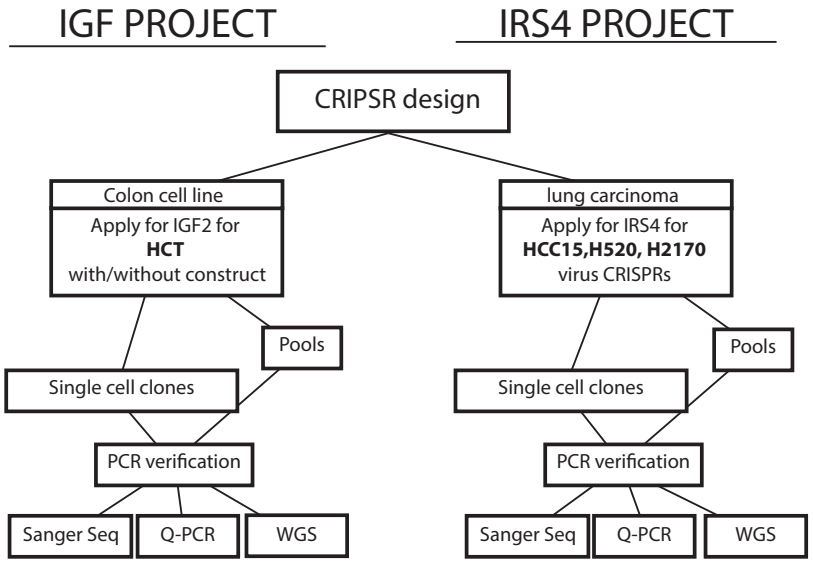


FIGURE A.2: Screening plan for recapitulating rearrangements for *IGF2* and *IRS4*

Appendix B

Appendix Related to Chapter 3

B.1 Additional Tables

TABLE B.1: Clones sequenced for the identification of SCNAs

Gene	RPE WT	RPE$^{TP53^{-/-}}$	RPE$^{TP53^{-/-}, \text{ hyperploid}}$
NF2	4	4	4
PTPN14	4	4	4
TSC1	NA	NA	2
PTEN	NA	NA	4
AHR	4	4	3
FRYL	3	4	5
SERPINB2	NA	NA	2
C6orf222	NA	NA	4
RNF7	NA	4	4
TMEM184A	NA	NA	2

TABLE B.2: Gene ontology of most significant underrepresented genes. Underrepresented genes are significantly enriched for essential genes. Source Toppgene.

	P-value	FDR BH	Genes from Input	Genes in Annotation
translational termination	1.35E-07	1.93E-04	14	95
ribosome biogenesis	1.67E-07	1.93E-04	19	178
RNA processing	2.35E-07	2.06E-04	43	718
mRNA metabolic process	3.54E-07	2.59E-04	40	653
cell cycle	5.31E-05	5.28E-03	67	1611
protein transport	5.42E-05	5.28E-03	62	1457
cell division	5.56E-05	5.30E-03	39	779
translational initiation	6.08E-05	5.67E-03	15	179
mitotic cell cycle	6.48E-05	5.92E-03	44	927
protein folding	9.70E-05	8.68E-03	17	230

TABLE B.3: Pathways enriched from the no-selection growth screen top candidates. Source: Toppgene

Pathway			
Name	P-adjusted	Input genes	Annotated
Hippo signaling pathway	1.6E-04	9	154
Developmental Biology	2.2E-03	11	419
neddylation	2.4E-03	3	11
CRMPs in Sema3A signaling	5.4E-03	3	16
the XRE-AhR mediated of drug-metabolizing enzyme expression	5.4E-03	2	3
Inhibition of TSC complex formation by PKB	5.4E-03	2	3
Cell-Cell communication	5.4E-03	6	131
Regulation of Microtubule Cytoskeleton	5.4E-03	4	44
Ubiquitin mediated proteolysis	5.8E-03	6	137
Energy dependent regulation of mTOR by LKB1-AMPK	5.8E-03	3	18
Axon guidance mediated by semaphorins	6.2E-03	3	19
Regulation of Actin Cytoskeleton	6.9E-03	6	147
p53 pathway by glucose deprivation	7.4E-03	3	21
Regulation of actin cytoskeleton	7.4E-03	7	215
Integrated Breast Cancer Pathway	7.7E-03	6	154
Vibrio cholerae infection	7.7E-03	4	54
mTOR Signaling Pathway	8.6E-03	3	23
Signaling events mediated by VEGFR1 and VEGFR2	1.2E-02	4	63
Genes related to PIP3 signaling in cardiac myocytes	1.5E-02	4	67
PKB-mediated events	1.5E-02	3	29
AMPK signaling	1.5E-02	4	68
Adherens junction	1.8E-02	4	73
CXCR4-mediated signaling events	2.0E-02	4	76
Insulin/IGF pathway-protein kinase B signaling cascade	2.0E-02	3	34
Direct p53 effectors	2.0E-02	5	134
CHL1 interactions	2.2E-02	2	9

TABLE B.4: gRNAs used in study. The clone name indicates the gene it targets. The gRNAs were obtained from the GeCKO library except for clones with "CS" (custom designed).

Gene and clone name	gRNA sequence	Gene and clone name	gRNA sequence
PTPN14_4	CTAGCCGGCCTAGCTGTGCA	**C6orf222_1**	GCCGATGGCCGCTGACACAC
PTPN14_5	GAAATAGCACATACTCTCTG	**C6orf222_2**	CTCCTCGGAGCCGTGTTTCT
PTPN14_6	AGAGTATGTGCTATTTCCTA	**C6orf222_5**	CAGTACCATCCACTTTGATG
NF2_1	CCTGGCTTCTTACGCCGTCC	**RNF7_1**	GCGATACGTGCGCCATCTGC
NF2_5	ATTCCACGGGAAGGAGATCT	**RNF7_2**	CCACCGCGTTCCACTTCTTG
NF2_6	TGAGCCTACCTTGGCCTGGA	**RNF7_3**	TTCAGCTTGACATCTAAGAC
TSC1_4	TTTATCCATCCTCTCGTTAC	**PTPN14_CS**	GACCAGGTGATTCGGCTAGC
TSC1_5	ACCTTCGAGGGTCCAGTTCA	**NF2_CS**	TCGGATTTCATTCCACGGGA
TSC1_6	ATTCGTTAATCCTGTCCAAG	**TSC1_CS**	TTATCCATCCTCTCGTTACT
TSC2_4	GTGGCCTCAACAATCGCATC	**TSC2_CS**	GGCGGCATGACGCCTTTCCG
TSC2_5	CCAACGAAGACCTTCACGAA	**PTEN_CS**	TATCCAAACATTATTGCTAT
TSC2_6	AGCACGCAGTGGAAGCACTC	**FRYL_CS**	TTAGACTTTGTGCTAGACCG
PTEN_2	CCTACCTCTGCAATTAAATT	**AHR_CS**	GCCTCCGTTTCTTTCAGTAG
PTEN_4	ACAGATTGTATATCTTGTAA	**SERPINB2_CS**	ATGGAGCATCTCGTCCACCA
PTEN_5	ACGCCTTCAAGTCTTTCTGC	**TMEM184A_CS**	ATTTGGCAAATACCACGACG
FRYL_2	TACTTCGCACCTTGTTTGAC	**THRAP3_CS**	GGAGAAAAAGTCCTCTTCTA
FRYL_5	TATTGTCCAGATCTCTTCAG	**C6orf222_CS**	CCCCAAGAAACACGGCTCCG
FRYL_6	CTCTTGATGACATATTCACC	**RNF7_CS**	CGATACGTGCGCCATCTGCA
AHR_2	TTGCTGCTCTACAGTTATCC	**PTPN14_CS**	GACCAGGTGATTCGGCTAGC
AHR_5	TCCGTTTCTTTCAGTAGGGG	**NF2_CS**	CCTGGACGGCGTAAGAAGCC
AHR_6	AGTTGTCACTACAGATGCTT	**TSC1_CS**	TTATCCATCCTCTCGTTACT
SERPINB2_1	GATACCTGCAAAATCGCATC	**TSC2_CS**	GGCGGCATGACGCCTTTCCG
SERPINB2_2	GGGCAGCACCGAAGACCAGA	**PTEN_CS**	TATCCAAACATTATTGCTAT
SERPINB2_5	TGAGAAGTCTGCGAGCTTCC	**FRYL_CS**	TTAGACTTTGTGCTAGACCG
TMEM184A_1	GTCTGTCCTAGTGTCCGCAG	**AHR_CS**	GCCTCCGTTTCTTTCAGTAG
TMEM184A_2	CTTACTTGAAGTCCCCGTCG	**SERPINB2_CS**	ATGGAGCATCTCGTCCACCA
TMEM184A_3	GCGCTCCTACACCGTGCCAC	**TMEM184A_CS**	ATTTGGCAAATACCACGACG
THRAP3_1	GATCTTGAACGGCCTCGACG	**THRAP3_CS**	GGAGAAAAAGTCCTCTTCTA
THRAP3_2	TGGCCGGCTATCCTTAGAAG	**C6orf222_CS**	CCCCAAGAAACACGGCTCCG
THRAP3_3	GACTGCTTATAAAGCAGTCC	**RNF7_CS**	CGATACGTGCGCCATCTGCA

B.2 Additional Figures

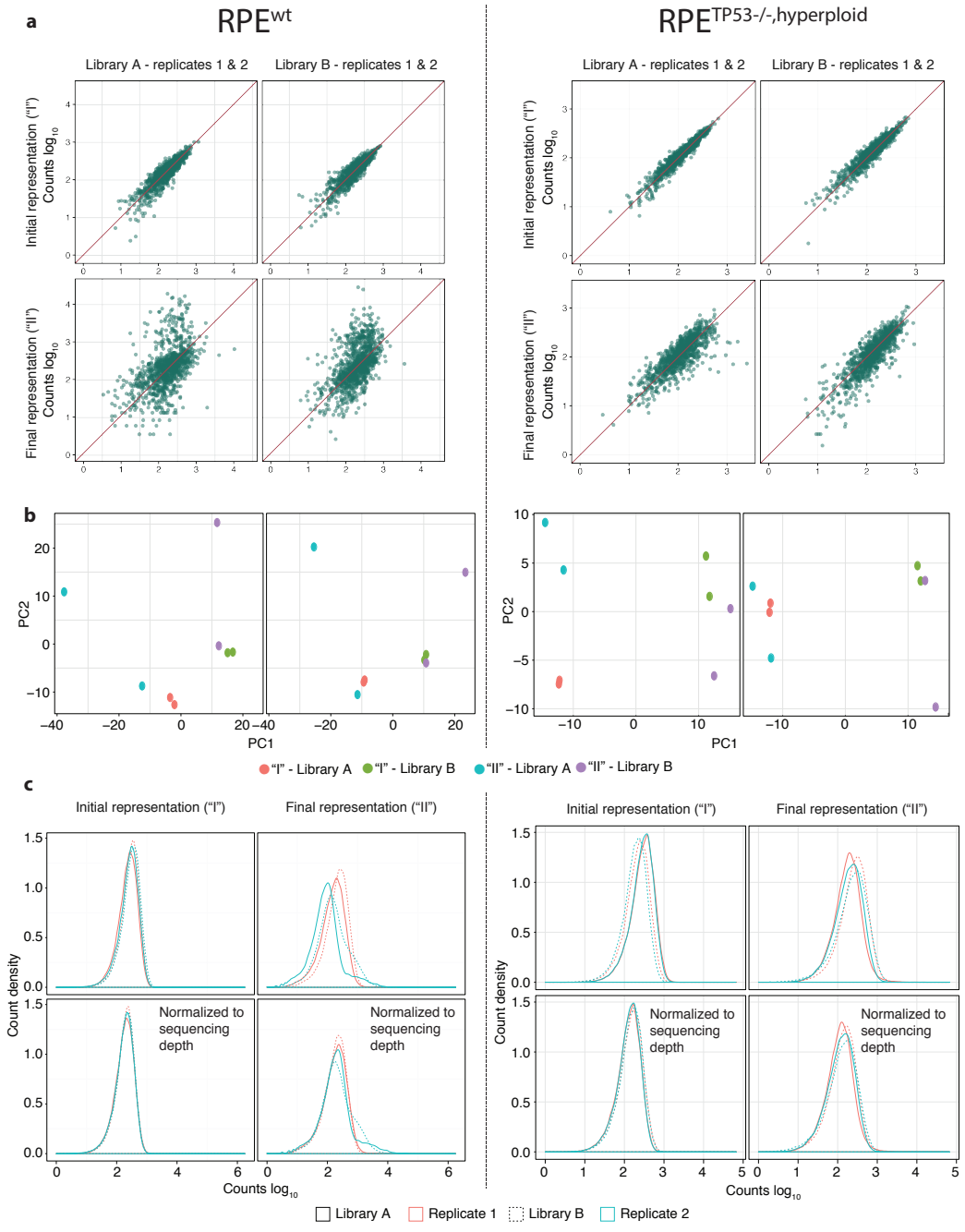


FIGURE B.1: Quality controls of the RPE^{WT} and $RPE^{TP53^{-/-},hyperploid}$ trasnformation screen. (**a**) The representation of each control guide in replicate was plotted against each other in order to evaluate the reproducibility of the screen. (**b**) Principle component analysis plots indicate that after normalization the replicates are similar. (**c**) Density plots before and after normalization. Replicates overlap after normalization.

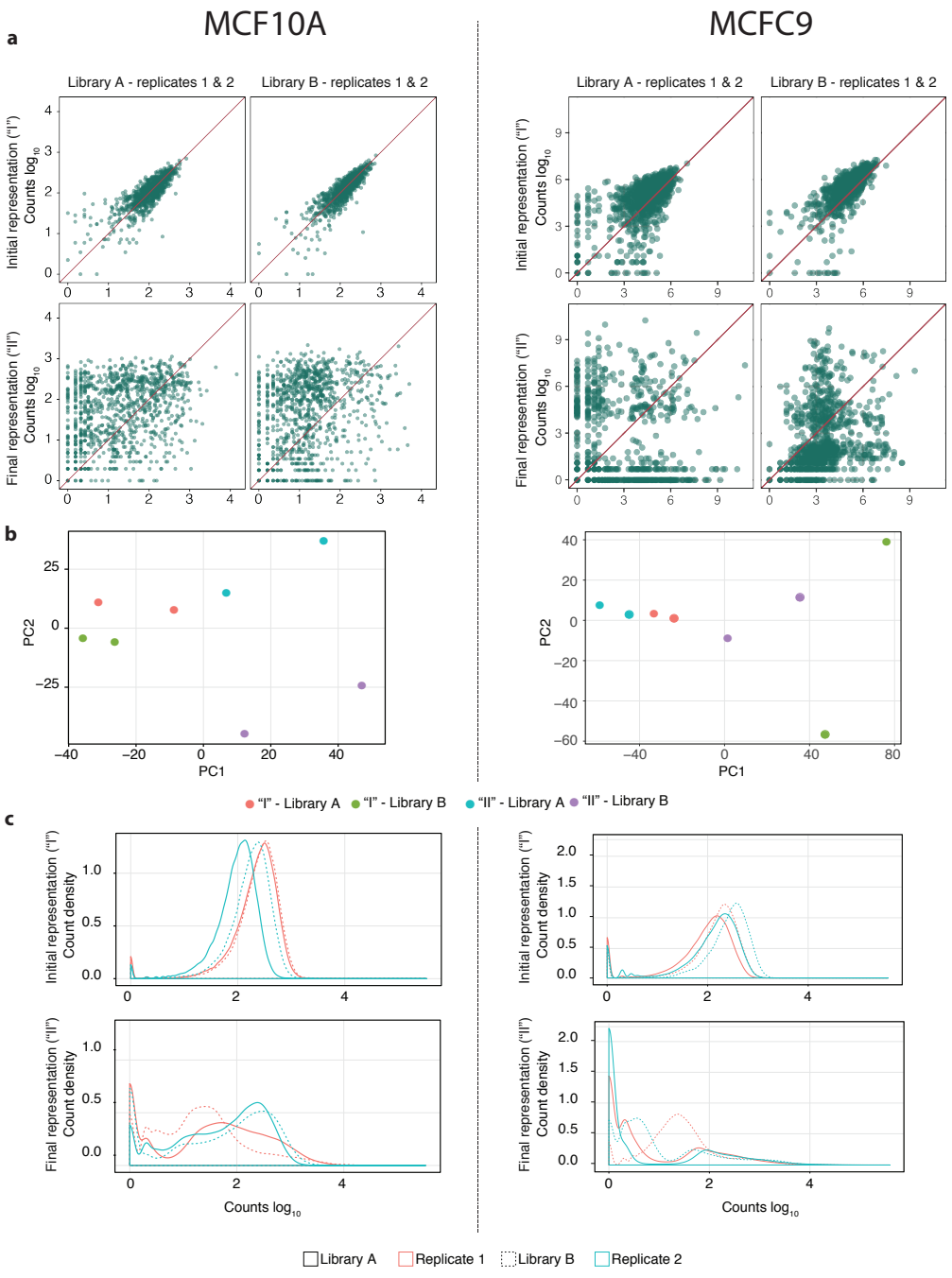


FIGURE B.2: Quality controls of the MCF10A^{WT} and MCF10A$^{TP53^{-/-},hyperploid}$ trasnformation screen. (**a**) The representation of each control guide in replicate was plotted against each other in order to evaluate the reproducibility of the screen. The final representation shows that the results are not reproducible. (**b**) Principle component analysis plots. (**c**) Density plots.

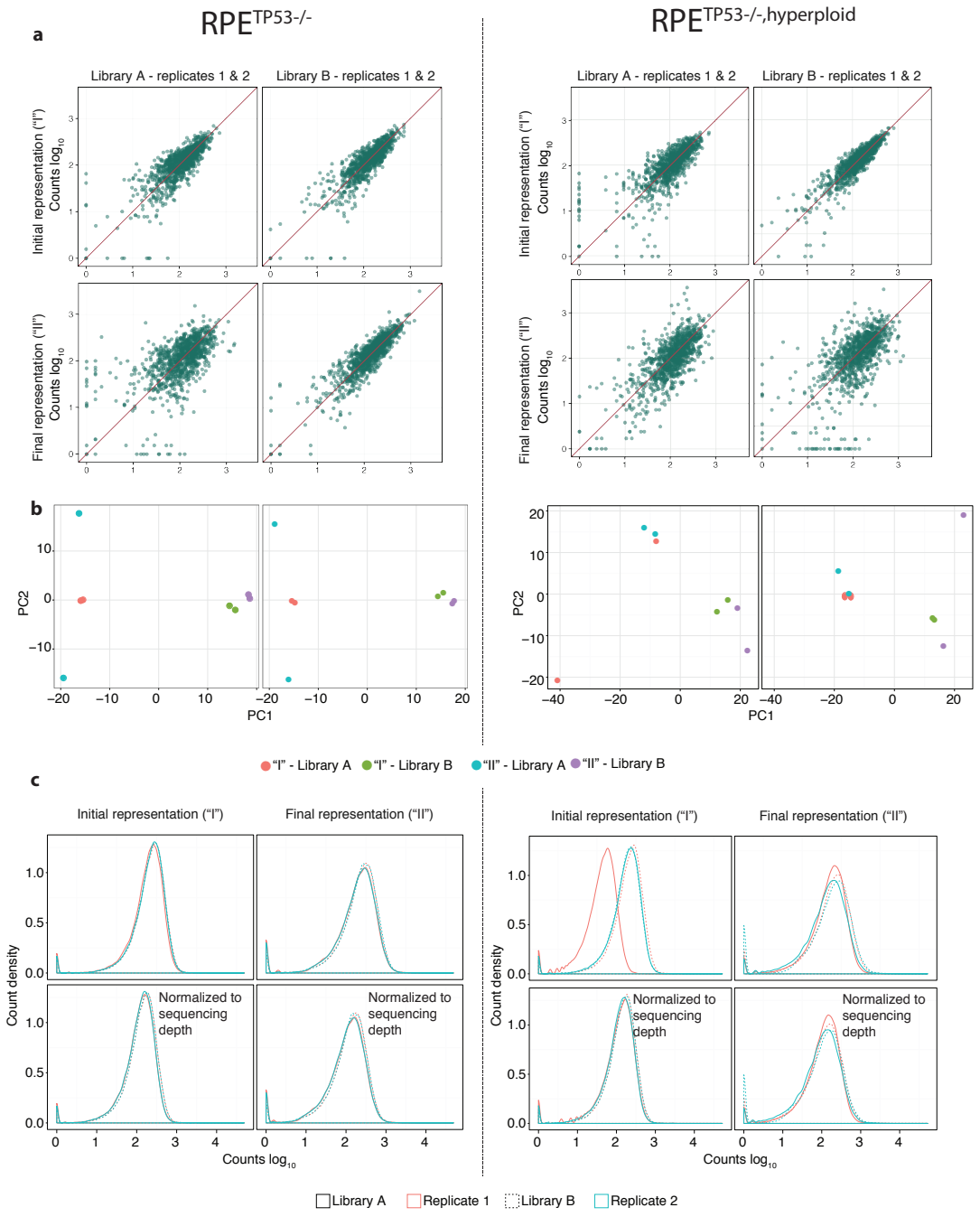


FIGURE B.3: Quality controls of the RPE^{WT} and $RPE^{TP53^{-/-},hyperploid}$ no-selection growth screen. (**a**) The representation of each control guide in replicate was plotted against each other in order to evaluate the reproducibility of the screen. (**b**) Principle component analysis plots indicate that after normalization the replicates are similar. (**c**) Density plots before and after normalization. Replicates overlap after normalization.

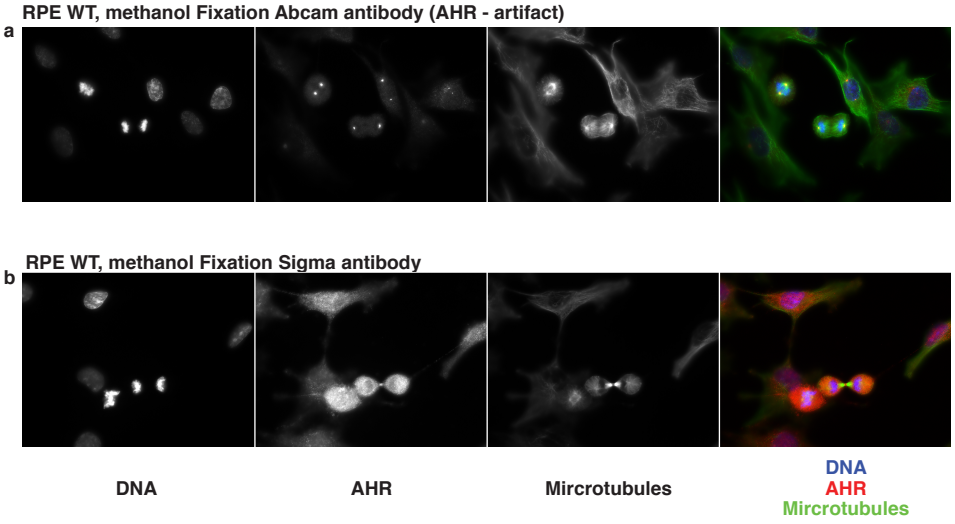


FIGURE B.4: Immunofluorescence of RPE$^{TP53^{-/-}}$ stained for DNA, AHR and microtubules. (**a**) AHR antibody is localizing in the centrosomes.Though this is also present in the knock out clones (data not shown), indicating that this staining is an artifact. (**b**) Testing the sigma antibody shows that there is increase expression of AHR in mitotic cells. This observation needs to still be verified with AHR knock out clones.

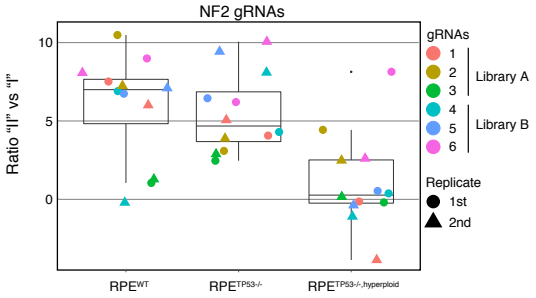


FIGURE B.5: Ratio of normalized counts of each gRNA for *NF2* in all cell lines in the transformation screen. Only two guides targeting *NF2* in the RPE$^{TP53^{-/-},hyperploid}$ give higher levels of gRNA representation.

Bibliography

Akeno N., Miller A. L., Ma X., and Wikenheiser-Brokamp K. A. p53 suppresses carcinoma progression by inhibiting mTOR pathway activation. *Oncogene*, 34(5):589–599, January 2015.

Albertson D. G. Gene amplification in cancer. *Trends in genetics : TIG*, 22(8):447–455, August 2006.

Andersson P., McGuire J., Rubio C., Gradin K., Whitelaw M. L., Pettersson S., Hanberg A., and Poellinger L. A constitutively active dioxin/aryl hydrocarbon receptor induces stomach tumors. *Proceedings of the national academy of sciences of the United States of America*, 99(15):9990–9995, July 2002.

Astsaturov I., Ratushny V., Sukhanova A., Einarson M. B., Bagnyukova T., Zhou Y., Devarajan K., Silverman J. S., Tikhmyanova N., Skobeleva N. et al. Synthetic lethal screen of an EGFR-centered network to improve targeted therapies. *Science signaling*, 3(140):ra67–ra67, 2010.

Barrangou R. Diversity of CRISPR-Cas immune systems and molecular machines. *Genome biology*, 16(1):247, 2015.

Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D. A., and Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(5819):1709–1712, March 2007.

Bates D., Mächler M., Bolker B., and Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

Beli P., Lukashchuk N., Wagner S. A., Weinert B. T., Olsen J. V., Baskcomb L., Mann M., Jackson S. P., and Choudhary C. Proteomic investigations reveal a role for RNA

processing factor THRAP3 in the DNA damage response. *Molecular cell*, 46(2):212–225, April 2012.

Bid H. K., Roberts R. D., Manchanda P. K., and Houghton P. J. RAC1: an emerging therapeutic option for targeting cancer angiogenesis and metastasis. *Molecular cancer therapeutics*, 12(10):1925–1934, October 2013.

Bieging K. T., Mello S. S., and Attardi L. D. Unravelling mechanisms of p53-mediated tumour suppression. *Nature reviews. Cancer*, 14(5):359–370, May 2014.

Blasco M. A. Telomeres and human disease: ageing, cancer and beyond. *Nature reviews. Genetics*, 6(8):611–622, August 2005.

Boch J., Scholze H., Schornack S., Landgraf A., Hahn S., Kay S., Lahaye T., Nickstadt A., and Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, 326(5959):1509–1512, December 2009.

Boggiano J. C., Vanderzalm P. J., and Fehon R. G. Tao-1 phosphorylates Hippo/MST kinases to regulate the Hippo-Salvador-Warts tumor suppressor pathway. *Developmental cell*, 21(5):888–895, November 2011.

Bonilla X., Parmentier L., King B., Bezrukov F., Kaya G., Zoete V., Seplyarskiy V. B., Sharpe H. J., McKee T., Letourneau A. et al. Genomic analysis identifies new drivers and progression pathways in skin basal cell carcinoma. *Nature genetics*, 48(4):398–406, April 2016.

Bosco E. E., Nakai Y., Hennigan R. F., Ratner N., and Zheng Y. NF2-deficient cells depend on the Rac1-canonical Wnt signaling pathway to promote the loss of contact inhibition of proliferation. *Oncogene*, 29(17):2540–2549, April 2010.

Cavenee W. K. and White R. L. The genetic basis of cancer. *Scientific American*, 272 (3):72–79, March 1995.

Chen B., Gilbert L. A., Cimini B. A., Schnitzbauer J., Zhang W., Li G.-W., Park J., Blackburn E. H., Weissman J. S., Qi L. S. et al. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*, 155(7):1479–1491, December 2013.

Chen H.-Y. and Chen R.-H. Cullin 3 Ubiquitin Ligases in Cancer Biology: Functions and Therapeutic Implications. *Frontiers in oncology*, 6:113, 2016.

Chen S., Sanjana N. E., Zheng K., Shalem O., Lee K., Shi X., Scott D. A., Song J., Pan J. Q., Weissleder R. et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*, 160(6):1246–1260, March 2015.

Chiasson-MacKenzie C., Morris Z. S., Baca Q., Morris B., Coker J. K., Mirchev R., Jensen A. E., Carey T., Stott S. L., Golan D. E. et al. NF2/Merlin mediates contact-dependent inhibition of EGFR mobility and internalization via cortical actomyosin. *The Journal of cell biology*, 211(2):391–405, October 2015.

Chien J., Staub J., Avula R., Zhang H., Liu W., Hartmann L. C., Kaufmann S. H., Smith D. I., and Shridhar V. Epigenetic silencing of TCEAL7 (Bex4) in ovarian cancer. *Oncogene*, 24(32):5089–5100, July 2005.

Collins B. C., Gillet L. C., Rosenberger G., Röst H. L., Vichalkovski A., Gstaiger M., and Aebersold R. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nature methods*, 10(12):1246–1253, December 2013.

Cooper S. E., Hodimont E., and Green C. M. A fluorescent bimolecular complementation screen reveals MAF1, RNF7 and SETD3 as PCNA-associated proteins in human cells. *Cell Cycle*, 14(15):2509–2519, August 2015.

Cosme-Blanco W., Shen M.-F., Lazar A. J. F., Pathak S., Lozano G., Multani A. S., and Chang S. Telomere dysfunction suppresses spontaneous tumorigenesis in vivo by initiating p53-dependent cellular senescence. *EMBO reports*, 8(5):497–503, May 2007.

Couzens A. L., Knight J. D. R., Kean M. J., Teo G., Weiss A., Dunham W. H., Lin Z.-Y., Bagshaw R. D., Sicheri F., Pawson T. et al. Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Science signaling*, 6(302):rs15–rs15, November 2013.

Croce C. M. Oncogenes and cancer. *The New England journal of medicine*, 358(5): 502–511, January 2008.

Croucher D. R., Saunders D. N., Lobov S., and Ranson M. Revisiting the biological roles of PAI2 (SERPINB2) in cancer. *Nature reviews. Cancer*, 8(7):535–545, July 2008.

Curto M., Cole B. K., Lallemand D., Liu C.-H., and McClatchey A. I. Contact-dependent inhibition of EGFR signaling by Nf2/Merlin. *The Journal of cell biology*, 177(5):893–903, June 2007.

d'Adda di Fagagna F., Reaper P. M., Clay-Farrace L., Fiegler H., Carr P., Von Zglinicki T., Saretzki G., Carter N. P., and Jackson S. P. A DNA damage checkpoint response in telomere-initiated senescence. *Nature*, 426(6963):194–198, November 2003.

Datsenko K. A., Pougach K., Tikhonov A., Wanner B. L., Severinov K., and Semenova E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature communications*, 3:945, 2012.

Davis H. E., Morgan J. R., and Yarmush M. L. Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes. *Biophysical chemistry*, 97(2-3):159–172, June 2002.

Denison M. S. and Nagy S. R. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annual review of pharmacology and toxicology*, 43(1):309–334, 2003.

Dixon J. R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., Hu M., Liu J. S., and Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, May 2012.

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. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology*, September 2014.

Dominguez A. A., Lim W. A., and Qi L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nature reviews. Molecular cell biology*, 17(1):5–15, January 2016.

Donehower L. A., Harvey M., Slagle B. L., McArthur M. J., Montgomery C. A., Butel J. S., and Bradley A. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature*, 356(6366):215–221, March 1992.

Doudna J. A. and Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, 346(6213):1258096–1258096, November 2014.

Draviam V. M., Stegmeier F., Nalepa G., Sowa M. E., Chen J., Liang A., Hannon G. J., Sorger P. K., Harper J. W., and Elledge S. J. A functional genomic screen identifies a

role for TAO1 kinase in spindle-checkpoint signalling. *Nature cell biology*, 9(5):556–564, May 2007.

Epinat J.-C., Arnould S., Chames P., Rochaix P., Desfontaines D., Puzin C., Patin A., Zanghellini A., Pâques F., and Lacroix E. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic acids research*, 31 (11):2952–2962, June 2003.

Ernst J., Kheradpour P., Mikkelsen T. S., Shoresh N., Ward L. D., Epstein C. B., Zhang X., Wang L., Issner R., Coyne M. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49, May 2011.

Eskiocak U., Kim S. B., Ly P., Roig A. I., Biglione S., Komurov K., Cornelius C., Wright W. E., White M. A., and Shay J. W. Functional parsing of driver mutations in the colorectal cancer genome reveals numerous suppressors of anchorage-independent growth. *Cancer research*, 71(13):4359–4365, July 2011.

Eulalio A., Behm-Ansmant I., and Izaurralde E. P bodies: at the crossroads of post-transcriptional pathways. *Nature reviews. Molecular cell biology*, 8(1):9–22, January 2007.

Evers B., Jastrzebski K., Heijmans J. P. M., Grernrum W., Beijersbergen R. L., and Bernards R. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nature biotechnology*, 34(6):631–633, June 2016.

Fan Y., Boivin G. P., Knudsen E. S., Nebert D. W., Xia Y., and Puga A. The aryl hydrocarbon receptor functions as a tumor suppressor of liver carcinogenesis. *Cancer research*, 70(1):212–220, January 2010.

Fisher R., Pusztai L., and Swanton C. Cancer heterogeneity: implications for targeted therapeutics. *British journal of cancer*, 108(3):479–485, February 2013.

Franken N. A. P., Rodermond H. M., Stap J., Haveman J., and van Bree C. Clonogenic assay of cells in vitro. *Nature protocols*, 1(5):2315–2319, 2006.

Freedman V. H. and Shin S. I. Cellular tumorigenicity in nude mice: correlation with cell growth in semi-solid medium. *Cell*, 3(4):355–359, December 1974.

Fridman J. S. and Lowe S. W. Control of apoptosis by p53. *Oncogene*, 22(56):9030–9040, December 2003.

Fritz W. A., Lin T.-M., Cardiff R. D., and Peterson R. E. The aryl hydrocarbon receptor inhibits prostate carcinogenesis in TRAMP mice. *Carcinogenesis*, 28(2):497–505, February 2007.

Gamsjaeger R., Liew C. K., Loughlin F. E., Crossley M., and Mackay J. P. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends in biochemical sciences*, 32(2):63–70, February 2007.

Ganem N. J., Cornils H., Chiu S.-Y., O'Rourke K. P., Arnaud J., Yimlamai D., Théry M., Camargo F. D., and Pellman D. Cytokinesis failure triggers hippo tumor suppressor pathway activation. *Cell*, 158(4):833–848, August 2014.

Gersbach C. A. and Perez-Pinera P. Activating human genes with zinc finger proteins, transcription activator-like effectors and CRISPR/Cas9 for gene therapy and regenerative medicine. *Expert opinion on therapeutic targets*, 18(8):835–839, August 2014.

Gilbert L. A., Larson M. H., Morsut L., Liu Z., Brar G. A., Torres S. E., Stern-Ginossar N., Brandman O., Whitehead E. H., Doudna J. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2):442–451, July 2013.

Gilbert L. A., Horlbeck M. A., Adamson B., Villalta J. E., Chen Y., Whitehead E. H., Guimaraes C., Panning B., Ploegh H. L., Bassik M. C. et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, 159(3):647–661, October 2014.

Gobeil S., Zhu X., Doillon C. J., and Green M. R. A genome-wide shRNA screen identifies GAS1 as a novel melanoma metastasis suppressor gene. *Genes & development*, 22(21): 2932–2940, November 2008.

Gröschel S., Sanders M. A., Hoogenboezem R., de Wit E., Bouwman B. A. M., Erpelinck C., van der Velden V. H. J., Havermans M., Avellino R., van Lom K. et al. A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. *Cell*, 157(2):369–381, April 2014.

Hamaratoglu F., Willecke M., Kango-Singh M., Nolo R., Hyun E., Tao C., Jafar-Nejad H., and Halder G. The tumour-suppressor genes NF2/Merlin and Expanded act through Hippo signalling to regulate cell proliferation and apoptosis. *Nature cell biology*, 8(1):27–36, January 2006.

Hanahan D. and Weinberg R. A. The hallmarks of cancer. *Cell*, 100(1):57–70, January 2000.

Hanahan D. and Weinberg R. A. Hallmarks of cancer: the next generation. *Cell*, 144 (5):646–674, March 2011.

Hariharan I. K. Energy stress tames the Hippo pathway. *Nature cell biology*, 17(4): 362–363, April 2015.

Hart T., Chandrashekhar M., Aregger M., Steinhart Z., Brown K. R., MacLeod G., Mis M., Zimmermann M., Fradet-Turcotte A., Sun S. et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, 163(6): 1515–1526, December 2015.

Harvey K. and Tapon N. The Salvador-Warts-Hippo pathway - an emerging tumour-suppressor network. *Nature reviews. Cancer*, 7(3):182–191, March 2007.

Hast B. E., Cloer E. W., Goldfarb D., Li H., Siesser P. F., Yan F., Walter V., Zheng N., Hayes D. N., and Major M. B. Cancer-derived mutations in KEAP1 impair NRF2 degradation but not ubiquitination. *Cancer research*, 74(3):808–817, February 2014.

Hasty P., Sharp Z. D., Curiel T. J., and Campisi J. mTORC1 and p53: clash of the gods? *Cell Cycle*, 12(1):20–25, January 2013.

Hasumi H., Baba M., Hasumi Y., Lang M., Huang Y., Oh H. F., Matsuo M., Merino M. J., Yao M., Ito Y. et al. Folliculin-interacting proteins Fnip1 and Fnip2 play critical roles in kidney tumor suppression in cooperation with Flcn. *Proceedings of the national academy of sciences of the United States of America*, page 201419502, March 2015.

Hayashi M. T., Cesare A. J., Rivera T., and Karlseder J. Cell death during crisis is mediated by mitotic telomere deprotection. *Nature*, 522(7557):492–496, June 2015.

Henson J. D., Neumann A. A., Yeager T. R., and Reddel R. R. Alternative lengthening of telomeres in mammalian cells. *Oncogene*, 21(4):598–610, January 2002.

Hnisz D., Abraham B. J., Lee T. I., Lau A., Saint-André V., Sigova A. A., Hoke H. A., and Young R. A. Super-enhancers in the control of cell identity and disease. *Cell*, 155 (4):934–947, November 2013.

Hnisz D., Weintraub A. S., Day D. S., Valton A.-L., Bak R. O., Li C. H., Goldmann J., Lajoie B. R., Fan Z. P., Sigova A. A. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science (New York, N.Y.)*, page aad9024, March 2016.

Hoadley K. A., Yau C., Wolf D. M., Cherniack A. D., Tamborero D., Ng S., Leiserson M. D. M., Niu B., McLellan M. D., Uzunangelov V. et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, 158(4):929–944, August 2014.

Horn S., Figl A., Rachakonda P. S., Fischer C., Sucker A., Gast A., Kadel S., Moll I., Nagore E., Hemminki K. et al. TERT promoter mutations in familial and sporadic melanoma. *Science (New York, N.Y.)*, 339(6122):959–961, February 2013.

Horowitz J. M., Yandell D. W., Park S. H., Canning S., Whyte P., Buchkovich K., Harlow E., Weinberg R. A., and Dryja T. P. Point mutational inactivation of the retinoblastoma antioncogene. *Science (New York, N.Y.)*, 243(4893):937–940, February 1989.

Hovestadt V., Jones D. T. W., Picelli S., Wang W., Kool M., Northcott P. A., Sultan M., Stachurski K., Ryzhova M., Warnatz H.-J. et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature*, 510(7506):537–541, June 2014.

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. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9):827–832, September 2013.

Hu G. and Luo J. A primer on using pooled shRNA libraries for functional genomic screens. *Acta biochimica et biophysica Sinica*, 44(2):103–112, February 2012.

Huang F. W., Hodis E., Xu M. J., Kryukov G. V., Chin L., and Garraway L. A. Highly recurrent TERT promoter mutations in human melanoma. *Science (New York, N.Y.)*, 339(6122):957–959, February 2013.

Huber F., Meurer M., Bunina D., Kats I., Maeder C. I., Stefl M., Mongis C., and Knop M. PCR Duplication: A One-Step Cloning-Free Method to Generate Duplicated Chromosomal Loci and Interference-Free Expression Reporters in Yeast. *PloS one*, 9 (12):e114590, 2014.

Ikeda M., Chiba S., Ohashi K., and Mizuno K. Furry protein promotes aurora A-mediated Polo-like kinase 1 activation. *The Journal of biological chemistry*, 287(33): 27670–27681, August 2012.

Ikuta T., Tachibana T., Watanabe J., Yoshida M., Yoneda Y., and Kawajiri K. Nucleocytoplasmic shuttling of the aryl hydrocarbon receptor. *Journal of biochemistry*, 127 (3):503–509, March 2000.

Jaramillo M. C. and Zhang D. D. The emerging role of the Nrf2-Keap1 signaling pathway in cancer. *Genes & development*, 27(20):2179–2191, October 2013.

Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., and Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096):816–821, August 2012.

Johnson R. and Halder G. The two faces of Hippo: targeting the Hippo pathway for regenerative medicine and cancer treatment. *Nature reviews. Drug discovery*, 13(1): 63–79, January 2014.

Kaelin W. G. The concept of synthetic lethality in the context of anticancer therapy. *Nature reviews. Cancer*, 5(9):689–698, September 2005.

Kannan K., Kaminski N., Rechavi G., Jakob-Hirsch J., Amariglio N., and Givol D. DNA microarray analysis of genes involved in p53 mediated apoptosis: activation of Apaf-1. *Oncogene*, 20(26):3449–3455, June 2001.

Kim N. W., Piatyszek M. A., Prowse K. R., Harley C. B., West M. D., Ho P. L., Coviello G. M., Wright W. E., Weinrich S. L., and Shay J. W. Specific association of human telomerase activity with immortal cells and cancer. *Science (New York, N.Y.)*, 266 (5193):2011–2015, December 1994.

Kim Y. G., Cha J., and Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the national academy of sciences of the United States of America*, 93(3):1156–1160, February 1996.

König R., Chiang C.-y., Tu B. P., Yan S. F., DeJesus P. D., Romero A., Bergauer T., Orth A., Krueger U., Zhou Y. et al. A probability-based approach for the analysis of large-scale RNAi screens. *Nature methods*, 4(10):847–849, October 2007.

Kraft K., Geuer S., Will A. J., Chan W. L., Paliou C., Borschiwer M., Harabula I., Wittler L., Franke M., Ibrahim D. M. et al. Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell reports*, 10(5):833–839, February 2015.

Land H., Parada L. F., and Weinberg R. A. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature*, 304(5927):596–602, August 1983.

Lane D. P. Cancer. p53, guardian of the genome. *Nature*, 358(6381):15–16, July 1992.

Laplante M. and Sabatini D. M. mTOR signaling in growth control and disease. *Cell*, 149(2):274–293, April 2012.

Leighton P. A., Saam J. R., Ingram R. S., Stewart C. L., and Tilghman S. M. An enhancer deletion affects both H19 and Igf2 expression. *Genes & development*, 9(17): 2079–2089, September 1995.

Li J., Shou J., Guo Y., Tang Y., Wu Y., Jia Z., Zhai Y., Chen Z., Xu Q., and Wu Q. Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *Journal of molecular cell biology*, 7(4):284–298, August 2015.

Li W., Cooper J., Zhou L., Yang C., Erdjument-Bromage H., Zagzag D., Snuderl M., Ladanyi M., Hanemann C. O., Zhou P. et al. Merlin/NF2 loss-driven tumorigenesis linked to CRL4(DCAF1)-mediated inhibition of the hippo pathway kinases Lats1 and 2 in the nucleus. *Cancer cell*, 26(1):48–60, July 2014a.

Li W., Xu H., Xiao T., Cong L., Love M. I., Zhang F., Irizarry R. A., Liu J. S., Brown M., and Liu X. S. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome biology*, 15(12):554, 2014b.

Liao Y., Smyth G. K., and Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research*, 41(10):e108–e108, May 2013.

Lin C. Y., Erkek S., Tong Y., Yin L., Federation A. J., Zapatka M., Haldipur P., Kawauchi D., Risch T., Warnatz H.-J. et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature*, 530(7588):57–62, February 2016.

Liu G., Parant J. M., Lang G., Chau P., Chavez-Reyes A., El-Naggar A. K., Multani A., Chang S., and Lozano G. Chromosome stability, in the absence of apoptosis, is critical for suppression of tumorigenesis in Trp53 mutant mice. *Nature genetics*, 36(1):63–68, January 2004.

Ludlow L. B., Schick B. P., Budarf M. L., Driscoll D. A., Zackai E. H., Cohen A., and Konkle B. A. Identification of a mutation in a GATA binding site of the platelet glycoprotein Ibbeta promoter resulting in the Bernard-Soulier syndrome. *The Journal of biological chemistry*, 271(36):22076–22080, September 1996.

Luo B., Cheung H. W., Subramanian A., Sharifnia T., Okamoto M., Yang X., Hinkle G., Boehm J. S., Beroukhim R., Weir B. A. et al. Highly parallel identification of essential genes in cancer cells. *Proceedings of the national academy of sciences of the United States of America*, 105(51):20380–20385, December 2008.

Luo J., Emanuele M. J., Li D., Creighton C. J., Schlabach M. R., Westbrook T. F., Wong K.-K., and Elledge S. J. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell*, 137(5):835–848, May 2009.

Ly P., Eskiocak U., Parker C. R., Harris K. J., Wright W. E., and Shay J. W. RNAi screening of the human colorectal cancer genome identifies multifunctional tumor suppressors regulating epithelial cell invasion. *Cell research*, 22(11):1605–1608, November 2012.

Lynch T. J., Bell D. W., Sordella R., Gurubhagavatula S., Okimoto R. A., Brannigan B. W., Harris P. L., Haserlat S. M., Supko J. G., Haluska F. G. et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *The New England journal of medicine*, 350(21): 2129–2139, May 2004.

Maciejowski J., Li Y., Bosco N., Campbell P. J., and de Lange T. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell*, 163(7):1641–1654, December 2015.

MacLachlan T. K. and El-Deiry W. S. Apoptotic threshold is lowered by p53 transactivation of caspase-6. *Proceedings of the National Academy of Sciences of the United States of America*, 99(14):9492–9497, July 2002.

Makarova K. S., Wolf Y. I., Alkhnbashi O. S., Costa F., Shah S. A., Saunders S. J., Barrangou R., Brouns S. J. J., Charpentier E., Haft D. H. et al. An updated evolutionary

classification of CRISPR-Cas systems. *Nature reviews. Microbiology*, 13(11):722–736, November 2015.

Mansour M. R., Abraham B. J., Anders L., Berezovskaya A., Gutierrez A., Durbin A. D., Etchin J., Lawton L., Sallan S. E., Silverman L. B. et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science (New York, N.Y.)*, 346(6215):1373–1377, December 2014.

Mardin B. R., Drainas A. P., Waszak S. M., Weischenfeldt J., Isokane M., Stütz A. M., Raeder B., Efthymiopoulos T., Buccitelli C., Segura-Wang M. et al. A cell-based model system links chromothripsis with hyperploidy. *Molecular systems biology*, 11(9): 828–828, 2015.

Martincorena I. and Campbell P. J. Somatic mutation in cancer and normal cells. *Science (New York, N.Y.)*, 349(6255):1483–1489, September 2015.

McCoy M. S., Bargmann C. I., and Weinberg R. A. Human colon carcinoma Ki-ras2 oncogene and its corresponding proto-oncogene. *Molecular and cellular biology*, 4(8): 1577–1582, August 1984.

Mertens F., Johansson B., Fioretos T., and Mitelman F. The emerging complexity of gene fusions in cancer. *Nature reviews. Cancer*, 15(6):371–381, June 2015.

Mihaylova M. M. and Shaw R. J. The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. *Nature cell biology*, 13(9):1016–1023, September 2011.

Mitelman F., Johansson B., and Mertens F. The impact of translocations and gene fusions on cancer causation. *Nature reviews. Cancer*, 7(4):233–245, April 2007.

Mori S., Chang J. T., Andrechek E. R., Matsumura N., Baba T., Yao G., Kim J. W., Gatza M., Murphy S., and Nevins J. R. Anchorage-independent cell growth signature identifies tumors with metastatic potential. *Oncogene*, 28(31):2796–2805, August 2009.

Murray I. A., Patterson A. D., and Perdew G. H. Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nature reviews. Cancer*, 14(12):801–814, December 2014.

Nik-Zainal S., Alexandrov L. B., Wedge D. C., Van Loo P., Greenman C. D., Raine K., Jones D., Hinton J., Marshall J., Stebbings L. A. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*, 149(5):979–993, May 2012.

Northcott P. A., Lee C., Zichner T., Stütz A. M., Erkek S., Kawauchi D., Shih D. J. H., Hovestadt V., Zapatka M., Sturm D. et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature*, 511(7510):428–434, July 2014.

Nuñez J. K., Lee A. S. Y., Engelman A., and Doudna J. A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*, 519(7542):193–198, March 2015.

Ohtake F., Fujii-Kuriyama Y., and Kato S. AhR acts as an E3 ubiquitin ligase to modulate steroid receptor functions. *Biochemical pharmacology*, 77(4):474–484, February 2009.

Okada T., Lopez-Lago M., and Giancotti F. G. Merlin/NF-2 mediates contact inhibition of growth by suppressing recruitment of Rac to the plasma membrane. *The Journal of cell biology*, 171(2):361–371, October 2005.

Ooi A., Dykema K., Ansari A., Petillo D., Snider J., Kahnoski R., Anema J., Craig D., Carpten J., Teh B.-T. et al. CUL3 and NRF2 mutations confer an NRF2 activation phenotype in a sporadic form of papillary renal cell carcinoma. *Cancer research*, 73 (7):2044–2051, April 2013.

Orr-Weaver T. L., Szostak J. W., and Rothstein R. J. Yeast transformation: a model system for the study of recombination. *Proceedings of the national academy of sciences of the United States of America*, 78(10):6354–6358, October 1981.

Park S. and Lehner B. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. *Molecular systems biology*, 11(7):824–824, July 2015.

Parker R. and Sheth U. P bodies and the control of mRNA translation and degradation. *Molecular cell*, 25(5):635–646, March 2007.

Parsons S. J. and Parsons J. T. Src family kinases, key regulators of signal transduction. *Oncogene*, 23(48):7906–7909, October 2004.

Peifer M., Hertwig F., Roels F., Dreidax D., Gartlgruber M., Menon R., Krämer A., Roncaioli J. L., Sand F., Heuckmann J. M. et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature*, 526(7575):700–704, October 2015.

Perez-Pinera P., Kocak D. D., Vockley C. M., Adler A. F., Kabadi A. M., Polstein L. R., Thakore P. I., Glass K. A., Ousterout D. G., Leong K. W. et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods*, 10(10): 973–976, October 2013.

Petroski M. D. and Deshaies R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nature reviews. Molecular cell biology*, 6(1):9–20, January 2005.

Pombo A. and Dillon N. Three-dimensional genome architecture: players and mechanisms. *Nature reviews. Molecular cell biology*, 16(4):245–257, April 2015.

Poon C. L. C., Lin J. I., Zhang X., and Harvey K. F. The sterile 20-like kinase Tao-1 controls tissue growth by regulating the Salvador-Warts-Hippo pathway. *Developmental cell*, 21(5):896–906, November 2011.

Puga A., Barnes S. J., Dalton T. P., Chang C. y., Knudsen E. S., and Maier M. A. Aromatic hydrocarbon receptor interaction with the retinoblastoma protein potentiates repression of E2F-dependent transcription and cell cycle arrest. *The Journal of biological chemistry*, 275(4):2943–2950, January 2000.

Puga A., Ma C., and Marlowe J. L. The aryl hydrocarbon receptor cross-talks with multiple signal transduction pathways. *Biochemical pharmacology*, 77(4):713–722, February 2009.

Pugh R. J., Slee J. B., Farwell S. L. N., Li Y., Barthol T., Patton W. A., and Lowe-Krentz L. J. Transmembrane Protein 184A Is a Receptor Required for Vascular Smooth Muscle Cell Responses to Heparin. *The Journal of biological chemistry*, 291(10):5326–5341, March 2016.

Qu B. H., Karas M., Koval A., and LeRoith D. Insulin receptor substrate-4 enhances insulin-like growth factor-I-induced cell proliferation. *The Journal of biological chemistry*, 274(44):31179–31184, October 1999.

Rattan R., Narita K., Chien J., Maguire J. L., Shridhar R., Giri S., and Shridhar V. TCEAL7, a putative tumor suppressor gene, negatively regulates NF-kappaB pathway. *Oncogene*, 29(9):1362–1373, March 2010.

Ren K., Yuan J., Yang M., Gao X., Ding X., Zhou J., Hu X., Cao J., Deng X., Xiang S. et al. KCTD10 is involved in the cardiovascular system and Notch signaling during early embryonic development. *PloS one*, 9(11):e112275, 2014.

Ribeiro P. S., Josué F., Wepf A., Wehr M. C., Rinner O., Kelly G., Tapon N., and Gstaiger M. Combined functional genomic and proteomic approaches identify a PP2A complex as a negative regulator of Hippo signaling. *Molecular cell*, 39(4):521–534, August 2010.

Rizzo P., Osipo C., Foreman K., Golde T., Osborne B., and Miele L. Rational targeting of Notch signaling in cancer. *Oncogene*, 27(38):5124–5131, September 2008.

Roadmap Epigenomics Consortium, Kundaje A., Meuleman W., Ernst J., Bilenky M., Yen A., Heravi-Moussavi A., Kheradpour P., Zhang Z., Wang J. et al. Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539):317–330, February 2015.

Roberts C. W. M., Leroux M. M., Fleming M. D., and Orkin S. H. Highly penetrant, rapid tumorigenesis through conditional inversion of the tumor suppressor gene Snf5. *Cancer cell*, 2(5):415–425, November 2002.

Saliba J., Saint-Martin C., Di Stefano A., Lenglet G., Marty C., Keren B., Pasquier F., Valle V. D., Secardin L., Leroy G. et al. Germline duplication of ATG2B and GSKIP predisposes to familial myeloid malignancies. *Nature genetics*, 47(10):1131–1140, October 2015.

Samant R. S., Clarke P. A., and Workman P. E3 ubiquitin ligase Cullin-5 modulates multiple molecular and cellular responses to heat shock protein 90 inhibition in human cancer cells. *Proceedings of the national academy of sciences of the United States of America*, 111(18):6834–6839, May 2014.

Sancar A., Lindsey-Boltz L. A., Unsal-Kaçmaz K., and Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual review of biochemistry*, 73(1):39–85, 2004.

Sanjana N. E., Shalem O., and Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nature methods*, 11(8):783–784, July 2014.

Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., Preibisch S., Rueden C., Saalfeld S., Schmid B. et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682, July 2012.

Schuler M., Bossy-Wetzel E., Goldstein J. C., Fitzgerald P., and Green D. R. p53 induces apoptosis by caspase activation through mitochondrial cytochrome c release. *The Journal of biological chemistry*, 275(10):7337–7342, March 2000.

Shalem O., Sanjana N. E., Hartenian E., Shi X., Scott D. A., Mikkelsen T. S., Heckl D., Ebert B. L., Root D. E., Doench J. G. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)*, 343(6166):84–87, January 2014.

Silva J. M., Marran K., Parker J. S., Silva J., Golding M., Schlabach M. R., Elledge S. J., Hannon G. J., and Chang K. Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science (New York, N.Y.)*, 319(5863):617–620, February 2008.

Soucy T. A., Dick L. R., Smith P. G., Milhollen M. A., and Brownell J. E. The NEDD8 Conjugation Pathway and Its Relevance in Cancer Biology and Therapy. *Genes & cancer*, 1(7):708–716, July 2010.

Stankiewicz P. and Lupski J. R. Genome architecture, rearrangements and genomic disorders. *Trends in genetics : TIG*, 18(2):74–82, February 2002.

Starheim K. K., Gromyko D., Evjenth R., Ryningen A., Varhaug J. E., Lillehaug J. R., and Arnesen T. Knockdown of human N alpha-terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization. *Molecular and cellular biology*, 29(13):3569–3581, July 2009.

Stehelin D., Fujita D. J., Padgett T., Varmus H. E., and Bishop J. M. Detection and enumeration of transformation-defective strains of avian sarcoma virus with molecular hybridization. *Virology*, 76(2):675–684, February 1977.

Stephens P. J., Greenman C. D., Fu B., Yang F., Bignell G. R., Mudie L. J., Pleasance E. D., Lau K. W., Beare D., Stebbings L. A. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 144(1):27–40, January 2011.

Stewart-Ornstein J. and Lahav G. Dynamics of CDKN1A in Single Cells Defined by an Endogenous Fluorescent Tagging Toolkit. *Cell reports*, 14(7):1800–1811, February 2016.

Stracquadanio G., Wang X., Wallace M. D., Grawenda A. M., Zhang P., Hewitt J., Zeron-Medina J., Castro-Giner F., Tomlinson I. P., Goding C. R. et al. The importance of p53 pathway genetics in inherited and somatic cancer genomes. *Nature reviews. Cancer*, 16(4):251–265, April 2016.

Tateishi K., Omata M., Tanaka K., and Chiba T. The NEDD8 system is essential for cell cycle progression and morphogenetic pathway in mice. *The Journal of cell biology*, 155(4):571–579, November 2001.

Tomas A., Futter C. E., and Eden E. R. EGF receptor trafficking: consequences for signaling and cancer. *Trends in cell biology*, 24(1):26–34, January 2014.

Tsai S. Q. and Joung J. K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nature reviews. Genetics*, 17(5):300–312, April 2016.

Tsai S. Q., Zheng Z., Nguyen N. T., Liebers M., Topkar V. V., Thapar V., Wyvekens N., Khayter C., Iafrate A. J., Le L. P. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology*, 33(2):187–197, February 2015.

Untergasser A., Nijveen H., Rao X., Bisseling T., Geurts R., and Leunissen J. A. M. Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research*, 35(Web Server issue):W71–4, July 2007.

Urnov F. D., Rebar E. J., Holmes M. C., Zhang H. S., and Gregory P. D. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*, 11(9):636–646, September 2010.

van de Werken H. J. G., Landan G., Holwerda S. J. B., Hoichman M., Klous P., Chachik R., Splinter E., Valdes-Quezada C., Oz Y., Bouwman B. A. M. et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature methods*, 9(10): 969–972, October 2012.

Vousden K. H. and Prives C. Blinded by the Light: The Growing Complexity of p53. *Cell*, 137(3):413–431, May 2009.

Wang T., Wei J. J., Sabatini D. M., and Lander E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science (New York, N.Y.)*, 343(6166):80–84, January 2014a.

Wang W., Huang J., Wang X., Yuan J., Li X., Feng L., Park J.-I., and Chen J. PTPN14 is required for the density-dependent control of YAP1. *Genes & development*, 26(17): 1959–1971, September 2012.

Wang X., Fu A. Q., McNerney M. E., and White K. P. Widespread genetic epistasis among cancer genes. *Nature communications*, 5:4828, 2014b.

Wang Y., Zheng Y., Luo F., Fan X., Chen J., Zhang C., and Hui R. KCTD10 interacts with proliferating cell nuclear antigen and its down-regulation could inhibit cell proliferation. *Journal of cellular biochemistry*, 106(3):409–413, February 2009.

Weischenfeldt J., Symmons O., Spitz F., and Korbel J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature reviews. Genetics*, 14(2):125–138, February 2013.

Westbrook T. F., Martin E. S., Schlabach M. R., Leng Y., Liang A. C., Feng B., Zhao J. J., Roberts T. M., Mandel G., Hannon G. J. et al. A genetic screen for candidate tumor suppressors identifies REST. *Cell*, 121(6):837–848, June 2005.

Wilson K. E., Li Y.-W., Yang N., Shen H., Orillion A. R., and Zhang J. PTPN14 forms a complex with Kibra and LATS1 proteins and negatively regulates the YAP oncogenic function. *The Journal of biological chemistry*, 289(34):23693–23700, August 2014.

Wilson K. E., Yang N., Mussell A. L., and Zhang J. The Regulatory Role of KIBRA and PTPN14 in Hippo Signaling and Beyond. *Genes*, 7(6):23, 2016.

Winter J., Breinig M., Heigwer F., Brügemann D., Leible S., Pelz O., Zhan T., and Boutros M. caRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens. *Bioinformatics (Oxford, England)*, page btv617, October 2015.

Wolf J., Müller-Decker K., Flechtenmacher C., Zhang F., Shahmoradgoli M., Mills G. B., Hoheisel J. D., and Boettcher M. An in vivo RNAi screen identifies SALL1 as a tumor suppressor in human breast cancer with a role in CDH1 regulation. *Oncogene*, 33(33): 4273–4278, August 2014.

Wright A. V., Nuñez J. K., and Doudna J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell*, 164(1-2):29–44, January 2016.

Xu X., Qiao W., Linke S. P., Cao L., Li W. M., Furth P. A., Harris C. C., and Deng C. X. Genetic interactions between tumor suppressors Brca1 and p53 in apoptosis, cell cycle and tumorigenesis. *Nature genetics*, 28(3):266–271, July 2001.

Yates L. R. and Campbell P. J. Evolution of the cancer genome. *Nature reviews. Genetics*, 13(11):795–806, November 2012.

Yin Y. and Shen W. H. PTEN: a new guardian of the genome. *Oncogene*, 27(41): 5443–5453, September 2008.

Yoshihara K., Wang Q., Torres-Garcia W., Zheng S., Vegesna R., Kim H., and Verhaak R. G. W. The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*, 34(37):4845–4854, September 2015.

Zack T. I., Schumacher S. E., Carter S. L., Cherniack A. D., Saksena G., Tabak B., Lawrence M. S., Zhsng C.-Z., Wala J., Mermel C. H. et al. Pan-cancer patterns of somatic copy number alteration. *Nature genetics*, 45(10):1134–1140, October 2013.

Zender L., Xue W., Zuber J., Semighini C. P., Krasnitz A., Ma B., Zender P., Kubicka S., Luk J. M., Schirmacher P. et al. An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell*, 135(5):852–864, November 2008.

Zhang C.-Z. and Pellman D. From Mutational Mechanisms in Single Cells to Mutational Patterns in Cancer Genomes. *Cold Spring Harbor symposia on quantitative biology*, 80:117–137, 2015.

Zhao B., Li L., Wang L., Wang C.-Y., Yu J., and Guan K.-L. Cell detachment activates the Hippo pathway via cytoskeleton reorganization to induce anoikis. *Genes & development*, 26(1):54–68, January 2012.

Zhu J., Sammons M. A., Donahue G., Dou Z., Vedadi M., Getlik M., Barsyte-Lovejoy D., Al-awar R., Katona B. W., Shilatifard A. et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature*, 525(7568):206–211, September 2015a.

Zhu J., Adli M., Zou J. Y., Verstappen G., Coyne M., Zhang X., Durham T., Miri M., Deshpande V., De Jager P. L. et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*, 152(3):642–654, January 2013.

Zhu K., Liu Q., Zhou Y., Tao C., Zhao Z., Sun J., and Xu H. Oncogenes and tumor suppressor genes: comparative genomics and network perspectives. *BMC genomics*, 16 Suppl 7(Suppl 7):S8, 2015b.