Functional genomics and compound mode-of-action screening in haploid human cells

A thesis submitted for the degree of

Doctor of Philosophy in Clinical Medicine

Trinity Term 2017

Supervisor: Dr Sebastian Nijman

Bianca Gapp

Wolfson College

University of Oxford

Bianca Gapp
Wolfson College

Doctor of Philosophy
Trinity Term 2017

Abstract

Functional genomics and compound mode-of-action screening in haploid human cells

More than a decade after the completion of the human genome project, the function of a large number of genes remains to be elucidated. Forward and reverse genetic approaches have proven to be powerful tools to study gene function and have provided insights into fundamental biological processes. Furthermore, functional genetic screening can lead to a better understanding of the action of endogenous and exogenous stimuli such as hormones or drugs on biological systems. Thus far, systematic and unbiased studies have largely been limited to model organisms. However, complex disease-relevant genotypes and phenotypes cannot be studied in entirety in lower organisms creating a need for systematic approaches in human cells. This thesis describes a series of studies using forward and reverse genetic approaches combined with state-of-the-art technology in haploid human cells. The first chapter describes the development of a quantitative phenotypic read-out using a novel application of RNA-sequencing that allows the functional annotation of genes in signalling pathways. The presented data demonstrate that the employed shallow RNA-sequencing method is scalable and suitable as a read-out for reverse genetic screening. The second chapter focuses on the implementation of this method in a large reverse genetic study in human cells to functionally annotate tyrosine kinases in signalling pathways upon stimulation with a set of ten polypeptides and small molecules. The screens revealed known and unexpected interactions between different signalling molecules and pathways, validating the technical approach in a biological context. The third chapter presents a pilot study describing the set-up of a forward genetic technique for compound mode-of-action screening using a pooled human mutant cell line collection. The chemical genetic approach displayed sufficient sensitivity and allowed to monitor thousands of gene-drug interactions simultaneously. Together, this thesis combines elements to advance technological and biological aspects of functional genomics and chemical genetics.

Declaration

I hereby declare that the thesis that I am submitting is entirely my own work and describes my own research with the exception of the computational analysis and the methods section, which describes this analysis. I have acknowledged the work that has been carried out by others in the text and figure legends. The data presented in chapter 2 and 3 have been previously published in Gapp et al., 2016. Figures taken or adapted from this manuscript are indicated in the respective figure legends. Chapter 4 constitutes unpublished results and figures.

Word count: 35210

Acknowledgements

I dedicate my thesis to my family and my partner, Anne. I would like to take this opportunity to thank my parents, Anita and Karl, for their unfailing support and love. I would also like to thank my sister, Michele, for her support and for always being there for me. I would like to thank especially my partner and best friend Anne, for keeping me healthy and happy, for his constant support and encouragement as well as his patience and always finding the right words. I could not have done it without you!

I thank my supervisor Sebastian Nijman, for his guidance and scientific as well as personal support over the past years. I am truly grateful to have worked with and learned from such an inspirational scientist. I would like to thank Tomasz Konopka, for his supervision, guidance and patience during our work together. I am extremely grateful and consider myself very lucky to be supervised by both of you. I thank you for challenging me, nurturing my scientific skills and making me the scientist I am today.

I thank Vineet Dalal, for working with me during the final stages of my PhD. For his support, dedication and positive nature. Special thanks goes to Barbara Mair for always lending a helping hand, having an open ear and her support to stay strong during challenging times. You have kept me sane in times of need and I feel privileged having worked with you. You were the most amazing colleagues and friends I could have possibly imagined.

I would like to thank past and present Nijman lab members for their constant support; Especially Chandan Seth Nanda, Roberta Baronio, Erica De Zan and Stefania Militi for their support and help and for providing a fun working environment.

I would like to thank my collaborators: Thomas Penz, Tilmann Bürckstümmer, Christoph Bock and Stefan Kubicek. I am grateful to have worked with you and learned from you.

Finally, I would like to thank Boehringer Ingelheim for funding the research that is the basis of this thesis.

Table of contents

List of Figures

List of Tables

Abbreviations

AATK	Apoptosis-associated tyrosine kinase
ABC	ATP binding cassette
ABL	ABL proto-oncogene 1
ABL1	ABL proto-oncogene 1
ABL2	ABL proto-oncogene 2
ACTA	Activin a
ACTB	Activin b
ACVR1B	Activin A receptor type 1B
ACVR2B	Activin A receptor type 2B
AGC	Protein kinase A, G and C
AIMP1	Aminoacyl tRNA synthetase complex interacting multifunctional protein 1
AKT	AKT serine/threonine kinase 1
APHD	Aphidicolin
APOL4	Apolipoprotein L4
ARMC8	Armadillo repeat containing 8
ATCC	American type culture collection
ATP	Adenosine triphosphate
AXL	AXL receptor tyrosine kinase
BCR	B cell receptor
BCR-ABL	Breakpoint cluster region – ABL fusion gene/protein
BMP	Bone morphogenetic protein
BMP2	Bone morphogenetic protein 2
BMP13	Bone morphogenetic protein 13
BMPR1A	Bone morphogenetic protein receptor type 1A
BMPR2	Bone morphogenetic protein receptor type 2
BNIP3	BCL2 interacting protein 3
bp	Base pair
BRAF	B-Raf proto-oncogene
BSA	Bovine serum albumin
BsmBI	*Bacillus stearothermophilus* restriction endonuclease
CALM2	Calmodulin 2
CALM3	Calmodulin 3
CAMK	CaM-activated kinase
CAMP	Camptothecin
Cas	CRISPR-associated protein
Cas9	CRISPR-associated protein Cas9
CCK4	Colon carcinoma kinase 4
CD4	Cluster of differentiation 4
CDC25C	Cell division cycle 25C

CDK	Cyclin dependent kinase
cDNA	Complementary DNA
CHPT1	Choline phosphotransferase 1
CK1	Casein kinase 1
CLK	CDC like kinase
CMGC	CDK, MAPK, GSK and CLK kinase group
CML	Chronic myeloid leukaemia
CMV	Cytomegalovirus
CO_2	Carbon dioxide
COMMD1	Copper metabolism domain containing 1
COMT	Catechol-O-methyltransferase
Cre	Causes recombination
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPR/Cas	Clustered regularly interspaced short palindromic repeats/CRISPR-associated endonuclease
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/CRISPR-associated endonuclease Cas9
crRNA	CRISPR RNA
CSK	C-Src tyrosine kinase
CTNNB1	Catenin beta 1
CXCL9	C-X-C motif chemokine ligand 9
DAG	Diacylglycerol
dCas9	Deficient Cas9
dCas9-KRAB	Deficient Cas9 – Krueppel-associated box
dCas9-VP64	Deficient Cas9 – tetramer of herpes simplex virus protein VP16
DDIT4	DNA damage inducible transcript 4
DDR1	Discoidin domain receptor tyrosine kinase 1
DDR2	Discoidin domain receptor tyrosine kinase 2
DE	Differential expression
DFOM	Deferoxamine
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double-strand breaks
dsRNA	Double stranded RNA
ECL	Enhanced chemiluminescence
eGFP	Enhanced green fluorescent protein
EGFR	Epidermal growth factor receptor
eHAP	Engineered-HAPloid; near-haploid human cell line derived from HAP1
EMS	Ethyl methanesulfonate

ENU	N-ethyl-N-nitrosourea
EPHA2	EPH receptor A2
EPHA3	EPH receptor A3
EPHA4	EPH receptor A4
EPHA6	EPH receptor A6
EPHB1	EPH receptor B1
EPHB3	EPH receptor B3
EPHB4	EPH receptor B4
EPHB6	EPH receptor B6
ERAD	Endoplasmic reticulum-associated degradation
ERBB2	ERBB2 receptor tyrosine kinase 2
ERBB3	ERBB2 receptor tyrosine kinase 3
ES	Embryonic stem cell
FAM78A	Family with sequence similarity 78 member A
FBS	Foetal bovine serum
FC	Fold change
FER	FER tyrosine kinase
FES	FES proto-oncogene tyrosine kinase
FGF	Fibroblast growth factor
FGF1	Fibroblast growth factor 1
FGF18	Fibroblast growth factor 18
FGFR	Fibroblast growth factor receptor
FGFR1	Fibroblast growth factor receptor 1
FGFR2	Fibroblast growth factor receptor 2
FGFR3	Fibroblast growth factor receptor 3
FGFR4	Fibroblast growth factor receptor 4
FLT1	Fms related tyrosine kinase 1
FLT4	Fms related tyrosine kinase 4
FokI	*Flavobacterium okeanokoites* restriction endonuclease
FPKM	Fragments per kilobase million
FRS2	FGFR substrate 2
FRY	FRY microtubule binding protein
FYN	FYN proto-oncogene
GAB1	GRB2-associated binding protein 1
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GBP3	Guanylate binding protein 3
GDF7	Growth differentiation factor 7
GDF11	Growth differentiation factor 11
gDNA	Genomic DNA
GFP	Green fluorescent protein

GLCT	Galectin 1
GLM	General linear models
GNRHR2	Gonadotropin releasing hormone receptor 2
GO	Gene ontology
GRB2	Growth factor receptor-bound 2
GRIK1	Glutamate ionotropic receptor kainite type subunit 1
gRNA	Guide RNA
GSK	Glycogen synthase kinase
GSNAP	Genomic short-read nucleotide alignment program
GWAS	Genome-wide association study
HAP1	Near-haploid human cell line
HCK	HCK proto-oncogene, Src family tyrosine kinase
HCl	Hydrochloric acid
HDR	Homology-directed repair
HEK293T	Human embryonic kidney 293T cell line
HER2	Human epidermal growth factor receptor 2
HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5
HERG	Heregulin
hg19	Human genome version 19
HIF1A	Hypoxia-inducible factor 1-alpha
HiSeq	Illumina HiSeq sequencing system
HIV	Human immunodeficiency virus
HRP	Horseradish peroxidase
HSV-TK	Herpes simplex virus thymidine kinase
HTR6	5-Hydroxytryptamine receptor 6
HU	Hydroxyurea
IC10	Inhibitory concentration of 10 %
i.e.	id est
IFI44	Interferon induced protein 44
IFI44L	Interferon induced protein 44 like
IFNa	Interferon alpha
IFNb	Interferon beta
IFNg	Interferon gamma
IFNL2	Interleukin 28A
IGF1R	Insulin like growth factor 1 receptor
IL13	Interleukin 13
IL16	Interleukin 16
IMDM	Iscove’s Modified Dulbeccos’s Medium
Indel	Insertion or deletion
INS	Insulin

INSR	Insulin receptor
IONM	Ionomycin
IP3	Inositol-3,4,5-triphosphate
IRF9	Interferon regulatory factor 9
ITK	IL2 inducible T-cell kinase
JAK	Janus kinase
JAK1	Janus kinase 1
JAK2	Janus kinase 2
KBM7	Near-haploid human cell line
KO	Knock-out
KRAB	Krueppel-associated box
LACT	Lactic acid
LMTK2	Lemur tyrosine kinase 2
LMTK3	Lemur tyrosine kinase 3
LOC81691	Exonuclease NEF-Sp
loxP	Locus of crossover in phage P1
LTK	Leukocyte receptor tyrosine kinase
LTR	Long-terminal repeats
LYN	LYN proto-oncogene
MAPK	Mitogen-activated protein kinase
MATK	Megakaryocyte-associated tyrosine kinase
MERTK	MER proto-oncogene tyrosine kinase
MET	MET proto-oncogene receptor tyrosine kinase
miRNA	Micro RNA
MoA	Mode-of-action
mRNA	Messenger RNA
NCS	Neocarzinostain
NEB	New England Biolabs
NHEJ	Non-homologous end joining
nt	Nucleotide
NTRK1	Neurotrophic receptor tyrosine kinase 1
NUP62CL	Nucleoporin 62 C-terminal like
pA	Polyadenylation signal
PAM	Protospacer-associated motif
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDGFR	Platelet-derived growth factor receptor
PDGFRA	Platelet-derived growth factor receptor alpha
PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
PI3K	Phosphoinositide 3 kinase

PIP2	Phosphatidylinositol-4,5-bisphosphate
PKC	Protein kinase C
PLCγ	Phospholipase Cγ
PMA	Phorbol myristate acetate
pPol	Retroviral DNA polymerase
PTB	Phosphotyrosine binding
PTK2B	Protein tyrosine kinase 2 beta
PVDF	Polyvinylidene difluoride
qRT-PCR	Quantitative real-time polymerase chain reaction
RAB1A	Ras-related protein rab-1A
RAF	Raf-1 proto-oncogen
RAS	Rat sarcoma protein
RESV	Resveratrol
RET	Ret proto-oncogene
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
ROR1	Receptor tyrosine kinase like orphan receptor 1
ROR2	Receptor tyrosine kinase like orphan receptor 2
ROTN	Rotenone
RSK2	Ribosomal S6 kinase 2
RSP1	R-spondin 1
RT	Room temperature
RTK	Receptor tyrosine kinase
RTP4	Receptor transporter protein 4
RYK	Receptor-like tyrosine kinase
SA	Splice acceptor
SDS	Sodium dodecyl sulfate
SETDB1	SET domain bifurcated 1
SH2	Src homology
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SIRT1	Sirtuin 1
SLC30A7	Solute carrier family 30 member 7
SLK	STE20 like kinase
SNP	Single-nucleotide polymorphism
SOCS4	Suppressor of cytokine signaling 4
SOS	Son of sevenless
SRC	Rous sarcoma oncogene cellular homolog
SSBP2	Single-stranded DNA binding protein 2

STAT	Signal transducer and activator of transcription
STAT1	Signal transducer and activator of transcription 1
STC1	Stanniocalcin 1
STE	Serine/threonine kinase
STYK1	Serine/threonine/tyrosine kinase 1
SYK	Spleen tyrosine kinase
TALE	Transcription activator-like effector
TALEN	Transcription activator-like effector nuclease
TATDN3	TatD DNase domain containing 3
TGFb	Transforming growth factor beta
TILLING	Target-induced local lesions in genomes
TK	Tyrosine kinase
TKL	Tyrosine kinase-like
TNFa	Tumor necrosis factor alpha
TNK1	Tyrosine kinase non receptor 1
TNK2	Tyrosine kinase non receptor 1
tracrRNA	trans-activating crRNA
trRNA	trans-activating crRNA
t-SNE	t-distributed stochastic neighbour embedding
TUNC	Tunicamycin
TYK2	Tyrosine kinase 2
TYRO3	TYRO3 protein tyrosine kinase
UK	United Kingdom
USD	United States dollar
UV	Ultraviolet
VEGFB	Vascular endothelial growth factor b
VEGFR	Vascular endothelial growth factor receptor
VP64	Tetramer of herpes simplex virus protein VP16
VSV-G	Vesicular stomatitis virus glycoprotein
WNT	Wnt family member
WNT3A	Wnt family member 3A
WNT7A	Wnt family member 7A
WT	Wild-type
YES1	YES proto-oncogene 1
ZFN	Zinc finger nucleases

Chapter 1. Introduction

1.1. Genetic screens

The identification of novel drug targets and the understanding of the mode-of-action (MoA) of compounds is a constant battle for improving treatment regimens and clinical outcome. Despite considerable progress in the management of human diseases, clinical successes have been out-weighed by setbacks resulting from the limited understanding of cellular circuitries that underlie healthy and diseased states and drug action.

More than a decade after the completion of the human genome project still a large number of genes, and hence potential drug targets, remain to be assigned to a particular function (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001). Additionally, the lack of understanding of the MoA of compounds has hampered the market entry of new therapeutics (Bunnage et al., 2015). As a consequence, the interest in tools suitable for the functional annotation of genes and elucidation of compound MoA has increased significantly in the post-genomic era (Nijman, 2015; Yi et al., 2017).

Genetic screening has become a valuable tool to elucidate molecular mechanisms underlying biological processes and such approaches have identified effective targets for the treatment of various malignancies. Genetic screens investigate the consequences of genetic perturbations (e.g. gene knock-out, mutation, over-expression) on a given phenotype and can be used to identify causal connections and infer wild-type gene action (Lehner, 2013; Shalem et al., 2015). In combination with small molecules, genetic screening can be exploited to identify drug targets or affected pathways thereby highlighting the MoA of compounds, hence this area of research has been termed chemical genetics (Nijman, 2015).

Elucidating how genes underpin biological processes using high-throughput screening approaches has led to a better understanding of the genetic wiring of normal and malignant cells and the identification of therapeutic targets for the

treatment of human diseases. Chemical genetic screens have indeed helped to unravel compound MoA, resistance mechanisms, drug synergisms and off-target effects for certain small molecules (Lehner, 2013; Mohr et al., 2014; Shalem et al., 2015; Nijman, 2015).

Although substantial progress has been made, our understanding of gene function and the MoA of effective therapeutics is far from complete. Thus, there is a need for the development of innovative screening approaches to understand cellular circuitries and small molecule MoA.

The following introductory sections describe different genetic screening approaches for functional genomics and compound MoA studies. Subsequently, section 1.2. provides an overview of technical approaches for forward and reverse genetics and section 1.3. summarises applications in model organisms and cultured cells.

1.1.1. Traditional genetic screening

Traditional genetic screening can be divided in forward and reverse genetic screening techniques (Figure 1.1). Forward genetic screening seeks to identify genes underlying a particular phenotype. For this type of screening, genetic modifications are introduced at random in cells or whole organisms. Mutants displaying the phenotype of interest are then identified and the affected genes analysed. The advantage of a forward genetic approach is that it offers an unbiased way to uncover genes underpinning a particular process (Shalem et al., 2015). Forward genetic screens have elucidated various aspects of biology ranging from genes important for embryonic development to signalling pathways underlying human health and disease (St Johnston, 2002; Jorgensen and Mango, 2002; Kaletta and Hengartner, 2006). Although the generation of mutants for forward genetic screening can be fast and inexpensive, the mapping of genes underlying an aberrant phenotype can be tedious.

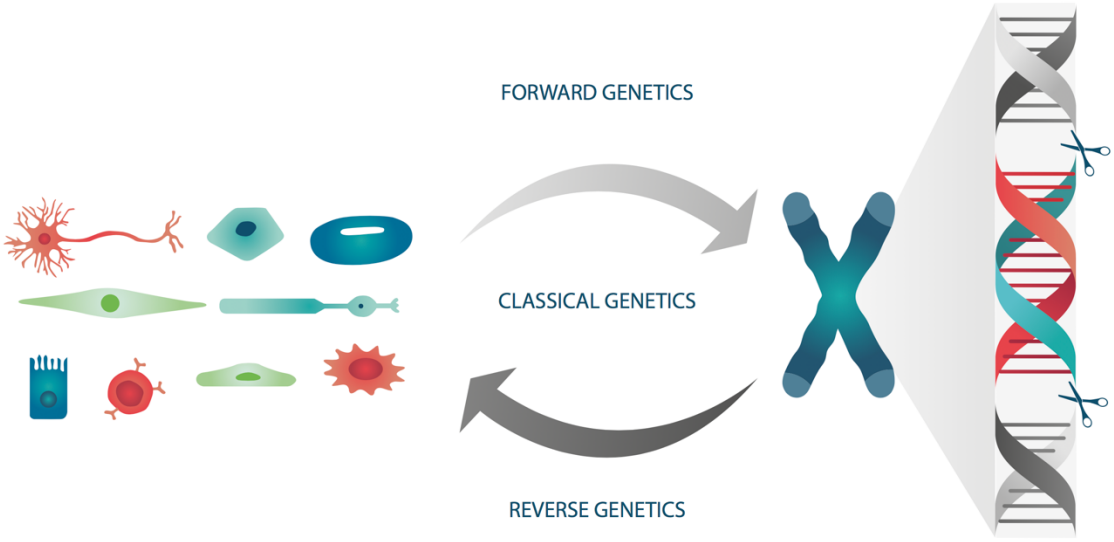


Figure 1.1 Forward and reverse genetics. Genetic screening approaches link genotypes with phenotypes and can be divided in forward and reverse techniques. Forward genetics uncovers genes responsible for a particular phenotype. Reverse genetics investigates phenotypes as a consequence of a specific genetic modification (e.g. gene knock-out, point mutation etc.).

In contrast to forward genetic screens, reverse genetic approaches aim to characterise phenotypes of a mutant panel harbouring predetermined genetic alterations (Shalem et al., 2015). The advantage is that specific genetic alterations are studied in a defined background and that the analysis of many phenotypes can yield a deep understanding of the function of individual genes. Reverse genetic approaches have been instrumental for uncovering genes important in development, physiology and various malignancies (Adams and Sekelsky, 2002; Kaletta and Hengartner, 2006; Kersten et al., 2016).

However, the generation of mutant panels can be cumbersome and the phenotypes analysed are often based on hypotheses about the phenotypic effect of a particular genetic alteration.

1.1.2. Genetic modifier screening

Besides the above described traditional or monogenic screening techniques, genetic modifier screens can be performed to shed light on the function of genes. These screens are carried out in a sensitised background (e.g. knock-out, mutation, over-expression) and second-site modifications are screened for genes that enhance or suppress a particular phenotype (St Johnston, 2002; Jorgensen and Mango, 2002). Because genetic modifier screens aim to identify genes or gene sets that contribute to the manifestation of a particular phenotype and often rely on non-targeted mutagenesis, they could be considered as a subclass of forward genetic screening. Enhancer screens investigate mutations that result in a more pronounced phenotype of a particular mutant, while suppressor screens identify mutations that make the mutant phenotype less severe. Such genetic modifier screens frequently identify genes that play a role in the same pathway or impinge on the same process and are powerful for the study of genetic interactions and pathways (St Johnston, 2002; Jorgensen and Mango, 2002; Boutros and Ahringer, 2008). For example, genetic modifier screens using the fruit fly *Drosophila melanogaster* as a model organism have elucidated various aspects of receptor tyrosine kinase signalling (Simon et al., 1991; Rogge et al., 1991).

Within this framework, synthetic lethality and synthetic viability screens can be considered extreme versions of enhancer and suppressor screens, respectively. Synthetic viability screening refers to the identification of modifications that rescue an otherwise lethal phenotype, whereas synthetic lethality screening describes the identification of combinations of perturbations that result in a severe growth impairment or death (Fece de la Cruz et al., 2015).

1.1.3. Chemical genetic screening

Genetic screening, in combination with small molecules, can be an effective way to identify interactions between genes and compounds that might become useful for

therapeutic intervention. Chemical genetic screening reveals components in the pathway(s) required for drug action or direct drug targets, thereby determining the MoA of compounds (Nijman, 2015). The major advantage is that the identified interactions or drug targets might be directly translatable into clinical applications. For instance, many studies have leveraged chemical genetic screening to elucidate the MoA of chemical agents or to understand drug resistance mechanisms and off-target effects (Kaletta and Hengartner, 2006; Nijman, 2015). Drawbacks for compound screens include that potent and cell-permeable small molecules do not exist for all proteins and that the off-target activity of small molecules can confound the interpretation of phenotypes. Chemical genetic screening in human cells is further detailed in chapter 4.

1.2. Genetic screening technology

The following section of this chapter summarises traditional and state-of-the-art technologies used for forward and reverse genetic screening. Genetic screening techniques can be divided in (i) unbiased non-targeted mutagenesis approaches enabling the identification of unknown genes underlying a particular phenotype (forward genetics), (ii) target-selected mutagenesis techniques where genes are mutagenised at large-scale but only selected mutants are screened (reverse genetics), (iii) gene-targeted mutagenesis methods where only selected genes are targeted and modified (reverse genetics).

1.2.1. Non-targeted mutagenesis approaches

Initially, genetic studies relied on rare, spontaneously arising mutations resulting from DNA replication errors or the incorrect repair of genetic lesions. This made early genetic studies slow, tedious and reliant on simple model organisms that reproduce fast and in large numbers. The advent of non-targeted mutagenesis techniques using

radiation, chemicals or the insertion of DNA sequences, has allowed the generation of mutants at much higher frequency and eventually enabled unbiased large-scale forward genetic and saturation screens also in higher organisms.

1.2.1.1. Radiation mutagenesis

Already in the late 1920s it was shown that the use of x-ray results in a dramatically increased mutation rate in fruit flies (Muller, 1927; Muller, 1930). Since then, ionising radiation (e.g. x-ray and γ-ray) and short-wave ultraviolet (UV) light have been used for the random introduction of large numbers of mutations in diverse model organisms (Chakrabarti et al., 1983; Walker and Streisinger, 1983; Stewart et al., 1991).

Ionising radiation induces different types of DNA damage, ranging from single nucleotide modifications to double strand breaks (DSBs), causing large multigene deletions, translocations and other gross chromosomal aberrations (Breimer, 1988). The use of short-wave UV light results in pyrimidine dimer formation that impedes transcription and replication, ultimately also leading to deletions, translocations and duplications (Bauer et al., 2015). Although radiation mutagenesis offers an effective approach for the introduction of modifications in the DNA of cells and whole organisms, the occurrence of large deletions and complex genomic rearrangements encompassing more than one gene at a time can complicate the mapping and interpretation of mutations underlying a particular phenotype. Due to this drawback, radiation based mutagenesis approaches have been largely replaced by methods that allow the linking of particular phenotypes to lesions in a single gene, such as chemical or insertional mutagenesis techniques.

1.2.1.2. Chemical mutagenesis

The discovery that feeding of ethyl methanesulfonate (EMS) results in high mutagenic loads in *D. melanogaster* marked the beginning of chemical mutagenesis (Lewis and Bacher, 1968) and has since impacted genetic screening dramatically.

There are several chemical agents creating different mutagenic spectra, the most potent and widely used are EMS and N-ethyl-N-nitrosourea (ENU). Both are alkylating agents that modify individual DNA bases and thereby lead to single nucleotide substitutions. While EMS most commonly results in guanine alkylation, ENU can add an ethyl group to all four bases, which interferes with DNA replication and frequently leads to the introduction of point mutations (Sega, 1984; Shibuya and Morimoto, 1993). EMS- or ENU-induced single nucleotide changes can create missense or nonsense mutations and hence gain-of-function, reduced function or loss-of-function of genes. The different types of genetic lesions induced by chemical agents can be used to study genes whose complete loss-of-function would be lethal or to characterise functional domains required for protein interactions or biochemical activity (Housden et al., 2017).

Advantages of chemical mutagenesis approaches include easy administration of mutagens and the efficient generation of point mutations, offering a convenient approach for generating large mutant pools and performing saturation screens. The first saturation screen aiming to uncover all genes underlying a particular phenotype was carried out in *D. melanogaster* and identified genes regulating embryonic patterning (Nüsslein-Volhard and Wieschaus, 1980). One drawback of chemical mutagenesis is that the molecular effect of the generated point mutations can be difficult to pinpoint, as some lead to the inactivation of a gene while others result in an activation. Hence, follow-up experiments have to carefully address all these options. Another complication can be that the mapping of causal mutations has been expensive and time consuming due to the requirement of laborious linkage analysis through crossing schemes with characterised lines (St Johnston, 2002; Jorgensen

and Mango, 2002; Housden et al., 2017). The advent of massively parallel sequencing for the analysis of chemically mutagenised populations (Sarin et al., 2008; Forment et al., 2017) and insertional mutagenesis approaches have helped to overcome difficulties associated with the mapping of mutants and substantially increased the speed with which mutated genes can be identified (see section below).

1.2.1.3. Insertional mutagenesis

Insertional mutagenesis techniques exploit the integration of genetic elements to achieve gene disruption in a high-throughput manner. The use of so-called gene trap vectors, for instance, in combination with viruses or DNA transposons for delivery has been widely applied to introduce insertional mutations in a near-random fashion. Conventional gene trap vectors encode for a splice acceptor, a promoterless reporter gene and a polyadenylation signal (Figure 1.2). Such vectors can be introduced into the genome of target cells by retroviral infection or the use of transposons. If located in the correct orientation in or near the coding region of a gene, the splice acceptor allows the inserted fragment to be fused to an endogenous gene. The encoded polyadenylation signal then causes transcription to be prematurely aborted at this position. Aberrant splicing and the subsequent formation of a truncated fusion protein result in the functional inactivation of the gene. At the same time, the expression of the reporter protein enables the detection of cells or organisms harbouring inactivating gene trap insertions. If gene trap vectors insert in the antisense orientation, gene disruption is only conferred when the vector is integrated in the coding region of a gene, thereby directly disrupting the reading frame. Insertions that occur in promoters or regulatory regions can lead besides the loss-of-function of a gene also to the over- or mis-expression of genes (Stanford et al., 2001; Ivics et al., 2009).

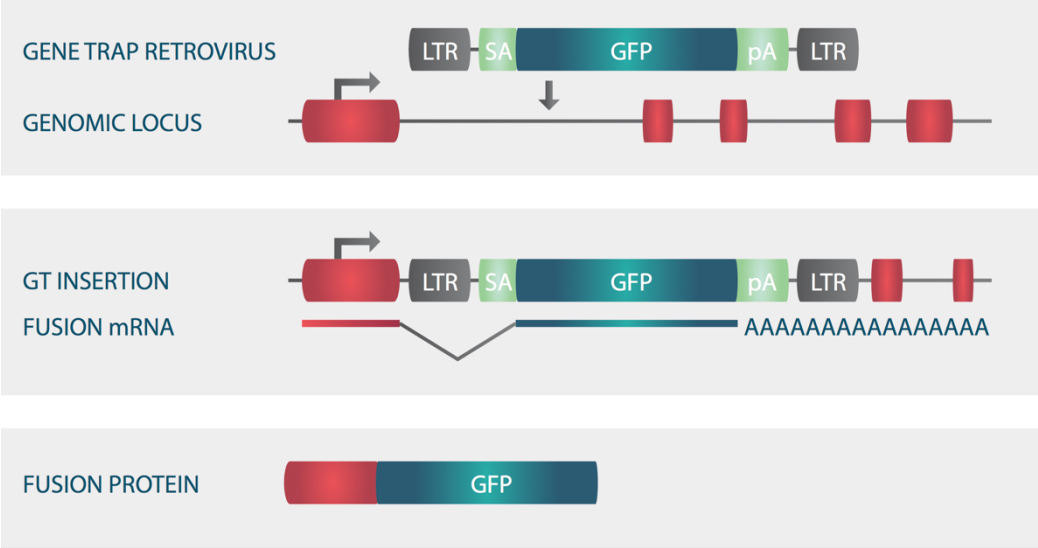


Figure 1.2 Retroviral gene trap vector. Conventional gene trap (GT) vectors consist of a splice acceptor (SA), a promoterless reporter gene (e.g. GFP) and a polyadenylation signal (pA) flanked by retroviral long-terminal repeats (LTR). Packaging of the gene trap vector into a retrovirus and the subsequent infection of target cells results in the near-random integration of the vector in the host genome. Gene trap insertions in an exon (sense and antisense integrations) or intron (sense integrations only) of a coding gene lead to premature termination of transcription due to the polyadenylation signal and aberrant splicing as a consequence of the activity of the splice acceptor. The expression of a truncated mRNA and protein result in the functional disruption of the targeted gene and the expression of the reporter allows the detection of mutants. Adapted from Stanford et al., 2001.

Large-scale insertional mutagenesis screens have identified genes involved in cancer development, embryonic development, host-pathogen interactions and the MoA of compounds (Gaiano et al., 1996; Golling et al., 2002; Kool and Berns, 2009; Carette et al., 2009; Carette et al., 2011a; Carette et al., 2011b; Jae et al., 2013; Jacobson et al., 2013; Jae et al., 2014; Winter et al., 2014; Pillay et al., 2016; Staring et al., 2017). Apart from forward genetic studies, gene trap approaches have been also employed to assemble collections of knock-out animals or isogenic human cells (Cooley et al., 1988; Zambrowicz et al., 1998; Amsterdam et al., 1999; Skarnes et al., 2004; Bürckstümmer et al., 2013). The use of gene trap vectors in combination with

near-haploid human cell lines and the application of insertional mutagenesis techniques to generate large knock-out collections are discussed in more detail in section 2.1.2.

Compared to radiation and chemical mutagenesis, the key advantage of insertional mutagenesis methods is that the inserted fragment does not only inactivate the gene but also directly provides a DNA tag that can be used for the rapid identification of the disrupted gene. Furthermore, large-scale screens comprising hundreds of independent mutants that harbour insertions in the same gene but at different positions allow the calling of differences in the investigated phenotypes with high statistical confidence (Nijman, 2015). A disadvantage of this method is that gene trap integrations mediated by viruses or transposases are not completely randomly distributed over the genome but are biased towards certain integration “hot spots”. Retroviral delivery displays a bias towards the insertion at the 5’-end of actively transcribed genes, whereas transposon mediated insertions show a tendency toward the integration near the transposon donor site (“local hopping”) (Wu et al., 2003; Ivics et al., 2009).

Each of the described non-targeted mutagenesis approaches (radiation, chemical and insertional mutagenesis) comes with their own strengths and weaknesses. A general drawback of non-targeted mutagenesis screens, however, is that the resulting mutants are typically heterozygotes harbouring a mutation only in one allele, while the other one is wild-type. As a consequence, only dominant alleles and haploinsufficient loci result in a detectable phenotype, whereas the inactivation of recessive alleles does not. The unmasking of recessive phenotypes in model organisms requires cumbersome intercrossing of the progeny of heterozygous mutants to obtain homozygously mutant animals (Shalem et al., 2015; Housden et al., 2017). In cultured mammalian cells, the use of near-haploid or Bloom helicase deficient cells, which have an increased mitotic recombination rate, has allowed the generation of homozygous mutant alleles and enabled screens for recessive

phenotypes (Guo et al., 2004; Carette et al., 2009; Elling et al., 2011; Forment et al., 2017). Strategies for the deconvolution of randomly mutagenised pools to generate mutant collections with defined genetic alterations are described in the sections below.

1.2.2. Target-selected mutagenesis

The generation of targeted knock-out collections in multicellular organisms has been challenging due to the lack of methods to specifically manipulate their genomes. Target-selected mutagenesis approaches that rely on random mutagenesis followed by the screening for modifications in desired genes to generate mutant repertoires has been used to circumvent these difficulties.

Chemical mutagenesis strategies have been combined with PCR based approaches or target induced local lesions in genomes (TILLING) to characterise modifications induced by chemical agents: To detect chemically induced insertions or deletions, regions of interest are amplified by PCR and analysed by gel electrophoresis and sequencing. To identify point mutations, PCR products are denatured and annealed to allow the formation of heteroduplexes between mutant and wild-type strands. The resulting heteroduplexes can be treated with endonucleases that cleave mismatches, and the resulting products can again be analysed by denaturing gel electrophoresis and sequencing (Stemple, 2004). Such approaches have enabled the assembly of large mutant panels from chemically mutagenised populations in various systems like plants, worms and vertebrates (Jansen et al., 1997; Liu et al., 1999; Till et al., 2003; Wienholds et al., 2003; Gilchrist et al., 2006).

Another frequently used target-selected mutagenesis strategy exploits insertional mutagenesis, followed by the isolation of individuals harbouring insertions at defined locations in the genome. Insertion sites of individual mutants can be mapped by using frequently cutting restriction enzymes, followed by ligation, inverse PCR with primers that bind the inserted sequence and Sanger sequencing. To map the

insertion sites with an increased throughput, combinatorial pooling strategies and deep sequencing have been used (Stanford et al., 2001). Compared to chemical mutagenesis and subsequent PCR amplification of selected genes, the advantage of insertional mutagenesis is that uncharacterised genes can also be assessed.

Insertional mutagenesis techniques have been used to generate large knock-out collections of mutant individuals, mouse embryonic stem (ES) cells as well as human cells (Amsterdam et al., 1999; Spradling et al., 1999; Kuromori et al., 2004; Skarnes et al., 2004; Austin et al., 2004; Auwerx et al., 2004; Duverger et al., 2007; International Mouse Knockout Consortium et al., 2007; Bellen et al., 2011; Vallin et al., 2012; Bürckstümmer et al., 2013). Chapter 4 of this thesis details how this technique has been employed to generate a large mutant collection of human cells to study the MoA of selected compounds.

1.2.3. RNA interference

Another way to overcome hurdles associated with generating homozygous mutants in diploid organisms or cultured cells is the use of RNA interference (RNAi) to silence the expression of genes in a targeted manner. The discovery that double-stranded RNA is able to interfere with gene expression in *Caenorhabditis elegans* (Fire et al., 1998) was soon followed by the demonstration that RNAi systems can be used to down-regulate gene expression in a variety of other organisms and cultured cells (Hamilton and Baulcombe, 1999; Hammond et al., 2000; Elbashir et al., 2001; Lewis et al., 2002; McCaffrey et al., 2002; Xia et al., 2002). The development of large-scale RNAi libraries has enabled high-throughput forward and reverse genetic approaches, which was a breakthrough for the study of gene function particularly in cultured mammalian cells (Boutros and Ahringer, 2008; Mohr et al., 2010).

RNAi mediated gene silencing can be achieved by the delivery of synthetic long double-stranded RNAs (dsRNAs) or small interfering RNAs (siRNAs), as well as the expression of small hairpin RNAs (shRNAs) from plasmids and viral vectors (Figure

1.3). Long dsRNAs and shRNAs require processing by the cytoplasmic enzyme Dicer which generates 21-23 nucleotide-long double-stranded siRNA duplexes that are incorporated into the RNA-induced silencing complex (RISC). siRNAs do not require pre-processing and can be loaded into the RISC directly. Subsequently, one strand of the siRNA duplex (passenger strand) is cleaved and the other strand (guide strand) is paired to a complementary mRNA sequence by the RISC complex. Perfect complementarity of the guide strand to a particular mRNA sequence results in the targeted degradation of the mRNA, whereas partial complementarity leads to translational repression (Wilson and Doudna, 2013).

Besides RNAi, other methods have been developed that modify gene expression in a targeted manner. For example, the method of choice for reverse genetics in zebrafish is the use of morpholino antisense technology to achieve the sequence-dependent knock-down of genes. Morpholinos are synthetic RNA or DNA oligonucleotides with a unique backbone chemistry. Following injection into zygotes, morpholinos interfere with gene expression by base pairing to complementary mRNA sequences and thus repress their translation similar to RNAi (Nasevicius and Ekker, 2000).

As another example, the CRISPR/Cas9 system (see 1.2.4.2. and 2.1.1.) has been adapted for the targeted modulation of gene expression in a variety of organisms and human cells. Nuclease-deficient Cas9 (dCas9) proteins (Qi et al., 2013) have been coupled to various effector domains that interfere with transcription. Directing dCas9 fusion proteins to a target sequence through an associated short guide RNA (gRNA) results in the sequence-dependent transcriptional repression of targeted genes (Gilbert et al., 2013; Gilbert et al., 2014).

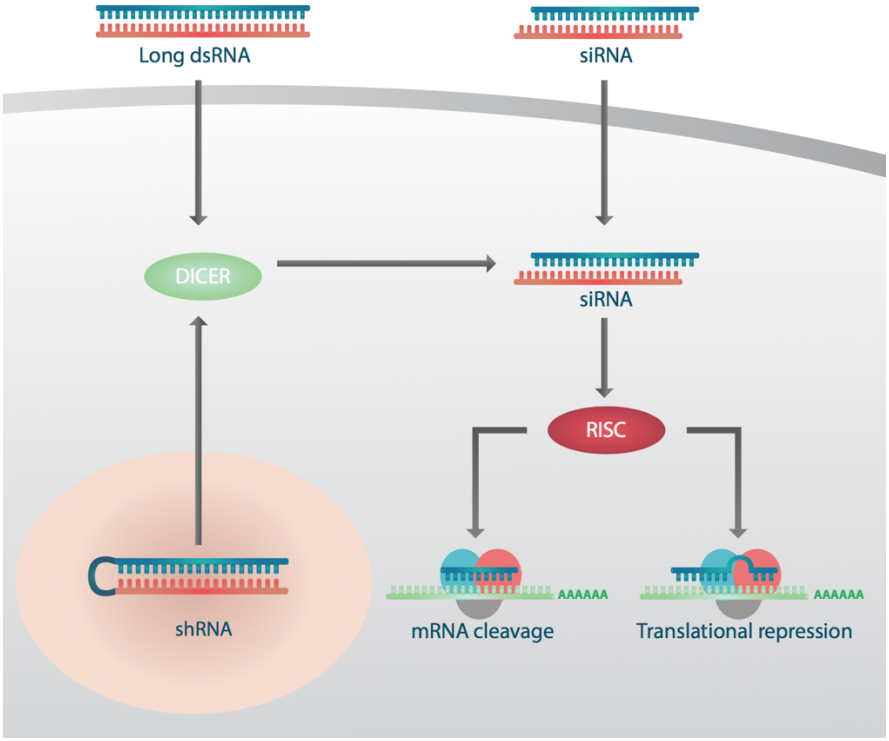


Figure 1.3 Strategies for RNAi mediated gene silencing. Sequence-specific gene knock-down can be achieved by exogenously provided long dsRNAs (500-1000 nucleotides) and short siRNAs (21-23 nucleotides) or the endogenous expression of shRNAs. Long dsRNAs and shRNAs are cleaved by the endonuclease Dicer into 21-23 nucleotide-long siRNA duplexes containing symmetric 3'-overhangs of two nucleotides. Processed or exogenously provided siRNAs are subsequently incorporated into the RISC complex. One strand of the siRNA duplex is degraded, the other strand associates with a homologous mRNA. Perfect complementarity triggers the sequence-specific degradation of the mRNA, partial complementarity results in translational repression. Adapted from Bobbin and Rossi, 2016.

Although RNAi opened up new possibilities for large-scale screening in a wide range of organisms and human cells, RNAi based screens have suffered from high amounts of false positives and false negatives, as well as from poor reproducibility. Furthermore, incomplete knock-down of the targeted mRNA and unintended off-target effects are problematic, resulting in the unspecific silencing of non-target mRNAs (Kaelin, 2012). Intervention at the DNA level by targeted genome

engineering techniques (see below) offers a way around the drawbacks of RNAi based methods.

1.2.4. Targeted genome engineering

The advent of tools to introduce specific modifications into DNA has enabled the engineering of cells, tissues and organisms for the reverse genetic investigation of gene function. Conventional gene targeting methods rely on the removal or replacement of specific genes by homologous recombination, whereas newer technologies exploit programmable nucleases for the targeted engineering of genomes.

1.2.4.1. Classical homologous recombination techniques

The first site-specific genetic manipulation tools relied on homologous recombination between a targeting vector and an endogenous gene of interest. Vectors containing large homology regions to a particular gene and a selectable marker allow the removal or replacement of a gene at its normal chromosomal location, resulting in selective gene disruption. Homologous recombination based techniques can be also used to tag endogenous genes, generate fusion proteins or introduce specific mutations (Capecchi, 2005). Disrupting gene function by homologous recombination has been widely used and has proven particularly powerful in yeast and mouse model systems.

Compared to other model organisms, the targeted genetic manipulation of *Saccharomyces cerevisiae* is relatively straightforward due to its highly efficient homologous recombination machinery and the fact that only a single allele needs to be disrupted if the yeast is used in its haploid state (Forsburg, 2001). As a consequence, *S. cerevisiae* was the first eukaryote for which precise gene knock-outs could be generated and subsequently, deletion collections targeting every open reading frame of non-essential genes were constructed (Winzeler et al., 1999;

Forsburg, 2001). Temperature-sensitive mutants have been created to investigate the effect of essential genes under permissive/non-permissive temperature conditions (Ben-Aroya et al., 2008; Li et al., 2011). Such collections have allowed the functional annotation of genes and have provided insight in genetic and gene-environment interactions (Giaever et al., 2002; Hillenmeyer et al., 2008; Costanzo et al., 2010; Giaever and Nislow, 2014; Costanzo et al., 2016).

Targeted genetic alterations by homologous recombination became possible in mice due to the development of mouse ES cell technology (Evans and Kaufman, 1981; Martin, 1981). Using ES cells instead of whole organisms, together with the targeted introduction of selection markers, allowed the generation of modified ES cells. Manipulated ES cells can be subsequently injected into wild-type blastocysts to obtain germline chimeric animals used for the creation of transgenic mice (Capecchi, 2005). This technology made it possible to selectively knock-out any gene of interest and evaluate its role in mouse development and physiology (Austin et al., 2004; International Mouse Knockout Consortium et al., 2007; White et al., 2013; Brown and Moore, 2012).

Since then, more versatile second generation gene targeting systems have been developed to create point mutations, deletions, inversions, duplications and translocations, for instance, the Cre/loxP technology. Cre is a recombinase enzyme from the P1 bacteriophage that catalyses the recombination between loxP sites. The crossing of transgenic Cre recombinase-expressing mouse lines with lines containing loxP sites flanking a particular gene of interest allows Cre mediated excision or inversion of the “floxed” target gene (Lewandoski, 2001). Several Cre transgenic lines have been established to enable the efficient gene modification in various cell types and developmental stages. Large collections of mutant mouse lines and conditional knock-out mice for the analysis of gene function have been generated using this technique (International Mouse Knockout Consortium et al., 2007; White et al., 2013; Kersten et al., 2016).

Despite the tremendous success of the described methods for studying gene function, drawbacks do still exist: One limitation is that higher organisms do not have an efficient homologous recombination machinery, making technologies that rely on it time consuming, labour-intense and expensive. In addition, Cre expression in conditional mice can occur in unintended tissues or developmental stages, or Cre mediated excision can be incomplete, thereby confounding the interpretation of phenotypes. More recently, targeted genetic manipulation techniques, as described in the section below, allow to overcome hurdles associated with homologous recombination based approaches.

1.2.4.2. Programmable nucleases

Genome engineering methods such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats/CRISPR-associated (CRISPR/Cas) developed over the past decade have dramatically advanced reverse genetic applications in model organisms and human cells. The key advantage of programmable nucleases is that they induce site-specific double strand breaks (DSBs) that can be repaired by non-homologous end joining (NHEJ) or the homology-directed repair (HDR) pathway. DSB repair via the NHEJ pathway is error-prone and frequently leads to the generation of small insertions and deletions (indels) that result in frameshift mutations and gene disruption if located in the coding region of a gene, thereby creating a functional gene knock-out. Repair of DSB lesions via the HDR pathway can be exploited to generate targeted knock-ins if an appropriate repair template with homology arms for the desired integration site is provided. Such insertions may be employed to introduce or correct specific mutations or for the insertion of a reporter sequence. Programmable nucleases can be also used to generate DSBs at more than one site at a time, thereby triggering various chromosomal rearrangements including large deletions, inversions or translocations (Kim and Kim, 2014; Boettcher and McManus, 2015).

ZFNs and TALENs have a modular structure that comprises a sequence-specific DNA binding and a non-specific nuclease domain (Figure 1.4). While ZFNs use zinc-finger domains, TALENs use transcription activator-like effector (TALE) domains for the site-specific binding of DNA. The DNA-binding modules are fused to the endonuclease domain of the bacterial restriction enzyme FokI to form chimeric proteins that enable the targeted generation of DSBs. For the effective DNA cleavage by the FokI endonuclease, ZFNs and TALEN must dimerise. While ZFNs bind nucleotide triplets, TALENs bind single nucleotides (Urnov et al., 2010; Bogdanove and Voytas, 2011). Various model organisms including nematodes, flies, fish and mice have been modified using ZFNs and TALENs (Bibikova et al., 2002; Morton et al., 2006; Meng et al., 2008; Doyon et al., 2008; Cui et al., 2011; Wood et al., 2011; Bedell et al., 2012; Zu et al., 2013). However, disadvantages of these programmable nucleases include a limited targeting range and that they require the time-consuming design and generation of customized fusion proteins (Kim and Kim, 2014). The CRISPR/Cas9 gene editing tool, described briefly below and in more detail in section 2.1.1., has allowed to overcome the hurdles associated with ZFNs and TALENs.

CRISPR/Cas systems are RNA based microbial defence mechanisms that recognise and cleave DNA from invading bacteriophages and plasmids (Horvath and Barrangou, 2010; Bhaya et al., 2011). Repurposing of the CRISPR/Cas9 system for the targeted modification of eukaryotic genomes has allowed editing of a wide variety of organisms and human cells with hitherto unseen efficiency and precision. The Cas9 endonuclease is directed to a target sequence through an associated short guide RNA (gRNA) by base-pairing, and Cas9 mediated cleavage occurs in close proximity to the protospacer adjacent motif (PAM) leading to the targeted generation of DSBs (Figure 1.5) (Wright et al., 2016).

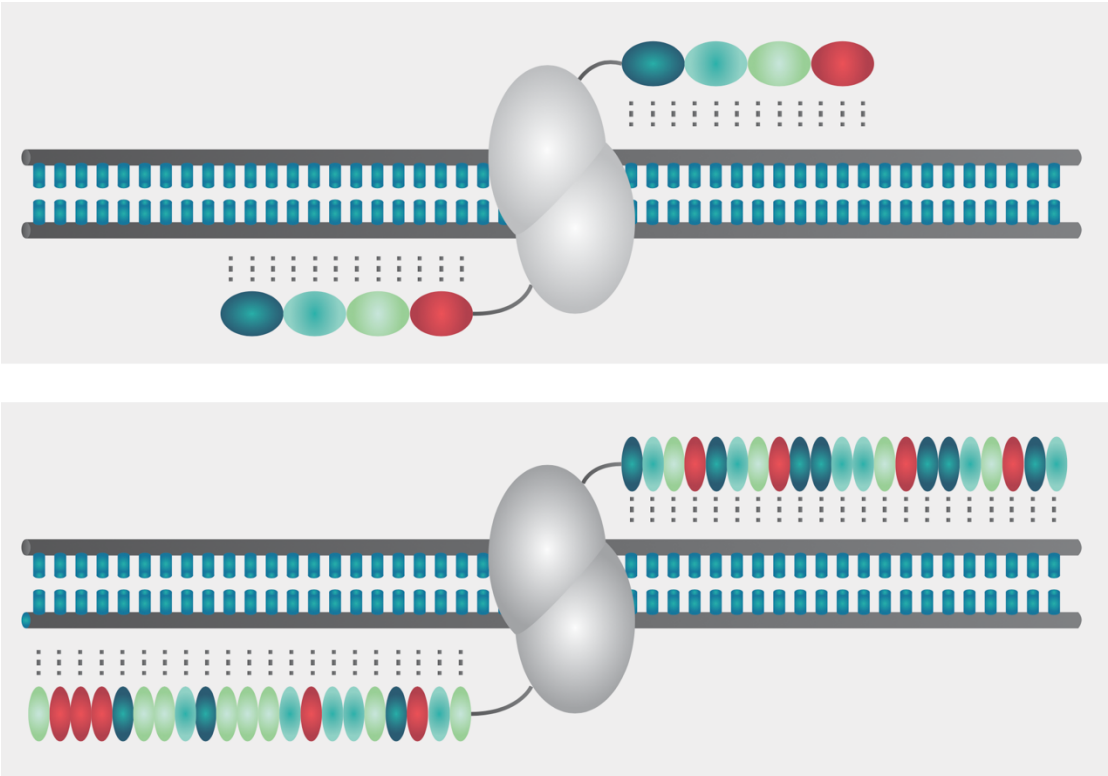


Figure 1.4 Structure of ZFNs and TALENs. ZFNs (upper panel) and TALENs (lower panel) consist of a sequence-specific DNA-binding and a non-specific nuclease domain that introduces DSBs. The DNA binding domains of ZFNs are composed of zinc-fingers, the binding domains of TALENs comprise transcription activator-like effectors (TALEs). While 3-6 zinc-fingers make up a ZFN subunit that recognises a 9-18 bp DNA stretch, TALENs are composed of 33-35 amino acid repeats typically recognising 30-40 bp sequences. The DNA-binding modules are fused to the FokI endonuclease to form chimeric proteins that dimerise and enable the targeted generation of DSBs. Adapted from Kim and Kim, 2014 and Haimovich et al., 2015.

The CRISPR/Cas9 system has been used for a wide variety of applications from targeted gene editing to large-scale forward genetic screens to correcting genetic diseases in mice and cultured human cells (Schwank et al., 2013; Wu et al., 2013; Wang et al., 2014; Shalem et al., 2014; Koike-Yusa et al., 2014; Zhou et al., 2014; Gilbert et al., 2014; Yin et al., 2014; Konermann et al., 2015; Long et al., 2016;

Nelson et al., 2016; Tabebordbar et al., 2016; Yang et al., 2016). Modifications and further applications of this system are detailed in section 2.1.1.

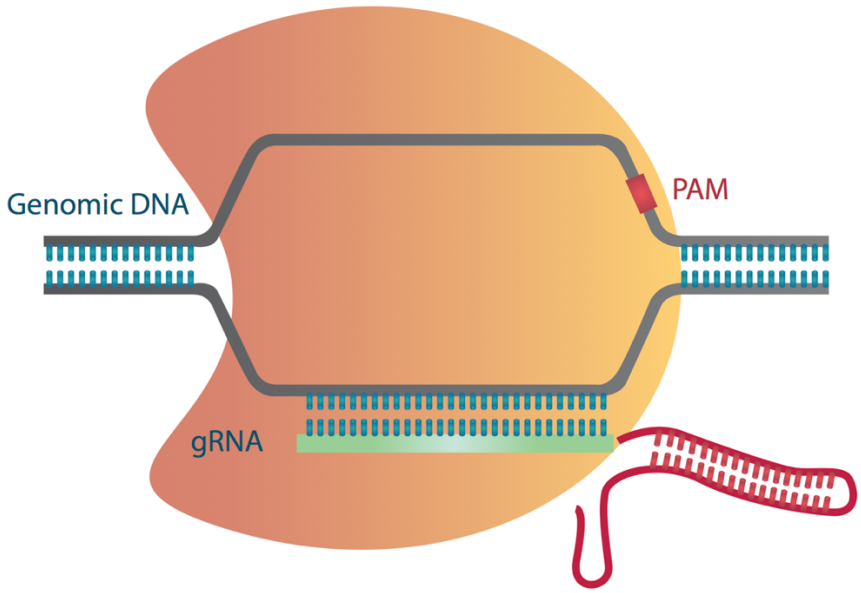


Figure 1.5 The CRISPR/Cas9 system. The Cas9 endonuclease is directed to a specific genetic locus through an associated 20 nucleotide guide RNA (gRNA). The presence of a protospacer adjacent motif (PAM) results in the site-specific cleavage of DNA and the generation of a DSB. Adapted from Haimovich et al., 2015.

In contrast to ZFNs and TALENs, the CRISPR/Cas9 system only requires the design and generation of a 20 nucleotide gRNA to achieve target specificity, which is much cheaper and easier by comparison. Another advantage is that nearly every sequence in a genome can be targeted (Kim and Kim, 2014). Because of the versatility and precision of its application, the CRISPR/Cas9 system is the current method of choice for targeted genome engineering.

Like the other programmable nucleases, the CRISPR/Cas9 system also suffers from unintended off-target alterations introduced at loci that display homology to the target sites (Fu et al., 2013; Hsu et al., 2013). Modifications have been made in order to reduce off-target effects, for example by using truncated gRNAs (Fu et al., 2014),

"paired nickases" (Ran et al., 2013a; Cho et al., 2014), or improved computational methods to predict potential off-target sites more accurately (Hsu et al., 2013; Doench et al., 2016).

1.3. Applications of genetic screens

Genetic screens in model organisms and cultured cells have provided fundamental insights in various biological processes. Applications and biological findings arising from genetic screening in yeast, flies, worms, fish, mice and cultured cells are summarised in the sections below.

1.3.1. Genetic screening in model organisms

Forward and reverse genetic screening in model systems has identified genes underpinning a wide range of biological processes. One of the major eukaryotic model organisms for linking genotype to phenotype via genetic screening is the budding yeast *Saccharomyces cerevisiae*. The key advantage of this organism is that it has a haploid and diploid life cycle. Using haploid cells in combination with insertional mutagenesis or targeted genetic manipulation approaches circumvents difficulties associated with genetic screens in diploid organisms, i.e. the necessity of inactivating both alleles of a gene to obtain a phenotype (Forsburg, 2001). Combined with the fact that *S. cerevisiae* displays high rates of homologous recombination, this made the budding yeast the first eukaryote for which precise gene knock-outs were constructed (Rothstein, 1983). As mentioned previously, deletion collections targeting every open reading frame of non-essential genes and temperature-sensitive mutant collections for essential genes (Winzeler et al., 1999; Ross-Macdonald et al., 1999; Ben-Aroya et al., 2008; Li et al., 2011) have been used to study gene function under various environmental conditions and to elucidate the MoA of small molecules (Giaever et al., 2002, Hillenmeyer et al., 2008; Ben-Aroya et al., 2008; Li et al., 2011;

Nijman, 2015). Screening of double-deletion mutants provided insight into global genetic interaction networks and the characterisation of the basic wiring of yeast cells (Costanzo et al., 2010; Costanzo et al., 2016).

Genetic screening in the fruit fly *Drosophila melanogaster* has been mainly focused on developmental processes, since many mutations result in a phenotype that can be easily observed, such as defective eye, wing or limb development (St Johnston, 2002). For instance, chemical mutagenesis screens have allowed to identify genes underpinning embryonic patterning (Nüsslein-Volhard and Wieschaus, 1980) or differentiation (Simon et al., 1991; Rogge et al., 1991). Transposon mediated insertional mutagenesis and the subsequent derivation of individuals harbouring insertions at defined genomic locations, as well as the use of targeted recombination based systems and RNAi, have been used to study the function of genes in flies in more detail (Adams and Sekelsky, 2002; Boutros and Ahringer, 2008). Forward and reverse genetic approaches in *D. melanogaster* have shed light on various signal transduction pathways including Wnt, Hedgehog or Notch signalling (Pires-daSilva and Sommer, 2003). Many of the pathways characterised in *Drosophila* were later found to be important in human biology and pathology, ranging from metabolic disorders to cancer and metastasis (St Johnston, 2002; Pandey and Nichols, 2011).

The nematode *Caenorhabditis elegans* is another popular model organism for the investigation of developmental processes. The advantage of this model system is that the fate of all cells during development has been mapped (Sulston and Horvitz, 1977; Sulston et al., 1983). Thus, genetic manipulation allows to interrogate the role of particular genes by tracking the affected cells in the developmental process. For example, forward genetic screens in *C. elegans* have provided insight into Notch and RAS signalling, apoptosis and the MoA of several drugs (Jorgensen and Mango, 2002; Kaletta and Hengartner, 2006). The development of target-selected mutagenesis approaches and several methods for the effective delivery of dsRNA has allowed reverse genetic investigations in *C. elegans* (Fire et al., 1998; Jansen et

al., 1997; Liu et al., 1999). Studying the phenotypic consequences of specific genetic alterations has identified genes involved in embryonic development, cell signalling or pathologies relevant in the human system, such as neurodegeneration or genetic disease (Kaletta and Hengartner, 2006).

As for *D. melanogaster* and *C. elegans*, genetic screens in the zebrafish *Danio rerio* have been mainly leveraged to study developmental processes. Non-targeted mutagenesis screens and target-selected mutagenesis have provided insight into organogenesis or embryonic patterning, uncovered genes essential for embryonic development and have allowed the assembly of extensive mutant collections (Haffter et al., 1996; Driever et al., 1996; Gaiano et al., 1996; Amsterdam et al., 1999; Golling et al., 2002; Wienholds et al., 2002). Morpholino technology has been applied to study the function of genes during development and to model human genetic disease (Doitsidou et al., 2002; North et al., 2009; Neugebauer et al., 2009). In addition, small molecule screens have been carried out to repurpose existing drugs for novel applications or to identify the MoA of poorly characterised compounds (North et al., 2007; Yu et al., 2008; Yeh et al., 2009; Rihel et al., 2010).

Compared to other model organisms, mice (*Mus musculus*) are phylogenetically more similar to humans and have hence been a popular system for biomedical research. A major breakthrough for targeted genetic manipulation was the development of homologous recombination based methods in ES cells (Evans and Kaufman, 1981; Martin, 1981) used for the assembly of large collections of mutant mouse lines and conditional knock-out mice as described before (Austin et al., 2004; Auwerx et al., 2004; International Mouse Knockout Consortium et al., 2007; Gondo, 2008; Brown and Moore, 2012). These efforts have enabled the analysis of gene function in mouse development and physiology, and have provided insight into various human diseases (White et al., 2013; Gondo, 2008; Brown and Moore, 2012; Kersten et al., 2016). The use of programmable nucleases for genome engineering has accelerated and facilitated the modelling of different diseases such as cancer

(Carbery et al., 2010; Cui et al., 2011; Wang et al., 2013; Heckl et al., 2014; Xue et al., 2014; Maddalo et al., 2014; Zuckermann et al., 2015; Zhang et al., 2015). Traditional forward genetic and modifier screens have identified sets of genes underlying viability and human disease, and have suggested novel therapeutic targets (Hrabe de Angelis et al., 2000; Nolan et al., 2000; Hoebe et al., 2003; Carpinelli et al., 2004; Buchovecky et al., 2013).

These examples illustrate that genetic screens in model organisms can generate valuable insights, despite the advances of screening in a human cell system, as described below.

1.3.2. Genetic screening *in vitro*

The fact that cells can be cultured under well-controlled conditions *in vitro* together with their fast growth rate and easy genetic manipulation have allowed cultured cells to become the most widely used biological system for carrying out genetic screens. In contrast to genetic screening using whole organisms that are preferred for studying developmental processes or behaviour, *in vitro* screens allow a “reductionist approach” that can be used to investigate cellular processes in detail, thereby providing unique insight into molecular mechanisms.

RNAi based screens in cultured *Drosophila* and mouse cells have suggested genes involved in host-pathogen interactions or cancer biology. The use of RNAi in human cells has, for example, helped to gain insight into diverse biological processes, ranging from cell division to pathogen responses to cancer development (Boutros and Ahringer, 2008; Mohr et al., 2010). Despite these successes, the incomplete and non-specific knock-down of gene expression has limited the utility of RNAi based approaches for forward and reverse genetic approaches in cultured cells (Kaelin, 2012).

As pointed out previously, insertional mutagenesis employed in near-haploid cells and CRISPR/Cas9 based screens provided a way around these limitations. Insertional mutagenesis approaches in haploid mammalian cells have identified genes involved in host-pathogen interactions or the action of toxins and compounds (Carette et al., 2009; Elling et al., 2011; Carette et al., 2011a; Carette et al., 2011b; Birsoy et al., 2013; Jae et al., 2013; Jacobson et al., 2013; Elling and Penninger, 2014; Jae et al., 2014; Winter et al., 2014; Blomen et al., 2015; Wang et al., 2015; Pillay et al., 2016; Staring et al., 2017). More recently, large-scale forward genetic screens using CRISPR/Cas9 have uncovered essential genes and genes involved in drug resistance (Shalem et al., 2014; Wang et al., 2014; Zhou et al., 2014). The use of gene trapping and CRISPR/Cas9 in human cells is further detailed in sections 2.1.1. and 2.1.2.

While forward genetic investigations in cultured cells have revealed genes underlying diverse biological processes in exquisite detail, complementary reverse genetic approaches have been suffering from difficulties associated with the generation of large sets of targeted mutants. Now, the power of targeted genome engineering using CRISPR/Cas9 in human cell systems for the first time allows systematic and unbiased reverse genetic approaches. A human knock-out collection generated with CRISPR/Cas9 technology is used in chapter 3 to set up and validate a functional genetic screening tool for large-scale reverse genetic investigations in a human cell system.

1.4. Objectives of this thesis

Aim 1: Establishment of a functional gene annotation tool in human cells

Forward genetic screening has been employed extensively in human cells using RNAi, gene trap, and CRISPR approaches. In contrast, reverse genetic approaches have been limited to arrayed RNAi screens and typically interrogating only a single

phenotype, such as cell viability or changes in a particular signal transduction pathway. Thus, deep phenotyping of gene mutants has been largely restricted to model organisms. This impedes fundamental human biology research and hinders the understanding of diseases. In addition, mutations can be species specific and cannot be modelled in other organisms. There is thus an urgent need for a general, scalable and easily accessible method for reverse genetics in human cells.

Chapter 2 and 3 of this thesis aim to establish a large-scale reverse genetic screening platform and provide a proof-of-concept study for a scalable and generic approach to quantify phenotypes in human cells. Chapter 2 describes the set-up of a phenotypic profiling platform based on RNA-sequencing. For the implementation of this transcriptome based screening method, transcriptional profiles of haploid human cells in response to a diverse range of external stimuli are collected. Chapter 3 employs the established phenotypic profiling platform for systematic and unbiased functional gene annotation in human cells. Parallel reverse genetic screens with ten different stimuli are conducted to functionally annotate 55 non-essential tyrosine kinases in signalling pathways via transcriptional signature comparisons of engineered haploid isogenic cell lines. By comparing transcriptional signatures of stimulated wild-type cells to isogenic kinase knock-out cell lines, the implication of particular kinases in specific pathways is demonstrated.

Together, chapter 2 and 3 demonstrate that multiplexed RNA-sequencing is a robust and sensitive method to study cellular perturbations and describe a scalable method for large-scale reverse genetic screening in human cells.

Aim 2: Set-up of a pooled drug sensitivity and compound MoA platform in human cells

Variable patient response to treatment is a prevalent concern in the application of drugs to treat human diseases. Better understanding of gene-drug interactions and compound MoA are essential to increase the knowledge of how compounds affect

cellular physiology and to ultimately fine-tune clinical treatment. Chemical screening approaches provide a strategy to identify gene-drug interactions and MoA of compounds by comparing responses of different mutant cell lines to treatments.

Chapter 4 details the establishment of a pooled chemical genetic approach to investigate thousands of potential gene-drug interactions in an isogenic human cell line model. The aim of this chapter is to identify genes underlying drug sensitivity by monitoring the viability of 4958 isogenic cell lines upon drug application. Another goal is to generate different response profiles for chemical agents with different MoAs, which could allow to elucidate unknown MoAs by comparing the profiles of compounds with unknown mechanisms to profiles of drugs with established MoA. This chapter highlights two pilot experiments to (i) assess the sensitivity for detecting the drop-out of single clones from a complex pool, which is a technical requirement for this approach to work, and (ii) evaluate the consequences of drug application on mutant cell viability. In summary, chapter 4 presents a pooled screening strategy that is sensitive enough to detect individual gene-drug interactions and that can be potentially exploited for MoA studies in human cells.

Chapter 2. A platform for large-scale cell profiling by multiplexed RNA-sequencing

2.1. Introduction

This chapter provides an overview of current cutting-edge technologies such as the CRISPR/Cas9 system, haploid genetic approaches in human cells and technological advances in massively parallel sequencing. The results part of this chapter describes an application combining all these techniques and details the establishment of a phenotypic profiling method based on multiplexed RNA-sequencing that is robust and scalable.

2.1.1. Genome editing using CRISPR/Cas9

The rise of CRISPR/Cas for the targeted engineering of genomes has revolutionised applications in model organisms and human cells. CRISPR loci and Cas genes are components of the adaptive immune system of some archaea and bacteria orchestrating the response to foreign genetic material (Horvath and Barrangou, 2010; Bhaya et al., 2011). Invading DNA from viruses or plasmids is fragmented and small pieces (spacers) are incorporated into short repeat sequences within the CRISPR locus. Transcription of this repeat-spacer array and subsequent RNA processing result in the generation of small CRISPR RNAs (crRNAs) that can bind and guide Cas family endonucleases to a complementary target sequence. New infections are thus defended through the targeted degradation of foreign DNA mediated by the crRNA/Cas complex (Wright et al., 2016).

The CRISPR locus was first described in *E. coli* in 1987 (Ishino et al., 1987) and its function in adaptive immunity was determined 20 years later by showing that *S. thermophilus* can become resistant to a bacteriophage by integrating a part of the phage genome into its CRISPR locus (Barrangou et al., 2007). Of the six CRISPR systems described so far, type I-III are best studied. In contrast to the type I and III systems that rely on several Cas effector endonucleases for the pre-processing of crRNAs and cutting of foreign DNA, the type II CRISPR system requires only the

Cas9 protein which can catalyse both steps. This has rendered the type II system the most widely used CRISPR/Cas system in genome engineering applications. In order to effectively target and cleave DNA, Cas9 requires the crRNA together with a second trans-activating crRNA (tracrRNA or trRNA) that is partially complementary to the crRNA. Furthermore, Cas9 requires a PAM motif located 3' of the crRNA complementary sequence. Cas9 exerts its endonuclease function in close proximity to the PAM sequence, resulting in the generation of DSBs (Makarova et al., 2011; Wright et al., 2016).

Further simplification of the type II system of *S. pyogenes* was achieved by combining the crRNA and trRNA to a single guide RNA (gRNA), which turned out to be as effective as the two separate RNAs (Jinek et al., 2012). A series of publications in 2013 has demonstrated that the introduction of the prokaryotic CRISPR/Cas9 system into eukaryotic cells also results in the generation of targeted DSBs and site-specific genome editing (Cong et al., 2013; Mali et al., 2013a; Cho et al., 2013; Hwang et al., 2013; Jinek et al., 2013; DiCarlo et al., 2013; Yang et al., 2013; Wang et al., 2013; Gratz et al., 2013; Friedland et al., 2013).

As outlined previously, DSBs can be repaired by two distinct repair pathways in eukaryotes: NHEJ and HDR (Figure 2.1). Repair of DSBs by NHEJ is error-prone and can lead to the introduction of small indels, thereby disrupting the targeted locus. Indels in coding regions resulting in a frameshift and a premature stop codon can hence lead to gene inactivation. On the other hand, DSBs can be accurately repaired by HDR, given that an appropriate template is provided (e.g. a sister chromatid). This pathway can be exploited to introduce specific mutations or knock-ins at a particular locus by supplying a repair template (Kim and Kim, 2014).

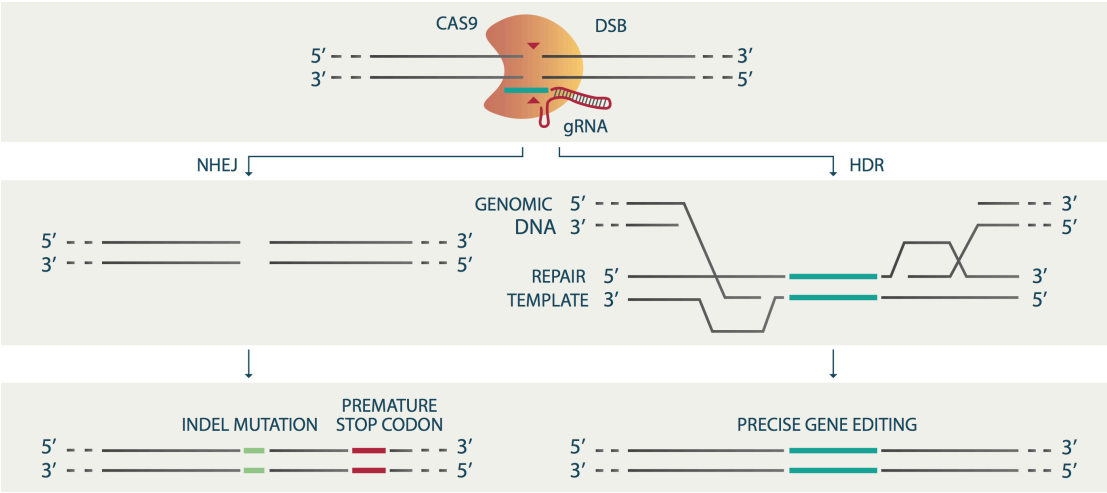


Figure 2.1 Gene editing of eukaryotic cells using the CRISPR/Cas9 system. The Cas9 effector endonuclease is targeted to a locus of interest through an associated 20 nt gRNA. The presence of a PAM sequence allows the sequence-specific cleavage of the targeted DNA. The resulting DSBs can be repaired by NHEJ or HDR. The processing and re-joining of DSBs by the NHEJ repair machinery occurs with low fidelity and frequently leads to the formation of indels. If present in a coding region of the targeted gene, the introduced indels can result in a premature stop codon and hence, a non-functional protein. The NHEJ pathway can be exploited for the generation of gene knock-outs. Repair via the HDR pathway relies on DSB repair via homologous recombination which is more precise due to the usage of an exact copy as a template (e.g. a sister chromatid). This pathway can be used to introduce specific mutations or knock-ins if a homologous repair template is provided. Adapted from Ran et al., 2013b.

The CRISPR/Cas9 system has been further modified to create a mutant Cas9 version that cleaves only one DNA strand (“nickase”), thereby forcing cells to use the HDR machinery, which results in decreased indel formation (Cong et al., 2013). Furthermore, nuclease-deficient Cas9 (dCas9) proteins (Qi et al., 2013) have been fused to a variety of effector domains including VP64, KRAB or eGFP (Perez-Pinera et al., 2013; Maeder et al., 2013; Mali et al., 2013b; Gilbert et al., 2013; Chen et al., 2013). These can be used to activate, repress or visualise genes in a highly specific manner. More recently, the CRISPR/Cas9 system has been adapted for optogenetic

or chemical control to tightly regulate Cas9 activity through light or exogenous molecules (Nihongaki et al., 2015a; Nihongaki et al., 2015b; Liu et al., 2016; Maji et al., 2017). The unprecedented ease, precision and versatility of the CRISPR/Cas9 system have solved many problems associated with other genome editing methods (e.g. ZFN, TALEN) and RNAi (see section 1.2.3. and 1.2.4.2.) (Kaelin, 2012; Boettcher and McManus, 2015). However, the CRISPR/Cas9 technology also suffers from unwanted off-target effects, albeit to a lesser extent (Fu et al., 2013; Hsu et al., 2013). CRISPR/Cas9 can tolerate up to five mismatches between gRNA and target locus but still lead to DSBs if a PAM sequence is present. Substantial effort has been made to modify the CRISPR/Cas9 system to have fewer off-target sites, e.g. by using truncated gRNAs (Fu et al., 2014), “paired nickases” (Ran et al., 2013a; Cho et al., 2014) or improved computational methods to predict potential off-target sites more accurately (Hsu et al., 2013; Doench et al., 2016).

To date, the CRISPR/Cas9 system has been utilised for modifying the genome of a vast variety of human cells and model organisms (Cong et al., 2013; Mali et al., 2013a; Cho et al., 2013; Hwang et al., 2013; Jinek et al., 2013; DiCarlo et al., 2013; Yang et al., 2013; Wang et al., 2013; Gratz et al., 2013; Friedland et al., 2013; Niu et al., 2014). In only four years since its first application for genome engineering, the CRISPR/Cas9 system has already enabled the generation of large mutant cell line collections, the introduction of specific point mutations and tags into the genome, or large genomic deletions and rearrangements by using two gRNAs (Sander and Joung, 2014; Essletzbichler et al., 2014; Lackner et al., 2015; Gapp et al., 2016). Furthermore, the assembly of genome-wide gRNA libraries has allowed to perform whole genome knock-out screens, as well as screening with dCas9-VP64 or dCas9-KRAB fused libraries (Wang et al., 2014; Shalem et al., 2014; Koike-Yusa et al., 2014; Zhou et al., 2014; Gilbert et al., 2014; Konermann et al., 2015). The CRISPR/Cas9 system has also been leveraged to correct genetic diseases such as cystic fibrosis, dominant cataract disorder, Duchenne muscular dystrophy or rare liver

disorders in postnatal mouse models and/or cultured human cells (Schwank et al., 2013; Wu et al., 2013; Yin et al., 2014; Long et al., 2016; Nelson et al., 2016; Tabebordbar et al., 2016, Yang et al., 2016). Other studies showed the potential application of eradicating latent HIV-1 copies in human CD4+ T cell genomes (Kaminski et al, 2016).

The adaptation of the CRISPR/Cas system for targeted genome engineering has thus set the stage for a new era in molecular biology and has already revolutionised many aspects of basic and translational research as well as industry. Further improvements on existing tools and the development of novel applications will continue to expand the gene perturbation toolbox for investigators.

2.1.2. Haploid genetics in human cells

A fundamental drawback of functional genomic studies in diploid organisms and human cells is that the disruption of one allele is often not sufficient to result in an observable phenotype. This is further complicated by the fact that many human cell lines are aneuploid and harbour several copies of genes and chromosomes. The targeted editing of multiple alleles at the same time is technically challenging even for the highly efficient CRISPR/Cas9 system (Barrangou et al., 2015).

The discovery that some tumour cells display a near-haploid karyotype (Kessous et al., 1980; Mayne and Maher, 1989; Drouin et al., 1993; Safavi et al., 2013) and the successful derivation of cell lines thereof have allowed to overcome technical difficulties associated with the editing of aneuploid genomes and have paved the way for haploid genetics in human cells. For instance, the KBM7 cell line was derived from a chronic myeloid leukaemia (CML) patient where an isolated myeloid subpopulation displayed a karyotype haploid for every chromosome except for chromosome 8 and a part of chromosome 15 (Andersson et al., 1995; Kotecki et al., 1999) (Figure 2.2). KBM7 cells also harbour a reciprocal translocation of a part of chromosome 9 with 22 (“Philadelphia chromosome”) by which the ABL kinase gene

is fused to the BCR gene and becomes constitutively active. The resulting fusion protein interferes with several cellular processes such as proliferation, survival, differentiation or adhesion and contributes to the malignant transformation in CML. KBM7 cells are highly dependent on the fusion protein produced by the BCR-ABL translocation and undergo apoptosis upon application of a BCR-ABL inhibitor (Scappini et al., 2004).

A re-programming attempt of near-haploid KBM7 cells yielded a fibroblast-like cell line termed HAP1 as a by-product (Carette et al., 2010). In contrast to KBM7 cells, HAP1 cells have lost the second copy of chromosome 8 and the expression and dependency on the BCR-ABL signalling cascade. More recently, CRISPR/Cas9 was used to delete the heterozygous fragment of chromosome 15 integrated on chromosome 19 to generate the first fully haploid human cell line (eHAP) (Essletzbichler et al., 2014).

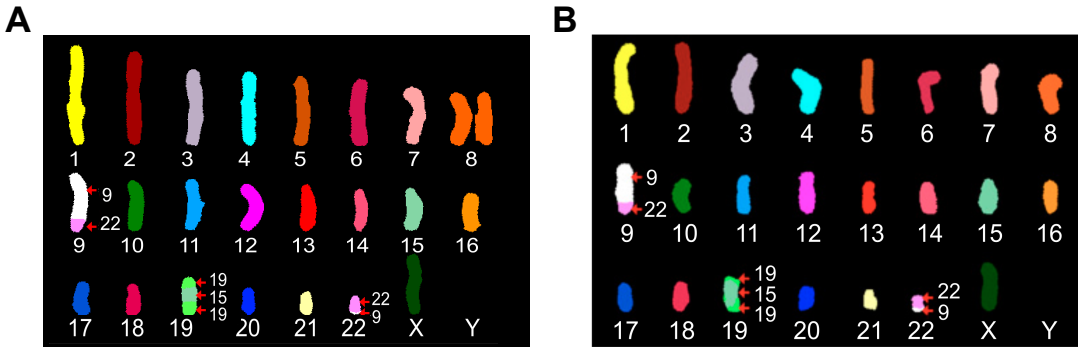


Figure 2.2 Karyotype of the human KBM7 and HAP1 cell lines. A. Spectral karyotype of KBM7 cells showing the presence of one copy of each chromosome, with the exception of chromosome 8 and a part of chromosome 15. The integration of the diploid part of chromosome 15 into chromosome 19 and the reciprocal translocation between chromosome 9 and 22 can be observed. B. Spectral karyotype of HAP1 cells. As in A, except for only one copy of chromosome 8. The KBM7 karyotype picture was kindly provided by Horizon Discovery Ltd, the HAP1 karyotype picture was taken from Essletzbichler et al., 2014.

Near-haploid cells in combination with non-targeted mutagenesis approaches or CRISPR/Cas9 provide a powerful tool for forward genetic screening in human cells. This is because only a single allele needs to be disrupted in order to result in a full knock-out and hence a loss-of-function phenotype. Due to these benefits, many studies have utilised KBM7 and HAP1 cells in combination with a retroviral gene trap vector encoding for a fluorescent reporter, a splice acceptor site and a polyadenylation signal (Carette et al., 2009; Carette et al., 2011a; Carette et al., 2011b; Jae et al., 2013; Jacobson et al., 2013; Jae et al., 2014; Winter et al., 2014; Blomen et al., 2015; Wang et al., 2015; Pillay et al., 2016; Staring et al., 2017). Upon infection and integration of the gene trap cassette into the genome of near-haploid cells, genes are aberrantly spliced and transcription is prematurely aborted. As detailed in chapter 1, retroviral gene trap vectors can be inserted in two ways: sense or antisense. If integrated in the sense orientation, gene disruption can be achieved via the insertion in intronic and exonic regions due to the activity of the splice acceptor and polyadenylation signal. Insertions in the antisense orientation result in a loss-of-function mutant only when integrated in the coding region of a gene, thereby effectively disrupting the reading frame (Stanford et al., 2001). The quantification of sense and antisense insertions in a mutagenised population under standard growth conditions or upon perturbation has allowed the investigation of a multitude of biological questions. For instance, positive selection screens have been employed to determine the MoA of compounds or biological agents as well as to investigate host-pathogen interactions (Carette et al., 2009; Carette et al., 2011a; Carette et al., 2011b; Jae et al., 2013; Jacobson et al., 2013; Jae et al., 2014; Winter et al., 2014; Pillay et al., 2016; Staring et al., 2017). Negative selection screens were carried out to uncover genes essential for cell growth and for the identification of synthetic lethal interactions (Blomen et al., 2015; Wang et al., 2015). More recently, haploid human cells have been combined with CRISPR/Cas9 technology (section 2.1.1.) and screens have been performed to give insight into gene essentiality or endoplasmic

reticulum-associated degradation (ERAD) (Wang et al., 2015; Timms et al., 2016). Another study reported the derivation of haploid human embryonic stem cells from oocytes (Sagi et al., 2016). The authors furthermore succeeded in differentiating these cells into neurons, cardiomyocytes or pancreatic cells while retaining their fully haploid state, and demonstrated their use for forward genetic screening. This potentially offers an exciting new avenue for forward genetics in different haploid cell types.

Additionally, gene trapping and targeted gene editing were used to assemble large collections of isogenic KBM7 and HAP1 knock-out cells (Bürckstümmer et al., 2013; Gapp et al., 2016). These collections for the first time allow large-scale reverse genetic approaches, so far limited to model organisms. Chapter 3 of this thesis describes the first proof-of-concept study for transcriptome based reverse genetic screening in haploid human cells.

In summary, the derivation of haploid human cells combined with insertional mutagenesis or CRISPR/Cas9 technology has enabled unprecedented possibilities for human cell genetics.

2.1.3. Technological advances in massively parallel sequencing

Tremendous technological progress over the past ten years has allowed massively parallel sequencing to become a routine technique in virtually every field of biology. Massively parallel sequencing approaches allow the cataloguing of genomes, transcriptomes or epigenomes and the investigation of genomic features that underlie complex phenotypes and human disease. Three major advances have fuelled this variety of sequencing applications: i) the development of library preparation protocols that do not require a cloning step, ii) the massive parallelisation of sequencing reactions and iii) the direct detection of sequenced reads, discussed in more detail below (Van Dijk et al., 2014).

Current DNA and RNA-sequencing platforms require the standardised preparation of samples. This process is termed “library preparation”. Such library preparation protocols usually involve DNA fragmentation, reverse transcription for RNA samples, the ligation of specific adapters and a PCR amplification step. The final libraries are blunt-ended DNA molecules containing the DNA template of interest flanked by adapter sequences that are used by the sequencing platform. The development and continuous improvement of such library preparation protocols has significantly improved the yield and throughput of the sample preparation for massively parallel sequencing. This was achieved by reducing the number of steps in the library preparation process (e.g. Illumina’s Nextera technology allows DNA fragmentation, fragment end-polishing and adapter ligation in a single step), improving the purification on beads and the use of polymerases that are less prone to biased amplification during the PCR step (Van Dijk et al., 2014; Goodwin et al., 2016). Further improvements in whole genome amplification and miniaturisation of reaction volumes now allow sequencing from single cells (Gawad et al., 2016).

For performing the actual sequencing reaction, a variety of systems have been developed. The choice of the appropriate sequencing platform depends on the application, e.g. PacBio generates very long reads and is ideal for *de novo* genome assembly, while Illumina generates short reads at a high throughput and is, among other applications, frequently used for expression profiling. Furthermore, some platforms (e.g. MinION by Oxford Nanopore Technologies) are able to sequence single DNA molecules, while others (e.g. Illumina) are not sensitive enough and require the amplification of the initial templates on beads or a solid surface (Figure 2.3). This amplification step guarantees a favourable ratio of signal over noise and allows, together with improved sequencing chemistries, base-calling software and throughput in data handling, a massive parallelisation of short-read sequencing. In this way, millions of sequencing reactions can be run and processed simultaneously. Further improvements in flow cell architecture, polymerase efficiency and camera

speed have resulted in an unprecedented throughput and the lowest cost per base sequenced so far, allowing the sequencing of a human genome for less than 1000 USD (Van Dijk et al., 2014; Goodwin et al., 2016).

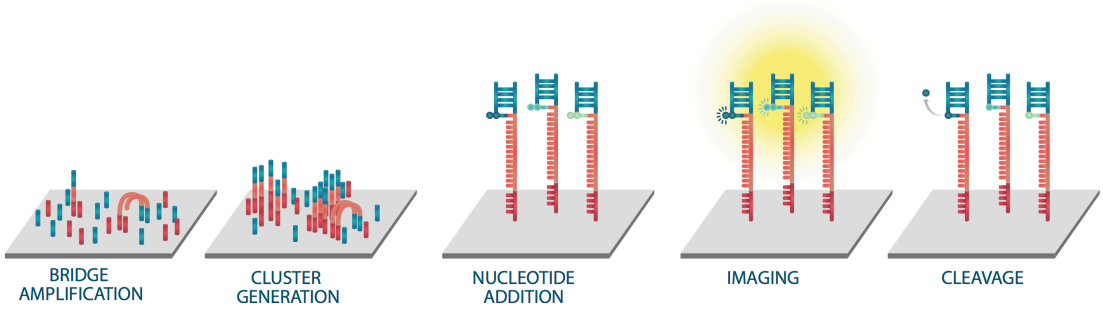


Figure 2.3 Illumina bridge amplification and sequencing by synthesis. The fragmented DNA template of interest, flanked by Illumina-compatible adapters is hybridised to primers immobilised on a solid surface of a flow cell. Bridge amplification takes place by the interaction of the unbound adapter with compatible primers in close proximity and subsequent PCR amplification, resulting in the sequencing cluster generation. Sequencing is carried out in a cyclic manner by the addition of primers, fluorescently labelled and terminally blocked nucleotides (i.e. preventing further DNA synthesis beyond this point) and a DNA polymerase. During each cycle, one of the four added nucleotides binds to the complementary base of the template and the fluorescent colours emitted by the incorporated bases are recorded by a camera. Fluorophores are subsequently cleaved, removed from the flow cell and the 3' OH-group at the last nucleotide is regenerated, thereby allowing the addition of a new nucleotide in the following position. The cycle of nucleotide addition, imaging and OH-group regeneration is then repeated. Adapted from Goodwin et al., 2016.

Massively parallel sequencing is widely used for the characterisation of whole genomes. After sequencing of the main model organisms and the human genome had been completed by 2001 (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001), a plethora of genomes of different species have been added to the growing list. Furthermore, massively parallel sequencing

approaches have enabled population genomic studies that require a high number of samples (1000 Genomes Project Consortium et al., 2010; Sudmant et al., 2015; UK10K Consortium et al., 2015; Gudbjartsson et al., 2015; Lek et al., 2016; Narasimhan et al., 2016; Saleheen et al., 2017). Large catalogues of healthy and diseased human genomes have been assembled as well. These include the “Cancer Genome Atlas”, the “Cancer genome project” or the “1000 Genome project” (Cancer Genome Atlas Research Network et al., 2013; 1000 Genomes Project Consortium et al., 2010; 1000 Genomes Project Consortium et al., 2015), which aim to provide a rich list of mutations and sequence variations (e.g. SNPs) that occur in cancer patients and the healthy human population. Several initiatives in different countries have been launched, aiming to increase the large-scale analysis of human genomes, with the “100.000 Genome Project” in the UK (Siva, 2015) constituting the biggest effort so far.

In contrast to whole genome sequencing, whole exome and targeted sequencing techniques analyse only a specific part of the genome and thereby allow increased sequencing throughput and depth at a lower cost. The majority of the variants linked with human disease are found in coding regions and these techniques can hence be used as diagnostic tools in the clinic (Van Dijk et al., 2014; Lohr et al., 2014).

Massively parallel sequencing has also transformed whole transcriptome analysis. RNA-sequencing has effectively replaced microarray approaches due to its much increased sensitivity and specificity (Wang et al., 2009). Another advantage of sequencing over arrays is that a single protocol can be employed for RNA analysis of any species. RNA-sequencing can be used to sequence the entire RNA molecule, to investigate transcriptional start or termination sites, to study RNA structure or as a read-out for screening (Ozsolak and Milos, 2011; Gapp et al., 2016; Dixit et al., 2016; Adamson et al., 2016; Datlinger et al., 2017).

Standard RNA-sequencing protocols that are designed for sequencing the entire transcript body can be used to investigate splicing patterns, gene fusions, allelic expression or single nucleotide variants (Ozsolak and Milos, 2011). Although these protocols are valuable for the characterisation of transcripts in detail, the relatively low sequencing throughput (usually around 10 samples on a single Illumina HiSeq2000 lane) makes it a cost-intense read-out for screening. In order to increase the throughput and ultimately exploit RNA-sequencing as a quantitative read-out for large-scale reverse genetic screening, the results sections of chapter 2 and 3 describe the development and application of a 3' based library preparation protocol that facilitates a high degree of multiplexing (Gapp et al., 2016). This protocol specifically amplifies the 3' end of each transcript by using a primer that is complementary to the poly-A tail of each mRNA in combination with a random primer. Sequencing only the 3' end of each transcript allows a higher degree of multiplexing (48 or more samples on a single Illumina HiSeq2000 lane) because only a shorter portion of the transcript has to be sequenced, which in turn reduces the complexity of the library and thus the need for deeper sequencing. Ultimately, this reduces the cost per sample sequenced, making it an attractive unbiased read-out for screening. A disadvantage of this approach is that it does not provide information on the entire RNA molecule and hence cannot be used to study the sequence or splicing of transcripts in detail.

Another exciting application is the use of massively parallel sequencing to study the DNA and RNA of single cells. Single cell sequencing has been used to investigate the diversity of microbial ecosystems or the heterogeneity of normal and diseased human tissues (especially cancer) (Gawad et al., 2016). More recently, single cell RNA-sequencing has been employed to investigate molecular circuitries in immune cells or the unfolded protein response (Dixit et al., 2016; Adamson et al., 2016; Jaitin et al., 2016; Datlinger et al., 2017).

Advances in massively parallel sequencing have propelled the high-throughput analysis of DNA and RNA samples and led to a hitherto unmet robustness and quality of sequencing data. Sequencing platforms and data analysis pipelines continue to evolve and new systems (e.g. MinION by Oxford Nanopore Technologies or DNA sequencers by Quantum Biosystems), allowing the sequencing of single molecules in real time, are becoming available (Di Ventra and Taniguchi, 2016). Together, massively parallel sequencing will continue to transform basic and translational research and will enable the analysis and understanding of increasingly complex phenotypes.

The experimental parts of chapters 2 and 3 combine state-of-the-art technologies (CRISPR/Cas9 and RNA-sequencing) with haploid genetics to explore the feasibility of using transcriptional profiling for reverse genetics read-out in human cells. Chapter 2 aims to establish a phenotypic profiling method based on multiplexed RNA-sequencing that is scalable and suitable for large-scale reverse genetic screening. Chapter 3 employs the experimental and computational framework for a proof-of-concept reverse genetic screen to functionally annotate tyrosine kinases in cellular pathways.

2.2. Establishment of a phenotypic profiling platform

2.2.1. Experimental design and set-up of a multiplexed RNA-sequencing platform

An insightful strategy to connect genotypes to phenotypes in lower organisms (i.e., yeast) has involved transcriptional profiling of mutant cells (DeRisi et al., 1997; Hughes et al., 2000). In particular, the comparison of transcriptional signatures induced upon genetic, chemical or environmental perturbations have enabled the

functional annotation of genes at large scale (Holstege et al., 1998; Chua et al., 2006; Lamb et al., 2006; Hu et al., 2007; van Wageningen et al., 2010; Lenstra et al., 2011; Kemmeren et al., 2014). However, the feasibility of such an approach in higher organisms, including human cells, has hitherto not been demonstrated. The aim of this part of the study is to establish a transcriptional profiling platform based on highly multiplexed RNA-sequencing that can be exploited to obtain a quantitative unbiased read-out in human cells.

To benchmark highly multiplexed RNA-sequencing in this setting, 70 diverse polypeptides and small molecules (Appendix Table 6.3) were selected based on the expression of their reported target genes (FPKM > 1). In order to collect and assess expression profiles upon perturbation with this set of stimuli, human near-haploid HAP1 cells were seeded to 12-well plates in medium containing 10% FBS. 36 h post seeding, cells were cultured under reduced serum conditions to minimise the influence of agents contained in the culture media for 16 h. HAP1 cells were stimulated with the recombinant polypeptides and small molecules for 6 h at 70% cell confluency (Figure 2.4A and Appendix Table 6.3). Following RNA isolation and library preparation, 48 unstimulated and stimulated samples were multiplexed on a single Illumina HiSeq lane. The employed 3’ based library preparation protocol was co-developed with Lexogen GmbH and was designed to facilitate a high degree of multiplexing. The protocol is now commercially available as QuantSeq 3’ mRNA-Seq library prep kit.

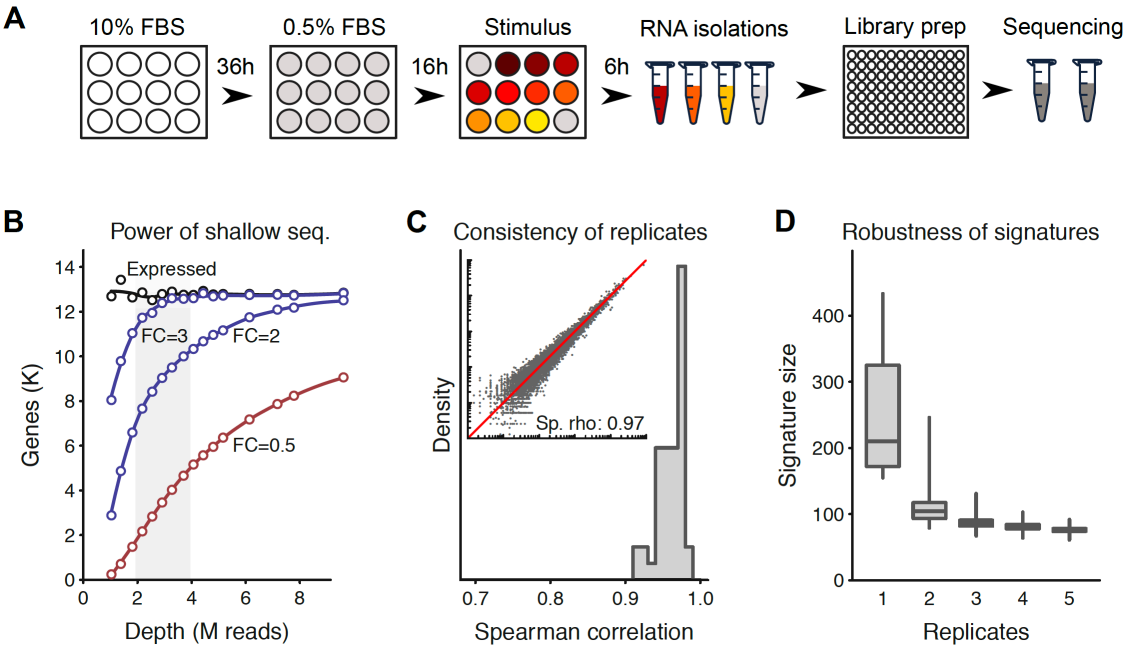


Figure 2.4 Multiplexed RNA-sequencing. A. Experimental approach for stimulation experiments. HAP1 cells were seeded to 12-well plates and cultured in IMDM+10 % FBS for 36 h. Cells were subsequently grown in IMDM+0.5 % FBS for 16 h and stimulated with polypeptides or small molecules for 6 h. Total RNA was isolated and prepared for massively parallel sequencing using QuantSeq 3´mRNA-Seq library prep protocol. 48 libraries were pooled and sequenced on a single Illumina HiSeq2000 lane. B. Modelling of gene expression analysis in dependency on sequencing depth. Circles indicate synthetic samples generated by subsampling of 24 unstimulated wild-type expression profiles. Black line indicates the number of genes detected (cut-off used is transcripts per million reads above 1), coloured lines show the number of genes scored as differentially expressed if the expression changes by a factor of 0.5, 2 or 3. C. Spearman correlations between replicates of expression profiles measured with multiplexed RNA-sequencing. Inset indicates expression values of a representative pair of replicates. D. Modelling of transcriptional signature robustness. Stimulation experiments with FGF18 were carried out in 8 replicates and used to model the signature robustness with increasing replicate numbers. Transcriptional signatures of unstimulated and FGF18 stimulated samples were compared and the signature size of a subset of replicates was determined. Signature size calculations were repeated 20 times. Boxplots show quartiles (minimum, 25, 50, 75 %, and maximum) of the repeated calculations. Adapted from Gapp et al., 2016.

Multiplexing of 48 samples on one Illumina HiSeq2000 lane results in shallow RNA-sequencing with $3x10^6$ reads per sample on average. Highly multiplexed RNA-sequencing allows the detection of more than 95% of all genes expressed in HAP1 wild-type cells and the quantification of more than 70% of the genes if their expression changes by a factor of two (Figure 2.4B). Furthermore, Spearman correlations between replicates of expression profiles were high (Sp. rho: 0.97), indicating that the shallow sequencing approach is highly reproducible (Figure 2.4C). The expression profiles after stimulation were compared to unstimulated samples to determine differentially expressed genes that are specifically altered in response to a particular agent. Concordance across replicate samples was subsequently used to define stimulus-specific gene sets (transcriptional signatures). FGF18 stimulations carried out in eight replicates were used to model the robustness of the computed transcriptional signatures depending on the number of replicates (Figure 2.4D). This showed that defined transcriptional signatures were reproducible when based only on two replicates, with the majority of signature genes called in both samples.

2.2.2. Transcriptional signatures of HAP1 cells upon stimulation

Of the 70 tested stimuli, 34 yielded a transcriptional response in HAP1 wild-type cells after 6 h (Figure 2.5). Transcriptional signatures comprised between two and 240 up- and down-regulated genes. Genes scored as differentially expressed and grouped into transcriptional signatures were generally up- rather than down-regulated. The reason for this observation might be attributable to sampling after 6 h, where the different kinetics of mRNA transcription and turn-over play a role. While transcription is a relatively fast process that can be completed within minutes after stimulation, mRNA turn-over relies on transcript stability which can range from several minutes to several hours (de Nadal et al., 2011).

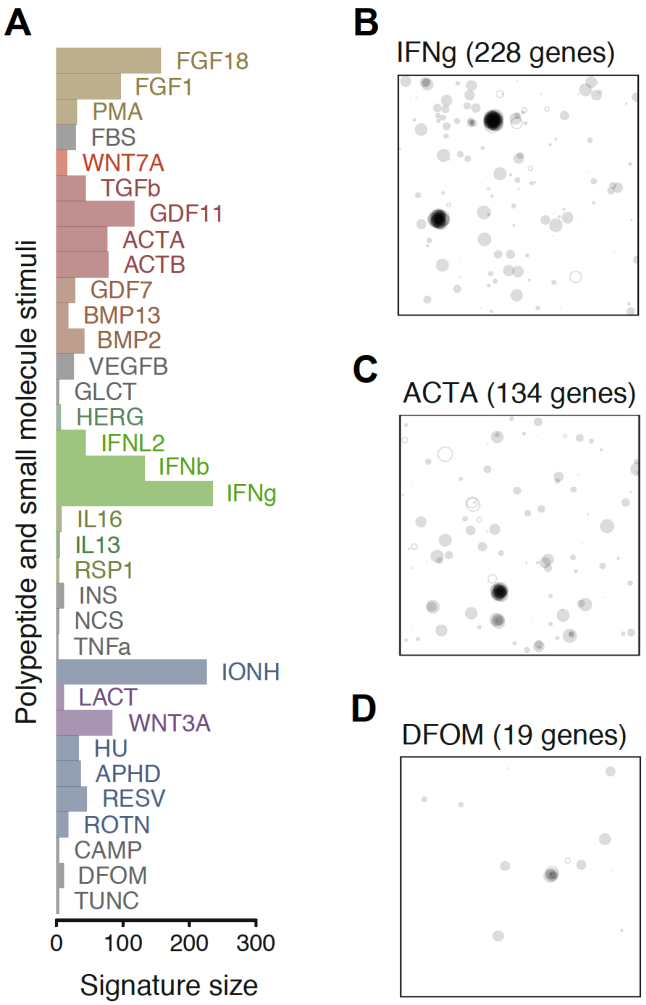


Figure 2.5 Transcriptional signatures of HAP1 wild-type cells upon stimulation. A. Transcriptional signatures induced upon stimulation with polypeptides and small molecules. Differentially expressed genes are defined by comparing expression profiles of stimulated and unstimulated samples. Concordance across replicate samples was used to define transcriptional signatures specific for a particular stimulus. Signature sizes for 34 different stimuli are indicated. B. Visualisation of differentially expressed genes upon stimulation with IFNg. Differentially expressed genes between IFNg stimulated and unstimulated samples are scored based on effect sizes (fold change). Individual up- and down-regulated genes of one replicate are represented as dots; size corresponds to magnitude of fold change. C. As in B for stimulation with ACTA. D. as in B for stimulation with DFOM. Adapted from Gapp et al., 2016.

Examples of one large (228 genes induced by interferon gamma), medium (134 genes induced by activin a) and small (19 genes induced by deferoxamine) transcriptional signature are depicted in Figure 2.5B-D. Failure to detect a transcriptional response for the other tested stimuli could be due to sub-optimal dosing, timing, assay sensitivity or true unresponsiveness of HAP1 cells.

Gene ontology (GO) analysis of the transcriptional signatures revealed associations that were in accordance with the previously described roles of the applied stimuli. For instance, treatment with transforming growth factor beta (TGFb) super-family members (TGFb, activins, bone morphogenetic proteins, growth differentiation factors) resulted in an enrichment of processes including development, differentiation and morphogenesis, fitting with the known function of these polypeptides (Schmierer and Hill, 2007).

The most significant GO terms for the tested interferons were linked to viral and immune responses as well as to the interferon signalling pathway, confirming their well-established roles in antiviral immunity (Schneider et al., 2014).

Top-ranking GO associations in response to fibroblast growth factors (FGFs) were cellular development and cell differentiation, which is concordant with the described functions of the FGF family (Turner and Grose, 2010).

2.3. A platform for reverse genetic screening by highly multiplexed RNA-sequencing

2.3.1. Clustering of transcriptional signatures of stimulated HAP1 cells

Different methods can be applied to detect similarities between transcriptional signatures. One way to assess similarities is based on the Jaccard index, which is defined as the size of the intersection of two sample sets divided by the size of their union. Clustering of stimuli induced transcriptional signatures based on the Jaccard

distance (1 - Jaccard index) revealed similarities in the expression profiles between cells stimulated with polypeptides of the same family (Figure 2.6A).

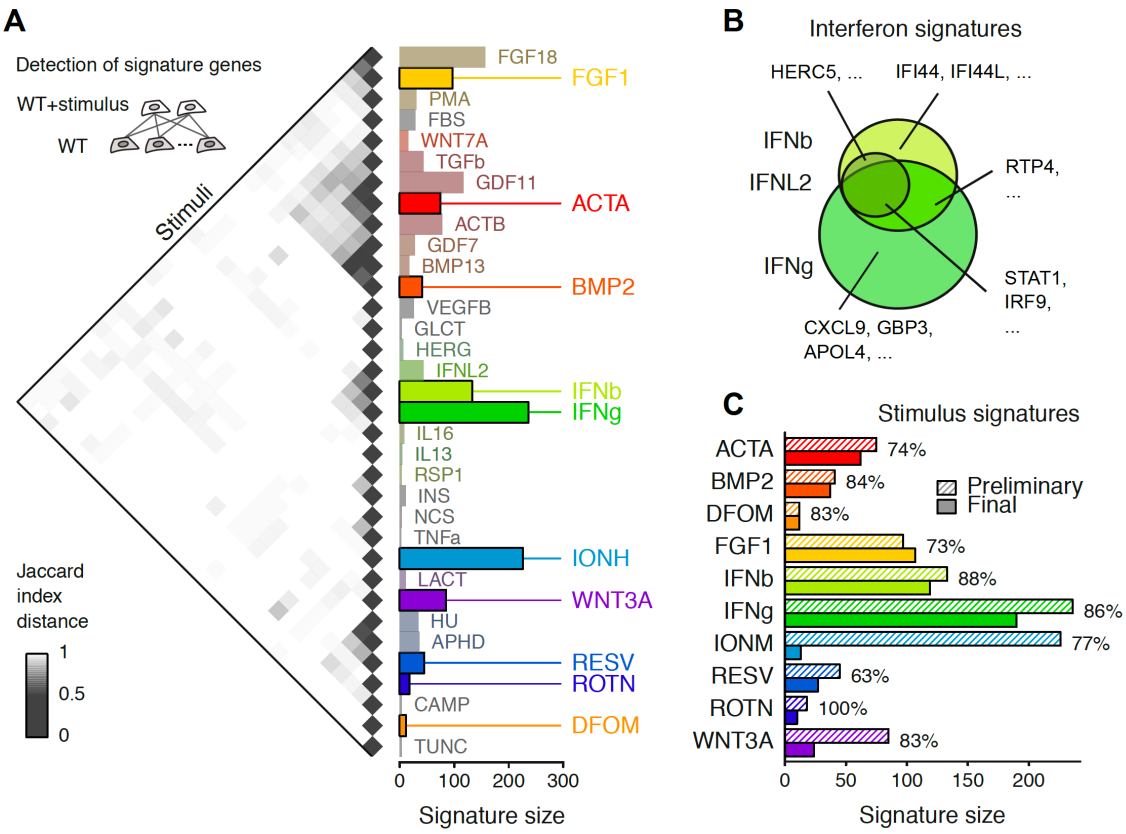


Figure 2.6 Selection of ten stimuli for reverse genetic screening. A. Clustering of transcriptional signatures induced by polypeptides and small molecules. Genes are scored as differentially expressed by comparing transcriptional responses of control samples and samples treated with different stimuli. Transcriptional signatures are determined by consensus of replicate samples. Clustering based on the Jaccard distance indicates similarities between transcriptional signatures, bar graphs indicate signature sizes. Solid colours depict the ten stimuli selected for reverse genetic screening. B. Comparison of transcriptional signatures of different interferons. Transcriptional signature genes shared and specific to IFNb, IFNL2 and IFNg are displayed. C. Transcriptional signature sizes for the ten stimuli selected for reverse genetic screening. Signature sizes for initial and final stimulation experiments are shown, the overlap (size of intersection divided by the size of the final set) between the collected transcriptional signatures is indicated. Adapted from Gapp et al., 2016.

For example, activins (ACTA, ACTB), bone morphogenetic proteins (BMP2, BMP13), growth differentiation factors (GDF7, GDF11) and TGFb belong to the TGFb super-family and clustered together. More specifically, polypeptides signalling via activin receptors (ACTA, ACTB, GDF11) formed one sub-cluster and stimuli binding to BMP receptors (BMP2, BMP13, GDF7) formed another. Also, the transcriptional signatures induced by related interferons (IFNa, IFNb, IFNg) were similar based on the Jaccard distance and clustered together. Although some signature genes were shared between all three tested interferons, others were specific to only one stimulus (Figure 2.6B). This indicates that the shallow sequencing approach does not only capture gross differences in expression profiles upon perturbation with different stimuli but also allows to detect small differences even for related stimuli.

From the complete set of stimuli, six polypeptides and four small molecules were selected based on diversity, reproducibility and transcriptional signature size and used for transcriptome based reverse genetic screening (chapter 3). The polypeptides were activin a (ACTA), bone morphogenetic protein 2 (BMP2), fibroblast growth factor 1 (FGF1), interferon beta (IFNb), interferon gamma (IFNg) and wingless-type family member 3A (WNT3A). The four selected small molecules were deferoxamine (DFOM), ionomycin (IONM), resveratrol (RESV) and rotenone (ROTN). In order to confirm and strengthen the confidence in the transcriptional signatures for these ten agents, stimulation experiments were repeated under the same experimental conditions. With the exception of the IONM signature, the previously defined transcriptional signatures could be confirmed in independent experiments (Figure 2.6C). The reason for the discrepancy between the initial and final IONM signature was a change in the concentration due to toxicity observed at the initial higher concentration.

GO analysis of the induced transcriptional signatures also confirmed the findings from the benchmarking experiment using 70 stimuli and was in accordance with the described biological functions of the ten stimuli. For example, the top-scoring GO

enrichment for BMP2 was the BMP signalling pathway, and response to hypoxia was the second most significant association for DFOM.

2.3.2. Transcriptional signature comparisons between wild-type and knock-out cells

As the described experimental set-up was intended for use as a tool for reverse genetic screening, it was necessary to demonstrate the feasibility of using transcriptional profiling in combination with defined gene knock-outs. To this end, hypoxia-inducible factor 1-alpha (HIF1A) and beta-catenin (CTNNB1) knock-outs were generated using CRISPR/Cas9 technology and stimulated with DFOM and WNT3A, two agents reported to trigger the respective signalling pathways.

HIF1A is a subunit of the HIF1 transcription factor that acts as a master regulator for the transcription of hypoxia-responsive genes (Semenza, 2014). DFOM stabilises HIF1A by inhibiting the iron-dependent prolyl hydroxylases that target HIF1A for degradation. HIF1A knock-out cells should therefore show a dampened hypoxic response compared to HAP1 wild-type cells. Indeed, comparison of the transcriptional signatures between stimulated and unstimulated wild-type and knock-out cells revealed an abrogated transcriptional response in HIF1A knock-out cells (Figure 2.7A).

The WNT/CTNNB1 pathway plays an essential role in both embryonic development and adult life. Upon activation of the WNT pathway e.g. by WNT3A, CTNNB1 translocates to the nucleus and functions as a transcriptional co-activator for the expression of WNT responsive genes (Clevers and Nusse, 2012). Given that CTNNB1 is a key component of this signal transduction pathway, CTNNB1 knock-out cells should have impaired WNT signalling. This prediction could indeed be confirmed by comparing transcriptional signatures between CTNNB1 knock-out and wild-type cells (Figure 2.7B).

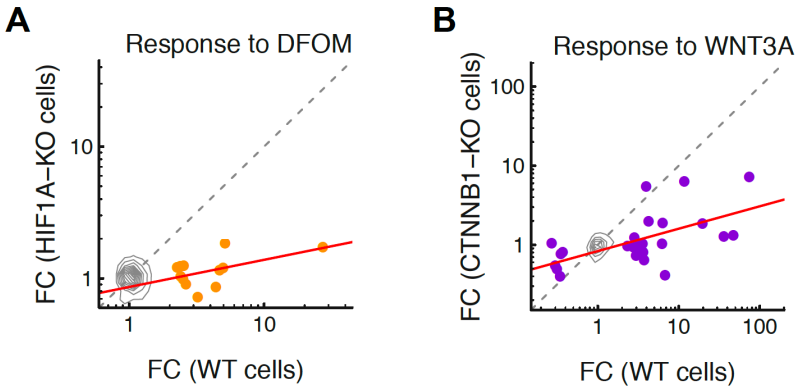


Figure 2.7 HIF1A and CTNNB1 knock-outs display a dampened transcriptional response to DFOM and WNT3A stimulation. A. Transcriptional responses to DFOM in HAP1 wild-type and HIF1A knock-out cells. Fold changes (FC) of the expression profiles between unstimulated and stimulated samples are plotted. Dots show transcriptional signature genes, contour indicates genes not differentially expressed. Gray dashed line is the diagonal, red line is the linear fit based on transcriptional signature genes. B. As in A for transcriptional response to WNT3A in wild-type and CTNNB1 knock-out cells. Adapted from Gapp et al., 2016.

Together, these results demonstrate that the established experimental and computational pipeline allows the connection of genes to pathways and that it can be therefore used for transcriptome based reverse genetic screening.

2.4. Discussion

This chapter describes the establishment of a phenotypic profiling platform based on shallow RNA-sequencing. In order to increase the throughput and achieve a high degree of multiplexing, a library preparation protocol was employed that only amplifies the 3' end of mRNAs. For benchmarking this platform, transcriptional profiles of HAP1 cells in response to a diverse range of external stimuli were collected.

The presented results show that shallow RNA-sequencing offers a generic phenotypic profiling method that allows to study cellular perturbations in human cells at high throughput. High correlation between technical and biological replicates indicates the robustness of the shallow sequencing approach. Furthermore, it was demonstrated that highly multiplexed RNA-sequencing can capture not only general responses to perturbation but is also sensitive enough to reveal subtle differences in signalling response between related stimuli. Due to the continuous technological progress of massively parallel sequencing technology since the beginning of this study, it would now be possible to multiplex up to 96 samples while maintaining the same sequencing depth as is presented here for 48 samples. The first library preparation protocols allowing to multiplex even up to 384 different samples on a single sequencing lane are becoming available. Further improvements in sequencing technology will allow to increase the throughput beyond that in the near future, making shallow RNA-sequencing an even more powerful approach. Importantly, shallow RNA-sequencing is widely applicable to any cell type or tissue and requires only small amounts of input material, making it suitable for the analysis of precious samples (e.g. biopsies).

Recently, also other studies have employed the 3' based library preparation protocol established during this study. Albeit at lower throughput and smaller scale, these studies have measured consequences of drug inhibition or shRNA knock-down in human cell lines (Zhu et al., 2016; Brägelmann et al., 2017; Hogg et al., 2017), characterised genetic screening hits and polyadenylation patterns in mice (Forment et al., 2017; Xiao et al., 2016), investigated RNA decay in flies (Reimao-Pinto et al., 2016) or interrogated the role of miRNAs in the development of the vascular system in fish (Kasper et al., 2017).

Besides the suitability for efficient transcriptional profiling in human cells, the presented phenotypic profiling platform has some potential limitations. While

multiplexing of 48 (or more) samples on a single HiSeq2000 lane increases the throughput and reduces sequencing costs, shallow RNA-sequencing gives reduced resolution compared to deeper sequencing protocols; i.e., changes in genes expressed at low level may not be captured by shallow RNA-sequencing. A biological rather than technical limitation of the established method is that cellular responses that only affect transcription transiently or not at all cannot be assessed.

Transcriptional profiling of HIF1A and CTNNB1 knock-out cells in response to specific stimuli revealed reduced expression of signature genes compared to stimulated wild-type cells. This shows that shallow RNA-sequencing can be used to link genes to pathways in human cells. It has been previously demonstrated that transcriptional signatures can connect genes with phenotypes in model organisms and human cells using microarray technology (DeRisi et al., 1997; Holstege et al., 1998; Hughes et al., 2000; Chua et al., 2006; Lamb et al., 2006; Hu et al., 2007; van Wageningen et al., 2010; Lenstra et al., 2011; Kemmeren et al., 2014). Advantages of RNA-sequencing over array-based technologies include the increased throughput and precision by which expression levels can be determined. Furthermore, microarray technologies are based on previously selected query transcripts that are immobilised on a chip, whereas RNA-sequencing is an unbiased method and does not depend on array design (Wang et al., 2009).

With minor modifications of the experimental set-up, highly multiplexed RNA-sequencing will also be suited to survey diverse cellular perturbations (e.g. growth conditions or responses to different environments) or improve compound MoA studies previously performed in yeast and human cells using microarrays (Hughes et al., 2000; Lamb et al., 2006). Another interesting application would be to study dynamic stimulus responses in a high-throughput manner, by sampling several time points in order to chart regulatory networks in the cell, for example the signalling

mechanisms underlying pathogen responses (Bar-Joseph et al., 2012). Highly multiplexed RNA-sequencing can be useful to monitor transcriptional responses at large scale to understand a multitude of processes to investigate both basic biology and disease mechanisms. As an example for such an application, shallow RNA-sequencing was used for large-scale reverse genetics in a human cell system, as discussed in chapter 3.

In summary, multiplexed RNA-sequencing is a highly robust and sensitive method to study cellular perturbations and provides a scalable and suitable method for reverse genetic screening in human cells.

Chapter 3. Parallel reverse genetic screening of human knock-out cells

3.1. Introduction

Using the experimental platform described in chapter 2, a proof-of-concept reverse genetic screen aiming to annotate tyrosine kinases in different signalling pathways was designed. This chapter provides an introduction to these kinases, their roles in cellular signalling cascades and highlights their importance as therapeutic targets. The results section of this chapter describes the proof-of-concept study for transcriptome based reverse genetic screening of tyrosine kinases in detail.

3.1.1. Introduction to tyrosine kinases

Tyrosine kinases (TKs) have fundamental roles in a wide range of cellular processes including cell division, differentiation and migration. With 90 members, TKs comprise the second largest group among the kinase family and can be classified into receptor tyrosine kinases (RTKs) and cytoplasmic TKs (Manning et al., 2002). RTKs are high-affinity cell surface receptors that bind polypeptide ligands, whereas cytoplasmic TKs are effector or adapter proteins transmitting signals within the cytoplasm and to the nucleus. By transferring phosphate from ATP to intramolecular tyrosine residues (autophosphorylation) or to other target proteins, they modulate protein activity, interaction and subcellular localisation (Lemmon and Schlessinger, 2010).

Aberrant regulation of mutations in TK genes are associated with a variety of human pathologies such as cancer (Huang et al., 2014; Fleuren et al., 2016), chronic inflammatory and autoimmune diseases (Clark et al., 2014; Stark et al., 2015; Futosi and Mocsai, 2016), Alzheimer's disease (Nygaard et al., 2014) and diabetes mellitus (Fountas et al., 2015). The central role of TKs and their association with human diseases makes them attractive pharmaceutical targets (Rask-Andersen et al., 2014; Fleuren et al., 2016). Small molecule inhibitors and monoclonal antibodies targeting aberrantly regulated or mutated kinases have been identified and successfully applied particularly in the field of oncology (Wu et al., 2015; Gharwan and Groninger,

2016). More recently, TK inhibitors have become available or are being evaluated in clinical trials also for the treatment of inflammatory or metabolic conditions and Alzheimer's disease (Nygaard et al., 2014; Rask-Andersen et al., 2014; Fountas et al., 2015; Schwartz et al., 2016).

Often the kinases that have been targeted therapeutically represent well-characterised cytoplasmic TKs and RTKs. Hence, potentially effective drug targets remain unexploited due to the limited knowledge on this important class of enzymes (Rask-Andersen et al., 2014). Gaining insight into the cellular functions of so far poorly characterised kinases could thus affect fundamental as well as translational research.

3.1.2. Tyrosine kinases and signalling

TKs are essential components of numerous intra- and intercellular signalling pathways responding to a wide range of external stimuli. They can bind polypeptide ligands directly (RTKs) or act further down-stream (cytoplasmic TKs) by transducing the signal to the nucleus upon binding of an external stimulus, leading to the activation of a stimulus-specific transcriptional response (Manning et al., 2002).

There are 58 RTKs encoded in the human genome. Well-characterised subfamilies include the epidermal growth factor receptors (EGFRs), fibroblast growth factor receptors (FGFRs) or vascular endothelial growth factor receptors (VEGFRs). RTKs are high-affinity cell surface receptors for a variety of growth factors, hormones and cytokines. All RTKs share a similar architecture: an extracellular ligand binding domain, a single-pass transmembrane domain, a cytoplasmic kinase domain and a regulatory c-terminal domain. RTK activation by ligand-binding to the extracellular domain leads to receptor dimerisation, autophosphorylation of the kinase domain(s) and a conformational change that stabilises the active kinase. The phosphorylated tyrosine residues of RTKs provide a recognition and binding surface for down-stream adapter, scaffold and effector proteins containing Src homology (SH2) or

phosphotyrosine binding (PTB) domains. These signalling components then transmit the signal from the ligand-activated RTK further until it reaches the nucleus to trigger a ligand-specific transcriptional programme (Lemmon and Schlessinger, 2010).

A well-studied example for a RTK signalling pathway is the FGF pathway which is important for development, angiogenesis, differentiation and proliferation. Since both the FGF1 stimulus and FGFR knock-outs are used for reverse genetic screening in this study (see results part of this chapter), this signalling cascade is used here as a representative example. With 18 FGF ligands and 4 FGFRs, the FGF signalling pathway is highly complex and context dependent. A simplified schematic of the FGF signalling pathway is displayed in Figure 3.1. Upon binding of an FGF ligand to an FGFR, receptors dimerise and transphosphorylate each other. The phosphorylation of tyrosine residues allows the binding of PTB domain-containing adapter proteins such as FGFR substrate 2 (FRS2) which is largely specific to FGFRs. Subsequent FRS2 phosphorylation by the activated FGFR enables the recruitment of other adapter proteins such as son of sevenless (SOS), growth factor receptor-bound 2 (GRB2) and GRB2-associated binding protein 1 (GAB1), thereby activating the down-stream RAS-RAF-MAPK and PI3K-AKT pathways (Turner and Grose, 2010).

Besides FRS2, also the SH2 domain-containing phospholipase C gamma (PLCγ) can directly bind to activated FGFRs. This results in the hydrolysis of phosphatidylinositol-4,5-bisphosphate (PIP2) to inositol-3,4,5-triphosphate (IP3) and diacylglycerol (DAG) which activates protein kinase C (PKC). PKC then phosphorylates RAF, thereby further activating the RAF-MAPK pathway. Depending on the cellular context, several other pathways involving p38 MAPK, signal transducer and activators of transcription (STAT) or ribosomal S6 kinase 2 (RSK2) can be activated by signalling through FGFRs as well. Together the activated pathways drive target gene expression, thereby regulating central cellular processes such as differentiation, proliferation or survival (Turner and Grose, 2010).

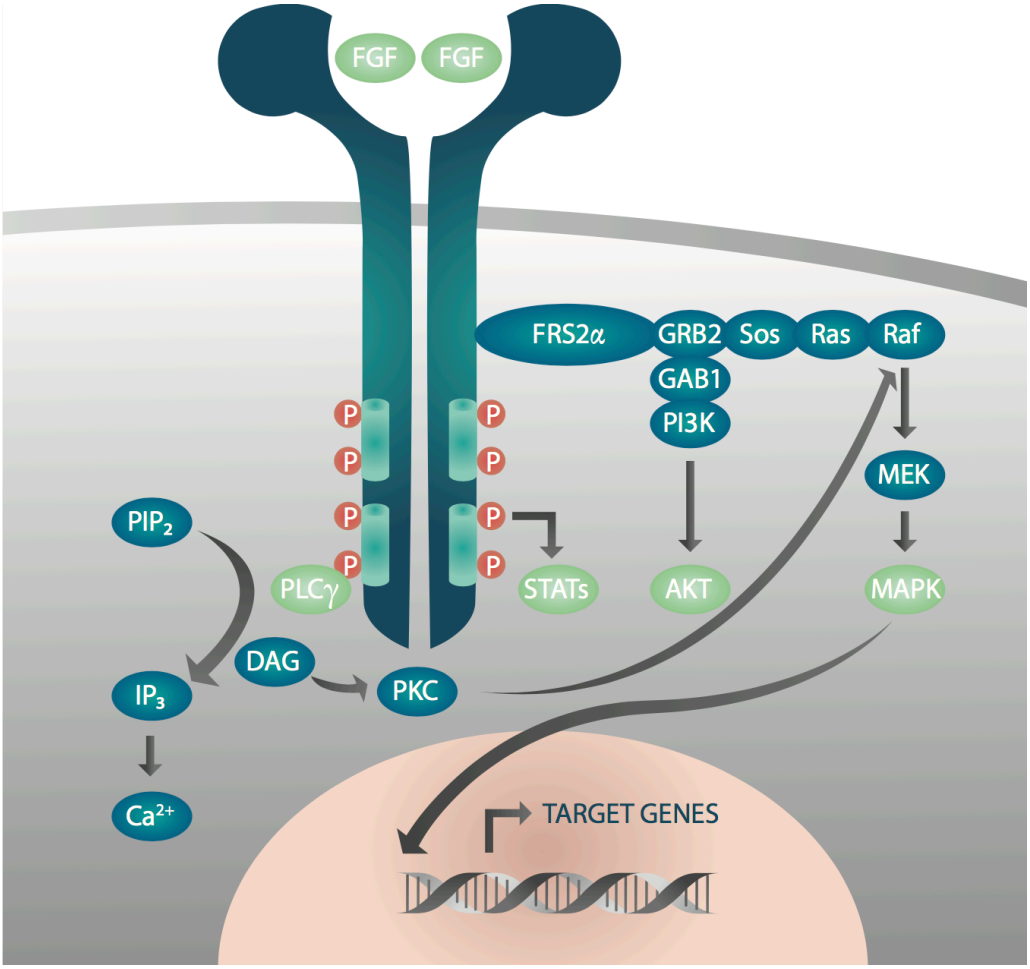


Figure 3.1 The FGF signalling pathway. FGF binding to the extracellular domains of FGFRs leads to receptor dimerisation, conformational changes and subsequent transphosphorylation of the intracellular FGFR domains. This allows the binding of several adapter proteins such as FRS2α, GRB2, SOS or GAB1, activating the RAS-RAF-MAPK pathway and PI3K-AKT pathway as well as STATs and PLCγ. Signal transduction to the nucleus leads to the expression of FGFR pathway specific target genes. Adapted from Turner and Grose, 2010.

In contrast to RTKs, cytoplasmic TKs lack the extracellular and transmembrane domains and only consist of a cytoplasmic domain. Cytoplasmic TKs comprise 32 members and can be found either associated with a receptor, in the cytoplasm not associated with the membrane, or in the nucleus. Cytoplasmic TKs are activated by binding to a receptor, phosphorylation by other kinases or the dissociation of an inhibitor. Once activated, cytoplasmic TKs can be bound by down-stream signalling

components via SH2 domains. Eventually, signal transmission leads to alterations of the expression of target genes. Cytoplasmic TKs play essential roles in proliferation, differentiation or adhesion and are critical components for the control of the immune system. Famous cytoplasmic TKs include Rous sarcoma oncogene cellular homolog (SRC), Janus family tyrosine kinase (JAK) or Abelson leukemia oncogene cellular homolog (ABL) (Manning et al., 2002).

As an example of signalling via cytoplasmic TKs, the JAK-STAT pathway is described (Figure 3.2), since interferon ligands that activate this signalling cascade and JAK kinase knock-outs are included in this study's genetic screens. The JAK-STAT pathway is crucial for the regulation of immune responses and comprises four JAKs and seven STATs. JAKs are receptor-associated cytoplasmic TKs that can bind to a variety of surface receptors. Some cytokine receptors lack an intracellular tyrosine kinase domain and therefore depend on JAKs to initiate cytoplasmic signalling events. Cytokine binding to a JAK-associated receptor leads to receptor dimerisation and a conformational change that activates JAKs. These subsequently phosphorylate each other, as well as the receptors they are associated with, allowing the binding of SH2 domain-containing STATs. Subsequent phosphorylation of the STAT transcription factors by JAKs results in STAT dimerisation and translocation into the nucleus. STATs can then activate the expression of cytokine-regulated genes in the nucleus (O'Shea et al., 2015).

Given the essential role of TKs in diverse signalling cascades where they control various biological processes, mutated or aberrantly regulated TKs significantly contribute to carcinogenesis and other human diseases. As a consequence, several TK specific inhibitors and antibodies have been developed for therapeutic intervention and are discussed in the section below.

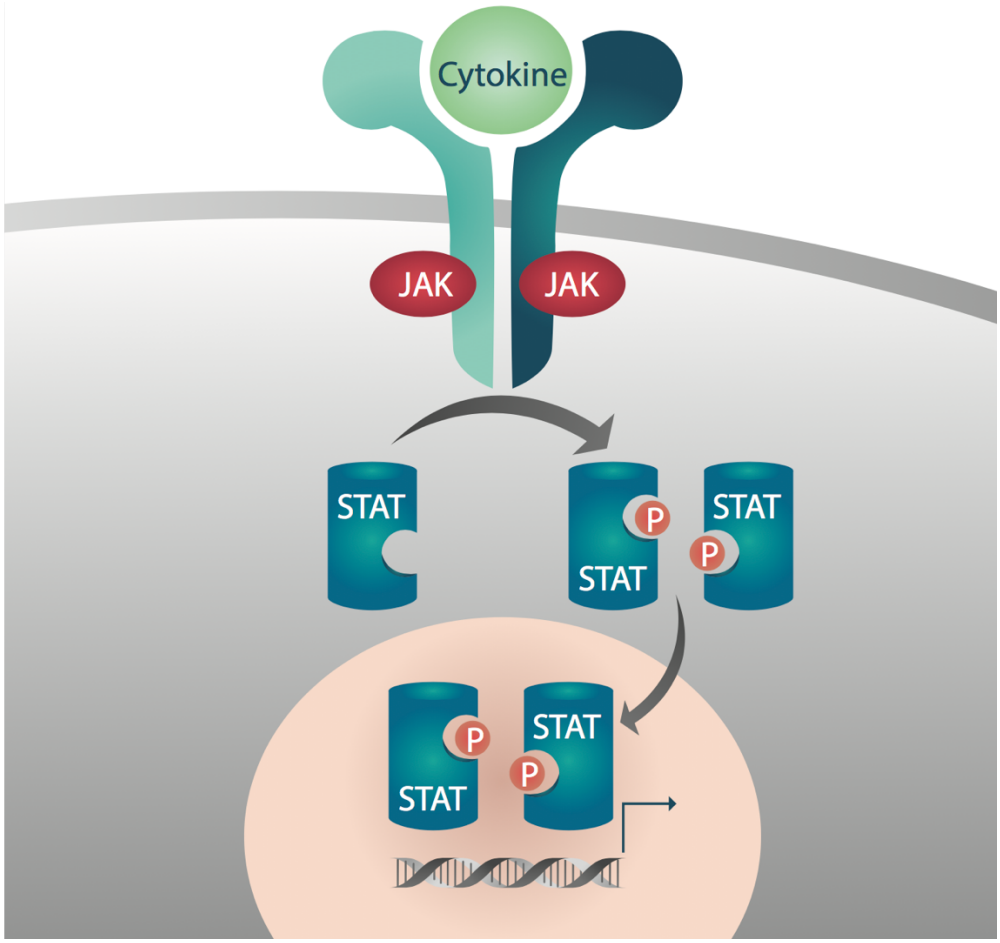


Figure 3.2 The JAK-STAT signalling pathway. Cytokine binding to the extracellular domains of JAK associated receptors results in receptor dimerisation, conformational changes and subsequent JAK activation. JAKs, in turn, phosphorylate the intracellular domains of the activated receptors and each other. This provides a binding surface for STAT transcription factors that are then phosphorylated, leading to dimerisation and translocation to the nucleus. STATs then regulate the expression of a set of target genes. Adapted from Shuai and Liu, 2003.

3.1.3. Tyrosine kinase inhibitors

The development of TK inhibitors has been a major advance in cancer treatment, providing a specific alternative to broadly acting chemotherapeutics. Since the approval of the first TK inhibitor to target BCR-ABL in CML (Druker et al., 2001), 15 other small molecules and monoclonal antibodies have entered the clinic for the

targeted treatment of various cancers. Many more are currently in clinical or pre-clinical trials to assess their potential as anticancer agents (Fleuren et al., 2016).

TK inhibitors can bind reversibly or irreversibly to their target protein. Reversible TK inhibitors fall into five groups and act either by competing with ATP or allosterically by inhibiting the catalytic activity of the TK: i) ATP-competitive inhibitors that recognise and bind the ATP-binding site when the kinase is in an active confirmation; ii) ATP-competitive inhibitors that target the ATP-binding site when the kinase is in a non-active confirmation; iii) allosteric inhibitors that bind outside but in close proximity to the ATP-binding site and change the conformation of the ATP-binding pocket; iv) allosteric inhibitors that bind away from the ATP-binding site; v) inhibitors with multiple binding modes. Irreversible TK inhibitors usually bind close to the ATP-binding pocket and thereby block the ATP-binding site (Wu et al., 2015).

Inhibitors can be specific for RTKs, cytoplasmic TKs, or can target multiple kinases. Some have even been designed to only target a mutated kinase variant, making these inhibitors very potent and highly selective. However, off-target effects are often observed towards the normal unmutated kinase or related other kinases. On the other hand, such “polypharmacology” may contribute to the efficacy of the small molecules and might be superior to selective inhibitors in some cases, since often multiple TK pathways are activated simultaneously in pathological signalling, especially in cancer (Hopkins, 2008). Although, treatment with selective TK inhibitors can result in complete remissions in some cases, drug resistance is almost inevitable, ultimately leading to progressive disease and poor survival. This can be due to the selection of pre-existing or *de novo* clonal variants that are resistant to the targeted agents (Huang et al., 2014). Resistance mechanisms include point mutations in the inhibitor binding site, gene amplification or over-expression, compensating alterations in the affected signalling pathways and modulation of drug influx/efflux mediated e.g. by ABC transporter over-expression. Better understanding of resistance mechanisms is crucial to develop next generation TK inhibitors that are

highly effective and overcome currently known resistance mechanisms (Lovly and Shaw, 2014).

A different strategy to target RTKs in a highly specific manner is by using monoclonal antibodies. Again, this strategy has been mainly employed in the treatment of cancer. High specificity is achieved by targeting antigens that are differentially expressed in tumour cells or mutated versions that are not expressed on non-cancerous cells (tumour-associated antigens) (Weiner, 2015; Gharwan and Groninger, 2016). Targeting the extracellular domain of an RTK with a monoclonal antibody blocks the ligand-binding site and thereby inhibits the dimerisation and activation of the RTK itself and down-stream signalling. Binding of a monoclonal antibody to an RTK can furthermore result in receptor internalisation, induction of apoptosis by antibody-dependent cell-mediated cytotoxicity, complement-mediated cytotoxicity and antibody-dependent phagocytosis by innate immune cells (Weiner, 2015). The first approved therapeutic antibody was directed against HER2 (trastuzumab in 1998), a TK frequently over-expressed in breast cancer (Hudis, 2007). Other examples of successfully applied antibodies are cetuximab and panitumumab which are directed against EGFR to treat metastatic colorectal cancer (Gharwan and Groninger, 2016). The approval of some monoclonal antibodies has been extended to the treatment of other cancers in addition to the initial indication after proven effective there. For example, since 2010 trastuzumab is also approved for the treatment of HER2-positive metastatic stomach or gastroesophageal cancers (Lordick and Janjigian, 2016). Furthermore, antibodies with improved bioavailability and activity have continued to enter clinical trials. Antibodies targeting different RTKs including members of the MET proto-oncogene receptor tyrosine kinase (MET), insulin like growth factor 1 receptor (IGF1R), platelet derived growth factor receptor (PDGFR) or

the FGFR subfamilies have also been developed and are currently being tested in clinical trials (Fauvel and Yasri, 2014; Fleuren et al., 2016).

Other exciting applications of targeted RTK therapy include monoclonal antibodies conjugated to drugs, cytotoxins or radioactive molecules, thereby combining the highly specific properties of antibodies with potent cytotoxic agents. The first approved antibody-compound conjugate was ado-trastuzumab-emtansine for the treatment of advanced HER2-positive breast cancer and metastatic breast cancer. After binding of trastuzumab to the HER2 RTK, emtansine is released and inhibits microtubule assembly (Weiner, 2015).

As with small molecule inhibitors, patients can become resistant to antibody treatments as well. Resistance mechanisms include mutations of the antibody binding site, compensation of the targeted pathway by upregulation of down-stream components or the activation of an alternative pathway (Weiner, 2015; Gharwan and Groninger, 2016).

Although notable examples for the clinical exploitation of deregulated kinases exist, many remain unexploited due to limited understanding of the cellular functions of the key signalling molecules. The following results section describes a proof-of-concept study for the functional annotation of TKs aiming to gain more insight into this important class of enzymes.

3.2. Parallel reverse genetic screening of kinase knock-out cells

3.2.1. A collection of CRISPR/Cas9 knock-out cell lines

The emergence of CRISPR/Cas9 technology for the first time allows the targeted generation of knock-out panels and hence, systematic reverse genetic investigations in human cells. A commercially available human knock-out collection generated with

CRISPR/Cas9 covers all non-essential kinases and is employed in this study. Using the TK subset of this targeted knock-out collection in combination with the phenotypic screening platform described in the previous chapter should enable to test the feasibility of using transcriptome based screening as a functional gene annotation tool in a human cell system.

The isogenic TK knock-out panel was generated at Horizon Genomics GmbH using CRISPR/Cas9 technology. gRNAs were designed for all TK genes (Appendix Table 6.1) that are expressed (FPKM>0.1) and non-essential in HAP1 wild-type cells (Essletzbichler et al., 2014; Blomen et al., 2015). Of the 56 TKs that met these criteria, 55 could be successfully engineered to contain frameshift mutations, resulting in a functional TK knock-out (Figure 3.3).

In addition, some non-TK genes that could potentially serve as controls were selected for knock-out based on literature, as well as RNA expression levels and essentiality as before (Essletzbichler et al., 2014; Blomen et al., 2015). For instance, activins and BMPs signal via the corresponding activin and BMP receptors. Compared to wild-type cells, activin A receptor type 1B (ACVR1B) and type 2B (ACVR2B) knock-outs were expected to show a reduced transcriptional response upon ACTA stimulation, and bone morphogenetic protein receptor type 1A (BMPR1A) and type 2 (BMPR2) knock-out cells were predicted to serve as positive controls for BMP2 stimulation (Schmierer and Hill, 2007). Similarly, knock-outs for the calcium-binding proteins calmodulin 2 (CALM2) and calmodulin 3 (CALM3) were generated as potential controls for the calcium-modulating agent IONM, and histone deacetylase sirtuin 1 (SIRT1) knock-outs were considered as positive controls for the stimulation with the histone deacetylase activator RESV. HIF1A and CTNNB1 knock-out cells, already shown to yield a reduced transcriptional response upon DFOM and WNT3A stimulation compared to wild-type cells (chapter 2), were included and screened against a panel of ten stimuli as well. All clones harbouring frameshift mutations predicted to lead to a functional knock-out of the targeted TK and control

genes were confirmed by Sanger sequencing for quality control (Appendix Table 6.2). Clones harbouring frameshift indels in coding regions were then expanded and used for genetic screening.

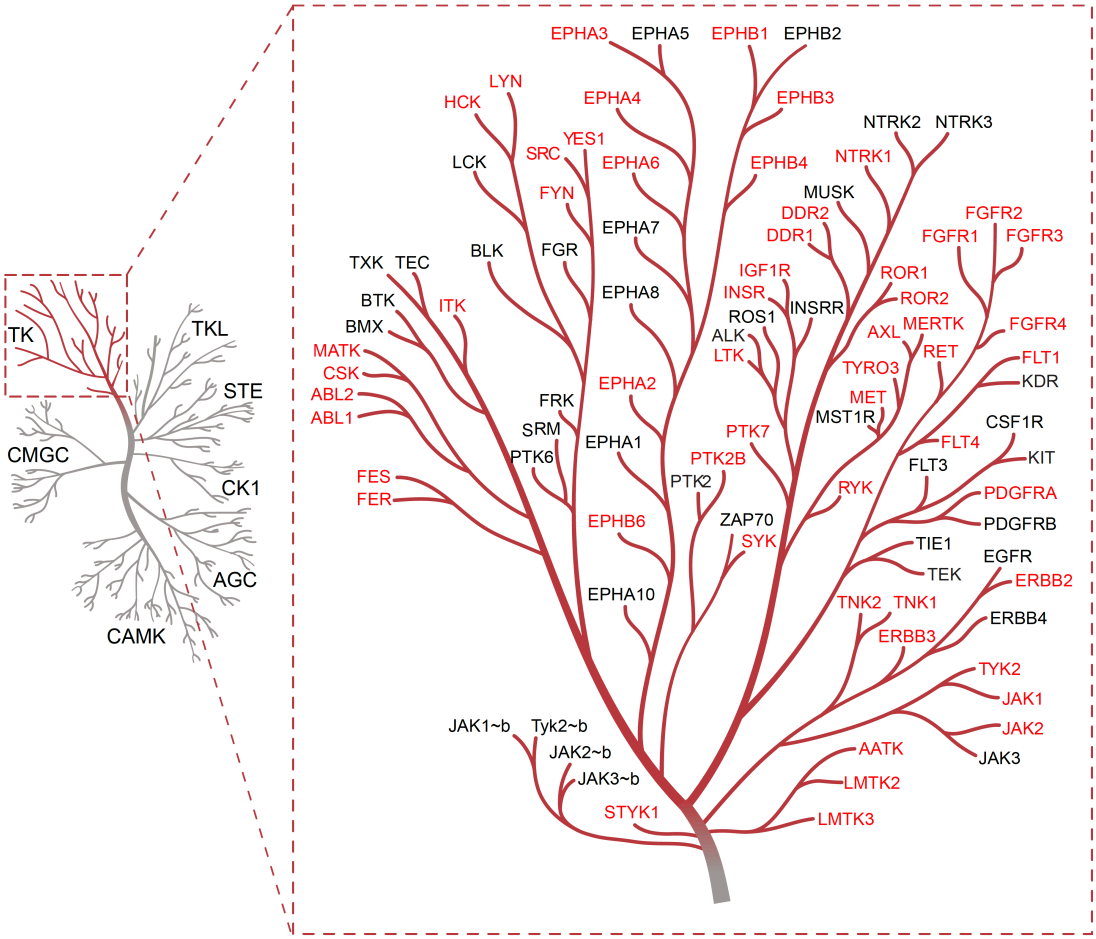


Figure 3.3 An isogenic TK knock-out panel. Schematic of the human kinome tree (left) with a zoom-in into the TK branch (right). Abbreviations on the left indicate kinase groups, abbreviations on the right designate TK genes. Red represents expressed and non-essential TKs selected for CRISPR/Cas9 engineering, black indicates TKs not selected for editing. Taken from Gapp et al., 2016.

3.2.2. Experimental design of the parallel reverse genetic approach and consistency of replicates

To test if transcriptional signature comparisons between HAP1 wild-type and knock-out cell lines can link TKs to signalling pathways in a large-scale setting, knock-out cells were screened against a panel of ten stimuli. Six polypeptides and four small molecules were selected based on diversity, reproducibility and transcriptional signature from the benchmarking experiments presented in chapter 2. As described, polypeptides included activin a (ACTA), bone morphogenetic protein 2 (BMP2), fibroblast growth factor 1 (FGF1), interferon beta (IFNb), interferon gamma (IFNg) and wingless-type family member 3A (WNT3A). The four selected small molecules were deferoxamine (DFOM), ionomycin (IONM), resveratrol (RESV) and rotenone (ROTN).

In total, 64 isogenic knock-out cell lines (55 TK and nine control knock-outs) were processed in batches of four cell lines with ten stimuli and unstimulated controls, handling 96 samples at a time (Figure 3.4A). Parallel screening of the ten stimuli and unstimulated controls allowed signature comparisons within the same batch in the majority of cases. By maintaining replicates and knock-out samples within one batch, most computational analyses could be carried out without the need for batch correction, which can have confounding effects on the expression analysis. Further advantages of this screening design are its scalability and modularity, allowing easy expansion of transcriptome based reverse genetic screening beyond the analysis of TKs.

All stimulation experiments were performed in a 12-well format under reduced serum conditions (0.5 % FBS) in duplicates. Libraries were generated in 96-well plates and 48 samples pooled per Illumina HiSeq2000 lane (Figure 3.4A). In total, 1700 transcriptional signatures of stimulated and unstimulated cell lines were collected. Spearman correlations between replicate samples were high, indicating robust performance of the established assay also in a large screening setting (Figure 3.4B).

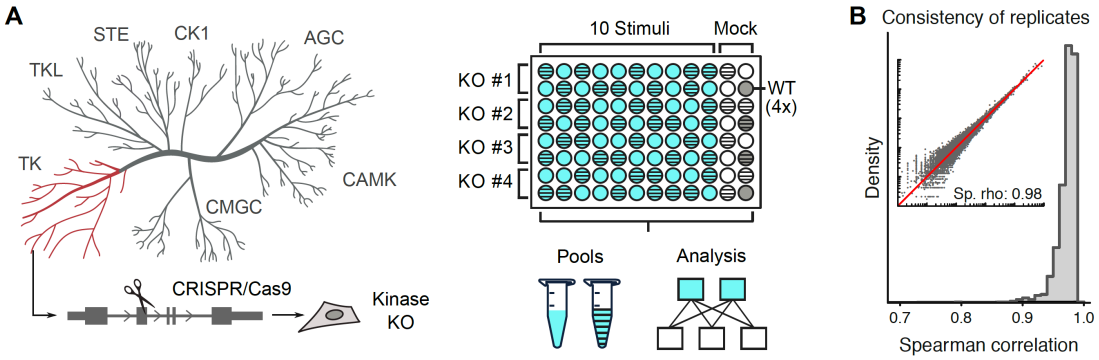


Figure 3.4 Parallel reverse genetic screening in haploid human cells. A. Experimental approach and data analysis for parallel reverse genetic screening. 64 tyrosine kinase (TK) and control knock-outs were generated using CRISPR/Cas9 technology. Library preparations were carried out in batches consisting of four knock-out cell lines stimulated with ten polypeptides and small molecules together with unstimulated knock-out and wild-type controls in a 96-well format. 48 samples were sequenced on one Illumina HiSeq2000 lane. Transcriptional signature genes were determined by comparing stimulated and unstimulated samples from the same batch. Abbreviations on the left indicate kinase groups. WT, wild-type; KO, knock-out. B. Spearman correlations between replicates of expression profiles of stimulated and unstimulated wild-type and knock-out sample measured by RNA-sequencing. Inset depicts expression values of a representative pair of replicates. Adapted from Gapp et al., 2016.

3.2.3. Clustering of wild-type and knock-out cell lines by stimulus

After ensuring the technical validity and robustness of the screen, transcriptional signatures were computed (see chapters 2 and 6) and a clustering approach was employed. As expected, t-distributed stochastic neighbour embedding (t-SNE) clustering of the transcriptional signature genes of wild-type and knock-out cell lines revealed eleven distinct clusters: ten stimuli clusters for the ten parallel screens and one cluster with unstimulated samples (Figure 3.5). Most knock-outs clustered together with wild-type samples treated with the same stimulus, suggesting that the majority of stimulated knock-outs displayed a transcriptional response comparable to stimulated wild-type cells. This result was expected, since most of the TKs screened

were not predicted to play a role in one of the ten pathways investigated, and hence, the respective signalling pathway should not be differentially affected between knock-out and wild-type. However, the fact that some stimulated knock-outs resembled a transcriptional profile comparable to unstimulated control samples could indicate that the investigated TK takes part in a particular signalling cascade, as the knock-out fails to elicit a transcriptional response upon stimulation.

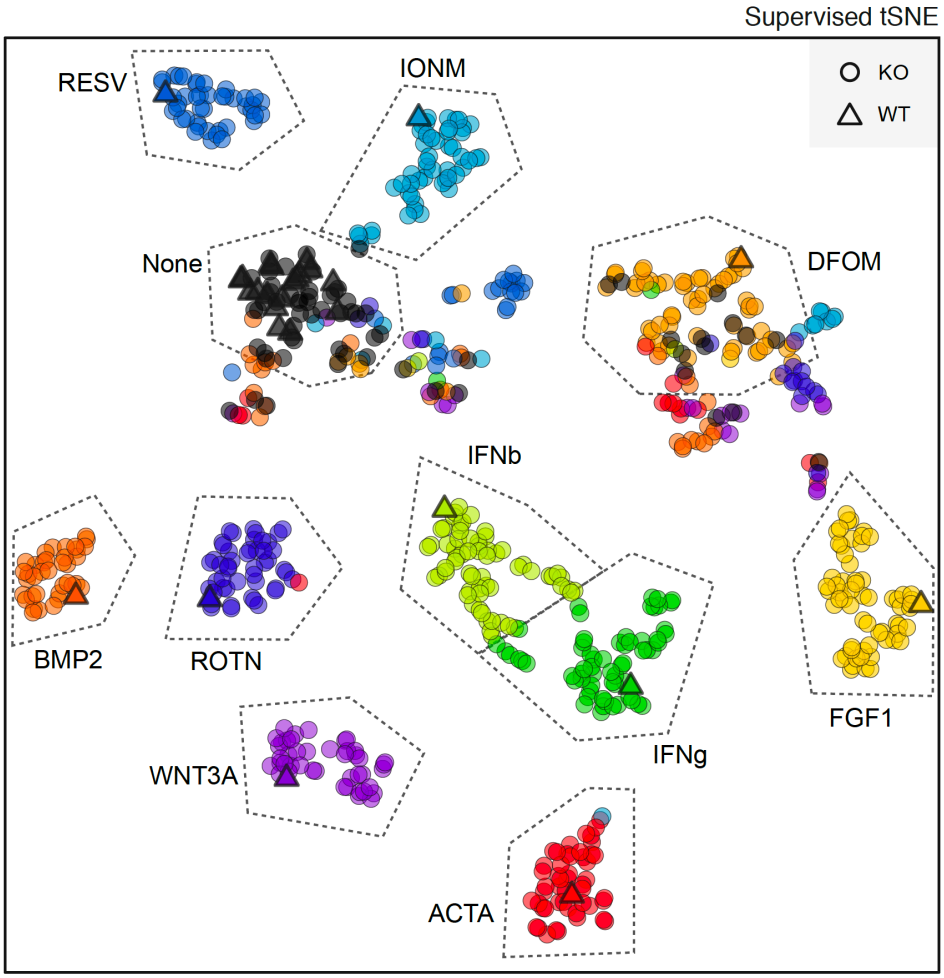


Figure 3.5 Clustering of the parallel reverse genetic screens. Supervised t-distributed stochastic neighbour embedding (tSNE) clustering of stimulated and unstimulated wild-type and knock-out cells. The ten polypeptides and small molecules used for screening are indicated. Triangles depict average wild-type transcriptional responses measured in triplicates, circles show average of the 64 knock-out responses measured in duplicates. Taken from Gapp et al., 2016.

In contrast to the other polypeptides and small molecules, the DFOM cluster was less defined, unexpectedly revealing that around 19% of unstimulated knock-out clones displayed a transcriptional response comparable to DFOM stimulation. This observation is discussed in more detail in section 3.2.6.

3.2.4. Technical covariates affecting the measurement of transcriptional responses

In order to avoid the scoring of false positive interactions between genes and pathways, technical parameters that could affect the assessment of transcriptional signature overlaps between stimulated knock-out and wild-type cells were investigated. Analysis of the 1700 signature gene sets showed weak correlations between transcriptional signature overlaps and technical covariates such as RNA concentration and sequencing depth (Appendix Figure 3.1). A negative correlation of the transcriptional signature overlap with RNA concentration suggests that the confluency of the collected samples could have influenced the transcriptional response to some of the tested stimuli. A positive correlation with sequencing depth is most likely a result of increased assay sensitivity when samples are sequenced deeper. In order to minimise the scoring of false positive interactions due to these technical confounding effects, general linear models (GLM) were created to correct for such effects. GLMs were used to fit the transcriptional signature overlap as a function of RNA concentration and sequencing depth. Residuals from these models were used to rank the transcriptional responses of individual knock-out cells to the ten stimuli. Small scores of stimulated knock-out cells suggest a dampened transcriptional response compared to wild-type cells, large response scores indicate an increased transcriptional response.

3.2.5. Knock-out cell responses to stimulation

Scoring the transcriptional responses of all knock-out cells using GLMs as described above revealed three response patterns: i) wild-type-like response to all stimuli, ii) dampened response to one or two specific stimuli, iii) reduced transcriptional response across many stimuli. Most knock-out cell lines fell in the first category showing a response pattern comparable to wild-type cells as observed before (Figure 3.5). 15 knock-outs (BMPR1A, CCK4, CTNNB1, EPHA4, EPHB4, ERBB2, FGFR1, FGFR3, HIF1A, JAK1, LMTK3, MET, PDGFRA, ROR2 and YES1) showed an abrogated transcriptional response to specific stimuli, and five knock-outs (ABL1, EPHB1, IGF1R, MERTK and TYK2) displayed a generally dampened response pattern (Appendix Figure 3.2). Knock-out cell lines were ranked based on the magnitude of the observed expression changes of signature genes upon stimulation compared to wild-type cells (Appendix Figure 3.3). Knock-outs showing the most reduced transcriptional response and therefore the smallest stimulus response score were selected for follow-up validation (see section 3.3.).

The response of all 64 knock-out cell lines upon stimulation with the ten polypeptides and small molecules is summarised in Appendix Figure 3.2 and Appendix Figure 3.3.

3.2.6. Hypoxic-like state as a passenger effect in screened knock-out cell lines

Unexpectedly, twelve out of 64 unstimulated knock-out cell lines used in the reverse genetic screen displayed a transcriptional signature comparable to DFOM-stimulated wild-type and knock-out cells (Figure 3.5 and Figure 3.6A). This included TKs belonging to various subfamilies with a seemingly random pattern and suggested that around 19% of the knock-outs had an activated hypoxic response under normoxic conditions. To test whether this hypoxic-like state was directly driven by the inactivation of TK genes, independent knock-out clones were generated for FGFR3 and PDGFRA as representative examples. gRNAs targeting a different exon (Table

6.1) than in the original cell lines were used to exclude potential off-target or clone-specific effects.

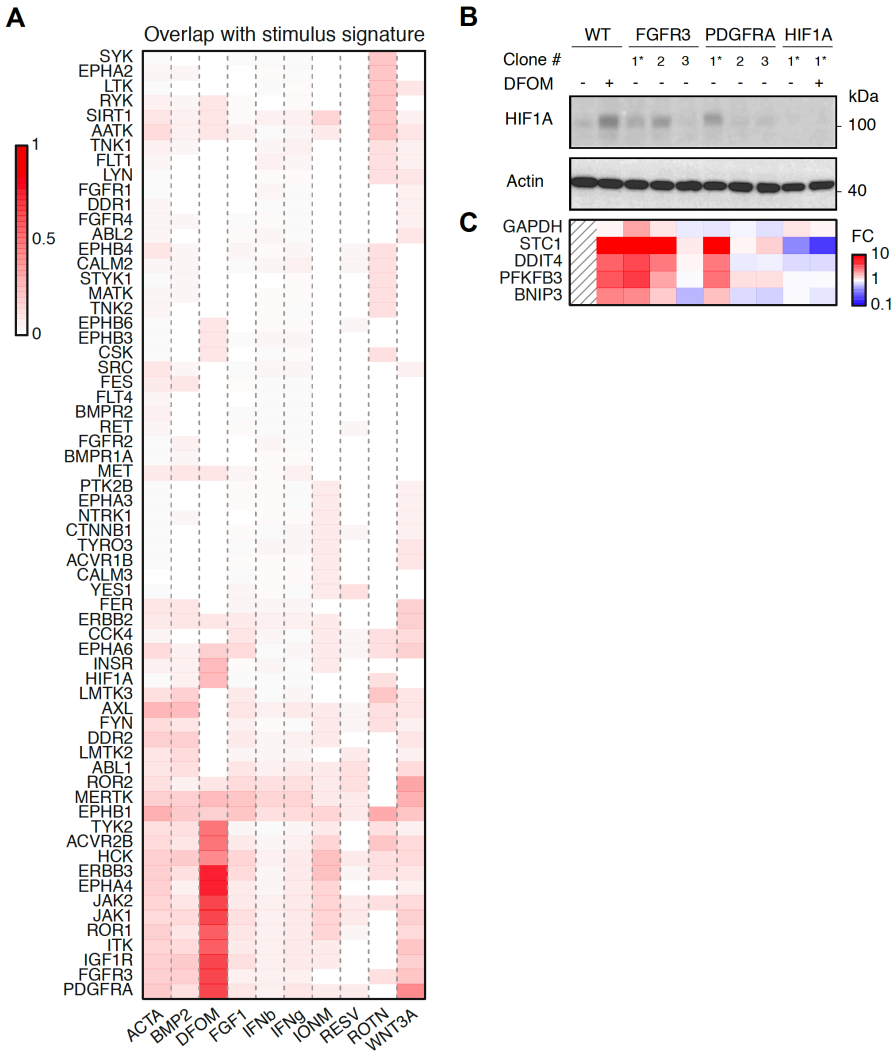


Figure 3.6 Unstimulated knock-out cell lines display a hypoxic-like response. A. Comparison between transcriptional profiles of the 64 knock-out cell lines used for reverse genetic screening with the transcriptional signatures induced in wild-type cells upon stimulation. Knock-out cell lines on the y-axis are clustered according to the overlap between the unstimulated profiles and stimulus-induced signatures. B. HIF1A expression levels in wild-type and knock-out cell lines used for reverse genetic screening and independently generated knock-out lines. Asterisks indicate knock-out cell lines used for screening. Actin levels serve as a loading control. C. qRT-PCRs of four hypoxia responsive genes of cell lines as in B. Expression in unstimulated wild-type cells was used for normalisation and fold change (FC) calculation, GAPDH was measured as a housekeeping control. Taken from Gapp et al., 2016.

Consistent with an activated hypoxia response, the original FGFR3 and PDGFRA knock-out cell lines used for reverse genetic screening indeed displayed increased HIF1A protein levels and increased expression of hypoxia-responsive genes, as confirmed by Western blot and qRT-PCR, respectively (Figure 3.6B and C). However, this effect was not observed in all the independently generated knock-out clones. This suggests that the activated hypoxic response detected in the reverse genetic screen is a frequently occurring passenger effect.

3.3. Transcriptional profiling of kinase knock-outs links genotypes to pathways

3.3.1. Comparison of stimulus responses measured by RNA-sequencing and qRT-PCR

To assess whether the identified gene-stimulus interactions were reproducible in independent experiments and using a different technology for the read-out, the top three knock-out/stimulus pairs showing the smallest response score for every screen (Appendix Figure 3.3) were selected and validated by qRT-PCR (Figure 3.7 and Appendix Figure 3.4). Two to five signature genes displaying a reduced expression level after stimulation (at least two-fold) in RNA-sequencing in knock-outs compared to wild-type were selected for qRT-PCR measurements. Stimulation experiments of the TK knock-out and control cell lines used in the screen along with wild-type controls were performed as described for the reverse genetic screen. Expression changes of signature genes between wild-type and selected knock-outs as measured by RNA-sequencing and qRT-PCR were quantitatively consistent (Figure 3.7). Even small differences in stimulus response were consistent between the two assays. Among the validated hits were interactions previously described in the literature such as HIF1A KO with DFOM, CTNNB1 KO with WNT3A, JAK1 KO with IFNb and IFNg, as well as FGFR family member KOs with FGF1, validating the transcriptome based

reverse genetic approach (discussed in more detail in sections 3.3.2. and 3.3.3.). In addition, a number of unexpected, novel interactions were found (see section 3.3.4.). In summary, of 33 gene-stimulus combinations tested, 26 could be validated in independent stimulation experiments using the same knock-out clones as in the genetic screen and using qRT-PCR as an independent read-out.

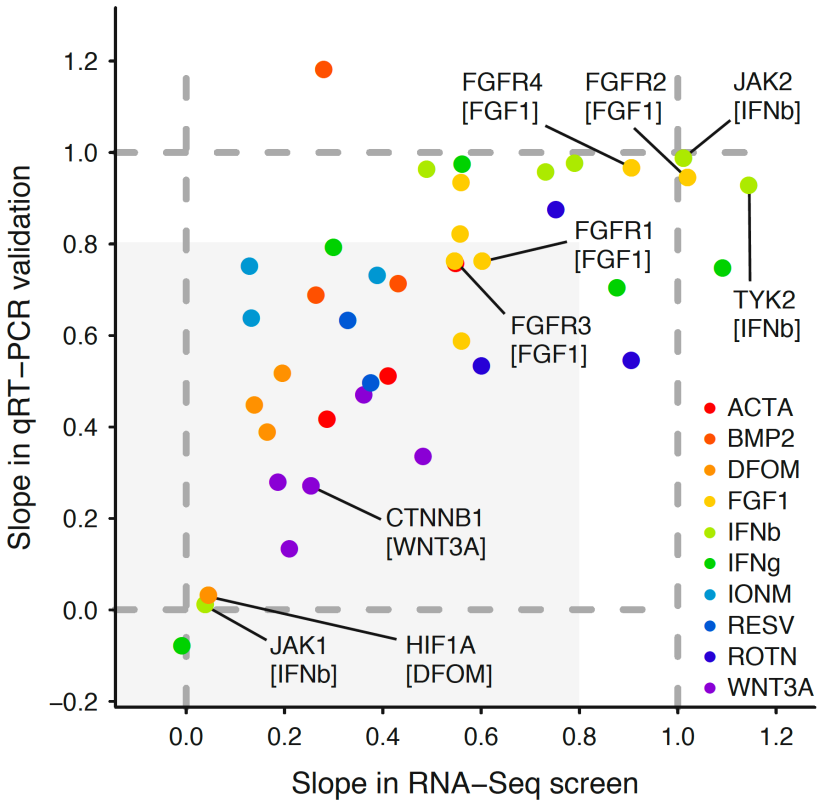


Figure 3.7 qRT-PCR validations of the top interactions identified in the ten reverse genetic screens. Comparison of stimulus responses between wild-type and knock-out cells. x-axis gives the slope of the best fit line for transcriptional signature gene expression in wild-type vs knock-out cells as determined by RNA-sequencing, y-axis shows the slope of the best fit line of the same genes in wild-type vs knock-out cells as measured by qRT-PCR. Dashed lines are guides indicating a hypothetical equal response of wild-type and knock-out cells and fully unresponsive knock-out clones. Gray shaded area indicates that both assays show a reduced transcriptional response in knock-out compared to wild-type cells. Selected gene-stimulus interactions are indicated. Taken from Gapp et al., 2016.

3.3.2. Response of JAK family member knock-out cells to IFN stimulation

The receptor-associated JAK family mediates signalling via a variety of surface receptors including interferon type I and type II receptors (Rane and Reddy, 2000). In HAP1 cells, three family members (JAK1, JAK2, TYK2) are expressed and non-essential and were therefore included as knock-out cell lines in the screen. Since JAK1 and TYK2 are known to be critical for interferon type I (IFNb) signalling, and JAK1 and JAK2 for type II (IFNg) signalling (Rane and Reddy, 2000), stimulation of the JAK knock-out panel with IFNb and IFNg was expected to yield a reduced transcriptional response upon stimulation in all JAK family knock-outs. Surprisingly, only JAK1 knock-out cells showed a completely abrogated transcriptional response upon stimulation with IFNb and IFNg, while the other family members responded similarly to wild-type cells (Figure 3.8). Interferon type I and type II signalling in HAP1 cells therefore seems to be strictly dependent on JAK1, at least under the experimental conditions used. This observation was confirmed in independent experiments using either RNA-sequencing or qRT-PCR as a read-out (Figure 3.8C-D and Figure 3.9). However, it cannot be excluded that JAK2 and TYK2 impinge on interferon signalling at a different time point or under a different experimental condition.

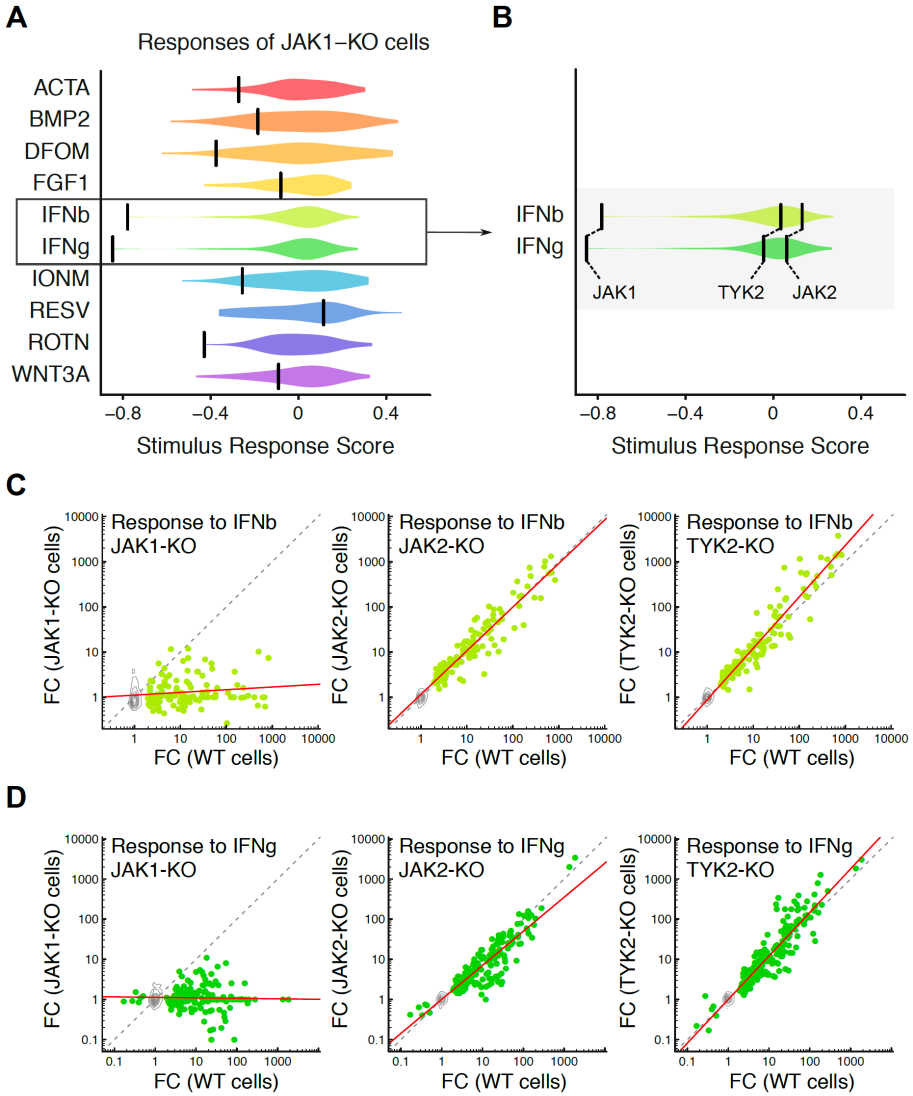


Figure 3.8 JAK1 knock-out cells are insensitive to IFNb and IFNg stimulation. A. JAK1 knock-out cell response across the ten reverse genetic screens. Stimulus response scores represent transcriptional signature overlaps between wild-type and knock-out cells corrected for technical parameters using general linear models. Black bar indicates the stimulus response score for JAK1 knock-out cells. B. As in A for JAK family member responses to IFNb and IFNg. C. Transcriptional responses of wild-type and JAK family knock-outs upon stimulation with IFNb. Fold change (FC) in gene expression between stimulated and unstimulated wild-type cells is plotted against fold change in gene expression between stimulated and unstimulated knock-out cells. Dots represent transcriptional signature genes, contour indicates genes not differentially expressed. Gray dashed line shows the diagonal, red line is the best fit line for signature genes. D. As in C for response of JAK family members to IFNg. WT, wild-type; KO, knock-out. Adapted from Gapp et al., 2016.

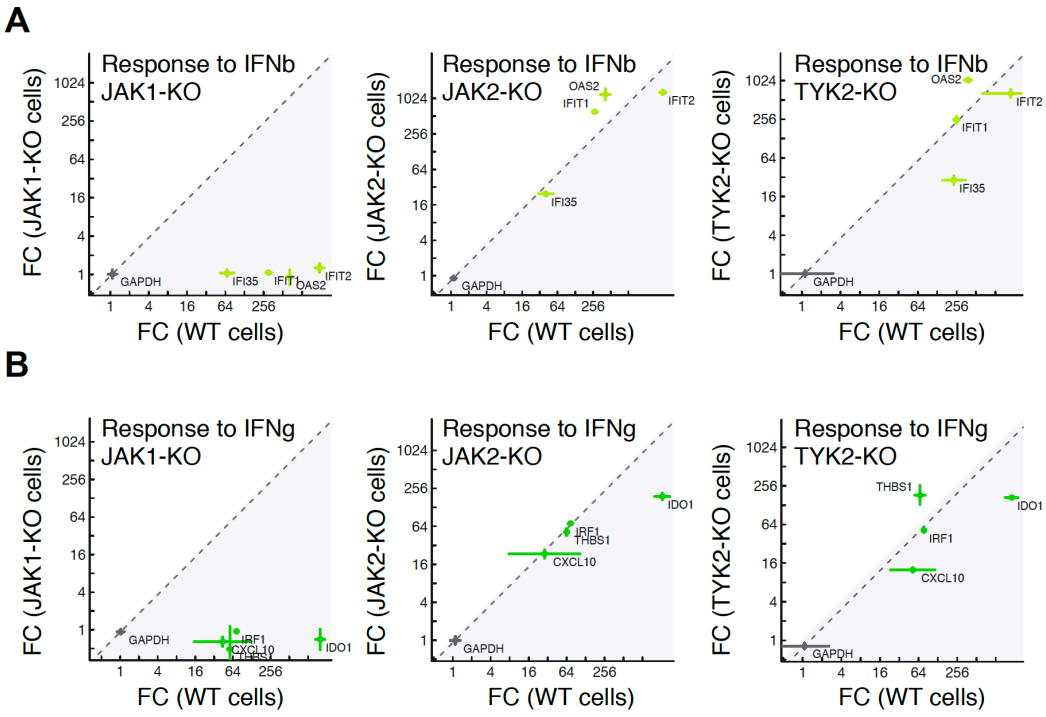


Figure 3.9 qRT-PCR validation of JAK family knock-outs. A. Transcriptional response of JAK1, JAK2 and TYK2 knock-out cells upon stimulation with IFNb. Fold change (FC) of four transcriptional signature genes (green) and GAPDH (gray) upon stimulation with IFNb are compared between wild-type and indicated knock-out cells. Error bars represent the standard deviation of three technical replicates. B. As in A for IFNg stimulus. WT, wild-type; KO, knock-out.

3.3.3. Response of FGFR family member knock-out cells to FGF1 stimulation

Multiplexed RNA-sequencing also revealed small differences in the highly redundant FGF pathway, where FGF1 can bind to and signal via four different FGFRs in a context-dependent manner (Figure 3.10) (Raju et al., 2014). While FGFR1 and FGFR3 knock-out cells showed an attenuated transcriptional response to FGF1, FGFR2 and FGFR4 behaved like wild-type cells. As all four receptors are highly expressed (FPKM: 51.9 (FGFR1), 15.8 (FGFR2), 3.7 (FGFR3), 11.5 (FGFR4)), this result could indicate that the response to FGF1 in HAP1 cells is mediated via FGFR1 and FGFR3. FGFR1 knock-outs showed a uniform reduction in the signature gene expression, while this was not the case for FGFR3 knock-outs.

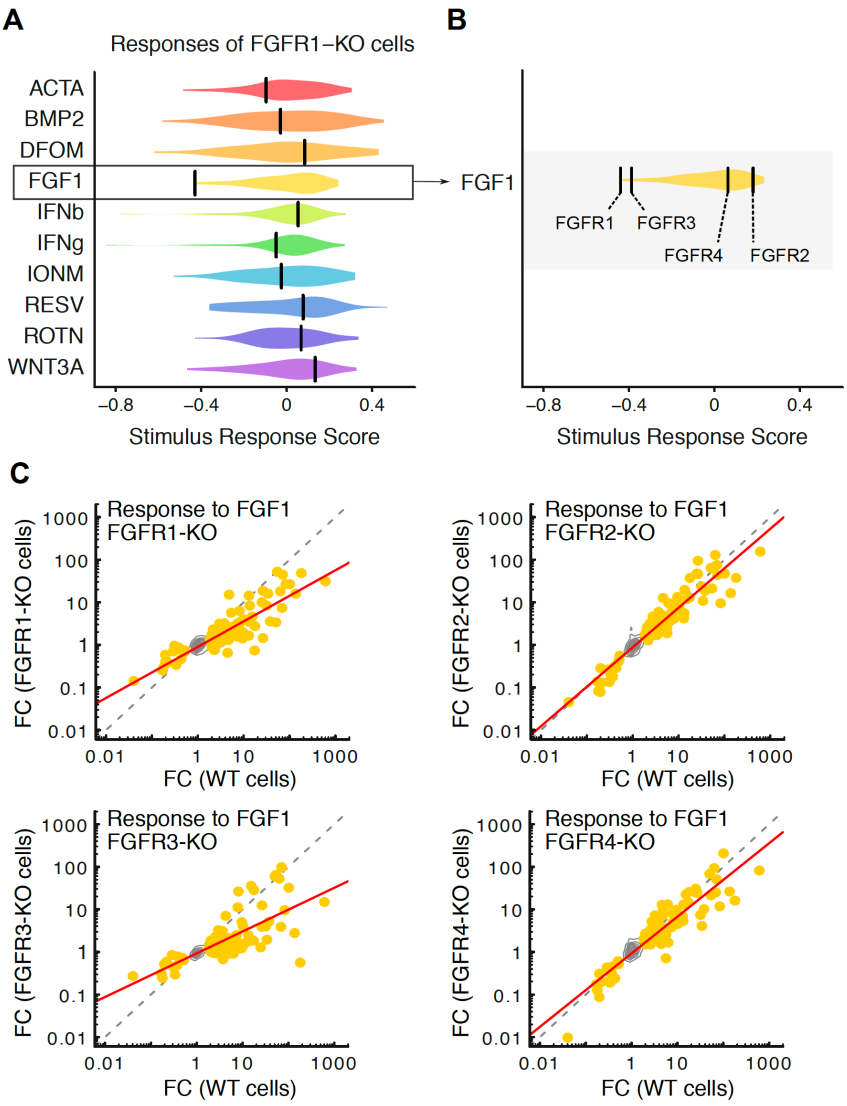


Figure 3.10. FGFR1 and FGFR3 knock-outs show a dampened transcriptional response upon FGF1 stimulation. A. FGFR1 knock-out cell response across the ten reverse genetic screens. Stimulus response scores represent transcriptional signature overlaps between wild-type and knock-out cells corrected for technical parameters using general linear models. Black bar indicates the stimulus response score for FGFR1 knock-out cells. B. As in A for FGFR family member responses to FGF1. C. Transcriptional response of wild-type and FGFR family knock-outs upon stimulation with FGF1. Fold change (FC) in gene expression between stimulated and unstimulated wild-type cells is plotted against fold change in gene expression between stimulated and unstimulated knock-out cells. Dots represent transcriptional signature genes, contour indicates genes not differentially expressed. Gray dashed line indicates the diagonal, red line is the best fit line for signature genes. WT, wild-type; KO, knock-out. Adapted from Gapp et al., 2016.

The dampened transcriptional response of FGFR1 and FGFR3 knock-out cells in response to FGF1 stimulation could be reproduced in independent stimulation experiments and with a qRT-PCR read-out (Figure 3.11). These results illustrate the complexity of FGF1 signalling events and show that the established transcriptional profiling platform (chapter 2) can be used to study even a highly redundant pathway.

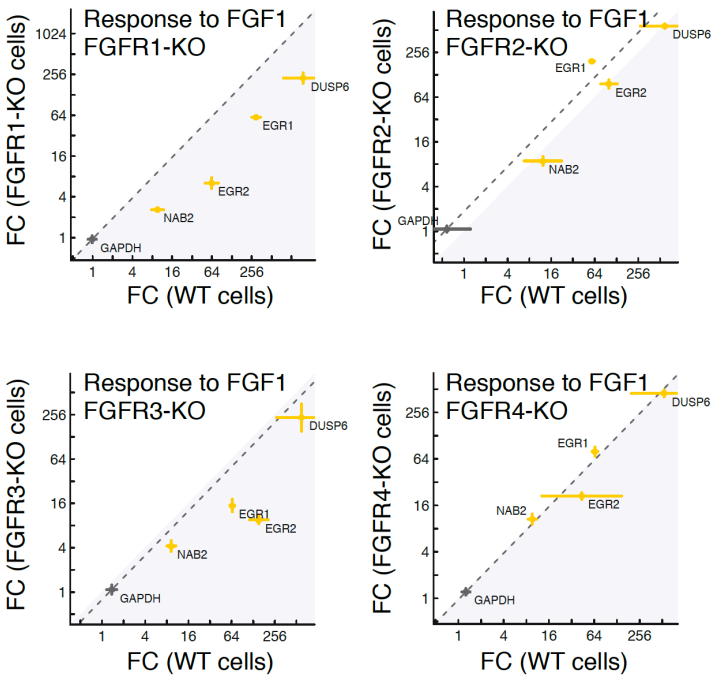


Figure 3.11 qRT-PCR validation of FGFR knock-outs. Transcriptional response of FGFR family knock-out cells upon stimulation with FGF1. Fold change (FC) of four transcriptional signature genes (yellow) and GAPDH (gray) upon stimulation with FGF1 is compared between wild-type and indicated knock-out cells. Error bars represent the standard deviation of three technical replicates. WT, wild-type; KO, knock-out.

The fact that the presented reverse genetic approach could be used to link JAK1 with interferon signalling and FGFR1 and FGFR3 with FGF signalling confirms that transcriptional profiling can be used to functionally annotate genes in pathways even in a large screening setting.

3.3.4. Validation using independently generated knock-outs

As mentioned previously, several gene-stimulus connections not yet or poorly described in literature were detected and validated by qRT-PCR. Among them were three TKs that showed an interaction with one stimulus each as well as five TKs that displayed an interaction with many stimuli (Appendix Figure 3.4 and Table 3.1). A diminished transcriptional response of the TK knock-outs compared to wild-type cells upon stimulation could be indicative for the involvement of these TKs in the stimulated pathways.

Table 3.1 Candidate gene-stimulus interactions

Cell line	Gene	Stimulus
CCK4	Colon carcinoma kinase 4	RESV
ERBB2	ErB-B2 receptor tyrosine kinase 2	IONM
FGFR3	Fibroblast growth factor receptor 3	WNT3A
PDGFRA	Platelet derived growth factor receptor alpha	DFOM, ROTN
EPHB1	Ephrin receptor B1	ACTA, IFNg, WNT3A
IGF1R	Insulin like growth factor 1 receptor	ACTA, DFOM, FGF1, WNT3A
MERTK	MER proto-oncogene tyrosine kinase	IFNb, IONM, RESV
TYK2	Tyrosine kinase 2	ACTA, BMP2, FGF1

In order to exclude potential off-target or clone-specific effects, independent knock-out clones were generated for eight genes using a lentiviral CRISPR construct. gRNAs were designed to target CCK4, ERBB2, PDGFRA, FGFR3, EPHB1, IGF1R, MERTK and TYK2 (Table 3.1 and Table 6.1). Compared to the TK knock-out clones used in the reverse genetic screen, different exons were targeted to generate truly independent gene knock-outs. Indels introduced using CRISPR/Cas9 were confirmed by Sanger sequencing. Two clonal cell lines harbouring frameshift mutations could be obtained for each of the targeted genes except for ERBB2 where only one knock-out cell line could be established. Stimulation experiments of the original and new

knock-out cell lines as well as wild-type controls were performed in parallel as described for the screen. Transcriptional response was measured by qRT-PCR and fold changes upon stimulation were compared between knock-outs and wild-types (Figure 3.12 and Appendix Figure 3.5).

Only the interaction between FGFR3 and WNT3A could be confirmed in both independently generated clones (Figure 3.12). One independent FGFR3 knock-out clone showed an abrogated response that was more pronounced than in the screening clone, the other one displayed only a slight reduction comparable to the original clone.

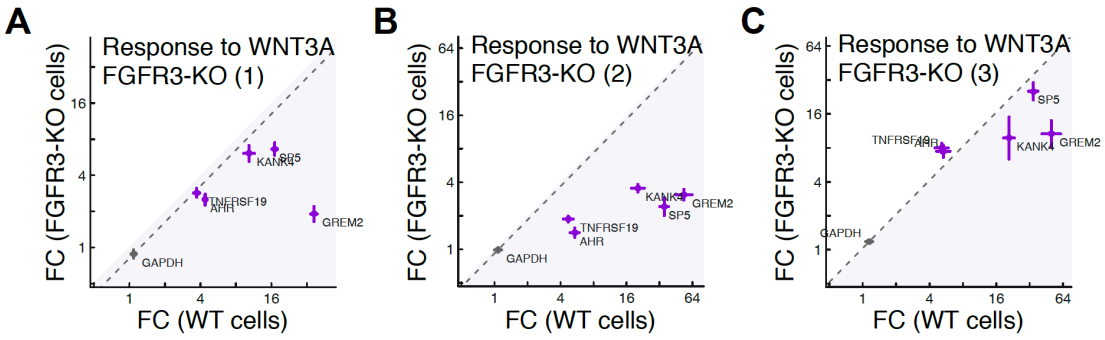


Figure 3.12 qRT-PCR validation of FGFR3 knock-outs in response to WNT3A. A. Transcriptional response of FGFR3 knock-out cells used for reverse genetic screening upon stimulation with WNT3A. Fold change (FC) of five transcriptional signature genes (purple) and GAPDH (gray) upon stimulation is compared between wild-type and knock-out cells. Error bars represent the standard deviation of three technical replicates. B. and C. as in A for an independently generated knock-out clone. WT, wild-type; KO, knock-out.

A dampened transcriptional response of CCK4 knock-out cells in response to RESV (Appendix Figure 3.5A) and PDGFRA knock-outs with ROTN was only seen in one of the additional clones (Appendix Figure 3.5D). The interactions between TYK2 and BMP2 as well as FGF1 were validated only in one of the tested additional clones and were much less pronounced for both stimuli than in the screening clone (Appendix

Figure 3.5Q-R). Thus, except for the interaction between FGFR3 and WNT3A, the observed reduction in transcriptional response for the tested knock-out clones (Table 3.1) does not seem to be causally driven by the respective gene deficiency. This failure to confirm the suggested gene-stimulus interactions in almost all independently generated clones rather hints towards substantial variability in the stimulus responses between different clonal cell lines. However, a definite conclusion is non-trivial to draw from the available data because the differences between the transcriptional response in wild-type and knock-outs observed in the screen were quantitatively small. Thus, they might be potentially difficult to robustly validate in a cellular system that is notoriously noisy. Further aspects of reproducibility in this setting will be discussed below.

3.4. Discussion

This chapter demonstrates the suitability of the established shallow RNA-sequencing platform (see chapter 2) as a large-scale gene annotation tool in haploid human cells. Parallel reverse genetic screens with ten different stimuli across 64 knock-out cell lines were conducted to functionally characterise kinase genes in signalling pathways via transcriptional profiling.

The parallel screening strategy presented in this chapter has several desirable features. First, most transcriptional signature comparisons were between samples within the same batch, circumventing the need for explicit batch correction. Further advantages of this screening design are its scalability and modularity. These allow expansion of transcriptome based reverse genetic screening beyond the analysis of TKs, for example to functionally annotate all non-essential kinases or other enzyme families and eventually all genes for which knock-outs can be generated.

The comparison of knock-out to wild-type cells revealed a number of technical parameters that introduced a bias in the scoring of reduced transcriptional responses. For instance, in some cases the measured RNA concentrations correlated with the amount of transcriptional signature overlap between wild-type and knock-outs. To minimise the detection of false positive interactions as a result of such confounding effects, the computational analysis employed GLMs to correct for these technical variations. These results highlight the importance of recording technical parameters and controlling for confounding effects.

By comparing transcriptional signatures of stimulated wild-type cells to isogenic TK knock-out cell lines, the involvement of particular kinases in specific pathways was demonstrated and known and unexpected interactions were revealed.

Using an isogenic cell line model covering all non-essential and expressed TKs constitutes a unique toolset for studying the impact of genotype on phenotype. The generation of TK knock-outs by CRISPR/Cas9 in contrast to partial RNAi knock-down allowed the systematic investigation of the contribution of this important class of enzymes to particular pathways.

The presented results indicate that, as expected, most knock-out cell lines responded similarly to wild-type cells upon stimulation. Only few knock-outs showed a dampened transcriptional response, suggesting the involvement of a particular TK in the stimulated pathway. For example, the loss of the cytoplasmic TK JAK1, but not the other family members JAK2 or TYK2, completely abrogated the response to interferon beta and gamma signals. This was surprising, as all three JAK family members are linked to interferon signalling (Rane and Reddy, 2000), and suggests a distinct function of JAK1 compared to the other two family members, at least in HAP1 cells.

Another example of an unexpected finding in a well-studied pathway was the diminished transcriptional response to FGF1 in knock-outs of the corresponding

receptors FGFR1 and FGFR3 but not FGFR2 and FGFR4. This suggests that the response to FGF1 is predominantly mediated by only two out of four FGFRs. Furthermore, only FGFR3 knock-out cells, but none of the other FGFR family members, displayed a dampened transcriptional response to WNT3A. Several studies have highlighted extensive crosstalk between RTKs and the WNT/CTNNB1 pathway and some studies indicate the involvement of FGFRs in the activation of the WNT/CTNNB1 pathway (Dailey et al., 2005; Krejci et al., 2012; Buchtova et al., 2015). This crosstalk between FGFR3 and WNT/CTNNB1 signalling seems to be conserved in HAP1 cells.

These findings highlight that transcriptional profiling can generate new hypotheses even for well-studied signalling cascades like JAK/STAT and FGF, and demonstrate that it can be used to study seemingly redundant signalling modules or crosstalk between pathways.

Subsequent validation of previously undescribed interactions between TKs and certain stimuli was less successful. Poor reproducibility between the knock-out clones used for reverse genetic screening and independently generated clones indicates tremendous variability of transcriptional responses between cell lines derived from single cell clones. This is supported by single cell sequencing studies, from which the magnitude of cellular heterogeneity within a pool of seemingly equivalent cells has become increasingly evident (Navin, 2015; Gawad et al., 2016). Even before, clonal heterogeneity has been noted as a phenomenon that can complicate *in vitro* experiments but also underlies disease, drug resistance and evolution (Navin, 2015; Yi et al., 2017).

The heterogeneity in the transcriptional response measured in this study could derive from genetic or epigenetic differences between different clones, which could have a larger impact in haploid genomes (Leeb and Wutz, 2013). Another indication for clonal variability is the slightly activated hypoxic response that is only seen in some

of the isogenic cell lines. In this case, it is tempting to speculate that the observed hypoxic passenger effect is selected for: the isogenic cell lines are generated by seeding single cells into separate wells; they are subsequently grown until macroscopic colonies emerge, creating a hypoxic environment for cells within such a colony. Hence, cells that are able to cope better in hypoxic conditions might have a growth advantage, possibly explaining why clones with a slightly activated hypoxic response make up around 20% of all clones used in the screen.

One drawback of a collection derived from single cell clones is that the generation and isolation of individual mutants is a laborious process. This could be circumvented by generating a polyclonal population with CRISPR/Cas9, introducing a range of different lesions in the target gene in different cells (Shalem et al., 2015). Although this approach would allow to generate mutant cell lines more effectively, a problem of such a mixed mutant population of cells is that it cannot be considered a true, complete knock-out cell line, as some cells will harbour wild-type alleles. At the same time, the use of a polyclonal population should enable to overcome the disadvantages of heterogeneity associated with single cell derived cell lines.

Another improvement to the presented strategy could be the generation of mutant clones or populations by using multiple unique gRNAs targeting the same gene. This should increase the yield of true positive interactions, reduce false positives and speed up follow-up validations. Considering only hits that show the same phenotype with different gRNAs should allow the identification of high confidence interactions.

Finally, an alternative to the laborious and expensive qRT-PCR validations carried out in this study would be targeted RNA-sequencing. Sequencing a focused set of transcripts (e.g. transcriptional signature genes) would enable validation experiments with increased throughput and lower cost (Van Dijk et al., 2014).

Taken together, the presented project is a proof-of-concept study for a transcriptome based screening method with the potential of becoming a high-throughput reverse genetic screening tool for functional gene annotation in human cells. Here, a collection of haploid knock-out cell lines is used to demonstrate the feasibility of the transcriptional screening approach. However, the presented strategy should be applicable to other human cell systems as well, even though the generation of knock-outs with CRISPR/Cas9 in diploid cells may be less efficient. In addition, shallow RNA-sequencing may perform less well using other diploid cell systems (Gapp et al., 2016). Slight adjustments to the number of samples sequenced in parallel should allow to compensate for the potential confounding effects of sequencing transcriptomes of diploid versus haploid cells.

Transcriptome based screening should also allow to assess phenotypes not only of knock-outs but also of engineered point mutants or chemical agents. In addition to functionally annotating genes in pathways, the presented strategy could be employed to identify negative regulators that result in increased signalling when perturbed. More generally, hit validation or targeted hypothesis testing could be further areas of application. For example, advances in sequencing technologies have resulted in finding associations of mutations with human diseases on large scales but causal links are missing. This is especially true in the field of genome-wide association studies, which associate germline genetic variants with diseases (Lowe and Reddy, 2015). In these cases, the presented method could address the need to validate and test these associations, provided that the relevant reference/control signatures are available.

Collectively, this study presents, tests and validates a functional genetic screening tool based on transcriptional profiling that is generic and will allow the study of a plethora of other cellular systems and biological questions.

Chapter 4. High-throughput barcode-sequencing for pooled drug sensitivity screening in human cells

4.1. Introduction

This chapter provides an introduction to pooled screening approaches and further elaborates on compound MoA screening in human cells in addition to chapter 1. The results section of this chapter describes the establishment of a pooled chemical genetic approach to investigate thousands of potential gene-drug interactions in an isogenic human cell line model.

4.1.1. Pooled screening approaches in cultured human cells

Although arrayed screening approaches are extremely valuable, as they allow the investigation of complex phenotypes (e.g. morphology, transcriptional signatures), they can be tedious and expensive, since every sample (e.g. knock-out cell line) has to be handled separately. In contrast, pooled screening techniques allow the investigation of thousands of samples across many genetic backgrounds and conditions simultaneously. Pooled screens usually assess phenotypes that are robust and cost-effectively detectable, such as growth and viability or “simple” reporter based assays. Another advantage of pooled screens over arrayed approaches is that all cells are cultured under uniform conditions. Furthermore, pooled screening offers easier scalability and can be performed faster and cheaper (Mohr et al., 2014; Shalem et al., 2015).

Generally, pooled screens start with the packaging of a mutagen library (RNAi, gene trap, CRISPR) into a virus and subsequent infection of a population of target cells under conditions where each transduced cell will harbour one mutagenic element (siRNA, shRNA, gene trap insertion, gRNA) on average. Cells exhibiting the phenotype of interest are then selected, the mixed population is subjected to massively parallel sequencing and enriched or depleted clones are identified (Mohr et al., 2014; Hartenian and Doench, 2015; Shalem et al., 2015; Nijman, 2015).

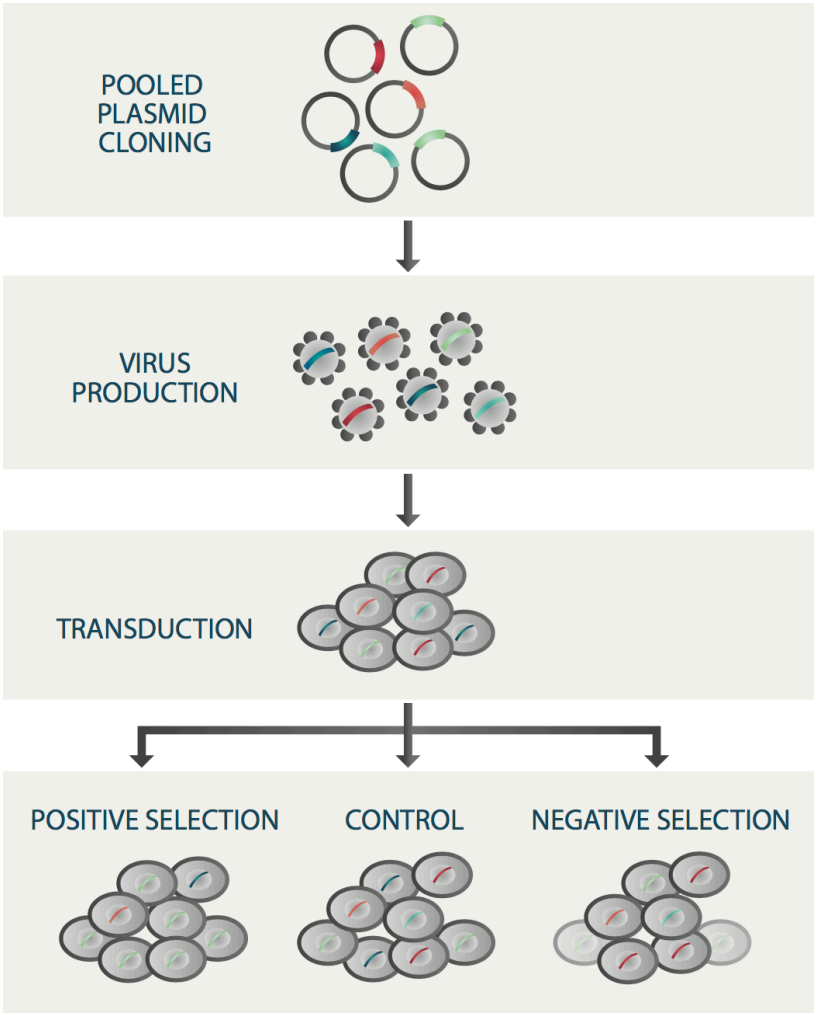


Figure 4.1 Pooled screening. Mutagen libraries are constructed in a pooled fashion, packaged into viruses and a cell population is infected to harbour one integration per cell. Following positive or negative selection, viral integration sites are amplified by PCR and products are subjected to massively parallel sequencing. Comparing cell populations upon perturbation to untreated control samples allows to identify genes underlying a particular phenotype. Adapted from Hartenian and Doench, 2015 and Shalem et al., 2015.

Pooled genetic or compound screens can be carried out in the form of positive or negative selection screens. Positive selection implies the identification of mutants refractory to treatment while negative selection or “drop-out” screens are based on determining the absence of mutants after a certain treatment. Both types of screens are extremely valuable for the identification of true positive hits, but drop-out screens

are technically more challenging, since most cell lines in a pool will not respond to a particular treatment and only few clones will be absent (or reduced) from the complex mixture. Hence, control samples and deep sequencing are required to confidently call the absence of a few samples from a complex mixture of cells. Other important considerations for pooled screening approaches are the appropriate library representation to achieve low signal-to-noise ratios and that some interactions may be missed (false negatives), for example due to the insufficient knock-down or knock-out of genes (Mohr et al., 2014; Hartenian and Doench, 2015; Shalem et al., 2015).

The combination of haploid cells with insertional mutagenesis provides an attractive possibility for pooled screening (see section 2.1.2.). A pooled knock-out population is generated using retroviral integration of a gene trap vector in a quasi-random manner, and gene trap insertions are recovered by massively parallel sequencing (Carette et al., 2009; Carette et al., 2011a; Blomen et al., 2015; Wang et al., 2015). Pooled genome-wide CRISPR screening in haploid, diploid or polyploid cancer cells is achieved by packaging pooled gRNA plasmid libraries into viruses and subsequent infection of target cells (Wang et al., 2014; Shalem et al., 2014; Hart et al., 2015; Hartenian and Doench, 2015; Shalem et al., 2015). Pooled screens can also be carried out using focused libraries targeting gene families such as kinases or functional protein domains (Wang et al, 2014; Shi et al., 2015). Screens with focused libraries can be performed in a smaller scale, thereby using fewer cells and less reagents.

Large-scale analysis of pooled cell lines is facilitated by the introduction of molecular barcodes that serve as a tag to monitor the presence of each line in a mixed population. Massively parallel sequencing of the barcodes allows to monitor the response of mixed populations to a perturbation, thereby enabling the investigation of gene combinations or gene-drug interactions in a pooled fashion. This approach is

very versatile, since it can be combined with insertional mutagenesis (gene trap, transposon), CRISPR libraries (knock-out, gene activation or repression), cDNA libraries, small molecule inhibitors or RNAi (Nijman, 2015). Recently, pooled barcoded screening was applied to study molecular circuitries in immune cells or to investigate the unfolded protein response by combining CRISPR/Cas9 with single cell RNA-sequencing (Dixit et al., 2016; Adamson et al., 2016).

Pooled screening approaches in human cells have proven extremely valuable to identify genes essential for viability or offer candidates for anticancer drug development (Gilbert et al., 2014; Blomen et al., 2015; Wang et al., 2015; Hart et al., 2015). The experimental part of this chapter utilises a barcoded human knock-out collection for pooled compound MoA screening.

4.1.2. Compound mode-of-action (MoA) screening in human cells

Genetic screening in combination with small molecules (chemical genetics) allows the identification of drug targets and MoA of compounds. As described in chapter 1, target identification and the better understanding of MoAs aids to uncover applications for approved and novel drugs, explain resistance mechanisms or to reduce side effects. Chemical genetics can be also leveraged to functionally annotate genes or to identify novel gene-drug interactions that might be exploited for therapeutic intervention (Nijman, 2015).

Chemical genetic screening can be carried out in arrayed or pooled formats and typically involves cell lines harbouring genetic alterations or deliberately mutagenised populations and subsequent application of compounds for several days. The phenotype (e.g. viability, morphology, transcriptional signatures) upon drug treatment is determined relative to samples treated with the drug solvent (“vehicle”) as controls (Grimm, 2004). One advantage of using drugs for screening is that the findings might be more easily translated into clinical applications. A disadvantage is that it is not

possible to perform fully unbiased, genome/proteome-wide compound screens because specific, potent and cell-permeable small molecules do not exist for all proteins. Another shortcoming is that many small molecules are unspecific and hit several targets, which can complicate the interpretation and follow-up experiments (Jones and Bunnage, 2017).

Nonetheless, chemical genetics originally pioneered in yeast (Ho et al., 2011) has provided ample insight into compound MoA. Chemical genetic screens can reveal the direct drug target or modifiers of the pathway affected by drug application. RNAi and gain-of-function screens have been extensively used to investigate the MoA of anticancer agents (Nijman, 2015). Also haploid genetic screens have been used for chemical screening and have, for example, identified genes underlying the MoA of the adjuvant Leu-Leu-OMe or the cytotoxic agents 3-bromopyruvate and YM-155 (Jacobson et al., 2013; Birsoy et al., 2013; Winter et al., 2014). Similarly, CRISPR/Cas9 screens have been performed to investigate the action of small molecules and have led, for instance, to the discovery of novel mechanisms underlying BRAF inhibitor resistance (Shalem et al., 2014).

Screens relying on the principle that compounds with a similar MoA tend to show similar phenotypes have been employed to reveal compound MoA, resistance mechanisms, drug synergisms and off-target effects (Nijman, 2015). Several studies have shown the feasibility of gaining insight through such “guilt-by-association” screens using transcriptional profiles, genetic interaction patterns or imaging as read-outs (Perlman et al., 2004; Lamb et al., 2006; Muellner et al., 2011; Jiang et al., 2011; Breinig et al., 2015; Martins et al., 2015).

Together, genetic screening and chemical genetics in human cells have provided valuable insight into the molecular mechanisms underlying healthy and diseased states and have identified effective targets for the treatment of various diseases.

However, in order to improve and ultimately guide therapy development and clinical applications, a more in depth understanding of cellular circuitries and compound MoAs is required.

Hence, the following section describes the development of a pooled forward genetics approach with the goal to establish a method to explore novel gene-drug interactions and MoAs of uncharacterised compounds.

4.2. A platform for pooled forward genetic screening by highly multiplexed barcode-sequencing

4.2.1. A collection of 4958 barcoded knock-out cell lines

Insertional mutagenesis screens in human near-haploid cells have enabled unbiased forward genetic screens and have provided insight into the action of compounds and bacterial toxins (Carette et al., 2009; Carette et al., 2011a; Jacobson et al., 2013; Birsoy et al., 2013; Winter et al., 2014). However, such screens require large cell numbers and as a result have been mainly carried out in the form of positive selection screens. Chemical genetic screens with pooled knock-out collections consisting of characterised barcoded clones require less cells and allow easy tracking of individual clones in a complex pool (Nijman, 2015). Using a collection of barcoded human knock-out cells in combination with compounds should enable cost-effective positive and negative selection screens and allow the efficient assessment of compound MoA.

For the implementation of a highly multiplexed pooled screening platform to investigate thousands of gene-drug interactions simultaneously, a defined KBM7 knock-out pool was assembled. 4958 commercially available isogenic KBM7 knock-out clones were obtained. This collection comprises clones harbouring inactivating mutations of 2144 different genes, i.e. some genes are represented more than once. Individual KBM7 knock-outs had been generated using a retroviral gene trap

mutagenesis approach and were tagged with a molecular barcode flanked by universal primer binding sites to enable the analysis (Bürckstümmer et al., 2013). In this study, all clones were sequentially pooled together in equal ratios. Finally, this pooled cell collection was employed to monitor the viability of thousands of knock-outs in response to drug treatment simultaneously, using massively parallel sequencing as a read-out.

4.2.2. Generation and validation of HSV-TK expressing cell lines

A high assay sensitivity is crucial for pooled screening approaches to accurately determine the presence or absence of single clones in a mixed population. The latter is more challenging to detect, because only a minor fraction of clones will be absent from the pool after treatment. To this end, two barcoded clones were engineered to ectopically express the herpes virus thymidine kinase (HSV-TK). The non-toxic pro-drug ganciclovir is metabolised into a cytotoxic agent only by HSV-TK but not the human thymidine kinase homologs and therefore specifically kills HSV-TK expressing cells (Ardiani et al., 2012). KBM7 wild-type and HSV-TK expressing cells were treated with ganciclovir for five days and then their viability was compared to untreated controls (Figure 4.2). Both HSV-TK expressing clones were exquisitely sensitive to ganciclovir, while wild-type cells remained unaffected at the tested concentrations.

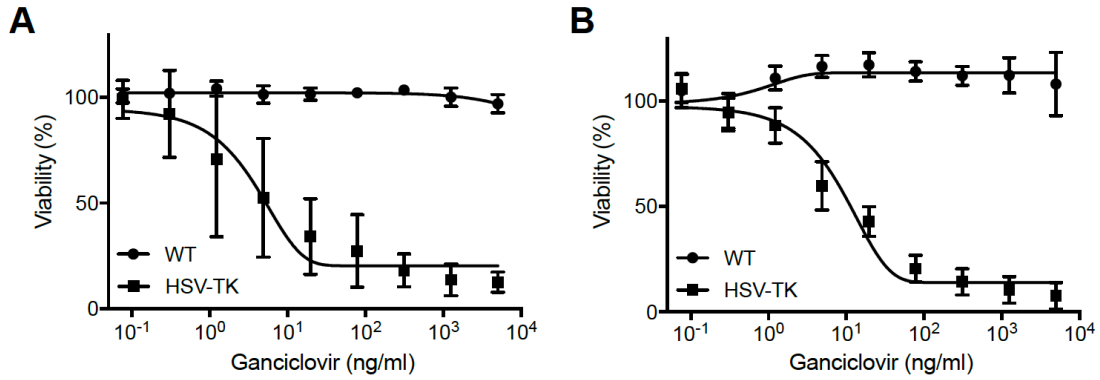


Figure 4.2 HSV-TK expressing KBM7 cells are sensitive to ganciclovir. A. KBM7 wild-type and HSV-TK expressing cells treated with different concentrations of ganciclovir for five days. Viability upon treatment is normalised to untreated controls. Error bars represent the standard deviation of three technical replicates. B. As in A for a different HSV-TK expressing cell line.

4.2.3. A small-scale pilot study to evaluate assay sensitivity

In order to assess the feasibility of detecting the drop-out of single clones from a complex pool, the two HSV-TK expressing clones were added to the 4958 clones pool at the same ratio as the other clones. Each clone was represented with 1000 cells on average. The pool was then treated with ganciclovir for eight days, gDNA was isolated from day 0 and day 8 and barcodes were amplified. Barcode composition and therefore clone abundance was determined by deep sequencing. Comparison of the ganciclovir treated pool with untreated cells showed that both HSV-TK expressing clones were less abundant upon ganciclovir treatment after eight days (Figure 4.3).

Hence, the presented pooled screening approach is sufficiently sensitive to precisely monitor drop-out of single clones from a pooled population. Therefore, the approach was considered suitable for drug sensitivity screening. This method should allow to investigate thousands of gene-drug interactions simultaneously, thereby enabling compound MoA studies.

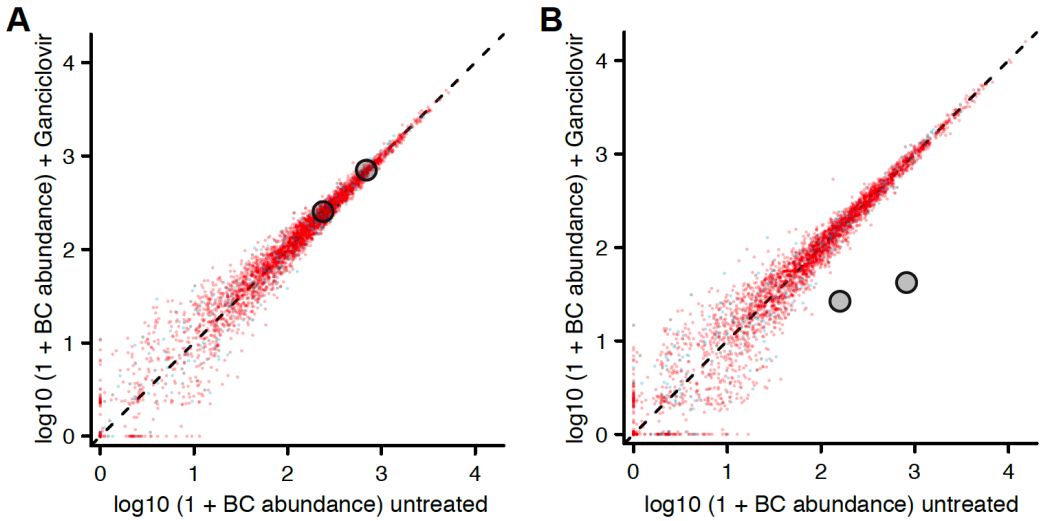


Figure 4.3 Drop-out of HSV-TK expressing KBM7 cells from a complex pool. Comparison of barcode (BC) abundance of the 4958 KBM7 clone pool and two HSV-TK expressing clones between ganciclovir treated and untreated samples at day 0. Red dots indicate knock-out clones, gray dots depict HSV-TK expressing clones. B. As in A for clone abundance at day 8.

4.3. Drug sensitivity screening using highly multiplexed barcode-sequencing

4.3.1. Selection of ten compounds for drug sensitivity screening

Having established a technical framework with ganciclovir as an "ideal" drug allowed testing for the effects of a panel of ten compounds and one drug combination. Seven of these ten compounds were selected based on two studies in *S. cerevisiae* (Lum et al., 2004; Hillenmeyer et al., 2008) that indicated that certain drugs have broad effects and affect many yeast deletion strains ("frequent hitters"), while others show a specific interaction profile with one or very few mutants. Testing such broad effect drugs should result in widespread changes in the clone abundance profile and, therefore, seven of these drugs were selected as candidates for a small pilot study, namely doxorubicin, everolimus, hydroxyurea, mycophenolic acid, methotrexate, camptothecin and irinotecan. The other three drugs (flutamide, phenprocoumon and nilotinib) and the drug combination (flutamide and phenprocoumon) stemmed from a

collaboration with Stefan Kubicek's and Giulio Superti-Furga's research groups at the Centre for Molecular Medicine (CeMM) in Vienna and were expected to show specific rather than broad gene-drug interaction patterns (Licciardello et al., 2017).

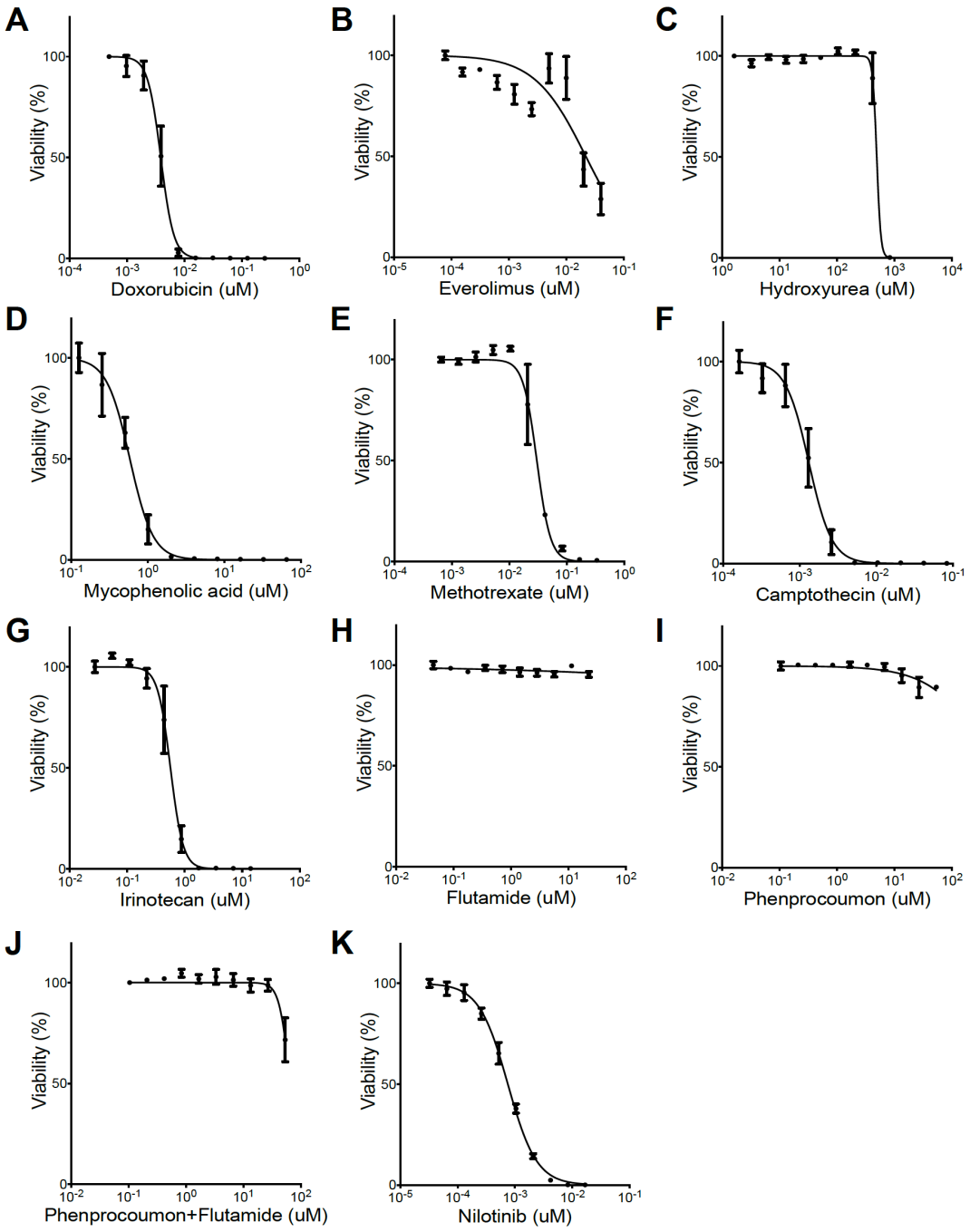


Figure 4.4 Drug sensitivity of KBM7 wild-type cells to different compounds. A-K. KBM7 wild-type cells were treated with different drug concentrations as indicated for eight days. Viability upon compound treatment is normalised to untreated controls. Error bars represent the standard deviation of three technical replicates.

In a first step, different concentrations of the ten drugs and the drug combination were tested on KBM7 wild-type cells. Dose response experiments were performed in 96-well plates and drugs were applied for eight days to define the concentration that results in killing of 10% of the cells (IC_{10}). Except for flutamide all drug applications resulted in a specific response profile that allowed the IC_{10} determination (Figure 4.4). IC_{10} concentrations were subsequently re-tested in a 6-well format to more closely mimic the screening setting. This showed the expected minimal reduction in viability compared to untreated controls (Appendix Figure 4.1).

4.3.2. Pooled forward genetic screening to unravel compound MoAs

In the screening setting, KBM7 knock-out pools were treated with the ten compounds and the drug combination at the previously determined IC_{10} for eight days. The treatment, as expected, induced minimal viability effects on the pools, confirming the determined screening concentrations and compound activity. Subsequently, genomic DNA was isolated from compound-treated and untreated samples taken at day 8, barcodes were amplified by PCR and samples were subjected to deep sequencing. A gene was scored as a hit when the abundance of the respective knock-out clone changed by a factor of at least 4, comparing treated and untreated samples. This comparative analysis revealed small differences between control and treatment (Figure 4.5 and Appendix Table 4.1).

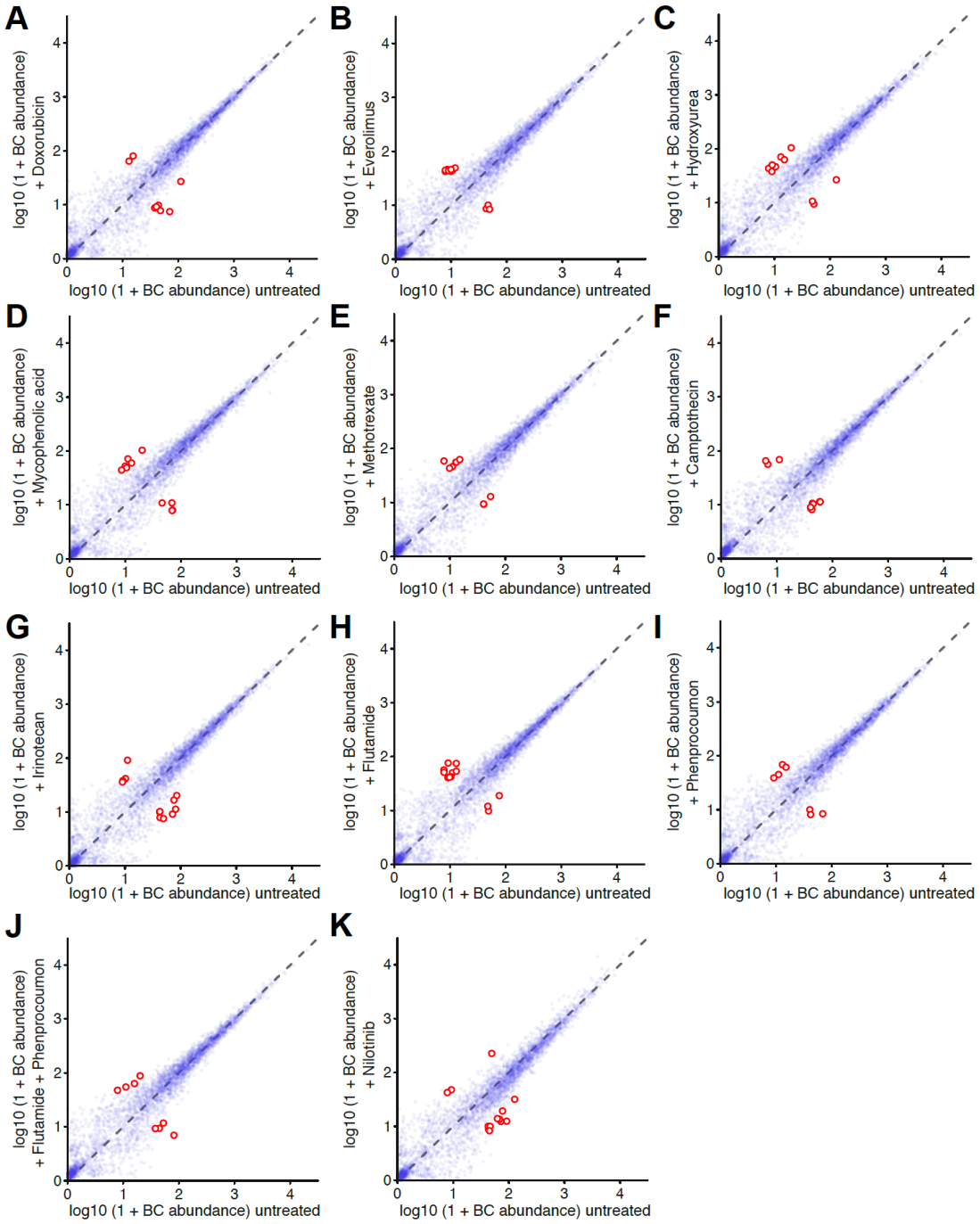


Figure 4.5 Clone abundance of 4958 KBM7 cells upon drug treatment. A-K. Comparison of barcode (BC) abundance in the 4958 KBM7 clone pool between compound-treated and untreated samples at day 8 as indicated. Red circles highlight clones with significantly altered representation after drug treatment as compared to the control. The affected genes are summarised in Appendix Table 4.1.

However, only one of the hits (FRY = FRY microtubule binding protein) overlapped between the two topoisomerase inhibitors camptothecin and irinotecan (Pommier, 2006). Treatment with these inhibitors was expected to yield similar knock-out response profiles, since they act via the same MoA. Of the four KBM7 knock-out clones present in the 4958 clones pool that harbour inactivating gene trap insertions in the FRY gene, only one scored as a hit with increased abundance, which hints towards a false positive gene-drug interaction. This is further supported by the fact that FRY is also more enriched upon treatment with the unrelated drugs methotrexate, phenprocoumon, mycophenolic acid and flutamide, as well as the combination treatment.

Furthermore, single treatment with flutamide, an androgen receptor antagonist, and phenprocoumon, a vitamin K epoxide reductase inhibitor (Ufer, 2005; Chen et al., 2009), and the combination treatment was expected to yield an overlap in the response profile. However, except for FRY and GRIK1 (glutamate ionotropic receptor kainite type subunit 1; only one GRIK1 knock-out clone was present in the pool), which are also scored as hits in response to unrelated drugs, no interactions were shared between the single and combination treatments.

Despite most of the targeted genes being represented by multiple independent clones in the KBM7 pool, in most cases only one clone scored as a hit (e.g. FRY, COMMD1, COMT etc.). Moreover, some knock-out clones display altered abundance in response to treatment with several unrelated drugs (e.g. FRY, GRIK1, FAM78A etc.). These findings suggest that most identified gene-drug interactions are false positives. Removal of these low-confidence interactions results in one to four potential true positive interactions for each tested compound and none for the combination treatment. However, the remaining candidate interactions are represented by only one knock-out clone in the pool, reducing the confidence in these potential true positives.

Failure to detect larger differences in clone abundance upon treatment with compounds could be due to sub-optimal dosing despite careful prior assessment. Most likely, the chosen compound concentrations of an IC_{10} was too low to elicit detectable differences in the abundance of different knock-out clones.

In summary, this study showed that the pooled screening approach is sufficiently sensitive to monitor drop-outs of single cells from the pooled population, but the suitability for compound MoA screening remains to be demonstrated. In principle and once the technical hurdles have been overcome, this approach should be suitable for interrogating thousands of gene-drug interactions simultaneously and should enable compound MoA studies.

4.4. Discussion

This chapter describes the set-up of a chemical genetic approach to investigate compound sensitivity in an isogenic human cell line model. By measuring barcode abundance of 4958 pooled knock-out cell lines upon drug application, thousands of gene-drug interactions were monitored simultaneously.

Chemical genetic screens with pooled mutant collections of characterised barcoded cell lines offer a convenient opportunity for assessing drug sensitivity in a high-throughput manner (Muellner et al., 2011; Nijman, 2015; Smida et al., 2016). Simultaneous tracking of thousands of isogenic knock-out cell lines upon drug application enables cost-effective positive and negative selection screens and can reveal compound MoA.

Nowadays, such mutant or knock-out cell line collections can be efficiently generated by CRISPR/Cas9. As shown and discussed in the previous chapter, CRISPR/Cas9 based approaches can yield panels of cell lines harbouring complete knock-outs (Gapp et al., 2016). During the course of this study, several publications have

highlighted the feasibility of performing pooled CRISPR/Cas9 screens in human cells (Wang et al., 2014; Shalem et al., 2014; Hart et al., 2015; Hartenian and Doench, 2015; Shalem et al., 2015). However, at the start of this project, this technology was not yet available at this scale and, hence, an isogenic human knock-out collection generated by an insertional mutagenesis approach was employed (Bürckstümmer et al., 2013). In contrast to CRISPR/Cas9, the inactivation of genes by insertional mutagenesis relies on the non-targeted integration of a gene trap cassette in the haploid genome of KBM7 cells. Depending on the exact location of the inserted gene trap cassette, not all transcript variants may be interrupted, and as a consequence, not all isogenic cell lines represent complete knock-outs. Hence, probably not all clones present in the 4958 isogenic cell line pool can show a phenotype upon drug application even when an insertion is present. However, cell line pools created with CRISPR/Cas9 may suffer from similar constraints, in that not all cells in the pool harbour inactivating mutations, as a fraction of indels generated by Cas9 are expected to maintain the reading frame.

The generation of single cell derived knock-out clones using CRISPR/Cas9 technology to generate a collection of characterised, barcoded human cells would allow to overcome both hurdles, incomplete knock-out within and across cells. As discussed in chapter 3, such an undertaking would be a time and cost-intense effort and should ideally comprise independent knock-out cell lines for the same gene generated by multiple unique gRNAs targeting different parts of the gene. This would help to avoid scoring false positive interactions arising from the heterogeneity between different single cell derived clones. Although the generation of such a collection would be an enormous undertaking, it would have high value for pooled screening approaches in human cells and would benefit the biomedical community in general. Thus, it is imaginable and desirable that in the near future, a larger research consortium, a commercial party and/or a joint public-private partnership will create and offer extensive collections of ready-to-use barcoded knock-out clones.

The results of this chapter demonstrate that the presented chemical genetic approach is highly sensitive and that it could, in principle, be used to monitor thousands of gene-drug interactions in parallel. High assay sensitivity is imperative for pooled screening approaches to accurately determine the increased abundance and especially the selective absence of single barcodes in a large and complex mixed population. The drop-out of two clones ectopically expressing HSV-TK could be detected specifically upon ganciclovir treatment of the whole cell line pool. This demonstrates that the established screening strategy is sufficiently sensitive to detect the negative selection of single clones in a complex pool despite the considerable amount of experimental noise, i.e. biological and technical fluctuations, that are inherent in such screens. These results show that the established screening approach would be suitable for collecting compound response profiles at high-throughput. Mapping of gene-drug interactions at large-scale should allow to link compound action with specific cellular processes as will be discussed below.

In this pilot study, however, no major differences in cell viability across the mutant pool could be detected upon application of compounds with different MoAs to the mutagenised cell line collection. Minimal changes in barcode abundance following treatment with ten different drugs is most likely due to inappropriate dosing so that clone abundance was not substantially affected. Despite these setbacks in this first application test, it is conceivable that only minor adjustments in drug concentration would suffice to successfully use this experimental set-up for the investigation of compound MoA in a high-throughput and cost-effective manner.

Desirable extensions of the presented screening assay would include an increased resolution of the generated dataset by using multiple doses or extending the scope with larger panels of compounds and isogenic cell lines. As previously demonstrated in yeast (Giaever et al., 2004; Lum et al., 2004; Parsons et al., 2006; Hillenmeyer et

al., 2008; Hillenmeyer et al., 2010; Lee et al., 2014), the collection of large data sets from many chemical genetic screens can be used to investigate the MoA of compounds based on the “guilt-by-association” principle. These approaches rely on the concept that compounds with a similar MoA tend to show similar drug response profiles across a mutant cell population. With larger datasets containing hundreds of drug response profiles, it should become feasible to compare profiles of compounds with unknown MoA to profiles of drugs with established MoA. Although some studies have demonstrated the feasibility of “guilt-by-association” screens in mammalian cells (Muellner et al., 2011; Jiang et al., 2011; Martins et al., 2015), large-scale investigations have not yet been completed in human cells. Adjusting the experimental conditions to accommodate a larger scope could leverage the established approach for such yeast-like screens and pave the way for large-scale “guilt-by-association” screening in a human cell system.

In summary, this chapter describes a small-scale pilot study for pooled drug sensitivity and MoA screening using an isogenic human cell line collection. The presented strategy is highly sensitive and allows the detection of individual gene-drug interactions in a complex setting. The pooled screening approach using a collection of barcoded human cell lines can be potentially exploited for MoA studies in human cells and could provide a new powerful tool in the chemical genetic screening repertoire that should expand the capacity to determine the MoA of compounds in a fast and cost-effective manner.

Chapter 5. Final discussion

5.1. Synopsis

Despite the constant progress in the treatment of human diseases, major challenges remain due to the incomplete understanding of the molecular underpinnings of human biology (what to target) and drug action (how to target). For example, inhibiting deregulated kinases with small molecules and monoclonal antibodies has revolutionised cancer therapy by offering a highly specific alternative to the broadly acting chemotherapeutics (Fleuren et al., 2016). However, the cellular functions of many kinases remain poorly explored despite their biological and pharmaceutical importance as a class of enzymes and drug targets (Weiner, 2015; Fleuren et al., 2016; Gharwan and Groninger, 2016). Similarly, limited understanding of the precise MoA of compounds has impeded the repurposing of approved compounds for new applications and the development of novel therapeutics (Bunnage et al., 2015).

Another fundamental problem that hampers successful clinical applications, with cancer treatment as the prime example, is the emergence of drug resistance, which ultimately leads to progressive disease and poor survival (Zahreddine and Borden, 2013). Refractoriness to treatment can be intrinsic due to resistance mechanisms present in tumour cells prior to treatment, or acquired, when tumour cells initially respond to a particular therapy and resistance develops over time. Resistance mechanisms of targeted agents include point mutations in the drug binding site, target gene amplification or over-expression, compensatory alterations in the affected signalling pathway or modulation of the drug influx/efflux (Zahreddine and Borden, 2013; Huang et al., 2014; Lovly and Shaw, 2014).

Advanced understanding of these mechanisms is crucial to develop new inhibitors that are highly effective and less prone to drug resistance. Another option to address the emergence of drug resistance is by combination therapy, which has been in clinical use with conventional chemotherapeutics for decades. Recently, it has also

been applied by combining targeted therapeutics with chemotherapy or other targeted agents (Kummar et al., 2010; Gotwals et al., 2017). Increasing the number of potent targeted therapeutics will allow to expand the range of combinatorial treatments to be tested and clinically applied and to eventually reduce drug resistance.

Unravelling the function of all human genes and elucidating the causal underpinnings of complex diseases and drug resistance, together with revealing the action of small molecules, could significantly improve treatment regimens and clinical outcome. Meeting these challenges will be a laborious long-term goal for the research community as a whole, as individual genes and compounds would need to be assessed by custom assays to determine their function at a molecular level. There is thus a need for powerful, high-throughput systems to complete these pictures. In this thesis, two proof-of-concept studies are described for the functional annotation of genes and the assessment of compound MoA in haploid human systems, aiming to provide tools for improved understanding of gene and drug action.

5.2. Genetic screening as a high-throughput tool to assign genes with functions in human cells

Genetic screens in model organisms and cultured cells have revealed the wiring underlying diverse biological processes in exquisite detail, have proven essential to delineate the MoA of compounds and have identified drug targets for the effective treatment of various diseases (Mohr et al., 2014; Shalem et al., 2015; Nijman, 2015). Initially, genetic studies relied on spontaneously arising mutations and later on the use of non-targeted mutagenesis approaches, until the development of RNAi and targeted gene manipulation techniques enabled systematic forward and reverse

genetic screens with defined genetic mutants (Lehner, 2013; Mohr et al., 2014; Shalem et al., 2015; Housden et al., 2017).

Although forward genetic screening technologies have been used to investigate a plethora of biological processes in human cells, complementary reverse genetic approaches have received less attention due to technical hurdles. For example, RNAi, being the predominant method for reverse genetic screening in human cells for two decades, has suffered from substantial off-target effects and incomplete protein depletion (Kaelin, 2012). Repurposing of the CRISPR/Cas9 system for the targeted modification of the human genome and target-selected approaches to generate characterised knock-out cell lines based on insertional mutagenesis employed in near-haploid cells have provided a way around these limitations (Bürckstümmer et al., 2013; Wright et al., 2016). Using isogenic human knock-out cell lines has greatly enlarged the spectrum of possible applications, as demonstrated by the approaches presented in this thesis. Chapter 3 employs a human knock-out collection generated with CRISPR/Cas9 (Gapp et al., 2016) and chapter 4 utilises a pooled knock-out collection comprising characterised, barcoded human cell lines generated via insertional mutagenesis (Bürckstümmer et al., 2013).

Finding causal connections between genotype and phenotype can be challenging when using, for instance, compendia of established non-isogenic (cancer) cell lines, as multiple genetic alterations can co-occur in the same cell line (Garnett et al., 2012; Barretina et al., 2012; Basu et al., 2013; Seashore-Ludlow et al., 2015; Iorio et al., 2016). In contrast, isogenic collections allow the establishment of causal links between a particular phenotype and a specific genetic alteration (e.g. knock-out) due to the presence of only one defined mutation at a time and therefore offer an ideal starting point for reverse genetic investigations in human cells.

However, one drawback of working with isogenic single knock-out collections is that in some cases the inactivation of one non-essential gene might not be sufficient to induce a phenotype due to genetic or functional redundancy of the targeted gene.

The simplest form of genetic redundancy includes the presence of two alleles in diploid organisms or gene duplications. Functional redundancy of non-allelic or structurally unrelated genes represents a more complex mechanism, where the function of one gene in a particular pathway can be compensated by the function of another gene in a redundant parallel pathway (Masel and Siegal, 2009). Indeed, studies in model organisms and human cells have demonstrated that the inactivation of single genes often does not affect a particular phenotype (e.g. viability, morphology), indicating that the function of many genes can be compensated by others (Winzeler et al., 1999; Giaever et al., 2002; Blomen et al., 2015; Wang et al., 2015). However, most single mutants are susceptible to the inactivation or inhibition of an additional non-essential gene, allowing to map gene-gene or gene-drug interactions at large scale (Hillenmeyer et al., 2008; Costanzo et al., 2010; Nichols et al., 2011; Laufer et al., 2013; Blomen et al., 2015; Costanzo et al., 2016).

Single knock-out collections in human cells have become available only recently and comprise only a fraction of all non-essential human genes (Bürckstümmer et al., 2013; Gapp et al., 2016). The ultimate goal should be to generate collections of double knock-outs or even higher-order mutant collections to eventually assign every human gene with a function and globally understand the wiring of a human cell. As detailed in chapter 4, such collections should ideally comprise characterised barcoded knock-outs and should be generated with CRISPR/Cas9 technology. Although it would be a massive undertaking, generating large-scale human double or higher-order knock-out collections would have tremendous value for pooled screening approaches in human cells and would benefit the biomedical community in general.

5.3. Transcriptome based screening to link human genes with functions

In this study, transcriptional profiling has been employed to perform the first transcriptome based reverse genetic screen in human cells. Indeed, this technology has offered a novel opportunity to link genes with functions in model organisms (DeRisi et al., 1997; Holstege et al., 1998; Hughes et al., 2000; Chua et al., 2006; Hu et al., 2007; van Wageningen et al., 2010; Lenstra et al., 2011; Kemmeren et al., 2014). Chapter 2 of this thesis describes the establishment of a phenotypic profiling method based on multiplexed RNA-sequencing. The presented results show that the established transcriptome based screening strategy is sufficiently sensitive to measure the transcriptional response of haploid human cells to a variety of cellular perturbations (e.g. small molecules and polypeptides) and that it is suitable for linking genes to pathways. Chapter 3 demonstrates that the implemented experimental and computational framework is scalable and applicable to large-scale reverse genetic investigations in a human cell system.

In particular, systematic functional annotation of tyrosine kinases (TKs), as used in this screen, has the potential to affect many aspects of basic and translational research, since kinases play fundamental roles in various biological processes and are frequently altered in human diseases (Fleuren et al., 2016). The findings that various TKs participate in interferon, FGF1 and WNT3A signalling is in concordance with the previously established functions of these enzymes and validates the transcriptome based screening approach for linking genes with pathways (Turner and Grose, 2010; O'Shea et al., 2015). A logical extension of the presented reverse genetic screens would be the functional characterisation of all non-essential kinases, which could have a profound impact on biomedical research and clinical application.

As a reverse genetic tool, the presented screening platform should allow easy expansion even beyond the analysis of kinases to other gene families and eventually,

all genes for which knock-outs can be generated. Furthermore, transcriptome based screening should allow to chart phenotypes not only for knock-outs, but also for other cellular perturbations, such as engineered point mutations or chemical agents. Such screens could greatly advance the efforts to link genotype to phenotype in human cells. Moreover, the described functional genomics approach is generic, which should allow to address similar biological questions also in other cellular systems and model organisms.

Taken together, chapters 2 and 3 describe a proof-of-concept study for a transcriptome based screening method with the potential to pave the way for a high-throughput reverse genetic tool for functional gene annotation in human cells and other biological systems.

5.4. Chemical-genetic screening in human cells for high-throughput compound MoA determination

Besides the critical need to advance the understanding of the function of human genes and how their malfunctions cause disease, a more in depth understanding of the MoA of approved and novel compounds is needed to improve treatment regimens and ultimately guide therapy. Currently, the clinical development of novel compounds suffers from high attrition rates. A lack of accurate preclinical target validation methods has been identified as the main cause for the failure of drugs in late clinical stages (Bunnage et al., 2015).

In molecular terms, determining the MoA of compounds, assessing potential off-target activities and/or revealing genotype-specific effects early in drug development has remained challenging (Schenone et al., 2013; Bunnage et al., 2015). As a consequence, effort is wasted on compounds that are ultimately not suitable for clinical use due to various reasons including poor pharmacogenetics or limited efficacy and safety. Establishing methods to evaluate and select early-stage

therapeutic agents that are more likely to pass through these bottlenecks will speed up drug discovery. As a tool for addressing the first step on this path, genetic screening in combination with small molecules provides an efficient way to identify gene-drug interactions at large scale and can thereby elucidate the MoA of compounds or assess potential off-target effects (Nijman, 2015).

Chapter 4 describes the development of a pooled forward genetic approach to explore thousands of gene-drug interactions and suggests that this set-up could be used to study MoAs of uncharacterised compounds at large scale.

The implemented pooled screening approach is sufficiently sensitive to monitor drop-outs of single clones from a complex mixture of nearly 5000 clones and can measure thousands of gene-drug interactions in parallel. Mapping of chemical-genetic interactions across large human knock-out collections should improve compound classification and aid the assignment of compounds with a MoA. Furthermore, the identification of novel gene-drug interactions has the potential to provide new avenues for therapeutic intervention.

Although not formally demonstrated, the established chemical genetic strategy should allow to assess the MoA of compounds based on the “guilt-by-association” principle as previously shown in yeast (Giaever et al., 2004; Lum et al., 2004; Parsons et al., 2006; Hillenmeyer et al., 2008; Hillenmeyer et al., 2010; Lee et al., 2014) and mammalian systems on a smaller scale (Perlman et al., 2004; Lamb et al., 2006; Muellner et al., 2011; Jiang et al., 2011; Breinig et al., 2015; Martins et al., 2015). Large-scale “guilt-by-association” screens have not yet been successfully performed in human cells.

Minor technical adjustments (described in chapter 4) to the presented pooled chemical genetic approach should allow this approach to become suitable for studying drug action in a human cell system at high throughput. Such an assay would have great potential to address the critical need for a systematic and high-throughput

tool to assess compound MoA, off-target activities and genotype-specific effects in a preclinical setting.

Together, the systematic characterisation of compound MoA in diverse genetic backgrounds using complementary approaches, such as transcriptional profiling or microscopy-based high-content screening, will accelerate the delivery of potent and specific genotype-stratified therapeutics.

5.5. Concluding remarks and future perspectives

Over the past century, genetics has developed from the investigation of spontaneously arising mutations in human and/or model organisms to large-scale screening approaches in cultured human cells. To date, genetic screening sits at the intersection of many disciplines, including cellular and molecular biology, functional genomics and bioinformatics.

In summary, this thesis identifies three areas in biomedical research instrumental to understanding human biology, and therefore medical intervention, which can be investigated with modern screening methods: the function of genes, the interaction of genes with compounds (MoA) and the processes underlying drug resistance.

Because of the sheer number of possible interactions and combinations and the existence of redundancy, buffering and feedback mechanisms, cellular circuitries are inherently complex. Unravelling such complex interplays requires powerful tools, systems and methods, as well as the bridging of diverse disciplines from molecular biology to bioinformatics. Screens that are carefully designed, combined with state-of-the-art technologies and robust bioinformatic pipelines can generate data on an unprecedented scale. Different screening approaches (including design and analysis workflows) will create the toolbox of the future biomedical researcher.

This thesis adds to this toolbox with the set-up of two strategies to disentangle the function of genes and compounds in a human cell system. The established screening techniques provide novel high-throughput tools for the systematic functional annotation of human genes, the identification of effective targets suitable for therapeutic intervention and a more in depth understanding of the MoA of compounds.

In the near future, genetic screening in human cells using various CRISPR/Cas9 tools will become further embedded within basic and applied biomedical research. Using the CRISPR/Cas9 system in a temporally controlled manner or utilising nuclease deficient Cas9 (dCas9) tethered to various effector domains will move beyond the proof-of-concept screening and allow forward and reverse genetic investigations at a hitherto unmet scale in human cells.

Improvements in culturing systems and differentiation protocols will allow the screening of patient's tumour cells and to transdifferentiate one cell type into another, enabling the use of the "correct" cell type or tissue to model diseases and test therapeutic responses. At the same time, the continuous increase in performance of massively parallel sequencing will allow the more widespread use of omics approaches as a read-out for screening and for dissecting increasingly complex phenotypes.

Together, these advances will yield unprecedented functional insight into the human genome and will enable the assessment of phenotypes with unmet resolution in any cell type of interest. Eventually, this will allow the scientific community to systematically assign every gene with a particular function, as well as to connect non-coding regulatory elements to their target genes and to a specific phenotype. After the completion of the human genome project, such a vast phenotype knowledge base will be the next major milestone in understanding human biology.

In combination with chemical agents, large-scale mapping of gene-drug interactions will increase the clinical potential of compounds with different mechanisms of binding and inhibition. Systematic mapping of the complex interplay between genes, compounds and cellular pathways in diverse genetic backgrounds using complementary approaches such as transcriptional profiling, high-content screening or chemical mutagenesis in haploid systems will accelerate the development of potent and specific genotype-stratified therapeutics. Ultimately, it will become possible to predict how aberrant cellular circuitries lead to disease phenotypes and how those can be targeted therapeutically.

This thesis is mainly focused on the “How to” of genetic screening. Throughout the chapters, it demonstrates how well-designed screens can elucidate gene function and compound action. It advocates the meticulous development of a toolkit of universally applicable, robust and scalable screening methods and the availability of mutant human cell collections to use with the established screening approaches. The long-term result of a large number of such screens would be a broad knowledge base in which all gene functions are detailed - a biological code book. With this atlas in hand, charting of gene-drug interactions and increased understanding of the MoA of compounds will allow the translation of this knowledge to clinical applications and, ultimately, to the patient’s bedside.

Chapter 6. Material and methods

6.1. Cell lines

The near-haploid human cell line HAP1 was provided by T. Brummelkamp. HAP1 knock-out cell lines were purchased from Horizon Genomics or generated as described below. The human embryonic kidney cell line HEK293T was obtained from the American Type Culture Collection (ATCC). The near-haploid human KBM7 cell line and knock-out pools were ordered from Horizon Genomics. HAP1 and KBM7 cells were propagated in Iscove's Modified Dulbeccos's Medium (IMDM+GlutaMAX, Invitrogen GIBCO) supplemented with 10 % foetal bovine serum (FBS, Invitrogen GIBCO), 100 u/ml penicillin and 100 µg/ml streptomycin (Sigma-Aldrich). HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM+GlutaMAX, Invitrogen GIBCO) supplemented with 10 % FBS, 100 u/ml penicillin and 100 µg/ml streptomycin (Sigma-Aldrich). All cell lines were maintained in a humidified incubator at 37 °C with 5 % CO_2.

HAP1 knock-out cell lines used for reverse genetic screening were generated at Horizon Genomics and are listed in Appendix Table 6.1. guide RNAs (gRNA) targeting non-essential (Blomen et al., 2015) and expressed (Essletzbichler et al., 2014) human kinases (Manning et al., 2002) were designed using http://crispr.mit.edu/. Cell lines were obtained by transient transfection of plasmids encoding for the specific gRNA and *S. pyogenes* Cas9 (pX165 obtained from the Zhang lab) using Turbofectin (Origene). Transfected cells were selected with 20 µg/ml blasticidin for 24 h. Clonal cell lines were obtained by plating cells in a limiting dilution.

Additional HAP1 knock-out cell lines targeting the same kinase genes but different exons were generated by lentiviral transduction (Table 6.1). gRNAs were designed as described above using http://crispr.mit.edu/. Oligonucleotides encoding the

specific gRNAs were cloned into the lentiCRISPR v2 backbone vector (Addgene #52961) expressing *S. pyogenes* Cas9. Lentivirus was produced by calcium phosphate mediated transfection into HEK293T producer cells (see below). Transduced HAP1 cells were selected with 0.5 µg/ml puromycin for 3 d and clonal cell lines were generated by limiting dilution.

Table 6.1 HAP1 knock-out cell lines for validation experiments

Gene	Group	Family	Clone	gRNA.sequence	Indel	Exon
TYK2	TK	JAKA	36-8	CCCAACGGGCTTACTGCCCC	1bp del	3
TYK2	TK	JAKA	2840-2	AGACGCCATGCACGAAGGCC	20bp del	14
TYK2	TK	JAKA	3415-10	CTTGGGCCTGAGCATCGAAG	8bp del	4
TYK2	TK	JAKA	3415-11	CTTGGGCCTGAGCATCGAAG	5bp del	4
ACVR2B	TKL	STKR	2842-7	TTTGAGCCAGGGAAACCTCC	2bp del	8
ACVR2B	TKL	STKR	3427-2	CACAGCAGCAGAAGTACACC	2bp del	3
ACVR2B	TKL	STKR	3427-9	CACAGCAGCAGAAGTACACC	5bp del	3
CCK4	TK	CCK4	SC1	TGGCTGGATCTCAGCTTCCG	1bp ins	3
CCK4	TK	CCK4	SC2	TGGCTGGATCTCAGCTTCCG	20bp del	3
EPHB1	TK	EPH	SC1	TACTGCAGAGCTGGGCTGGA	1bp ins	2
EPHB1	TK	EPH	SC2	TACTGCAGAGCTGGGCTGGA	1bp ins	2
ERBB2	TK	EGFR	SC1	CCAGGTGGTGCAGGGAAACC	1bp del	5
FGFR3	TK	FGFR	SC1	CAGCAGGAGCAGTTGGTCTT	19bp del	3
FGFR3	TK	FGFR	SC2	CAGCAGGAGCAGTTGGTCTT	14bp del	3
IGF1R	TK	INSR	SC1	TGCCGCCACTACTACTATGC	1bp ins	3
IGF1R	TK	INSR	SC2	TGCCGCCACTACTACTATGC	1bp ins; 5bp del	3
MERTK	TK	AXL	SC1	GCCAGAGCGCGGGGAGGAAG	10bp del	1
MERTK	TK	AXL	SC2	GCCAGAGCGCGGGGAGGAAG	1bp ins	1
PDGFRA	TK	PDGFR	SC1	CAGCCTAAGACCAGGAACGC	10bp del	2
PDGFRA	TK	PDGFR	SC2	CAGCCTAAGACCAGGAACGC	2bp del	2
ROR2	TK	ROR	SC1	GCGGCGGCCGCCCAGACGGC	23bp del	1
ROR2	TK	ROR	SC2	GCGGCGGCCGCCCAGACGGC	1bp ins; 1bp del	1

Genomic DNA (gDNA) of clonal HAP1 knock-out cells was isolated using DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's specifications. 500 ng were amplified by PCR (synthesis cycle: 2 min at 94 °C, (30 sec at 94 °C, 30 sec at 55 °C, 55 sec at 68 °C) x 38), 7 min 68 °C) using GoTaq DNA polymerase (Promega) and primers flanking the gRNA target site (Appendix Table 6.2). PCR products were purified according to the QIAquick PCR purification kit (Qiagen) protocol and analysed by Sanger sequencing. Clones harbouring frameshift mutations were selected and stored for use.

KBM7 knock-out pools used for pooled compound MoA screening were purchased from Horizon Genomics. Individual KBM7 knock-outs were generated using a gene

trap insertional mutagenesis approach and limiting dilution as described previously (Bürckstümmer et al., 2013). In total, 288 knock-out pools (up to 20 barcoded individual knock-outs per pool) were ordered and further pooled into four sub-pools (approx. 1250 uniquely barcoded knock-outs per sub-pool). Pooling was carried out by combining $5x10^6$ cells of each KBM7 knock-out pool into one sub-pool.

6.2. Plasmids and reagents

Plasmids encoding gRNAs or cDNAs used for retro- and lentivirus production are summarised in Table 6.2. Sense and antisense gRNAs were annealed using T4 ligation buffer (NEB) and cloned into a lentiCRISPR v2 backbone vector (Addgene #52961) using the *BsmBI* restriction enzyme and T4 DNA ligase (NEB). Gateway cloning of cDNAs from pDONR223 into pBABE destination vectors was performed following manufacturer's instructions (Invitrogen). All vectors were verified by Sanger sequencing.

Table 6.2 Plasmids

Plasmid	Type	Insert	Species	Source
pDONR223-HSV-TK	Gateway donor vector	HSV-TK	Herpes simplex	Insert derived from pAL119-TK Addgen plasmid #21911
pBabeNeo-GW-Empty	Expression vector, retroviral	None	NA	Nijman lab plasmid #819
pBabeNeo-GW-HSV-TK	Expression vector, retroviral	HSV-TK	Herpes simplex	Generated for this study
lentiCRISPR-v2	Expression vector, lentiviral	Cas9	Synthetic	Addgene plasmid #52961
lentiCRISPR-v2-CCK4-ex3-gRNA	Expression vector, lentiviral	Cas9; CCK4 gRNA	Synthetic; Homo sapiens	Generated for this study
lentiCRISPR-v2-EPHB1-ex2-gRNA	Expression vector, lentiviral	Cas9; EPHB1 gRNA	Synthetic; Homo sapiens	Generated for this study
lentiCRISPR-v2-ERBB2-ex5-gRNA	Expression vector, lentiviral	Cas9; ERBB2 gRNA	Synthetic; Homo sapiens	Generated for this study
lentiCRISPR-v2-FGFR3-ex3-gRNA	Expression vector, lentiviral	Cas9; FGFR3 gRNA	Synthetic; Homo sapiens	Generated for this study
lentiCRISPR-v2-IGF1R-ex3-gRNA	Expression vector, lentiviral	Cas9; IGF1R gRNA	Synthetic; Homo sapiens	Generated for this study
lentiCRISPR-v2-MERTK-ex1-gRNA	Expression vector, lentiviral	Cas9; MERTK gRNA	Synthetic; Homo sapiens	Generated for this study
lentiCRISPR-v2-PDGFRA-ex2-gRNA	Expression vector, lentiviral	Cas9; PDGFRA gRNA	Synthetic; Homo sapiens	Generated for this study
lentiCRISPR-v2-ROR2-ex1-gRNA	Expression vector, lentiviral	Cas9; ROR2 gRNA	Synthetic; Homo sapiens	Generated for this study

Ligation mixtures (20 to 100 ng DNA) were added to competent *E. coli* DH5alpha, F10, or Stbl3 and heat shock was performed at 42 °C for 30 sec. Transformed *E. coli* were selected using 100 µg/ml ampicillin (AppliChem). Mini- and midi-preps were performed according to QIAprep spin prep kit protocol (Qiagen).

Recombinant polypeptides and small molecules used for stimulation experiments are summarised in Appendix Table 6.3. Polypeptides were diluted following manufacturer's instructions in water, 0.1 % BSA, 0.1 % acetic acid, 10 mM sodium citrate (pH 3), 5 mM sodium phosphate (pH 8 or 7.2), or 10 mM acetic acid. Polypeptide stocks were prepared in PBS containing 0.1% BSA. Small molecules were diluted in water, DMSO or 20 mM MES buffer (pH 5.5).

6.3. Retro- and lentivirus production

Retro/lentivirus production was carried out by calcium phosphate mediated transfection of retro- and lentiviral vectors into HEK293T producer cells. Derivatives of pBABE were used for retrovirus production, lentiCRISPR v2 was used for lentivirus production (Table 6.2). $5x10^5$ HEK293T cells were co-transfected with packaging and envelope vectors (pPol or dR8.91 and VSV-G) in 158 µl HEPES buffered saline (pH 7.0) and 8.3 µl 2.5 M $CaCl_2$. pCMV-GFP served as a transfection control. Medium was replaced 24 h post transfection and viral supernatants were collected at 36, 48 and 60 h post transfection. Pooled viral supernatants were centrifuged at 2000 rpm for 5 min at 4 °C and stored at -80 °C.

6.4. Transduction of target cells

HAP1 cell transductions were performed for at least 12 h using 20-50 % viral supernatant and 7 µg/ml polybrene (Sigma-Aldrich) in a 12-well format. KBM7 transductions were carried out using 50 % viral supernatant and 8 µg/ml protamine

sulphate (Sigma-Aldrich) in a 24-well format. Spin infections were performed at 2000 rpm for 30 min at 30 °C and repeated up to 3 times. Infected cell lines were selected with 0.5 µg/ml puromycin (Sigma-Aldrich) or 1 mg/ml neomycin (Invitrogen, Gibco). Over-expression of cDNAs and gRNA mediated knock-out was verified via Western blotting, qRT-PCR, and/or Sanger sequencing.

6.5. Stimulation experiments

Polypeptide and small molecule stimuli (Appendix Table 6.3) were applied on $2x10^5$ HAP1 cells seeded in 12-well plates. Cells were washed twice with PBS 36 h post seeding and cultured under reduced serum conditions (IMDM supplemented with 0.5 % FBS and 100 u/ml penicillin and 100 µg/ml streptomycin) for 16 h. Stimulations with polypeptides or small molecules were carried out for 6 h. Samples were washed twice with cold PBS before storage at -80 °C.

6.6. Dose response experiments

Dose response assays in 96-well plates were performed with $1x10^4$ KBM7 cells per well. Compounds (Table 6.3) were provided by S. Kubicek (CeMM) as 10 mM stocks in DMSO. Serial dilutions were applied for 8 d and viability was determined using CellTiter-Glo Luminescent Cell Viability Assay (Promega) following manufacturer's specification with the exception that the CellTiter-Glo reagent was diluted 1:7 in PBS. Luminescence was read on a Victor X3 2030 Multilabel Reader (PerkinElmer). Viability upon drug application was normalised to a DMSO control and plotted against drug concentration to determine the inhibitory concentration of 10 % (IC_{10}).

Dose response experiments in 6-well plates were carried out with $5x10^6$ KBM7 cells treated with compounds (IC_{10}) or vehicle (DMSO) for 8 d. Cells were counted with a Casy cell counter (Roche) and fresh compounds were added every three days.

Viability of compound-treated wells was normalised to a DMSO control and plotted against the time in culture.

Table 6.3 Compounds

Compound	Description	IC_{10}	Source
Doxorubicin	DNA intercalator; topoisomerase inhibitor	980 pM	S. Kubicek
Everolimus	Non-specific serine-threonine protein kinase inhibitor	150 pM	S. Kubicek
Hydroxyurea	Ribonucleoside-diphosphate reductase inhibitor	6.3 µM	S. Kubicek
Mycophenolic acid	Bacterial inosine-5-monophosphate dehydrogenase inhibitor	250 nM	S. Kubicek
Methotrexate	Dihydrofolate reductase inhibitor	7.5 nM	S. Kubicek
Camptothecin	Topoisomerase inhibitor	650 pM	S. Kubicek
Irinotecan	Topoisomerase inhibitor	150 nM	S. Kubicek
Flutamide	Androgen receptor antagonist	3.7 µM	S. Kubicek
Phenprocoumon	Vitamin K epoxide reductase inhibitor	7.4 µM	S. Kubicek
Flutamide+Phenprocoumon	Androgen receptor antagonist+vitamin K epoxide reductase inhibitor	3.7 µM + 7.4 µM	S. Kubicek
Nilotinib	Tyrosine kinase inhibitor	1 nM	S. Kubicek
Ganciclovir	Nucleoside analog; pro-drug	NA	Merck Millipore

6.7. Reverse genetic screens

For the reverse genetic screens, 64 HAP1 knock-out cell lines (Appendix Table 6.1) were screened in batches of four lines against ten stimuli (Table 6.4) along with unstimulated knock-out and wild-type controls. Stimulation experiments (see above) were performed for 6 h in duplicates and carried out in parallel. RNAs of 96 samples at a time were isolated and libraries for RNA-sequencing prepared using the QuantSeq 3' mRNA-Seq library prep kit (Lexogen) (see below).

Table 6.4 Stimuli for reverse genetic screening

Stimulus	Name	Description	Conc.	Vendor
ACTA	Activin A	Modulator of development	4 ng/ml	Peprotech
BMP2	Bone morphogenetic protein 2	Modulator of development	220 ng/ml	Peprotech
DFOM	Deferoxamine	Hypoxia inducing agent	50 uM	Sigma Aldrich
FGF1	Fibroblast growth factor 1	Growth factor	40 ng/ml	Peprotech
IFNb	Interferon beta	Cytokine	18 ng/ml	pbl assay science
IFNg	Interferon gamma	Cytokine	40 ng/ml	pbl assay science
IONM	Ionomycin	Modulator of calcium	0,1 uM	Sigma Aldrich
RESV	Resveratrol	Modulator of metabolism	10 uM	Sigma Aldrich
ROTN	Rotenone	Modulator of reactive oxygen species	2,5 uM	Sigma Aldrich
WNT3A	Wingless-type MMTV integration site family member 3A	Modulator of development	200 ng/ml	RD systems

6.8. Pooled compound MoA screening

Pooled compound MoA screening was carried out by screening 4958 barcoded KBM7 knock-out cell lines against a panel of ten compounds (Table 6.3). Four KBM7 knock-out sub-pools comprising of approximately 1250 clones each were pooled together to obtain the 4958 knock-out collection used for screening. $5x10^6$ KBM7 pooled knock-out cells (each barcode represented with 1000 cells on average) were seeded to 6-well plates and treated with compounds (IC_{10}) for 8 d in duplicates. Cells were split and fresh compounds were applied every 3 d. At each passage $5x10^6$ cells were maintained to preserve the barcode representation. Compounds were provided by S. Kubicek (CeMM) as 10 mM stocks in DMSO and further diluted to 1000 x IC_{10} stocks. Pellets of $5x10^6$ treated and untreated cells were collected on day 0 and day 8. gDNAs were isolated with DNeasy Blood & Tissue kit (Qiagen) and samples subjected to barcode-sequencing (see below).

For the benchmarking experiment, two barcoded KBM7 clones were modified to ectopically express *Herpes simplex* virus thymidine kinase (HSV-TK) by retroviral transduction (see above). 3000 cells of each of the barcoded HSV-TK expressing control clones were spiked into a $1.5x10^7$ knock-out pool such that each barcode was represented by 3000 cells on average. Ganciclovir (Merck Millipore) treatment was performed for 8 d at 1 µg/ml. Pellets were taken and samples were prepared for sequencing as described above.

6.9. RNA-sequencing

Total RNAs of stimulated and unstimulated HAP1 cells were isolated using RNeasy Mini kit protocol (Qiagen) according to manufacturer's instructions. 500 ng total RNA was used for library generation with QuantSeq 3' mRNA-Seq library prep kit (Lexogen). Libraries were generated according to the manufacturer's T-fill protocol

for 96 samples except for the adjustment of 13 instead of 12 PCR cycles (Synthesis cycle: 30 sec at 98 °C, (10 sec at 98 °C, 20 sec at 65 °C, 30 sec at 72 °C) x 13, 1 min at 72 °C). Libraries were quantified using Qubit dsDNA HS assay on a Qubit 2.0 Fluorometric Quantitation System (Life Technologies) and 48 libraries at a time were pooled in equal concentrations for sequencing on a single Illumina HiSeq2000 lane. The final library pools were quality controlled with an Experion DNA 1K analysis kit on an Experion electrophoresis system (Bio-Rad). T-fill reactions on the cBot were carried out as described previously (Wilkening et al., 2013) with the exception that the T-fill solution was provided in a primer tube strip. Clusters were generated using the protocol SR Amp Lin Block TubeStripHyp v8.0.xml. Sequencing was performed with 50 bp single read v3 chemistry on a HiSeq2000 flow cell (Illumina).

6.10. Barcode-sequencing

gDNAs were isolated with DNeasy Blood & Tissue kit (Qiagen) following manufacturer's specifications. Barcodes were amplified from 4 µg gDNA (Synthesis cycle: 2 min at 94 °C, (30 sec at 94 °C, 30 sec at 60 °C, 40 sec at 68 °C) x 23), 7 min 68 °C) using GoTaq DNA polymerase (Promega) and primers flanking the barcode sequence (Table 6.5) as previously described (Bürckstümmer et al., 2013). Primers contained the Illumina sequencing adapters and multiplexing tags required for Illumina sequencing. PCR products were purified with QIAquick PCR purification kit protocol (Qiagen) and Agencourt AMPure XP beads (Beckman Coulter) according to Illumina TruSeq RNA prep protocol. Libraries were quantified using Qubit dsDNA HS assay on a Qubit 2.0 Fluorometric Quantitation System (Life Technologies). 12 libraries were pooled together in equal concentrations and the size distribution of the final libraries was assessed using Experion DNA 1K analysis kit and an Experion electrophoresis system (Bio-Rad). 24 libraries were sequenced using 50 bp paired-end chemistry and custom primers on 2 HiSeq2000 lanes (Illumina).

Table 6.5 Primers/adapters for barcode-sequencing

Name	Sequence
HG181	AATGATACGGCGACCACCGAGATCTACACAGAACTCGTCAGTTCCACCAC
HG1424	CAAGCAGAAGACGGCATACGAGATCCATACGGTACCACAGTGACTGGAGTTCAGACGTGTGC
HG1425	CAAGCAGAAGACGGCATACGAGATGGACAGTCCGTTGTTGTGACTGGAGTTCAGACGTGTGC
HG1426	CAAGCAGAAGACGGCATACGAGATGTGGTCTACTGATGTGTGACTGGAGTTCAGACGTGTGC
HG1427	CAAGCAGAAGACGGCATACGAGATCGCCGTTCTAACATCGTGACTGGAGTTCAGACGTGTGC
HG1428	CAAGCAGAAGACGGCATACGAGATATAGTAAGTGGATCCGTGACTGGAGTTCAGACGTGTGC
HG1429	CAAGCAGAAGACGGCATACGAGATAGAGTGTTCACCATGGTGACTGGAGTTCAGACGTGTGC
HG1430	CAAGCAGAAGACGGCATACGAGATGTTTATTACCGACTCGTGACTGGAGTTCAGACGTGTGC
HG1431	CAAGCAGAAGACGGCATACGAGATCCGGTGCATACCGGTGTGACTGGAGTTCAGACGTGTGC
HG1432	CAAGCAGAAGACGGCATACGAGATACGTGGGTTACATATGTGACTGGAGTTCAGACGTGTGC
HG1433	CAAGCAGAAGACGGCATACGAGATGACATTTTGAAAGTCGTGACTGGAGTTCAGACGTGTGC
HG1434	CAAGCAGAAGACGGCATACGAGATATCGATCCCTGTTCCGTGACTGGAGTTCAGACGTGTGC
HG1435	CAAGCAGAAGACGGCATACGAGATGGCACGTGATGGTATGTGACTGGAGTTCAGACGTGTGC
HG1436	CAAGCAGAAGACGGCATACGAGATACGTTATAGCATTTCGTGACTGGAGTTCAGACGTGTGC
HG1437	CAAGCAGAAGACGGCATACGAGATGCTATCATAAGTTCAGTGACTGGAGTTCAGACGTGTGC
HG1438	CAAGCAGAAGACGGCATACGAGATTAGGAGGCCTAGCCGGTGACTGGAGTTCAGACGTGTGC
HG1439	CAAGCAGAAGACGGCATACGAGATCTCTGGTGCGTAAATGTGACTGGAGTTCAGACGTGTGC
HG1440	CAAGCAGAAGACGGCATACGAGATCACGTTCACAGGTGAGTGACTGGAGTTCAGACGTGTGC
HG1441	CAAGCAGAAGACGGCATACGAGATAAGCTTTGAGCTCTCGTGACTGGAGTTCAGACGTGTGC

6.11. Western blotting

HAP1 cells were lysed in 4x SDS sample buffer (320 mM Tris-HCl pH 6.8, 40 % glycerol, 16 µg/ml bromphenol blue, 8 % SDS, 10 % 2-mercaptoethanol) and heated to 95 °C for 10 min. Whole cell lysates were separated on a NuPAGE 4-12 % Bis-Tris Gel (Invitrogen) at 130 V for 1.5 h and wet blotted onto a polyvinylidene difluoride (PVDF, Amersham Hybond-P, GE Healthcare) membrane at 400 mA for 2 h. PVDF membranes were blocked with 0.2 % Tropix I-Block (Applied Biosystems) in PBS containing 0.1 % Tween-20 for 1 h. Primary antibodies were applied in a 1:1000 or 1:2000 dilution in iBlock containing 0.1 % Tween-20 and 0.05 % sodium azide for 12 h at 4 °C. Primary antibodies used were: mouse anti-HIF1A (1:2000) from BD Biosciences (610959) and rabbit anti-actin (1:1000) from Sigma-Aldrich (A2066). Membranes were washed with PBS containing 0.1 % Tween-20 and incubated with goat-anti-rabbit (Biorad) or goat-anti-mouse (Biorad) HRP-conjugated secondary antibodies (1:10000 dilution in 0.2 % iBlock and 0.1 % Tween) for 1 h at RT. Western lightning Plus-ECL (PerkinElmer) was used for signal detection. Western blots were imaged on a Universal Hood III machine (Biorad).

6.12. Quantitative Real Time PCR

Total RNAs were isolated using the RNeasy Mini kit protocol (Qiagen) according to manufacturer's specifications. DNase digest was performed using the TURBO DNase kit (Ambion) and cDNA synthesis of 500 ng to 1 µg total RNA was carried out using the RevertAid Reverse Transcriptase kit (Fermentas) according to manufacturer's instructions (synthesis cycle for the reverse transcription: 10 min at 25 °C, 60 min at 42 °C and 10 min at 70 °C). qRT-PCRs were carried out in triplicates on a StepOnePlus system with 25 to 50 ng cDNA and 500 nM forward and reverse primers (Appendix Table 6.4) using KAPA ABI Prism SYBR Fast (KAPABIOSYSTEMS) or PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) according to manufactuer's protocols (synthesis cycle: 3 min at 95 °C, (3 sec at 95 °C, 30 sec at 60 °C) x 40). GAPDH was used as an endogenous control.

The computational analyses described below (sections 6.13. and 6.14.) were set up and performed by Tomasz Konopka. The computational analysis pipeline has been previously published in Gapp et al., 2016, the sections below were copied from this manuscript.

6.13. RNA-sequencing data processing and alignment

Unaligned reads in fastq format were trimmed of adapter sequence AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC using Cutadapt (v.1.2.1) and then partitioned using TriageTools (Fimereli et al., 2013) (v0.2.2) to select long (--length 35), high-quality (--quality 9), and sequence-complex (--lzw 0.33) reads. Selected reads were aligned using GSNAP (Wu & Nacu, 2010) (v2014-02-28) onto a custom genome index (gmap_build -k 14 -q 2) based on hg19 supplemented with

ERCC92 transcript sequences. Expression estimates on Gencode V19 genes were collected from the alignments using Exp3p. This procedure implements read counting on gene bodies and normalises by total sequencing depth; because the RNA-sequencing protocol is designed to capture one read per transcript through the polyadenylated tail, expression normalisation does not include the length of the gene body (taken from Gapp et al., 2016).

6.14. Expression analysis

Analysis was performed in a series of modules built around a custom toolkit, ExpCube. The analysis was split into two parts. The first part consisted of the analysis of four 96-well plates representing the stimulus discovery phase of the project. The second part was an extension to the entire data set (20x 96-well plates). Expression data from all samples were gathered into one object. This object included central estimates as well as intervals for each gene in each sample. Common steps in expression analysis are normalisation and batch correction. However, examining profiles of unstimulated wild-type HAP1 cells and controls showed that various implementations of these steps highlighted parts of the signal and hid others, making it difficult to select a unified scheme for the entire screen. Furthermore, the experimental design was such that most intended comparisons were between samples within a single batch, mitigating the need for explicit batch correction. For these reasons, the central expression values were not adjusted. Instead, within-plate and across-plate variation for unstimulated wild-type samples (of which there are four or more replicates per plate) was used to adjust uncertainty intervals. For each gene, quantiles among replicates in each plate and quantiles among group averages across plates were computed. This empirical variability was compared to the base Poisson intervals and was allowed to obtain a rescaling factor for each gene's Poisson interval. For 98 % of the genes, the across-plate variability was larger than

the within-plate variability (median ratio equal to 1.8), indicating the importance of replicates of wild-type controls in each of the library preparation plates. This interval rescaling operation was applied to all samples in the screen. Thus, empirical data on reproducibility of comparable samples from across the screen were incorporated into the expression profiles of all other samples (taken from Gapp et al., 2016).

Differential expression (DE) was scored based on effect sizes (fold changes) and uncertainty levels (z-scores, defined as differences in central expression values divided by a joint estimate of interval size). Outlier samples due to failed sequencing were excluded from the analysis. Considering groups A and B, each sample in group A was scored against each sample in group B, one gene at a time. A score of +1 was set for a z-score > 1.75 and a fold change > 1.75, a score of 0 was set for a z-score < 1.25 and a fold-change < 1.25, and a linear gradient of scores was set for intermediate cases (negative values for down-regulation). Through the z-score component, this approach penalised inconsistent/unreliable genes whose intervals were substantially modified in the previous step. A group-level DE score was obtained using the mean of the sample-level scores. By construction, these scores lie in [-1, 1] and carry the same interpretation independently of the number of samples per group, albeit with some variability with very few replicates. A gene was declared to be in a signature if the score was > 0.7 (taken from Gapp et al., 2016).

For the power analysis, raw data from 24 replicates of unstimulated wild-type cells from the stimulus discovery phase were pooled. The pool was then subset into bins of varying size and the alignment and expression calling pipeline was applied on each bin. From these expression profiles, the number of genes with expression above 1 transcript per million reads were computed. Hypothetical profiles with genes over- or under-expressed by various fold changes were created, and the criteria to call differential expression were applied. The number of genes called in this analysis reflects the sensitivity of the method to identify expression changes under uncertainty

due to low coverage and biological variability. This calculation is presented in the ExpCube package vignette (taken from Gapp et al., 2016).

For stimulus selection in the discovery phase, stimuli whose group signature contained at least two replicates and at least two signature genes were considered. Clustering of stimuli was performed using a Jaccard index distance between signature gene sets. Gene set enrichment analysis was performed by comparing signature genes with a background set of expressed genes in HAP1 cells using the topGO package (Alexa et al., 2006). In the screening phase, overlaps for each stimulus and each mutant cell line with the expected responses in wild-type cells were compared. Expression profiles were collapsed onto the ten selected signatures, and tSNE (van der Maaten and Hinton, 2008) clustering was performed based on Euclidean distances between groups using the dimensionally reduced data. For more detailed analysis, overlaps were correlated with technical features revealing unintentional relations with RNA concentration and depth. To correct for these effects, linear models of the form $O = aR + bD$, where O denotes overlap, R is average RNA concentration (ng per µl), D is average sequencing depth (millions of reads), and a and b are coefficients, were set up. A stimulus response score was defined as the residuals between observed and modelled overlap. Extreme values of this score identify outlying cell lines, i.e. mutants showing abnormal response given cell density and sequencing performance (taken from Gapp et al., 2016).

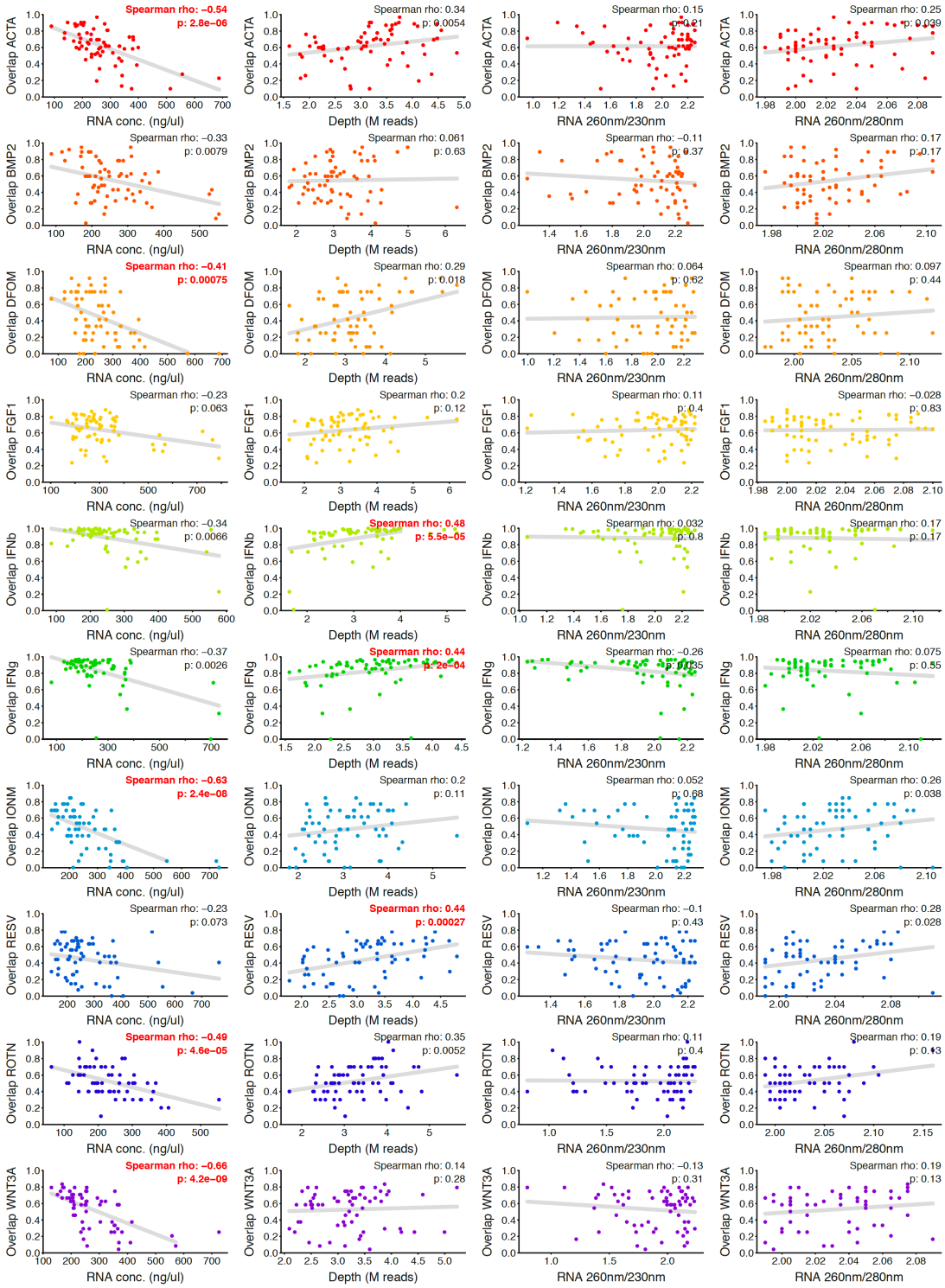
For comparison between RNA-sequencing and qRT-PCR data, slopes of best-fit lines between knock-out (KO) and wild-type (WT) responses plotted on logarithmic scales were computed. Linear fits on log axes suggest a model where KO-response is a power of the WT-response. The linear fit is a convenient summary of the overall patterns with few fitted parameters. In the case of RNA-sequencing data, the best-fit line was computed using signature genes with one outlier removed. In the case of

qRT-PCR, the line was fit using four signature genes and GAPDH (taken from Gapp et al., 2016).

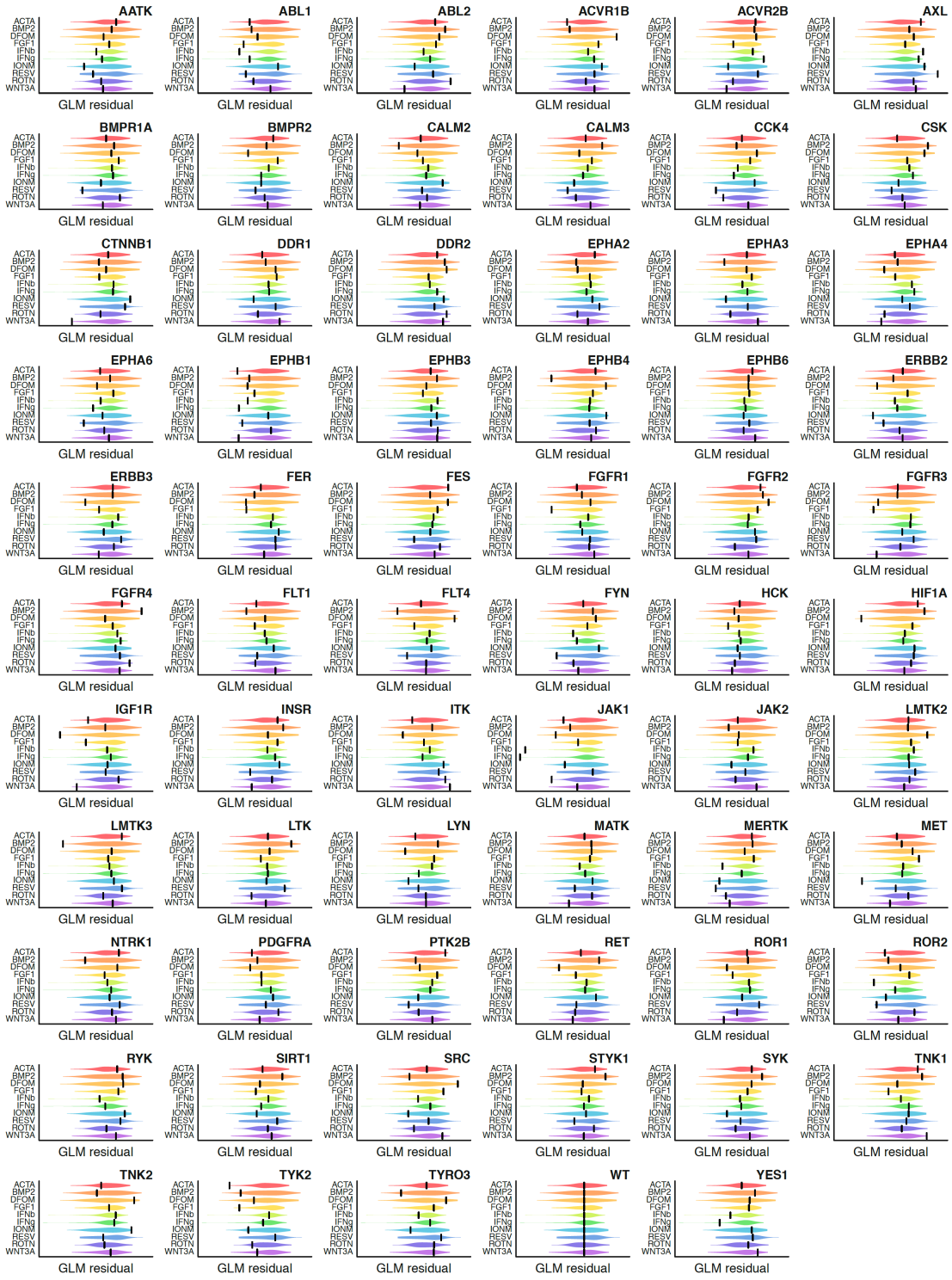
6.15. Availability

All raw sequencing data have been deposited in the European Nucleotide Archive under accession ERP012914. Exp3p software is available at https://github.com/tkonopka/Exp3p (v0.1). ExpCube software is available at https://github.com/tkonopka/ExpCube. Additional code, data files, and processed expression values are available at https://zenodo.org/record/51842 (taken from Gapp et al., 2016).

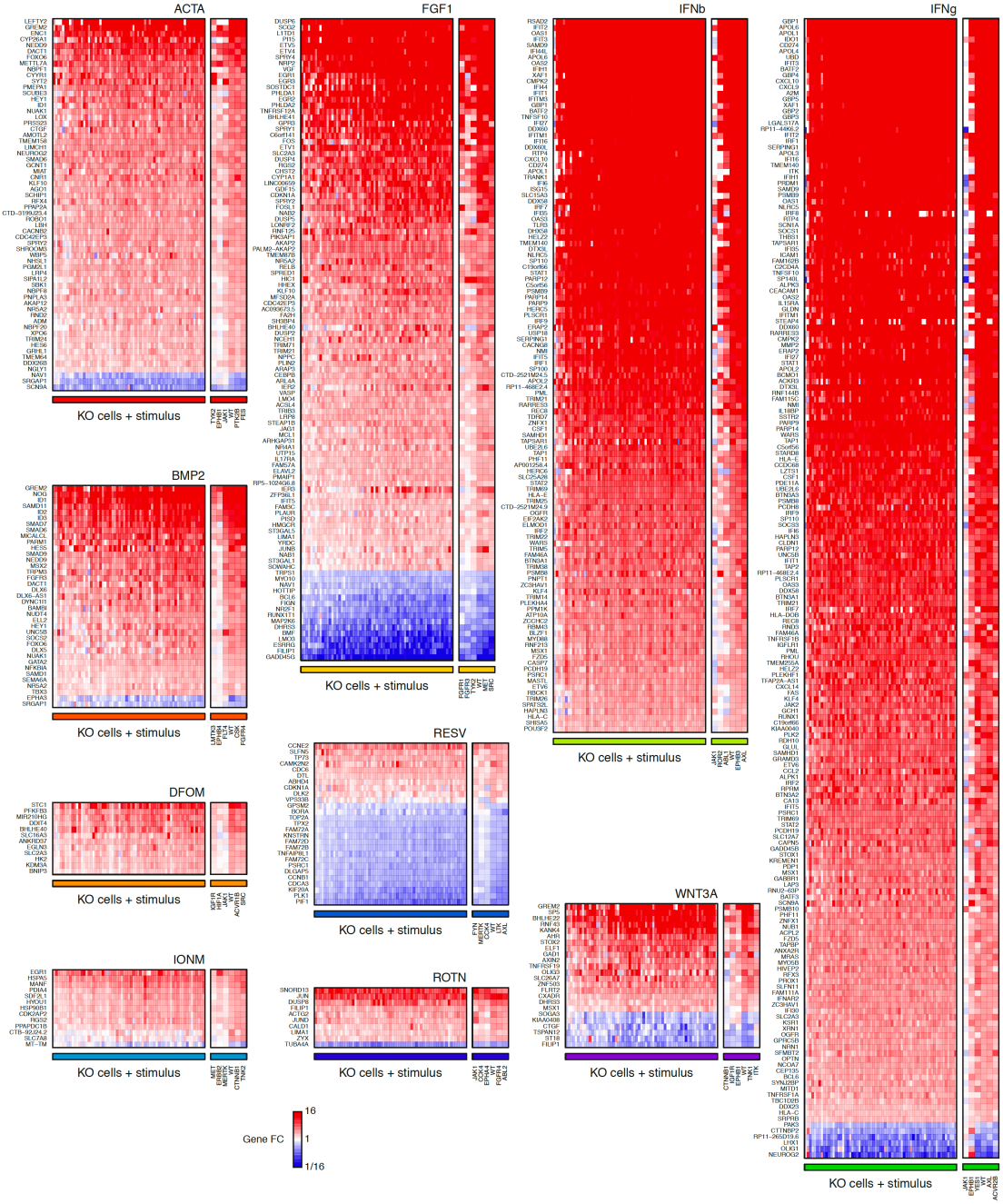
Chapter 7. Appendix



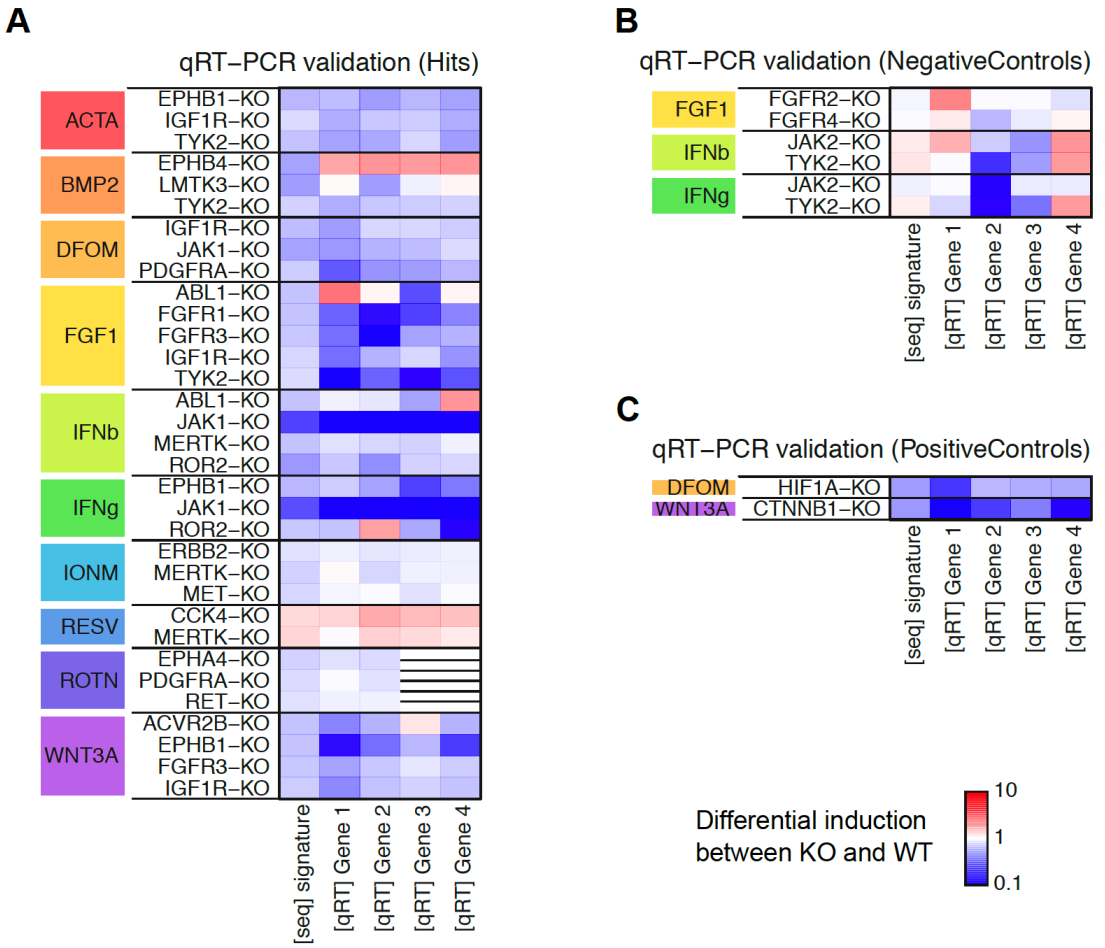
Appendix Figure 3.1 Correlations between transcriptional signature overlaps and technical parameters. The overlap between the transcriptional signatures of wild-type and knock-out cells across all ten stimuli is shown as a function of RNA concentration, sequencing depth and RNA quality (light absorbance of RNA samples at 260 nm vs 230 and 260 vs 280). Spearman correlations are indicated, correlation coefficients marked in red suggest significance. Taken from Gapp et al., 2016.



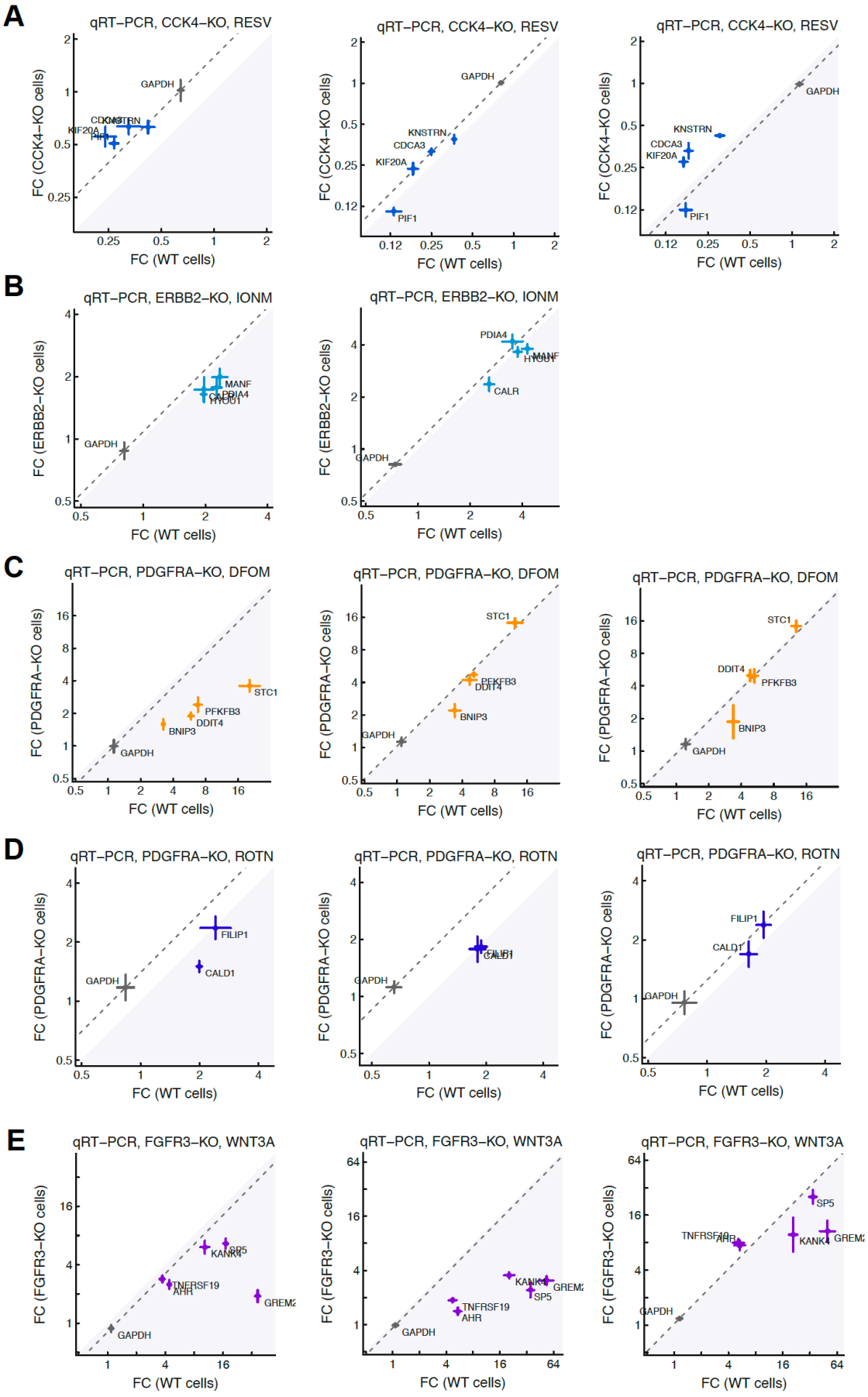
Appendix Figure 3.2 Knock-out cell line responses upon perturbation with ten stimuli. The 64 TK and control knock-outs used for reverse genetic screening are displayed as individual plots and indicated with a black bar. Each violin depicts the distribution of response scores of the entire knock-out panel across one of the ten parallel screens. Scores are overlaps of the transcriptional signatures between knock-out and wild-type samples corrected for technical parameters with general linear models (GLMs). Taken from Gapp et al., 2016.

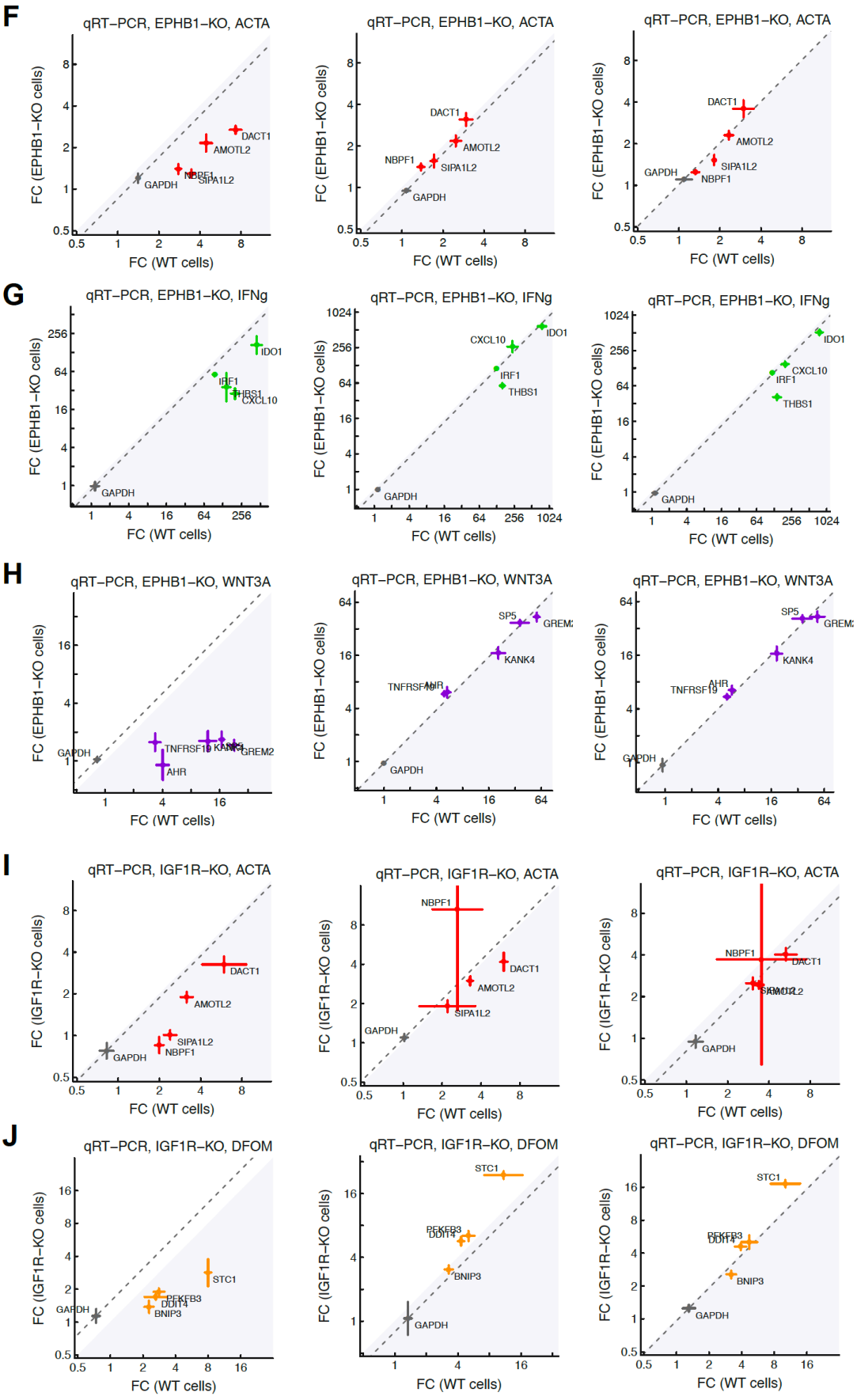


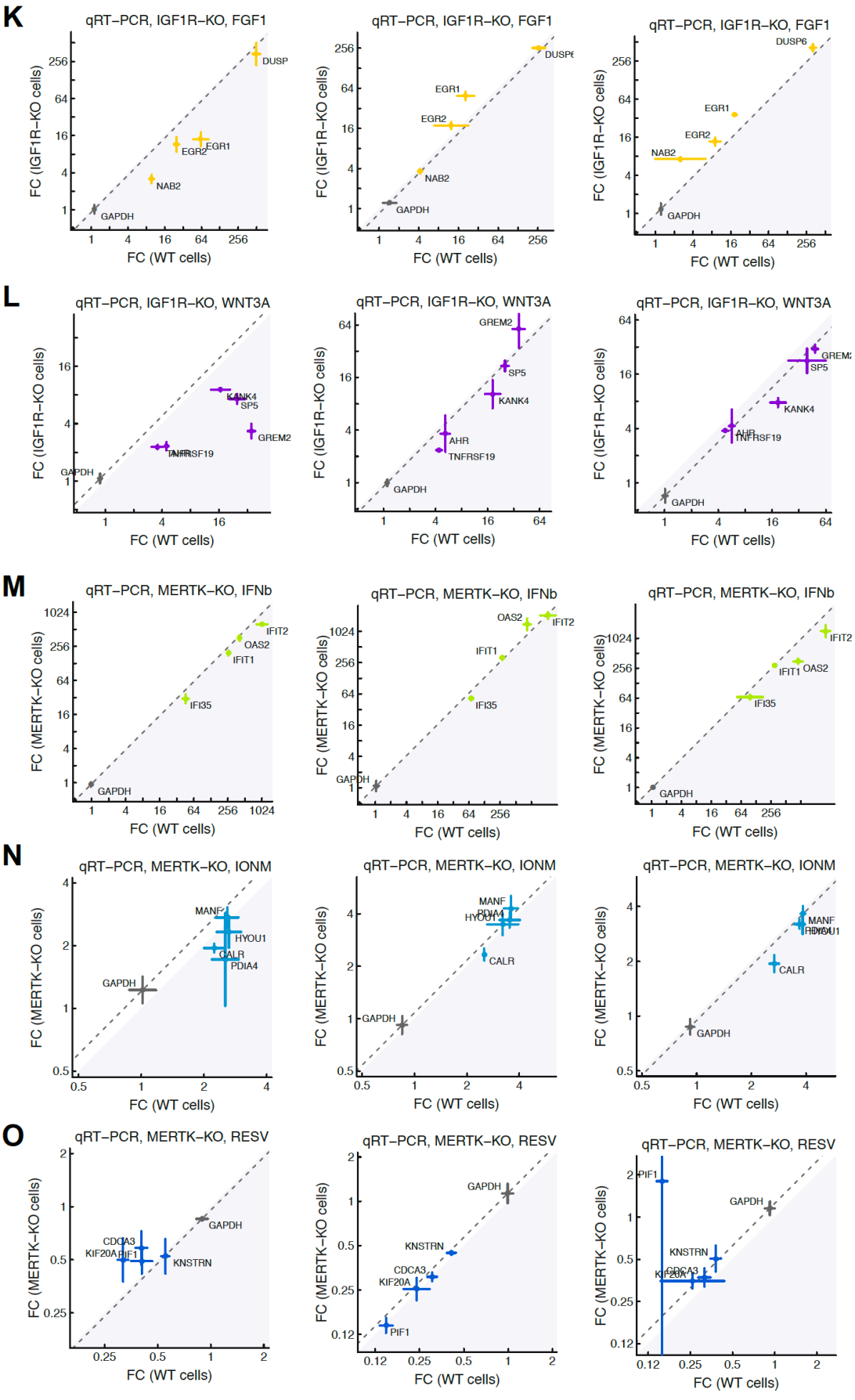
Appendix Figure 3.3 Transcriptional signatures of wild-type and knock-out cells. Transcriptional signatures of the 64 TK and control knock-outs as well as wild-type cells are shown after treatment with the ten stimuli. Transcriptional signature genes are displayed on the y-axis, wild-type and knock-out cell responses are shown and ranked based on their response score on the x-axis. Zoom-in on the right of each heatmap depicts the three knock-out cell lines with the smallest stimulus response score, the transcriptional response of wild-type cells and the two knock-out cells with the largest score upon stimulation. WT, wild-type; KO, knock-out; FC, fold change. Taken from Gapp et al., 2016.

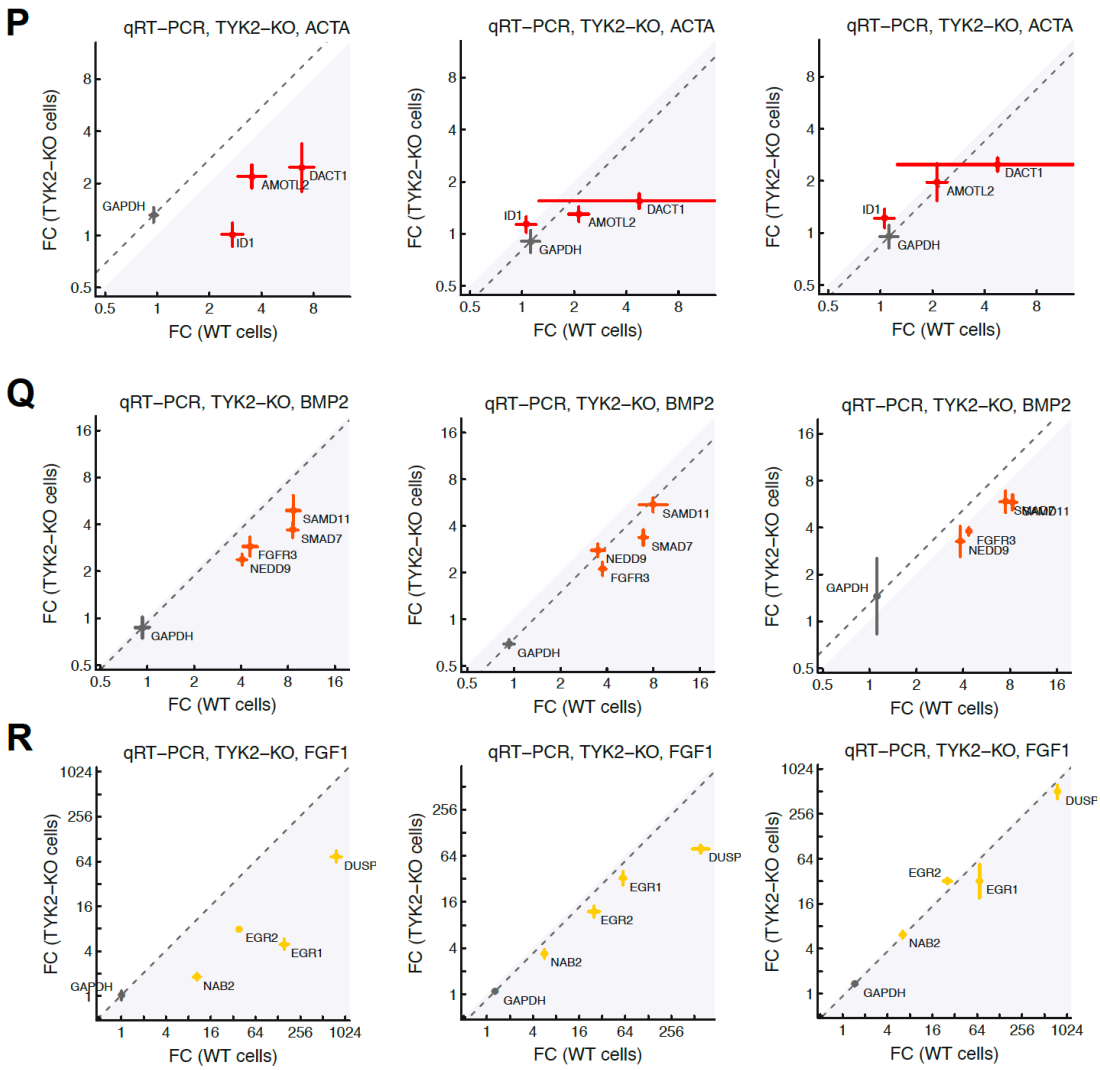


Appendix Figure 3.4 qRT-PCR validations of selected knock-out cell lines. A. Expression of selected transcriptional signature genes in wild-type vs knock-out cells across the ten stimuli as measured by RNA-sequencing and qRT-PCR. The first column indicates the average transcriptional response across signature genes determined by RNA-sequencing, the subsequent four columns indicate gene expression of four representative genes from each signature determined by qRT-PCR. B. As in A for negative interactions. C. As in A for positive interactions. Crossed-out fields indicate missing data. Taken from Gapp et al., 2016.

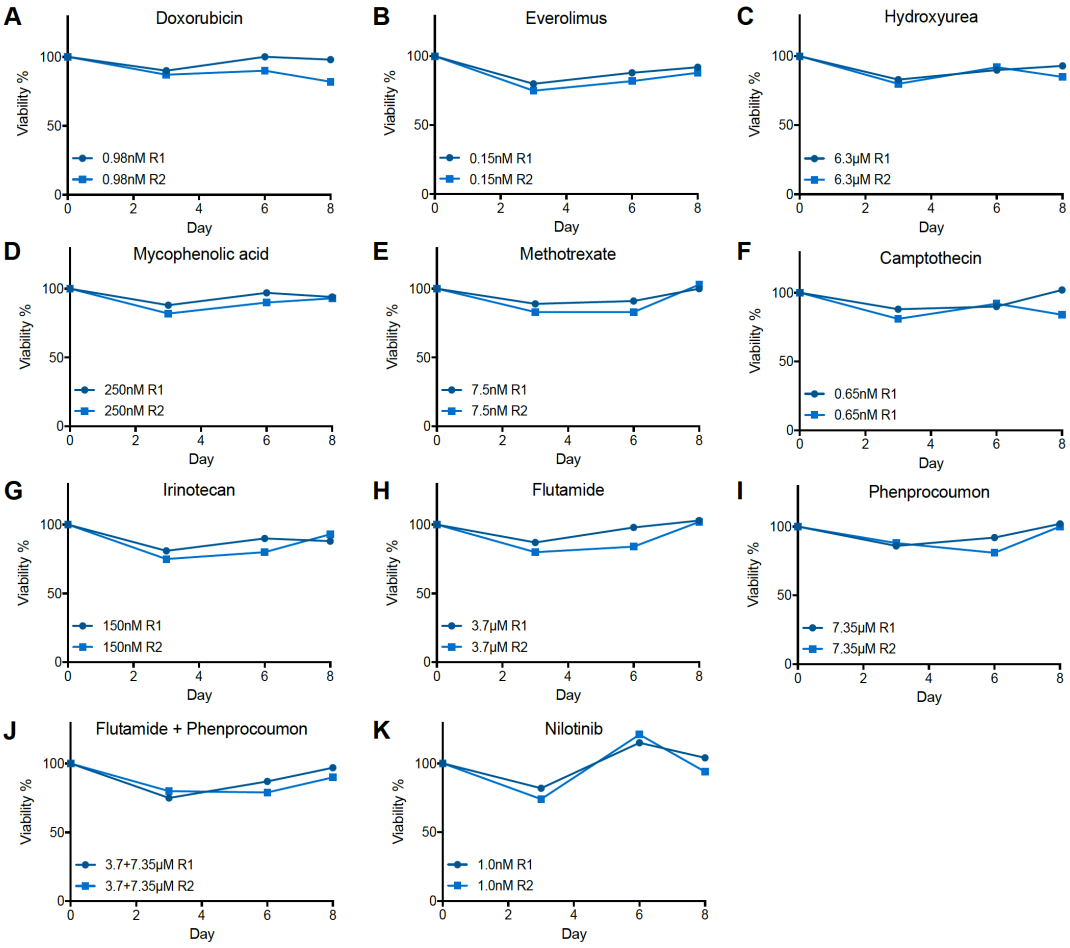

A
qRT–PCR, CCK4–KO, RESV
FC (CCK4–KO cells)
FC (WT cells)
GAPDH
KIF20A
CDCA3
KNSTRN
PIF1
B
qRT–PCR, ERBB2–KO, IONM
FC (ERBB2–KO cells)
FC (WT cells)
GAPDH
MANF
PDIA4
HYOU1
CALR
C
qRT–PCR, PDGFRA–KO, DFOM
FC (PDGFRA–KO cells)
FC (WT cells)
GAPDH
STC1
PFKFB3
DDIT4
BNIP3
D
qRT–PCR, PDGFRA–KO, ROTN
FC (PDGFRA–KO cells)
FC (WT cells)
GAPDH
FILIP1
CALD1
E
qRT–PCR, FGFR3–KO, WNT3A
FC (FGFR3–KO cells)
FC (WT cells)
GAPDH
SP5
KANK4
TNFRSF19
AHR
GREM2


F
qRT–PCR, EPHB1–KO, ACTA
FC (EPHB1–KO cells)
FC (WT cells)
DACT1
AMOTL2
NBPF1
SIPA1L2
GAPDH
G
qRT–PCR, EPHB1–KO, IFNg
IDO1
IRF1
THBS1
CXCL10
H
qRT–PCR, EPHB1–KO, WNT3A
TNFRSF19
KANK4
AHR
SP5
GREM2
I
qRT–PCR, IGF1R–KO, ACTA
FC (IGF1R–KO cells)
J
qRT–PCR, IGF1R–KO, DFOM
STC1
PPP1R15A
BNIP3


K
qRT–PCR, IGF1R–KO, FGF1
FC (IGF1R–KO cells)
FC (WT cells)
DUSP
EGR2
EGR1
NAB2
GAPDH
qRT–PCR, IGF1R–KO, FGF1
FC (IGF1R–KO cells)
FC (WT cells)
EGR1
EGR2
NAB2
GAPDH
qRT–PCR, IGF1R–KO, FGF1
FC (IGF1R–KO cells)
FC (WT cells)
DUSP6
EGR1
EGR2
NAB2
GAPDH
L
qRT–PCR, IGF1R–KO, WNT3A
FC (IGF1R–KO cells)
FC (WT cells)
KANK4
SP5
GREM2
GAPDH
qRT–PCR, IGF1R–KO, WNT3A
FC (IGF1R–KO cells)
FC (WT cells)
GREM2
SP5
KANK4
AHR
TNFRSF19
GAPDH
qRT–PCR, IGF1R–KO, WNT3A
FC (IGF1R–KO cells)
FC (WT cells)
SP5
KANK4
TNFRSF19
GAPDH
M
qRT–PCR, MERTK–KO, IFNb
FC (MERTK–KO cells)
FC (WT cells)
IFIT2
OAS2
IFIT1
IFI35
GAPDH
qRT–PCR, MERTK–KO, IFNb
FC (MERTK–KO cells)
FC (WT cells)
OAS2
IFIT2
IFIT1
IFI35
GAPDH
qRT–PCR, MERTK–KO, IFNb
FC (MERTK–KO cells)
FC (WT cells)
IFIT2
IFIT1
OAS2
IFI35
GAPDH
N
qRT–PCR, MERTK–KO, IONM
FC (MERTK–KO cells)
FC (WT cells)
MANF
HYOU1
CALR
PDIA4
GAPDH
qRT–PCR, MERTK–KO, IONM
FC (MERTK–KO cells)
FC (WT cells)
MANF
CALR
GAPDH
qRT–PCR, MERTK–KO, IONM
FC (MERTK–KO cells)
FC (WT cells)
CALR
GAPDH
O
qRT–PCR, MERTK–KO, RESV
FC (MERTK–KO cells)
FC (WT cells)
GAPDH
KNSTRN
qRT–PCR, MERTK–KO, RESV
FC (MERTK–KO cells)
FC (WT cells)
GAPDH
KNSTRN
CDCA3
KIF20A
PIF1
qRT–PCR, MERTK–KO, RESV
FC (MERTK–KO cells)
FC (WT cells)
PIF1
GAPDH
KNSTRN



Appendix Figure 3.5 qRT-PCR validation of selected interactions. A-R. Transcriptional responses of indicated knock-out cells upon stimulation with one of the ten stimuli. Fold changes (FC) of two to five transcriptional signature genes (coloured) and GAPDH (gray) upon stimulation are compared between wild-type and knock-out cells. Error bars represent the standard deviation of three technical replicates. Left panels show knock-out clones used for reverse genetic screening. Middle and right panels display two independently generated knock-out clones. WT, wild-type; KO, knock-out.



Appendix Figure 4.1 Validation of IC_{10} concentrations of different drugs on KBM7 wild-type cells. A-K. KBM7 wild-type cells treated with the IC_{10} concentration of ten drugs and one drug combination as indicated for eight days. Viability upon treatment is normalised to untreated controls.

Appendix Table 4.1 Compound MoA screening hits

Compound	Gene KOs increased viability	Gene KOs decreased viability
Doxorubicin	ETV6, LOC349160	AKNA, ARHGAP22, C12orf42, CHPT1, FAM78A, LOC283887
Everolimus	CTSE, EEPD1, GLYAT, GRIK1, RARA, ST3GAL3, TK2, UTRN	COMT, PRKCE, SSBP2
Hydroxyurea	CDKN2C, COMMD1, GRIK1, LOC349160, NAIF1, RARA, SH3BP5, UTRN	FAM78A, GNRHR2, STK10
Mycophenolic acid	COMMD1, CTSE, EPB41, FRY, RARA, TK2	C12orf42, GPR114, NUP62CL, WIPF1
Methotrexate	FRY, GRIK1, INPP4A, LOC349160, NAIF1, TK2	ANKS1A, SETDB1, SLK
Camptothecin	FRY, SOCS4, XYLT1	ANKS1A, COMT, CSK, MAPK6, P2RX1, SMYD3, SRCRB4D
Irinotecan	AKAP13, FRY, SH3BP5, UTRN	ARMC8, B3GNTL1, CADPS2, FNBP1L, INPP4B, TRPM3, TUBGCP3, ZNF521
Flutamide	CNOT8, EVI2B, NF1, FRY, GLYAT, GRIK1, LOC100216545, LOC349160, RAB1A, TATDN3	C20orf94, FAM78A, LOC81691
Phenprocoumon	FRY, LOC349160, NAIF1, SH3BP5	LOC100652927, LOC283887, SLC30A
Flutamide+Phenprocoumon	COMMD1, FRY, GRIK1, RPTOR	CD79A, COMT, FNBP1L, PARVG
Nilotinib	AKAP13, EVI2B, NF1, GRIK1	AIMP1, ARHGAP22, C20orf94, CDC25C, COMT, HOTTIP, HTR6, LDLRAD4, PRKCE, WIPF1

Appendix Table 6.1 HAP1 knock-out cell lines for reverse genetic screening

Gene	Group	Subfamily	Clone	gRNA sequence	Indel	Exon
AATK	TK	LMR	354-20	GAGTGGCAGGACGTACACGT	7bp del	4
ABL1	TK	ABL	243-16	TGGGGCTGGATAATGGAGCG	11bp del	3
ABL2	TK	ABL	29163-14	AGTTCGCTCTAAGAATGGGC	10bp del	3
AXL	TK	AXL	695-12	CGTAGCACTAATGTTCTCAG	1bp ins	8
CCK4	TK	CCK4	276-9	CGGGCCCGGTACATGTGTAC	7bp del	2
CSK	TK	CSK	198-3	GGTTCCACGGCAAGATCACA	13bp del	6
DDR1	TK	DDR	1244-13	CCCCCTAGGTTGTGGCGCAT	2bp del	8
DDR2	TK	DDR	694-27	ATCTCTTGGCGGAACCGTCA	1bp del	5
EPHA2	TK	EPH	222-11	CAACGTGATGTCTGGCGACC	16bp del	3
EPHA3	TK	EPH	295-11	TGCTCTGTTCTCGACAGCTT	17bp del	1
EPHA4	TK	EPH	223-7	AACACCGAGATCCGGGATGT	7bp del	3
EPHA6	TK	EPH	143-5	GTGATAGCCACCGCCGCTGT	10bp del	8
EPHB1	TK	EPH	699-28	GACGATGGAGATAACATTGC	14bp del	5
EPHB3	TK	EPH	707-29	GTTGTCATCACAGCGTGAGC	11bp del	5
EPHB4	TK	EPH	1278-11	GAATGTCAAGACGCTGCGTC	13bp del	4
EPHB6	TK	EPH	46-15	CTCAGCCTGACGGTAGTAAA	1bp ins	4
ERBB2	TK	EGFR	342-2	TTCGAAGCTGCAGCTCCCGC	10bp del	6
ERBB3	TK	EGFR	211-30	ACCATTGCCCAACCTCCGCG	4bp del	3
FER	TK	FER	1349-2	CAGTATGTATTGGCGTTGAA	1bp ins	4
FES	TK	FER	307-16	GGCCGAGCTTCGTCTACTGG	38bp del	2
FGFR1	TK	FGFR	1427-12	CAAGGTCGGGGACGGCCTAG	16bp del	1
FGFR2	TK	FGFR	203-10	TCTCCGCAGGGGGATACGTT	13bp del	8
FGFR3	TK	FGFR	1346-24	GCCAACCAGACGGCGGTGCT	5bp del	7
FGFR4	TK	FGFR	257-7	CAGAATCTCACCTTGATTAC	8bp del	2
FLT1	TK	VEGFR	358-10	TGATGTTAGGTGACGTAACC	5bp del	4
FLT4	TK	VEGFR	1246-39	GTGTGGGCTGAGTTTAACTC	5bp del	6
FYN	TK	SRC	22023-8	TGAACTCTTCGTCTCATACG	13bp del	1
HCK	TK	SRC	8-13	ACACTGTCCTGTGTACGTGC	20bp del	2
IGF1R	TK	INSR	184-15	TGTGGGGAATAAGCCCCCAA	7bp del	2
INSR	TK	INSR	730-14	ATATCCGGAACAACCTCACT	22bp del	2
ITK	TK	TEC	11-4	GACGATCTTCAAAGTATGCC	1bp ins	1
JAK1	TK	JAKA	12-1	GGAAGTGATCTTCTATCTGT	16bp del	3
JAK2	TK	JAKA	13-10	CCTCCATTTCTGTCATCGTA	1bp ins	3
LMTK2	TK	LMR	708-8	CTCTTGGGAGAGATTTACAC	5bp del	5
LMTK3	TK	LMR	702-4	ATGAGGACCACAGCGTAGGG	1bp ins	3

LTK	TK	ALK	357-13	CTGGCTCCAAGATACTAGGC	13bp del	2
LYN	TK	SRC	1664-23	GTAGCCTTGTACCCCTATGA	22bp del	3
MATK	TK	CSK	154-10	CTCGCATTTGGTGATACACT	2bp del	3
MERTK	TK	AXL	15332-14	GGTCCCGGGAATAGCGGGTA	2bp del	2
MET	TK	MET	26-18	CCTCGTGCTCCTGTTTACCT	82bp ins	2
NTRK1	TK	TRK	1259-15	CGGCAGCCAATGAGACCGTG	32bp del	8
PDGFRA	TK	PDGFR	236-7	CAGCCTAAGACCAGGAACGC	176bp ins	2
PTK2B	TK	FAK	350-12	TAAAGTTGGGCACGTTACGC	20bp del	2
RET	TK	RET	1368-11	ATTCGGGAGAACCGACCCCC	2bp del	3
ROR1	TK	ROR	1343-15	AAGACTCCATATAGACGGTG	8bp del	5
ROR2	TK	ROR	336-4	CGGGCCGTCCTGCCCATCAA	5bp del	2
RYK	TK	RYK	8066-9	TCTGCTAGTGTTTTTGTCCA	2bp ins	5
SRC	TK	SRC	348-23	CCCTCTATGACTATGAGTCT	8bp del	5
STYK1	TK	TK-Uni	70-9	GTCCCTAGGTGGAGGAACAG	11bp del	5
SYK	TK	SYK	6898-17	CGCCAGAGCCGCAACTACCT	29bp del	2
TNK1	TK	ACK	177-2	TGGGCCTAAGTCTAAGAACT	26bp del	3
TNK2	TK	ACK	2454-26	ACGCAAGTCGTGGATGAGTA	13bp del	3
TYK2	TK	JAKA	36-3	CCCAACGGGCTTACTGCCCC	1bp del	3
TYRO3	TK	AXL	115-15	GGTGAAGCTGACAGTGTCTC	2bp ins	2
YES1	TK	SRC	22696-7	TAGTGGGTTCTGCTCCATAA	1bp del	2
ACVR1B	TKL	STKR	254-9	AAAGTGGAGCTGGTCCCTGC	5bp del	2
ACVR2B	TKL	STKR	196-21	GCCCTTCTTCACGAGCTCGA	8bp del	2
BMPR1A	TKL	STKR	30151-5	GCCGGACAATAGAATGTTGT	5 bp del	4
BMPR2	TKL	STKR	76-11	GTTCCATTCTGAATTGAGGG	4bp ins	3
CALM2	Other	NA	1217-12	GAACTGTAATGAGATCTCTT	4bp del	3
CALM3	Other	NA	1214-3	GAACGGGACCATTGACTTCC	7bp del	4
CTNNB1	Other	NA	1032-12	GAAAAGCGGCTGTTAGTCAC	25bp del	3
HIF1A	Other	NA	1165-10	TTCTTTACTTCGCCGAGATC	1bp ins	2
SIRT1	Other	NA	1008-2	TCAATATCAAACATCGCTTG	23bp del	4

Appendix Table 6.2 Primers for Sanger sequencing validation

Gene	Clone	Forward primer	Reverse primer
AATK	354-20	CCTTCCTTGACCTGCGAGAAG	CACTCCAGAGCTGAGTCTTCC
ABL1	243-16	AGAGAAAGACAGCAGAAGTGATCTT	CTCGTACACCTCCCCGTACT
ABL2	29163-14	ATCTGCAGTGGTATTGATCCTGTAG	CTAGGTGAAAAGCTACGAGTCCTT
AXL	695-12	TCCCCAGAATGTGAATGTGA	CCCCTTTCACTCCTTTACCC
CCK4	276-9	ACACCTGCACTTACTGGACG	ATTCTGCCTGGCCCGTTTAA
CSK	198-3	GGGTACCAAACTCAGCCTCA	TTGCTGGCATGGTACATGAT
DDR1	1244-13	AGAAGTTTCGGTGGGTGATG	TGCTCCATCCCACATAGTCA
DDR2	694-27	CTGTGGCAAGAACCCAAAAT	CTTTCCCCACAGCCTTGATA
EPHA2	222-11	CCTGCAGGCTCACAACCTAA	TCCTCCACGTTCAGCTTCAC
EPHA3	295-11	CCTCCTTATCTCCAGTGTCAAACTT	AAAAAGGCACATCTTTCCTTCACTC
EPHA4	223-7	GACGTATCAGCCCCTGTGAT	TGCTTTGGATCTTTGCAGTGGG
EPHA6	143-5	TAATAGTGAAATGTCACCCAGGGAA	ACCTCAGGTGTGCTTTGTTATATCT
EPHB1	699-28	ACTTACCGGCTTGGTTTGTG	GTCTTTCCCTTGGCACACAT
EPHB3	707-29	CAGTGGTCCCTTAGCAGGTC	AGGAGTCTGCCTGGTGAAGA
EPHB4	1278-11	AAGATTCTGGAACGGGAACC	AGAGCAAGCCTCCATTTCAA
EPHB6	46-15	ATGTCCCCTTCCCTATAACCTCT	CTCTCGTCTGCTGCAATTGTG
ERBB2	342-2	CTCATCGCTCACAACCAAGTGA	GAATTATATGTGCTGACGCAAGCTA
ERBB3	211-30	GGCAACAAGAGCGTAACTCC	TTGCTTAGGGAGCAATGGACCT
FER	1349-2	CTCCTTGGCAACCTTTGTGT	TAACCCAGTCCAGGGGGTAT
FES	307-16	GAGAGGAAGGAGAAGGCAGATACAG	CAGTCCATCCTGACCCTACAGT
FGFR1	1427-12	GTTTATAACCAGACCAGAACCAACC	CTCTAACTGCAGAACTGGGATGT
FGFR2	203-10	TGGACCCCATTGATCCAAGCA	CTTGAGAATGGTCGTCGCCT
FGFR3	1346-24	CCATCTCCTGGCTGAAGAAC	CCTAGACCCAAATCCTCACG
FGFR4	257-7	GGAGCTGACAGTAGCCCTTG	GGTGTGTCCAGTAGGGTGCT
FLT1	358-10	ATTAAGGGAGAATGACGTGACTGTG	ATGTTACAGGAAAGTCCACTGAGAA
FLT4	1246-39	CTGTACCTGCAGTGCGAGAC	CCACACAGAACTCTCCAGCA
FYN	22023-8	CAATTGCCAAAAGATTTAAGGGTGG	TAAAGAAGCAACAAAACTGACGGAG
HCK	8-13	ACAGCTCTACCTGTTACATTTCTGT	CATTGTCACTTCAGCTGCTATCATT
IGF1R	184-15	TACCTCCACATCCTGCTCATC	GCAAGTGTCTAGCCCAGGAG
INSR	730-14	CCTTTCTCCTTCTGCACCTG	AAGATGACCAGCGCGTAGTT
ITK	11-4	GCCATTAGCAAAGGCTGTATGTATT	TTAGAACTGAAATCACCTAGCACCA
JAK1	12-1	ATTGTACAGATGTGGGGAAATGAGA	CACTCCCTTTTGCCATAAGAACTTT
JAK2	13-10	CTTTGTGTATGAAGGGTGCCTTATC	AGATTTCCCAAGGGAATGGTAAAGA
LMTK2	708-8	CAAGCTTTTGGTGAGGTGGT	TGCAATTGCTCACAAAGCAT

LMTK3	702-4	GGATTTACTGCCCCCTTCTC	GCCTGCAGCTGAGAGCTAGT
LTK	357-13	ATATCCTATTTCCACCCGCCTTAAC	TAGAAAACAGCCAAGACCCCTC
LYN	1664-23	CTCTTAGTGCTTCCTCTCATCCTTT	TTACTCCTCCAGGACTTTCATCTTC
MATK	154-10	TACCAGCTCTTGTTCTGCCA	ATCTCCGTGATGTCCTCCCTTT
MERTK	15332-14	GAATTCTGCTTATCTTTTCCTGGGG	TATGTGTCCAACTGTGTGTTTGAAG
MET	26-18	CTATGTAAAAGTCCAGTTGGGAAGC	TGAAGTTGGGAAGCTGATACTTCAT
NTRK1	1259-15	CCTCCTGCTGTTGCTCTTTC	CTCCTTACCCTCTCCCCAAG
PDGFRA	236-7	CAGAAGGTTTTGGCTTCAGG	AACTGCCACTGGAGAGCATT
PTK2B	350-12	TATAGCATTTGTGTCTGTCTTGGAG	TTTGACCAGTTTGAAGTTTTTCCCA
RET	1368-11	CTACCTCAAGGTCTTCCTGTCAC	GAATAAACCAGATGGCTTGTGTCAA
ROR1	1343-15	CCTGGTACGGTGAAAGCACT	TGACTTTCTGGGGGTGTAGG
ROR2	336-4	CTCCAAGTCAAGAAACACAGGAATC	AAGCAAATTAAAGGCTCCTGAAAGG
RYK	8066-9	CAGTCTACAAAGCACCAGCAAAATA	CTGAGTTCCAGCAATGTTCTGATAA
SRC	348-23	CTTTGAAGTCCCACCACCCAG	CTTCACTGAACCTGACTGTGTCTTT
STYK1	70-9	AGTGTTCATATTGGCTCGAAAGATG	AATTTGTTTCAGGGTCTGCTTTCC
SYK	6898-17	ACCAAAGTTCTCTGTCTCTGTCTTG	ATGGTGTAGTGGTGTGCCTT
TNK1	177-2	CCTTTTCTTCCCTGTAAGTCTCTCC	CAAAACCTCCAAGGATCTGAAACAG
TNK2	2454-26	GTTGGGGTTGATTTCATTGC	GAACGTCTTTGCTTGGCTTC
TYK2	36-3	AGACTCACCAACTTTATGTGCAATG	ACCCTGTCCAATAAATGTGGAAAAC
TYRO3	115-15	AATGGGAACTGTTTACAGAGACAGA	TTCTCAGATACCCAACTCCAACTG
YES1	22696-7	GATATGAACTTGGCACCACTGAAAA	AAGTCCAGCCATTAAATACAGACCT
ACVR1B	254-9	GTTTGGAAAAGGACCTGCAA	TTCTGGTCCCTCCTCACATC
ACVR2B	196-21	CCAGCTGGCACACGTAAATT	GCTACCATCTCAAGAGCCCC
BMPR1A	30151-5	ACATATAAAATTTGCAGGCCCCTTT	TAGGTAGCCTGGTTACTGTAAAACC
BMPR2	76-11	GCAAAACTGTTTCATAGCTTACACG	GATTTCAGGTGTTTGCCAGATGTAA
CALM2	1217-12	ATTGTGCCATTACCTGAAATGGTTT	ATTAATCTGTGTATTTGGATTGTGGT
CALM3	1214-3	TATGATCAATGAGGTGGATGCAGAT	GACCTCCTGGAAGTGGTAGCTC
CTNNB1	1032-12	CTTGGCTGTCTTTCAGATTTGACTT	TACCAGCTACTTGTTCTTGAGTGAA
HIF1A	1165-10	ACATGGCATCTTCTAATCCTTCTGT	AACATTGCGACCACCTTCTAAAAAT
SIRT1	1008-2	TTTCTGAAATAAGGGTAGGGTTGTA	GGTGGCTGTATGTGAACAAAATACT
TYK2	36-8	AGACTCACCAACTTTATGTGCAATG	ACCCTGTCCAATAAATGTGGAAAAC
TYK2	2840-2	GAAATAGGCATAGGTTGGGGTGTAG	GGACCCTAGTCACCATGACATC
TYK2	3415-10	GAAATAAAACCTGCAGGAAGGAGG	GTGAAGGCAGACATTGGGTCT
TYK2	3415-11	GAAATAAAACCTGCAGGAAGGAGG	GTGAAGGCAGACATTGGGTCT
ACVR2B	2842-7	GTGGGAGTTGGATCATGATGTTAAG	GAATCTCTACATTTCCCAGCTGACT
ACVR2B	3427-2	GAGTGGGAGATAAAGGACCAAGAAG	CTCACACTTCCAGAGCCAGACTAC
ACVR2B	3427-9	GAGTGGGAGATAAAGGACCAAGAAG	CTCACACTTCCAGAGCCAGACTAC
CCK4	SC1	TGCTGCATGCTTGTGCAAAT	GCCGGCATAGGGCAATATCT
CCK4	SC2	TGCTGCATGCTTGTGCAAAT	GCCGGCATAGGGCAATATCT
EPHB1	SC1	GCATAATCTTGGCGGGCTTC	CTGGTGCTGTCTTTGCATGG
EPHB1	SC2	GCATAATCTTGGCGGGCTTC	CTGGTGCTGTCTTTGCATGG
ERBB2	SC1	CTCTACTTCCCAAGCACGCA	CCATCCAGACATGACCTCGG
FGFR3	SC1	ATCTGGGTCAGAGCCTTCCT	AATGACCAGAGAGACCCCCA
FGFR3	SC2	ATCTGGGTCAGAGCCTTCCT	AATGACCAGAGAGACCCCCA
IGF1R	SC1	GTGGGGGTGAGGATTTCGTAG	TGCACCCGGATGAAACAACC
IGF1R	SC2	GTGGGGGTGAGGATTTCGTAG	TGCACCCGGATGAAACAACC
MERTK	SC1	CAGGTTCGGGACGTCCATCT	CTCGTCGGAACTTCAGGACT
MERTK	SC2	CAGGTTCGGGACGTCCATCT	CTCGTCGGAACTTCAGGACT
PDGFRA	SC1	CAGAAGGTTTTGGCTTCAGG	AACTGCCACTGGAGAGCATT
PDGFRA	SC2	CAGAAGGTTTTGGCTTCAGG	AACTGCCACTGGAGAGCATT
ROR2	SC1	TCGGGCGAGATGCGAATG	TTCAAGCGTTCCGAGGCG
ROR2	SC2	TCGGGCGAGATGCGAATG	TTCAAGCGTTCCGAGGCG

Appendix Table 6.3 Stimuli

Stimulus	Name	Description	Conc.	Vendor
5FU	Fluorouracil	DNA damaging agent	50 uM	Sigma Aldrich
ACTA	Activin A	Modulator of development	4 ng/ml	Peprotech
ACTB	Activin B	Modulator of development	8 ng/ml	Peprotech
APHD	Aphidicolin	DNA damaging agent	4 uM	Sigma Aldrich
bNGF	beta nerve growth factor	Growth factor	4 ng/ml	Peprotech
BDNF	Brain-derived neurotrophic factor	Growth factor	4000 ng/ml	Peprotech
BMP13	Bone morphogenic protein 13	Modulator of development	1000 ng/ml	Peprotech
BMP2	Bone morphogenic protein 2	Modulator of development	220 ng/ml	Peprotech
BROM	Bromopyruvic acid	Modulator of metabolism	4 uM	Sigma Aldrich
CAMP	Camptothecin	DNA damaging agent	0,004 uM	Sigma Aldrich
CSPL	Cisplatin	DNA damaging agent	200 uM	Sigma Aldrich
DFOM	Deferoxamine	Hypoxia inducing agent	50 uM	Sigma Aldrich
DMSO	Dimetyl sulfoxide	Solvent	0,1 %	Sigma Aldrich

DOXR	Doxorubicin	DNA damaging agent	0,04 uM	Sigma Aldrich
ENKP	Enkephalin	Ligand	40 ng/ml	Sigma Aldrich
EPO	Erythropoietin	Cytokine	1200 ng/ml	Sigma Aldrich
FasL	Fas Ligand	Cytokine	4 ng/ml	Enzo Life Sciences
FBS	Foetal bovine serum	Mixture including growth factors	10 %	Gibco
FGF1	Fibroblast growth factor 1	Growth factor	40 ng/ml	Peprotech
FGF18	Fibroblast growth factor 18	Growth factor	40 ng/ml	Peprotech
FGF23	Fibroblast growth factor 23	Growth factor	40 ng/ml	Peprotech
FOLL	Follistatin	Modulator of metabolism	1600 ng/ml	Peprotech
gAcrp30	Adiponectin	Modulator of metabolism	4000 ng/ml	Peprotech
GLCT	Galectin 1	Growth factor	10000 ng/ml	Biovision
GDF11	Growth differentiation factor 11	Differentiation factor	400 ng/ml	Peprotech
GDF7	Growth differentiation factor 7	Differentiation factor	200 ng/ml	Peprotech
HERG	Heregulin	Growth factor	2 ng/ml	Peprotech
HGF	Hepatocyte growth factor	Growth factor	160 ng/ml	Peprotech
HU	Hydroxyurea	DNA damaging agent	800 uM	Sigma Aldrich
IFNa	Interferon alpha	Cytokine	12 ng/ml	eBioscience
IFNb	Interferon beta	Cytokine	18 ng/ml	pbl assay science
IFNbH	Interferon beta (low dose)	Cytokine	1,8 ng/ml	pbl assay science
IFNg	Interferon gamma	Cytokine	40 ng/ml	pbl assay science
IGF1	Insulin-like growth factor 1	Growth factor	8 ng/ml	Peprotech
IL10	Interleukin 10	Cytokine	8 ng/ml	Peprotech
IL13	Interleukin 13	Cytokine	4 ng/ml	Peprotech
IL16	Interleukin 16	Cytokine	400 ng/ml	Peprotech
IL17A	Interleukin 17A	Cytokine	8 ng/ml	Peprotech
IL17D	Interleukin 17D	Cytokine	200 ng/ml	Peprotech
IL22	Interleukin 22	Cytokine	4 ng/ml	Peprotech
IFNL2	Interleukin 28A	Cytokine	20 ng/ml	Peprotech
IL31	Interleukin 31	Cytokine	20 ng/ml	Peprotech
IL36b	Interleukin 36b	Cytokine	40 ng/ml	Peprotech
INS	Insulin	Modulator of metabolism	400 ng/ml	Gibco
IONH	Ionomycin (high dose)	Modulator of calcium	0,4 uM	Sigma Aldrich
IONM	Ionomycin	Modulator of calcium	0,1 uM	Sigma Aldrich
JAG1	Jagged 1	Modulator of differentiation	20 ng/ml	Life technologies
LACT	Lactic acid	Modulator of metabolism	1200 ng/ml	Sigma Aldrich
LPA	Lipoprotein	Modulator of metabolism	10 uM	Sigma Aldrich
LPS	Lipopolysaccharides	Modulator of metabolism	440 ng/ml	Sigma Aldrich
METF	Metformin	Modulator of metabolism	39 uM	Sigma Aldrich
MYTM	Mytomycin	DNA damaging agent	160 uM	Sigma Aldrich
NCS	Neocarzinostain	DNA damaging agent	1 ng/ml	Sigma Aldrich
NOGG	Noggin	Modulator of development	12 ng/ml	Peprotech
PARQ	Paraquat	Modulator of reactive oxygen species	40 uM	Sigma Aldrich
PDGFAA	Platelet-derived growth factor alpha	Growth factor	12 ng/ml	Peprotech
PGE2	Prostaglandin E2	Modulator of metabolism	2000 ng/ml	Enzo Life Sciences
PGF2a	Prostaglandin F2a	Modulator of metabolism	2000 ng/ml	Enzo Life Sciences
PEDF	Pigment epithelium derived factor	Modulator of differentiation	120 ng/ml	Peprotech
PIGF1	Human placenta growth factor	Growth factor	40 ng/ml	Peprotech
PMA	Phorbol myristate acetate	Modulator of calcium	60 uM	Sigma Aldrich
RESV	Resveratrol	Modulator of metabolism	10 uM	Sigma Aldrich
ROTN	Rotenone	Modulator of reactive oxygen species	2,5 uM	Sigma Aldrich
RSP1	R-spondin 1	Modulator of development	50 ng/ml	Peprotech
SHH	Sonic Hedgehog	Modulator of development	4000 ng/ml	Peprotech
TGFb	Transforming growth factor beta	Growth factor	0,2 ng/ml	Peprotech
TNFa	Tumor necrosis factor	Cytokine	0,2 ng/ml	Peprotech
TUNC	Tunicamycin	DNA damaging agent	0,02 uM	Sigma Aldrich
VEGFB	Vascular endothelial growth factor C	Growth factor	8000 ng/ml	Peprotech
VEGFC	Vascular endothelial growth factor B	Growth factor	2000 ng/ml	Peprotech
WNT1	Wingless-type MMTV integration site family member 1	Modulator of development	10 ng/ml	Peprotech
WNT3	Wingless-type MMTV integration site family member 3A	Modulator of development	200 ng/ml	RD systems
WNT7A	Wingless-type MMTV integration	Modulator of development	240 ng/ml	Peprotech

	site family member 7A			
WNTR	WNT 3A with R-spondin 1	Modulator of development	NA	NA

Appendix Table 6.4 Primers for qRT-PCR

Gene	Stimulus	Forward primer	Reverse primer
ID1	ACTA	ACACAAGATGCGATCGTCC	GGAATCCGAAGTTGGAACC
DACT1	ACTA	CTGAGGCCTGGTCTTCACAG	GGCGACCTTGAGTCTCTCAG
NBPF1	ACTA	TCGCGTAACTTCCCATTCACAG	GAGGCGTAGGTGAGGGTCGC
AMOTL2	ACTA	TCTCTCGCTCCAGCTTCTCT	CAAGCAAGACACAGGAGGC
SIPA1L2	ACTA	TCCCATAGAAGAATTTGCGG	TGGAAGTGCCCAGAGAAAAC
SMAD7	BMP2	CCAGGCTCCAGAAGAAGTTG	CCAACTGCAGACTGTCCAGA
SAMD11	BMP2	CTGCTTCCGACGCTGAC	ATGATCCCCCTCATCACCTC
NEDD9	BMP2	CCTTTTGGTATCTGGATGGG	ACCATCAGTGCAGAGAAGCA
FGFR3	BMP2	ACAGCTCAGCTCCACAGCAT	GAGTCCTTGGGGACGGAG
STC1	DFOM	CAGGCTTCGGACAAGTCTGT	ACAGCAAGCTGAATGTGTGC
DDIT4	DFOM	TGGCACACAAGTGTTCATCC	CCTGGACAGCAGCAACAGT
PFKFB3	DFOM	CCACCAAAAAGCCTCGCATC	AGTCAGCGCTGCTCCG
BNIP3	DFOM	AATAGAAACCGAGGCTGGAA	GATGCAGGAGGAGAGCCTG
EGR1	FGF1	GGAAAAGCGGCCAGTATAGG	AGCCCTACGAGCACCTGAC
EGR2	FGF1	AGCAAAGCTGCTGGGATATG	GTGACCATCTTTCCCAATGC
DUSP6	FGF1	AGACACCACAGTTCTTGCCC	GCCGAATCTCTTTGAGAACG
NAB2	FGF1	AGTAGGAAAGGAGGTTGGCG	GACCCTGCAGCCCAGAC
IFIT1	IFNb	GAGGAGCCTGGCTAAGCAAA	GCTCCAGACTATCCTTGACCTG
IFI35	IFNb	CCCACAGCCTCATCTTGAGT	TCTGAAGCCTCAGCTCTTGC
IFIT2	IFNb	CAAGGAGTTTTCTCCCTCCA	CGAACAGCTGAGAATTGCAC
OAS2	IFNb	TGTTTTCCGTCCATAGGAGC	CTGATCGACGAGATGGTGAA
IRF1	IFNg	ACCTCTGCCTTCTTCCCTCT	CATCCGAGTGATGGGCATGT
IDO1	IFNg	TCAACTCTTTCTCGAAGCTGG	GATGAAGAAGTGGGCTTTGC
CXCL10	IFNg	GCAGGTACAGCGTACAGTTC	CAGCAGAGGAACCTCCAGTC
THBS1	IFNg	TTGCCACAGCTCGTAGAACA	CAATGCCACAGTTCCTGATG
MANF	IONM	TCACAGATCTTCTCCACAGGG	CAAGAGGCAAAGAGAATCGG
PDIA4	IONM	TGGTCAACACAAGCGTGACT	AGACTACGAGGGCTCCAGAA
HYOU1	IONM	CTGCTCATTGAAAAGCCCA	CTGCAGATCCGGGGAGTAG
CALR	IONM	TTGTCAAAGATGGTGCCAGA	CACCCAGAAATTGACAACCC
KNSTRN	RESV	TGGAAACATCGACTTCTTTGG	GCCTCGTTACGATGACCAGT
KIF20A	RESV	CAAGGGCCTAACCCTCAAGT	GATTTGGGGTCTGTGGTACG
CDCA3	RESV	GCTCTGGGGGAAGATTTGAT	GATCCCCGCTCTCCTACTCT
PIF1	RESV	TCCGAGTCCTCATATTCCCC	TTGTTCATCTGAGTGTCGGC
FILIP1	ROTN	AACACCGGTCACAACGTCAT	CTGCCACTACTGGCTTTCCT
CALD1	ROTN	TCTGTGCAGAAAAGCAGTGG	GGCTTTGTAGGTTTTGCGCT
SP5	WNT3A	CCGATGCGGCTACAGGT	CTGCAGGCCTTTCTCCAG
AHR	WNT3A	AGTTATCCTGGCCTCCGTTT	TCAGTTCTTAGGCTCAGCGTC
TNFRSF19	WNT3A	GCGATCTTCACGAGGTTGAC	CGGCTTTCAAGACATGGAGT
KANK4	WNT3A	GGGTGTGCGCCTAGCTG	TGACCCTCAGTGTCTCAGGC
GREM2	WNT3A	CAGGGAAAGCTTCCAGAACA	CCCGCTGTGGACTAGAGAAC

Chapter 8. Bibliography

Adams MD, Sekelsky JJ. 2002. From sequence to phenotype: reverse genetics in Drosophila melanogaster. *Nat Rev Genet* 3:189-98

Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, et al. 2016. A Multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* 167:1867-82

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22:1600-7

Amsterdam A, Burgess S, Golling G, Chen W, Sun Z, et al. 1999. A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev* 13:2713-24

Andersson BS, Collins VP, Kurzrock R, Larkin DW, Childs C, et al. 1995. KBM-7, a human myeloid leukemia cell line with double Philadelphia chromosomes lacking normal c-ABL and BCR transcripts. *Leukemia* 9:2100-8

Ardiani A, Johnson AJ, Ruan H, Sanchez-Bonilla M, Serve K, Black ME. 2012. Enzymes to die for: exploiting nucleotide metabolizing enzymes for cancer gene therapy. *Curr Gene Ther* 12:77-91

Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, et al. 2004. The knockout mouse project. *Nat Genet* 36:921-4

Auwerx J, Avner P, Baldock R, Ballabio A, Balling R, et al. 2004. The European dimension for the mouse genome mutagenesis program. *Nat Genet* 36:925-7

Bar-Joseph Z, Gitter A, Simon I. 2012. Studying and modelling dynamic biological processes using time-series gene expression data. *Nat Rev Genet* 13:552-64

Barrangou R, Birmingham A, Wiemann S, Beijersbergen RL, Hornung V, Smith A. 2015. Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference. *Nucleic Acids Res* 43:3407-19

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709-12

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, et al. 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483:603-7

Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, et al. 2013. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 154:1151-61

Bauer NC, Corbett AH, Doetsch PW. 2015. The current state of eukaryotic DNA base damage and repair. *Nucleic Acids Res* 43:10083-101

Bedell VM, Wang Y, Campbell JM, Poshusta TL, Starker CG, et al. 2012. In vivo genome editing using a high-efficiency TALEN system. *Nature* 491:114-8

Bellen HJ, Levis RW, He Y, Carlson JW, Evans-Holm M, et al. 2011. The Drosophila gene disruption project: progress using transposons with distinctive site specificities. *Genetics* 188:731-43

Ben-Aroya S, Coombes C, Kwok T, O'Donnell KA, Boeke JD, Hieter P. 2008. Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. *Mol Cell* 30:248-58

Bhaya D, Davison M, Barrangou R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45:273-97

Bibikova M, Golic M, Golic KG, Carroll D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* 161:1169-75

Birsoy K, Wang T, Possemato R, Yilmaz OH, Koch CE, et al. 2013. MCT1-mediated transport of a toxic molecule is an effective strategy for targeting glycolytic tumors. *Nat Genet* 45:104-8

Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, et al. 2015. Gene essentiality and synthetic lethality in haploid human cells. *Science* 350:1092-6

Bobbin ML, Rossi JJ. 2016. RNA interference (RNAi)-based therapeutics: Delivering on the promise? *Annu Rev Pharmacol Toxicol* 56:103-22

Boettcher M, McManus MT. 2015. Choosing the right tool for the job: RNAi, TALEN, or CRISPR. *Mol Cell* 58:575-85

Bogdanove AJ, Voytas DF. 2011. TAL effectors: customizable proteins for DNA targeting. *Science* 333:1843-6

Boutros M, Ahringer J. 2008. The art and design of genetic screens: RNA interference. *Nat Rev Genet* 9:554-66

Bragelmann J, Klumper N, Offermann A, von Massenhausen A, Bohm D, et al. 2017. Pan-cancer analysis of the mediator complex transcriptome identifies CDK19 and CDK8 as therapeutic targets in advanced prostate cancer. *Clin Cancer Res* 23:1829-40

Breimer LH. 1988. Ionizing radiation-induced mutagenesis. *Br J Cancer* 57:6-18

Breinig M, Klein FA, Huber W, Boutros M. 2015. A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells. *Mol Syst Biol* 11:846

Brenner S. 1974. The genetics of Caenorhabditis elegans. *Genetics* 77:71-94

Brown SD, Moore MW. 2012. Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Dis Model Mech* 5:289-92

Buchovecky CM, Turley SD, Brown HM, Kyle SM, McDonald JG, et al. 2013. A suppressor screen in Mecp2 mutant mice implicates cholesterol metabolism in Rett syndrome. *Nat Genet* 45:1013-20

Buchtova M, Oralova V, Aklian A, Masek J, Vesela I, et al. 2015. Fibroblast growth factor and canonical WNT/beta-catenin signaling cooperate in suppression of chondrocyte differentiation in experimental models of FGFR signaling in cartilage. *Biochim Biophys Acta* 1852:839-50

Bunnage ME, Gilbert AM, Jones LH, Hett EC. 2015. Know your target, know your molecule. *Nat Chem Biol* 11:368-72

Bürckstümmer T, Banning C, Hainzl P, Schobesberger R, Kerzendorfer C, et al. 2013. A reversible gene trap collection empowers haploid genetics in human cells. *Nat Methods* 10:965-71

Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, et al. 2013. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45:1113-20

Capecchi MR. 2005. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* 6:507-12

Carbery ID, Ji D, Harrington A, Brown V, Weinstein EJ, et al. 2010. Targeted genome modification in mice using zinc-finger nucleases. *Genetics* 186:451-9

Carette JE, Guimaraes CP, Varadarajan M, Park AS, Wuethrich I, et al. 2009. Haploid genetic screens in human cells identify host factors used by pathogens. *Science* 326:1231-5

Carette JE, Guimaraes CP, Wuethrich I, Blomen VA, Varadarajan M, et al. 2011a. Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat Biotechnol* 29:542-6

Carette JE, Pruszak J, Varadarajan M, Blomen VA, Gokhale S, et al. 2010. Generation of iPSCs from cultured human malignant cells. *Blood* 115:4039-42

Carette JE, Raaben M, Wong AC, Herbert AS, Obernosterer G, et al. 2011b. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* 477:340-3

Carpinelli MR, Hilton DJ, Metcalf D, Antonchuk JL, Hyland CD, et al. 2004. Suppressor screen in Mpl-/- mice: c-Myb mutation causes supraphysiological production of platelets in the absence of thrombopoietin signaling. *Proc Natl Acad Sci U S A* 101:6553-8

Chakrabarti S, Streisinger G, Singer F, Walker C. 1983. Frequency of gamma-Ray induced specific locus and recessive lethal mutations in mature germ cells of the zebrafish, Brachydanio rerio. *Genetics* 103:109-23

Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, et al. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155:1479-91

Chen Y, Clegg NJ, Scher HI. 2009. Anti-androgens and androgen-depleting therapies in prostate cancer: new agents for an established target. *Lancet Oncol* 10:981-91

Cho SW, Kim S, Kim JM, Kim JS. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31:230-2

Cho SW, Kim S, Kim Y, Kweon J, Kim HS, et al. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* 24:132-41

Chua G, Morris QD, Sopko R, Robinson MD, Ryan O, et al. 2006. Identifying transcription factor functions and targets by phenotypic activation. *Proc Natl Acad Sci U S A* 103:12045-50

Clark JD, Flanagan ME, Telliez JB. 2014. Discovery and development of Janus kinase (JAK) inhibitors for inflammatory diseases. *J Med Chem* 57:5023-38

Clevers H, Nusse R. 2012. Wnt/beta-catenin signaling and disease. *Cell* 149:1192-205

Cong L, Ran FA, Cox D, Lin S, Barretto R, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819-23

Cooley L, Kelley R, Spradling A. 1988. Insertional mutagenesis of the Drosophila genome with single P elements. *Science* 239:1121-8

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, et al. 2010. The genetic landscape of a cell. *Science* 327:425-31

Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, et al. 2016. A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353

Cui X, Ji D, Fisher DA, Wu Y, Briner DM, Weinstein EJ. 2011. Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat Biotechnol* 29:64-7

Dailey L, Ambrosetti D, Mansukhani A, Basilico C. 2005. Mechanisms underlying differential responses to FGF signaling. *Cytokine Growth Factor Rev* 16:233-47

Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, et al. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* 14:297-301

de Nadal E, Ammerer G, Posas F. 2011. Controlling gene expression in response to stress. *Nat Rev Genet* 12:833-45

DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278:680-6

Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, et al. 2005. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169:1915-25

DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. 2013. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res* 41:4336-43

Di Ventra M, Taniguchi M. 2016. Decoding DNA, RNA and peptides with quantum tunnelling. *Nat Nanotechnol* 11:117-26

Dixit A, Parnas O, Li B, Chen J, Fulco CP, et al. 2016. Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167:1853-66

Dobbelstein M, Sorensen CS. 2015. Exploiting replicative stress to treat cancer. *Nat Rev Drug Discov* 14:405-23

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 34:184-91

Doitsidou M, Reichman-Fried M, Stebler J, Koprunner M, Dorries J, et al. 2002. Guidance of primordial germ cell migration by the chemokine SDF-1. *Cell* 111:647-59

Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, et al. 2008. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* 26:702-8

Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, et al. 1996. A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* 123:37-46

Drouin V, Viguie F, Debesse B. 1993. Near-haploid karyotype in a squamous cell lung carcinoma. *Genes Chromosomes Cancer* 7:209-12

Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, et al. 2001. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* 344:1031-7

Duverger Y, Belougne J, Scaglione S, Brandli D, Beclin C, Ewbank JJ. 2007. A semi-automated high-throughput approach to the generation of transposon insertion mutants in the nematode Caenorhabditis elegans. *Nucleic Acids Res* 35:e11

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411:494-8

Elling U, Penninger JM. 2014. Genome wide functional genetics in haploid cells. *FEBS Lett* 588:2415-21

Elling U, Taubenschmid J, Wirnsberger G, O'Malley R, Demers SP, et al. 2011. Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* 9:563-74

Essletzbichler P, Konopka T, Santoro F, Chen D, Gapp BV, et al. 2014. Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res* 24:2059-65

Evans MJ, Kaufman MH. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292:154-6

Fauvel B, Yasri A. 2014. Antibodies directed against receptor tyrosine kinases: current and future strategies to fight cancer. *MAbs* 6:838-51

Fece de la Cruz F, Gapp BV, Nijman SM. 2015. Synthetic lethal vulnerabilities of cancer. *Annu Rev Pharmacol Toxicol* 55:513-31

Fimereli D, Detours V, Konopka T. 2013. TriageTools: tools for partitioning and prioritizing analysis of high-throughput sequencing data. *Nucleic Acids Res* 41:e86

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 391:806-11

Fisher E, Scambler P. 1994. Human haploinsufficiency--one for sorrow, two for joy. *Nat Genet* 7:5-7

Fleuren ED, Zhang L, Wu J, Daly RJ. 2016. The kinome 'at large' in cancer. *Nat Rev Cancer* 16:83-98

Forment JV, Herzog M, Coates J, Konopka T, Gapp BV, et al. 2017. Genome-wide genetic screening with chemically mutagenized haploid embryonic stem cells. *Nat Chem Biol* 13:12-4

Forsburg SL. 2001. The art and design of genetic screens: yeast. *Nat Rev Genet* 2:659-68

Fountas A, Diamantopoulos LN, Tsatsoulis A. 2015. Tyrosine kinase inhibitors and diabetes: A novel treatment paradigm? *Trends Endocrinol Metab* 26:643-56

Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA. 2013. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nat Methods* 10:741-3

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, et al. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 31:822-6

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32:279-84

Fuchs F, Pau G, Kranz D, Sklyar O, Budjan C, et al. 2010. Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Mol Syst Biol* 6:370

Futosi K, Mocsai A. 2016. Tyrosine kinase signaling pathways in neutrophils. *Immunol Rev* 273:121-39

Gaiano N, Amsterdam A, Kawakami K, Allende M, Becker T, Hopkins N. 1996. Insertional mutagenesis and rapid cloning of essential genes in zebrafish. *Nature* 383:829-32

Gapp BV, Konopka T, Penz T, Dalal V, Burckstummer T, et al. 2016. Parallel reverse genetic screening in mutant human cells using transcriptomics. *Mol Syst Biol* 12:879

Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, et al. 2012. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483:570-5

Gawad C, Koh W, Quake SR. 2016. Single-cell genome sequencing: current state of the science. *Nat Rev Genet* 17:175-88

1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, et al. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061-73

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, et al. 2015. A global reference for human genetic variation. *Nature* 526:68-74

Gharwan H, Groninger H. 2016. Kinase inhibitors and monoclonal antibodies in oncology: clinical implications. *Nat Rev Clin Oncol* 13:209-27

Giaever G, Chu AM, Ni L, Connelly C, Riles L, et al. 2002. Functional profiling of the Saccharomyces cerevisiae genome. *Nature* 418:387-91

Giaever G, Flaherty P, Kumm J, Proctor M, Nislow C, et al. 2004. Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc Natl Acad Sci U S A* 101:793-8

Giaever G, Nislow C. 2014. The yeast deletion collection: a decade of functional genomics. *Genetics* 197:451-65

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, et al. 2014. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159:647-61

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, et al. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154:442-51

Gilchrist EJ, O'Neil NJ, Rose AM, Zetka MC, Haughn GW. 2006. TILLING is an effective reverse genetics technique for Caenorhabditis elegans. *BMC Genomics* 7:262

Golling G, Amsterdam A, Sun Z, Antonelli M, Maldonado E, et al. 2002. Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat Genet* 31:135-40

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17:333-51

Gotwals P, Cameron S, Cipolletta D, Cremasco V, Crystal A, et al. 2017. Prospects for combining targeted and conventional cancer therapy with immunotherapy. *Nat Rev Cancer* 17:286-301

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, et al. 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 194:1029-35

Grimm S. 2004. The art and design of genetic screens: mammalian culture cells. *Nat Rev Genet* 5:179-89

Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, et al. 2015. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* 47:435-44

Guo G, Wang W, Bradley A. 2004. Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. *Nature* 429:891-5

Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, et al. 1996. The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development* 123:1-36

Haimovich AD, Muir P, Isaacs FJ. 2015. Genomes by design. *Nat Rev Genet* 16:501-16

Hamilton AJ, Baulcombe DC. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286:950-2

Hammond SM, Bernstein E, Beach D, Hannon GJ. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* 404:293-6

Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, et al. 2015. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163:1515-26

Hartenian E, Doench JG. 2015. Genetic screens and functional genomics using CRISPR/Cas9 technology. *FEBS J* 282:1383-93

Hartwell LH, Culotti J, Reid B. 1970. Genetic control of the cell-division cycle in yeast. I. Detection of mutants. *Proc Natl Acad Sci U S A* 66:352-9

Heckl D, Kowalczyk MS, Yudovich D, Belizaire R, Puram RV, et al. 2014. Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat Biotechnol* 32:941-6

Hillenmeyer ME, Ericson E, Davis RW, Nislow C, Koller D, Giaever G. 2010. Systematic analysis of genome-wide fitness data in yeast reveals novel gene function and drug action. *Genome Biol* 11:R30

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, et al. 2008. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320:362-5

Ho CH, Piotrowski J, Dixon SJ, Baryshnikova A, Costanzo M, Boone C. 2011. Combining functional genomics and chemical biology to identify targets of bioactive compounds. *Curr Opin Chem Biol* 15:66-78

Hoebe K, Du X, Georgel P, Janssen E, Tabeta K, et al. 2003. Identification of Lps2 as a key transducer of MyD88-independent TIR signalling. *Nature* 424:743-8

Hogg SJ, Vervoort SJ, Deswal S, Ott CJ, Li J, et al. 2017. BET-Bromodomain inhibitors engage the host immune system and regulate expression of the immune checkpoint ligand PD-L1. *Cell Rep* 18:2162-74

Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, et al. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95:717-28

Hopkins AL. 2008. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 4:682-90

Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327:167-70

Housden BE, Muhar M, Gemberling M, Gersbach CA, Stainier DY, et al. 2017. Loss-of-function genetic tools for animal models: cross-species and cross-platform differences. *Nat Rev Genet* 18:24-40

Hrabe de Angelis M, Nicholson G, Selloum M, White JK, Morgan H, et al. 2015. Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. *Nat Genet* 47:969-78

Hrabe de Angelis MH, Flaswinkel H, Fuchs H, Rathkolb B, Soewarto D, et al. 2000. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* 25:444-7

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31:827-32

Hu Z, Killion PJ, Iyer VR. 2007. Genetic reconstruction of a functional transcriptional regulatory network. *Nat Genet* 39:683-7

Huang M, Shen A, Ding J, Geng M. 2014. Molecularly targeted cancer therapy: some lessons from the past decade. *Trends Pharmacol Sci* 35:41-50

Hudis CA. 2007. Trastuzumab-mechanism of action and use in clinical practice. *N Engl J Med* 357:39-51

Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, et al. 2000. Functional discovery via a compendium of expression profiles. *Cell* 102:109-26

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, et al. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31:227-9

International Human Genome Sequencing Consortium, Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921

International Mouse Knockout Consortium, Collins FS, Rossant J, Wurst W. 2007. A mouse for all reasons. *Cell* 128:9-13

Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, et al. 2016. A landscape of pharmacogenomic interactions in cancer. *Cell* 166:740-54

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* 169:5429-33

Ivics Z, Li MA, Mates L, Boeke JD, Nagy A, et al. 2009. Transposon-mediated genome manipulation in vertebrates. *Nat Methods* 6:415-22

Jacobson LS, Lima H, Jr., Goldberg MF, Gocheva V, Tsiperson V, et al. 2013. Cathepsin-mediated necrosis controls the adaptive immune response by Th2 (T helper type 2)-associated adjuvants. *J Biol Chem* 288:7481-91

Jae LT, Raaben M, Herbert AS, Kuehne AI, Wirchnianski AS, et al. 2014. Lassa virus entry requires a trigger-induced receptor switch. *Science* 344:1506-10

Jae LT, Raaben M, Riemersma M, van Beusekom E, Blomen VA, et al. 2013. Deciphering the glycosylome of dystroglycanopathies using haploid screens for lassa virus entry. *Science* 340:479-83

Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, et al. 2016. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* 167:1883-96 e15

Jansen G, Hazendonk E, Thijssen KL, Plasterk RH. 1997. Reverse genetics by chemical mutagenesis in Caenorhabditis elegans. *Nat Genet* 17:119-21

Jiang H, Pritchard JR, Williams RT, Lauffenburger DA, Hemann MT. 2011. A mammalian functional-genetic approach to characterizing cancer therapeutics. *Nat Chem Biol* 7:92-100

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816-21

Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. *Elife* 2:e00471

Jones LH, Bunnage ME. 2017. Applications of chemogenomic library screening in drug discovery. *Nat Rev Drug Discov* 16:285-96

Jorgensen EM, Mango SE. 2002. The art and design of genetic screens: Caenorhabditis elegans. *Nat Rev Genet* 3:356-69

Kaelin, WG Jr. 2012. Use and abuse of RNAi to study mammalian gene function. *Science* 337:421-2

Kaletta T, Hengartner MO. 2006. Finding function in novel targets: C. elegans as a model organism. *Nat Rev Drug Discov* 5:387-98

Kaminski R, Chen Y, Fischer T, Tedaldi E, Napoli A, et al. 2016. Elimination of HIV-1 genomes from human T-lymphoid cells by CRISPR/Cas9 gene editing. *Sci Rep* 6:22555

Kasper DM, Moro A, Ristori E, Narayanan A, Hill-Teran G, et al. 2017. MicroRNAs establish uniform traits during the architecture of vertebrate embryos. *Dev Cell* 40:552-65

Kemmeren P, Sameith K, van de Pasch LA, Benschop JJ, Lenstra TL, et al. 2014. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* 157:740-52

Kersten K, de Visser KE, van Miltenburg MH, Jonkers J. 2016. Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol Med* 9:137-53

Kessous A, Colombies P, Pris J, Clement D. 1980. Near haploid cell line in lymphoid blast crisis of Ph1-positive chronic myeloid leukemia. *Cancer Res* 40:1354-9

Kim H, Kim JS. 2014. A guide to genome engineering with programmable nucleases. *Nat Rev Genet* 15:321-34

Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Mdel C, Yusa K. 2014. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* 32:267-73

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, et al. 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517:583-8

Kool J, Berns A. 2009. High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. *Nat Rev Cancer* 9:389-99

Kotecki M, Reddy PS, Cochran BH. 1999. Isolation and characterization of a near-haploid human cell line. *Exp Cell Res* 252:273-80

Krejci P, Aklian A, Kaucka M, Sevcikova E, Prochazkova J, et al. 2012. Receptor tyrosine kinases activate canonical WNT/beta-catenin signaling via MAP kinase/LRP6 pathway and direct beta-catenin phosphorylation. *PLoS One* 7:e35826

Kummar S, Chen HX, Wright J, Holbeck S, Millin MD, et al. 2010. Utilizing targeted cancer therapeutic agents in combination: novel approaches and urgent requirements. *Nat Rev Drug Discov* 9:843-56

Kuromori T, Hirayama T, Kiyosue Y, Takabe H, Mizukado S, et al. 2004. A collection of 11 800 single-copy Ds transposon insertion lines in Arabidopsis. *Plant J* 37:897-905

Lackner DH, Carre A, Guzzardo PM, Banning C, Mangena R, et al. 2015. A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nat Commun* 6:10237

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, et al. 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929-35

Laufer C, Fischer B, Billmann M, Huber W, Boutros M. 2013. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat Methods* 10:427-31

Lee AY, St Onge RP, Proctor MJ, Wallace IM, Nile AH, et al. 2014. Mapping the cellular response to small molecules using chemogenomic fitness signatures. *Science* 344:208-11

Leeb M, Wutz A. 2013. Haploid genomes illustrate epigenetic constraints and gene dosage effects in mammals. *Epigenetics Chromatin* 6:41

Lehner B. 2013. Genotype to phenotype: lessons from model organisms for human genetics. *Nat Rev Genet* 14:168-78

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536:285-91

Lemmon MA, Schlessinger J. 2010. Cell signaling by receptor tyrosine kinases. *Cell* 141:1117-34

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, et al. 2011. The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell* 42:536-49

Lewandoski M. 2001. Conditional control of gene expression in the mouse. *Nat Rev Genet* 2:743-55

Lewis DL, Hagstrom JE, Loomis AG, Wolff JA, Herweijer H. 2002. Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. *Nat Genet* 32:107-8
Lewis EB, Bacher F. 1968. Methods of feeding ethyl methane sulphonate (EMS) to Drosophila males. *Drosoph Inf Serv* 43:193

Li Z, Vizeacoumar FJ, Bahr S, Li J, Warringer J, et al. 2011. Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat Biotechnol* 29:361-7

Licciardello MP, Ringler A, Markt P, Klepsch F, Lardeau CH, et al. 2017. A combinatorial screen of the CLOUD uncovers a synergy targeting the androgen receptor. *Nature chemical biology* 13:771-8

Liu KI, Ramli MN, Woo CW, Wang Y, Zhao T, et al. 2016. A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. *Nat Chem Biol* 12:980-7

Liu LX, Spoerke JM, Mulligan EL, Chen J, Reardon B, et al. 1999. High-throughput isolation of Caenorhabditis elegans deletion mutants. *Genome Res* 9:859-67

Lohr JG, Adalsteinsson VA, Cibulskis K, Choudhury AD, Rosenberg M, et al. 2014. Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol* 32:479-84

Long C, Amoasii L, Mireault AA, McAnally JR, Li H, et al. 2016. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* 351:400-3

Lordick F, Janjigian YY. 2016. Clinical impact of tumour biology in the management of gastroesophageal cancer. *Nat Rev Clin Oncol* 13:348-60

Lovly CM, Shaw AT. 2014. Molecular pathways: resistance to kinase inhibitors and implications for therapeutic strategies. *Clin Cancer Res* 20:2249-56

Lowe WL, Jr., Reddy TE. 2015. Genomic approaches for understanding the genetics of complex disease. *Genome Res* 25:1432-41

Lum PY, Armour CD, Stepaniants SB, Cavet G, Wolf MK, et al. 2004. Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell* 116:121-37

Maddalo D, Manchado E, Concepcion CP, Bonetti C, Vidigal JA, et al. 2014. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* 516:423-7

Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* 10:977-9

Maji B, Moore CL, Zetsche B, Volz SE, Zhang F, et al. 2017. Multidimensional chemical control of CRISPR-Cas9. *Nat Chem Biol* 13:9-11

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, et al. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9:467-77

Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, et al. 2013b. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 31:833-8

Mali P, Yang L, Esvelt KM, Aach J, Guell M, et al. 2013a. RNA-guided human genome engineering via Cas9. *Science* 339:823-6

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. 2002. The protein kinase complement of the human genome. *Science* 298:1912-34

Martin GR. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* 78:7634-8

Martins MM, Zhou AY, Corella A, Horiuchi D, Yau C, et al. 2015. Linking tumor mutations to drug responses via a quantitative chemical-genetic interaction map. *Cancer Discov* 5:154-67

Masel J, Siegal ML. 2009. Robustness: mechanisms and consequences. *Trends Genet* 25:395-403

Mayne KM, Maher EJ. 1989. Near-haploid cell line in megakaryoblastic transformation of Philadelphia-positive chronic myeloid leukemia. *Cancer Genet Cytogenet* 39:133-6

McCaffrey AP, Meuse L, Pham TT, Conklin DS, Hannon GJ, Kay MA. 2002. RNA interference in adult mice. *Nature* 418:38-9

Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* 26:695-701

Mohr S, Bakal C, Perrimon N. 2010. Genomic screening with RNAi: results and challenges. *Annu Rev Biochem* 79:37-64

Mohr SE, Smith JA, Shamu CE, Neumuller RA, Perrimon N. 2014. RNAi screening comes of age: improved techniques and complementary approaches. *Nat Rev Mol Cell Biol* 15:591-600

Morton J, Davis MW, Jorgensen EM, Carroll D. 2006. Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. *Proc Natl Acad Sci USA* 103:16370-5

Muellner MK, Uras IZ, Gapp BV, Kerzendorfer C, Smida M, et al. 2011. A chemical-genetic screen reveals a mechanism of resistance to PI3K inhibitors in cancer. *Nat Chem Biol* 7:787-93

Muller HJ. 1927. Artificial Transmutation of the Gene. *Science* 66:84-7

Muller HJ. 1930. Types of visible variations induced by x-rays in Drosophila. *J Genet* 22: 299-334

Narasimhan VM, Hunt KA, Mason D, Baker CL, Karczewski KJ, et al. 2016. Health and population effects of rare gene knockouts in adult humans with related parents. *Science* 352:474-7

Nasevicius A, Ekker SC. 2000. Effective targeted gene 'knockdown' in zebrafish. *Nature genetics* 26:216-20

Navin NE. 2015. The first five years of single-cell cancer genomics and beyond. *Genome Res* 25:1499-507

Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, et al. 2016. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* 351:403-7

Neugebauer JM, Amack JD, Peterson AG, Bisgrove BW, Yost HJ. 2009. FGF signalling during embryo development regulates cilia length in diverse epithelia. *Nature* 458:651-4

Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, et al. 2011. Phenotypic landscape of a bacterial cell. *Cell* 144:143-56

Nihongaki Y, Kawano F, Nakajima T, Sato M. 2015a. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* 33:755-60

Nihongaki Y, Yamamoto S, Kawano F, Suzuki H, Sato M. 2015b. CRISPR-Cas9-based photoactivatable transcription system. *Chem Biol* 22:169-74

Nijman SM. 2015. Functional genomics to uncover drug mechanism of action. *Nat Chem Biol* 11:942-8

Niu Y, Shen B, Cui Y, Chen Y, Wang J, et al. 2014. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* 156:836-43

Nolan PM, Peters J, Strivens M, Rogers D, Hagan J, et al. 2000. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet* 25:440-3

Norbury C, Nurse P. 1992. Animal cell cycles and their control. *Annu Rev Biochem* 61:441-70

North TE, Goessling W, Peeters M, Li P, Ceol C, et al. 2009. Hematopoietic stem cell development is dependent on blood flow. *Cell* 137:736-48

North TE, Goessling W, Walkley CR, Lengerke C, Kopani KR, et al. 2007. Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature* 447:1007-11

Nurse P. 1975. Genetic control of cell size at cell division in yeast. *Nature* 256:547-51

Nüsslein-Volhard C, Wieschaus E. 1980. Mutations affecting segment number and polarity in Drosophila. *Nature* 287:795-801

Nygaard HB, van Dyck CH, Strittmatter SM. 2014. Fyn kinase inhibition as a novel therapy for Alzheimer's disease. *Alzheimers Res Ther* 6:8

O'Shea JJ, Schwartz DM, Villarino AV, Gadina M, McInnes IB, Laurence A. 2015. The JAK-STAT pathway: impact on human disease and therapeutic intervention. *Annu Rev Med* 66:311-28

Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12:87-98

Pandey UB, Nichols CD. 2011. Human disease models in Drosophila melanogaster and the role of the fly in therapeutic drug discovery. *Pharmacol Rev* 63:411-36

Parsons AB, Lopez A, Givoni IE, Williams DE, Gray CA, et al. 2006. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* 126:611-25

Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, et al. 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* 10:973-6

Perlman ZE, Slack MD, Feng Y, Mitchison TJ, Wu LF, Altschuler SJ. 2004. Multidimensional drug profiling by automated microscopy. *Science* 306:1194-8

Pillay S, Meyer NL, Puschnik AS, Davulcu O, Diep J, et al. 2016. An essential receptor for adeno-associated virus infection. *Nature* 530:108-12

Pires-daSilva A, Sommer RJ. 2003. The evolution of signalling pathways in animal development. *Nat Rev Genet* 4:39-49

Pommier Y. 2006. Topoisomerase I inhibitors: camptothecins and beyond. *Nat Rev Cancer* 6:789-802

Prakash S, Sung P, Prakash L. 1993. DNA repair genes and proteins of Saccharomyces cerevisiae. *Annu Rev Genet* 27:33-70

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, et al. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152:1173-83

Raju R, Palapetta SM, Sandhya VK, Sahu A, Alipoor A, et al. 2014. A network map of FGF-1/FGFR signaling system. *J Signal Transduct* 2014:962962

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, et al. 2013a. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154:1380-9

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013b. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8:2281-308

Rane SG, Reddy EP. 2000. Janus kinases: components of multiple signaling pathways. *Oncogene* 19:5662-79

Rask-Andersen M, Masuram S, Schioth HB. 2014. The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu Rev Pharmacol Toxicol* 54:9-26

Reimao-Pinto MM, Manzenreither RA, Burkard TR, Sledz P, Jinek M, et al. 2016. Molecular basis for cytoplasmic RNA surveillance by uridylation-triggered decay in Drosophila. *EMBO J* 35:2417-34

Rihel J, Prober DA, Arvanites A, Lam K, Zimmerman S, et al. 2010. Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. *Science* 327:348-51

Rogge RD, Karlovich CA, Banerjee U. 1991. Genetic dissection of a neurodevelopmental pathway: Son of sevenless functions downstream of the sevenless and EGF receptor tyrosine kinases. *Cell* 64:39-48

Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, et al. 1999. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402:413-8

Rothstein RJ. 1983. One-step gene disruption in yeast. *Methods in enzymology* 101:202-11

Safavi S, Forestier E, Golovleva I, Barbany G, Nord KH, et al. 2013. Loss of chromosomes is the primary event in near-haploid and low-hypodiploid acute lymphoblastic leukemia. *Leukemia* 27:248-50

Sagi I, Chia G, Golan-Lev T, Peretz M, Weissbein U, et al. 2016. Derivation and differentiation of haploid human embryonic stem cells. *Nature* 532:107-11

Saleheen D, Natarajan P, Armean IM, Zhao W, Rasheed A, et al. 2017. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* 544:235-9

Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 32:347-55

Sarin S, Prabhu S, O'Meara MM, Pe'er I, Hobert O. 2008. Caenorhabditis elegans mutant allele identification by whole-genome sequencing. *Nat Methods* 5:865-7

Scappini B, Gatto S, Onida F, Ricci C, Divoky V, et al. 2004. Changes associated with the development of resistance to imatinib (STI571) in two leukemia cell lines expressing p210 Bcr/Abl protein. *Cancer* 100:1459-71
Schenone M, Dancik V, Wagner BK, Clemons PA. 2013. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* 9:232-40

Schmierer B, Hill CS. 2007. TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility. *Nat Rev Mol Cell Biol* 8:970-82

Schneider WM, Chevillotte MD, Rice CM. 2014. Interferon-stimulated genes: a complex web of host defenses. *Annu Rev Immunol* 32:513-45

Schwank G, Koo BK, Sasselli V, Dekkers JF, Heo I, et al. 2013. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* 13:653-8

Schwartz DM, Bonelli M, Gadina M, O'Shea JJ. 2016. Type I/II cytokines, JAKs, and new strategies for treating autoimmune diseases. *Nat Rev Rheumatol* 12:25-36

Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, et al. 2015. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov* 5:1210-23

Sega GA. 1984. A review of the genetic effects of ethyl methanesulfonate. *Mutat Res* 134:113-42

Semenza GL. 2014. Oxygen sensing, hypoxia-inducible factors, and disease pathophysiology. *Annu Rev Pathol* 9:47-71

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343:84-7

Shalem O, Sanjana NE, Zhang F. 2015. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* 16:299-311

Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. 2015. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* 33:661-7

Shibuya T, Morimoto K. 1993. A review of the genotoxicity of 1-ethyl-1-nitrosourea. *Mutat Res* 297:3-38

Shuai K, Liu B. 2003. Regulation of JAK-STAT signalling in the immune system. *Nat Rev Immunol* 3:900-11

Simon MA, Bowtell DD, Dodson GS, Laverty TR, Rubin GM. 1991. Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the sevenless protein tyrosine kinase. *Cell* 67:701-16

Siva N. 2015. UK gears up to decode 100,000 genomes from NHS patients. *Lancet* 385:103-4

Skarnes WC, von Melchner H, Wurst W, Hicks G, Nord AS, et al. 2004. A public gene trap resource for mouse functional genomics. *Nat Genet* 36:543-4

Smida M, Fece de la Cruz F, Kerzendorfer C, Uras IZ, Mair B, et al. 2016. MEK inhibitors block growth of lung tumours with mutations in ataxia-telangiectasia mutated. *Nat Commun* 7:13701

Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, et al. 1999. The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. *Genetics* 153:135-77

St Johnston D. 2002. The art and design of genetic screens: Drosophila melanogaster. *Nat Rev Genet* 3:176-88

Stanford WL, Cohn JB, Cordes SP. 2001. Gene-trap mutagenesis: past, present and beyond. *Nat Rev Genet* 2:756-68

Staring J, von Castelmur E, Blomen VA, van den Hengel LG, Brockmann M, et al. 2017. PLA2G16 represents a switch between entry and clearance of Picornaviridae. *Nature* 541:412-6

Stark AK, Sriskantharajah S, Hessel EM, Okkenhaug K. 2015. PI3K inhibitors in inflammation, autoimmunity and cancer. *Curr Opin Pharmacol* 23:82-91

Stemple DL. 2004. TILLING--a high-throughput harvest for functional genomics. *Nat Rev Genet* 5:145-50

Stewart HI, Rosenbluth RE, Baillie DL. 1991. Most ultraviolet irradiation induced mutations in the nematode Caenorhabditis elegans are chromosomal rearrangements. *Mutat Res* 249:37-54

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, et al. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature* 526:75-81

Sulston JE, Horvitz HR. 1977. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev Biol* 56:110-56

Sulston JE, Schierenberg E, White JG, Thomson JN. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol* 100:64-119

Tabebordbar M, Zhu K, Cheng JK, Chew WL, Widrick JJ, et al. 2016. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* 351:407-11

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, et al. 2003. Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res* 13:524-30

Timms RT, Menzies SA, Tchasovnikarova IA, Christensen LC, Williamson JC, et al. 2016. Genetic dissection of mammalian ERAD through comparative haploid and CRISPR forward genetic screens. *Nat Commun* 7:11786

Turner N, Grose R. 2010. Fibroblast growth factor signalling: from development to cancer. *Nat Rev Cancer* 10:116-29

Ufer M. 2005. Comparative pharmacokinetics of vitamin K antagonists: warfarin, phenprocoumon and acenocoumarol. *Clin Pharmacokinet* 44:1227-46

UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, et al. 2015. The UK10K project identifies rare variants in health and disease. *Nature* 526:82-90

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. 2010. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11:636-46

Vallin E, Gallagher J, Granger L, Martin E, Belougne J, et al. 2012. A genome-wide collection of Mos1 transposon insertion mutants for the C. elegans research community. *PLoS One* 7:e30482

van der Maaten LJP, Hinton GE. 2008. Visualizing high-dimensional data using t-SNE. *JMLR* 9:2579-2605

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends Genet* 30:418-26

van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, et al. 2010. Functional overlap and regulatory links shape genetic interactions between signaling pathways. *Cell* 143:991-1004
Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. *Science* 291:1304-51

Walker C, Streisinger G. 1983. Induction of mutations by gamma-rays in pregonial germ cells of zebrafish embryos. *Genetics* 103:125-36

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, et al. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153:910-8

Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, et al. 2015. Identification and characterization of essential genes in the human genome. *Science* 350:1096-101

Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343:80-4

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57-63

Weiner GJ. 2015. Building better monoclonal antibody-based therapeutics. *Nat Rev Cancer* 15:361-70

White JK, Gerdin AK, Karp NA, Ryder E, Buljan M, et al. 2013. Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* 154:452-64

Wienholds E, Schulte-Merker S, Walderich B, Plasterk RH. 2002. Target-selected inactivation of the zebrafish rag1 gene. *Science* 297:99-102

Wienholds E, van Eeden F, Kosters M, Mudde J, Plasterk RH, Cuppen E. 2003. Efficient target-selected mutagenesis in zebrafish. *Genome Res* 13:2700-7

Wilkening S, Pelechano V, Jarvelin AI, Tekkedil MM, Anders S, et al. 2013. An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res* 41:e65

Wilson RC, Doudna JA. 2013. Molecular mechanisms of RNA interference. *Annu Rev Biophys* 42:217-39

Winter GE, Radic B, Mayor-Ruiz C, Blomen VA, Trefzer C, et al. 2014. The solute carrier SLC35F2 enables YM155-mediated DNA damage toxicity. *Nat Chem Biol* 10:768-73

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, et al. 1999. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* 285:901-6

Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, et al. 2011. Targeted genome editing across species using ZFNs and TALENs. *Science* 333:307

Wright AV, Nunez JK, Doudna JA. 2016. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* 164:29-44

Wu P, Nielsen TE, Clausen MH. 2015. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol Sci* 36:422-39

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26:873-81

Wu X, Li Y, Crise B, Burgess SM. 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300:1749-51

Wu Y, Liang D, Wang Y, Bai M, Tang W, et al. 2013. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* 13:659-62

Xia H, Mao Q, Paulson HL, Davidson BL. 2002. siRNA-mediated gene silencing in vitro and in vivo. *Nat Biotechnol* 20:1006-10

Xiao MS, Zhang B, Li YS, Gao Q, Sun W, Chen W. 2016. Global analysis of regulatory divergence in the evolution of mouse alternative polyadenylation. *Mol Syst Biol* 12:890

Xue W, Chen S, Yin H, Tammela T, Papagiannakopoulos T, et al. 2014. CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* 514:380-4

Yaish P, Gazit A, Gilon C, Levitzki A. 1988. Blocking of EGF-dependent cell proliferation by EGF receptor kinase inhibitors. *Science* 242:933-5

Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154:1370-9

Yang Y, Wang L, Bell P, McMenamin D, He Z, et al. 2016. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat Biotechnol* 34:334-8

Yeh JR, Munson KM, Elagib KE, Goldfarb AN, Sweetser DA, Peterson RT. 2009. Discovering chemical modifiers of oncogene-regulated hematopoietic differentiation. *Nature chemical biology* 5:236-43

Yi S, Lin S, Li Y, Zhao W, Mills GB, Sahni N. 2017. Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nat Rev Genet*

Yin H, Xue W, Chen S, Bogorad RL, Benedetti E, et al. 2014. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol* 32:551-3

Young DW, Bender A, Hoyt J, McWhinnie E, Chirn GW, et al. 2008. Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat Chem Biol* 4:59-68

Yu PB, Hong CC, Sachidanandan C, Babitt JL, Deng DY, et al. 2008. Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism. *Nature chemical biology* 4:33-41

Zahreddine H, Borden KL. 2013. Mechanisms and insights into drug resistance in cancer. *Front Pharmacol* 4:28

Zambrowicz BP, Friedrich GA, Buxton EC, Lilleberg SL, Person C, Sands AT. 1998. Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* 392:608-11

Zhang L, Jia R, Palange NJ, Satheka AC, Togo J, et al. 2015. Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS One* 10:e0120396

Zhou Y, Zhu S, Cai C, Yuan P, Li C, et al. 2014. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509:487-91

Zhu H, Bengsch F, Svoronos N, Rutkowski MR, Bitler BG, et al. 2016. BET Bromodomain Inhibition Promotes Anti-tumor Immunity by Suppressing PD-L1 Expression. *Cell Rep* 16:2829-37

Zu Y, Tong X, Wang Z, Liu D, Pan R, et al. 2013. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods* 10:329-31

Zuckermann M, Hovestadt V, Knobbe-Thomsen CB, Zapatka M, Northcott PA, et al. 2015. Somatic CRISPR/Cas9-mediated tumour suppressor disruption enables versatile brain tumour modelling. *Nat Commun* 6:7391