Title
Using Functional Genomics to Understand Physiological Processes and Injury Response in Human Erythroid Cells

Author
Sobh, Amin

Publication Date
2018

Peer reviewed|Thesis/dissertation

Using Functional Genomics to Understand Physiological Processes and Injury Response in Human Erythroid Cells

By

Amin Sobh

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Comparative Biochemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Christopher Vulpe, Chair
Professor Andreas Stahl
Professor James Olzmann
Professor Jen Chywan Wang

Spring 2018

Abstract

Using Functional Genomics to Understand Physiological Processes and Injury Response in Human Erythroid Cells

by

Amin Sobh

Doctor of Philosophy in Comparative Biochemistry

University of California, Berkeley

Professor Christopher Vulpe, Chair

The vital roles of red blood cells (RBCs) underscore the importance of functional and homeostatic cellular processes in erythroid precursors. Iron acquisition and trafficking by erythroid cells is a key physiological process for their survival, proliferation and differentiation. In contrast, exposure to hematotoxicants can deplete erythroid precursors or disrupt erythroid differentiation. Complete elucidation of mechanisms relevant to a biological process or response to an injury in erythroid cells can improve our understanding of the pathophysiology and etiology of RBC disorders. Interrogating the genome in the context of a cellular process is a key strategy to understand underlying molecular mechanisms. Comprehensive identification of pathways governing physiological as well as injurious cellular processes has become more feasible due to the emergence of genome-wide functional genomic approaches. Recently, the CRISPR-Cas9 system has revolutionized the field of functional genomics as it enabled simple, efficient and cost-effective platforms for large-scale genetic screening. In this series of studies, CRISPR-based genome-wide loss-of function screening was employed to identify genes and pathways contributing to erythroid key processes including iron uptake, heme trafficking and erythroid differentiation, as well as response to hematotoxicants.

Chapter 1 reviews the various CRISPR-Cas9 screening platforms and their applications in functional genomics. Different screening strategies using large-scale loss-of-function and gain-of-function approaches are compared in terms of their effectiveness, feasibility, and suitability to study particular cellular processes. Steps of pooled screening, data analysis methods and hit validation strategies are described and critical technical considerations are addressed. In addition, the power of CRISPR-based genetic screening in different research fields including basic cell biology, cancer, drug discovery, pharmacogenomics and toxicogenomics is demonstrated. Finally, limitations of the current CRISPR screening platforms are revised and strategies for improvement are proposed.

In chapter 2, a genome-wide CRISPR-based screen designed to study iron uptake in erythroid cells is described. The physiological iron form, transferrin-bound iron (TBI), is taken up by erythroid cells through the transferrin receptor (TfR) by endocytosis. Non-transferrin bound iron (NTBI), which is present in the circulation during iron overload, can also be acquired by

erythroid cells. Uncontrolled cellular NTBI influx can result in iron toxicity. To identify molecular determinants of TBI and NTBI uptake, a genome-wide loss-of-function screen was performed in human K562 erythroleukemic cells which can utilize both TBI and NTBI to grow and proliferate. The screen revealed multiple genes whose disruption resulted in defective growth when either TBI or NTBI is the exclusive source of iron. Unsurprisingly, TBI uptake candidates included the transferrin receptor (TfR1) in addition to several components of the endocytic pathway. Follow up studies on one candidate, CCDC115, confirmed its role in cellular TBI uptake. CCDC115 is a V-ATPase assembly factor and its role in endosomal acidification likely underlies its function in transferrin iron influx.

In chapter 3, a genome-wide knockout screen investigating cellular mechanisms of heme trafficking and heme-induced erythroid differentiation is documented. Under physiological conditions, heme is synthesized in the mitochondria of erythroid cells where it plays structural as well as functional roles in erythroid differentiation. Additionally, erythroid cells are capable of taking up extracellular heme which might be utilized in pathological forms of erythropoiesis or could induce toxicity in the absence of other erythropietic signals. Heme treatment of K562 erythroleukemic cells induces erythroid differentiation resulting in a proliferation block. The screen was designed to identify genes whose inactivation alleviates the heme-induced proliferation block. The identified candidates include components of the clathrin-mediated endocytosis and vesicle acidification pathways in addition to epigenetic and RNA processing regulators. Intriguingly, the V-ATPase assembly factor CCDC115 was among the top candidates. Analysis of CCDC115 deficient cells unveiled an unprecedented role of CCDC115 in cellular heme uptake. Gene products whose loss results in heme-induced toxicity were also identified and included the heme oxygenase HMOX2, other detoxification enzymes and members of the ABC transporter family.

Chapter 4 reports identification of cellular mechanisms that modulate erythroid cell sensitivity to arsenic trioxide (ATO), a potent hematotoxicant and an effective anti-leukemic agent. A genome-wide loss-of-function CRISPR-based screen revealed novel molecular components influencing susceptibility of K562 erythroleukemic cells to ATO. Many of the candidates identified in the primary screen were simultaneously validated in a secondary screening approach. Functional enrichment analysis of validated genes revealed multiple pathways/processes implicated in the cellular response to ATO. The most significant pathway controls biosynthesis of selenocysteine, the 21st amino acid, and its incorporation into selenoproteins. Inactivation of components of this pathway resulted in notable cellular resistance to ATO. Based on the screening results, two models explaining the role of selenocysteine metabolism in ATO toxicity were proposed: the compromised thioredoxin reductase model and the arsenic-selenium-glutathione export model. Intriguingly, selenium pre-treatment of cells exhibited a protective effect against ATO cytotoxicity.

In chapter 5, susceptibility of erythroid cells to another hematotoxicant, acetaldehyde, was studied. Acetaldehyde, the primary product of alcohol metabolism and an endogenous metabolite, is a potential carcinogen and its well-established genotoxicity is thought to underlie bone marrow failure in Fanconi Anemia. A genome-wide knockout screen in erythroleukemic K562 cells identified several determinants of sensitivity to acetaldehyde. Consistent with the documented role of aldehydes in DNA damage, multiple identified candidate genes encode DNA repair enzymes. The top candidate gene encodes the tumor suppressor OVCA2, whose function

is unknown. The uncovered role of OVCA2 in detoxifying acetaldehyde was validated in a secondary screen and by individual disruption of the *OVCA2* gene. Interestingly, OVCA2 deficient cells displayed increased accumulation of the acetaldehyde-derived DNA adduct N2-ethylidene-2G.

Overall, these studies demonstrated the effectiveness of CRISPR/Cas9 loss-of-function genetic screening in deciphering mechanisms relevant to erythroid cell survival, proliferation, differentiation and response to injury.

Table of Contents

Dedication

This dissertation is dedicated to my wife, who truly believed in me and supported me in every aspect of my life, and to my mother, the woman whose love and devotion made me the person I am today.

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Appendix 2

List of Tables

Chapter 1

Appendix 3

List of Abbreviations

ABC - ATP-binding cassette
Ach - Acetaldehyde
APL - Acute promyelocytic leukemia
ATO - Arsenic trioxide
ATP - Adenosine triphosphate
BER - Base excision repair
Cas9 - CRISPR-associated protein-9
CERES - Computational correction of copy-number effect in CRISPR-Cas9 essentiality screens
CPM - Counts per million
CRISPR - Clustered regularly interspaced short palindromic repeats
CRISPR KO - CRISPR Knockout
CRISPRa - CRISPR activation
CRISPRi - CRISPR interference
Ctrl - Control
DAVID - Database for Annotation, Visualization and Integrated Discovery
DMEM – Dulbecco's Modified Eagle's Medium
DPC - DNA-protein crosslink
DSB - Double strand break
FA - Fanconi anemia
FACS - Fluorescence activated cell sorting
FBS - Fetal bovine serum
FC - Fold change
FDR - False discovery rate
Fe (III) Citrate - Ferric citrate
GeCKO V2 - Genome-scale CRISPR knockout version 2
GO - Gene ontology
GSH - Glutathione
GTP - Guanosine triphosphate
Hem - Hmein
HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Holo-Tf - Holo-transferrin
HR - Homologous recombination
IARC – International Agency for Research on Cancer
ICL - Interstrand crosslink
LC/MS - Liquid Chromatography/Mass Spectrometry
MAGeCK - Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout
MOI - Multiplicity of infection
N2-Et-2G - N2-ethyl-2'-deoxyguanosine
N2-Ethylidene-2G - N2-ethylidene-2'-deoxyguanosine
NADPH - Nicotinamide adenine dinucleotide phosphate
NER - Nucleotide Excision Repair
NHEJ - Non-homologous end joining
NTBI - Non-transferrin bound iron

NTC - Non-targeting control
PCR - Polymerase chain reaction
PS - Penicillin/Streptomycin
PVDF - Polyvinylidene difluoride
qPCR - Quantitative PCR
RIGER - RNAi Gene Enrichment Ranking
RNAi - RNA interference
RNP - Ribonucleoprotein
ROS - Reactive oxygen species
RPMI - Roswell Park Memorial Institute
SD - Standard deviation
SDS - Sodium dodecyl sulfate
Se - Selenium
Sec - Selenocysteine
SECIS - Selenocysteine insertion sequence
SecTRAP - Selenium-compromised thioredoxin reductase-derived apoptotic protein
Ser - Serine
sgRNA - single guide RNA
shRNA - short hairpin RNA
STRING - Search tool for the retrieval of interacting genes/proteins
TBI - Transferrin-bound iron
TC-NER - Transcription coupled nucleotide excision repair
Tf - Transferrin
TfR - Transferrin receptor
TLS - Translesion synthesis
Trx - Thioredoxin
TrxR - Thioredoxin reductase
V-ATPase - Vacuolar type H+ -ATPAse
Veh - Vehicle
WT - Wild-type
ZnMP - Zinc Mesoporphyrin

Acknowledgements

First, I would like to express my deep gratitude to my advisor Chris Vulpe for his mentorship, guidance and support. I have learned a lot from him during the period I spent in his lab and I had the chance to implement my own ideas and develop many research skills. I am grateful for all the opportunities given to me to participate in many scientific conferences and workshops.

I am also grateful to my other dissertation committee members Andreas Stahl, James Olzmann and Jen Chywan (Wally) Wang for their time, guidance and enlightenment.

Many thanks to Luoping Zhang and Martyn Smith from UC Berkeley school of public health for the collaboration and support.

I would like to extend my gratitude to the collaborators at other universities, Mitchell Knutson and Supak Jenkitkasemwong from the University of Florida, Silvia Balbo and Alessia Stornetta from the University of Minnesota, who helped me with some experiments.

I am thankful to my past and present colleagues at the Vulpe laboratory for their help and cooperation. Special thanks to Rola Zeidan Sobh (my colleague and my wife), Nader Ahmadieh, Gulce Naz Yazici, Jie Zhou and Mani Tagmount.

I would like to thank my previous advisor, Zouhair Attieh, for believing in me and pushing me to follow my dream.

Finally, I am grateful to my wonderful family, relatives and friends for their endless love and uplifting conversations. Rola: Thank you for always being my inspiration. Maha: Thank you for the love and prayers. Suheil: Thank you for having faith in me and always motivating me. Amal: Thank you for making everything easier. Lobna: Thank you for always being there for me. Reem, Randa, Bilal and Bilal: Thank you for all the encouragement. I really appreciate the moral support you all provided for me during this period. Your presence in my life is priceless.

Chapter 1

Functional Genomics in the CRISPR-Cas9 Era

Development of functional genomics

For any cellular process to occur, a signature network of gene-encoded proteins functions in a highly coordinated manner. Despite the huge progress in annotating genomes, the extent of our knowledge of biological functions of protein coding genes is still far from completion. Forward and reverse genetic studies largely contributed to the current established relationships between various biological functions and genomes. Functional genomics tools, which allow comprehensive investigation of a biological process through genome-wide perturbations, have been widely used in a variety of eukaryotic models. High-throughput genetic screens, performed in different organisms including budding yeast (*Saccharomyces cerevisiae*), fruit flies (*Drosophila melanogaster*) and nematodes (*Caenorhabditis elegans*), have led to various discoveries of genetic functions [reviewed in (Perrimon et al., 2010; Scherens and Goffeau, 2004; Yamamoto-Hino and Goto, 2013)]. Importantly, methods resulting in loss or gain of a specific gene function were also developed in mammalian cells and employed in high-throughput approaches to better understand physiological and pathological cellular processes. Gain-of-function genetic screening using expression cloning and arrayed cDNA libraries has unveiled key molecular components implicated in a variety of biological processes (Seed, 1995). Similarly, large-scale loss-of function interrogation based on RNA interference (RNAi) identified key factors involved in physiological and injurious cellular activities (Luo et al., 2009; Mercer et al., 2012; Neumann et al., 2010; Pelkmans et al., 2005) in addition to revealing molecular components with diagnostic and therapeutic relevance to cancer (Mullenders and Bernards, 2009). Recently, genome-editing tools have revolutionized the field of functional genomics. The emergence of the clustered regularly interspaced short palindromic repeat (CRISPR)-CRISPR associated protein 9 (Cas9) system provided a versatile tool for genetic perturbation that permits loss-of-function as well as gain-of function genome-wide screening in mammalian cells (Konermann et al., 2015; Shalem et al., 2014). Due to its simplicity, flexibility, high efficiency, wide accessibility and relatively low cost, the CRISPR-Cas9 tool rapidly became the first choice for functional genomics studies.

CRISPR-Cas9 system enables different forms of genetic perturbation

The CRISPR-Cas9 system is an efficient genome-editing tool derived from components of the prokaryotic adaptive antiviral immunity (Jinek et al., 2012; Ran et al., 2013). Cas9 is an RNA-guided nuclease that can be recruited to a target genomic region by a small RNA molecule (single guide RNA or sgRNA) through complementary base-pairing. Co-expression of Cas9 and a gene-specific sgRNA in a cell generates a double-stranded break (DSB) in the target locus that is fixed by the mutagenic non-homologous end joining (NHEJ) DNA repair system (Jinek et al., 2013). Cas9-induced mutagenesis promotes knockout of the target gene with remarkable efficiency (Shalem et al., 2014). Precise sequence editing in the target locus can also be generated through homologous recombination (HR) by simultaneously introducing an appropriate DNA donor with Cas9 and the targeting sgRNA (Mali et al., 2013). In addition to sequence modifications, the CRISPR-Cas9 system has been repurposed for regulation of target gene expression by generating a catalytically inactive Cas9 (dCas9) with retained ability to be

guided by a sgRNA (Qi et al., 2013). Accordingly, the CRISPR interference (CRISPRi) system, where dCas9 is fused with a transcriptional repression domain (KRAB), was introduced and demonstrated robust transcriptional repression of target genes (Gilbert et al., 2013). Alternatively, fusion of dCas9 to the VP64 transcriptional activation domain led to development of the RNA-guided transcriptional activation system or CRISPRa (Maeder et al., 2013; Perez-Pinera et al., 2013). Other CRISPR-based transcriptional activation systems including the synergistic activation mediator (SAM) and dCas9-SunTag-VP64 were developed and displayed a more robust induction of target gene expression (Konermann et al., 2015; Tanenbaum et al., 2014). Further, fusing dCas9 to histone-modifying enzymes such as the histone acetyltransferase p300 or the histone demethylase LSD1 enabled generating effective CRISPR-based epigenetic activators and repressors respectively (Hilton et al., 2015; Kearns et al., 2015). A major difference between the CRISPR knockout and the CRISPR regulatory systems is the position of sgRNA binding on genetic loci. While the Cas9 nuclease must be recruited to the coding region of a gene to edit a target exon, dCas9 fused to a regulatory effector needs to be recruited to a regulatory genetic region in proximity to the transcriptional start site.

CRISPR-based functional genomic screening platforms

The diversity of CRISPR-based genetic perturbation approaches offers multiple platforms for loss-of-function as well as gain-of-function genetic screening. Pooled CRISPR-based screening approaches are widely-used in functional genomic studies. Generating a pooled sgRNA library for high-throughput screening involves massively parallel synthesis of a pool of gene-specific oligonucleotides and cloning into sgRNA expression vectors (Shalem et al., 2014). A pooled screening approach requires lentiviral delivery of the sgRNA library into cells at a low multiplicity of infection (MOI). Low-MOI transduction results in a pool of infected cells, where a single viral vector is delivered and stably integrated into the genome of each cell (Shalem et al., 2015). The Cas9 nuclease or dCas9-effector can be co-expressed from the same sgRNA expression vector or from a separate vector (Sanjana et al., 2014). In a pooled screen, selection for a particular phenotype (cell proliferation/viability or reporter fluorescence) is carried out, and the readout is provided by quantifying each gene-specific guide sequence in the pool by deep sequencing (Agrotis and Ketteler, 2015). One limitation of pooled screening with large genome-wide sgRNA libraries is the requirement of massive numbers of cells to provide sufficient library representation. As an alternative, a sgRNA library targeting a defined gene set that is relevant to the research question can be generated and used in pooled screening. A less commonly used CRISPR-based functional screening platform is the arrayed screening. Arrayed CRISPR libraries targeting the whole genome are commercially available as multiple functional sets. Although genome-wide screens with arrayed CRISPR libraries can be of high cost and require the use of automation, screening in an arrayed micro-well plate format can provide a broad range of simply-assayable cellular phenotypes and does not require next generation sequencing for the readout (Agrotis and Ketteler, 2015; Shalem et al., 2015). Focused CRISPR-based sgRNA libraries have been effectively utilized in arrayed screening (Henser-Brownhill et al., 2017; Sredni Simone Treiger et al., 2017).

Loss-of function screens can be performed using the CRISPR-mediated gene knockout platform. Unlike RNAi-based methods, CRISPR knockout leads to complete disruption of the gene product thus resulting in more robust phenotypes during genetic screening. Different designs of pooled genome-wide knockout sgRNA libraries targeting human and mouse genes have been

generated and successfully used in screening (Doench et al., 2016; Hart et al., 2017; Ma et al., 2015; Park et al., 2017; Sanjana et al., 2014; Tzelepis et al., 2016; Wang et al., 2015). In addition to differences in sgRNA designs, these libraries vary in size mainly due to different numbers of sgRNAs targeting each gene. CRISPR-mediated knockdown of gene expression (CRISPRi) can also be implemented in loss-of-function screening. While gene knockdown by RNAi occurs post-transcriptionally, CRISPRi represses gene expression at the level of transcription (Unniyampurath et al., 2016). In genome-wide approaches, similar to shRNA (RNAi-based), CRISPRi results in generating a homogenous phenotype in the entire population of cells expressing the same sgRNA (Boettcher and McManus, 2015). Indeed, this property provides an advantage over CRISPR-mediated knockout screening where a phenotypically heterogeneous population of cells expressing the same sgRNA could be generated due to variable NHEJ-mediated mutagenesis. CRISPRi demonstrated a remarkable sensitivity and resulted in stronger loss-of-function phenotypes compared to shRNA when used in large-scale screening (Gilbert et al., 2014; Qi et al., 2013).

Large-scale gain-of-function interrogation is also feasible using CRISPR-based methods. The improvement of CRISPRa methods has led to the emergence of multiple effective transcriptional activation tools that were successfully employed in genome-wide gain-of-function screening (Gilbert et al., 2014; Horlbeck et al.; Joung et al., 2017a, 2017b; Konermann et al., 2015). Unlike the cDNA overexpression libraries which have been the only available tool for gain-of-function screening, CRISPRa libraries are easy to prepare and induce expression from endogenous genes (Shalem et al., 2015). CRISPR-based methods also enabled high throughput epigenome editing screening. Such approach demonstrated a promising potential for identifying epigenetic regulatory elements in genomic loci (Klann et al., 2017).

Design strategies and technical considerations for CRISPR-based pooled screening

In a pooled screening approach involving large-scale genetic perturbation, phenotype-based selection is necessary to identify perturbations that are relevant to a studied biological process. In general, three selectable cellular phenotypes can be exploited in pooled approaches: cell proliferation/survival, expression of a cell surface marker or altered fluorescence of an artificial cellular reporter (Figure 1.1). A genetic perturbation could induce a change in cell proliferation/survival of a specific cell type, or under a certain condition, resulting in a relative enrichment (positive selection) or depletion (negative selection) of the corresponding cell in the pool (Shalem et al., 2014; Wang et al., 2014). On the other hand, genetic perturbations that can lead to changes in the expression of a cell surface marker or the fluorescence intensity of an artificial cellular reporter can be selected by fluorescence associated cell sorting (FACS) (Doench et al., 2014; Wu et al., 2016). To identify a genetic perturbation underlying a phenotype, a stable and unique label must exist for each genetic perturbation. In CRISPR-based approaches, such a label is generated by the use of low MOI lentiviral delivery of sgRNA expression vectors, which leads to the integration of a single gene-specific sgRNA vector in the host genome of each cell in the pool. Consequently, the identity of a genetic perturbation can be revealed by identifying the short oligonucleotide sequence encoding the gene-specific portion (protospacer) of the sgRNA. Parallel sequencing of all protospacer sequences in a pool can be performed by next generation sequencing (Hartenian Ella and Doench John G., 2015). Several applications of pooled CRISPR-based genome-wide knockout screening *in vitro* can be implemented using different selection strategies. One simple application is using a negative

selection screen to identify genes that are essential for the growth of a particular cell type (Wang et al., 2015). In negative selection screening, the gene loss leads to depletion of the corresponding cells from the cellular pool. Additional negative selection applications involve screening in isogenic cell lines to reveal synthetic-lethal interactions with a key oncogene and hence identify novel cancer therapeutic targets (Martin et al., 2017). Alternatively, positive selection screening can be performed to identify genes whose loss confer resistance to a drug (Shalem et al., 2014). Such screens require the use of a lethal dose of the drug to ensure enrichment of resistant mutants in the cell population. Both positive and negative selections can be applied when a sub-lethal dose of a chemical stressor is used in the screen. Accordingly, genes whose loss confer increased or decreased cellular sensitivity to a studied chemical can be identified from the same screen (Xia et al., 2016). Pooled CRISPR screens can be also performed *in vivo* by injecting immunocompromised mice with a tumorigenic cellular library. Such screens allow identification of genes whose loss can induce cancer metastasis (Chen et al., 2015) or sensitize tumors to a therapeutic agent (Manguso et al., 2017).

There are numerous factors that should be considered during large-scale screening using the CRISPR-Cas9 system. One important factor is the choice of the cell type in which screening will be performed. Choosing the cell type is largely dependent on the research question and the tissue of interest. In screens where cell proliferation is the selective phenotype, the choice of a cell type is limited to cells with a proliferative capacity. In certain cases, a screen can be initially performed in a proliferative cell line model that is suitable for screening and the hits can be validated in the cell type of interest. Alternatively, manipulating the cell type of interest to obtain another selectable phenotype, such as modulation of a fluorescent reporter, might be necessary. Another factor to be considered in genetic screening is the power of selection. A strong selection can maximize the difference in abundance of cells, with a genetic perturbation relevant to the selectable phenotype, between experimental and control conditions. Parameters affecting selection include doubling time of the cells used in the screen, duration of the screen (number of cell doublings) and strength of the applied selective pressure (dose of a drug or chemical). Some chemicals/drugs can be very toxic and a continuous treatment with a selective dose for the period of the screen could result in undesirable reductions in cell numbers. A pulse treatment-recovery strategy is an alternative approach in these situations. Optimizing all these parameters prior to screening is recommended. One more factor to consider when dealing with large-scale CRISPR screening is maintaining an optimal representation of the used library. Maintaining sufficient representation of large genome-wide libraries during a screen could be challenging due to the requirement of massive cell numbers in each culture condition. It is recommended to maintain a cell number that is at least 500 fold the library size throughout all the stages of a screen (Joung et al., 2017b).

Computational analysis methods for pooled CRISPR screens

Following a pooled library screen, gene-specific protospacer sequences (sgRNAs) that are used to label a specific gene perturbation must be identified and quantified by next generation sequencing. The relative abundance of each sequence in the pool is compared across control and experimental samples to assess enrichment or dropout. For example, a comparison can be made across two different time points: an initial time point (T0) prior to selection and a final time point (T) which corresponds to the time at which the screen is stopped. Alternatively, a comparison between a treatment and a control condition at the same time point is performed. Preparation of

samples for sequencing involves PCR-amplification of protospacer sequences from genomic DNA collected from each sample (control or experimental) using primers that allow incorporation of appropriate sequencing adapters. Despite the complexity of a sgRNA library, conserved regions (from the library vector) in the sequencing reads decrease the library diversity and can create detection problems during next generation sequencing (Head et al., 2014). Using a cocktail of primers with staggers or shuffled sequences can increase the library diversity. The sequencing output for each sample is a file of raw sequences (FASTQ) that does not include any computational information. Multiple steps are required to analyze FASTQ files obtained from sgRNA library sequencing. These steps include sequence quality assessment, read trimming to obtain the 20-bp protospacer, alignment of the reads with the sgRNA library, counting and normalization of reads, analysis of read count distribution, variance estimation, differential sgRNA abundance analysis and gene ranking (Spahn et al., 2017). Although each of these steps can be applied separately, tools that can perform all these steps in a defined workflow have been developed. These include command-line based tools like MAGeCK (Li et al., 2014) and caRpools (Winter et al., 2016) or user-friendly platforms that can be run as a web-service such as CRISPRAnalyzeR (Winter et al., 2017), CRISPRcloud (Jeong et al., 2017) and PinAPL-PY (Spahn et al., 2017). Differential abundance of individual sgRNAs can be analyzed using algorithms that are commonly used for RNAseq such as DESeq2 (Love et al., 2014) and edgeR (Dai et al., 2014). Other algorithms including RIGER (Luo et al., 2008), MAGeCK (Li et al., 2014) and STARS (Doench et al., 2016) take into consideration differential abundance of multiple sgRNAs targeting each gene to perform gene ranking. CRISPRAnalyzeR can simultaneously run multiple analysis methods and reveal overlaps of hits among the different methods (Winter et al., 2017). Functional classification of gene sets obtained from CRISPR screens could be required to derive a biological interpretation. Online tools for gene set enrichment and pathway analysis include Gene Ontology (GO) enrichment analysis (Ashburner et al., 2000), PANTHER (Mi et al., 2017), DAVID (Huang et al., 2007; Sherman et al., 2007), and Reactome (Fabregat et al., 2018). Interaction networks between gene products can be visualized using the STRING database (Szklarczyk et al., 2017).

Validating hits from genome-wide CRISPR screens

Like all high-throughput experiments, large-scale CRISPR screens can generate false positive results and therefore validation of hits is necessary. False positives can be generated as a result of biases that might occur at any level during a screen. The stochastic noise of next generation sequencing is probably a major source of undesirable biases. Such noise is much more observed for reads with lower counts where the high variance can result in untrue large fold changes (Anders and Huber, 2010). Off-target effects of sgRNAs could also generate false positives as the phenotype of interest could be generated from unintended genetic perturbations (Tsai and Joung, 2016). Technical considerations that can minimize false positive hits in a screen include performing multiple replicates per screen, evaluating screen noise by analyzing non-targeting sgRNAs, sequencing at maximum possible depth, using multiple analysis platforms, removing low count reads from the analysis and using stringent cut-offs for candidate selection. In CRISPR knockout screens, false positive hits can arise from sgRNAs targeting genes within genomic amplification regions, where a Cas9-induced site-specific DNA damage can arrest cell proliferation resulting in depletion of the corresponding sgRNAs (Sheel and Xue, 2016). Accordingly, CRISPRi would be the method of choice for loss-of-function screening in cell lines with genomic amplifications.

One efficient strategy to simultaneously validate hits obtained from a primary genome-wide screen is secondary screening using a custom library (validation library) (Doench, 2018). Secondary screening involves generating a validation library enriched for sgRNAs targeting candidate genes identified in a primary screen. In addition to the target gene set, such libraries must include several sgRNAs targeting control genes as well as non-targeting sgRNAs. The relatively small number of genes in a validation library compared to a genome-wide library allows including much more sgRNAs per gene while maintaining a small library size (Doench, 2018). Increasing the number of sgRNAs targeting each gene would allow validating a hit with multiple sgRNAs. Moreover, secondary screening using a small library is practically more feasible, cost effective and can be performed in various cell types and using different selection phenotypes. Secondary screening not only eliminates false positives but can also minimize false negatives (Doench, 2018). When secondary screening is to be performed, candidate selection from the primary screen can be less stringent and a single sgRNA showing a phenotype is sufficient to consider the corresponding gene a candidate. In addition to validating a phenotype with multiple sgRNAs, further confidence in screening hits can be provided through functional classification. Identifying multiple hits in a screen that are involved in the same cellular pathway/process or are subunits of a cellular complex is a strong indication that these hits are not false positives.

While simultaneous enrichment or depletion of multiple sgRNAs targeting the same gene increases confidence in a hit, validation of the effect of a genetic perturbation on the phenotype of interest using targeted approaches is still required (Joung et al., 2017b). Genetic validation can be performed using targeted CRISPR-based perturbation of a candidate gene in the same cell line, and assaying for the screen phenotype. For example, if cell proliferation/survival is used for phenotype selection in the screen, validation can be performed using cell viability assays. When validating a phenotype for a candidate gene, it is generally recommended to target the gene with multiple sgRNAs (Doench, 2018). It is also critical to confirm a genetic perturbation by genotyping (CRISPR knockout), measuring transcript levels (CRISPRa/i) or measuring protein levels (CRISPR knockout, CRISPRa/i) (Joung et al., 2017b). If an identified candidate gene is druggable, the corresponding drug can be used to confirm the screening phenotype. Further validation of a candidate gene can be attained if the phenotype can be rescued by reversing the genetic perturbation. In doxycycline-inducible CRISPRa/i systems, repression or activation of a target gene can be easily reversed by doxycycline removal (Gilbert et al., 2014; Guo et al., 2017). Alternatively, a CRISPR knockout can be complemented by expressing an sgRNA-resistant cDNA encoding the missing gene (Park et al., 2017).

CRISPR screening applications in diverse research fields

The emergence of CRISPR-based functional genomic screening tools was accompanied by a burst in applications in numerous fields of biological research. Loss-of-function and gain-of function screens using the CRISPR-Cas systems were applied to answer questions about fundamental cellular and molecular mechanisms, study complex diseases including cancer, reveal mechanisms of drug resistance, identify novel drug targets and understand mechanisms of susceptibility and resistance to chemical substances.

Applications in basic cell biology

Several fundamental cellular processes have been functionally investigated by genetic screens using the CRISPR-Cas9 system. In addition to confirming previously recognized mechanisms

underlying these processes, CRISPR screens comprehensively revealed novel molecular determinants and regulatory cellular pathways contributing to each process. Genome-wide CRISPR knockout parallel screening in multiple human cell lines have extended the list of core fitness genes in addition to identifying cell-type specific fitness genes (Hart et al., 2015). A positive selection loss-of-function screen based on cell death identified novel key components of oxidative phosphorylation, the major pathway for cellular ATP production (Arroyo et al., 2016). In addition, a genome-wide CRISPR screen in mouse primary dendritic cells unveiled unknown positive and negative regulators of lipopolysaccharide-stimulated induction of tumor necrosis factor (TNF), and provided insight into regulatory networks relevant to innate immunity (Parnas et al., 2015). In a recent study, CRISPR screening uncovered novel regulators of Hedgehog signaling which is involved in embryonic development and stem cell homeostasis (Pusapati et al., 2018). Moreover, determinants of mitophagy were also investigated in a CRISPR screen where multiple steady-state regulators of the key mitophagy-promoting protein PARKIN were identified (Potting et al., 2018).

Applications in cancer research and drug target discovery

Several CRISPR screens were performed to identify genes implicated in cancer cell growth, progression and metastasis. Genome-wide screening in acute myeloid leukemia (AML) cell lines and a non-AML reference revealed a set of AML-specific essential genes (Tzelepis et al., 2016). In another study, loss-of-function screening in 14 different human AML cell lines revealed genes with differential essentiality across the cell lines and uncovered synthetic-lethal partners of Ras oncoprotein (Wang et al., 2017). A genome-wide CRISPR screen in prostate cancer cells revealed an RNA processing factor as a requirement for prostate cancer growth (Fei et al., 2017). In addition to *in vitro* approaches, CRISPR-based pooled screening *in vivo* was demonstrated as an effective tool for cancer research. A landmark study, where a tumorigenic CRISPR knockout cellular pool was transplanted into immunocompromised mice, identified genes involved in tumor growth and metastasis *in vivo* (Chen et al., 2015). Moreover, genes whose loss enhances or inhibits KRAS-dependent colorectal tumor growth *in vivo* were identified by a CRISPR knockout screen in isogenic xenografts harboring wild type or mutant KRAS (Yau et al., 2017). CRISPR-mediated transcriptional activation also provides an *in vivo* screening platform to interrogate large-scale gene activation for studying cancer progression and response to treatment. A CRISPR activation screen *in vivo*, using a sgRNA library targeting components of DNA damage response, showed that transcriptional activation of the checkpoint kinase (CHEK2) slows the progression of B-cell acute lymphoblastic leukemia (B-ALL) and increases sensitivity to chemotherapy (Braun et al., 2016). Identification of novel therapeutic targets is a hallmark of CRISPR screening applications in cancer research. In cancer, selective targets are often those whose modulation generates synthetic lethal interactions with an oncogenic mutation. Alternatively, a therapeutic target can be a component whose perturbation increases sensitivity to another anticancer agent. Accordingly, multiple studies employing CRISPR screening uncovered cellular targets with potential therapeutic implications in specific cancer types or in combination with other drugs. PREX1 was identified by CRISPR-based loss-of-function screening as a potential therapeutic target in cancers driven by oncogenic Ras (Wang et al., 2017). Additionally, CRISPR screening demonstrated that RNF43-mutant pancreatic tumors are selectively sensitive to WNT-FZD 5 signaling inhibition (Steinhart et al., 2017). Furthermore, several CRISPR screens performed *in vitro* and *in vivo* identified cellular components that can be potentially targeted in combination with other existing therapies to treat different types of cancer (Cobb et

al., 2016; Manguso et al., 2017; Ruiz et al., 2016; Sun et al., 2018). A CRISPR double-gene-knockout screening approach, designed to study pairwise synthetic lethal interactions in a set of druggable genes, was successfully employed to reveal novel cancer drug combinations (Han et al., 2017). Extending the combinatorial CRISPR screening approach to study interactions within multiple sets of druggable genes can lead to the discovery of unexpected effective therapeutic combinations. This strategy would allow repurposing existing drugs for cancer therapy as an alternative of laborious *de novo* drug discovery.

Applications in pharmacogenomics and toxicogenomics

In addition to identifying cellular therapeutic targets, CRISPR-based screening can help establish mechanisms of action as well as genetic determinants of enhanced or decreased response to existing drugs. Identifying a target of a given drug using CRISPR knockout screening is challenging because the complete ablation of a target gene occurs in both control and treatment conditions. However, CRISPRa/i-based screening can be an effective method to reveal direct drug targets, as a knockdown or activation of the target would synergize with or protect against the drug respectively (Anderson et al., 2015). Accordingly, a combined CRISPRa/i screening approach revealed microtubule destabilization as a mechanism of action of the myelodysplastic syndrome drug Rigosertib (Jost et al., 2017). Recently, a powerful CRISPR knockout-based screening strategy, named CRISPR-induced resistance in essential genes (CRISPRres), was developed to simultaneously identify targets of a given drug and mutations in these targets leading to drug resistance (Neggers et al., 2018). CRISPRres involves the use of tiling sgRNA libraries targeting a set of relevant essential genes. For CRISPRres screening, enrichment of mutations occurs at two levels: (1) since the target gene is essential, only mutations where the protein function is maintained are enriched; (2) using a lethal dose of the studied drug, only mutants that are resistant to the drug are enriched. Target genes can be identified by sequencing sgRNA protospacers while mutations giving rise to drug resistance are identified by sequencing the target locus. Using this approach nicotinamide phosphoribosyltransferase (NAMPT) was identified as the main target of the anticancer agent KPT-9274 (Neggers et al., 2018). Furthermore, CRISPR-based screening approaches successfully uncovered mechanisms affecting sensitivity to various anticancer agents including 6-thioguanine (Koike-Yusa et al., 2014; Wang et al., 2014), etoposide (Wang et al., 2014), vemurafenib (Shalem et al., 2014), sorafenib (Sun et al., 2018) and ATR inhibitors (Ruiz et al., 2016) as well as immunotherapy (Manguso et al., 2017).

Mechanisms of toxicity as well as determinants of susceptibility to toxic substances can be elucidated by CRISPR screens using the same strategies described for pharmacogenomics applications. Several CRISPR-based loss-of function and gain of function screens were performed to study biological toxins including a chimeric cholera/diphtheria fusion toxin (Gilbert et al., 2014), *Clostridium septicum* alpha toxin (Koike-Yusa et al., 2014), anthrax toxin (Zhou et al., 2014), and ricin (Horlbeck et al., 2016). These studies demonstrated the effectiveness of CRISPR-based genetic screening in identifying genes and pathways that modulate cellular susceptibility to toxic agents. Despite the adverse health effects resulting from exposure to ubiquitous toxic chemicals, mechanisms of toxicity as well as cellular processes affecting susceptibility to a large number of these chemicals are poorly characterized. So far, only few CRISPR-based screens studying chemical toxicants were reported. A loss-of-function screen in liver cells identified genes whose disruption alter sensitivity to triclosan, a ubiquitous

antimicrobial agent (Xia et al., 2016). Another CRISPR-based knockout screen using a metabolism-focused sgRNA library provided novel mechanistic insight into the toxicity of paraquat, a herbicide known to increase the risk of Parkinson's disease (Reczek et al., 2017). It was shown that inactivation of each of the top three positive selection candidate genes (*POR, ATP7A, SLC45A4*) abolished paraquat-induced oxidative stress. The screen further revealed a role of cellular copper homeostasis in paraquat cytotoxicity.

Limitations of the current CRISPR screening platforms and future directions

Despite the effectiveness and versatility of use in a wide range of research applications, several limitations exist for CRISPR-based screening in general and for specific CRISPR screening platforms in particular. Although off-target effects are a major concern in CRISPR biology, sgRNA off-target activity does not pose a real issue for CRISPR-based screening. Rationally, in a screen, the chance that multiple sgRNAs targeting the same gene would generate the phenotype of interest due to off-target perturbations is very low. Since multiple sgRNAs per gene are often required for candidate selection or validation, false-positive rates due to off-target effects are expected to be minimal. In genomic amplification regions, sgRNAs targeting amplified genes can result in CRISPR-mediated lethality that is independent on a target gene inactivation (Sheel and Xue, 2016). Accordingly, in CRISPR knockout negative selection screens performed to identify essential genes, all sgRNAs targeting an amplified gene would be depleted even if the gene is not essential. Although using CRISPRi instead of CRISPR knockout would mitigate such false-positive results, incomplete disruption of an essential gene may not result in robust cellular depletion. A computational method named CERES, that demonstrated notable correction of copy number effect, has recently emerged as a promising tool to reduce false positives in genetic vulnerability screens (Meyers et al., 2017). One true limitation of CRISPR-based screening is that sgRNAs are picked computationally and hence their on-target activities are only predicted. Efforts to maximize on-target activity and minimize off-target effects improved the quality of sgRNA libraries to a certain extent (Doench et al., 2016). Poor on-target activities of sgRNAs in a genome-wide library can result in false-negatives. Increasing the number of sgRNAs per gene could increase the chance of uncovering all genes involved in a particular phenotype. However, screening with larger-size sgRNA libraries poses an issue in terms of feasibility and cost. False negatives can also arise from compensatory effects of genes with functional redundancy. Since functional redundancy is usually among genes of the same family, using sgRNAs that target conserved coding regions could potentially reduce false negatives caused by functional redundancy (Mao et al., 2016).

Some limitations of CRISPR-based screening are specific to a particular screening platform or even to a screening strategy. CRISPR knockout screening is an effective approach for studying mechanisms modulating sensitivity to a given drug or toxicant. However, identifying a mechanism of action often requires revealing a direct target of the studied substance. In a knockout screen, complete loss of an essential target typically leads to similar depletion phenotypes between control pools and pools treated with the targeting substance. This poses a limitation for CRISPR knockout screening in identifying direct drug/chemical targets. In contrast, in CRISPRi screens, the targeting substance could synergize with the decreased levels of the target, leading to depletion of the corresponding cells only in the treated pool. For certain genes, a differential phenotype can only be detected when the gene function is completely lost, which imposes a disadvantage for CRISPRi screens. On the other hand, compensatory

mechanisms which can result in false negatives are likely activated upon complete gene knockout but not transcriptional repression (Rossi et al., 2015). Complementary loss-of-function screening using both CRISPR knockout and CRISPRi would be a good strategy for comprehensive mechanistic understanding of the response to a drug or toxicant (Rosenbluh et al., 2017). In most CRISPR knockout screening systems, early expression of Cas9 and the library sgRNAs leads to early gene disruption. Inactivation of certain essential genes can result in depletion of the corresponding cells during the antibiotic selection stage, prior to applying the selection stressor. Using an inducible Cas9 expression system for screening would allow timely expression of Cas9 and can highlight the role of essential genes in the response to a stressor. Unlike positive selection screening, negative selection maintains the library complexity at the end of a screen. Therefore, detecting a true depletion under the selective condition is challenging and requires higher library representation, more screening replicates and more sequencing depth. Positive selection of dead cells using Annexin V purification is a promising strategy to overcome technical challenges of negative selection screens where cell death is the selectable phenotype (Arroyo et al., 2016). For negative selection screens involving a chemical stressor, a depletion phenotype due to a particular genetic perturbation is not necessarily specific to the used stressor. Further validation with other stressors is required to rule out a general susceptibility phenotype.

Current CRISPR-based screening platforms offer promising tools to comprehensively interrogate genomes. Several strategies are to be considered to increase the power of CRISPR screens in the future. First, more potent sgRNA libraries can be constructed by implementing the most up-to-date algorithms and enriching for experimentally-validated sgRNAs. Second, screens using focused sgRNA libraries should be performed instead of genome-wide libraries when a research question is particular. Focused libraries allow increasing the number of sgRNAs per gene while maintaining a manageable library size. Third, inducible systems should be more utilized in CRISPR screening systems to allow proper time control of selection initiation. Finally, a standard, widely-accepted data analysis workflow should be assigned to analyze CRISPR screening datasets.

Figures

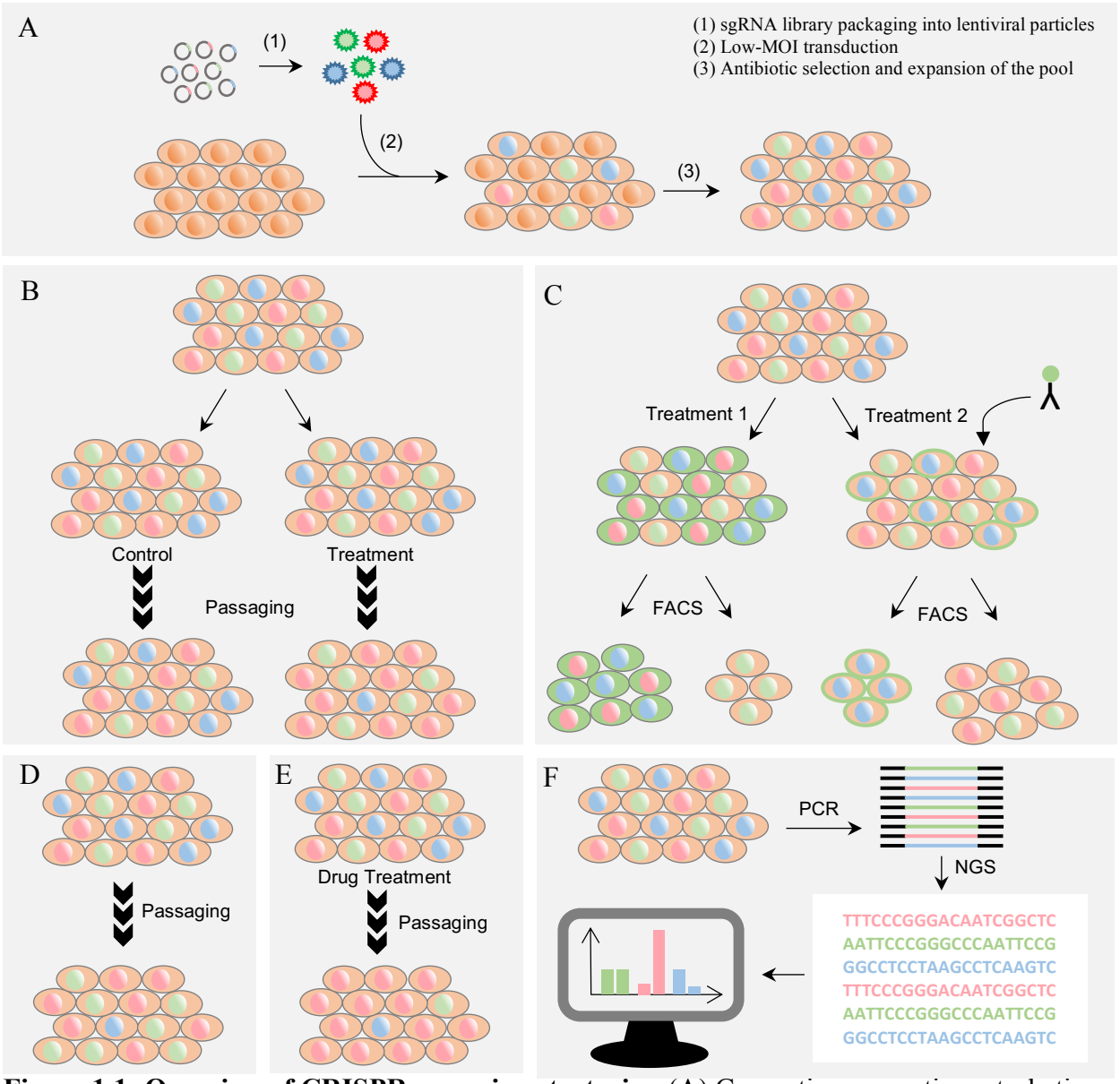

Figure 1.1: Overview of CRISPR screening strategies. (A) Generating a genetic perturbation cellular pool. (B) Positive and negative selection screening using differential proliferation/ survival under a stressor treatment for phenotype selection. Comparison can be made between control and treatment pools at the same time point. (C) FACS-based negative (left) and positive (right) selection of genetic perturbations modulating a fluorescent signal. Treatment 1 induces a fluorescent intracellular reporter while treatment 2 represses a cell surface marker that can be detected by a fluorescent antibody. (D) Negative selection screening to identify genes essential for cell growth. (E) Positive selection screening to identify genetic perturbations conferring resistance to a drug. In (D) and (E), comparisons can be made between two time points (i.e before and after the assigned passages of the cellular pools). (F) Quantifying genetic perturbations through amplification of gene-specific protospacers and counting them by next generation sequencing (NGS).

Tables

Table 1.1: Choice of CRISPR screening platform, selection strategy and validation method.

Study type	Screening Platform	Screening Strategy	Validation Method
Identification of Essential genes	CRISPR KO CRISPRi	Negative selection screening based on cell survival (Comparison between initial and final time points)	- Secondary screening - Individual gene knockout/ knockdown followed by cell viability assays
Cell differentiation	CRISPR KO CRISPRa/i	- Positive selection based on differentiation associated proliferation block (if applicable) - Positive or negative selection by FACS based on expression of a differentiation marker (intracellular fluorescent reporter or cell surface marker)	- Secondary screening - Individual gene disruption/ activation followed by cell viability assays or expression analysis of the differentiation marker
Gene expression regulation	CRISPR KO CRISPRa/i	Positive or negative selection by FACS in a reporter cell line	- Secondary screening - Individual gene disruption/ activation followed by gene expression analysis
Mechanisms of drug resistance	CRISPR KO CRISPRi	Positive selection screening based on cell survival (Comparison between initial and final time points or between treatment and control samples at the screen endpoint)	- Secondary screening - Individual gene knockout/ knockdown followed by cell viability assays
Chemical susceptibility /resistance	CRISPR KO CRISPRa/i	Positive and negative selection screening based on cell survival (Comparison between treatment and control samples at the screen endpoint)	- Secondary screening - Individual gene disruption/ activation followed by cell viability assays
Synthetic-lethal interactions with a drug or a gene mutation	CRISPR KO CRISPRi	Negative selection screening based on cell survival. Isogenic cell lines are required to study the effect of a gene mutation (Comparison between initial and final time points or between treatment and control samples at the screen endpoint)	- Secondary screening - Individual gene disruption/ activation followed by cell viability assays - If the candidate gene is druggable, the drug can be used to validate the synthetic-lethal phenotype

References

Agrotis, A., and Ketteler, R. (2015). A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. Front. Genet. *6*.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anderson, D.J., Le Moigne, R., Djakovic, S., Kumar, B., Rice, J., Wong, S., Wang, J., Yao, B., Valle, E., Kiss von Soly, S., et al. (2015). Targeting the AAA ATPase p97 as an Approach to Treat Cancer through Disruption of Protein Homeostasis. Cancer Cell *28*, 653–665.

Arroyo, J.D., Jourdain, A.A., Calvo, S.E., Ballarano, C.A., Doench, J.G., Root, D.E., and Mootha, V.K. (2016). A Genome-wide CRISPR Death Screen Identifies Genes Essential for Oxidative Phosphorylation. Cell Metab. *24*, 875–885.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Boettcher, M., and McManus, M.T. (2015). Choosing the Right Tool for the Job: RNAi, TALEN or CRISPR. Mol. Cell *58*, 575–585.

Braun, C.J., Bruno, P.M., Horlbeck, M.A., Gilbert, L.A., Weissman, J.S., and Hemann, M.T. (2016). Versatile in vivo regulation of tumor phenotypes by dCas9-mediated transcriptional perturbation. Proc. Natl. Acad. Sci. *113*, E3892–E3900.

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

Cobb, L.P., Soderquist, R., Berchuck, A., Gaillard, S., and Wood, K. (2016). CRISPR-CAS9-negative selection screen identifies genes that sensitize high-grade serous ovarian cancer cell lines to carboplatin, paclitaxel, and olaparib. Gynecol. Oncol. *141*, 44–45.

Dai, Z., Sheridan, J.M., Gearing, L.J., Moore, D.L., Su, S., Wormald, S., Wilcox, S., O'Connor, L., Dickins, R.A., Blewitt, M.E., et al. (2014). edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. F1000Research *3*, 95.

Doench, J.G. (2018). Am I ready for CRISPR? A user's guide to genetic screens. Nat. Rev. Genet. *19*, 67–80.

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

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res. *46*, D649–D655.

Fei, T., Chen, Y., Xiao, T., Li, W., Cato, L., Zhang, P., Cotter, M.B., Bowden, M., Lis, R.T., Zhao, S.G., et al. (2017). Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. Proc. Natl. Acad. Sci. *114*, E5207–E5215.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442–451.

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

Guo, J., Ma, D., Huang, R., Ming, J., Ye, M., Kee, K., Xie, Z., and Na, J. (2017). An inducible CRISPR-ON system for controllable gene activation in human pluripotent stem cells. Protein Cell *8*, 379–393.

Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A., and Bassik, M.C. (2017). Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat. Biotechnol. *35*, 463–474.

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

Hart, T., Tong, A., Chan, K., Leeuwen, J. van, Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and Design of Genome-wide CRISPR/Cas9 Knockout Screens. BioRxiv 117341.

Hartenian Ella, and Doench John G. (2015). Genetic screens and functional genomics using CRISPR/Cas9 technology. FEBS J. *282*, 1383–1393.

Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., and Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. BioTechniques *56*, 61-passim.

Henser-Brownhill, T., Monserrat, J., and Scaffidi, P. (2017). Generation of an arrayed CRISPR-Cas9 library targeting epigenetic regulators: from high-content screens to in vivo assays. Epigenetics *12*, 1065–1075.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR/Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Horlbeck, M.A., Gilbert, L.A., Villalta, J.E., Adamson, B., Pak, R.A., Chen, Y., Fields, A.P., Park, C.Y., Corn, J.E., Kampmann, M., et al. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. ELife *5*, e19760.

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

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. *8*, R183.

Jeong, H.-H., Kim, S.Y., Rousseaux, M.W.C., Zoghbi, H.Y., and Liu, Z. (2017). CRISPRcloud: a secure cloud-based pipeline for CRISPR pooled screen deconvolution. Bioinforma. Oxf. Engl. *33*, 2963–2965.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816–821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. ELife *2*.

Jost, M., Chen, Y., Gilbert, L.A., Horlbeck, M.A., Krenning, L., Menchon, G., Rai, A., Cho, M.Y., Stern, J.J., Prota, A.E., et al. (2017). Combined CRISPRi/a-Based Chemical Genetic Screens Reveal that Rigosertib Is a Microtubule-Destabilizing Agent. Mol. Cell *68*, 210-223.e6.

Joung, J., Engreitz, J.M., Konermann, S., Abudayyeh, O.O., Verdine, V.K., Aguet, F., Gootenberg, J.S., Sanjana, N.E., Wright, J.B., Fulco, C.P., et al. (2017a). Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. Nature *548*, 343–346.

Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017b). Genome-scale CRISPR-Cas9 Knockout and Transcriptional Activation Screening. Nat. Protoc. *12*, 828–863.

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9 -histone demethylase fusion. Nat. Methods *12*, 401–403.

Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2017). CRISPR–Cas9 epigenome editing enables high-throughput

screening for functional regulatory elements in the human genome. Nat. Biotechnol. *35*, 561–568.

Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M.D.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat. Biotechnol. *32*, 267–273.

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

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

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*.

Luo, B., Cheung, H.W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., Hinkle, G., Boehm, J.S., Beroukhim, R., Weir, B.A., et al. (2008). Highly parallel identification of essential genes in cancer cells. Proc. Natl. Acad. Sci. U. S. A. *105*, 20380–20385.

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

Ma, H., Dang, Y., Wu, Y., Jia, G., Anaya, E., Zhang, J., Abraham, S., Choi, J.-G., Shi, G., Qi, L., et al. (2015). A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death. Cell Rep. *12*, 673–683.

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

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-Guided Human Genome Engineering via Cas9. Science *339*, 823–826.

Manguso, R.T., Pope, H.W., Zimmer, M.D., Brown, F.D., Yates, K.B., Miller, B.C., Collins, N.B., Bi, K., LaFleur, M.W., Juneja, V.R., et al. (2017). *In vivo* CRISPR screening identifies *Ptpn2* as a cancer immunotherapy target. Nature *547*, 413–418.

Mao, L., Liu, C., Wang, Z., Niu, X., Xue, L., Zhou, Z., Cai, Z., Yu, M., Li, Y., Wu, D., et al. (2016). A genome-wide loss-of-function screening method for minimizing false-negatives caused by functional redundancy. Cell Res. *26*, 1067–1070.

Martin, T.D., Cook, D.R., Choi, M.Y., Li, M.Z., Haigis, K.M., and Elledge, S.J. (2017). A role for mitochondrial translation in promotion of viability in K-Ras mutant cells. Cell Rep. *20*, 427–438.

Mercer, J., Snijder, B., Sacher, R., Burkard, C., Bleck, C.K.E., Stahlberg, H., Pelkmans, L., and Helenius, A. (2012). RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. Cell Rep. *2*, 1036–1047.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nat. Genet. *49*, 1779–1784.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. *45*, D183–D189.

Mullenders, J., and Bernards, R. (2009). Loss-of-function genetic screens as a tool to improve the diagnosis and treatment of cancer. Oncogene *28*, 4409–4420.

Neggers, J.E., Kwanten, B., Dierckx, T., Noguchi, H., Voet, A., Bral, L., Minner, K., Massant, B., Kint, N., Delforge, M., et al. (2018). Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. Nat. Commun. *9*, 502.

Neumann, B., Walter, T., Hériché, J.-K., Bulkescher, J., Erfle, H., Conrad, C., Rogers, P., Poser, I., Held, M., Liebel, U., et al. (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature *464*, 721–727.

Park, R.J., Wang, T., Koundakjian, D., Hultquist, J.F., Lamothe-Molina, P., Monel, B., Schumann, K., Yu, H., Krupzcak, K.M., Garcia-Beltran, W., et al. (2017). A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. Nat. Genet. *49*, 193–203.

Parnas, O., Jovanovic, M., Eisenhaure, T.M., Herbst, R.H., Dixit, A., Ye, C.J., Przybylski, D., Platt, R.J., Tirosh, I., Sanjana, N.E., et al. (2015). A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. Cell *162*, 675–686.

Pelkmans, L., Fava, E., Grabner, H., Hannus, M., Habermann, B., Krausz, E., and Zerial, M. (2005). Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. Nature *436*, 78–86.

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

Perrimon, N., Ni, J.-Q., and Perkins, L. (2010). In vivo RNAi: Today and Tomorrow. Cold Spring Harb. Perspect. Biol. *2*.

Potting, C., Crochemore, C., Moretti, F., Nigsch, F., Schmidt, I., Manneville, C., Carbone, W., Knehr, J., DeJesus, R., Lindeman, A., et al. (2018). Genome-wide CRISPR screen for PARKIN

regulators reveals transcriptional repression as a determinant of mitophagy. Proc. Natl. Acad. Sci. *115*, E180–E189.

Pusapati, G.V., Kong, J.H., Patel, B.B., Krishnan, A., Sagner, A., Kinnebrew, M., Briscoe, J., Aravind, L., and Rohatgi, R. (2018). CRISPR Screens Uncover Genes that Regulate Target Cell Sensitivity to the Morphogen Sonic Hedgehog. Dev. Cell *44*, 113-129.e8.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

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

Reczek, C.R., Birsoy, K., Kong, H., Martínez-Reyes, I., Wang, T., Gao, P., Sabatini, D.M., and Chandel, N.S. (2017). A CRISPR screen identifies a pathway required for paraquat-induced cell death. Nat. Chem. Biol. *13*, 1274–1279.

Rosenbluh, J., Xu, H., Harrington, W., Gill, S., Wang, X., Vazquez, F., Root, D.E., Tsherniak, A., and Hahn, W.C. (2017). Complementary information derived from CRISPR Cas9 mediated gene deletion and suppression. Nat. Commun. *8*, 15403.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., and Stainier, D.Y.R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature *524*, 230–233.

Ruiz, S., Mayor-Ruiz, C., Lafarga, V., Murga, M., Vega-Sendino, M., Ortega, S., and Fernandez-Capetillo, O. (2016). A Genome-wide CRISPR Screen Identifies CDC25A as a Determinant of Sensitivity to ATR Inhibitors. Mol. Cell *62*, 307–313.

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

Scherens, B., and Goffeau, A. (2004). The uses of genome-wide yeast mutant collections. Genome Biol. *5*, 229.

Seed, B. (1995). Developments in expression cloning. Curr. Opin. Biotechnol. *6*, 567–573.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science *343*, 84–87.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nat. Rev. Genet. *16*, 299–311.

Sheel, A., and Xue, W. (2016). Genomic amplifications cause false positives in CRISPR screens. Cancer Discov. *6*, 824–826.

Sherman, B.T., Huang, D.W., Tan, Q., Guo, Y., Bour, S., Liu, D., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007). DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics *8*, 426.

Spahn, P.N., Bath, T., Weiss, R.J., Kim, J., Esko, J.D., Lewis, N.E., and Harismendy, O. (2017). PinAPL-Py: A comprehensive web-application for the analysis of CRISPR/Cas9 screens. Sci. Rep. *7*, 15854.

Sredni Simone Treiger, Suzuki Mario, Yang Jian-Ping, Topczewski Jacek, Bailey Anders W., Gokirmak Tufan, Gross Jeffrey N., Andrade Alexandre, Kondo Akihide, Piper David R., et al. (2017). A functional screening of the kinome identifies the Polo-like kinase 4 as a potential therapeutic target for malignant rhabdoid tumors, and possibly, other embryonal tumors of the brain. Pediatr. Blood Cancer *64*, e26551.

Steinhart, Z., Pavlovic, Z., Chandrashekhar, M., Hart, T., Wang, X., Zhang, X., Robitaille, M., Brown, K.R., Jaksani, S., Overmeer, R., et al. (2017). Genome-wide CRISPR screens reveal a Wnt–FZD5 signaling circuit as a druggable vulnerability of *RNF43*-mutant pancreatic tumors. Nat. Med. *23*, 60–68.

Sun, W., He, B., Yang, B., Hu, W., Cheng, S., Xiao, H., Yang, Z., Wen, X., Zhou, L., Xie, H., et al. (2018). Genome-wide CRISPR screen reveals SGOL1 as a druggable target of sorafenib-treated hepatocellular carcinoma. Lab. Invest. 1.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45*, D362–D368.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Tsai, S.Q., and Joung, J.K. (2016). Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. Nat. Rev. Genet. *17*, 300–312.

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O.M., Mupo, A., Grinkevich, V., Li, M., Mazan, M., et al. (2016). A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. *17*, 1193–1205.

Unniyampurath, U., Pilankatta, R., and Krishnan, M.N. (2016). RNA Interference in the Age of CRISPR: Will CRISPR Interfere with RNAi? Int. J. Mol. Sci. *17*.

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

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101.

Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell *168*, 890-903.e15.

Winter, J., Breinig, M., Heigwer, F., Brügemann, D., Leible, S., Pelz, O., Zhan, T., and Boutros, M. (2016). caRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens. Bioinforma. Oxf. Engl. *32*, 632–634.

Winter, J., Schwering, M., Pelz, O., Rauscher, B., Zhan, T., Heigwer, F., and Boutros, M. (2017). CRISPRAnalyzeR: Interactive analysis, annotation and documentation of pooled CRISPR screens. BioRxiv 109967.

Wu, Y., Zhou, L., Wang, X., Lu, J., Zhang, R., Liang, X., Wang, L., Deng, W., Zeng, Y.-X., Huang, H., et al. (2016). A genome-scale CRISPR-Cas9 screening method for protein stability reveals novel regulators of Cdc25A. Cell Discov. *2*, 16014.

Xia, P., Zhang, X., Xie, Y., Guan, M., Villeneuve, D.L., and Yu, H. (2016). Functional Toxicogenomic Assessment of Triclosan in Human HepG2 Cells Using Genome-Wide CRISPR-Cas9 Screening. Environ. Sci. Technol. *50*, 10682–10692.

Yamamoto-Hino, M., and Goto, S. (2013). In Vivo RNAi-Based Screens: Studies in Model Organisms. Genes *4*, 646–665.

Yau, E.H., Kummetha, I.R., Lichinchi, G., Tang, R., Zhang, Y., and Rana, T.M. (2017). Genome-Wide CRISPR Screen for Essential Cell Growth Mediators in Mutant KRAS Colorectal Cancers. Cancer Res. *77*, 6330–6339.

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature *509*, 487–491.

Chapter 2

Using CRISPR-Cas9 Genome-wide Loss-of-function screening to Study Mechanisms of Iron Uptake in Erythroid Cells

Introduction

Being essential for several vital biological processes, iron is a key element for almost all living organisms. At the cellular level, iron is involved in DNA synthesis and repair (Puig et al., 2017), transcriptional and post-transcriptional regulation (Pantopoulos, 2004; Shah and Xie, 2014), translation (Han et al., 2001), energy metabolism (Oexle et al., 1999) and xenobiotic metabolism (Gilardi and Nardo, 2017). Systemically, iron is essential for hemoglobin formation and oxygen transport in the blood (Abbaspour et al., 2014). In addition to the numerous physiological functions, excess iron can cause cellular damage and organ dysfunction through the formation of reactive oxygen species (ROS) (Kohgo et al., 2008). Hence, mechanisms of iron homeostasis must be tightly regulated to maintain normal cellular iron levels.

Cellular iron uptake is a key step in iron homeostasis. Cells take up different forms of non-heme iron that can exist in the circulation. The major physiological form of iron in the circulation is transferrin-bound iron (TBI) (Coates, 2014). TBI can be taken up by the well-characterized receptor-mediated endocytosis pathway that involves binding of holo-transferrin (holo-Tf) to the transferrin receptor TfR1 (Hentze et al., 2010). Under conditions of iron overload, non-transferrin bound iron (NTBI), a potentially toxic form of iron, appears in the circulation (Cabantchik et al., 2005; Jacobs et al., 2005). The chemical composition of NTBI is poorly characterized but it has been suggested that ferric citrate is the predominant species (Hider, 2002). Cellular NTBI uptake takes place mostly in the liver and is mediated by the divalent metal transporter DMT1 and the zinc transporter ZIP14 (Liuzzi et al., 2006). Another transporter, ZIP8 was shown to import NTBI into cells, but the role of ZIP8 in physiological or pathological iron metabolism is still unclear (Wang et al., 2012). Acquisition of NTBI through these transporters requires reduction of ferric iron (Fe^{3+}) to ferrous iron (Fe^{3+}) that can be facilitated by a membrane-bound reductase or extracellular ascorbate (Lane and Lawen, 2008). A study on iron uptake in T lymphocytes suggested that NTBI can be acquired as oligomeric ferric citrate in a mechanism that involves an unknown carrier protein (Arezes et al., 2013). A role of bulk endocytosis in cellular NTBI uptake has also been reported (Sohn et al., 2012). Unlike TBI uptake that is regulated by iron-dependent reduction of TfR1 expression via the iron responsive element (IRE)/iron regulatory proteins (IRP) system, cellular ingress of iron from NTBI does not appear to be controlled by excess iron in the cell (Brissot et al., 2012).

In erythroid precursors, TBI is considered the primary iron source that can provide sufficient iron for hemoglobin production under physiological conditions (Ponka, 1997). To acquire iron from TBI, erythroid cells use the classical receptor-mediated endocytosis pathway in a process that involves binding of diferric transferrin (Tf- Fe_2) to the transferrin receptor TfR1, clathrin-dependent internalization of the Tf-Fe_2/TfR1 complex, endosomal acidification and release of free iron, reduction of Fe^{3+} to Fe^{2+} by the endosomal reductase STEAP3 and transport of Fe^{2+} to the cytosol via the divalent metal transporter DMT1 (Muckenthaler et al., 2017). Efficient TBI uptake in erythroid precursors require endocytic recycling of TfR1, which involves the trafficking molecules Sec15l1, Vps35 and Snx3 (Chen et al., 2013). Although not utilized in

heme synthesis and hemoglobin production, NTBI present in the circulation under conditions of iron overload can be taken up by erythroid precursors and results in ineffective erythropoiesis due to accumulation of redox-active labile iron in these cells (Prus and Fibach, 2011). The mechanisms by which erythroid precursors import NTBI are still unclear. Studies on Belgrade rats indicated that DMT1 is involved in NTBI uptake by erythroid cells (Garrick et al., 1999). In human K562 erythroleukemic cells, enhanced iron uptake stimulated by iron deprivation, that is abrogated by inhibiting protein synthesis, is not associated with induction of DMT1 expression suggesting a role for another unknown transporter in erythroid NTBI uptake (Kovar et al., 2006).

While the mechanism of TBI uptake in erythroid cells is well-characterized, a comprehensive investigation of cellular determinants of TBI acquisition in erythroid cells has not been performed. In addition, identification of transporters and other cellular factors contributing to erythroid NTBI uptake is still required. In this study, we used genome-wide CRISPR-Cas9 loss-of-function screening in human erythroleukemic cells to identify genes whose disruption limits cellular growth when either TBI or NTBI is an exclusive source of iron. In addition to genes with a known role in cellular iron uptake, our screen revealed multiple novel candidates relevant to the process. We validated the role of one gene, *CCDC115,* in cellular TBI acquisition using multiple targeted approaches. Our work demonstrates the power of functional genomics in studying particular cellular processes.

Materials and methods

Cell culture in chemically-defined media and cell viability assays

K562 cells were initially cultured in RPMI 1640 (Thermo Fisher) supplemented with 10% fetal bovine serum (FBS; Corning) and 1% penicillin/streptomycin (PS; Thermo Fisher). Cells were adapted to a chemically-defined serum-free medium formulated as previously described (Neumannova et al., 1995) with some modifications. Adaptation was performed over the course of 4 days, each day reducing FBS by 2.5%. The chemically-defined medium is composed of iron-free RPMI 1640 with L-Glutamine (Thermo Fisher) supplemented with 1mM sodium pyruvate (Thermo Fisher), 15 mM HEPES (Thermo Fisher), 1% PS, 20 μM ethanolamine (Sigma), 5 μM hydrocortisone (Sigma), 20 μM L-ascorbic acid (Thermo Fisher), 50 nM copper sulfate ($CuSO_4$), 0.5 μM zinc sulfate ($ZnSO_4$), 50 nM sodium selenite (Na_2SeO_3), 500 μg/ml bovine serum albumin (BSA; Sigma) and 5 μg/ml insulin (insulin zinc solution; Thermo Fisher). Iron was supplemented as ferric citrate (Sigma) or holo-transferrin (Sigma) as required. Cells were cultured in a humidified incubator with 5% CO_2 at 37° C. Cell viability was assessed by measuring ATP levels using the CellTiter Glo luminescent cell viability assay kit (Promega) following the manufacturer's instructions. Briefly, K562 cells were seeded in the iron-free media at a density of 10^5 cells/ml (10^4 cells/well) in opaque 96-well cell culture plates. For iron-dependent cell growth assays, cells were treated with different concentrations of holo-transferrin (0-10 μg/ml) or ferric citrate (0-10 μM) for 72 hours. For assessment of mutant growth under different iron sources, cells were treated with 8 μg/ml holo-transferrin or 100 μM ferric citrate for 7 days. To measure the ATP content in each well, cells were mixed with 100 μL of the CellTiter Glo reagent and lysed on an orbital plate shaker for 2 minutes. Plates were incubated for 10 minutes at room temperature in the dark to stabilize the signal. Luminescent signals from all wells were read on a Synergy H1 microplate reader (BioTek Instruments).

Genome-wide and focused CRISPR/Cas9 libraries

See supplementary methods on page 111.

Lentiviral production and functional titration

See supplementary methods on page 111.

Genome-wide primary screening

GeCKO v2 library, packaged in lentiviral particles was transduced into $100x10^6$ K562 cells in 12-well plates using the same protocol described for viral titration. Each well, containing $2.5x10^6$ cells, was transduced with 10 μl of the GeCKO v2 virus which results in a MOI of 0.45 (determined from titration). Transduced cells from all wells were pooled and the non-infected cells were eliminated by puromycin selection (2μg/ml) for 7 days during which the infected cells were expanded. The chemically-defined medium used during puromycin selection and expansion of the cell library contained both transferrin-bound iron [TBI: holo-transferrin (8 μg/ml)] and non-transferrin bound iron [NTBI: ferric citrate (100 μM)]. To perform the screen, the pooled cell library was split into TBI and NTBI groups (in T225 flasks) with 2 replicates for each group. In the TBI group, holo-transferrin (8 μg/ml) was the sole iron source whereas in the NTBI group, cells were treated with ferric citrate (100 μM) as an exclusive source of iron. Cells were maintained under the described selective conditions for 14 days. Media were changed every 48 hours and at least $25x10^6$ cells were maintained in each replicate resulting in a representation of ~ 400-fold the library size. At the end of the screen, $25x10^6$ cells from each replicate were washed with PBS and the pellets were saved for DNA extraction.

Secondary screening
$10x10^6$ K562 cells were transduced with the validation library in a 12-well plate using the described lentiviral transduction protocol. Each well, containing $2.5x10^6$ cells, was transduced with 2.5 μl of the validation library virus which results in a MOI of 0.27 (determined from titration). The secondary screen was conducted by applying the same TBI and NTBI conditions described for the primary screen and for the same period of time (14 days). At least $2.5x10^6$ cells were maintained in each condition representing ~ 900-fold the library size. Screens were performed in T25 cell culture flasks and each condition was run in triplicate. At the end of the screen, $2.5x10^6$ cells from each replicate were washed with PBS and the pellets were saved for DNA extraction.

DNA extraction, library preparation and next generation sequencing
See supplementary methods on page 112.

Data processing and computational analysis
See supplementary methods on page 113.

Generation of *TFRC*$^{-/-}$ and *CCDC115*$^{-/-}$ K562 cells
Template assembly for guide sequences targeting *TFRC* or *CCDC115* (sequences shown in Table S2), sgRNA synthesis and Cas9-ribonucleoprotein (RNP) preparations were performed as previously described (Lingeman et al., 2017). Cas9-RNP complexes were delivered to cells using the 4D Nucleofector instrument with the SF Cell Line 4D-Nucleofector X Kit (Lonza). For each sample, 10 μl Cas9-RNP were mixed with $2.5x10^5$ K562 cells suspended in 20 μl SF solution and 25 μl of the mixture were transferred into the a Nucleocuvette well. Nucleofection was performed following the instrument's protocol for K562 cells. Transfected cells were resuspended in culture media, transferred to 12-well plates and recovered for 48 hours. DNA was extracted from each transfected pool using QuickExtract DNA extraction solution (Epicentre), the cleavage site was amplified by PCR and the cleavage efficiency of the used sgRNAs was determined using the T7 endonuclease I assay (NEB) following the manufacturer's protocol. To isolate individual clones, cells were seeded at a density of 25 cells/ml in ClonaCell-TCS semi solid media (Stem Cell technologies) supplemented with antibiotic/antimycotic solution (Thermo Fisher) and 25 μM ferric citrate for 14 days. Individual colonies were expanded in RPMI media (10% FBS, 1% PS) that also contains 25 μM ferric citrate. DNA was extracted from individual clones using the QuickExtract solution and target regions were amplified using Q5 Hot Start High-Fidelity 2X Master Mix (NEB). Clones with biallelic mutations in target regions were identified by *in vitro* digestion with Cas9 nuclease (NEB) following the manufacturer's protocol, and expanded for subsequent analyses. Fragment analysis for both T7 endonuclease and *in vitro* Cas9 digestion assays was performed on 2% agarose gel containing ethidium bromide. Primers used to amplify target regions for the T7 endonuclease and Cas9 *in vitro* digestion assays are shown in (Table S3).

Western blot analysis
Cells were lysed in RIPA buffer (Thermo Fisher) supplemented with protease inhibitor cocktail (Sigma). Proteins in each lysate were quantified using the Pierce BCA protein assay kit (Thermo Fisher). 50 μg of proteins were denatured in Laemmli sample buffer (Biorad) supplemented with 2-mercaptoethanol (Biorad) and run on 4–20% Mini-Protean TGX gels (Biorad). Proteins were

transferred to low fluorescence PVDF membranes using the Trans Blot Turbo RTA low LF mini kit and the Trans Blot Turbo transfer system (Biorad) following the manufacturer's protocol. Membranes were blocked in Odyssey Blocking Buffer (LICOR) for 1 hour and probed overnight with the primary antibodies diluted as follows: mouse anti-TfR1(Thermo Fisher; 1:1000), rabbit anti-CCDC115 (Sigma; 1:1000), mouse anti-α-Tubulin (LICOR; 1:1000), rabbit anti-β-Tubulin (LICOR; 1:1000). Membranes were washed and incubated for 1 hour with the IRDye 800CW-labeled goat anti-rabbit IgG and IRDye 680RD-labeled goat anti-mouse IgG secondary antibodies (LICOR) at a dilution of 1:15000. Membranes were scanned using Odyssey CLx instrument (LICOR).

RNA extraction, reverse transcription and real-time qPCR

RNA was extracted using the RNeasy Mini kit (Qiagen). Reverse transcription (rt) was performed from 0.5 µg RNA using the iScript cDNA synthesis kit (Biorad). Real-time qPCR was conducted using the Ssofast Evagreen Supermix (Biorad) on a CFX Connect real-time PCR instrument (Biorad) following the manufacturer's instructions. Data analysis was performed on the accompanying CFX manager software using the *ΔΔCT* method. Ribosomal protein L19 (*RPL19)* gene was used as a reference gene in qPCR assays. Amplification of RPL19 and TFRC cDNA was performed at an annealing temperature of 60°C and the used primers are shown in (Table S4).

^{59}Fe-Tf uptake assays

^{59}Fe-labeled transferrin (^{59}Fe-Tf) was prepared as previously described (Jenkitkasemwong et al., 2016). Cells grown in chemically-defined serum-free media were deprived from iron by culturing under iron-free conditions for 24 hours. ^{59}Fe-Tf treatments were performed in 12-well plates where 10^7 cells in each well were cultured in 1 ml medium containing 200 nM ^{59}Fe-Tf for 0, 3 and 24 hours. To remove membrane-bound transferrin, cells were washed once with TrypLE Express cell-dissociation solution (Thermo Fisher), followed by multiple washes with RPMI 1640 medium until the radioactivity in the supernatant reaches its minimum. For cell lysis, cells were suspended in a buffer containing 0.2M NaOH and 0.2% SDS and sonicated briefly. An aliquot from each lysate was used for protein quantification using BCA protein assay. ^{59}Fe radioactivity was measured in each lysate using a gamma counter (Perkin Elmer).

Results

Genome-wide screening reveals multiple candidate genes involved in cellular iron uptake

We used the CRISPR-Cas9 genome-wide knockout screening approach to study cellular mechanisms of transferrin-bound iron (TBI) and non-transferrin bound iron (NTBI) uptake in human erythroid cells. As a model for erythroid iron uptake, we used K562 cells that exhibited defective growth in iron-free media but can be rescued upon treatment with holo-transferrin (holo-Tf; TBI) or ferric citrate (Fe (III) citrate; NTBI) in a dose-dependent manner (Figure 2.1A, B). A pool of mutant K562 cells was generated by transducing cells with the GeCKO v2 sgRNA library and the cellular library was screened in the presence of different iron sources to identify mutants with an exclusive growth defect with either TBI or NTBI as the only iron source (Figure 2.2). To avoid loss of mutants whose growth is dependent on either of the two iron forms prior to the beginning of the screen, the cells were maintained in "complete medium", containing both TBI and NTBI, during transduction and puromycin selection. Throughout the screen, cells were cultured either in TBI (8 μg/ml holo-Tf) or NTBI (100 μM Fe (III) citrate) media. To assess the depletion of each mutant in one of the two conditions compared to the other, the abundance of each guide sequence, determined by next generation sequencing, was compared between TBI and NTBI pools. Since each gene is only targeted by 3 guide sequences in the library, a single sgRNA showing differential abundance between TBI and NTBI was sufficient to consider the corresponding gene a candidate. Guide sequences with sufficient representation (average log2 CPM > 2.5), that exhibit a differential abundance between TBI and NTBI with FDR < 0.1, were considered for candidate gene selection (Figure 2.3A). Using these criteria, we identified multiple genes that have a potential role in cellular iron uptake (Table S5). Mutants corresponding to 18 targeted genes exhibited lower abundance in NTBI compared to the TBI pools (FDR<0.05) consistent with a growth defect in NTBI (Figure 2.3B). Similarly, 24 mutants with defective growth only in TBI media were identified (Figure 2.3C). Supporting the validity of the screen, guides targeting *SLC11A1*, a gene encoding a non-transferrin iron transporter, and *TFRC*, which encodes the transferrin receptor TfR1 were depleted from the NTBI and TBI pools respectively (Figure 2.3B, C).

Secondary screening validates depletion of *TFRC* and *CCDC115* mutants in TBI-based medium

A customized sgRNA library (validation library; 8sgRNAs/gene) targeting candidate genes identified from the primary screen was used in secondary screening. The validation library also included several sgRNAs targeting non-candidate genes and multiple non-targeting sgRNAs. We transduced K562 cells with the validation library and screened the knockout pool for mutants with an exclusive growth defect either in the presence of TBI or NTBI as the only source of iron as we did previously for the primary screen. Abundance of each mutant was determined by quantifying the corresponding guide sequence using next generation sequencing, and compared among TBI and NTBI pools. Only a limited number of guides showed significant differential abundance (FDR <0.05) between TBI and NTBI in the secondary screen and the majority of these guides displayed lower abundance in TBI compared to NTBI (Table S6; Figure 2.4A). Following our criteria for candidate gene validation, where at least two guides per gene must show significant differential abundance, only two genes, *TFRC* and *CCDC115,* were validated in the secondary screen. For *TFRC,* 7 out of 8 guides showed lower abundance in TBI compared to NTBI and 6 of them were significant (Figure 2.4B). All *CCDC115* guide sequences displayed lower abundance in TBI compared to NTBI and two of them were significant (Figure 2.4C).

None of the NTBI uptake candidates identified in the primary screen were confirmed in our secondary screening.

CCDC115$^{-/-}$ cells display defective growth in TBI-based media

To confirm the role of CCDC115 in TBI-dependent cellular growth, we generated *CCDC115*$^{-/-}$ cells using the CRISPR-Cas9 gene knockout tool. We also generated *TFRC*$^{-/-}$ cells that were used as a control. For genomic editing of *TFRC* or *CCDC115* target sites, we transfected cells with gene-specific Cas9-RNP complexes. The on-target cleavage efficiency in the transfected pools was confirmed by T7 endonuclease assay (Figure 2.5A). Biallelic mutations in the *TFRC or CCDC115* target locus in isolated clones were revealed by *in vitro* Cas9 cleavage assays (Figure 2.5B). Loss of TfR1 and CCDC115 proteins in selected *TFRC*$^{-/-}$ and *CCDC115*$^{-/-}$ clones was confirmed by western blot analysis (Figure 2.5C). Interestingly, *CCDC115*$^{-/-}$ cells displayed a considerable increase in TfR1 levels compared to wild-type (WT) K562 cells cultured under the same conditions (Figure 2.5C). To assess the growth of *TFRC*$^{-/-}$ and *CCDC115*$^{-/-}$ cells in the different iron media, we treated cells in the chemically-defined iron-free media with holo-Tf (8μg/ml) or Fe (III) citrate (100 μM) for 7 days. Cell viability assays showed that all cells exhibited a growth defect in iron-free media but unlike wild-type cells that can be efficiently rescued with both holo-Tf and fe (III) citrate, *TFRC*$^{-/-}$ and *CCDC115*$^{-/-}$ cells displayed a defective growth in holo-Tf but not Fe (III) citrate media (Figure 2.5D).

CCDC115 is involved in cellular TBI uptake

Cellular iron loading is known to downregulate TfR1 expression by destabilizing its mRNA (Wang and Pantopoulos, 2011). Hence, we used TfR1 transcript levels as an indicator for intracellular iron levels in wild-type and *CCDC115*$^{-/-}$ cells cultured in the presence of different iron sources. While treatment with both TBI and NTBI resulted in decreased TfR1 mRNA levels in wild-type cells, only Fe (III) citrate (NTBI) but not holo-Tf (TBI) resulted in lower TfR1 mRNA levels in *CCDC115*$^{-/-}$ cells (Figure 2.6A). To study the effect of CCDC115 loss on TBI uptake, we labeled Tf with radioactive iron (^{59}Fe), cultured cells with ^{59}Fe-Tf and measured intracellular radioactivity at different time points. Similar to *TFRC*$^{-/-}$ cells, *CCDC115*$^{-/-}$ cells accumulated less ^{59}Fe compared to wild-type controls (Figure 2.6B).

Discussion

Erythroid cells are capable of taking up both transferrin-bound iron (TBI) and non-transferrin bound iron (NTBI). Iron acquisition from transferrin by erythroid precursors is a tightly regulated physiological process that is essential for hemoglobin production. The majority of key players in erythroid TBI uptake were identified but only little is known about the regulatory cellular components contributing to this process. Regulation of TBI utilization can occur at many levels including TfR1 expression, TfR1 trafficking and recycling, clathrin-mediated endocytosis, vesicle acidification, ferric iron reduction and endosomal efflux. NTBI, which is only available in the circulation during iron overload, can result in toxic effects when taken up by erythroid cells. While NTBI transporters were identified in certain tissues, mechanisms underlying erythroid NTBI uptake remain unclear. Identification of cellular components contributing to TBI and NTBI uptake in erythroid cells can reveal novel links between cellular iron homeostasis and other cellular processes and is crucial for better understanding of the etiology of certain anemias.

We employed a genome-wide functional screening approach using the CRIPSR-Cas9 knockout tool to identify indispensable determinants of TBI and NTBI uptake in human erythroid cells. The screen was performed in chemically-defined serum-free media where the desired form of iron was supplemented. The model cell line we used, K562, is an erythroleukemic cell line whose growth and proliferation are strongly dependent on iron and can acquire both TBI and NTBI (Figure 2.1). Our primary screen revealed multiple candidate genes with a potential role in TBI or NTBI acquisition by erythroid cells (Figure 2.3). However, only two hits were validated in our secondary screening approach. It is critical to mention that the screens we performed are negative selection screens where the readout depends on depletion due to defective cellular growth. Unlike positive selection screening, negative selection maintains the library complexity at the end of a screen, which renders detecting of partial depletion phenotypes under a selective condition a challenge. Accordingly, the lack of validation of a candidate in the secondary screen does not necessarily mean that the hit is a false positive. Secondary screening using more sensitive readout methods or validation of the phenotype by targeted gene knockout is required for confirming primary hits.

Although not validated in the secondary screen, a potential role of certain primary candidates in iron uptake is supported by evidence from the literature. Among the NTBI uptake candidate genes, we identified *SLC11A1*, which encodes the iron transporter Nramp1. Nramp1 is involved in iron transport across the phagolysosomal membrane and is involved in efficient iron recycling following erythrophagocytosis (Biggs et al., 2001; Soe-Lin et al., 2009). However, the potential role of Nramp1 in NTBI uptake by erythroid cells is opposed by the reported phagocyte-specific expression of the protein (Cellier, 2013). Additionally, the primary screen revealed components of vesicular trafficking, a key process for transferrin acquisition, as candidate determinants of TBI uptake. One of these candidate genes encodes VPS35, a retromer component that is involved in the recycling of the transferrin receptor (TfR1). VPS35 and its partner Snx3 have been shown to interact with TfR1 allowing its sorting to recycling endosomes (Chen et al., 2013). Another candidate gene, *DYNLRB2*, encodes a component of cytoplasmic dynein which is required for vesicular transport. Interestingly, a genome-wide association study of total iron binding capacity (TIBC) in humans revealed an association between serum TIBC levels and single nucleotide polymorphisms (SNPs) near the *DYNLRB2* gene (Li et al., 2015), thus supporting the role of DYNLRB2 in iron metabolism. Further, the screen identified the insulin receptor INSR as a

potential determinant of TBI uptake. In agreement with this finding, insulin has been shown to promote iron uptake in liver cells by inducing TfR1 expression (Biswas et al., 2013). It is important to mention that insulin was used as a supplement for the chemically-defined media used in our screen to promote cell proliferation under serum-free conditions. Consequently, it can be suggested that induction of iron uptake is one of the mechanisms underlying insulin-induced cell proliferation. Additional TBI uptake candidates that we revealed are cellular proteins and microRNAs (miRNAs) that were not previously linked to iron trafficking. We did not find common TBI uptake candidate genes between our screen and other functional screens studying transferrin endocytosis in other cell types (Collinet et al., 2010; Galvez et al., 2007).

Two of the candidate genes, *TFRC* and *CCDC115,* were validated in the secondary screen (Figure 2.4) and were considered for follow-up studies. *TFRC* encodes the transferrin receptor TfR1, which is an indispensable determinant of TBI uptake and its loss was expected to impair TBI-dependent cell proliferation. The other validated gene encodes the coiled-coil domain-containing protein 115 (CCDC115). The functional relevance of CCDC115 to TBI uptake is supported by its role as a vacuolar-ATPase (V-ATPase) assembly factor that is involved in endosomal acidification (Miles et al., 2017), a critical process for iron acquisition from transferrin. Similar to $TFRC^{-/-}$ cells, $CCDC115^{-/-}$ cells exhibited defective TBI-dependent growth (Figure 2.5D), indicating that CCDC115 is required for cellular TBI utilization. TfR1 expression is inversely related to cellular iron levels due to iron-dependent destabilization of TfR1 mRNA (Wang and Pantopoulos, 2011). Unlike NTBI treatment that decreased TfR1 mRNA levels, TBI treatment of $CCDC115^{-/-}$ cells had no effect on TfR1 expression (Figure 2.6A) suggesting an impairment of transferrin-mediated iron uptake. Further, a notable upregulation of TfR1 protein levels was observed in $CCDC115^{-/-}$ compared to wild-type cells cultured in a medium containing both TBI and NTBI (Figure 2.5C). In addition, comparable to $TFRC^{-/-}$ cells, $CCDC115^{-/-}$ cells displayed lower assimilation of ^{59}Fe from ^{59}Fe -labeled transferrin (Figure 2.6B). Collectively, these observations indicate that CCDC115 plays a role in TBI uptake. Consistent with this finding, it has been shown that disruption of CCDC115 impaired vesicle acidification and resulted in stabilizing HIF1α and IRP2, whose degradation is mediated by cellular iron (Miles et al., 2017). In addition to its role in vesicle acidification, CCDC115 was reported to be involved in protein glycosylation (Jansen et al., 2016). Since glycosylation of TfR1 is essential for trafficking to the cell surface, dimerization and Tf binding (Williams and Enns, 1991; Yang et al., 1993), we speculated that impaired TBI uptake in $CCDC115^{-/-}$ cells is a result of defective TfR1 glycosylation. However, western blot analysis did not detect any size-changing glycosylation errors in TfR1 in $CCDC115^{-/-}$ cells (Figure 2.5C). CCDC115 contribution to TBI acquisition is likely due to its role in endosomal acidification. Further studies are required to investigate the role of CCDC115 in erythroid iron uptake and erythropoiesis *in vivo*.

Our study provided novel insight into cellular iron uptake. Despite the inability to simultaneously validate many candidate genes, the secondary screen still revealed the genetic disruptions resulting in robust depletion phenotypes. More sensitive and reproducible screens would require improved designs where positive selection can be applied instead of negative selection. Overall, our work demonstrated that CRISPR-based genetic screening could be an effective strategy to study mechanisms of cellular iron homeostasis.

Figures

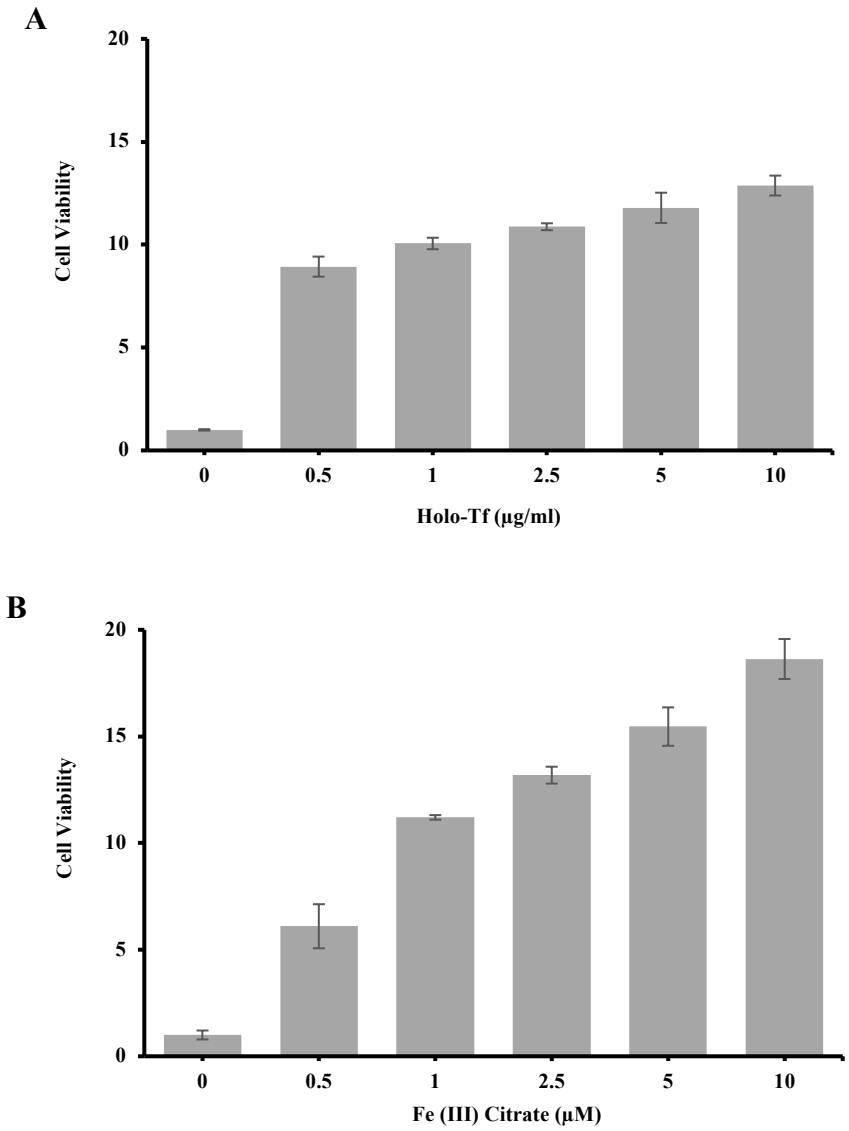

Figure 2.1: Iron-dependent growth of human erythroleukemic K562 cells. Cells grown in chemically-defined iron-free media were treated with multiple concentrations of holo-Tf (A) or ferric citrate (B). Cell viability was evaluated at 72 hours by CellTiter Glo assay. Cell viability is represented relative to iron-free conditions. Data are represented as mean ± standard deviation (n=3).

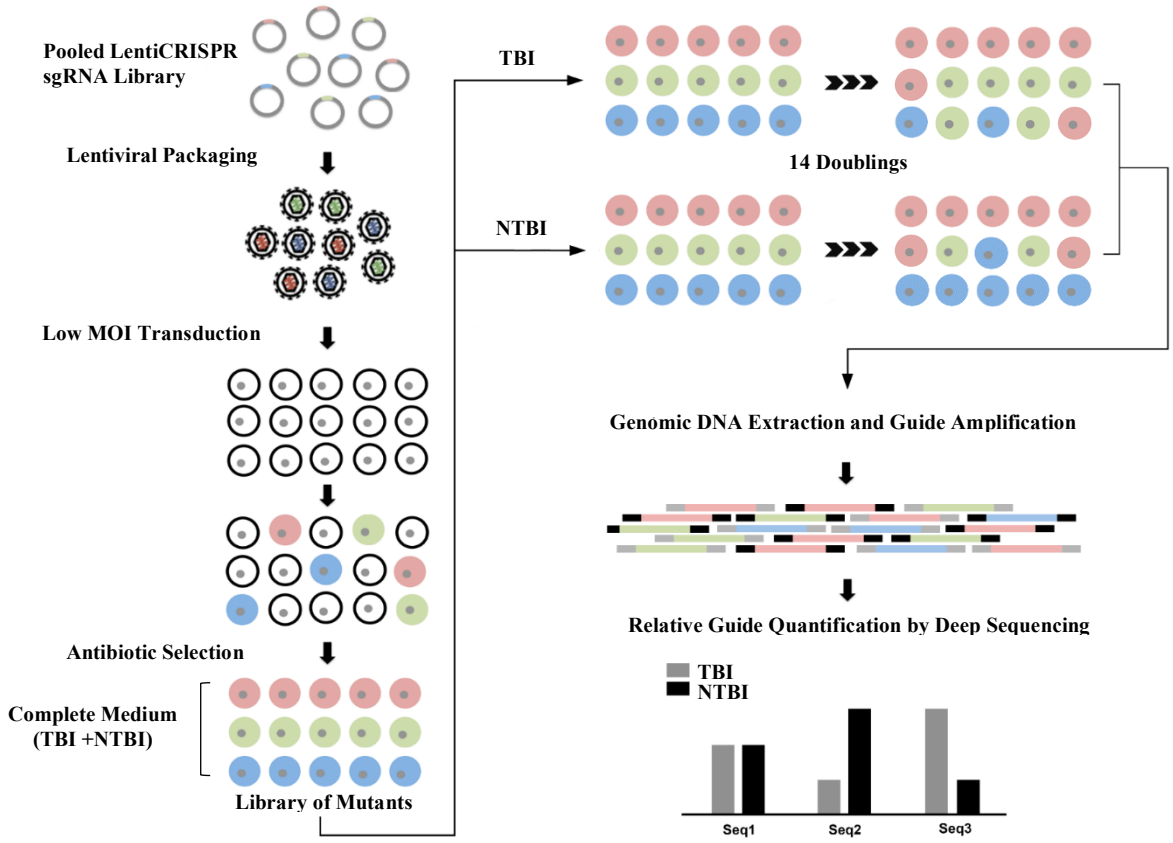

Figure 2.2: Workflow of the CRISPR-Cas9 genome-wide screening approach designed to study cellular iron uptake. Cells are transduced with the Cas9-sgRNA library at a low MOI to generate a library of mutants. Screening is performed in a chemically-defined serum-free media where either transferrin bound iron (TBI) or non-transferrin bound iron (NTBI) is the source of iron. Mutants with a defective uptake of a particular iron form are only depleted from the pool where this form is the exclusive source of iron. Guide sequences encoding all sgRNAs are PCR-amplified and quantified by deep sequencing. Each guide sequence is used as a barcode to label the corresponding mutant. Abundance of each guide sequence is compared between the TBI and NTBI pools.

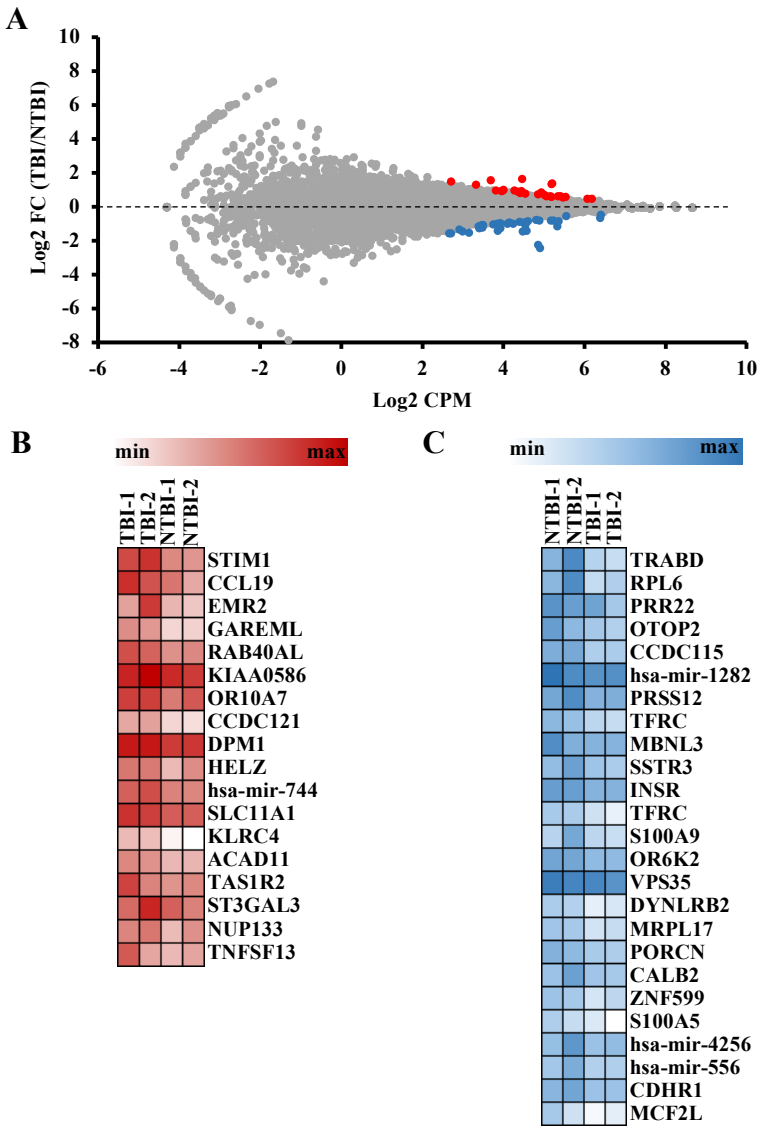

Figure 2.3: Primary screen reveals candidate genes involved in cellular iron uptake. (A) Scatter plot showing differential abundance of each guide sequence between TBI and NTBI. Fold changes (FC) are calculated by dividing the average normalized counts (n=2) for each guide in TBI to that in NTBI. Log2 FCs are plotted against the average abundance of each guide in the pool represented as Log2 counts per million (CPM). Differentially abundant guides with FDR < 0.1 and log2 CPM > 2.5 were selected as candidates. Guide sequences depleted in NTBI relative to TBI have Log2 FC > 0 and the top hits (FDR <0.1) are represented as red dots. Guide sequences depleted in TBI relative to NTBI have Log2 FC < 0 and the top hits (FDR <0.1) are represented as blue dots. (B, C) Heatmaps of the normalized counts of gene-specific guide sequences that are depleted in NTBI (B) or TBI (C) with FDR < 0.05. The screen was run in duplicate and the normalized counts for each replicate are shown in the heatmaps.

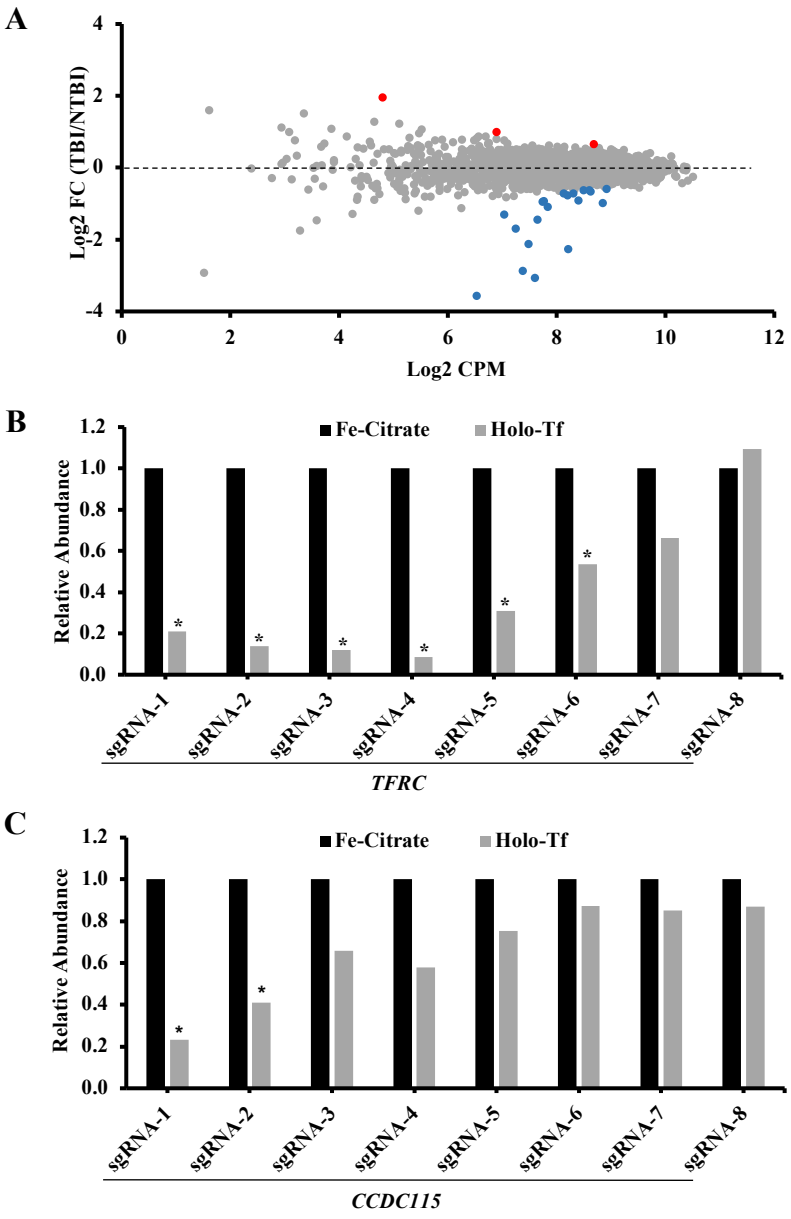

Figure 2.4: Secondary screen validates depletion of *TFRC* and *CCDC115* mutants in transferrin media. (A) Scatter plot showing differential abundance of each guide sequence between TBI and NTBI. Fold changes (FC) are calculated by dividing the average normalized counts (n=3) for each guide sequence in TBI to that in NTBI. Log2 FCs are plotted against the average abundance of each guide in the pool represented as Log2 counts per million (CPM). Differentially abundant guides with FDR < 0.05 were selected as candidates. Guide sequences depleted in NTBI relative to TBI have Log2 FC > 0 and the top hits (FDR <0.05) are represented as red dots. Guide sequences depleted in TBI relative to NTBI have Log2 FC < 0 and the top hits (FDR <0.05) are represented as blue dots. (B, C) Depletion of guide sequences targeting *TFRC* gene (B) and *CCDC115* gene (C) in TBI (Holo-Tf) relative to NTBI (Fe (III) Citrate). Bars represent abundance of each guide sequence relative to NTBI. Data are represented as mean (n=3). *= FDR < 0.05.

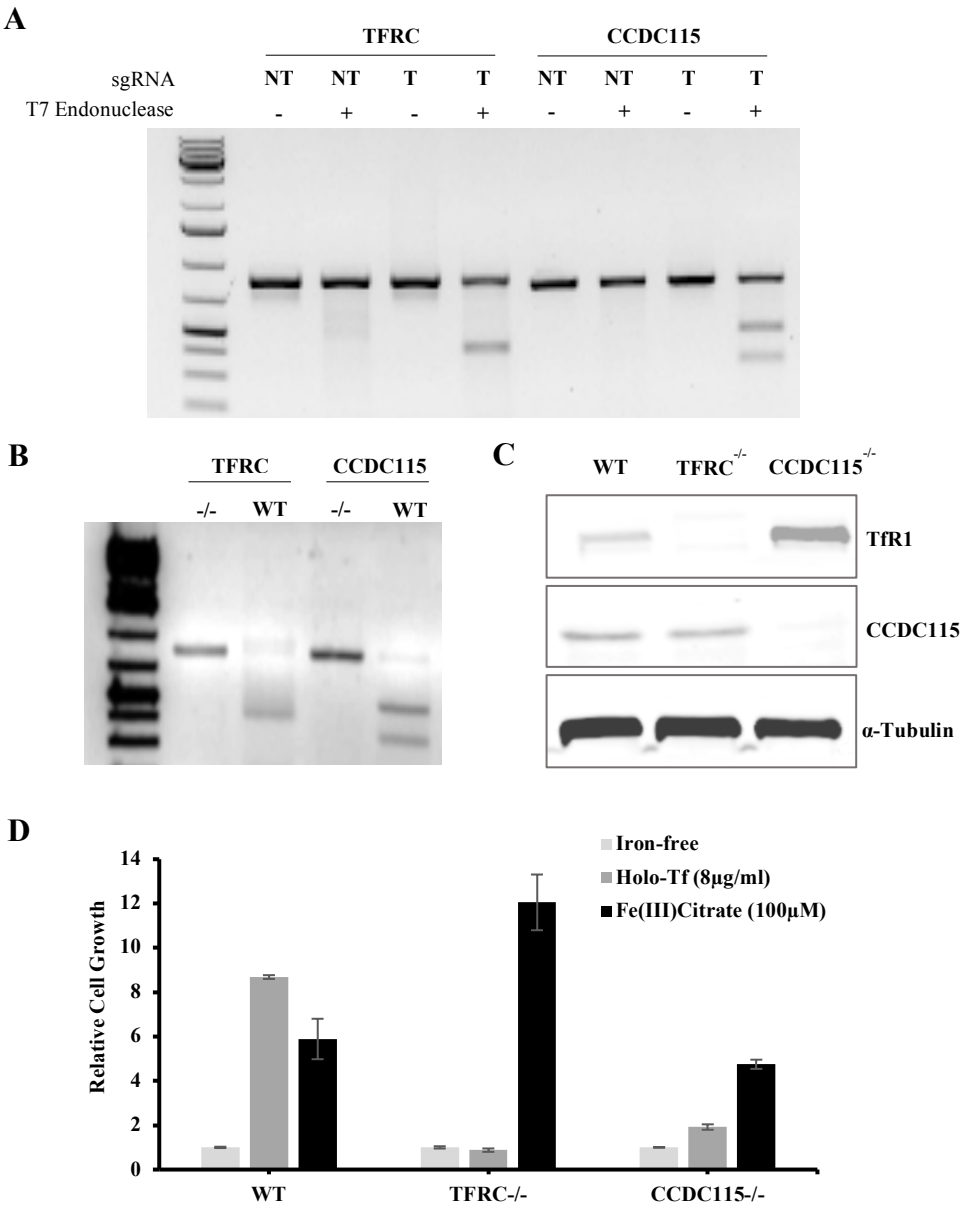

Figure 2.5: CCDC115 disruption inhibits transferrin-dependent cellular growth. (A) Validation of CRISPR-Cas9 editing of *TFRC* and *CCDC115* genes in K562 cells. Cells were nucleofected with Cas9 protein complexed with targeting (T) or non-targeting (NT) sgRNAs. Gene editing was assessed by T7 endonuclease assays. (B) Validation of biallelic editing of *TFRC* and *CCDC115* in K562 clones. Clones were isolated from K562 pools nucleofected with Cas9-RNPs targeting *TFRC* or *CCDC115*. Biallelic mutations were identified by *in vitro* Cas9 (IVC) digestion assays. IVC data of selected *TFRC*$^{-/-}$ and *CCDC115*$^{-/-}$ is shown. (C) Western blot analysis showing complete loss of TFR1 and CCDC115 in the selected clones. (D) Validation of defective growth of *TFRC* $^{-/-}$ and *CCDC115*$^{-/-}$ K562 clones in TBI. Cells grown in chemically-defined iron-free media were treated with 8 µg/ml holo-Tf or 100 µM ferric citrate. Cell viability was evaluated after 7 days by CellTiter Glo assay. Bars represent cellular growth relative to iron-free conditions. Data are represented as mean ± standard deviation (n=3).

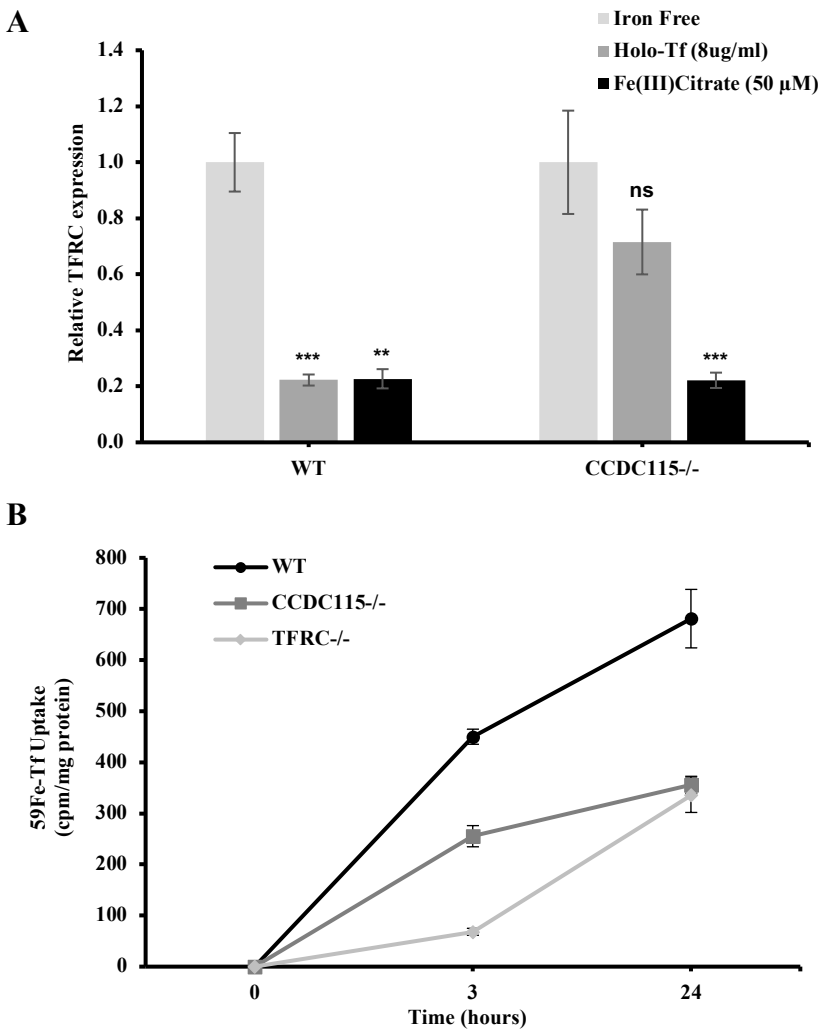

Figure 2.6: Inactivation of CCDC115 impairs cellular TBI uptake in K562 cells. (A) TFRC expression was evaluated by real-time qPCR. Cells grown in chemically-defined iron-free media were treated with 8 µg/ml holo-Tf or 100 µM ferric citrate for 48 hours. Bars represent TFRC expression relative to iron-free conditions. Data are represented as mean ± standard error (n=3). Statistical significance for the difference of TFRC expression between iron-treated and iron-free conditions was determined by Student's t-test, where ns = non-significant, **=$P < 0.01$, ***=$P < 0.001$. (B) Cellular ^{59}Fe-Tf uptake was evaluated by measuring radioactivity in cells treated with 200 nM Tf labeled with ^{59}Fe for the shown time points. Radioactivity was analyzed by gamma counting and normalized to the protein content in each sample. Data represents mean ± standard deviation (n=3).

References

Abbaspour, N., Hurrell, R., and Kelishadi, R. (2014). Review on iron and its importance for human health. J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci. *19*, 164–174.

Arezes, J., Costa, M., Vieira, I., Dias, V., Kong, X.L., Fernandes, R., Vos, M., Carlsson, A., Rikers, Y., Porto, G., et al. (2013). Non-Transferrin-Bound Iron (NTBI) Uptake by T Lymphocytes: Evidence for the Selective Acquisition of Oligomeric Ferric Citrate Species. PLOS ONE *8*, e79870.

Biggs, T.E., Baker, S.T., Botham, M.S., Dhital, A., Barton, C.H., and Perry, V.H. (2001). Nramp1 modulates iron homoeostasis in vivo and in vitro: evidence for a role in cellular iron release involving de-acidification of intracellular vesicles. Eur. J. Immunol. *31*, 2060–2070.

Biswas, S., Tapryal, N., Mukherjee, R., Kumar, R., and Mukhopadhyay, C.K. (2013). Insulin promotes iron uptake in human hepatic cell by regulating transferrin receptor-1 transcription mediated by hypoxia inducible factor-1. Biochim. Biophys. Acta BBA - Mol. Basis Dis. *1832*, 293–301.

Brissot, P., Ropert, M., Le Lan, C., and Loréal, O. (2012). Non-transferrin bound iron: A key role in iron overload and iron toxicity. Biochim. Biophys. Acta BBA - Gen. Subj. *1820*, 403–410.

Cabantchik, Z.I., Breuer, W., Zanninelli, G., and Cianciulli, P. (2005). LPI-labile plasma iron in iron overload. Best Pract. Res. Clin. Haematol. *18*, 277–287.

Cellier, M.F.M. (2013). Cell-Type Specific Determinants of NRAMP1 Expression in Professional Phagocytes. Biology *2*, 233–283.

Chen, C., Garcia-Santos, D., Ishikawa, Y., Seguin, A., Li, L., Fegan, K.H., Hildick-Smith, G.J., Shah, D.I., Cooney, J.D., Chen, W., et al. (2013). Snx3 Regulates Recycling of the Transferrin Receptor and Iron Assimilation. Cell Metab. *17*, 343–352.

Coates, T.D. (2014). Physiology and Pathophysiology of Iron in Hemoglobin-Associated Diseases. Free Radic. Biol. Med. *72*, 23–40.

Collinet, C., Stöter, M., Bradshaw, C.R., Samusik, N., Rink, J.C., Kenski, D., Habermann, B., Buchholz, F., Henschel, R., Mueller, M.S., et al. (2010). Systems survey of endocytosis by multiparametric image analysis. Nature *464*, 243–249.

Galvez, T., Teruel, M.N., Heo, W.D., Jones, J.T., Kim, M.L., Liou, J., Myers, J.W., and Meyer, T. (2007). siRNA screen of the human signaling proteome identifies the PtdIns(3,4,5)P3-mTOR signaling pathway as a primary regulator of transferrin uptake. Genome Biol. *8*, R142.

Garrick, L.M., Dolan, K.G., Romano, M.A., and Garrick, M.D. (1999). Non-transferrin-bound iron uptake in Belgrade and normal rat erythroid cells. J. Cell. Physiol. *178*, 349–358.

Gilardi, G., and Nardo, G.D. (2017). Heme iron centers in cytochrome P450: structure and catalytic activity. Rendiconti Lincei *28*, 159–167.

Han, A.-P., Yu, C., Lu, L., Fujiwara, Y., Browne, C., Chin, G., Fleming, M., Leboulch, P., Orkin, S.H., and Chen, J.-J. (2001). Heme-regulated eIF2α kinase (HRI) is required for translational regulation and survival of erythroid precursors in iron deficiency. EMBO J. *20*, 6909–6918.

Hentze, M.W., Muckenthaler, M.U., Galy, B., and Camaschella, C. (2010). Two to Tango: Regulation of Mammalian Iron Metabolism. Cell *142*, 24–38.

Hider, R.C. (2002). Nature of nontransferrin-bound iron. Eur. J. Clin. Invest. *32*, 50–54.

Jacobs, E.M.G., Hendriks, J.C.M., van Tits, B.L.J.H., Evans, P.J., Breuer, W., Liu, D.Y., Jansen, E.H.J.M., Jauhiainen, K., Sturm, B., Porter, J.B., et al. (2005). Results of an international round robin for the quantification of serum non-transferrin-bound iron: Need for defining standardization and a clinically relevant isoform. Anal. Biochem. *341*, 241–250.

Jansen, J.C., Cirak, S., van Scherpenzeel, M., Timal, S., Reunert, J., Rust, S., Pérez, B., Vicogne, D., Krawitz, P., Wada, Y., et al. (2016). CCDC115 Deficiency Causes a Disorder of Golgi Homeostasis with Abnormal Protein Glycosylation. Am. J. Hum. Genet. *98*, 310–321.

Jenkitkasemwong, S., Wang, C.-Y., and Knutson, M.D. (2016). Measurement of Transferrin- and Non-transferrin-bound Iron Uptake by Mouse Tissues. Bio-Protoc. *6*.

Kohgo, Y., Ikuta, K., Ohtake, T., Torimoto, Y., and Kato, J. (2008). Body iron metabolism and pathophysiology of iron overload. Int. J. Hematol. *88*, 7–15.

Kovar, J., Neubauerova, J., Cimburova, M., Truksa, J., Balusikova, K., and Horak, J. (2006). Stimulation of non-transferrin iron uptake by iron deprivation in K562 cells. Blood Cells. Mol. Dis. *37*, 95–99.

Lane, D.J.R., and Lawen, A. (2008). Non-transferrin Iron Reduction and Uptake Are Regulated by Transmembrane Ascorbate Cycling in K562 Cells. J. Biol. Chem. *283*, 12701–12708.

Li, J., Lange, L.A., Duan, Q., Lu, Y., Singleton, A.B., Zonderman, A.B., Evans, M.K., Li, Y., Taylor, H.A., Willis, M.S., et al. (2015). Genome-wide admixture and association study of serum iron, ferritin, transferrin saturation and total iron binding capacity in African Americans. Hum. Mol. Genet. *24*, 572–581.

Lingeman, E., Jeans, C., and Corn, J.E. (2017). Production of Purified CasRNPs for Efficacious Genome Editing. Curr. Protoc. Mol. Biol. *120*, 31.10.1-31.10.19.

Liuzzi, J.P., Aydemir, F., Nam, H., Knutson, M.D., and Cousins, R.J. (2006). Zip14 (Slc39a14) mediates non-transferrin-bound iron uptake into cells. Proc. Natl. Acad. Sci. U. S. A. *103*, 13612–13617.

Miles, A.L., Burr, S.P., Grice, G.L., and Nathan, J.A. (2017). The vacuolar-ATPase complex and assembly factors, TMEM199 and CCDC115, control HIF1α prolyl hydroxylation by regulating cellular iron levels. ELife *6*.

Muckenthaler, M.U., Rivella, S., Hentze, M.W., and Galy, B. (2017). A Red Carpet for Iron Metabolism. Cell *168*, 344–361.

Neumannova, V., Richardson, D.R., Kriegerbeckova, K., and Kovar, J. (1995). Growth of human tumor cell lines in transferrin-free, low-iron medium. Vitro Cell. Dev. Biol. - Anim. *31*, 625–632.

Oexle, H., Gnaiger, E., and Weiss, G. (1999). Iron-dependent changes in cellular energy metabolism: influence on citric acid cycle and oxidative phosphorylation. Biochim. Biophys. Acta *1413*, 99–107.

Pantopoulos, K. (2004). Iron metabolism and the IRE/IRP regulatory system: an update. Ann. N. Y. Acad. Sci. *1012*, 1–13.

Ponka, P. (1997). Tissue-Specific Regulation of Iron Metabolism and Heme Synthesis: Distinct Control Mechanisms in Erythroid Cells. Blood *89*, 1–25.

Prus, E., and Fibach, E. (2011). Uptake of Non-Transferrin Iron by Erythroid Cells.

Puig, S., Ramos-Alonso, L., María Romero, A., and Teresa Martínez-Pastor, M. (2017). The elemental role of iron in DNA synthesis and repair. Metallomics *9*, 1483–1500.

Shah, Y.M., and Xie, L. (2014). Hypoxia-Inducible Factors Link Iron Homeostasis and Erythropoiesis. Gastroenterology *146*, 630–642.

Soe-Lin, S., Apte, S.S., Andriopoulos, B., Andrews, M.C., Schranzhofer, M., Kahawita, T., Garcia-Santos, D., and Ponka, P. (2009). Nramp1 promotes efficient macrophage recycling of iron following erythrophagocytosis in vivo. Proc. Natl. Acad. Sci. *106*, 5960–5965.

Sohn, Y.-S., Ghoti, H., Breuer, W., Rachmilewitz, E., Attar, S., Weiss, G., and Cabantchik, Z.I. (2012). The role of endocytic pathways in cellular uptake of plasma non-transferrin iron. Haematologica *97*, 670–678.

Wang, J., and Pantopoulos, K. (2011). Regulation of cellular iron metabolism. Biochem. J. *434*, 365–381.

Wang, C.-Y., Jenkitkasemwong, S., Duarte, S., Sparkman, B.K., Shawki, A., Mackenzie, B., and Knutson, M.D. (2012). ZIP8 Is an Iron and Zinc Transporter Whose Cell-surface Expression Is Up-regulated by Cellular Iron Loading. J. Biol. Chem. *287*, 34032–34043.

Williams, A.M., and Enns, C.A. (1991). A mutated transferrin receptor lacking asparagine-linked glycosylation sites shows reduced functionality and an association with binding immunoglobulin protein. J. Biol. Chem. *266*, 17648–17654.

Yang, B., Hoe, M.H., Black, P., and Hunt, R.C. (1993). Role of oligosaccharides in the processing and function of human transferrin receptors. Effect of the loss of the three N-glycosyl oligosaccharides individually or together. J. Biol. Chem. *268*, 7435–7441.

Chapter 3

Investigation of Cellular Heme Trafficking and Heme-induced Erythroid Differentiation by Genome-scale CRISPR-Cas9 Knockout Screening

Introduction

Heme is a metalloporphyrin that is involved in multiple fundamental biological functions. The cellular activity of heme is mainly due to its function as a prosthetic group of key enzymes and other cellular proteins. Heme enzymes include the electron transport chain components cytochrome c and succinate dehydrogenase, drug metabolizing cytochrome P450 (CYP450) enzymes and antioxidant enzymes (catalase, peroxidase) (Reedy and Gibney, 2004). In addition, heme is an essential constituent of the oxygen carrier proteins hemoglobin and myoglobin (Gunsalus et al., 1977). Heme is also involved in cell signaling and other intracellular regulatory processes. For example, heme regulates gene expression through binding to transcription factor Bach1 (Igarashi and Sun, 2006). Protein synthesis can also be regulated by heme in a mechanism involving the heme-regulated eukaryotic initiation factor 2α (eIF2α) kinase (HRI) (Igarashi et al., 2008). Moreover, heme can regulate cellular iron homeostasis by oxidizing iron regulatory protein 2 (IRP2) resulting in its ubiquitination and proteasomal degradation (Ishikawa et al., 2005). Finally, heme itself can regulate its own intracellular levels in a tissue-specific manner. In non-erythroid cells, heme represses the 5-aminolevulinic acid synthase (ALAS1), the rate limiting enzyme in the heme biosynthesis pathway (Kolluri et al., 2005) while it induces the expression of the heme catabolizing enzyme heme oxygenase 1 (HMOX1) (Sun et al., 2002).

Mammalian cells can produce heme through a well-characterized biosynthetic pathway that requires glycine, succinyl-CoA and ferrous iron as substrates and involves mitochondrial as well as cytoplasmic enzymes (Schultz et al., 2010). Accordingly, transport of substrates, intermediates and heme itself across mitochondrial membranes is essential. Heme efflux from the mitochondria to the cytosol is controlled by the mitochondrial heme exporter protein FLVCR1b (Chiabrando et al., 2012). However, the mechanisms governing the transport of heme biosynthesis intermediates are poorly understood. Cells can also acquire heme from the surrounding environment but the mechanisms of heme import in mammalian cells are still unclear. The proton coupled folate transporter PCFT (formerly HCP1) was initially believed to be a mammalian intestinal heme importer (Shayeghi et al., 2005). However, the physiological role of PCFT in intestinal heme uptake is still debatable since folate was identified as its actual substrate (Yuan et al., 2013). HRG-1 is another mammalian transporter protein that was shown to import heme (Rajagopal et al., 2008). HRG-1 mainly localizes to endolysosomal compartments and was shown to transport heme from phagolysosomes during erythrophagocytosis in iron-recycling macrophages (White et al., 2013). Receptor-mediated endocytosis is a plausible route for cellular heme import (West and Oates, 2008). Heme bound to the plasma protein hemopexin was shown to be taken up via the LRP/CD91 cell surface receptor by receptor-mediated endocytosis (Hvidberg et al., 2005). Studies in intestinal and erythroid cells characterized cell surface proteins that can bind heme with high affinity (Galbraith et al., 1985; Gräsbeck et al., 1979; Majuri and Gräsbeck, 1987; Majuri R., 2009). Although these findings indicated the presence of a cell surface receptor for heme, the identity of such receptor is still unknown. Additional support for the role of endocytosis in heme acquisition is provided by the evidence that cellular heme uptake is

inhibited at a low temperature (Worthington et al., 2001), and upon inhibition of oxidative phosphorylation (Vaghefi et al., 2000). Accumulation of free heme in the cell induces oxidative stress and results in undesirable toxic effects (Kumar and Bandyopadhyay, 2005). The plasma membrane transporter FLVCR was shown to export free cellular heme and protect from its cytotoxic effects (Quigley et al., 2004). Another transporter that was shown to mediate cellular heme efflux is ABCG2, a member of the ATP-binding cassette (ABC) family of transporters (Desuzinges-Mandon et al., 2010). Unlike FLVCR1, evidence of a protective effect of ABCG2 *in vivo* is lacking. Alternative mechanisms of protection from heme toxicity involve heme degradation by heme oxygenase (HMOX) enzymes that can break down heme into biliverdin, ferrous iron and carbon monoxide (Kumar and Bandyopadhyay, 2005). HMOX1 is known to be induced by intracellular heme through a mechanism that involves the transcriptional repressor Bach1 (Sun et al., 2002).

In addition to its structural role as a component of hemoglobin, heme regulates erythroid differentiation at multiple levels. First, heme can stimulate its own synthesis through a positive feedback mechanism (Granick and Sassa, 1978), a process that ensures the production of adequate heme to meet the high demand during erythroid differentiation. Unlike ALAS1, the erythroid-specific ALAS (ALAS2) is not repressed by heme, but it is rather induced by heme through an IRP-dependent mechanism (Fujiwara and Harigae, 2015). Second, heme induces the expression of globin genes at the transcriptional levels by alleviating Bach1 repression of these genes (Tahara et al., 2004), or at the translational level by inhibiting the activity of HRI and subsequent activation of translation initiation factor eIF2α (Han et al., 2001). Third, heme-mediated activation of the antioxidant transcription factor Nrf2 provides an intracellular environment protected from heme-induced oxidative stress during erythroid differentiation (Chiabrando et al., 2014). It is widely accepted that under physiological conditions, *de novo* biosynthesis of heme in erythroid precursors provides sufficient amounts of heme to be utilized during erythropoiesis. However, erythroid cells are capable of taking up extracellular heme which might be utilized in pathological forms of erythropoiesis (Korolnek and Hamza, 2015) or could induce toxicity in erythroid precursors in the absence of other erythropietic signals.

Here, we used a CRISPR-Cas9-based functional genomic approach to comprehensively investigate cellular mechanisms involved in erythroid heme trafficking and heme-induced erythroid differentiation. We showed that receptor-mediated endocytosis and endosomal/endolysosomal acidification are essential components of heme uptake in erythroid cells. We further identified novel regulatory cellular factors/networks that could potentially contribute to heme-induced erythroid differentiation. Our work demonstrates that genome-wide screening using the CRISPR-Cas9 system is a powerful tool for understanding heme trafficking in mammalian cells.

Materials and methods

Cell viability assays

Cell viability assays were performed by measuring ATP levels using the CellTiter Glo luminescent cell viability assay kit (Promega) following the manufacturer's instructions. Hemin stock solution (4 mM) was prepared by dissolving 13 mg hemin (Sigma) in 200 µl NaOH (0.5 M), adding 250 µl Tris-HCl (1M, PH=7.5) and completing the volume to 5 ml with water. K562 cells were seeded at a density of 10^5 cells/ml (10^4 cells/well) in opaque 96-well cell culture plates and treated with 0, 30 or 60 µM hemin for 72 hours. Following the incubation period, cells were mixed with 100 µL of the CellTiter Glo reagent and lysed on an orbital plate shaker for 2 minutes. Plates were incubated for 10 minutes at room temperature in the dark to stabilize the signal. Luminescent signals from all wells were read on a Synergy H1 microplate reader (BioTek Instruments).

Genome-wide CRISPR/Cas9 library amplification

See supplementary methods on page 111.

Lentiviral production and functional titration

See supplementary methods on page 111.

Genome-wide CRISPR/Cas9 knockout screening

GeCKO v2 library, packaged in lentiviral particles was transduced into $100x10^6$ K562 cells in 12-well plates using the same protocol described for viral titration. Each well, containing $2.5x10^6$ cells, was transduced with 10 µl of the GeCKO v2 virus which results in a MOI of 0.45 (determined from titration). Transduced cells from all wells were pooled and the non-infected cells were removed by puromycin selection (2µg/ml) for 7 days during which the infected cells were expanded. To perform the screen, the pooled cell library was split into treatment (30 µM hemin) and vehicle (equivalent NaOH) conditions where at least $25x10^6$ cells were maintained in each condition resulting in a representation of ~ 400-fold the library size. Selection was applied for 7 days which corresponds to 7 K562 cell doublings and the media were changed every 48 hours. Screens were performed in T225 cell culture flasks and each condition was run in duplicate. At the end of the screen, $25x10^6$ cells from each replicate were washed with PBS and the pellets were stored at -80° C.

DNA extraction, library preparation and next generation sequencing

See supplementary methods on page 112.

Data processing and computational analysis

See supplementary methods on page 113.

Generation of *CCDC115*$^{-/-}$ K562 cells

Template assembly for the sgRNA sequence targeting *CCDC115* (sequence shown in Table S2), sgRNA synthesis and Cas9-ribonucleoprotein (RNP) preparation were performed as previously described (Lingeman et al., 2017). Cas9-RNP complexes were delivered to cells using the 4D Nucleofector instrument with the SF Cell Line 4D-Nucleofector X Kit (Lonza). 10 µl of Cas9-RNP were mixed with $2.5x10^5$ K562 cells suspended in 20 µl SF solution and 25 µl of the mixture were transferred into the Nucleocuvette well. Nucleofection was performed following the instrument's protocol for K562 cells. Transfected cells were resuspended in culture media and recovered for 48 hours. To assess the cleavage activity of the used sgRNA in the transfected

pool, DNA was extracted using the QuickExtract DNA extraction solution (Epicentre), the cleavage site was amplified by PCR and the cleavage efficiency was determined using the T7 endonuclease I assay (NEB) following the manufacturer's protocol. To isolate individual clones, cells were seeded at a density of 25 cells/ml in ClonaCell-TCS semi solid media (Stem Cell technologies) supplemented with antibiotic/antimycotic solution (Thermo Fisher) and 25 μM ferric citrate for 14 days. Individual colonies were expanded in RPMI media (10% FBS, 1% PS). DNA was extracted from individual clones using the QuickExtract solution and the target region was amplified using Q5 Hot Start High-Fidelity 2X Master Mix (NEB). Clones with biallelic mutations in the target region were identified by *in vitro* digestion with Cas9 nuclease (NEB) following the manufacturer's protocol, and expanded for subsequent analyses. Fragment analysis for both T7 endonuclease and *in vitro* Cas9 digestion assays was performed on 2% agarose gel containing ethidium bromide. Primer sequences for the amplification of the target region in *CCDC115* are shown in (Table S3).

Western blot analysis

Cells were lysed in RIPA buffer (Thermo Fisher) supplemented with protease inhibitor cocktail (Sigma). Proteins in each lysate were quantified using the Pierce BCA protein assay kit (Thermo Fisher). 50 μg of proteins were denatured in Laemmli sample buffer (Biorad) supplemented with 2-mercaptoethanol (Biorad) and run on 4–20% Mini-Protean TGX gels (Biorad). Proteins were transferred to low fluorescence PVDF membranes using the Trans Blot Turbo RTA mini LF transfer kit and the Trans Blot Turbo transfer system (Biorad) following the manufacturer's protocol. Membranes were blocked in Odyssey Blocking Buffer (LICOR) for 1 hour and probed overnight with the primary antibodies diluted as follows: rabbit anti-CCDC115 (Sigma; 1:1000), mouse anti-α-Tubulin (LICOR; 1:1000). Membranes were then washed and incubated for 1 hour with the IRDye 800CW-labeled goat anti-rabbit IgG and IRDye 680RD-labeled goat anti-mouse IgG secondary antibodies (LICOR) at a dilution of 1:15000. Membrane scanning was performed using the Odyssey CLx instrument (LICOR).

Real-time RT qPCR assays

Cells were treated with 30 μM hemin for 48 or 72 hours or with 1mM sodium butyrate (Sigma) for 72 hours. RNA was extracted from control and treated cells using the RNeasy Mini kit (Qiagen). Reverse transcription (rt) was performed from 0.5 μg RNA using the iScript cDNA synthesis kit (Biorad). Real-time qPCR was conducted using the Ssofast Evagreen Supermix (Biorad) on a CFX Connect real-time PCR instrument (Biorad) following the manufacturer's instructions. Data analysis was performed on the accompanying CFX manager software using the *ΔΔCT* method. Ribosomal protein L19 (*RPL19)* gene was used as a reference gene in qPCR assays. Amplification of RPL19, HBA1 and HBG1 cDNA was performed at an annealing temperature of 57°C and sequences of the primers used are listed in (Table S4).

Zinc (II) Mesoporphyrin (ZnMP) uptake assays

ZnMP (Frontier Scientific) was dissolved in DMSO to prepare a 4mM stock solution. ZnMP uptake assays were performed as previously described (O'Callaghan et al., 2010; Rajagopal et al., 2008) with some modifications. K562 cells were incubated with 5 μM ZnMP in uptake medium (25 mM HEPES, pH 7.4, 130 mM NaCl, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MgSO_4$, and 2.5 μM BSA) for 15 minutes. Following incubation, cells were centrifuged at 4°C and pellets were washed once with TrypLE Express cell-dissociation solution (Thermo Fisher) followed by 3 washes with ice-cold uptake medium containing 5% BSA to remove ZnMP bound to cell

surface proteins. Cells were then washed 3 times with ice-cold PBS and fixed with 2% paraformaldehyde. For fluorescence microscopy, cells were either loaded into a Cellometer cell counting chamber (Nexcelom) or transferred onto glass slides using a Cytospin 4 cytocentrifuge (Thermo Fisher). Cellular fluorescence was observed on an Olympus BX51 fluorescence microscope using the U-MWG2 filter set. Flow cytometry was performed on a LSRII cytometer (BD Biosciences) with excitation at 488 nm and emission at 575 nm and the results were analyzed using the accompanying FACSDiva 6.1.3 software.

Results

Heme induces erythroid differentiation in K562 cells

Hemin, the oxidized form of heme, has been previously reported to induce erythroid differentiation in K562 cells (Addya et al., 2004; Nakajima et al., 1997). Since cell differentiation is often associated with a proliferation block, we sought to study the effect of hemin on K562 cell proliferation. Hemin treatment resulted in a concentration-dependent drop in cell proliferation (Figure 3.1A). To verify heme-induced erythroid differentiation, we assessed the mRNA levels of globin alpha (HBA1) and gamma (HBG1) subunits in K562 cells upon hemin induction. Treatment with 30 μM hemin for 48 hours resulted in an increase in HBA1 and HBG1 transcript levels (Figure 3.1B).

Genome-wide CRISPR-Cas9 screening revealed multiple cellular and molecular components of heme trafficking and metabolism

We used CRISPR-Cas9 genome-wide knockout screening to study cellular mechanisms involved in hemin-induced erythroid differentiation of K562 cells. Since the induction of erythroid differentiation by hemin is accompanied by a notable reduction in cell proliferation, we designed a screen to identify genes whose loss-of-function alleviates the hemin-induced proliferation block. Such genes would be potentially involved in early stages of heme-induced erythroid differentiation or in cellular uptake/utilization of exogenous heme. We transduced K562 cells with the GeCKO v2 sgRNA library and screened the pooled cellular library under a selective hemin concentration (30 μM) or equivalent NaOH vehicle for 7 days (Figure 3.2). To assess the differential abundance of each mutant between hemin and vehicle treated pools at the end of the screen, the abundance of each guide sequence, determined by next generation sequencing, was compared between hemin and vehicle treated conditions. Individual guide sequences that exhibited differential abundance with FDR < 0.1 were considered for candidate gene selection (Figure 3.3A). Using this criterion, we identified 79 guide sequences that were enriched and 40 guide sequences that were depleted in hemin-treated pools compared to vehicle controls (Figure 3.3B, C, Table S7). Since each gene is targeted only by 3 guides, a single guide sequence displaying differential abundance between hemin and vehicle pools was sufficient to consider the corresponding gene a candidate. For some candidate genes, at least 2 guide sequences targeting the same gene were significantly enriched or depleted in the hemin pools (highlighted in figure 3.3B, C). An alternative gene ranking approach using MAGeCK revealed 20 candidate genes with FDR < 0.1 (Table S8). 80% of genes identified by MAGeCK analysis were already revealed as candidates from our primary analysis method.

Genes involved in heme uptake and heme-induced erythroid differentiation can be grouped into functional categories

Multiple protein-protein interactions were found within the list of gene products whose loss-of-function resulted in enrichment of the corresponding mutants in hemin treated pools (Figure 3.4A). Gene ontology (GO) analysis combined with literature-based manual curation revealed a role for multiple cellular processes in heme trafficking and metabolism as well as heme-induced erythroid differentiation (Figure 3.4B, Table S9). These processes include endocytosis, vesicular trafficking, endosomal acidification, RNA processing and epigenetic regulation. Notably, genes encoding the hematopoietic regulators RUNX1 and CBFB, the iron regulatory protein IRB2 and the heme biosynthesis component UROD were also identified as potential determinants of heme-induced erythroid differentiation. Among the genes whose loss led to depletion of the

corresponding mutants in the hemin pools are those encoding detoxifying enzymes (HMOX2, CYP26A1 and POR) and members of the ABC transporter family (ABCB10 and ABCF2) (Figure 3.4C).

Loss of the V-ATPase assembly factor CCDC115 leads to defective heme uptake

Results from our screen indicated that loss of endosomal acidification components leads to relative enrichment of the corresponding mutants in hemin compared to vehicle treated pools. The V-ATPase assembly factor CCDC115 was one of the top candidates since all 3 guide sequences targeting the encoding gene showed relatively increased abundance in the hemin pools (Tables S7, S8). To validate the role of CCDC115 in heme-induced erythroid differentiation, we generated *CCDC115*$^{-/-}$ K562 cells using the CRISPR-Cas9 gene knockout tool. Biallelic editing of the targeted region of *CCDC115* in the selected clone was validated by *in vitro* Cas9 digestion (Figure 3.5A). The complete loss of CCDC115 protein in *CCDC115*$^{-/-}$ cells was confirmed by western blot (Figure 3.5B). CCDC115 loss alleviated the differentiation-associated proliferation block induced by hemin (30 μM) and resulted in resistance to a cytotoxic hemin concentration (60 μM) (Figure 3.5C). In addition, *CCDC115*$^{-/-}$ cells exhibited impaired induction of expression of the globin genes *HBA1* and *HBG1* upon hemin treatment (Figure 3.5D, E). However, CCDC115 disruption did not abrogate the induction of *HBA1* by sodium butyrate (Figure 3.5F). To evaluate heme uptake, we treated cells with ZnMP, a fluorescent heme analog, and monitored cellular ZnMP accumulation by fluorescence microscopy. Unlike wild-type cells where the fluorescent signal is dispersed inside the cell, *CCDC115*$^{-/-}$ cells appear to accumulate ZnMP in an intracellular compartment (Figure 3.6A, S3). Cellular ZnMP uptake can be measured by flow cytometry as demonstrated by the time-dependent increase in the mean fluorescence intensities (MFI) upon ZnMP treatment (Figure S4). We did not detect a difference in MFI between wild-type and *CCDC115*$^{-/-}$ cells treated with ZnMP for 15 minutes (Figure 3.6B), which indicates that CCDC115 disruption does not affect the overall cellular ZnMP accumulation but rather impairs subcellular trafficking.

Discussion

Under physiological conditions, transferrin-bound iron is known to be the only iron source indispensable for erythropoiesis (Ponka, 1997). Erythroid cells produce heme through an efficient biosynthetic pathway where the acquired iron is utilized in the last step of heme synthesis (Dailey and Meissner, 2013). In addition to being a critical component of hemoglobin, synthesized heme is involved in several processes in erythroid cells (Fujiwara and Harigae, 2015). Different cell types, including erythroid cells are capable of taking up extracellular heme, but the mechanisms of cellular heme import are poorly understood. Heme uptake by erythroblasts appear to be dispensable for erythropoiesis under physiological conditions but could be required for pathological forms of erythropoiesis (Korolnek and Hamza, 2015). Free heme can result in cytotoxic effects through induction of oxidative stress (Kumar and Bandyopadhyay, 2005) and hence, its levels in erythroid cells and other cell types must be tightly regulated. Only little is known about the mechanisms involved in the delivery, intracellular trafficking, metabolism and export of heme in erythroid cells. Comprehensive identification of cellular components involved in heme trafficking and metabolism expands our understanding of erythropoiesis under normal and stress conditions.

In this study, multiple novel determinants of heme-induced erythroid differentiation and heme cytotoxicity were revealed using a CRISPR-based genome-wide functional screening approach. Consistent with previous reports (Addya et al., 2004; Alves et al., 2011), we demonstrated that heme induces erythroid differentiation in human erythroid K562 cells (Figure 3.1). The reduction in cell viability upon stimulation with 30 μM hemin was probably due to a slower proliferation rate associated with the erythroid lineage commitment. A similar reduction in cell proliferation of K562 cells in response to various stimuli that induce erythroid differentiation was previously reported (Ren et al., 2010; Takagaki et al., 2003; Toobiak et al., 2012). At a higher concentration (60 μM), the cytotoxicity of hemin likely contributes to further decrease in cell viability. Our genetic screen identified multiple candidate genes whose disruption alleviated the heme-induced decline in cell proliferation accompanying erythroid differentiation (Figure 3.3B). A related functional approach was performed in the heme auxotroph worms *Caenorhabditis elegans* and identified important components of heme homeostasis including the heme importer HRG-1 (Severance et al., 2010). However, despite being an excellent model for studying heme homeostasis, multiple *C. elegans* genes have no human homologs. Transcriptomic profiling during heme-induced differentiation of K562 cells revealed a list of heme responsive genes (Addya et al., 2004). We did not find common genes between the list of candidates identified in our screen and the list of the previously identified heme responsive genes.

Among the gene products whose loss alleviated the heme-induced proliferation block, we found several components of the clathrin-mediated endocytosis pathway, factors involved in intracellular vesicle trafficking in addition to multiple V-ATPase subunits and assembly factors (Figure 3.4B). These findings indicate that heme is taken up into erythroid cells through clathrin-mediated endocytosis and that endosomal acidification is required for efficient heme utilization. A role for receptor-mediated endocytosis in cellular heme uptake has been proposed by many studies [reviewed in (West and Oates, 2008)]. Heme binding proteins with unknown identity have been characterized on the surface of mammalian intestinal epithelial cells (Gräsbeck et al., 1979) and erythroleukemic cells (Galbraith et al., 1985; Majuri and Gräsbeck, 1987; Majuri R., 2009). Other reports demonstrated that the uptake of exogenous heme is temperature-dependent

(Worthington et al., 2001), requires ATP (Vaghefi et al., 2000) and involves trafficking to subcellular compartments (Yuan et al., 2016). Collectively, these studies support the receptor-mediated endocytosis model of heme uptake. The only discovered endocytic pathway for heme uptake involves the heme-binding protein hemopexin. The heme-hemopexin complex binds to the cell surface receptor LRP/CD91 and is internalized by receptor-mediated endocytosis (Hvidberg et al., 2005). K562 erythroleukemic cells have been shown to take up heme-hemopexin complexes in a temperature-dependent manner (Taketani et al., 1986). Accordingly, it is possible that heme uptake in K562 cells is mediated by the LRP/CD91 receptor. If exogenous heme is internalized by endocytosis, release of free intact heme in the endosome and its transport to the cytosol are required for cellular heme utilization. Our screen revealed multiple factors involved in endosomal acidification as determinants of heme uptake (Figure 3.4B). These results suggest that endosomal acidification following endocytosis could facilitate the release of free heme in the endosomal lumen. The heme transporter HRG-1 was shown to be responsible for the efflux of free heme from the phagolysosome of macrophages following erythrophagocytosis (White et al., 2013). The expression of HRG-1 in early erythroid cells (Korolnek and Hamza, 2015) and its localization to endosomes (Rajagopal et al., 2008) indicate that it could be involved in heme transport from the endosome to the cytosol in erythroid cells. Interestingly, HRG-1-mediated heme transport was shown to be dependent on endosomal acidification generated by V-ATPase activity (O'Callaghan et al., 2010). Neither the heme-hemopexin receptor LRP nor the heme transporter HRG-1 were identified in our screen as heme uptake candidates. One possible explanation could be poor on-target activity of all 3 sgRNAs in the library targeting each of LRP and HRG-1. Another plausible reason could be the presence of alternative compensatory components in K562 cells contributing to the process thus rendering these components dispensable for heme uptake. Further investigation is required to assess the role of LRP and HRG-1 in exogenous heme uptake by erythroid cells.

Components of cellular regulatory processes that are generally involved in cell differentiation were identified in our screen as determinants of heme-induced erythroid differentiation. The role of RNA processing, mainly mRNA splicing, in late erythropoiesis has been demonstrated in many studies (Heinicke et al., 2013; Loflin et al., 1999; Pimentel et al., 2014; Wojchowski, 2014; Yamamoto et al., 2009; Yang et al., 2005). Our screen identified several RNA processing factors whose loss potentially impairs heme-induced erythroid differentiation (Figure 3.4B). Intriguingly, one of the revealed RNA processing components, HNRNPD, which binds to and destabilizes AU-rich RNA-destabilizing element (ARE)-containing transcripts, has been shown to sequester into a heme-induced protein complex, and was suggested to be involved in stabilization of ARE-containing mRNAs during hemin-induced erythroid differentiation of K562 cells (Loflin et al., 1999). Multiple epigenetic regulatory mechanisms have been shown to play a role in erythroid differentiation (DeVilbiss et al., 2015; Wong et al., 2011; Wu et al., 2011; Yang et al.). We identified several epigenetic factors with a potential role in heme-induced erythroid differentiation (Figure 3.4B). Two of the identified components (SMARCD2 and ARID1A) are parts of the SWI/SNF chromatin remodeling complex. In support of this finding, a role of another component of the SWI/SNF complex, Brg1, in erythroid development has been reported (Bultman et al., 2005). In addition, two histone methyl transferases, KMT2B and CARM1, with previously documented roles in erythroid differentiation (Demers et al., 2007; Vu et al., 2013) were also revealed in our screen as potential regulators of heme-induced erythroid differentiation.

We identified two members of the core binding factor (CBF) transcription factor family, RUNX1 and CBFB, as determinants of heme-induced differentiation of K562 cells. Interestingly, RUNX1 and CBFB were shown to be indispensable for the differentiation of hematopoietic cells as both RUNX1 and CBFB deficient mice die *in utero* due to similar defective hematopoietic phenotypes (Ichikawa et al., 2013). CBFB binds to RUNX1 and stabilizes its binding to DNA (Yan et al., 2004). In opposition to our finding, RUNX1 was shown to repress the erythroid gene expression program and its expression is lost during erythroid maturation (Kuvardina et al., 2015; North et al., 2004). It is important to mention that our screen is designed to reveal components that are only involved in early stages of erythroid differentiation resulting in the differentiation-associated proliferation block. Hence, it is possible that the RUNX-1 is required for early stages of heme-induced erythroid differentiation that occur prior to the induction of expression of the erythroid-specific genes. Regulators of iron and heme homeostasis are key factors in erythroid differentiation. The iron regulatory protein IREB2 was identified in our screen as a candidate determinant for heme-induced erythroid differentiation. Notably, heme was shown to bind to and oxidize IREB2 resulting in its ubiquitination and proteasomal degradation (Ishikawa et al., 2005; Yamanaka et al., 2003). Accordingly, it can be speculated that the iron homeostatic function of IREB2 is a pre-requirement for heme-induced erythroid differentiation but it is dispensable during the differentiation process. We also identified a potential role for a component of the heme biosynthesis pathway, UROD, in heme-induced differentiation of K562 cells. Despite being an important erythroid factor due to its role in heme production (Dailey and Meissner, 2013), it is unclear how UROD could contribute to erythroid differentiation stimulated by exogenous heme.

Despite the importance of heme for erythropoiesis, heme accumulation in erythroid precursors can result in oxidative damage and induce apoptosis. Accordingly, cellular mechanisms that regulate intracellular levels of free heme or reduce its cytotoxicity are required in erythroid cells. We identified genes whose disruption sensitized K562 cells to hemin and resulted in depletion of the corresponding mutants from the hemin treated pools (Figure 3.3C). Few of the revealed genes encode detoxifying enzymes such as the heme catabolizing enzyme HMOX2, cytochrome P450 26A1 (CYP26A1) and the cytochrome P450 oxidoreductase (POR). It can be suggested that these enzymes protect erythroblasts from the toxicity of accumulating heme during erythroid differentiation and hence are critical for erythropoiesis. Intriguingly, HMOX2 expression in mouse erythroid precursors was shown to be predominant in the proerythroblast stage, where globin production is minimal, but it is lost during erythroid differentiation when globin protein levels are adequate (Doty et al., 2015). This finding further supports the role of HMOX2 in heme detoxification in early stages of erythroid differentiation. Other cellular components that we identified as determinants of heme resistance include the ATP-binding cassette (ABC) family members ABCB10 and ABCF2. ABCB10 is a mitochondrial transporter that is essential for erythropoiesis (Hyde et al., 2012; Tang Leilei et al., 2011). Although heme and its precursor ALA are candidate substrates of ABCB10, attempts to identify the true substrate identity generated conflicting results (Bayeva et al., 2013; Qiu et al., 2015; Seguin et al., 2017; Yamamoto et al., 2014). Hemin treatment of erythroid cells was shown to stimulate heme biosynthesis in the mitochondria (Granick and Sassa, 1978). Impaired mitochondrial heme efflux could result in mitochondrial dysfunction. Accordingly, a potential role of ABCB10 in mitochondrial heme export is a plausible explanation of the increased sensitivity to hemin upon ABCB10 disruption. The other identified ABC family member, ABCF2, is a less studied protein

with no associated transmembrane domains suggesting that it is not involved in transporting molecules across cell membranes (Kerr, 2004). A role of ABCF2 in erythropoiesis has not been reported. Interestingly, ABCF2 expression was shown to be regulated by the antioxidant transcription factor Nrf2 and it plays a critical role in chemoresistance (Bao et al., 2017). Therefore, a role of ABCF2 in reversing heme-induced oxidative damage can be proposed and could explain its role as a determinant of heme resistance. Proteins with a previously demonstrated role in mitochondrial and cytoplasmic heme export including FLVCR1 (Chiabrando et al., 2012; Quigley et al., 2004) and ABCG2 (Desuzinges-Mandon et al., 2010) were not identified in our screen as heme resistance determinants.

We were intrigued by the potential role of the endosomal acidification component, CCDC115, in heme uptake for several reasons: 1) *CCDC115* was one of the top-ranked candidate genes identified by our screen where all 3 sgRNAs targeting this gene resulted in an enrichment phenotype in hemin-treated pools; 2) CCDC115 is a V-ATPase assembly factor that is required for endosomal acidification (Miles et al., 2017), which we proposed as an essential process for erythroid heme uptake; 3) We have previously identified a role of CCDC115 in transferrin iron uptake in erythroid cells (chapter 2); 4) A role for CCDC115 in cellular heme uptake has not been previously reported. CCDC115 disruption in K562 cells alleviated hemin-induced proliferation block and conferred resistance to hemin toxicity (Figure 3.5C). *CCDC115*$^{-/-}$ K562 cells exhibited impaired induction of globin genes by hemin (Figure 3.5D, E) suggesting a defective differentiation toward the erythroid lineage. However, induction of globin genes by sodium butyrate was not abrogated upon CCDC115 disruption (Figure 3.5F), indicating that the defect in *CCDC115*$^{-/-}$ cells is at the level of utilization of exogenous heme. Microscopic analysis showed different intracellular trafficking profiles of the fluorescent heme analog ZnMP between wild-type and *CCDC115*$^{-/-}$ K562 cells (Figure 3.6A). While wild-type cells efficiently acquire ZnMP and spread it homogenously, *CCDC115*$^{-/-}$ cells seem to accumulate ZnMP in a subcellular compartment. We did not observe any difference in the overall cellular ZnMP uptake between wild-type and *CCDC115*$^{-/-}$ cells by flow cytometry (Figure 3.6B). Together, these results indicate that CCDC115 disruption does not affect internalization of extracellular heme, but rather impairs its subcellular trafficking following internalization. Consistent with an endocytosis-based model for heme uptake, and the role of CCDC115 in endosomal acidification, it can be speculated that in *CCDC115*$^{-/-}$ cells, ZnMP is trapped in an endosomal/endolysosomal compartment following internalization. Co-localization studies are further required to reveal the identity of the ZnMP accumulation region in these cells.

Our study provided novel insight into mechanisms of heme trafficking in erythroid cells as well as heme-induced erythroid differentiation. Indeed, comprehensive mechanistic studies *in vitro* and *in vivo* are further required to validate our findings. Yet, based on our results and other findings from literature, the following can be proposed: Erythroid cells take up heme from the surrounding environment via receptor-mediated endocytosis. Internalized heme is trafficked to endolysosomes where the increased acidity is required for its transport to the cytosol via a specialized transporter. An acidic microenvironment in endolysosomes could contribute to heme efflux by facilitating the dissociation of free heme from a bound receptor. Alternatively, a low luminal PH could be required for the function of the endolysosomal heme transporter. Acquired heme induces multiple signals that stimulate erythroid differentiation, in addition to its incorporation into globin proteins to form hemoglobin. Several regulatory and homeostatic cellular processes contribute to heme-induced erythroid differentiation. Excess heme is broken

down by heme oxygenase or exported out of the cell via heme exporter proteins. Cellular antioxidant mechanisms mitigate heme cytotoxicity during erythroid differentiation.

Figures

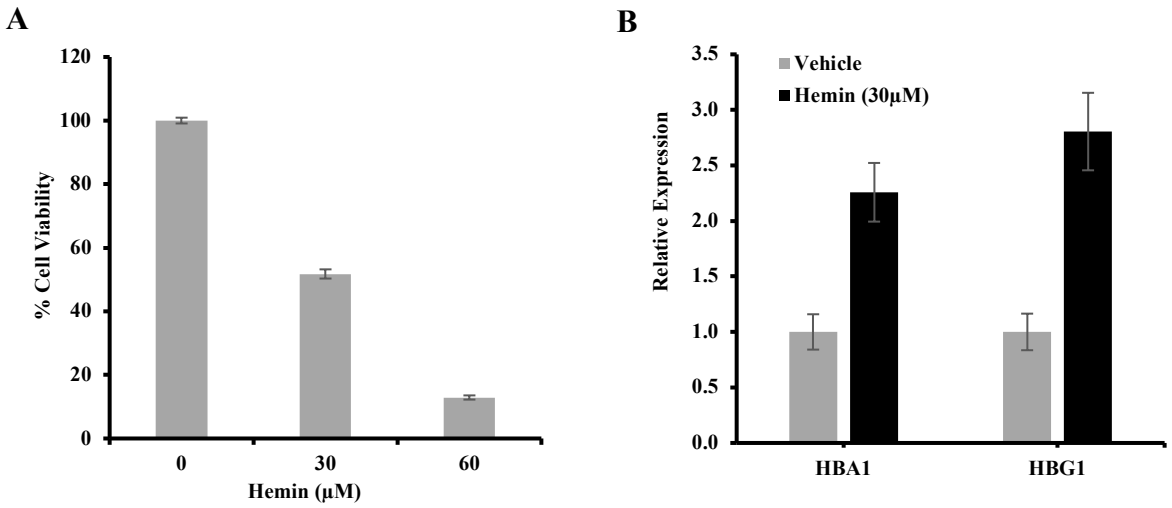

Figure 3.1: Hemin-induced differentiation of K562 erythroid cells. (A) Cells were treated with different concentrations of hemin for 72 hours and cell viability was evaluated by CellTiter Glo assays. Bars represent cell viability as percentage of vehicle-treated controls. Data are represented as mean ± standard deviation (n=3). (B). Cells were treated with 30 µM hemin or equivalent vehicle for 48 hours and the expression of globin genes (*HBA1* and *HBG1*) was measured by real-time qPCR. Bars represent expression of HBA1 or HBG1 relative to vehicle-treated controls. Data are represented as mean ± standard error (n=3).

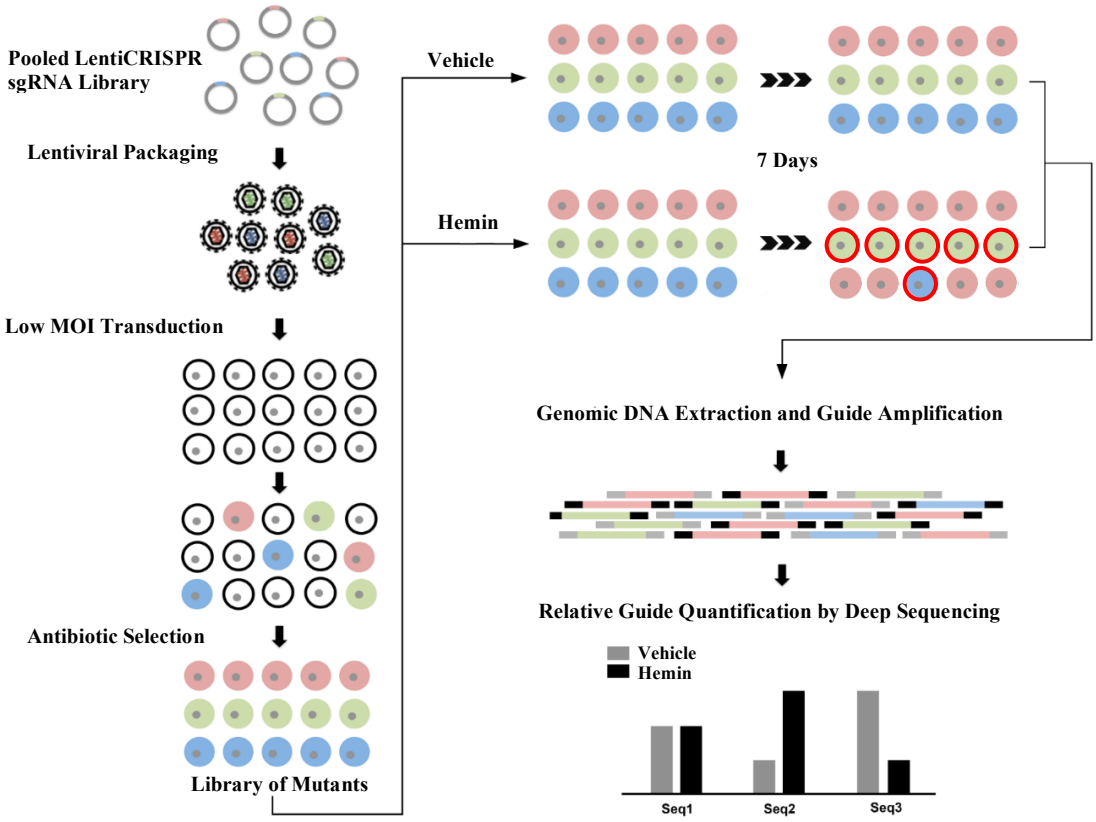

Figure 3.2: Workflow of the genome-wide CRISPR-Cas9 screen designed to study heme-induced erythroid differentiation. K562 cells are transduced with the Cas9-sgRNA library at a low MOI to generate a library of mutants. Screening is performed in media containing 30 μM hemin or equivalent vehicle to identify mutants that escape the differentiation-associated proliferation block (relatively enriched in hemin-treated pools) or those that have increased sensitivity to hemin toxicity (relatively depleted in hemin-treated pools). Red rims are used to represent differentiated cells in the pool. Guide sequences encoding all sgRNAs are PCR-amplified and quantified by deep sequencing. Each guide sequence is used as a barcode to label the corresponding mutant. Abundance of each guide sequence is compared between vehicle and hemin-treated pools.

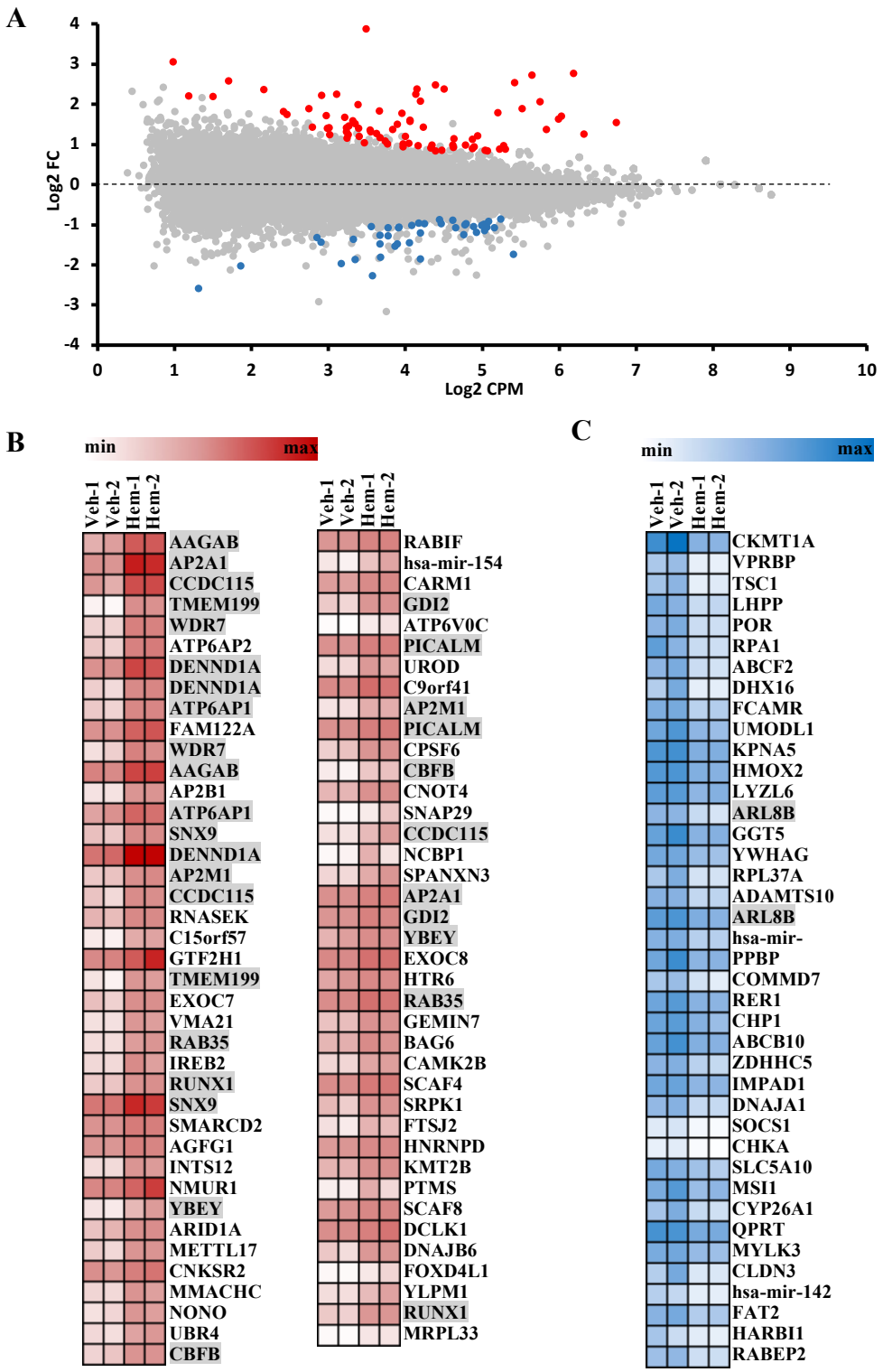

Figure 3.3: Identification of candidate genes involved in heme trafficking and metabolism. (A) Scatter plot showing differential abundance of each guide sequence between vehicle and hemin treated pools. Fold changes (FC) are calculated by dividing the average normalized counts (n=2) for each guide in hemin pools to that in vehicle pools. Log2 FCs are plotted against the

average abundance of each guide in all pools represented as Log2 counts per million (CPM). Guide sequences depleted in hemin relative to vehicle (Log2 FC < 0) with FDR <0.1 are represented as blue dots. Guide sequences enriched in hemin relative to vehicle (Log2 FC > 0) with FDR < 0.1 are represented as red dots. (B, C) Heatmaps of the normalized counts of gene-specific guide sequences that are enriched (B) or depleted (C) in hemin relative to vehicle pools with FDR < 0.1. Genes with at least 2 guide sequences exhibiting differential abundance are highlighted. The screen was run in duplicate and values for each vehicle (Veh) and hemin (Hem) replicate are represented.

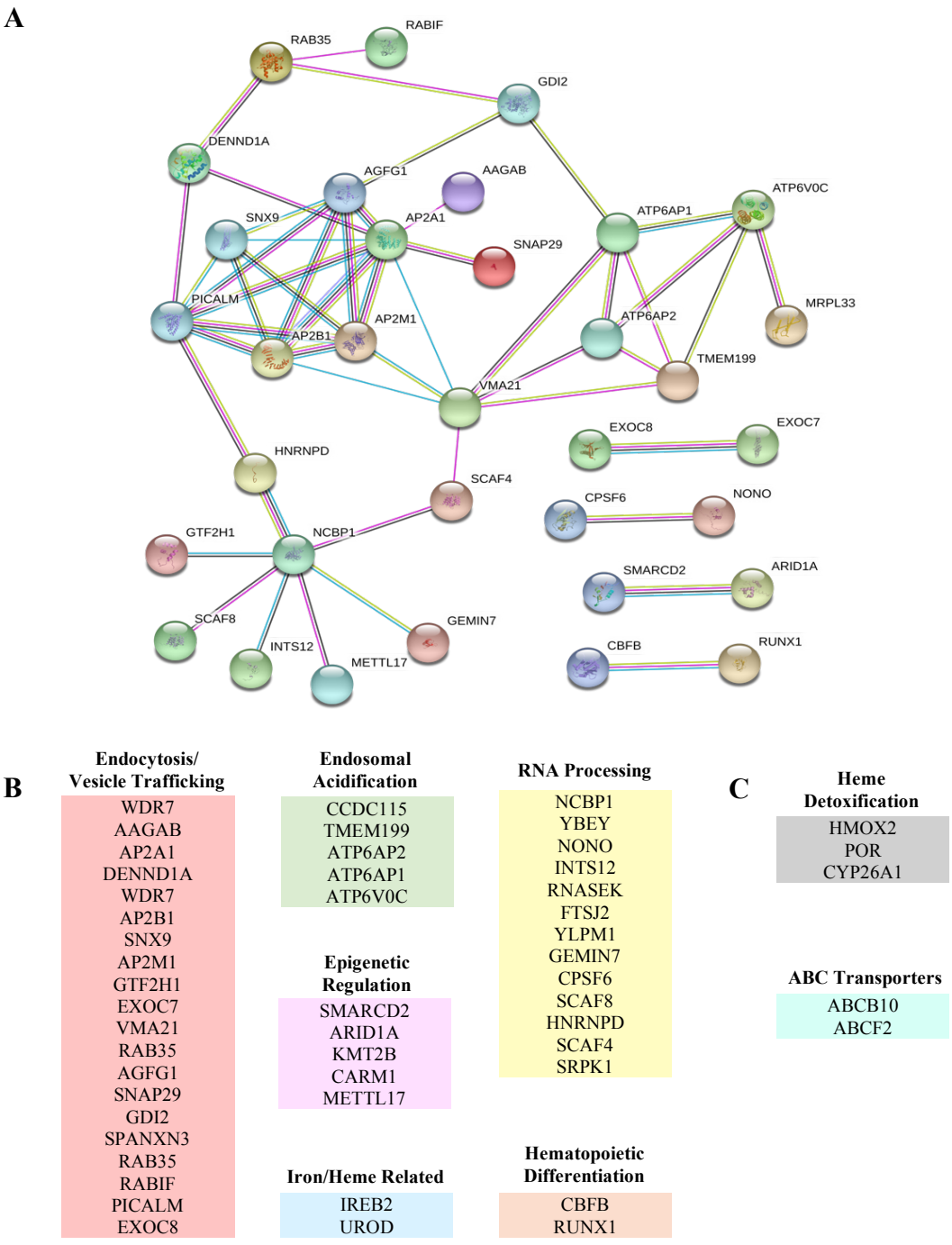

Figure 3.4: Interaction and functional classification of candidate genes from the hemin screen. (A) Interaction network of enrichment candidates (genes whose loss results in relative enrichment of the corresponding cells in hemin pools) integrated using the STRING protein-protein interaction database. (B) Functional classification of enrichment candidates. (C) Functional classification of depletion candidates (genes whose loss results in relative depletion of the corresponding cells in hemin pools). Genes were classified based on information from Gene Ontology (GO) enrichment analysis combined with manual literature-based curation.

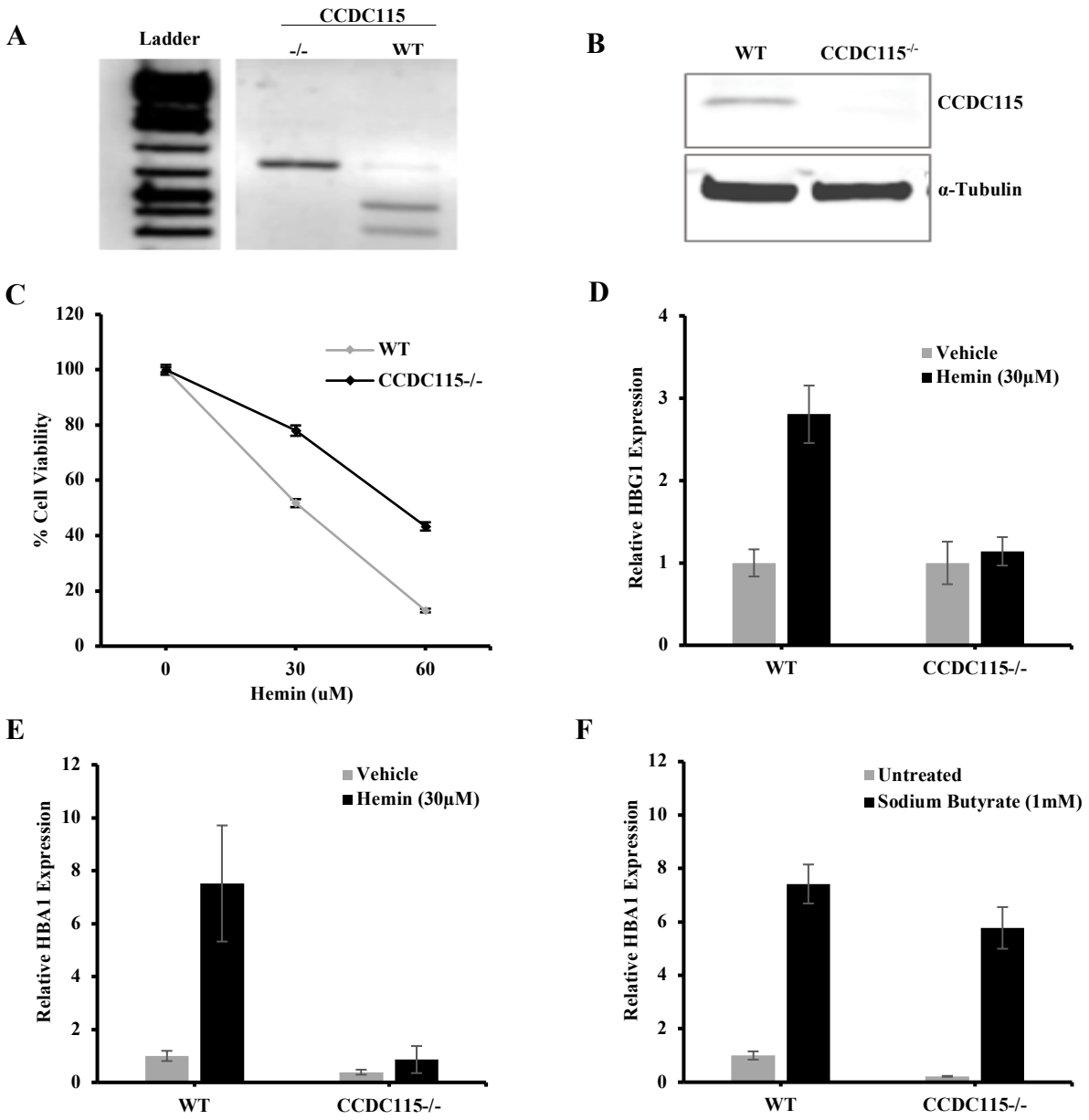

Figure 3.5: Effect of CCDC115 loss on hemin cytotoxicity and heme-induced expression of globin genes. (A) CRISPR-Cas9-induced biallelic editing of the *CCDC115* gene in the used K562 clone validated by *in vitro* Cas9 cleavage assay. (B) Validation of the complete loss of CCDC115 protein in *CCDC115*$^{-/-}$ cells by western blot. (C) Resistance of *CCDC115*$^{-/-}$ cells to hemin toxicity. Cells were treated with different concentrations of hemin for 72 hours and cell viability was evaluated by CellTiter Glo assay. Cell viability is represented as percentage of vehicle-treated controls. Data are represented as mean ± standard deviation (n=3). (D-F) Evaluation of the expression globin genes (*HBA1* and *HBG1*) by real-time qPCR. Cells were treated with 30 μM hemin or equivalent vehicle for 48 hours (D) or 72 hours (E) or 1mM sodium butyrate for 72 hours (F). Bars represent expression of HBA1 or HBG1 relative to controls. Data are represented as mean ± standard error (n=3).

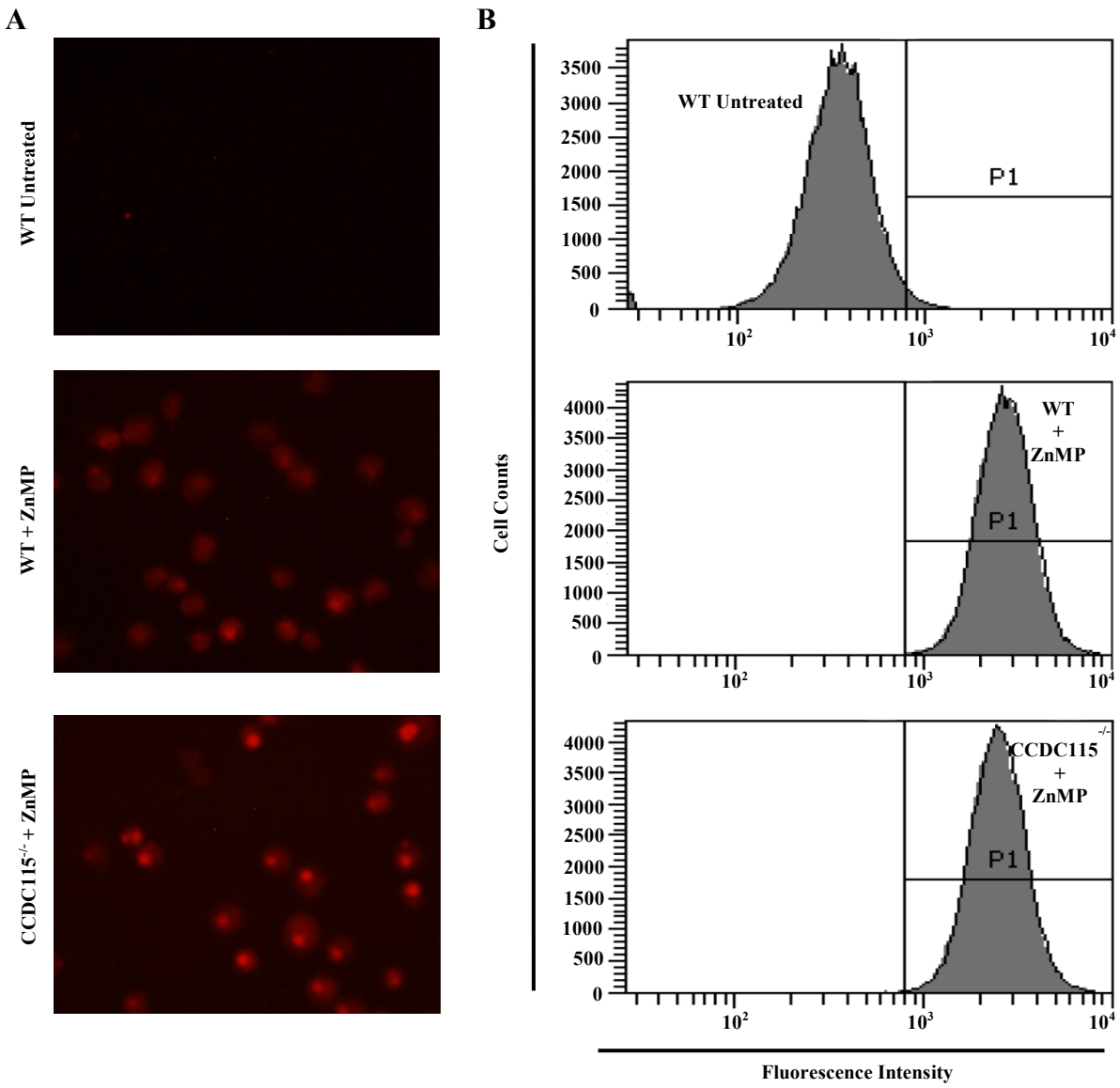

Figure 3.6: Zinc mesoporphyrin (ZnMP) uptake in WT and CCDC115$^{-/-}$ K562 cells. Cells were treated with the fluorescent heme analog ZnMP (5μM) for 15 minutes and excess ZnMP was removed by washing. (A) Evaluation of cellular ZnMP accumulation by fluorescence microscopy. Images were taken at 40X magnification. (B) Quantification of ZnMP uptake by flow cytometry. The gate (P1) is determined from wild-type cells treated with ZnMP for 15 minutes.

References

Addya, S., Keller, M.A., Delgrosso, K., Ponte, C.M., Vadigepalli, R., Gonye, G.E., and Surrey, S. (2004). Erythroid-induced commitment of K562 cells results in clusters of differentially expressed genes enriched for specific transcription regulatory elements. Physiol. Genomics *19*, 117–130.

Alves, L.R., Costa, E.S., Sorgine, M.H.F., Nascimento-Silva, M.C.L., Teodosio, C., Bárcena, P., Castro-Faria-Neto, H.C., Bozza, P.T., Orfao, A., Oliveira, P.L., et al. (2011). Heme-Oxygenases during Erythropoiesis in K562 and Human Bone Marrow Cells. PLOS ONE *6*, e21358.

Bao, L., Wu, J., Dodson, M., Vega, E.M.R. de la, Ning, Y., Zhang, Z., Yao, M., Zhang, D.D., Xu, C., and Yi, X. (2017). ABCF2, an Nrf2 target gene, contributes to cisplatin resistance in ovarian cancer cells. Mol. Carcinog. *56*, 1543–1553.

Bayeva, M., Khechaduri, A., Wu, R., Burke, M.A., Wasserstrom, J.A., Singh, N., Liesa, M., Shirihai, O.S., Langer, N.B., Paw, B.H., et al. (2013). ATP-Binding Cassette B10 Regulates Early Steps of Heme SynthesisNovelty and Significance. Circ. Res. *113*, 279–287.

Bultman, S.J., Gebuhr, T.C., and Magnuson, T. (2005). A Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in β-globin expression and erythroid development. Genes Dev. *19*, 2849–2861.

Chiabrando, D., Marro, S., Mercurio, S., Giorgi, C., Petrillo, S., Vinchi, F., Fiorito, V., Fagoonee, S., Camporeale, A., Turco, E., et al. (2012). The mitochondrial heme exporter FLVCR1b mediates erythroid differentiation. J. Clin. Invest. *122*, 4569–4579.

Chiabrando, D., Mercurio, S., and Tolosano, E. (2014). Heme and erythropoieis: more than a structural role. Haematologica *99*, 973–983.

Dailey, H.A., and Meissner, P.N. (2013). Erythroid Heme Biosynthesis and Its Disorders. Cold Spring Harb. Perspect. Med. *3*.

Demers, C., Chaturvedi, C.-P., Ranish, J.A., Juban, G., Lai, P., Morle, F., Aebersold, R., Dilworth, F.J., Groudine, M., and Brand, M. (2007). Activator-mediated Recruitment of the MLL2 Methyltransferase Complex to the β-globin Locus. Mol. Cell *27*, 573–584.

Desuzinges-Mandon, E., Arnaud, O., Martinez, L., Huché, F., Pietro, A.D., and Falson, P. (2010). ABCG2 Transports and Transfers Heme to Albumin through Its Large Extracellular Loop. J. Biol. Chem. *285*, 33123–33133.

DeVilbiss, A.W., Sanalkumar, R., Hall, B.D.R., Katsumura, K.R., Andrade, I.F. de, and Bresnick, E.H. (2015). Epigenetic Determinants of Erythropoiesis: Role of the Histone Methyltransferase SetD8 in Promoting Erythroid Cell Maturation and Survival. Mol. Cell. Biol. *35*, 2073–2087.

Doty, R.T., Phelps, S.R., Shadle, C., Sanchez-Bonilla, M., Keel, S.B., and Abkowitz, J.L. (2015). Coordinate expression of heme and globin is essential for effective erythropoiesis. J. Clin. Invest. *125*, 4681–4691.

Fujiwara, T., and Harigae, H. (2015). Biology of Heme in Mammalian Erythroid Cells and Related Disorders. BioMed Res. Int. *2015*.

Galbraith, R.A., Sassa, S., and Kappas, A. (1985). Heme binding to murine erythroleukemia cells. Evidence for a heme receptor. J. Biol. Chem. *260*, 12198–12202.

Granick, J.L., and Sassa, S. (1978). Hemin control of heme biosynthesis in mouse Friend virus-transformed erythroleukemia cells in culture. J. Biol. Chem. *253*, 5402–5406.

Gräsbeck, R., Kouvonen, I., Lundberg, M., and Tenhunen, R. (1979). An intestinal receptor for heme. Scand. J. Haematol. *23*, 5–9.

Gunsalus, I.C., Sligar, S.G., Nordlund, T., and Frauenfelder, H. (1977). Oxygen sensing heme proteins: monoxygenases, myoglobin and hemoglobin. Adv. Exp. Med. Biol. *78*, 37–50.

Han, A.-P., Yu, C., Lu, L., Fujiwara, Y., Browne, C., Chin, G., Fleming, M., Leboulch, P., Orkin, S.H., and Chen, J.-J. (2001). Heme-regulated eIF2α kinase (HRI) is required for translational regulation and survival of erythroid precursors in iron deficiency. EMBO J. *20*, 6909–6918.

Heinicke, L.A., Nabet, B., Shen, S., Jiang, P., Zalen, S. van, Cieply, B., Russell, J.E., Xing, Y., and Carstens, R.P. (2013). The RNA Binding Protein RBM38 (RNPC1) Regulates Splicing during Late Erythroid Differentiation. PLOS ONE *8*, e78031.

Hvidberg, V., Maniecki, M.B., Jacobsen, C., Højrup, P., Møller, H.J., and Moestrup, S.K. (2005). Identification of the receptor scavenging hemopexin-heme complexes. Blood *106*, 2572–2579.

Hyde, B.B., Liesa, M., Elorza, A.A., Qiu, W., Haigh, S.E., Richey, L., Mikkola, H.K., Schlaeger, T.M., and Shirihai, O.S. (2012). The mitochondrial transporter ABC-me (ABCB10), a downstream target of GATA-1, is essential for erythropoiesis *in vivo*. Cell Death Differ. *19*, 1117–1126.

Ichikawa, M., Yoshimi, A., Nakagawa, M., Nishimoto, N., Watanabe-Okochi, N., and Kurokawa, M. (2013). A role for RUNX1 in hematopoiesis and myeloid leukemia. Int. J. Hematol. *97*, 726–734.

Igarashi, K., and Sun, J. (2006). The Heme-Bach1 Pathway in the Regulation of Oxidative Stress Response and Erythroid Differentiation. Antioxid. Redox Signal. *8*, 107–118.

Igarashi, J., Murase, M., Iizuka, A., Pichierri, F., Martinkova, M., and Shimizu, T. (2008). Elucidation of the Heme Binding Site of Heme-regulated Eukaryotic Initiation Factor 2α Kinase

and the Role of the Regulatory Motif in Heme Sensing by Spectroscopic and Catalytic Studies of Mutant Proteins. J. Biol. Chem. *283*, 18782–18791.

Ishikawa, H., Kato, M., Hori, H., Ishimori, K., Kirisako, T., Tokunaga, F., and Iwai, K. (2005). Involvement of Heme Regulatory Motif in Heme-Mediated Ubiquitination and Degradation of IRP2. Mol. Cell *19*, 171–181.

Kerr, I.D. (2004). Sequence analysis of twin ATP binding cassette proteins involved in translational control, antibiotic resistance, and ribonuclease L inhibition. Biochem. Biophys. Res. Commun. *315*, 166–173.

Kolluri, S., Sadlon, T.J., May, B.K., and Bonkovsky, H.L. (2005). Haem repression of the housekeeping 5-aminolaevulinic acid synthase gene in the hepatoma cell line LMH. Biochem. J. *392*, 173–180.

Korolnek, T., and Hamza, I. (2015). Macrophages and iron trafficking at the birth and death of red cells. Blood *125*, 2893–2897.

Kumar, S., and Bandyopadhyay, U. (2005). Free heme toxicity and its detoxification systems in human. Toxicol. Lett. *157*, 175–188.

Kuvardina, O.N., Herglotz, J., Kolodziej, S., Kohrs, N., Herkt, S., Wojcik, B., Oellerich, T., Corso, J., Behrens, K., Kumar, A., et al. (2015). RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation. Blood *125*, 3570–3579.

Lingeman, E., Jeans, C., and Corn, J.E. (2017). Production of Purified CasRNPs for Efficacious Genome Editing. Curr. Protoc. Mol. Biol. *120*, 31.10.1-31.10.19.

Loflin, P., Chen, C.-Y.A., and Shyu, A.-B. (1999). Unraveling a cytoplasmic role for hnRNP D in the in vivo mRNA destabilization directed by the AU-rich element. Genes Dev. *13*, 1884–1897.

Majuri, R., and Gräsbeck, R. (1987). A rosette receptor assay with haem-microbeads. Demonstration of a haem receptor on K562 cells. Eur. J. Haematol. *38*, 21–25.

Majuri R. (2009). Heme-binding plasma membrane proteins of K562 erythroleukemia cells: Adsorption to heme-microbeads, isolation with affinity chromatography. Eur. J. Haematol. *43*, 220–225.

Miles, A.L., Burr, S.P., Grice, G.L., and Nathan, J.A. (2017). The vacuolar-ATPase complex and assembly factors, TMEM199 and CCDC115, control HIF1α prolyl hydroxylation by regulating cellular iron levels. ELife *6*, e22693.

Nakajima, O., Iwasaki, S., and Hashimoto, Y. (1997). Hemin-induced erythroid differentiation of human myeloleukemia K562 cell line and its modification by bioresponse modifiers. Cell. Mol. Biol. Noisy--Gd. Fr. *43*, 115–134.

North, T.E., Stacy, T., Matheny, C.J., Speck, N.A., and Bruijn, M.F.T.R. de (2004). Runx1 Is Expressed in Adult Mouse Hematopoietic Stem Cells and Differentiating Myeloid and Lymphoid Cells, But Not in Maturing Erythroid Cells. STEM CELLS *22*, 158–168.

O'Callaghan, K.M., Ayllon, V., O'Keeffe, J., Wang, Y., Cox, O.T., Loughran, G., Forgac, M., and O'Connor, R. (2010). Heme-binding Protein HRG-1 Is Induced by Insulin-like Growth Factor I and Associates with the Vacuolar H+-ATPase to Control Endosomal pH and Receptor Trafficking. J. Biol. Chem. *285*, 381–391.

Pimentel, H., Parra, M., Gee, S., Ghanem, D., An, X., Li, J., Mohandas, N., Pachter, L., and Conboy, J.G. (2014). A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis. Nucleic Acids Res. *42*, 4031–4042.

Ponka, P. (1997). Tissue-Specific Regulation of Iron Metabolism and Heme Synthesis: Distinct Control Mechanisms in Erythroid Cells. Blood *89*, 1–25.

Qiu, W., Liesa, M., Carpenter, E.P., and Shirihai, O.S. (2015). ATP Binding and Hydrolysis Properties of ABCB10 and Their Regulation by Glutathione. PLOS ONE *10*, e0129772.

Quigley, J.G., Yang, Z., Worthington, M.T., Phillips, J.D., Sabo, K.M., Sabath, D.E., Berg, C.L., Sassa, S., Wood, B.L., and Abkowitz, J.L. (2004). Identification of a Human Heme Exporter that Is Essential for Erythropoiesis. Cell *118*, 757–766.

Rajagopal, A., Rao, A.U., Amigo, J., Tian, M., Upadhyay, S.K., Hall, C., Uhm, S., Mathew, M.K., Fleming, M.D., Paw, B.H., et al. (2008). Haem homeostasis is regulated by the conserved and concerted functions of HRG-1 proteins. Nature *453*, 1127–1131.

Reedy, C.J., and Gibney, B.R. (2004). Heme Protein Assemblies. Chem. Rev. *104*, 617–650.

Ren, J.-G., Seth, P., Everett, P., Clish, C.B., and Sukhatme, V.P. (2010). Induction of Erythroid Differentiation in Human Erythroleukemia Cells by Depletion of Malic Enzyme 2. PLOS ONE *5*, e12520.

Schultz, I.J., Chen, C., Paw, B.H., and Hamza, I. (2010). Iron and Porphyrin Trafficking in Heme Biogenesis. J. Biol. Chem. *285*, 26753–26759.

Seguin, A., Takahashi-Makise, N., Yien, Y.Y., Huston, N.C., Whitman, J.C., Musso, G., Wallace, J.A., Bradley, T., Bergonia, H.A., Kafina, M.D., et al. (2017). Reductions in the mitochondrial ABC transporter Abcb10 affect the transcriptional profile of heme biosynthesis genes. J. Biol. Chem. *292*, 16284–16299.

Severance, S., Rajagopal, A., Rao, A.U., Cerqueira, G.C., Mitreva, M., El-Sayed, N.M., Krause, M., and Hamza, I. (2010). Genome-Wide Analysis Reveals Novel Genes Essential for Heme Homeostasis in Caenorhabditis elegans. PLoS Genet. *6*.

Shayeghi, M., Latunde-Dada, G.O., Oakhill, J.S., Laftah, A.H., Takeuchi, K., Halliday, N., Khan, Y., Warley, A., McCann, F.E., Hider, R.C., et al. (2005). Identification of an Intestinal Heme Transporter. Cell *122*, 789–801.

Sun, J., Hoshino, H., Takaku, K., Nakajima, O., Muto, A., Suzuki, H., Tashiro, S., Takahashi, S., Shibahara, S., Alam, J., et al. (2002). Hemoprotein Bach1 regulates enhancer availability of heme oxygenase-1 gene. EMBO J. *21*, 5216–5224.

Tahara, T., Sun, J., Nakanishi, K., Yamamoto, M., Mori, H., Saito, T., Fujita, H., Igarashi, K., and Taketani, S. (2004). Heme positively regulates the expression of beta-globin at the locus control region via the transcriptional factor Bach1 in erythroid cells. J. Biol. Chem. *279*, 5480–5487.

Takagaki, K., Katsuma, S., Horio, T., Kaminishi, Y., Hada, Y., Tanaka, T., Ohgi, T., and Yano, J. (2003). cDNA microarray analysis of altered gene expression in Ara-C-treated leukemia cells. Biochem. Biophys. Res. Commun. *309*, 351–358.

Taketani, S., Kohno, H., and Tokunaga, R. (1986). Receptor-mediated heme uptake from hemopexin by human erythroleukemia K562 cells. Biochem. Int. *13*, 307–312.

Tang Leilei, Bergevoet Saskia M., Bakker-Verweij Greet, Harteveld Cornelis L., Giordano Piero C., Nijtmans Leo, de Witte Theo, Jansen Joop H., Raymakers Reinier A. P., and van der Reijden Bert A. (2011). Human mitochondrial ATP-binding cassette transporter ABCB10 is required for efficient red blood cell development. Br. J. Haematol. *157*, 151–154.

Toobiak, S., Shaklai, M., and Shaklai, N. (2012). Carbon Monoxide Induced Erythroid Differentiation of K562 Cells Mimics the Central Macrophage Milieu in Erythroblastic Islands. PLOS ONE *7*, e33940.

Vaghefi, N., Guillochon, D., Bureau, F., Neuville, D., Lebrun, F., Arhan, P., and Bouglé, D. (2000). The effect of cysteine and 2,4-dinitrophenol on heme and nonheme absorption in a rat intestinal model. J. Nutr. Biochem. *11*, 562–567.

Vu, L.P., Perna, F., Wang, L., Voza, F., Figueroa, M.E., Tempst, P., Erdjument-Bromage, H., Gao, R., Chen, S., Paietta, E., et al. (2013). PRMT4 blocks myeloid differentiation by assembling a methyl-RUNX1-dependent repressor complex. Cell Rep. *5*, 1625–1638.

West, A.R., and Oates, P.S. (2008). Mechanisms of heme iron absorption: Current questions and controversies. World J. Gastroenterol. WJG *14*, 4101–4110.

White, C., Yuan, X., Schmidt, P.J., Bresciani, E., Samuel, T.K., Campagna, D., Hall, C., Bishop, K., Calicchio, M.L., Lapierre, A., et al. (2013). HRG1 is essential for heme transport from the phagolysosome of macrophages during erythrophagocytosis. Cell Metab. *17*, 261–270.

Wojchowski, D.M. (2014). Erythroid mRNA processing: a “splice of life.” Blood *124*, 474–475.

Wong, P., Hattangadi, S.M., Cheng, A.W., Frampton, G.M., Young, R.A., and Lodish, H.F. (2011). Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes. Blood *118*, e128–e138.

Worthington, M.T., Cohn, S.M., Miller, S.K., Luo, R.Q., and Berg, C.L. (2001). Characterization of a human plasma membrane heme transporter in intestinal and hepatocyte cell lines. Am. J. Physiol.-Gastrointest. Liver Physiol. *280*, G1172–G1177.

Wu, W., Cheng, Y., Keller, C.A., Ernst, J., Kumar, S.A., Mishra, T., Morrissey, C., Dorman, C.M., Chen, K.-B., Drautz, D., et al. (2011). Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. Genome Res. *21*, 1659–1671.

Yamamoto, M., Arimura, H., Fukushige, T., Minami, K., Nishizawa, Y., Tanimoto, A., Kanekura, T., Nakagawa, M., Akiyama, S., and Furukawa, T. (2014). Abcb10 Role in Heme Biosynthesis In Vivo: Abcb10 Knockout in Mice Causes Anemia with Protoporphyrin IX and Iron Accumulation. Mol. Cell. Biol. *34*, 1077–1084.

Yamamoto, M.L., Clark, T.A., Gee, S.L., Kang, J.-A., Schweitzer, A.C., Wickrema, A., and Conboy, J.G. (2009). Alternative pre-mRNA splicing switches modulate gene expression in late erythropoiesis. Blood *113*, 3363–3370.

Yamanaka, K., Ishikawa, H., Megumi, Y., Tokunaga, F., Kanie, M., Rouault, T.A., Morishima, I., Minato, N., Ishimori, K., and Iwai, K. (2003). Identification of the ubiquitin–protein ligase that recognizes oxidized IRP2. Nat. Cell Biol. *5*, 336–340.

Yan, J., Liu, Y., Lukasik, S.M., Speck, N.A., and Bushweller, J.H. (2004). CBFβ allosterically regulates the Runx1 Runt domain via a dynamic conformational equilibrium. Nat. Struct. Mol. Biol. *11*, 901–906.

Yang, G., Huang, S.-C., Wu, J.Y., and Benz, E.J. (2005). An erythroid differentiation-specific splicing switch in protein 4.1R mediated by the interaction of SF2/ASF with an exonic splicing enhancer. Blood *105*, 2146–2153.

Yang, J., Ma, J., Xiong, Y., Wang, Y., Jin, K., Xia, W., Chen, Q., Huang, J., Zhang, J., Jiang, N., et al. Epigenetic regulation of megakaryocytic and erythroid differentiation by PHF2 histone demethylase. J. Cell. Physiol.

Yuan, X., Fleming, M.D., and Hamza, I. (2013). Heme transport and erythropoiesis. Curr. Opin. Chem. Biol. *17*, 204–211.

Yuan, X., Rietzschel, N., Kwon, H., Nuno, A.B.W., Hanna, D.A., Phillips, J.D., Raven, E.L., Reddi, A.R., and Hamza, I. (2016). Regulation of intracellular heme trafficking revealed by subcellular reporters. Proc. Natl. Acad. Sci. *113*, E5144–E5152.

Chapter 4

Genome-scale CRISPR-Cas9 Loss-of-function Screens Identify Mechanisms Affecting Erythroid Cell Sensitivity to Arsenic Trioxide

Introduction

Arsenic is a ubiquitous toxicant that is associated with multiple health issues. The severity of arsenic-related health outcomes largely depends on the exposure. Acute arsenic exposure can result in severe digestive and neuronal complications while chronic exposure can lead to carcinogenesis (Ratnaike, 2003). Arsenic exists in organic and inorganic forms and is available in different valence states, but the trivalent and pentavalent states are the most common (Hughes et al., 2011). In human cells, inorganic arsenic is metabolized to monomethylarsenic acid (MMA) and dimethylarsenic acid (DMA) which have a lower reactivity (Vahter and Concha, 2001).

Hematotoxicity of arsenic has not been extensively investigated. A few studies in rats indicated that arsenic exposure affects certain hematological parameters (Dwivedi et al., 2014; Jalaludeen et al., 2016), but the arsenic-induced changes of these parameters were inconsistent between the reports. The most studied role of arsenic in hematology is its therapeutic effect in hematologic tumors, and in particular, acute promyelocytic leukemia (APL). Arsenic trioxide (As_2O_3; ATO) is currently one of the most potent agents in the treatment of APL. ATO is effective as a single agent, but its combination with all-trans retinoic acid (ATRA) results in high APL cure rates with minimal toxic effects (Lo-Coco et al., 2016).

Several mechanisms of action of arsenic, that can result in injurious cellular outcomes, have been reported in different cell types (Miller et al., 2002). In general, ATO and other trivalent arsenicals target sulfhydryl groups in proteins resulting in altered protein functions (Shen et al., 2013). In APL cells, ATO induces partial differentiation and apoptosis (Emadi and Gore, 2010). These outcomes are believed to result from ATO-induced degradation of PML-RARα, a fusion oncoprotein found in the vast majority of APL cases (Zhang et al., 2010). Furthermore, several *in vitro* studies demonstrated a therapeutic potential of ATO in other cancer types (Akao et al., 1999; Rousselot et al., 1999; Shen et al., 1999; Zhang et al., 1998). These studies indicate that other mechanisms, that are independent of PML-RARα degradation, could underlie ATO-induced apoptosis in cancer cells. Different proposed mechanisms of action of ATO include interference with key cell signaling pathways (Miller et al., 2002), disruption of the mitochondrial membrane potential, induction of oxidative stress and activation of caspase-3 (Cai et al., 2000; Kumar et al., 2014). Increased oxidative stress is a hallmark of ATO cytotoxicity in cancer cells. Accumulation of reactive oxygen species (ROS) is speculated to result from ATO-induced dysfunction of the mitochondria, where the direct action of ATO disrupts the inner mitochondrial membrane potential (Miller et al., 2002). Alternatively, oxidative damage can occur as a result of disruption of the thioredoxin (Trx) system, due to direct irreversible inhibition of the antioxidant selenoprotein thioredoxin reductase (TrxR) by ATO (Lu et al., 2007).

In addition to APL, ATO exhibited promising therapeutic effects in a variety of hematologic and solid tumors (Ding et al., 2014; Munshi et al., 2002; Wang et al., 2015). While the therapeutic effect of ATO in APL is largely dependent on PML-RARα degradation, its toxicity in other

cancer cell types is driven by alternative mechanisms. Identification of such mechanisms is critical to expand the therapeutic utility of ATO to a variety of malignancies. Comprehensive identification of the determinants of sensitivity to ATO could reveal synergistic targets for potentiating ATO in cancer treatment. In addition, deciphering cellular processes affecting susceptibility to ATO could contribute to understanding mechanism(s) of arsenic toxicity and developing strategies to mitigate its adverse effects.

In this study, we used a comprehensive functional genomic approach to study mechanisms of ATO cytotoxicity. The genome-wide CRISPR-Cas9 knockout tool was employed to generate a cellular pool where protein coding genes are globally disrupted. Mutants in the pool were screened in the presence of a sub-lethal dose of ATO to identify genes whose loss alters sensitivity to ATO. The cell line employed in this study, K562, is an erythroleukemic cell line derived from a chronic myelogenous leukemia (CML) patient. Using a cell line that doesn't express the PML-RARα allowed us to investigate mechanisms of ATO toxicity that are independent on degradation of this fusion oncoprotein. We identified novel pathways and individual cellular factors as determinants of vulnerability to ATO. This work provided new insights into mechanisms affecting susceptibility to arsenic toxicity.

Materials and methods

Cytotoxicity assays

Cell viability assays were performed using the CellTiter-Glo luminescent cell viability assay kit (Promega, Madison, WI) following the manufacturer's protocol. In brief, cells were seeded at a density of 10^5 cells/ml (10^4 cells/well) in opaque 96-well cell culture plates. Arsenic trioxide (ATO/As_2O_3; STREM Chemicals Inc.) stock of 10 mM was prepared by dissolving 20 mg of As_2O_3 in 300 µl NaOH (1M) and completing the volume to 10 ml with PBS. Cells were treated with 0, 0.5, 1, 1.5 and 2 µM ATO for 24, 48 and 72 hours. For experiments involving selenium pre-treatment, cells were pre-treated with 10 µM sodium selenite (Na_2SeO_3; MP Biomedicals) for 24 hours, washed with PBS, seeded in new plates and treated with 0, 0.5, 1, 1.5 and 2 µM ATO for 48 hours. At the end of each incubation period, 100 µL of the CellTiter Glo reagent was added to each well and the cells were lysed at an orbital plate shaker for 2 minutes. The plate was then incubated for 10 minutes at room temperature in the dark to stabilize the signal. Luminescent signals from all wells were read on a Synergy H1 microplate reader (BioTek Instruments, Winooski, VT).

Genome-wide and focused CRISPR/Cas9 libraries

See supplementary methods on page 111.

Lentiviral production and functional titration

See supplementary methods on page 111.

Genome-wide (primary) screening

GeCKO v2 library, packaged in lentiviral particles was transduced into $100x10^6$ K562 cells in 12-well plates using the same protocol described for viral titration. Each well, containing $2.5x10^6$ cells, was transduced with 10 µl of the GeCKO v2 virus which results in a MOI of 0.45 (determined from titration). Transduced cells from all wells were pooled and the non-infected cells were eliminated by puromycin selection (2µg/ml) for 7 days during which the infected cells were expanded. The obtained cellular library was split into treatment (1 µM ATO) and control (NaOH vehicle) conditions where at least $25x10^6$ cells were maintained resulting in a representation of ~ 400-fold the library size. Selection was applied for 7 days which corresponds to 7 K562 cell doublings and the media were replaced every 48 hours. Screens were performed in T225 cell culture flasks and each condition was run in duplicate. At the end of the screen, $25x10^6$ cells from each replicate were washed with PBS and the pellets were saved for DNA extraction.

Validation (secondary) screening

$10x10^6$ K562 cells were transduced with the validation library in a 12-well plate using the same described protocol. Each well, containing $2.5x10^6$ cells, was transduced with 2.5 µl of the validation library virus which results in a MOI of 0.27 (determined from titration). The same screening conditions described for the primary screen were applied for the validation screen. At least 2.5 $x10^6$ cells were maintained in each condition representing ~ 900-fold the library size. Screens were performed in T25 cell culture flasks and each condition was run in triplicate. At the end of the screen, $2.5x10^6$ cells from each replicate were washed with PBS and the pellets were saved for DNA extraction.

DNA extraction, library preparation and next generation sequencing

See supplementary methods on page 112.

Data processing and computational analysis
See supplementary methods on page 113.

Generation of *PSTK* knockout and control pools
sgRNAs targeting *PSTK* or a non-targeting (NTC) sgRNA (sequences shown in Table S2) were cloned into the CRISPR lentiviral backbone vector (lentiCRISPRv2) using Gibson assembly. To produce lentiviral particles, we co-transfected HEK293T cells in T25 culture flasks with 3.4µg of the target vector, 2.6µg of the packaging plasmid (psPAX2) and 1.7µg of the envelope plasmid (pMD2.G) using Lipofectamine 2000 and Plus reagent (Thermo Fisher Scientific) following the manufacturer's instructions. Viral solutions were collected 60 hrs post transfection, filtered through a Steriflip-HV 0.45 µm low protein binding PVDF membrane (Millipore), and concentrated 50 folds using Lenti-X Concentrator (Clontech) following the manufacturer's protocol. Cells were transduced with targeting or NTC vectors at a MOI < 0.5 and the transduced cells were enriched by puromycin selection. The obtained KO cellular pools were used in cytotoxicity assays to evaluate their sensitivity to As_2O_3.

Results

Genome-wide screening revealed novel candidate genes involved cellular response to ATO

We investigated cellular mechanisms affecting sensitivity to ATO in K562 cells using the CRIRPR-Cas9 genome-wide loss-of-function screening approach (Figure 4.1). The GeCKO v2 sgRNA library was used to generate a pool of mutant K562 cells that were screened to identify genes whose loss confers increased sensitivity or resistance to ATO. An ATO dose of 1μM can inhibit cell growth by ~30% (IC30) after 72 hrs as shown by cell proliferation assays (Figure 4.2A). This sub-lethal ATO dose was used as a stressor during the screen. At the end of the screen, differential growth of mutants between ATO and vehicle treated pools was determined by comparing the representation of the corresponding guide sequences, determined by next generation sequencing, between the two pools. Candidate selection was based on individual guide sequences that are sufficiently represented (average log2 CPM > 2.5) and displayed differential abundance between ATO and vehicle pools with FDR < 0.1 (Figure 4.2B). Using these criteria, we identified 102 candidate genes implicated in the toxic response to ATO (Table S10). To maximize the number of candidates for secondary screening, we implemented additional analysis methods for primary candidate selection. An analysis approach which involves summing all guide sequences targeting the same gene identified 97 candidate genes with FDR < 0.1 (Table S11). Among these, 49 candidates did not overlap with the list identified from the analysis approach based on individual sgRNAs (Figure 4.2C). Within the overlapping candidates, there were 34 genes whose disruption potentially confers resistance to ATO and 14 genes whose loss potentially increases cellular sensitivity to ATO (Figure 4.2D). A third, more stringent, analysis method using MAGeCK identified only 10 candidate genes with FDR < 0.1 (Table S12), 9 of which were already identified by the other two approaches (Figure 4.2C).

Secondary screening simultaneously validated multiple candidate genes

We generated a customized CRISPR-Cas9 knockout library (validation library) that is enriched for the primary hits identified by all the described analysis methods. In addition to primary candidate genes, the validation library targets a control set of genes that were not identified in the primary screen as candidates and includes several non-targeting sgRNAs that serve as negative controls. Each gene in the validated library is targeted with 6-8 sgRNAs. Secondary screening using the validation library was performed in K562 cells using the same conditions applied for the primary screen but on a smaller scale. Our criteria for hit validation required at least two guides per gene exhibiting significant depletion or enrichment in ATO compared to the vehicle conditions with FDR < 0.001. Using these criteria, we validated multiple primary candidates. Validated candidates included genes whose loss confers either higher sensitivity or resistance to ATO, with the latter being more prominent (Figure 4.3A; Table S13). The secondary screen validated 20 candidates that were initially revealed by both the summing and the individual guide-based analysis methods, 16 candidates that were exclusively revealed by the summing method and 7 candidates that were uniquely identified by the individual guide-based method. The vast majority of the validated genes were also revealed by MAGeCK analysis (Table S14, S15). Functional classification of validated genes uncovered multiple biological processes as determinants of cellular sensitivity to ATO (Figure 4.3B; Table S16).

Disruption of selenocysteine utilization protects against ATO cytotoxicity

Among the genes whose loss alleviated cellular sensitivity to ATO, we found functional clusters related to protein synthesis. The most significant pathway whose disruption resulted in ATO

resistance is selenocysteine (Sec) metabolism, which is critical for the production of selenoproteins (Figure 4.4A). Intriguingly, inactivation of any gene involved in Sec biosynthesis or its incorporation into selenoproteins conferred resistance to ATO (Figure 4.4B). All the Sec metabolism genes were revealed as candidates in the primary screen and were validated in the secondary screen. At least 5 guide sequences targeting each of these genes showed significant enrichment in ATO relative to vehicle treated pools with FDRs < 0.001 (Table S13). To further confirm that Sec is a determinant of sensitivity to ATO, we targeted the gene encoding phosphoseryl tRNA kinase (PSTK), a critical enzyme in Sec biosynthesis, in K562 cells using the CRISPR-Cas9 gene knockout tool. We generated two independent PSTK mutant pools by transducing cells with lentiCRISPR vectors expressing different PSTK sgRNAs and evaluated sensitivity to ATO by cell viability assays. PSTK mutant pools displayed enhanced ATO tolerance compared to control pools (Figure 4.4C). A mutual protective effect between arsenic and selenium has been reported in many studies (Carew and Leslie, 2010; Levander, 1977; Srivastava et al., 2010; Sun et al., 2014). One plausible explanation of the role of defective Sec biosynthesis/utilization in conferring ATO resistance is the increased availability of intracellular free selenium. To study the effect of selenium on ATO toxicity in K562 cells, we pre-treated cells with an exogenous selenium source (10 μM Na_2SeO_3) and assessed cell viability in response to multiple ATO doses. Interestingly, loading cells with selenium significantly improved cell viability in the presence of ATO (Figure 4.4D). In addition to selenoprotein synthesis, another translation-related process, diphthalmide biosynthesis, was also identified as a determinant of ATO sensitivity. Inactivation of two components in the diphthalmide biosynthesis pathway (DPH5, DPH6) resulted in resistance to ATO toxicity (Figure 4.4E).

Identification of oxidative stress response components and transport proteins influencing tolerance to ATO

ATO-induced ROS formation has been suggested to contribute to ATO toxicity in cancer cells. Consistent with the role of ATO in oxidative stress, we identified KEAP1, the Nrf2 inhibitor, as a determinant of sensitivity of K562 cells to ATO. A striking resistance to ATO upon loss of KEAP1 was demonstrated in the secondary screen, where all eight guide sequences targeting KEAP1 displayed a significant enrichment in ATO compared to vehicle pools that varies between the different guides (Figure 4.5A). Unexpectedly, loss-of-function of the thioredoxin-like protein TXNDC17, a potential antioxidant, also resulted in considerable resistance to ATO (Figure 4.5A). Few transporter proteins, whose disruption alters sensitivity to ATO, were identified in the primary screen and validated by our confirmatory approach. Loss of each of aquaporin 3 (AQP3), the zinc transporter ZNT1 (SLC31A1) and its regulator metal regulatory transcription factor 1 (MTF1) enhanced tolerance to ATO whereas disruption of the multidrug resistance protein ABCC1 remarkably increased sensitivity to ATO. At least three distinct guide sequences targeting each of these genes displayed significant differential abundance between ATO and vehicle treated pools, where different mutants of the same gene exhibited similar ATO sensitivity profiles (Figure 4.5B).

Identification of potential ATO synergy targets

In addition to the transporter protein ABCC1, we identified molecular components whose inactivation increases sensitivity to ATO. These include the ubiquitin conjugating enzyme E2H (UBE2H), myotrophin (MTPN), CCR4-not transcription complex subunit 2 (CNOT2), condensin-2 complex subunit D3 (NCAPD3) and dynein cytoplasmic light intermediate chain 1 (DYNC1LI1). At least two distinct guide sequences targeting each of these genes displayed a

significant depletion exclusively in the ATO pools, where different mutants of the same gene exhibited similar ATO sensitivity profiles (Figure 4.5C).

Discussion

While arsenic trioxide is a highly effective therapy for acute promyelocytic leukemia (APL), its clinical use is limited to this disease. The major mechanism underlying the sensitivity of APL cells to ATO is the degradation of PML-RARα, the APL-specific fusion oncoprotein (Lallemand-Breitenbach et al., 2012). However, multiple PML-RARα independent mechanisms have been shown to contribute to ATO-induced apoptosis in APL cells (Davison et al., 2002). Despite the extensive clinical trials studying the efficacy and safety of ATO and other arsenical drugs in the treatment of several hematological malignancies and solid tumors (Dilda and Hogg, 2007), the therapeutic use of arsenicals is exclusively confined to APL treatment. The absence of PML-RARα oncoprotein in other cancer types requires higher doses of ATO, which can result in multiple adverse effects. An alternative approach to potentiate ATO in cancer therapy is to target cellular processes interfering with ATO cytotoxicity in order to produce a synergistic effect. Hence, identification of the determinants of cellular sensitivity to ATO is the first step in developing therapeutic strategies for using ATO in combination with other drugs in cancer therapy.

In this study, we used genome-wide CRISPR-Ca9 loss-of-function screening to investigate cellular mechanisms affecting susceptibility to ATO in human K562 cells. To our knowledge, a comprehensive investigation of the role of genes in susceptibility of mammalian cells to arsenic by functional genomic screening has not been reported. A genome-wide functional RNA interference (RNAi) screen in human HEK293 cells focused exclusively on arsenic-induced endoplasmic reticulum (ER) stress and identified only 4 genes critical to the process (Oh et al., 2012). Genetic screens studying arsenic toxicity were performed in budding yeast (*Saccharomyces cerevisiae*), and identified multiple cellular components/pathways modulating sensitivity to arsenic (Dilda et al., 2008; J. Johnson et al., 2016; Zhou et al., 2009). Comparing results from the yeast studies to our results, we found that some cellular processes that are mechanistically relevant to arsenic toxicity in yeast also contribute to arsenic sensitivity in human cells. However, overlap between human homologs of the genes identified in the yeast screens and genes identified in our screen was minimal. Determinants of arsenic susceptibility identified in our study, which also contribute to arsenic sensitivity in yeast include oxidative stress (Dilda et al., 2008), histone modification (Zhou et al., 2009) and protein translation (J. Johnson et al., 2016). In addition to these processes, our screen revealed novel cellular pathways as determinants of sensitivity to ATO (Figure 4.3B).

Selenocysteine (Sec) metabolism was revealed by our screen as a key process influencing ATO toxicity. An overview of Sec biosynthesis and insertion into proteins is shown in Figure 4.4A. Production of Sec is initiated by charging the Sec tRNA with serine (Ser) which is converted to Sec in a multi-step process catalyzed by several enzymes (Schmidt and Simonović, 2012). Binding of the Sec $tRNA^{Sec}$ to the translational complex and subsequent insertion of Sec into the polypeptide chain requires a UGA codon and a downstream selenocysteine insertion sequence (SECIS) in the corresponding mRNA and is facilitated by the SECIS binding protein (SBP2) and the Sec-specific elongation factor (EFSEC) (Hoffmann and Berry, 2006). Interestingly, our results clearly demonstrate that the loss of any enzyme involved in Sec biosynthesis or factors involved in Sec incorporation into the polypeptide chain results in substantial ATO tolerance (Figure 4.4B, C). The observed variability in ATO resistance phenotypes between different sgRNAs targeting the same gene is possibly due to variable on-target activities of these sgRNAs which could result in different levels of loss of the protein's function. Our finding is apparently

inconsistent with the role of selenoproteins in protecting against oxidative damage (Arbogast and Ferreiro, 2009), especially since induction of oxidative stress is a major outcome of arsenic toxicity (Flora, 2011).

A well-known selenoprotein, thioredoxin reductase (TrxR), has been identified as an important molecular target of ATO (Lu et al., 2007). TrxR is a dimeric ubiquitous enzyme involved in the reduction of thioredoxin (Trx) through an NADPH-dependent mechanism (Holmgren, 1989). Mammalian TrxR contains an N-terminal active site that includes two cysteines (-Cys^{59}-Val-Asn-Val-Gly-Cys^{64}), and a C-terminal active site including a cysteine and a selenocysteine residue (-Gly-Cys^{497}-$SeCys^{498}$-Gly) which distinguishes the mammalian enzyme from yeast and bacterial homologs (Zhong and Holmgren, 2000; Zhong et al., 2000). The Sec residue of TrxR is essential for Trx reduction. The importance of Sec for ATO binding to TrxR is demonstrated by the lack of inhibition of the Sec-lacking bacterial TrxR by ATO, and the critical role of Sec in the inhibition of TrxR by other electrophilic compounds (Anestål et al., 2008; Lu et al., 2007). Binding of certain electrophilic compounds to TrxR inhibits Trx reduction, increases NADPH oxidase activity and converts the enzyme into an apoptosis-inducing prooxidant known as selenium compromised thioredoxin reductase-derived apoptotic protein or SeCTRAP (Anestål and Arnér, 2003; Anestål et al., 2008). These findings, in addition to our observation that defective Sec production and insertion into selenoproteins decrease cellular sensitivity to ATO, suggest that ATO targeting of TrxR and generation of SeCTRAPs could be one of the mechanisms underlying ATO cytotoxicity. Delivering Sec-deficient TrxRs to cells results in induction of apoptosis (Anestål and Arnér, 2003), indicating that defective Sec metabolism itself has deleterious effects on cells. It can be speculated that the toxicity of exogenous Sec-deficient TrxRs is caused by a dominant-negative mechanism in which Trx reduction by endogenous TrxRs is inhibited. A reduction in cell viability was also observed in PSTK mutant pools (data not shown), where endogenous Sec-deficient TrxRs could be generated. Defective Trx reduction likely underlies the poor cellular growth in these pools. However, PSTK mutants displayed resistance to ATO toxicity (Figure 4.4C). This indicates that Sec-deficient TrxRs, despite their negative effect on cellular growth, are protected from electrophile-derived conversion to Sec-TRAPs. Such a protective effect can exacerbate cytotoxicity induced by electrophiles targeting TrxRs. Increased NADPH oxidase activity of TrxR upon incubating the enzyme with ATO *in vitro* was not observed (Lu et al., 2007). However, it is possible that an electrophilic metabolite, produced by cellular metabolism of ATO, would compromise TrxR resulting in induction of NADPH oxidase activity of the enzyme. Additionally, it has been suggested that the apoptotic activity of the compromised TrxR requires an endogenous cellular substrate (Anestål et al., 2008). Intriguingly, the thioredoxin-like TrxR substrate TRP14 (TXNDC17), was revealed by our screen as a determinant of sensitivity to ATO (Figure 4.5 A). This suggests that TRP14 could potentially serve as an endogenous substrate for SecTRAPs. The role of TRP14 in the apoptotic function of SecTRAPs requires further investigation. A proposed model for ATO-derived SecTRAPs is shown in Figure 4.6B.

The protective interactions of selenium and arsenic have been demonstrated in several reports (Carew and Leslie, 2010; Levander, 1977; Srivastava et al., 2010; Sun et al., 2014). One mechanism underlying this mutual protection is the formation of the Seleno-bis (S-glutathionyl) arsinium ion $[(GS)_2AsSe]^-$, a compound that is actively excreted outside the cells through the multidrug resistance associated protein ABCC2 (MRP2) (Carew and Leslie, 2010). This detoxification process provides a different perspective for understanding the link between

selenocysteine metabolism and ATO toxicity. It can be speculated that the utilization of selenium for the production of selenocysteine leads to depletion of intracellular selenium levels. Alternatively, defective selenocysteine biosynthesis could result in accumulation of intracellular selenium that can be used in the formation of the $[(GS)_2AsSe]^-$ complex resulting in arsenic detoxification (Figure 4.6A). In agreement with this hypothesis, we showed that loading cells with selenium from an exogenous source protects against ATO toxicity (Figure 4.4D). We identified the multidrug transporter ABCC1 (MRP1) as a major sensitivity determinant for ATO (Figure 4.5B). Although ABCC1 is a candidate exporter of the $[(GS)_2AsSe]^-$ complex, studies on membrane vesicles prepared from human erythrocytes established $[(GS)_2AsSe]^-$ as a substrate for ABCC2 and not ABCC1 (Carew and Leslie, 2010). Further investigation is required to assess the role of ABCC1 in transporting the $[(GS)_2AsSe]^-$ complex in K562 cells.

Another identified translational process modulating sensitivity to ATO is diphthalmide biosynthesis. Diphthalmide is a post-translationally modified histidine found exclusively on the translation elongation factor 2 (eEF2) and is crucial for maintaining translational fidelity (Su et al., 2013). Diphthalmide is a target of ADP-ribosylating bacterial toxins and the disruption of diphthalmide biosynthesis results in resistance to these toxins (Liu and Leppla, 2003). To our knowledge, loss of diphthalmide has not been previously shown to confer resistance to other toxic compounds. Our screen demonstrates that inactivation of components of the diphthalmide biosynthesis pathway decreases sensitivity to ATO (Figure 4.4E). Since selenocysteine is encoded by UGA, which normally serves as a stop codon, translational fidelity is required for its insertion into selenoproteins (Zavacki et al., 2003). Consequently, ATO tolerance as a result of defective diphthalmide biosynthesis could be due to impaired selenocysteine incorporation.

Several mechanisms underlying arsenic-induced oxidative stress have been proposed. One mechanism involves direct interaction of an arsenic metabolite, dimethylarsine, with molecular oxygen to produce a peroxyl radical (Yamanaka et al., 1990). Arsenic-driven oxidative damage is often associated with a loss of mitochondrial membrane (MM) potential (Han et al., 2008; Hosseini et al., 2013; Park et al., 2000; Woo et al., 2002). Since the accumulation of intracellular reactive oxygen species (ROS) itself can promote MM damage, it is unclear whether the disruption of the MM potential is initiated by a direct action of arsenic on the mitochondria or it is secondary to ROS production via a different arsenic-driven mechanism. Certain arsenic species were proposed to induce oxidative stress by facilitating the release of iron from ferritin, which promotes iron catalyzed ROS production (Ahmad et al., 2000). Our study and other studies (Lin et al., 1999; Lu et al., 2007) indicated that arsenic can cause oxidative damage by targeting the antioxidant enzyme thioredoxin reductase. In agreement with the oxidative stress model of arsenic toxicity, we found that loss of KEAP1, the negative regulator of the antioxidant transcription factor Nrf2, conferred substantial resistance to ATO (Figure 4.5A). This finding is supported by the reported activation of the Nrf2-KEAP1 pathway by arsenical compounds (Lau et al., 2013), and the defensive role of Nrf2 against arsenic toxicity (Du et al., 2008).

Transporter proteins involved in cellular import or export of arsenical compounds are typically determinants of sensitivity or resistance to arsenic respectively. Our screen identified few transporter proteins whose disruption alters sensitivity of K562 cells to ATO (Figure 4.5B). Consistent with the reported role of aquaporins (AQPs) in importing trivalent arsenicals into mammalian cells (Rosen, 2002), we showed that loss of AQP3 decreased sensitivity of K562 cells to ATO. The revealed role of AQP3 in cellular ATO import is in line with a previous report

showing decreased cellular arsenic uptake upon AQP3 knockdown and increased accumulation of arsenic due to AQP3 overexpression (Lee et al., 2006). The zinc exporter ZnT1 (SLC30A1) was also revealed by our screen as a sensitivity determinant to ATO. Inactivation of ZnT1 or its transcriptional activator MTF-1 resulted in considerable ATO resistance. This finding suggests that cellular accumulation of zinc can result in arsenic tolerance. In support of this suggestion, a protective role of zinc against arsenic toxicity was reported in few studies (Kreppel et al., 1994; Milton et al., 2004). Multidrug resistance proteins (MRPs) play a critical role in cellular arsenic detoxification through the export of arsenic-glutathione complexes (Leslie, 2012). Our screen indicated that loss of MRP1 (ABCC1) renders cells vulnerable to ATO toxicity. The protective effect of ABCC1 against ATO is likely due to its role in exporting the arsenic triglutathione [$As(GS)_3$] complex out of the cell (Leslie et al., 2004).

Cellular components whose inactivation confers increased sensitivity to ATO could potentially serve as synergistic targets for potentiating ATO in cancer therapy. Few of these components (UBE2H, MTPN, CNOT2, NCAPD3, DYNC1LI1) were identified by our screen and validated by our confirmatory approach (Figure 4.5C). Interestingly, deletion of *UBC8*, the yeast homolog of the human *UBE2H* gene, conferred susceptibility of yeast cells to arsenic toxicity (Zhou et al., 2009). To our knowledge, no evidence supporting the role of any of the other candidates in cellular sensitivity to arsenic was reported.

Figures

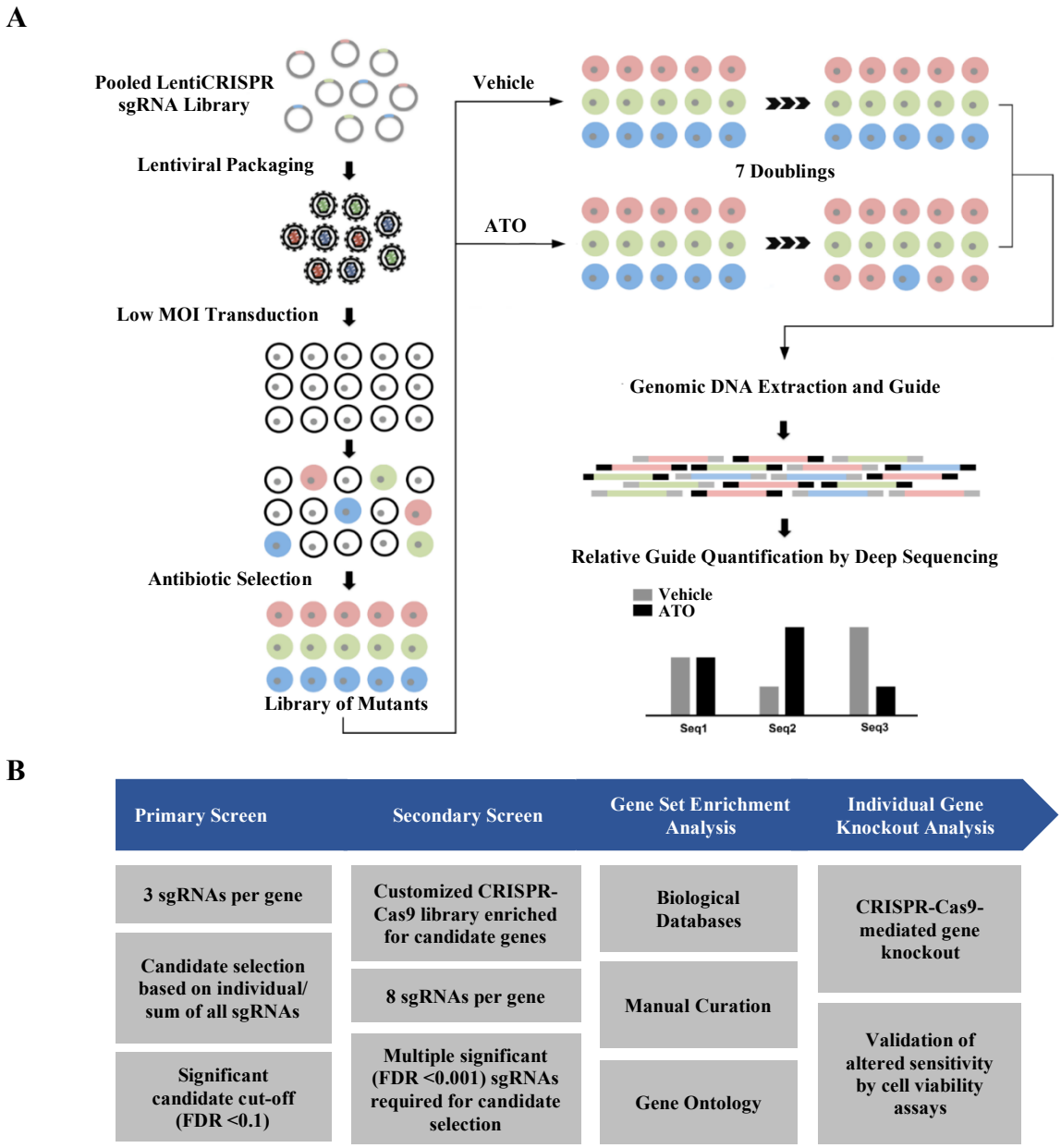

Figure 4.1: Overview of the CRISPR-based functional genomic approach designed to study ATO toxicity. (A) K562 cells are transduced with the Cas9-sgRNA library at a low MOI to generate a library of mutants. Screening is performed in media containing 1μM ATO or equivalent vehicle to identify mutants displaying growth advantage or disadvantage in the presence of ATO. Guide sequences encoding all sgRNAs are PCR-amplified and quantified by deep sequencing. Each guide sequence is used as a barcode to label the corresponding mutant. Abundance of each guide sequence is compared between vehicle and ATO-treated pools. (B) Study workflow showing screening strategies, candidate gene selection criteria, functional analysis and validation methods.

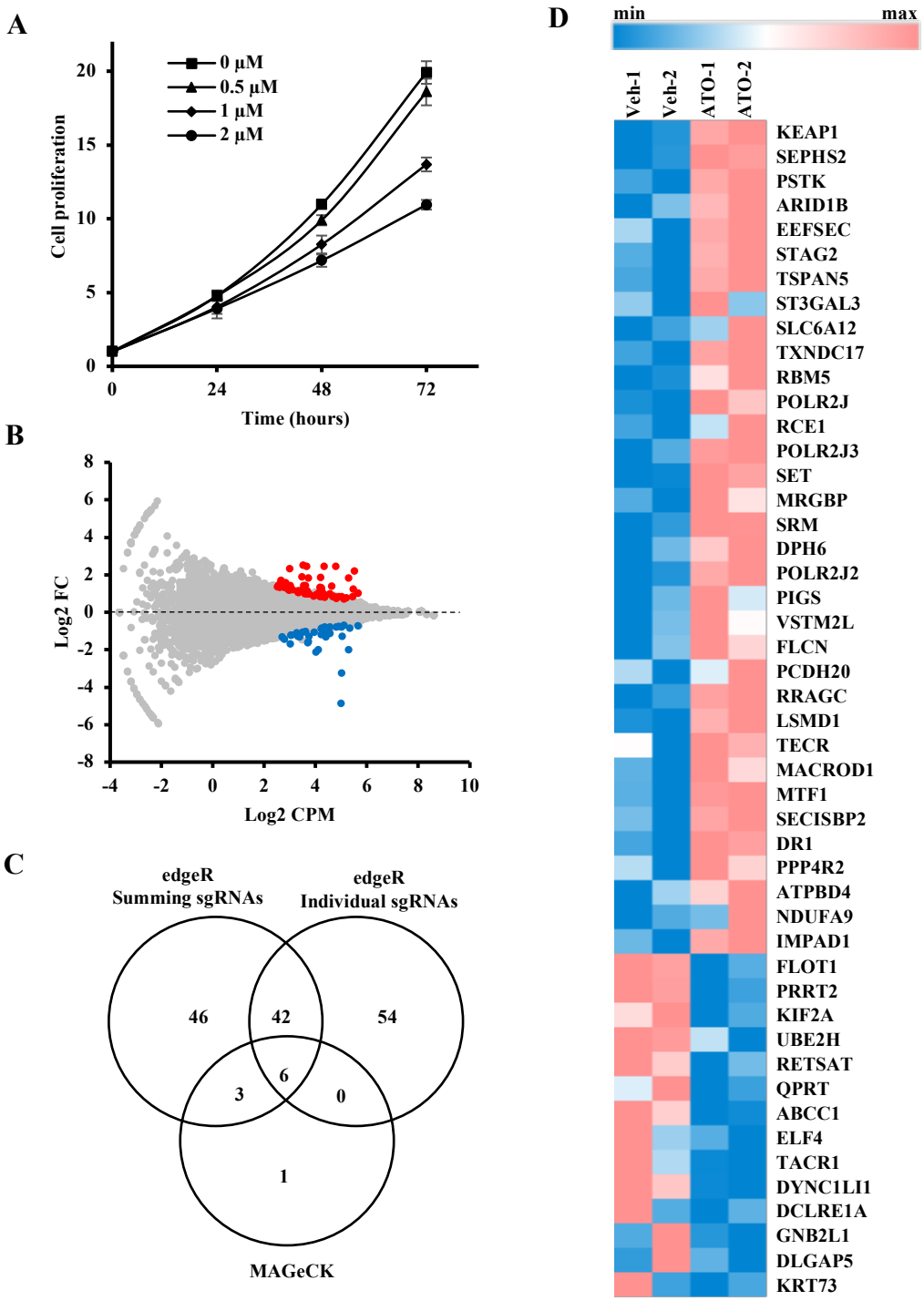

Figure 4.2: Identification of multiple candidate genes involved in ATO toxicity. (A) Cytotoxicity of ATO in K562 cells. Cells were treated with different concentrations of ATO and the quantity of viable cells was evaluated at 0, 24, 48 and 72 hours by CellTiter Glo assay. Values represent fold increase in the luminescent signal relative to the initial time point (T=0). Data are represented as mean ± standard deviation (n=3) (B) Scatter plot showing differential abundance of each guide sequence between ATO and vehicle treated pools. Fold changes (FC) are calculated by dividing the average normalized counts (n=2) for each guide in ATO pools to

that in vehicle pools. Log2 FCs are plotted against the average abundance of each guide in all pools represented as Log2 counts per million (CPM). Guide sequences enriched in ATO relative to vehicle (Log2 FC > 0) with FDR < 0.1 are represented as red dots. Guide sequences depleted in ATO relative to vehicle (Log2 FC < 0) with FDR <0.1 are represented as blue dots. (C) Venn diagram showing unique and common primary candidates identified by three different analysis methods. For each method, number of hits with FDR < 0.1 is shown. (D) Heatmap showing changes in the normalized sum of counts of all guide sequences targeting a gene between vehicle (Veh) and ATO pools. Only hits that were identified by both summing-based and individual guide-based analysis methods are shown. The screen was run in duplicate and the values corresponding to each vehicle and ATO replicate are represented in the heat map.

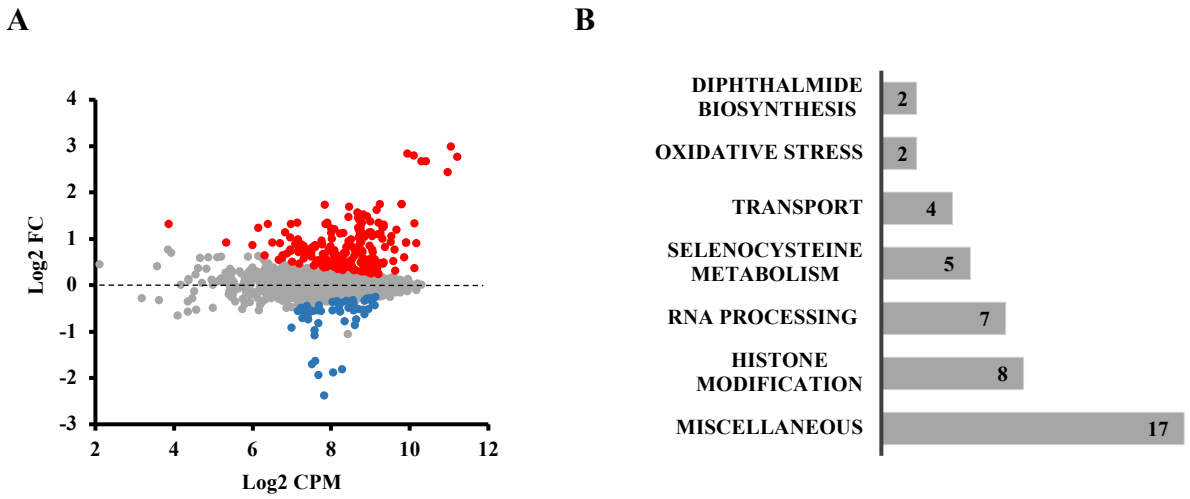

Figure 4.3: Simultaneous validation of multiple candidate genes involved in ATO toxicity by secondary screening. (A) Scatter plot showing differential abundance of each guide sequence between ATO and vehicle treated pools. Fold changes (FC) are calculated by dividing the average normalized counts (n=3) for each guide in ATO pools to that in vehicle pools. Log2 FCs are plotted against the average abundance of each guide in all pools represented as Log2 counts per million (CPM). Guide sequences enriched in ATO relative to vehicle (Log2 FC > 0) with FDR < 0.001 are represented as red dots. Guide sequences depleted in ATO relative to vehicle (Log2 FC < 0) with FDR <0.001 are represented as blue dots. (B) Functional classification of validated genes showing the different cellular processes involved in ATO toxicity. Genes were classified based on information from Gene Ontology (GO) enrichment analysis combined with manual literature-based curation. Bars represent the number of genes in each class (gene numbers are also shown on the bars).

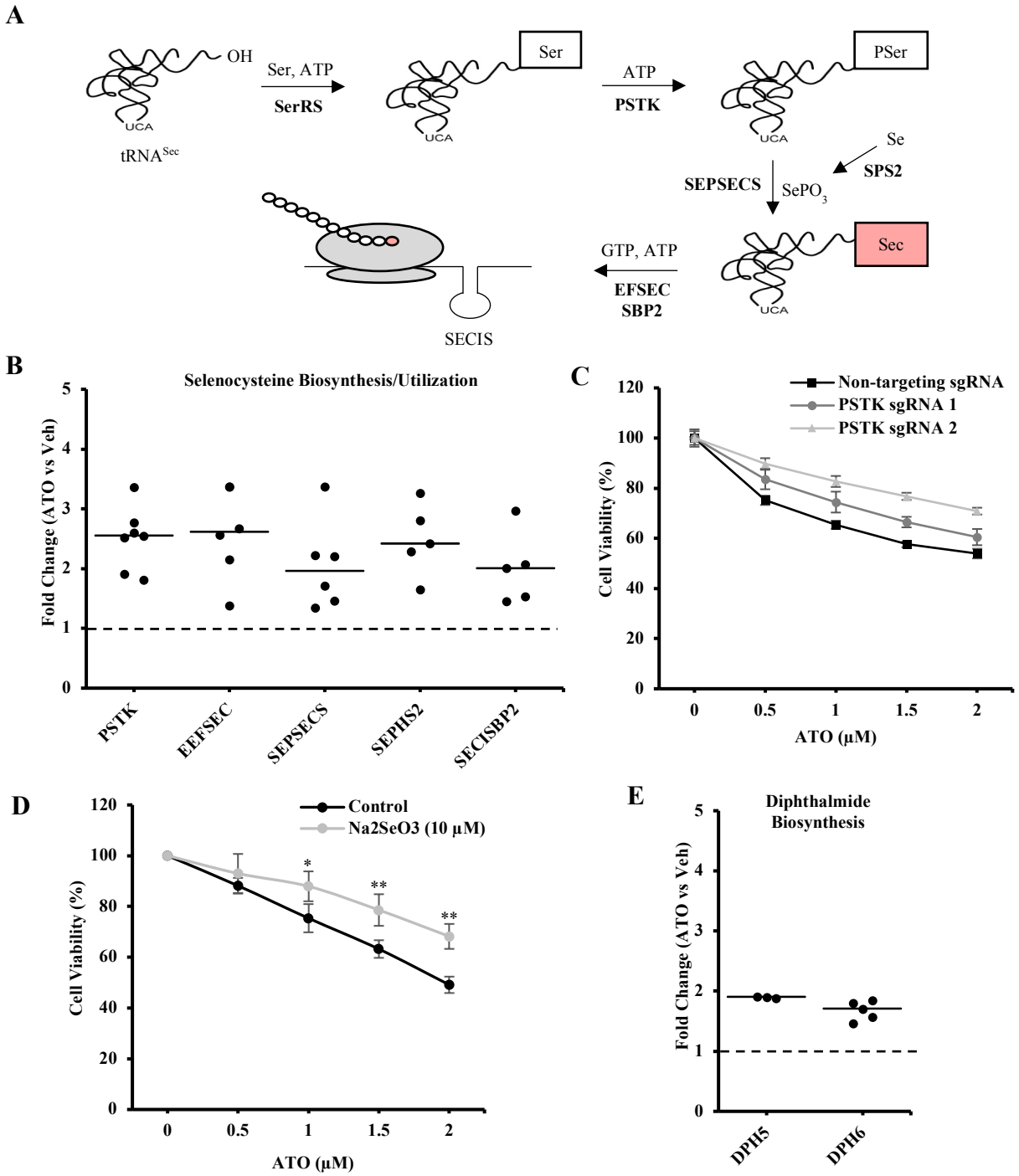

Figure 4.4: Disruption of selenocysteine biosynthesis and utilization decreases cellular susceptibility to ATO. (A) Overview of selenocysteine biosynthesis and incorporation into selenoproteins. The selenocysteine tRNA ($tRNA^{Sec}$) is initially charged with serine (Ser) by the seryl-tRNA synthetase (SerRS; *SARS* gene). Serine is then phosphorylated by the phosphoseryl tRNA kinase (PSTK; *PSTK* gene) to generate phosphoserine (PSer). Selenophosphate ($SePO_3$), synthesized from selenide by the action of selenophosphate synthetase (SPS2; *SEPHS2* gene), is used as the selenium donor for the conversion of PSer to selenocysteine (Sec), which is catalyzed

by the selenocysteine synthase (SEPSECS; *SEPSECS* gene). The selenocysteine-specific elongation factor (EFSEC; *EEFSEC* gene) and the Sec insertion sequence binding protein 2 (SBP2; *SECISBP2* gene) facilitate the insertion of Sec into the growing polypeptide chain in a mechanism that requires recognition of the Sec insertion sequence (SECIS) stem loop in the mRNA encoding selenoproteins. (B) Enrichment of guide sequences targeting components of the Sec biosynthesis/utilization pathway in ATO relative to vehicle controls. Each dot represents a unique guide sequence targeting the corresponding gene. Values correspond to fold enrichment of each guide in ATO relative to vehicle pools. Only enriched guides with FDR<0.001 are plotted. Lines represent median fold enrichment for all guides targeting each gene. (C) Decreased sensitivity of PSTK mutant pools to ATO. K562 cells were transduced with CRISPR/sgRNA vectors targeting PSTK or non-targeting control and non-transduced cells were depleted by puromycin selection. Cells were treated with different doses of ATO for 48 hours and cell viability was evaluated by CellTiter Glo assay. Cell viability is represented as percentage of vehicle-treated control. Data are represented as mean ± standard deviation (n=3) (D) Selenium-mediated protection against ATO toxicity. K562 cells were pre-treated with 10 μM sodium selenite for 24 hours, washed and treated with different doses of ATO for 48 hours. Cell viability was evaluated by CellTiter Glo assay. Cell viability is represented as percentage of vehicle control. Data are represented as mean of 3 independent experiments ± standard deviation. Statistical significance for the difference between pre-treated and control cells was determined by Student's t-test, where *=$P < 0.05$, **=$P < 0.01$. (E) Enrichment of guide sequences targeting components of the diphthalmide biosynthesis in ATO relative to vehicle controls. Each dot represents a unique guide sequence targeting the corresponding gene. Values correspond to fold enrichment of each guide in ATO relative to vehicle pools. Only enriched guides with FDR<0.001 are plotted. Lines represent median fold enrichment for all guides targeting each gene.

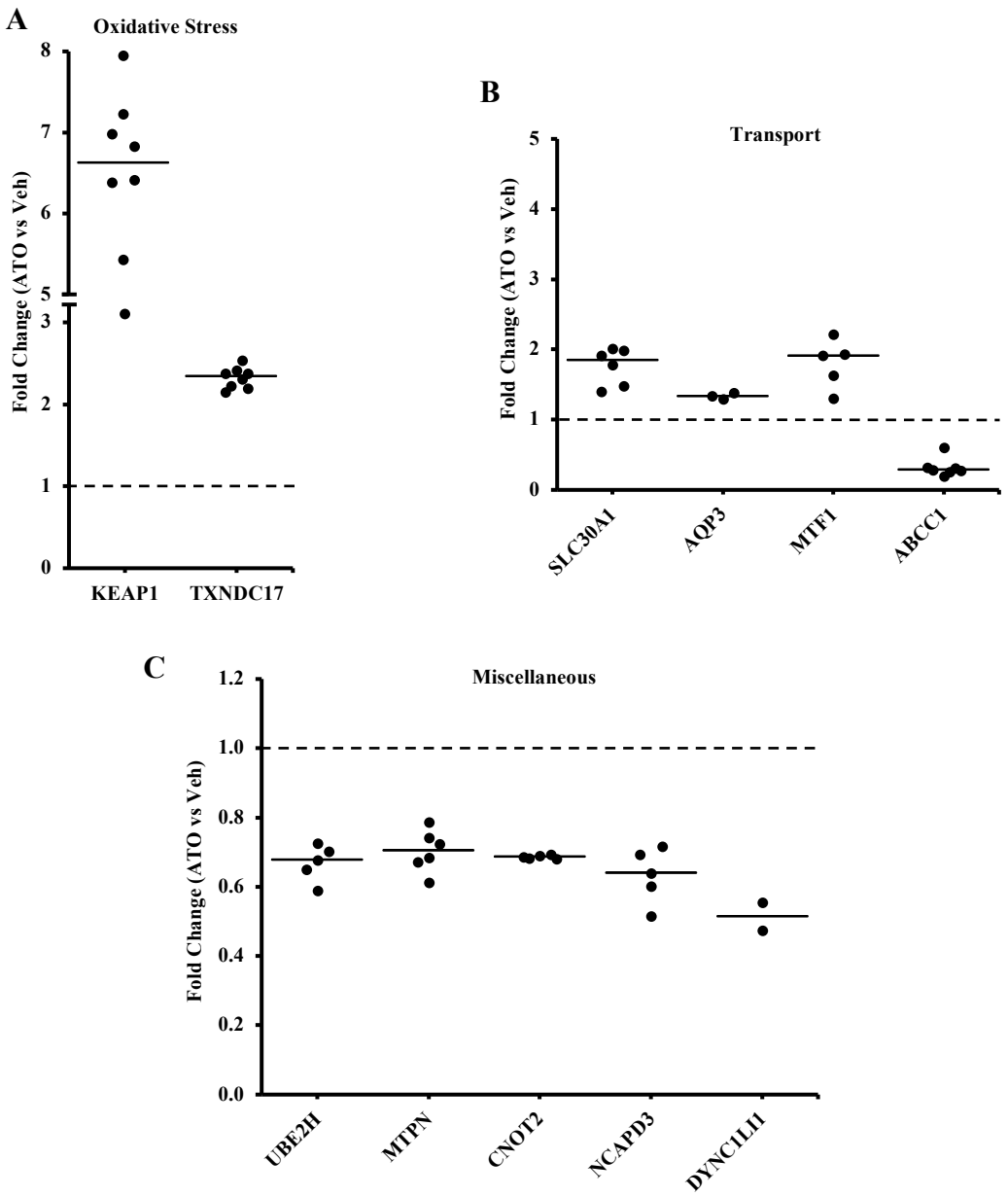

Figure 4.5: Cellular sensitivity to ATO is modulated by multiple cellular processes. (A) Enrichment of guide sequences targeting oxidative stress components in ATO relative to vehicle controls. (B) Differential abundance of guide sequences targeting various transport-related proteins between ATO and vehicle pools. (C) Depletion of guide sequences targeting miscellaneous cellular components in ATO relative to vehicle pools. Each dot represents a unique guide sequence targeting the corresponding gene. Values correspond to fold change of each guide in ATO relative to vehicle pools. Only guides showing differential abundance with FDR<0.001 are plotted. Lines represent median fold enrichment for all guides targeting each gene.

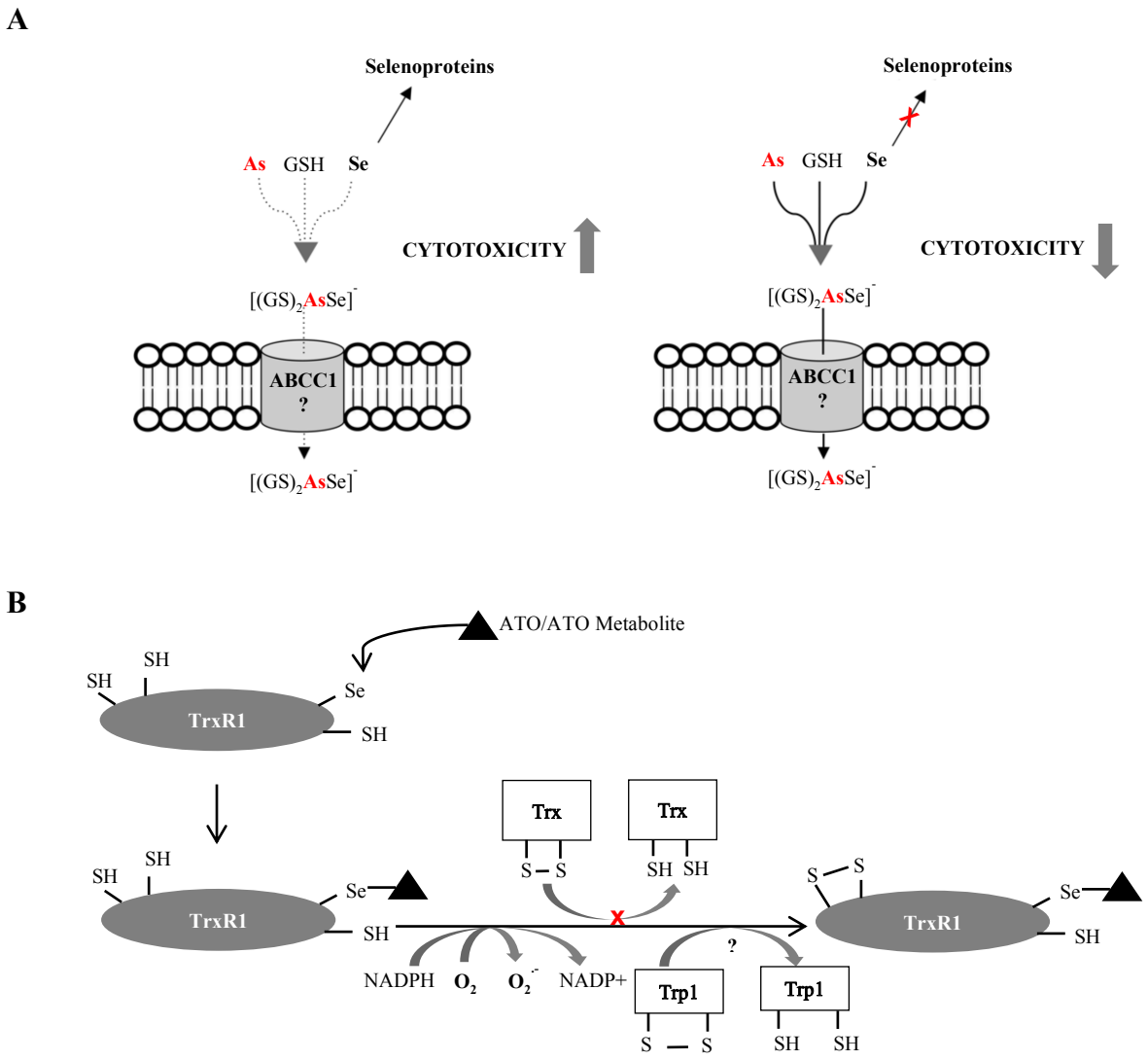

Figure 4.6: Proposed models illustrating potential roles of selenocysteine metabolism in ATO toxicity. (A) Arsenic-Glutathione-Selenium $[(GS)_2AsSe]^-$ export model. Arsenic forms a complex with selenium and glutathione that is exported out of the cell through multidrug resistance proteins. Utilization of selenium (Se) for selenocysteine biosynthesis and production of selenoproteins can deplete intracellular selenium levels leading to cellular accumulation of arsenic (As) which results in cytotoxicity (left). Disruption of selenocysteine biosynthesis/utilization can result in increased availability of intracellular selenium that detoxifies arsenic through the formation of the $[(GS)_2AsSe]^-$ complex (right). (B) Selenium Compromised Thioredoxin Reductase-derived Apoptotic Proteins (SeCTRAPs) model. ATO or its metabolite binding to the selenol group of Sec in the C-terminal active site of thioredoxin reductase (TrxR1) modifies the enzyme disrupting its ability to reduce thioredoxin (Trx). The modified enzyme gains an NADPH oxidase activity that produces superoxide. TRP14, the thioredoxin-related protein encoded by the *TXNDC17* gene is a potential endogenous substrate required for the NADPH oxidase activity of SeCTRAPs.

References

Ahmad, S., Kitchin, K.T., and Cullen, W.R. (2000). Arsenic Species That Cause Release of Iron from Ferritin and Generation of Activated Oxygen. Arch. Biochem. Biophys. *382*, 195–202.

Akao, Y., Nakagawa, Y., and Akiyama, K. (1999). Arsenic trioxide induces apoptosis in neuroblastoma cell lines through the activation of caspase 3 in vitro. FEBS Lett. *455*, 59–62.

Anestål, K., and Arnér, E.S.J. (2003). Rapid Induction of Cell Death by Selenium-compromised Thioredoxin Reductase 1 but Not by the Fully Active Enzyme Containing Selenocysteine. J. Biol. Chem. *278*, 15966–15972.

Anestål, K., Prast-Nielsen, S., Cenas, N., and Arnér, E.S.J. (2008). Cell Death by SecTRAPs: Thioredoxin Reductase as a Prooxidant Killer of Cells. PLOS ONE *3*, e1846.

Arbogast, S., and Ferreiro, A. (2009). Selenoproteins and Protection against Oxidative Stress: Selenoprotein N as a Novel Player at the Crossroads of Redox Signaling and Calcium Homeostasis. Antioxid. Redox Signal. *12*, 893–904.

Cai, X., Shen, Y.-L., Zhu, Q., Jia, P.-M., Yu, Y., Zhou, L., Huang, Y., Zhang, J.-W., Xiong, S.-M., Chen, S.-J., et al. (2000). Arsenic trioxide-induced apoptosis and differentiation are associated respectively with mitochondrial transmembrane potential collapse and retinoic acid signaling pathways in acute promyelocytic leukemia. Leukemia *14*, 262.

Carew, M.W., and Leslie, E.M. (2010). Selenium-dependent and -independent transport of arsenic by the human multidrug resistance protein 2 (MRP2/ABCC2): implications for the mutual detoxification of arsenic and selenium. Carcinogenesis *31*, 1450–1455.

Davison, K., Mann, K.K., and Miller, W.H. (2002). Arsenic trioxide: mechanisms of action. Semin. Hematol. *39*, 3–7.

Dilda, P.J., and Hogg, P.J. (2007). Arsenical-based cancer drugs. Cancer Treat. Rev. *33*, 542–564.

Dilda, P.J., Perrone, G.G., Philp, A., Lock, R.B., Dawes, I.W., and Hogg, P.J. (2008). Insight into the selectivity of arsenic trioxide for acute promyelocytic leukemia cells by characterizing Saccharomyces cerevisiae deletion strains that are sensitive or resistant to the metalloid. Int. J. Biochem. Cell Biol. *40*, 1016–1029.

Ding, D., Lim, K.S., and Eberhart, C.G. (2014). Arsenic trioxide inhibits Hedgehog, Notch and stem cell properties in glioblastoma neurospheres. Acta Neuropathol. Commun. *2*, 31.

Du, Y., Villeneuve, N.F., Wang, X.-J., Sun, Z., Chen, W., Li, J., Lou, H., Wong, P.K., and Zhang, D.D. (2008). Oridonin confers protection against arsenic-induced toxicity through activation of the Nrf2-mediated defensive response. Environ. Health Perspect. *116*, 1154–1161.

Dwivedi, N., Flora, G., Kushwaha, P., and Flora, S.J.S. (2014). Alpha-lipoic acid protects oxidative stress, changes in cholinergic system and tissue histopathology during co-exposure to arsenic-dichlorvos in rats. Environ. Toxicol. Pharmacol. *37*, 7–23.

Emadi, A., and Gore, S.D. (2010). Arsenic trioxide - An old drug rediscovered. Blood Rev. *24*, 191–199.

Flora, S.J.S. (2011). Arsenic-induced oxidative stress and its reversibility. Free Radic. Biol. Med. *51*, 257–281.

Han, Y.H., Kim, S.Z., Kim, S.H., and Park, W.H. (2008). Arsenic trioxide inhibits the growth of Calu-6 cells via inducing a G2 arrest of the cell cycle and apoptosis accompanied with the depletion of GSH. Cancer Lett. *270*, 40–55.

Hoffmann, P.R., and Berry, M.J. (2006). The importance of subcellular localization of SBP2 and EFsec for selenoprotein synthesis. In Selenium, (Springer, Boston, MA), pp. 73–82.

Holmgren, A. (1989). Thioredoxin and glutaredoxin systems. J. Biol. Chem. *264*, 13963–13966.

Hosseini, M.-J., Shaki, F., Ghazi-Khansari, M., and Pourahmad, J. (2013). Toxicity of Arsenic (III) on Isolated Liver Mitochondria: A New Mechanistic Approach. Iran. J. Pharm. Res. IJPR *12*, 121–138.

Hughes, M.F., Beck, B.D., Chen, Y., Lewis, A.S., and Thomas, D.J. (2011). Arsenic Exposure and Toxicology: A Historical Perspective. Toxicol. Sci. *123*, 305–332.

Jalaludeen, A.M., Ha, W.T., Lee, R., Kim, J.H., Do, J.T., Park, C., Heo, Y.T., Lee, W.Y., and Song, H. (2016). Biochanin A Ameliorates Arsenic-Induced Hepato- and Hematotoxicity in Rats. Molecules *21*, 69.

J. Johnson, A., Veljanoski, F., J. O'Doherty, P., S. Zaman, M., Petersingham, G., D. Bailey, T., Münch, G., Kersaitis, C., and J. Wu, M. (2016). Molecular insight into arsenic toxicity via the genome-wide deletion mutant screening of Saccharomyces cerevisiae. Metallomics *8*, 228–235.

Kreppel, H., Liu, J., Liu, Y., Reichl, F.X., and Klaassen, C.D. (1994). Zinc-Induced Arsenite Tolerance in Mice. Toxicol. Sci. *23*, 32–37.

Kumar, S., Yedjou, C.G., and Tchounwou, P.B. (2014). Arsenic trioxide induces oxidative stress, DNA damage, and mitochondrial pathway of apoptosis in human leukemia (HL-60) cells. J. Exp. Clin. Cancer Res. CR *33*, 42.

Lallemand-Breitenbach, V., Zhu, J., Chen, Z., and de Thé, H. (2012). Curing APL through PML/RARA degradation by As2O3. Trends Mol. Med. *18*, 36–42.

Lau, A., Whitman, S.A., Jaramillo, M.C., and Zhang, D.D. (2013). Arsenic-Mediated Activation of the Nrf2-Keap1 Antioxidant Pathway. J. Biochem. Mol. Toxicol. *27*, 99–105.

Lee, T.-C., Ho, I.-C., Lu, W.-J., and Huang, J. (2006). Enhanced Expression of Multidrug resistance-associated Protein 2 and Reduced Expression of Aquaglyceroporin 3 in an Arsenic-resistant Human Cell Line. J. Biol. Chem. *281*, 18401–18407.

Leslie, E.M. (2012). Arsenic–glutathione conjugate transport by the human multidrug resistance proteins (MRPs/ABCCs). J. Inorg. Biochem. *108*, 141–149.

Leslie, E.M., Haimeur, A., and Waalkes, M.P. (2004). Arsenic transport by the human multidrug resistance protein 1 (MRP1/ABCC1). Evidence that a tri-glutathione conjugate is required. J. Biol. Chem. *279*, 32700–32708.

Levander, O.A. (1977). Metabolic interrelationships between arsenic and selenium. Environ. Health Perspect. *19*, 159–164.

Lin, S., Cullen, W.R., and Thomas, D.J. (1999). Methylarsenicals and arsinothiols are potent inhibitors of mouse liver thioredoxin reductase. Chem. Res. Toxicol. *12*, 924–930.

Liu, S., and Leppla, S.H. (2003). Retroviral insertional mutagenesis identifies a small protein required for synthesis of diphthamide, the target of bacterial ADP-ribosylating toxins. Mol. Cell *12*, 603–613.

Lo-Coco, F., Di Donato, L., and Schlenk, R.F. (2016). Targeted Therapy Alone for Acute Promyelocytic Leukemia. N. Engl. J. Med. *374*, 1197–1198.

Lu, J., Chew, E.-H., and Holmgren, A. (2007). Targeting thioredoxin reductase is a basis for cancer therapy by arsenic trioxide. Proc. Natl. Acad. Sci. U. S. A. *104*, 12288–12293.

Miller, W.H., Schipper, H.M., Lee, J.S., Singer, J., and Waxman, S. (2002). Mechanisms of Action of Arsenic Trioxide. Cancer Res. *62*, 3893–3903.

Milton, A.G., Zalewski, P.D., and Ratnaike, R.N. (2004). Zinc protects against arsenic-induced apoptosis in a neuronal cell line, measured by DEVD-caspase activity. Biometals Int. J. Role Met. Ions Biol. Biochem. Med. *17*, 707–713.

Munshi, N.C., Tricot, G., Desikan, R., Badros, A., Zangari, M., Toor, A., Morris, C., Anaissie, E., and Barlogie, B. (2002). Clinical activity of arsenic trioxide for the treatment of multiple myeloma. Leukemia *16*, 1835–1837.

Oh, R.S., Pan, W.-C., Yalcin, A., Zhang, H., Guilarte, T.R., Hotamisligil, G.S., Christiani, D.C., and Lu, Q. (2012). Functional RNA Interference (RNAi) Screen Identifies System A Neutral Amino Acid Transporter 2 (SNAT2) as a Mediator of Arsenic-induced Endoplasmic Reticulum Stress. J. Biol. Chem. *287*, 6025–6034.

Park, W.H., Seol, J.G., Kim, E.S., Hyun, J.M., Jung, C.W., Lee, C.C., Kim, B.K., and Lee, Y.Y. (2000). Arsenic trioxide-mediated growth inhibition in MC/CAR myeloma cells via cell cycle arrest in association with induction of cyclin-dependent kinase inhibitor, p21, and apoptosis. Cancer Res. *60*, 3065–3071.

Ratnaike, R. (2003). Acute and chronic arsenic toxicity. Postgrad. Med. J. *79*, 391–396.

Rosen, B.P. (2002). Biochemistry of arsenic detoxification. FEBS Lett. *529*, 86–92.

Rousselot, P., Labaume, S., Marolleau, J.P., Larghero, J., Noguera, M.H., Brouet, J.C., and Fermand, J.P. (1999). Arsenic trioxide and melarsoprol induce apoptosis in plasma cell lines and in plasma cells from myeloma patients. Cancer Res. *59*, 1041–1048.

Schmidt, R.L., and Simonović, M. (2012). Synthesis and decoding of selenocysteine and human health. Croat. Med. J. *53*, 535–550.

Shen, S., Li, X.-F., Cullen, W.R., Weinfeld, M., and Le, X.C. (2013). Arsenic Binding to Proteins. Chem. Rev. *113*, 7769–7792.

Shen, Z.Y., Tan, L.J., Cai, W.J., Shen, J., Chen, C., Tang, X.M., and Zheng, M.H. (1999). Arsenic trioxide induces apoptosis of oesophageal carcinoma in vitro. Int. J. Mol. Med. *4*, 33–37.

Srivastava, D., Subramanian, R.B., Madamwar, D., and Flora, S.J.S. (2010). Protective effects of selenium, calcium, and magnesium against arsenic-induced oxidative stress in male rats. Arh. Hig. Rada Toksikol. *61*, 153–159.

Su, X., Lin, Z., and Lin, H. (2013). THE BIOSYNTHESIS AND BIOLOGICAL FUNCTION OF DIPHTHAMIDE. Crit. Rev. Biochem. Mol. Biol. *48*, 515–521.

Sun, H.-J., Rathinasabapathi, B., Wu, B., Luo, J., Pu, L.-P., and Ma, L.Q. (2014). Arsenic and selenium toxicity and their interactive effects in humans. Environ. Int. *69*, 148–158.

Vahter, M., and Concha, G. (2001). Role of metabolism in arsenic toxicity. Pharmacol. Toxicol. *89*, 1–5.

Wang, Y., Wang, L., Yin, C., An, B., Hao, Y., Wei, T., Li, L., and Song, G. (2015). Arsenic trioxide inhibits breast cancer cell growth via microRNA-328/hERG pathway in MCF-7 cells. Mol. Med. Rep. *12*, 1233–1238.

Woo, S.H., Park, I.-C., Park, M.-J., Lee, H.-C., Lee, S.-J., Chun, Y.-J., Lee, S.-H., Hong, S.-I., and Rhee, C.H. (2002). Arsenic trioxide induces apoptosis through a reactive oxygen species-dependent pathway and loss of mitochondrial membrane potential in HeLa cells. Int. J. Oncol. *21*, 57–63.

Yamanaka, K., Hoshino, M., Okamoto, M., Sawamura, R., Hasegawa, A., and Okada, S. (1990). Induction of DNA damage by dimethylarsine, a metabolite of inorganic arsenics, is for the major part likely due to its peroxyl radical. Biochem. Biophys. Res. Commun. *168*, 58–64.

Zavacki, A.M., Mansell, J.B., Chung, M., Klimovitsky, B., Harney, J.W., and Berry, M.J. (2003). Coupled tRNA(Sec)-dependent assembly of the selenocysteine decoding apparatus. Mol. Cell *11*, 773–781.

Zhang, W., Ohnishi, K., Shigeno, K., Fujisawa, S., Naito, K., Nakamura, S., Takeshita, K., Takeshita, A., and Ohno, R. (1998). The induction of apoptosis and cell cycle arrest by arsenic trioxide in lymphoid neoplasms. Leukemia *12*, 1383–1391.

Zhang, X.-W., Yan, X.-J., Zhou, Z.-R., Yang, F.-F., Wu, Z.-Y., Sun, H.-B., Liang, W.-X., Song, A.-X., Lallemand-Breitenbach, V., Jeanne, M., et al. (2010). Arsenic trioxide controls the fate of the PML-RARalpha oncoprotein by directly binding PML. Science *328*, 240–243.

Zhong, L., and Holmgren, A. (2000). Essential Role of Selenium in the Catalytic Activities of Mammalian Thioredoxin Reductase Revealed by Characterization of Recombinant Enzymes with Selenocysteine Mutations. J. Biol. Chem. *275*, 18121–18128.

Zhong, L., Arnér, E.S.J., and Holmgren, A. (2000). Structure and mechanism of mammalian thioredoxin reductase: The active site is a redox-active selenolthiol/selenenylsulfide formed from the conserved cysteine-selenocysteine sequence. Proc. Natl. Acad. Sci. U. S. A. *97*, 5854–5859.

Zhou, X., Arita, A., Ellen, T.P., Liu, X., Bai, J., Rooney, J.P., Kurtz, A.D., Klein, C.B., Dai, W., Begley, T.J., et al. (2009). A genome-wide screen in Saccharomyces cerevisiae reveals pathways affected by arsenic toxicity. Genomics *94*, 294–307.

Chapter 5

Identification of Molecular Determinants of Susceptibility to Acetaldehyde using CRISPR-Cas9-Genome-wide Knockout Screening in Human Erythroid Cells

Introduction

Acetaldehyde, an endogenous metabolite and the primary product of ethanol metabolism, is designated as Group 1 carcinogen by IARC (Secretan et al., 2009). Ethanol consumption is associated with increased risk of developing several forms of cancer, including esophageal (Brooks et al., 2009; Yokoyama and Omori, 2003) , liver (McKillop and Schrum, 2009), and breast cancer (Chen et al., 2011). Carcinogenic effects of acetaldehyde can arise from direct damage of the genome, through the formation of DNA adducts, or due to adduct-induced functional impairment of proteins that maintain genomic and epigenomic stability (Brooks and Zakhari, 2014). Acetaldehyde-derived DNA adducts include N(2)-ethylidene-2'-deoxyguanosine (N2-ethylidene-dG), which is the most prevalent, and N(2)-propano-2′-deoxyguanosine (PdG) (Setshedi et al., 2010). While the mutagenic potential of N2-ethylidene-dG is poorly studied, PdG was shown to contribute to DNA damage by inducing the formation of DNA-protein and DNA interstrand crosslinks, suggesting a role for this adduct in acetaldehyde-induced carcinogenesis (Brooks and Theruvathu, 2005). Cellular protein targets of acetaldehyde include erythrocyte membrane proteins, tubulin, ketosteroid reductase and the ethanol metabolizing enzyme CYP2E1(Niemelä, 1999). The exact mechanisms underlying toxicity and carcinogenesis of acetaldehyde, and the cellular processes modulating susceptibility to acetaldehyde and other aldehydes are still poorly understood (Brooks and Zakhari, 2014; Seitz and Stickel, 2007).

Hematotoxicity of acetaldehyde has been demonstrated in a few studies (Garaycoechea et al., 2012, 2018; Langevin et al., 2011; Parmar and D'Andrea, 2012; Venton et al., 2016). The genotoxic effect of acetaldehyde on hematopoietic stem cells (HSCs) was proposed as one of the mechanisms underlying bone marrow failure in Fanconi Anemia (FA), a blood disorder caused by defective DNA repair (Garaycoechea et al., 2012). It was shown that individual disruption of several genes in the FA DNA repair pathway results in increased susceptibility of cultured B cells to exogenous acetaldehyde (Langevin et al., 2011). The same study demonstrated that mice lacking *Fancd2,* a key gene in the FA DNA repair pathway, require normal activity of the acetaldehyde-catabolizing enzyme Aldh2 for development. In addition, $Aldh2^{-/-}Fancd2^{-/-}$ mice were shown to be highly susceptible to the toxicity of ethanol, the exogenous acetaldehyde precursor, and the majority of these mice develop leukemia. In humans, FA patients with ALDH2 deficiency, resulting from an *ALDH2* variant allele, showed a faster progression of bone marrow failure compared to control patients (Hira et al., 2013). In addition, induction of endogenous acetaldehyde accumulation, by pharmacological inhibition of ALDH2, is a promising therapeutic approach in the treatment of acute myeloid leukemia (AML), as it selectively targets leukemia stem cells (LSCs) (Venton et al., 2016). Collectively, these findings corroborate the hematotoxic effect of acetaldehyde and highlight the role of the FA DNA repair pathway in alleviating acetaldehyde toxicity.

In addition to the FA pathway, other DNA repair pathways have been shown to contribute to the protective effect against acetaldehyde-induced DNA damage in human cells. Disruption of Homologous Recombination (HR) components, including *BRCA1/2* and *RAD51*, increases

sensitivity to exogenous acetaldehyde treatment and endogenous acetaldehyde accumulation induced by pharmacological inhibition of ALDH2 (Tacconi et al., 2017). A role for the nucleotide excision repair (NER) pathway in repairing PdG adducts was demonstrated (Johnson et al., 1997), indicating that NER can also be involved in reversing acetaldehyde-induced DNA damage. Recently, a comprehensive analysis was performed in fission yeast to identify all the DNA repair pathways affecting sensitivity to acetaldehyde (Noguchi et al., 2016). This approach validated the protective roles of each of the FA, HR and NER pathways against acetaldehyde toxicity and revealed potential roles for base excision repair (BER), translesion DNA synthesis (TLS) and DNA-protein crosslink (DPC) repair in mitigating acetaldehyde toxicity (Brooks and Schuebel, 2016; Noguchi et al., 2016). To our knowledge, a similar large-scale genomic approach, investigating acetaldehyde susceptibility and tolerance mechanisms, in mammalian cells has not been performed. Elucidating these mechanisms in human cells is necessary for understanding the pathophysiology of hematologic and non-hematologic diseases related to alcohol consumption, as well as occupational exposure to acetaldehyde.

In this study, we used CRISPR-Cas9 loss-of-function screening to identify molecular determinants of susceptibility and resistance to acetaldehyde in human erythroleukemic K562 cells. A genome-wide pool of mutations was screened to identify those leading to altered sensitivity to a sub-lethal dose of acetaldehyde. Consistent with the reported role of aldehydes in DNA damage, multiple candidate genes, whose disruption conferred increased sensitivity to acetaldehyde, encode components of DNA repair. The screen revealed additional candidate genes implicated in acetaldehyde cytotoxicity, including a functionally uncharacterized tumor suppressor gene whose role in acetaldehyde tolerance was confirmed by CRISPR-base validation approaches.

Materials and methods

Cell culture and cell viability assays

Evaluation of cell viability was performed by measuring ATP levels using the CellTiter Glo luminescent cell viability assay kit (Promega) following the manufacturer's protocol. Briefly, K562 cells were seeded at a density of 10^5 cells/ml (10^4 cells/well) in opaque 96-well cell culture plates and exposed to multiple concentrations (0-50 mM) of acetaldehyde (Sigma) for 72 hours. To measure the ATP content in each well, cells were mixed with 100 µL of the CellTiter Glo reagent and lysed on an orbital plate shaker for 2 minutes. Plates were incubated for 10 minutes at room temperature in the dark to stabilize the signal. Luminescent signals from all wells were read on a Synergy H1 microplate reader (BioTek Instruments).

Genome-wide and focused CRISPR-Cas9 libraries

See supplementary methods on page 111.

Lentiviral production and functional titration

See supplementary methods on page 111.

Genome-wide (primary) screening

GeCKO v2 library, packaged in lentiviral particles was transduced into $100x10^6$ K562 cells in 12-well plates using the same protocol described for viral titration. Each well, containing $2.5x10^6$ cells, was transduced with 10 µl of the GeCKO v2 virus which results in a MOI of 0.45 (determined from titration). Transduced cells from all wells were pooled and the non-infected cells were eliminated by puromycin selection (2µg/ml) for 7 days during which the infected cells were expanded. The pooled cell library was split into treatment and control groups (in T225 flasks) with 2 replicates for each group. In the treatment group, cells were pulsed with 2.5 mM acetaldehyde for 24 hours followed by 48 hours recovery, with a total of three pulses. Control cells were left untreated and the media were changed following the same pattern used for the treatment group. At least $25x10^6$ cells were maintained in each replicate resulting in a representation of ~ 400-fold the library size. Using this strategy, selection was applied for 9 days which corresponds to 9 K562 doublings. At the end of the screen, $25x10^6$ cells from each replicate were washed with PBS and the pellets were saved for DNA extraction.

Secondary screening using the validation library

$10x10^6$ K562 cells were transduced with the validation library in a 12-well plate using the described lentiviral transduction protocol. Each well, containing $2.5x10^6$ cells, was transduced with 2.5 µl of the validation library virus which results in a MOI of 0.27 (determined from titration). Unlike the pulse/recovery treatment protocol applied for the primary screen, selection was performed under continuous exposure to 1 mM acetaldehyde for 7 days while control cells were left untreated. For each condition, the medium was changed every 48 hours and at least $2.5x10^6$ cells were maintained resulting in a representation of ~ 900-fold the library size. Screens were performed in T25 cell culture flasks and each condition was run in triplicate. At the end of the screen, $2.5x10^6$ cells from each replicate were washed with PBS and the pellets were saved for DNA extraction.

DNA extraction, library preparation and next generation sequencing

See supplementary methods on page 112.

Data processing and computational analysis

See supplementary methods on page 113.

Generation of *OVCA2* and control knockout pools

sgRNAs targeting *OVCA2* or a non-targeting control (NTC) sgRNA (sequences shown in Table S2) were cloned into the CRISPR lentiviral backbone vector lentiCRISPRv2 using Gibson assembly. To produce lentiviral particles, we co-transfected HEK293T cells in T25 culture flasks with 3.4μg of the target vector, 2.6μg of the packaging plasmid (psPAX2) and 1.7μg of the envelope plasmid (pMD2.G) using Lipofectamine 2000 and Plus reagent (Thermo Fisher Scientific) following the manufacturer's instructions. Viral solutions were collected 60 hrs post transfection, filtered through a Steriflip-HV 0.45 μm low protein binding PVDF membrane (Millipore), and concentrated 50 folds using Lenti-X Concentrator (Clontech) following the manufacturer's protocol. Cells were transduced with targeting or NTC vectors at a MOI < 0.5 and the transduced cells were enriched by puromycin selection. The obtained KO cellular pools were used in cytotoxicity assays to evaluate their sensitivity to acetaldehyde.

Measuring N2-ethylidene-2'-deoxyguanosine (N2-ethylidine-dG) DNA adduct levels

Evaluation of N2-ethylidene-dG adduct accumulation was performed using previously described methods (Balbo et al., 2016; Chen et al., 2007) with some modifications. Wild-type (WT), NTC and OVCA2 knockout K562 pools were either treated with 5mM acetaldehyde or left untreated for 48 hours. For each sample, DNA was extracted from $1x10^6$ cells using the Gentra Puregene Cell kit (Qiagen) following the manufacturer's protocol. DNA pellets were suspended in 100 μL water. [$^{15}N_5$]N2-ethyl-2'-deoxyguanosine ([$^{15}N_5$]N2-Et-dG) standard was prepared as previously described (Wang et al., 2006). 400 μL of 10 mM Tris/5 mM $MgCl_2$ buffer containing [$^{15}N_5$]N2-Et-dG (50 fmol) and sodium cyanoborohydride ($NaBH_3CN$, 30 mg) were added to DNA solutions, and the pH was adjusted to 7 with 0.12 N HCl. $NaBH_3CN$ in the buffer reduces N2-ethylidene-dG into the more stable N2-Et-dG. For DNA digestion, the mixtures were initially incubated with 626 units of DNase I (Sigma) overnight at room temperature. Additional 626 units of DNase I, 32.5 mU of phosphodiesterase I (Sigma) and 225 units of alkaline phosphatase (Roche) were added and the mixtures were incubated at 37 °C for 70 min and then overnight at room temperature. Enzymes were removed using Centrifree ultrafiltration device (Millipore) following the manufacturer's protocol. A 10 μl aliquot was removed from each sample for 2′-deoxyguanosine (dG) quantification. Samples were desalted and purified using a solid-phase extraction (SPE) cartridge [Strata-X 33 μm, 30 mg/ml (Phenomenex)]). The 70% CH_3OH fraction (1 mL) was collected, evaporated to dryness, dissolved in 250 μL H_2O, and purified using a mixed mode, anion exchange reversed phase extraction cartridge [Oasis MAX, 30 mg/cartridge (Waters)]. Adducts were eluted with 500 μL of 70% CH_3OH, and the solution was evaporated to dryness. Samples were reconstituted in 10 μL and 1 μL was analyzed by liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS/MS). A blank containing all the reagents without DNA was prepared to check for instrument baseline and possible contamination. Samples containing 50 μg calf thymus DNA (Worthington Biochemical) were processed similarly and used as controls to determine inter-day precision and accuracy. LC-ESI-MS/MS analysis was performed on an UltiMate 3000 Ultra-high-performance liquid chromatography (UHPLC) System (Thermo Fisher) with a 250 x 0.5 mm Luna C18 100A column (Phenomenex) and a TSQ Vantage (Thermo Fisher) triple quadrupole mass spectrometer. The solvent elution program was a 10 μL/minute gradient from 5 to 35% CH_3OH in 30 minutes, followed by a wash at 95% CH_3OH for 5 minutes and re-equilibration at 5% CH_3OH for 8 minutes. The ESI source was set in the positive ion mode as follows: voltage, 3.0 kV; heated ion transfer tube, 300 °C. The collision energy was 12 eV, and the Ar collision gas

pressure was 1.3 mTorr. Adducts were quantified by MS/MS with selected reaction monitoring (SRM) at m/z 296 → m/z 180 ($[M + H]^+ \rightarrow [BH^+]$) for N2-Et-dG, and at the corresponding transition m/z 301 → m/z 185 for $[^{15}N_5]$N2-Et-dG. A calibration curve was derived by injecting standard solutions prepared by mixing a 5 fmol of $[^{15}N_5]$N2-Et-dG with various amounts of N2 - Et-dG (0, 0.25, 0.5, 1, 2, 3, 4, 20, and 30 fmol). The amount of N2-Et-dG (fmol) is determined from the ratio of the peak area of N2-Et-dG to that of $[^{15}N_5]$N2-Et-dG and normalized to dG amounts which were quantified for each sample. Quantitation of dG was carried out on an UltiMate 3000 UHPLC System (Thermo Fisher) with a UV detector set at 254 nm. A 250 x 0.5 mm Luna C18 100A column (Phenomenex) was used with a flow rate of 10 µl/min and a gradient from 5% to 20% CH_3OH in H_2O over the course of 12 minutes followed by a 10 minutes hold at 20% CH_3OH. The column was then washed with 95% CH_3OH for 5 minutes and re-equilibrated to initial conditions for 15 minutes.

Results

Genome-wide screening revealed novel candidate genes involved in cellular response to acetaldehyde toxicity

We used the CRISPR-Cas9 genome-wide knockout screening approach to study mechanisms influencing the toxic response to acetaldehyde in human K562 erythroleukemic cells. A pool of mutants was generated by transducing cells with the GeCKO v2 sgRNA library. Screening was performed to identify genes whose loss alters sensitivity to acetaldehyde. Cell viability assays were carried out to determine a sub-lethal concentration of acetaldehyde to be used as a stressor in the screen. K562 cells exhibited a dose-dependent reduction in viability upon treatment with a range (0-50mM) of acetaldehyde concentrations for 48 hours (Figure 5.1B). However, even the lowest acetaldehyde concentration (2.5 mM), that leads to 15% reduction in cell viability after 48 hours, resulted in complete cell death after 7 days of exposure (data not shown). To overcome this problem, screening was performed by exposing cells to 2.5 mM acetaldehyde for 24 hours followed by a 48-hour recovery, with a total of 3 pulse-recovery cycles (9 days). This approach resulted in 50% reduction in cell viability in the treated pools by the end of the screen (data not shown). To assess the depletion and enrichment of each mutant in the acetaldehyde-treated pool, the representation of each guide sequence, determined by next generation sequencing, was compared between acetaldehyde-treated and control conditions. Since each gene is only targeted by 3 guide sequences in the library, a single guide sequence showing differential abundance between treatment and the control pools was sufficient to consider the corresponding gene a candidate. Sufficiently represented guide sequences (average log2 CPM > 2.5) displaying significant differential abundance between acetaldehyde and control pools (FDR <0.1) were considered for candidate gene selection (Figure 5.2A). Using these criteria, we identified 40 genes that are potentially involved in modulating the toxic response to acetaldehyde (Table S17). Guide sequences targeting 21 candidate genes were depleted from acetaldehyde-treated pools indicating increased sensitivity of the corresponding mutants to acetaldehyde (Figure 5.2B). In contrast, 19 of the identified mutants are potentially resistant to acetaldehyde as the guide sequences targeting the corresponding genes were enriched in the acetaldehyde pool (Figure 5.2C).

Secondary screening validated the protective role of DNA repair in acetaldehyde toxicity

A customized sgRNA library (validation library), that is enriched for hits identified by the genome-wide screen, was used in secondary screening. In the validation library, each gene is targeted with 6-8 sgRNAs. In addition to the primary candidate genes, the validation library targets multiple control genes, that were not identified by the primary screen as candidates, and includes several non-targeting sgRNAs that serve as negative controls. We transduced K562 cells with the validation library and screened the obtained pool for mutants with altered sensitivity to acetaldehyde. Unlike the primary screen where a pulse treatment-recovery approach with 2.5mM acetaldehyde was applied, secondary screening was performed by exposing cells to a lower acetaldehyde concentration (1mM) continuously for 7 days. The applied dose resulted in 30% reduction in cell viability in the treated pools compared to untreated controls at the end of the screen (data not shown). To assess the depletion and enrichment of each mutant in the acetaldehyde-treated pool, the representation of each guide sequence, determined by next generation sequencing, was compared between acetaldehyde-treated and control conditions. The secondary screen predominantly validated genes whose disruption increases sensitivity to acetaldehyde as the majority guide sequences displaying significant

differential representation (FDR < 0.05) exhibited lower abundance in acetaldehyde compared to control pools (log2 FC < 0) (Figure 5.3A; Table S18). Among the genes validated with multiple sgRNAs were those involved in DNA repair including *HELQ*, *ERCC8* and *UVSSA* in addition to *PPP4R2*, another DNA repair gene that was only revealed by the secondary screen. Guides corresponding to these genes were partially depleted from the acetaldehyde pools indicating increased sensitivity to acetaldehyde (Figure 5.3B).

Inactivation of OVCA2 confers increased sensitivity to acetaldehyde

One of the top candidate genes revealed in our screen encodes the tumor suppressor OVCA2. Increased sensitivity to acetaldehyde resulting from OVCA2 loss was validated in the secondary screen where all the guide sequences targeting the *OVCA2* gene were significantly depleted in acetaldehyde-treated pools compared to untreated controls (Figure 5.4A). To further confirm the role of OVCA2 in acetaldehyde tolerance, we targeted the *OVCA2* gene in K562 cells using the CRISPR-Cas9 gene knockout tool. We generated two independent OVCA2 mutant pools by transducing cells with lentiCRISPR vectors expressing different sgRNAs targeting *OVCA2* and evaluated sensitivity to acetaldehyde by cell viability assays. OVCA2 mutant pools displayed decreased tolerance to acetaldehyde compared to control pools transduced with a non-targeting sgRNA (Figure 5.4B).

OVCA2 is potentially involved in the repair of acetaldehyde-induced DNA adduct

Acetaldehyde exposure leads to the formation of DNA adducts. One of the major acetaldehyde DNA adducts, *N*2-ethylidene-2'-deoxyguanosine (N2-ethylidene-dG), is a product of the reaction of acetaldehyde with deoxyguanosine (Brooks and Zakhari, 2014). To study the effect of OVCA2 loss on N2-ethylidine-dG accumulation upon acetaldehyde exposure, we compared DNA adduct levels between OVCA2 mutant pools and control cells after treatment with 5mM acetaldehyde for 48 hours. Since N2-ethylidene-dG is unstable at the nucleoside level, a more stable form, N2-ethyl-2'-deoxyguanosine (N2-Et-dG), was derived by treating extracted DNA with the reducing agent sodium cyanoborohydride ($NaBH_3CN$). LC-MS/MS quantification showed accumulation of N2-Et-dG in all cell groups treated with acetaldehyde compared to untreated controls. Interestingly, OVCA2 knockout pools displayed higher adduct levels compared to wild-type and NTC pools (Figure 5.5).

Discussion

Acetaldehyde is a highly reactive byproduct of ethanol oxidation. Since acetaldehyde is produced during normal cellular metabolism, mechanisms that control its intracellular levels or reverse its effects in the cell are essential. Acetaldehyde-related adverse health effects, including hematological complications, can result from ethanol consumption, occupational exposure or defective cellular detoxification mechanisms. A few cellular processes, including metabolism by the aldehyde dehydrogenase (ALDH) and DNA damage response through the Fanconi Anemia-BRCA (FANC-BRCA) pathway, have been shown to mitigate acetaldehyde genotoxicity. However, comprehensive investigation of cellular processes influencing acetaldehyde toxicity is still required. Complete understanding of such processes and identification of novel determinants of susceptibility to acetaldehyde could ultimately lead to mitigating the risks associated with acetaldehyde exposure.

In this study, we employed the genome-wide CRISPR-Cas9 knockout screening tool to identify genes influencing susceptibility to acetaldehyde in human K562 erythroleukemic cells. To our knowledge, a genome-wide screen capturing molecular determinants of acetaldehyde sensitivity and tolerance in human cells has not yet been performed. Our primary screen revealed 21 genes that are potentially involved in cellular acetaldehyde tolerance and 19 candidate genes with a potential role in sensitizing cells to acetaldehyde (Figure 5.2B, C). Only a few of the candidate genes were validated by our secondary screening approach (Figures 5.3, 5.4). Non-validated genes are not necessarily false-positive hits. The limited number of validated genes could be due to the lower acetaldehyde concentration used in the secondary screen for selection. Such a low selective pressure in the screen would result in revealing only the genes that are indispensable for the stressor tolerance. Accordingly, the majority of hits validated by our secondary screen correspond to genes whose disruption increases sensitivity to acetaldehyde. A comprehensive functional genomics study of acetaldehyde toxicity in budding yeast (*Saccharomyces cerevisiae*) identified multiple determinants of acetaldehyde tolerance including 2 aldehyde dehydrogenases (ALD3 and ALD6), in addition to components of the oleic acid biosynthesis and pentose phosphate pathways (Matsufuji et al., 2008). Human homologs of the identified yeast genes were not revealed as candidates in our acetaldehyde screen. Despite the known protective role of aldehyde dehydrogenases, especially ALDH2, against acetaldehyde toxicity in humans and mice (Amanuma et al., 2015; Garaycoechea et al., 2012; Langevin et al., 2011), our screen did not reveal any role for aldehyde dehydrogenases in acetaldehyde tolerance in K562 cells. Consistently, disruption of ALDH2 in the chicken cell line DT40 did not result in increased sensitivity to acetaldehyde treatment (Langevin et al., 2011). These observations indicate that multiple compensatory mechanisms of acetaldehyde breakdown could be present in certain cell types.

Consistent with the established role of acetaldehyde in DNA damage (Brooks and Zakhari, 2014), the majority of genes confirmed by the secondary screen as determinants of acetaldehyde tolerance encode components of DNA repair (Figure 5.3B). Previous reports on DNA repair pathways that mitigate acetaldehyde genotoxicity revealed a major contribution of the FANC-BRCA pathway to acetaldehyde detoxification (Garaycoechea et al., 2012; Langevin et al., 2011; Tacconi et al., 2017). Mouse hematopoietic stem cells (HPCs) lacking *Fanca or Fancd2* genes display a considerable increase in sensitivity to acetaldehyde exposure (Garaycoechea et al., 2012). Similarly, chicken lymphoid DT40 cells lacking individual components of the FA

pathway are more vulnerable to acetaldehyde toxicity (Langevin et al., 2011). In addition, inactivation of FANCD2, BRCA-1 or BRCA-2 in non-small cell lung carcinoma H1299 cells, and colorectal adenocarcinoma DLD1 cells, leads to a substantial increase in sensitivity to acetaldehyde exposure (Tacconi et al., 2017). Our screen did not reveal any component of the FANC-BRCA pathway as determinant of acetaldehyde tolerance, probably due to functional redundancy between different components of this pathway in K562 cells. We identified novel DNA repair components whose disruption enhances cellular sensitivity to acetaldehyde. These include the helicase HELQ, which is involved in DNA interstrand crosslink (ICL) repair (Takata et al., 2013), and PP4R2, the regulatory subunit of protein phosphatase 4 (PP4) that is generally required for efficient DNA repair (Chowdhury et al., 2008; Herzig et al., 2017). It is possible that the role HELQ in ICL repair underlies the revealed tolerance to acetaldehyde. Consistent with this speculation, the reaction of acetaldehyde with DNA was shown to induce ICL formation (Liu et al., 2006; Wang et al., 2000). Additionally, the FANC-BRCA pathway, which is known to counteract acetaldehyde-induced DNA damage, is primarily involved in repairing ICLs (Kim and D'Andrea, 2012). Phosphorylation of the histone H2AX (which becomes γ-H2AX) is an early response to DNA ICLs that allows recruitment of multiple repair factors, including FANC-BRCA components, to the damaged loci (Clingen et al., 2008; Wang, 2007). A subsequent dephosphorylation of γ-H2AX by PP4 is required for efficient DNA repair (Chowdhury et al., 2008). Hence, PP4R2 could potentially contribute to the repair of acetaldehyde-induced ICLs consistent with a role in acetaldehyde tolerance. Additional validated genes encode ERCC8 and UVSSA, which are components of the transcription-coupled nucleotide excision repair (TC-NER) pathway (Fousteri and Mullenders, 2008; Nakazawa et al., 2012). A direct role for TC-NER in the removal of acetaldehyde-derived DNA adducts has not been reported. However, TC-NER has been recognized as an adduct-repairing pathway due to its role in detoxifying adduct-forming agents like cisplatin (Furuta et al., 2002). Collectively, our findings suggest that multiple DNA repair pathways can contribute to mitigating acetaldehyde genotoxicity. In line with this suggestion, a comprehensive study of DNA repair genes in fission yeast (*Schizosaccharomyces pombe*) revealed roles for components belonging to the FA, HR, NER, BER, TLS and DPC repair pathways in acetaldehyde tolerance (Noguchi et al., 2016).

We identified a novel acetaldehyde tolerance gene, *OVCA2*, whose inactivation results in a notable increase in sensitivity to acetaldehyde (Figure 5.4). *OVCA1* and *OVCA2* are ubiquitously expressed tumor suppressor candidates whose expression is considerably reduced in ovarian cancer as a result of allelic loss in the corresponding locus (Schultz et al., 1996). Unlike *OVCA1*, ectopic expression of *OVCA2* in ovarian cancer cells does not suppress cell proliferation (Bruening et al., 1999). OVCA2 expression is downregulated in response to retinoid treatment (Prowse et al., 2002). Biochemical characterization of OVCA2 demonstrated a serine hydrolase activity (Baxter et al., 2004), yet the biological function of OVCA2 has not been identified. Our results revealed for the first time a role for OVCA2 in acetaldehyde tolerance. Consistent with our finding, deletion of *FSH1*, the yeast homolog of *OVCA2*, exhibits defective growth in a medium containing ethanol, the precursor of acetaldehyde (Schlecht et al., 2014). Human *OVCA2* is located in a small region of chromosome 17 that is deleted in the vast majority of esophageal squamous cell carcinoma (ESCC) (Huang et al., 2000). Interestingly, ESCC incidence is strongly correlated to alcohol consumption and tobacco smoking, two of the most important acetaldehyde sources. (Freedman et al., 2007; Pandeya et al., 2013; Yang et al., 2017). More recently, a genome-wide association study (GWAS) identified multiple ESCC

susceptibility loci, including a locus on chromosome 17 that contains *OVCA2* (Wu et al., 2012). The same study verified the well-known association of the *ALDH2* locus to increased ESCC risk thus highlighting the role of acetaldehyde in this disease. Together, these findings suggest that *OVCA2* is an acetaldehyde detoxification gene whose loss increases susceptibility to ESCC.

Consistent with the genotoxic effect of acetaldehyde and the role of DNA repair in mitigating acetaldehyde toxicity, we speculated that a function of OVCA2 in DNA repair could underlie the tolerance to acetaldehyde. Acetaldehyde induce DNA damage mainly through the formation of DNA adducts. The major acetaldehyde-derived DNA adduct is N2-ethylidene-dG that can be experimentally quantified (Wang et al., 2006). We sought to study the effect of OVCA2 loss on cellular N2-ethylidene-dG levels. Our results show that OVCA2 disruption leads to increased accumulation of N2-ethylidene-dG in response to acetaldehyde treatment (Figure 5.5). This finding suggests that OVCA2 is involved in removal of acetaldehyde-derived DNA adducts as a mechanism of acetaldehyde detoxification. However, a role of OVCA2 in acetaldehyde metabolism is also possible, since inactivation of the acetaldehyde-metabolizing enzyme ALDH2 can also increase acetaldehyde-derived N2-ethylidene-dG levels (Amanuma et al., 2015). Hence, additional work is required to elucidate whether OVCA2-mediated acetaldehyde tolerance arises from acetaldehyde metabolism or DNA repair.

Figures

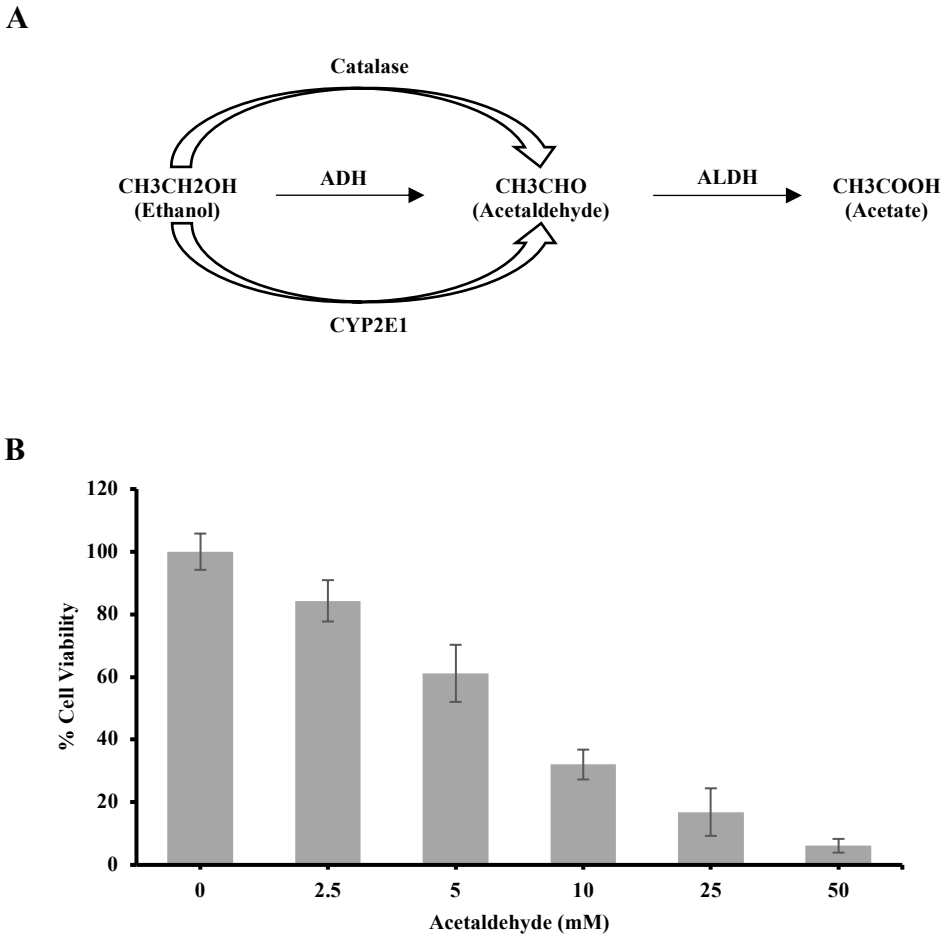

Figure 5.1: Acetaldehyde underlies alcohol-mediated cytotoxicity. (A) Overview of cellular ethanol metabolism. Transformation of ethanol to acetaldehyde can be catalyzed mainly by the alcohol dehydrogenase (ADH). The enzymes cytochrome P450 2E1 (CYP2E1) and Catalase can also break down ethanol to acetaldehyde. Further metabolism requires aldehyde dehydrogenase (ALDH) that converts acetaldehyde to acetate. (B) Cytotoxicity of acetaldehyde in K562 cells. Cells were treated with different concentrations of acetaldehyde for 72 hours and cell viability was evaluated by CellTiter Glo assay. Bars represent cell viability as percentage of untreated control. Data are represented as mean ± standard deviation (n=3).

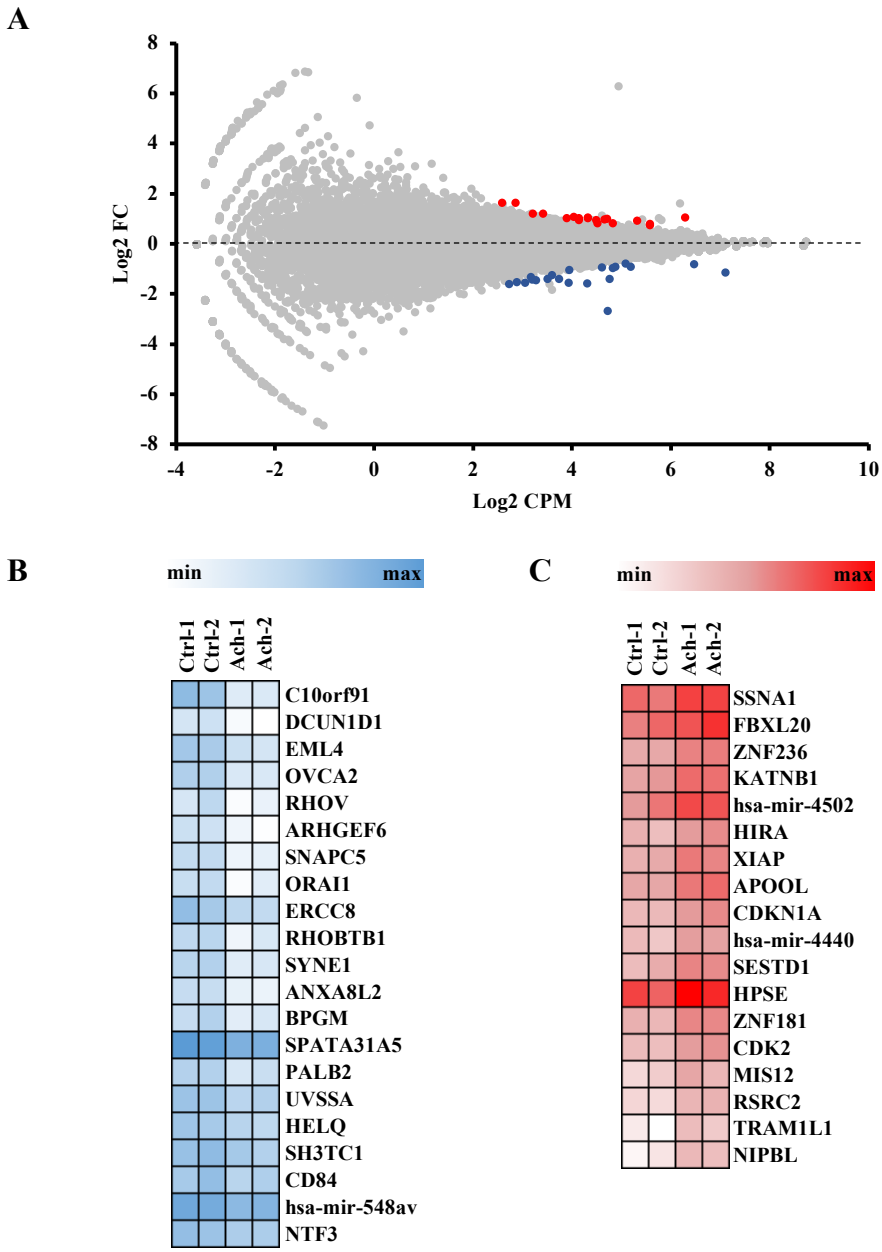

Figure 5.2: Identification of multiple candidate genes involved in acetaldehyde toxicity. (A) Scatter plot showing differential abundance of each guide sequence between acetaldehyde and control pools. Fold changes (FC) are calculated by dividing the average normalized counts (n=2) for each guide in acetaldehyde pools to that in control pools. Log2 FCs are plotted against the average abundance of each guide in all pools represented as Log2 counts per million (CPM). Guide sequences enriched in acetaldehyde relative to control (Log2 FC > 0) with FDR < 0.1 are represented as red dots. Guide sequences depleted in acetaldehyde relative to control (Log2 FC < 0) with FDR <0.1 are represented as blue dots. (B, C) Heatmaps of the normalized counts of gene-specific guide sequences that are depleted (C) or enriched (D) in acetaldehyde (Ach) compared to the untreated control (Ctrl). The screen was run in duplicate and the normalized counts for each replicate are shown in the Heatmaps.

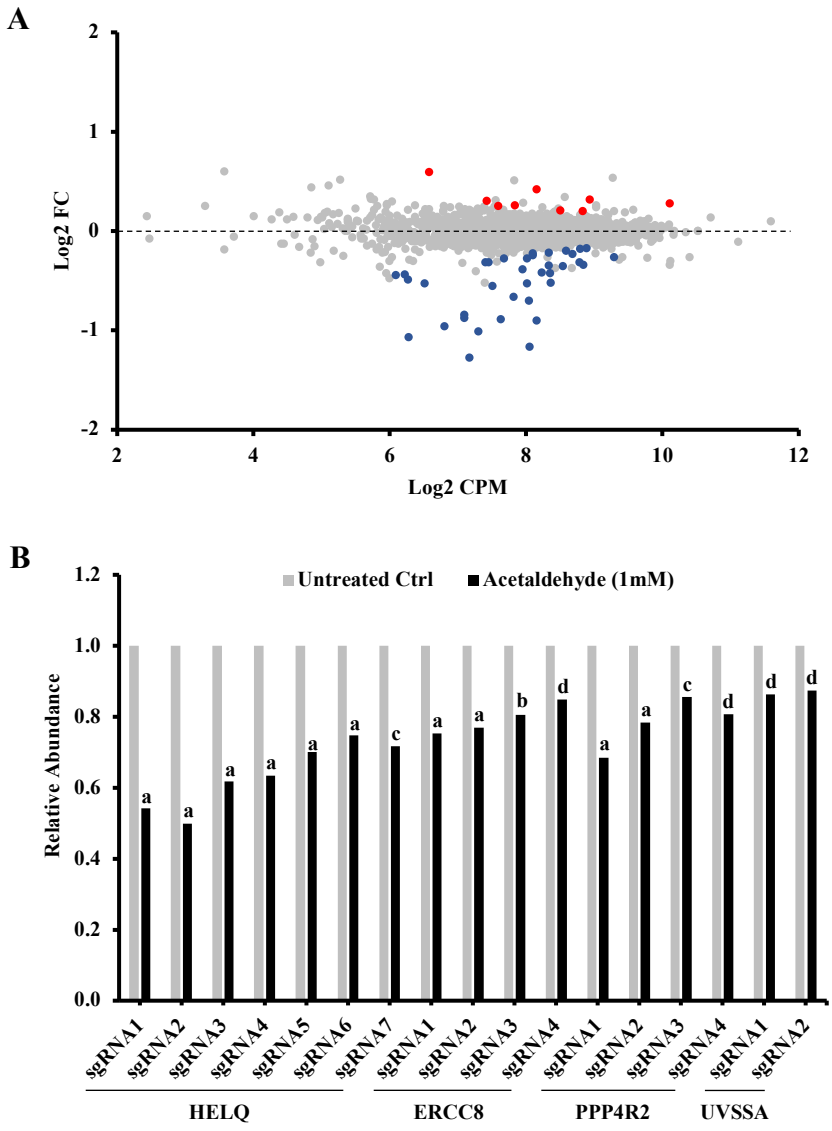

Figure 5.3: Secondary screening validates the growth disadvantage of DNA repair mutants in acetaldehyde pools. (A) Scatter plot showing differential abundance of each guide sequence between acetaldehyde and control pools. Fold changes (FC) are calculated by dividing the average normalized counts (n=3) for each guide in acetaldehyde pools to that in control pools. Log2 FCs are plotted against the average abundance of each guide in all pools represented as Log2 counts per million (CPM). Guide sequences enriched in acetaldehyde relative to control (Log2 FC > 0) with FDR < 0.2 are represented as red dots. Guide sequences depleted in acetaldehyde relative to control (Log2 FC < 0) with FDR <0.2 are represented as blue dots. (B) Partial depletion of guide sequences targeting DNA repair genes in acetaldehyde relative to control pools. Bars represent abundance of each guide sequence relative to untreated control. Data are represented as mean (n=3). FDRs corresponding to the differential abundance of each guide sequence are represented as follows: [a] = FDR < 0.001, [b] = FDR < 0.01, [c] = FDR < 0.1, [d] = FDR <0.2.

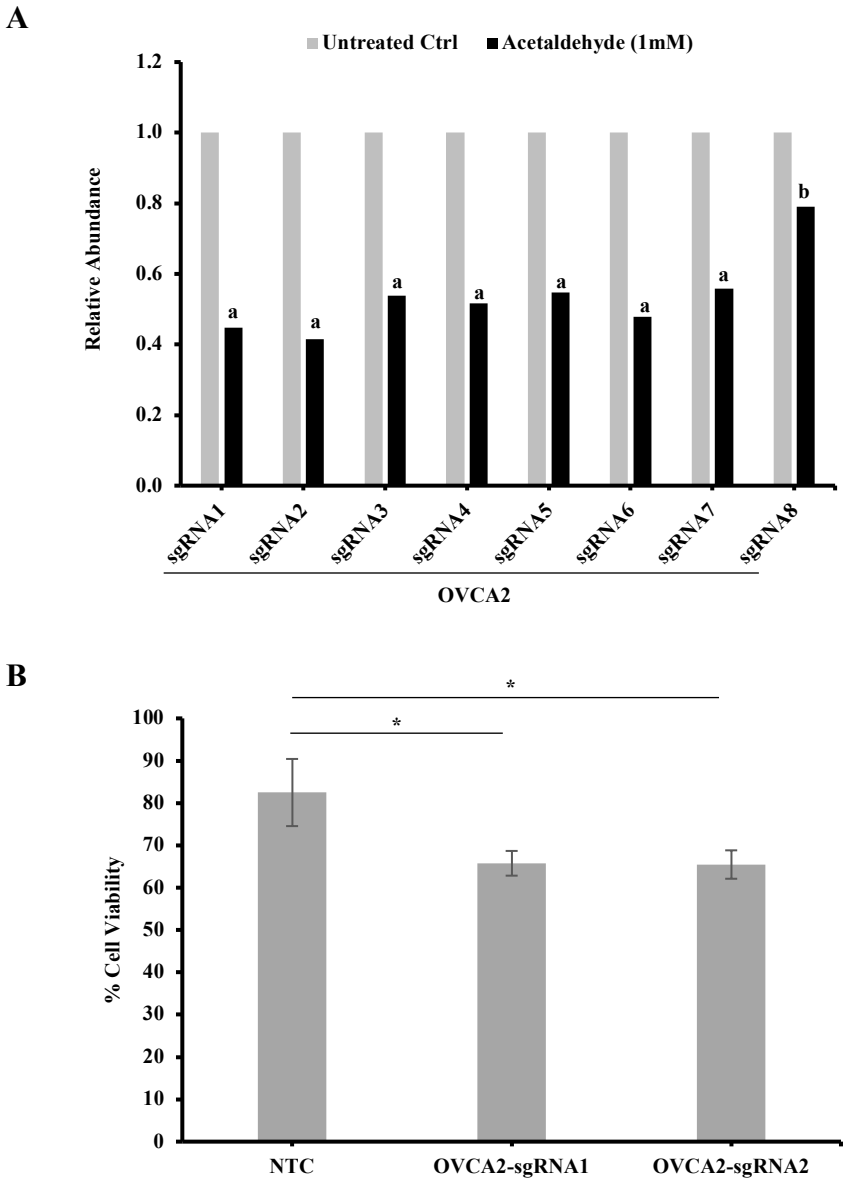

Figure 5.4: Disruption of OVCA2 decreases tolerance to acetaldehyde. (A) Depletion of guide sequences targeting OVCA2 in acetaldehyde relative to control pools (data extracted from the secondary screen). Bars represent abundance of each guide sequence relative to untreated control. Data are represented as mean (n=3). FDRs corresponding to the differential abundance of each guide sequence are represented as follows: [a] = FDR < 0.001, [b] = FDR < 0.01.
(B) OVCA2 inactivation results in increased sensitivity to acetaldehyde. K562 cells were transduced with CRISPR/sgRNA vectors targeting OVCA2 or non-targeting control (NTC) and non-transduced cells were depleted by puromycin selection. Cells were treated with 5mM acetaldehyde for 48 hours and cell viability was evaluated by CellTiter Glo assay. Cell viability is represented as percentage of untreated control. Data are represented as mean ± standard deviation (n=3). Statistical significance for the differences between NTC and OVCA2 mutant pools was determined by Student's t-test, where * = $P < 0.05$.

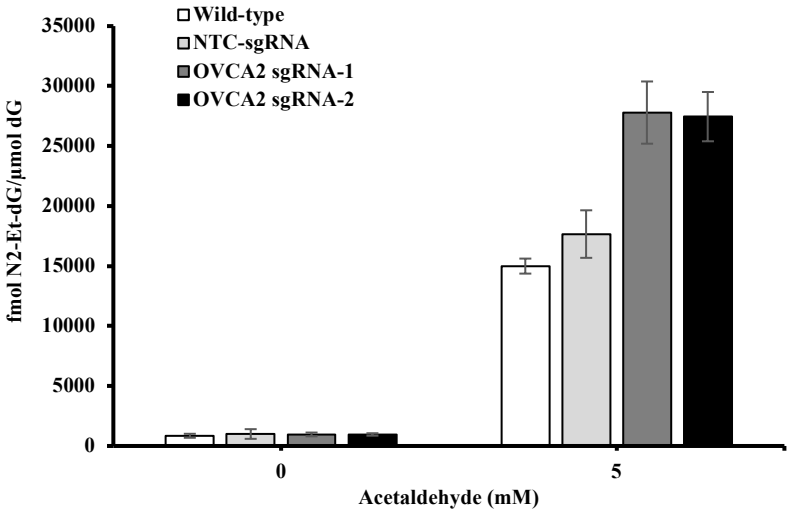

Figure 5.5: Increased accumulation of the acetaldehyde-derived DNA adduct N2-Et-dG upon OVCA2 disruption. K562 cells were transduced with CRISPR/sgRNA vectors targeting OVCA2 or non-targeting control (NTC) and non-transduced cells were depleted by puromycin selection. Cells were treated with 5mM acetaldehyde for 48 hours and untreated cells from each group served as controls. Levels of N2-Et-dG in DNA extracted from each sample were determined by LC-MS/MS. N2-ET-dG levels were normalized to the 2'-deoxyguanosine (dG) levels is each sample. Data are represented as mean ± standard deviation (n=3).

References

Amanuma, Y., Ohashi, S., Itatani, Y., Tsurumaki, M., Matsuda, S., Kikuchi, O., Nakai, Y., Miyamoto, S., Oyama, T., Kawamoto, T., et al. (2015). Protective role of ALDH2 against acetaldehyde-derived DNA damage in oesophageal squamous epithelium. Sci. Rep. *5*, 14142.

Balbo, S., Juanes, R.C., Khariwala, S., Baker, E.J., Daunais, J.B., and Grant, K.A. (2016). Increased levels of the acetaldehyde-derived DNA adduct N 2-ethyldeoxyguanosine in oral mucosa DNA from Rhesus monkeys exposed to alcohol. Mutagenesis *31*, 553–558.

Baxter, S.M., Rosenblum, J.S., Knutson, S., Nelson, M.R., Montimurro, J.S., Gennaro, J.A.D., Speir, J.A., Burbaum, J.J., and Fetrow, J.S. (2004). Synergistic Computational and Experimental Proteomics Approaches for More Accurate Detection of Active Serine Hydrolases in Yeast. Mol. Cell. Proteomics *3*, 209–225.

Brooks, P.J., and Schuebel, K. (2016). Timeless insights into prevention of acetaldehyde genotoxicity? Cell Cycle *16*, 308–309.

Brooks, P.J., and Theruvathu, J.A. (2005). DNA adducts from acetaldehyde: implications for alcohol-related carcinogenesis. Alcohol Fayettev. N *35*, 187–193.

Brooks, P.J., and Zakhari, S. (2014). Acetaldehyde and the genome: beyond nuclear DNA adducts and carcinogenesis. Environ. Mol. Mutagen. *55*, 77–91.

Brooks, P.J., Enoch, M.-A., Goldman, D., Li, T.-K., and Yokoyama, A. (2009). The Alcohol Flushing Response: An Unrecognized Risk Factor for Esophageal Cancer from Alcohol Consumption. PLOS Med. *6*, e1000050.

Bruening, W., Prowse, A.H., Schultz, D.C., Holgado-Madruga, M., Wong, A., and Godwin, A.K. (1999). Expression of OVCA1, a Candidate Tumor Suppressor, Is Reduced in Tumors and Inhibits Growth of Ovarian Cancer Cells. Cancer Res. *59*, 4973–4983.

Chen, L., Wang, M., Villalta, P.W., Luo, X., Feuer, R., Jensen, J., Hatsukami, D.K., and Hecht, S.S. (2007). Quantitation of an Acetaldehyde Adduct in Human Leukocyte DNA and the Effect of Smoking Cessation. Chem. Res. Toxicol. *20*, 108–113.

Chen, W.Y., Rosner, B., Hankinson, S.E., Colditz, G.A., and Willett, W.C. (2011). Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk. JAMA *306*, 1884–1890.

Chowdhury, D., Xu, X., Zhong, X., Ahmed, F., Zhong, J., Liao, J., Dykxhoorn, D.M., Weinstock, D.M., Pfeifer, G.P., and Lieberman, J. (2008). A PP4-phosphatase complex dephosphorylates γ-H2AX generated during DNA replication. Mol. Cell *31*, 33–46.

Clingen, P.H., Wu, J.Y.-H., Miller, J., Mistry, N., Chin, F., Wynne, P., Prise, K.M., and Hartley, J.A. (2008). Histone H2AX phosphorylation as a molecular pharmacological marker for DNA interstrand crosslink cancer chemotherapy. Biochem. Pharmacol. *76*, 19–27.

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

Freedman, N.D., Abnet, C.C., Leitzmann, M.F., Mouw, T., Subar, A.F., Hollenbeck, A.R., and Schatzkin, A. (2007). A prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. Am. J. Epidemiol. *165*, 1424–1433.

Furuta, T., Ueda, T., Aune, G., Sarasin, A., Kraemer, K.H., and Pommier, Y. (2002). Transcription-coupled Nucleotide Excision Repair as a Determinant of Cisplatin Sensitivity of Human Cells. Cancer Res. *62*, 4899–4902.

Garaycoechea, J.I., Crossan, G.P., Langevin, F., Daly, M., Arends, M.J., and Patel, K.J. (2012). Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. Nature *489*, 571.

Garaycoechea, J.I., Crossan, G.P., Langevin, F., Mulderrig, L., Louzada, S., Yang, F., Guilbaud, G., Park, N., Roerink, S., Nik-Zainal, S., et al. (2018). Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. Nature *553*, 171.

Herzig, J.K., Bullinger, L., Tasdogan, A., Zimmermann, P., Schlegel, M., Teleanu, V., Weber, D., Rücker, F.G., Paschka, P., Dolnik, A., et al. (2017). Protein phosphatase 4 regulatory subunit 2 (PPP4R2) is recurrently deleted in acute myeloid leukemia and required for efficient DNA double strand break repair. Oncotarget *8*, 95038–95053.

Hira, A., Yabe, H., Yoshida, K., Okuno, Y., Shiraishi, Y., Chiba, K., Tanaka, H., Miyano, S., Nakamura, J., Kojima, S., et al. (2013). Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. Blood *122*, 3206–3209.

Huang, J., Hu, N., Goldstein, A.M., Emmert-Buck, M.R., Tang, Z.Z., Roth, M.J., Wang, Q.H., Dawsey, S.M., Han, X.Y., Ding, T., et al. (2000). High frequency allelic loss on chromosome 17p13.3-p11.1 in esophageal squamous cell carcinomas from a high incidence area in northern China. Carcinogenesis *21*, 2019–2026.

Johnson, K.A., Fink, S.P., and Marnett, L.J. (1997). Repair of propanodeoxyguanosine by nucleotide excision repair in vivo and in vitro. J. Biol. Chem. *272*, 11434–11438.

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

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

Liu, X., Lao, Y., Yang, I.-Y., Hecht, S.S., and Moriya, M. (2006). Replication-coupled repair of crotonaldehyde/acetaldehyde-induced guanine-guanine interstrand cross-links and their mutagenicity. Biochemistry (Mosc.) *45*, 12898–12905.

Matsufuji, Y., Fujimura, S., Ito, T., Nishizawa, M., Miyaji, T., Nakagawa, J., Ohyama, T., Tomizuka, N., and Nakagawa, T. (2008). Acetaldehyde tolerance in Saccharomyces cerevisiae involves the pentose phosphate pathway and oleic acid biosynthesis. Yeast *25*, 825–833.

McKillop, I.H., and Schrum, L.W. (2009). Role of alcohol in liver carcinogenesis. Semin. Liver Dis. *29*, 222–232.

Nakazawa, Y., Sasaki, K., Mitsutake, N., Matsuse, M., Shimada, M., Nardo, T., Takahashi, Y., Ohyama, K., Ito, K., Mishima, H., et al. (2012). Mutations in UVSSA cause UV-sensitive syndrome and impair RNA polymerase IIo processing in transcription-coupled nucleotide-excision repair. Nat. Genet. *44*, 586–592.

Niemelä, O. (1999). Aldehyde-protein adducts in the liver as a result of ethanol-induced oxidative stress. Front. Biosci. J. Virtual Libr. *4*, D506-513.

Noguchi, C., Grothusen, G., Anandarajan, V., Martínez-Lage García, M., Terlecky, D., Corzo, K., Tanaka, K., Nakagawa, H., and Noguchi, E. (2016). Genetic controls of DNA damage avoidance in response to acetaldehyde in fission yeast. Cell Cycle *16*, 45–58.

Pandeya, N., Olsen, C.M., and Whiteman, D.C. (2013). Sex differences in the proportion of esophageal squamous cell carcinoma cases attributable to tobacco smoking and alcohol consumption. Cancer Epidemiol. *37*, 579–584.

Parmar, K., and D'Andrea, A.D. (2012). Stressed Out: Endogenous Aldehydes Damage Hematopoietic Stem Cells. Cell Stem Cell *11*, 583–584.

Prowse, A.H., Vanderveer, L., Milling, S.W.F., Pan, Z.-Z., Dunbrack, R.L., Xu, X.-X., and Godwin, A.K. (2002). OVCA2 is downregulated and degraded during retinoid-induced apoptosis. Int. J. Cancer *99*, 185–192.

Schlecht, U., Suresh, S., Xu, W., Aparicio, A.M., Chu, A., Proctor, M.J., Davis, R.W., Scharfe, C., and St Onge, R.P. (2014). A functional screen for copper homeostasis genes identifies a pharmacologically tractable cellular system. BMC Genomics *15*, 263.

Schultz, D.C., Vanderveer, L., Berman, D.B., Hamilton, T.C., Wong, A.J., and Godwin, A.K. (1996). Identification of Two Candidate Tumor Suppressor Genes on Chromosome 17p13.3. Cancer Res. *56*, 1997–2002.

Secretan, B., Straif, K., Baan, R., Grosse, Y., Ghissassi, F.E., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., et al. (2009). A review of human carcinogens—Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. Lancet Oncol. *10*, 1033–1034.

Seitz, H.K., and Stickel, F. (2007). Molecular mechanisms of alcohol-mediated carcinogenesis. Nat. Rev. Cancer *7*, 599–612.

Setshedi, M., Wands, J.R., and de la Monte, S.M. (2010). Acetaldehyde adducts in alcoholic liver disease. Oxid. Med. Cell. Longev. *3*, 178–185.

Tacconi, E.M., Lai, X., Folio, C., Porru, M., Zonderland, G., Badie, S., Michl, J., Sechi, I., Rogier, M., Matía García, V., et al. (2017). BRCA1 and BRCA2 tumor suppressors protect against endogenous acetaldehyde toxicity. EMBO Mol. Med. *9*, 1398–1414.

Takata, K., Reh, S., Tomida, J., Person, M.D., and Wood, R.D. (2013). Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs. Nat. Commun. *4*, 2338.

Venton, G., Pérez-Alea, M., Baier, C., Fournet, G., Quash, G., Labiad, Y., Martin, G., Sanderson, F., Poullin, P., Suchon, P., et al. (2016). Aldehyde dehydrogenases inhibition eradicates leukemia stem cells while sparing normal progenitors. Blood Cancer J. *6*, e469.

Wang, W. (2007). Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. Nat. Rev. Genet. *8*, 735–748.

Wang, M., McIntee, E.J., Cheng, G., Shi, Y., Villalta, P.W., and Hecht, S.S. (2000). Identification of DNA Adducts of Acetaldehyde. Chem. Res. Toxicol. *13*, 1149–1157.

Wang, M., Yu, N., Chen, L., Villalta, P.W., Hochalter, J.B., and Hecht, S.S. (2006). Identification of an Acetaldehyde Adduct in Human Liver DNA and Quantitation as N2-Ethyldeoxyguanosine. Chem. Res. Toxicol. *19*, 319–324.

Wu, C., Kraft, P., Zhai, K., Chang, J., Wang, Z., Li, Y., Hu, Z., He, Z., Jia, W., Abnet, C.C., et al. (2012). Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. Nat. Genet. *44*, 1090–1097.

Yang, X., Chen, X., Zhuang, M., Yuan, Z., Nie, S., Lu, M., Jin, L., and Ye, W. (2017). Smoking and alcohol drinking in relation to the risk of esophageal squamous cell carcinoma: A population-based case-control study in China. Sci. Rep. *7*, 17249.

Yokoyama, A., and Omori, T. (2003). Genetic polymorphisms of alcohol and aldehyde dehydrogenases and risk for esophageal and head and neck cancers. Jpn. J. Clin. Oncol. *33*, 111–121.

Main Conclusions

The series of studies performed here provided novel mechanistic insights into vital cellular processes and stress response in erythroid cells. Defects in iron and heme homeostasis as well as genetic vulnerabilities to external insults underlie numerous red blood cell disorders. Uncovering mechanistically-relevant determinants of physiological and injurious processes in red cell precursors could ultimately lead to developing therapeutic and risk-mitigating strategies for a variety of blood disorders. Loss-of-function genetic screens, using a CRISPR-based approach, were employed to comprehensively identify gene products involved in each of the studied processes. Effectiveness of CRISPR-based screens in interrogating genomes for understanding biological and pathological cellular activities was demonstrated in chapter 1.

Our study on iron uptake (chapter 2) identified the V-ATPase assembly factor CCDC115 as a determinant of transferrin iron uptake *in vitro*. Eukaryotic V-ATPase is a multi-subunit transmembrane complex that pumps protons across endosomal/lysosomal membranes resulting in luminal acidification (Marshansky et al., 2014). Diferric transferrin is taken up by clathrin-mediated endocytosis where endosomal acidification is required to release iron from transferrin facilitating its subsequent transport to the cytosol (Muckenthaler et al., 2017). Accordingly, the role of CCDC115 in transferrin iron uptake is likely due to its contribution to assembly of the V-ATPase complex and hence endosomal acidification. CCDC115 in not involved in non-transferrin bound iron uptake indicating that endosomal acidification is dispensable for acquiring this form of iron. Further studies are still required to investigate the role of CCDC115 in erythroid iron uptake *in vivo*.

In the heme study (chapter 3), multiple cellular processes contributing to heme-induced erythroid differentiation were revealed. Identification of a role for regulatory processes, such as epigenetic modification and RNA processing, in heme-induced erythroid differentiation is not surprising, since these processes are known to play key roles in cell differentiation (Liu et al., 2012; Wu and Sun, 2006). Since the cellular influx of exogenous heme is crucial for its action, defective heme uptake impairs the induction of erythroid differentiation upon heme treatment. Consequently, screening for determinants of heme-induced differentiation revealed mechanisms of heme uptake. Our study demonstrated that components of the clathrin-mediated endocytosis and endosomal/lysosomal acidification pathways are required for exogenous heme uptake. Intriguingly, the V-ATPase assembly factor CCDC115, which we established as pivotal for transferrin-bound iron uptake in chapter 2, was revealed as a determinant of heme acquisition. Loss of CCDC115 impaired heme-induced erythroid differentiation *in vitro.* Using a fluorescent heme analog, we demonstrated that CCDC115 loss leads to atypical heme accumulation in an intracellular region, which is likely an endolysosomal compartment. Therefore, similar to transferrin iron, heme is likely internalized by receptor-mediated endocytosis, where the role of CCDC115 in luminal acidification is required for its transport across the endolysosomal membrane.

The study reported in chapter 4 comprehensively uncovered determinants of sensitivity and resistance to the anti-leukemic agent arsenic trioxide (ATO) in erythroid cells. Disruption of gene products involved in selenocysteine metabolism, diphthalmide biosynthesis, oxidative stress, membrane transport, histone modification and RNA processing alters cellular sensitivity to ATO. The multidrug resistance protein ABCC1 is known to export arsenic-glutathione

conjugates and decrease arsenic cytotoxicity (Leslie, 2012). Hence, identification of ABCC1 as a determinant of ATO resistance demonstrates the effectiveness of our screen. Inactivation of KEAP1, the negative regulator of the antioxidant transcription factor Nrf2, conferred substantial cellular resistance to ATO. This finding indicates that ATO induces oxidative stress in erythroid cells that is alleviated through Nrf2-mediated antioxidant response. According to our results, biosynthesis and utilization of the 21st amino acid selenocysteine (Sec) render the cells vulnerable to ATO. One plausible explanation of this finding is the potential ATO-induced conversion of Sec-containing thioredoxin reductase (TrxR) into an apoptotic prooxidant enzyme. Evidences supporting this model can be found in different studies (Anestål and Arnér, 2003; Anestål et al., 2008; Lu et al., 2007). According to this model, the absence of a terminal Sec residue on TrxRs precludes the ATO-induced prooxidant function and protects against ATO-induced apoptosis. Therefore, Sec biosynthesis and incorporation into TrxRs promote cellular sensitivity to ATO. Another model for Sec-dependent sensitivity to ATO is based on the regulation of free intracellular selenium levels. Selenium can reduce arsenic cytotoxicity through formation of the Seleno-bis (S-glutathionyl) arsinium ion $[(GS)_2AsSe]^-$, a compound that is actively excreted outside the cell (Carew and Leslie, 2010). Thus, functional Sec biosynthesis and utilization could deplete free selenium resulting in impaired arsenic detoxification and vulnerability to ATO. This model is supported by our finding that selenium pre-treatment protects against ATO cytotoxicity.

In chapter 5, the documented study revealed genes whose disruption alters sensitivity of erythroid cells to acetaldehyde. Identification of DNA repair enzymes as determinants of acetaldehyde tolerance confirms the genotoxic effects of acetaldehyde. The identified enzymes belong to diverse DNA repair pathways indicating that acetaldehyde can induce various types of DNA damage. An unprecedented role for the tumor suppressor OVCA2 in acetaldehyde tolerance was revealed. OVCA2 loss resulted in higher sensitivity to acetaldehyde that is associated with increased accumulation of the acetaldehyde-derived DNA adduct N2-ethylidene-dG. Thus, a role for OVCA2 in DNA repair can be proposed. Additional work is required to determine whether OVCA2-mediated acetaldehyde tolerance is due to a role in acetaldehyde metabolism or DNA repair.

Through our work, we demonstrated how CRISPR-based genetic screens can be implemented to investigate a variety of biological processes. We provided novel insights into mechanisms governing or influencing physiological processes and response to injury in erythroid cells. While this work generated many hypotheses, only few follow-up studies were performed. Further studies are needed to comprehensively elucidate the molecular mechanisms underlying each of the studied processes.

References

Anestål, K., and Arnér, E.S.J. (2003). Rapid induction of cell death by selenium-compromised thioredoxin reductase 1 but not by the fully active enzyme containing selenocysteine. J. Biol. Chem. *278*, 15966–15972.

Anestål, K., Prast-Nielsen, S., Cenas, N., and Arnér, E.S.J. (2008). Cell Death by SecTRAPs: Thioredoxin Reductase as a Prooxidant Killer of Cells. PLOS ONE *3*, e1846.

Carew, M.W., and Leslie, E.M. (2010). Selenium-dependent and -independent transport of arsenic by the human multidrug resistance protein 2 (MRP2/ABCC2): implications for the mutual detoxification of arsenic and selenium. Carcinogenesis *31*, 1450–1455.

Leslie, E.M. (2012). Arsenic-glutathione conjugate transport by the human multidrug resistance proteins (MRPs/ABCCs). J. Inorg. Biochem. *108*, 141–149.

Liu, H., He, L., and Tang, L. (2012). Alternative splicing regulation and cell lineage differentiation. Curr. Stem Cell Res. Ther. *7*, 400–406.

Lu, J., Chew, E.-H., and Holmgren, A. (2007). Targeting thioredoxin reductase is a basis for cancer therapy by arsenic trioxide. Proc. Natl. Acad. Sci. U. S. A. *104*, 12288–12293.

Marshansky, V., Rubinstein, J.L., and Grüber, G. (2014). Eukaryotic V-ATPase: Novel structural findings and functional insights. Biochim. Biophys. Acta BBA - Bioenerg. *1837*, 857–879.

Muckenthaler, M.U., Rivella, S., Hentze, M.W., and Galy, B. (2017). A Red Carpet for Iron Metabolism. Cell *168*, 344–361.

Wu, H., and Sun, Y.E. (2006). Epigenetic Regulation of Stem Cell Differentiation. Pediatr. Res. *59*, 21R-25R.

Appendix 1

Supplementary Methods

Cell Culture

Human HEK293T and K562 cells lines were obtained from the biosciences divisional services cell culture facility, UC Berkeley (https://bds.berkeley.edu/facilities/cell-culture). HEK293T cells were cultured in DMEM (Thermo Fisher) supplemented with 10% fetal bovine serum (FBS; Corning) and 1% penicillin/streptomycin (PS; Thermo Fisher). K562 cells were cultured in RPMI 1640 (Thermo Fisher) supplemented with 10% FBS and 1% PS. Cells were cultured in a humidified incubator with 5% CO2 at 37° C.

Genome-wide and focused CRISPR/Cas9 libraries

The human genome-wide CRISPR knockout (GeCKO) version 2 sgRNA library cloned in LentiCRISPR v2 vector (Addgene # 1000000048, kindly deposited by Dr. Feng Zhang) was used for genome-wide screening. The GeCKO v2 library targets 19,050 protein-coding genes and 1,864 miRNAs and contains 1000 non-targeting sgRNAs with a total of 123,411 sgRNAs split into 2 half libraries A and B (Sanjana et al., 2014). For primary screening, we used half library A containing 65,383 sgRNAs with an average of 3 sgRNAs targeting each gene. For the focused (validation) library, sgRNA designs targeting each selected gene were picked up from the GeCKO v2 (half-libraries A and B) and the Brunello library (Doench et al., 2016). The validation library targets 307 genes (6-8 sgRNAs/gene) and contains 500 non-targeting sgRNAs, with a total of 2784 sgRNAs. Pooled custom oligonucleotides (79 bp) comprised of the 20 bp sgRNA sequence and the appropriate upstream (5'-cttGTGGAAAGGACGAAACACCg-3') and downstream (5'-gttttagagctaGAAAtagcaagttaaaataaggct-3') flanking sequences were synthesized by CustomArray pooled oligo synthesis service (CustomArray Inc., Bothell, WA). The obtained full-size oligos were PCR amplified, gel-purified and cloned into the LentiCRISPRv2 vector (Addgene # 52961) using Gibson assembly as previously described (Shalem et al., 2014). Both GeCKO v2 and validation libraries were transformed into Endura electrocompetent cells (Lucigen) using previously described protocols for library amplification (Sanjana et al., 2014; Shalem et al., 2014). Transformation efficiency for each library ensured sufficient representation of all constructs (~150-fold library size for GeCKO v2 and ~500-fold library size for validation library). Plasmid DNA was isolated from amplified colonies using the Maxiprep plasmid DNA purification kit (Qiagen).

Lentiviral production and functional titration

Lentivirus production was performed as previously described (Shalem et al., 2014), with minor modifications. Briefly, HEK293T cells cultured in a T225 flasks were co-transfected with 20 µg of the plasmid library, 15 µg of the packaging plasmid (psPAX2, Addgene # 12260) and 10 µg of the envelope plasmid (pMD2.G, Addgene # 12259). Media containing the virus were collected 60 hrs post transfection and filtered through a Steriflip-HV 0.45 µm low protein binding PVDF membrane (Millipore). The lentiviral supernatant was concentrated 50 folds using Lenti-X Concentrator (Clontech) following the manufacturer's protocol. Viral solutions were aliquoted and stored at -80° C until further use. To perform functional titration of the prepared viral solutions, K562 cells were suspended in transduction medium (RPMI 1640, 10% FBS, 1% PS + 8ug/ml polybrene) and seeded at a density of $1.5x10^6$ cells/ml in 12-well plates ($3x10^6$ cells per

well). Different volumes (0, 2.5, 5, 10, 15 and 20 μl) of the virus were mixed with the cell suspension in each well and the plates were centrifuged at 1000g for 2 hrs at 33° C. Transduced cells from each well were suspended in fresh media and recovered for 48 hours. For each transduction volume, cells were seeded in 96-well plate at a density of 10^5 cells/ml (10^4 cells/well; 100 μl) with or without puromycin (2 μg/ml) and maintained for 7 days during which 25 μl of cell suspension from each well were added to 75 μl of fresh media in a new replica plate every 48 hours. Following puromycin selection, cell viability in each condition was evaluated by CellTiter Glo and the multiplicity of infection (MOI) corresponding to each transduction volume was calculated by dividing the average luminescence signal from wells with puromycin by the average luminescence signal from wells without puromycin. A transduction volume corresponding to a MOI of 0.25-0.5 was used in the large-scale transduction (Figure S1).

DNA extraction, library preparation and next generation sequencing

Genomic DNA was isolated from $25x10^6$ cells (primary screen) using the Blood and Cell Culture DNA Midi kit (Qiagen) or $2.5x10^6$ cells (secondary screen) using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's protocols. Library preparation for next generation sequencing was performed as previously described (Sanjana et al., 2014), with minor modifications. For each sample, the pool of guide sequences was amplified from genomic DNA by high fidelity PCR using the Herculase II Fusion DNA Polymerase kit (Agilent). For the genome-wide screen, 150 μg genomic DNA were amplified for each sample (10μg genomic DNA/reaction; 15 reactions/sample). For the secondary screen, 15 μg genomic DNA were amplified for each sample (5μg genomic DNA/reaction; 3 reactions/sample). In addition to the appropriate amount of genomic DNA template, each PCR reaction contained 20 μl of the 5X reaction buffer, 500 nM of each of the forward and reverse primers, 1 mM dNTPs, 1 μl polymerase and an appropriate volume of water to reach a final volume of 100 μl. PCR was conducted with the following conditions: 95°C/2 min; 18 cycles of 95°C/20 s, 60°C/20 s, 72°C/30 s; followed by 72°C/3 min. Following amplification, reactions corresponding to the same sample were pooled. To prepare the samples for next generation sequencing, the obtained amplicons were further amplified using primers that include appropriate P5 and P7 illumina adapter sequences. In order to increase the diversity of the libraries, the forward primer used in the second PCR included a 5N shuffle sequence. To allow multiplexing of samples, multiple reverse primers were used in the second PCR and each primer contained a unique 8 bp index that was used to label each sample. For each PCR2 reaction, 5 μl of the first PCR product were used as a template. 7 PCR2 reactions per sample were performed for the genome-wide screen while a single PCR2 reaction was performed for each sample of the validation screen. PCR2 conditions and amplification protocol were similar to those used for PCR1 but 20 amplification cycles were applied instead of 18. For the genome-wide screen, PCR2 reactions corresponding to the same sample were pooled. Primers used in the first and second PCRs are shown in Table S1. The quality of the 358 bp PCR2 amplicon (shown in Figure S2A) was assessed on a 2% agarose gel (Figure S2B) and then using a 2100 bioanalyzer (Agilent). If necessary, unincorporated primers and non-specific products were removed from each sample using pippin prep (Sage Science). Following purification, individual samples labeled with different indices were quantified on a Qubit fluorometer (Thermo Fisher Scientific) and pooled in equimolar amounts. Pooled libraries were deep sequenced using the illumina Hiseq2500 platform (single read 50 bp) with a coverage exceeding 500 folds the size of each library.

Data processing and computational analysis

Raw FASTQ files were demultiplexed using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and processed to contain only the unique 20-bp guide sequences. To align the processed reads with the reference library, guide sequences from the library were assembled into a Burrows-Wheeler index using the Bowtie build-index function. Reads were aligned using the Bowtie aligner and the number of uniquely aligned reads for each guide was calculated. Individual guide counts or the sum of counts of all guides targeting each gene were used as an input into edgeR, where the counts were normalized using the upper-quartile method. Differential abundance of each guide or all guides targeting a gene between two different conditions was determined using the negative binomial generalized linear model (GLM) approach implemented in edgeR (Risso et al., 2014). False discovery rates (FDRs) were estimated to correct for multiple comparisons. Candidate selection was based on individual guide sequences displaying differential representation between the two conditions with FDR < 0.1. For primary screening, additional hits identified by summing all the guides targeting a gene, if any, were also considered for primary candidate selection. Prioritization of candidates for follow-up analysis was based on the number of individual guide sequences per gene that are differentially represented between the two conditions with FDR < 0.1. Alternatively, gene ranking was performed using Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout (MaGeCK), which combines ranks of guides targeting the same gene using a modified version of robust ranking aggregation (RRA) (Li et al., 2014). Interaction networks were integrated using the STRING protein-protein interaction database (http://string-db.org). Functional enrichments within the list of candidate genes were determined by Gene Ontology (GO) enrichment analysis implemented in the STRING database complemented with literature-based manual curation.

References

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

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

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902.

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

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science *343*, 84–87.

Appendix 2

Supplementary Figures

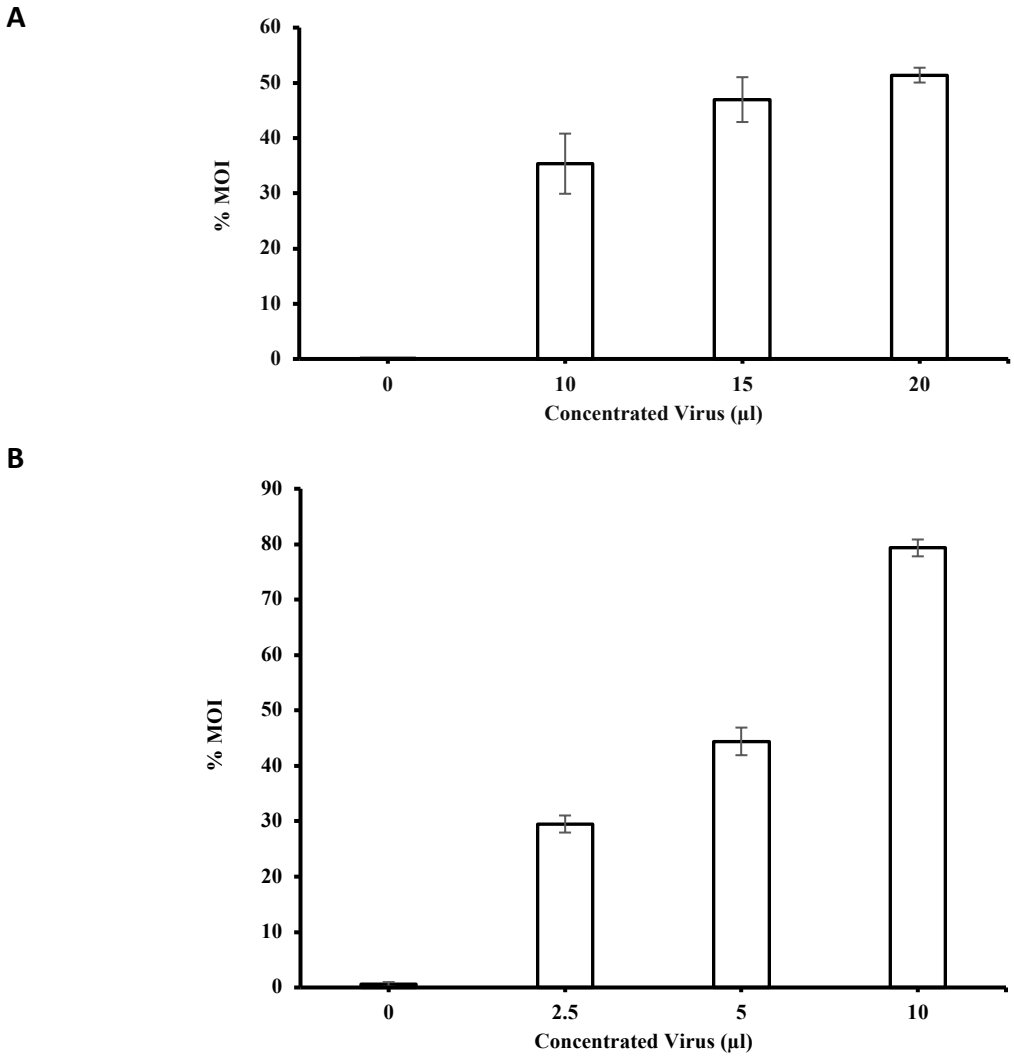

Figure S1: Functional titration of the CRISPR lentivirus libraries in K562 cells. (A) Titration of GeCKO v2 library. (B) Titration of the validation library. Cells transduced with different volumes of the viral solution were seeded in 96-well plates with or without puromycin for 7 days. Cell viability was measured by CellTiter Glo and the MOI corresponding to each transduction volume was determined by dividing the luminescent signal of replicates with puromycin by that of replicates without puromycin. Data are represented as mean ± standard deviation (n=3).

A

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**nnnnn**TCTTGTGGAAAGGAC
GAAACACCG**NNNNNNNNNNNNNNNNNNN**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT
CAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGAATTCGCTAGCTAGGTCTTGAAAGGAGTGGGAATTGGCTC
CGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAACAGTGCAGGGGAAAGAATAGTAGAAGATCGGAAGAGC
ACACGTCT*GAACTCCAGTCAC*NNNNNNNNATCTCGTATGCCGTCTTCTGCTTG

B

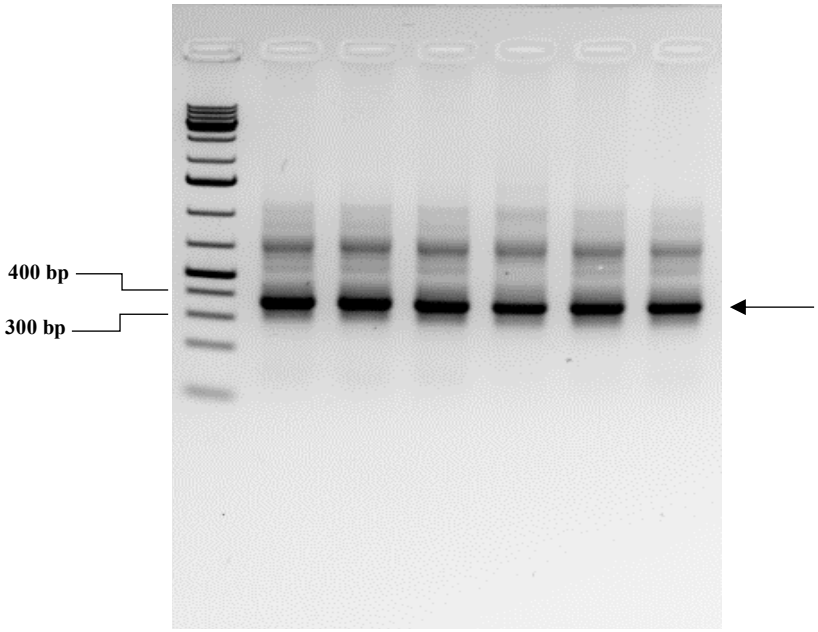

Figure S2: Detailed structure and quality check of PCR2 amplicons. (A) PCR2 amplicon sequence (size = 358 bp). Grey highlight = llumina P5 sequence + Rd1 seq primer; **nnnnn** = 5N shuffle sequence; red highlight = 50 bp read sequence; **NNNNNNNNNNNNNNNNNNN** = 20 bp guide sequence; blue highlight = Index seq primer; *NNNNNNNN* = 8 bp index (barcode); green highlight = Illumina P7 sequence. (B) Representative gel image of the 358 bp PCR2 amplicons.

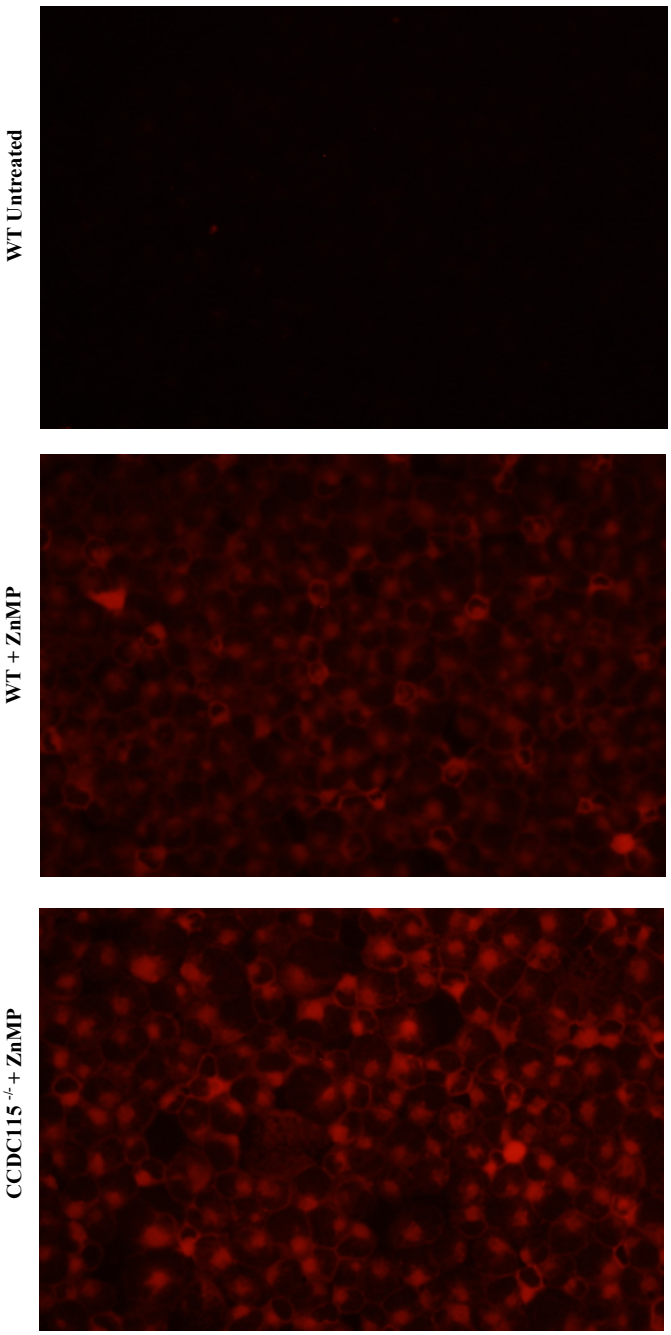

Figure S3: ZnMP uptake in wild-type (WT) and CCDC115$^{-/-}$ cells (cytospin preparations). Cells were treated with the fluorescent heme analog ZnMP (5μM) for 15 minutes and excess ZnMP was removed by washing. Immobilization of cells onto glass slides was performed by cytospin centrifugation. Levels of cellular fluorescence were evaluated by fluorescence microscopy. Images were taken at 40X magnification.

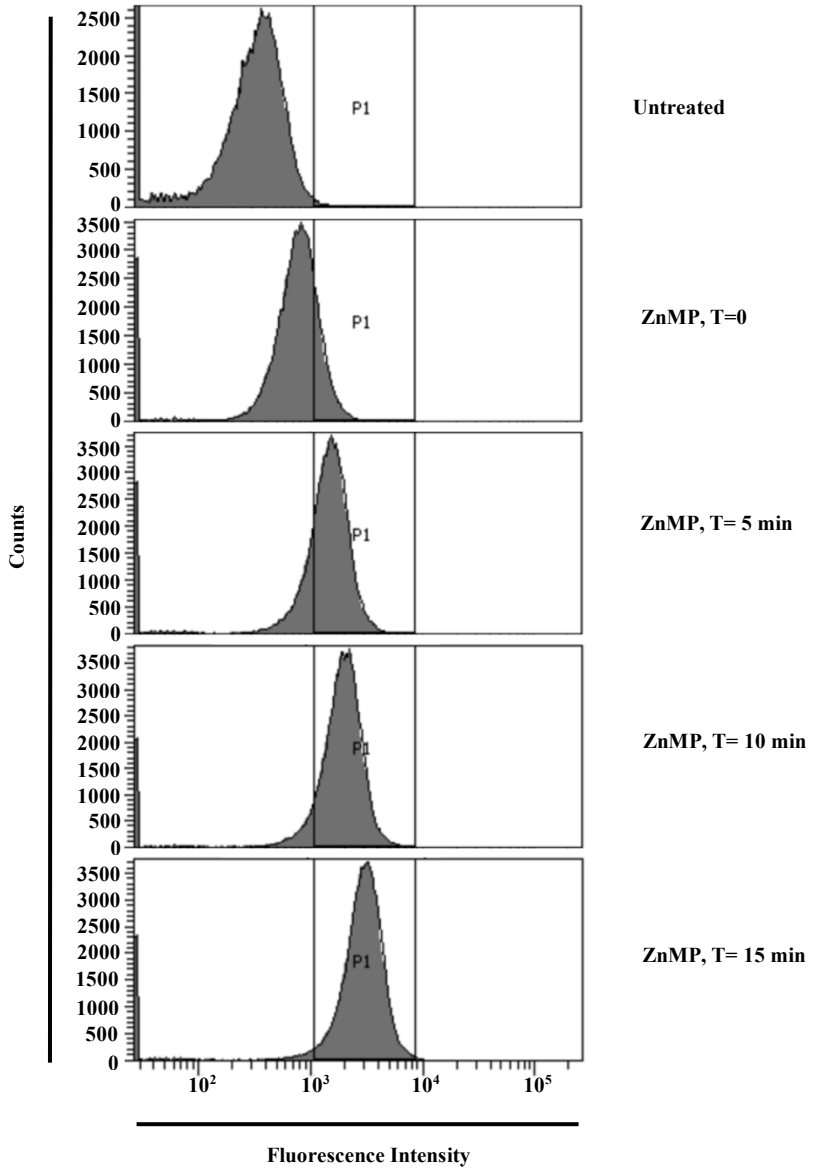

Figure S4: Time-dependent uptake of ZnMP in wild-type K562 cells. Cells were treated with the fluorescent heme analog ZnMP (5μM) for 0, 5, 10 and 15 minutes and excess ZnMP was removed by washing. Cellular ZnMP uptake was measured by flow cytometry. The gate (P1) is determined from cells treated with ZnMP for 15 minutes.

Appendix 3

Supplementary Tables

Table S1: List of PCR1 and PCR2 primers used for library preparation for next generation sequencing.

PCR1 Primers				
Forward	AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG			
Reverse	TCTACTATTCTTTCCCCTGCACTGTtgtgggcgatgtgcgctctg			
PCR2 Primers				
Forward Primer				
Illumina P5+ Rd1 seq primer			**5N shuffle**	**Priming Site**
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT			NNNNN	TCTTGTGGAAAGGACGAAACACCG
Reverse Primers				
#	**Illumina P7**	**8-bp Index**	**Index Seq Primer**	**Priming Site**
R1	CAAGCAGAAGACGGCATACGAGAT	AAGTAGAG	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	TCTACTATTCTTTCCCCTGCACTGT
R2		ACACGATC		
R3		CGCGCGGT		
R4		CATGATCG		
R5		CGTTACCA		
R6		TCCTTGGT		

Table S2: List of sequences of sgRNAs used for individual gene knockout.

Gene	**sgRNA #**	**sgRNA sequence (5' → 3')**
Non-targeting Control (NTC)	-	ACGGAGGCTAAGCGTCGCAA
TFRC	-	AAATTCATATGTCCCTCGTG
CCC115	-	AGTTCCTCACAGTCTACGTC
PSTK	1	ATCATACGCGACAACACCGA
PSTK	2	CTCTTGAGACCTGTTTACAG
OVCA2	1	GACACCAAGAGGATAAACCG
OVCA2	2	CTGAGGCACACGAGCTCGGC

Table S3: List of primers used to amplify target regions for T7 endonuclease and Cas9 in vitro digestion assays.

Gene	**Forward primer (5' → 3')**	**Reverse primer (5' → 3')**
TFRC	AGGAGCAGTTCTGCTTCAGG	CCTGAGGTCAAAAGTTCGAGA
CCDC115	CCCAGTTAATCATCCTGAGACA	TTATTTGCACGTTGCCTTTG

Table S4: List of primers used in real-time qPCR

Gene	**Forward primer (5' → 3')**	**Reverse primer (5' → 3')**
RPL19	TCCGCTGTGGCAAGAAGAAGGT	ACCGTCACAGGCTTGCGGAT
TFRC	CAATGGTTCT CCACCAAACA	GAGCCTGTGGGGAAGGG
HBA1	CAACTTCAAG CTCCTAAGCC	CTTAACGGTATTTGGAGGTCAG
HBG1	GCAGCTTGTCACAGTGCAGTTC	TGGCAAGAAGGT GCTGACTTC

Table S5: TBI/NTBI primary screen candidate genes ranked by FDR.
Candidate selection based on individual sgRNA sequences with log2 count per million (CPM) > 2.5 and showing differential abundance between TBI and NTBI with FDR < 0.1.

Gene	sgRNA sequence	log2CPM	log2FC	FDR
TRABD	GAAGCCTCTGACGGCACAAC	4.89414208	-2.424718889	9.47E-33
RPL6	GTACTCTCACCTTGCCCCTG	4.847034576	-2.216492026	2.75E-28
STIM1	CACAGTATACAATTGGACCG	5.188483669	1.384971819	1.68E-14
CCL19	AAGTTCCTCACGATGTACCC	5.174654559	1.372914016	3.50E-14
PRR22	CCGATCGTACCCATAGCGGG	5.320000045	-1.116920744	1.15E-09
EMR2	GAGAGCGAGAACACGTGTCA	4.459849209	1.645043016	1.43E-09
OTOP2	CATCTGGATGGCGGCCGTGG	4.558402875	-1.402816498	2.13E-08
CCDC115	AGTTCCTCACAGTCTACGTC	4.475529837	-1.417816308	4.52E-08
hsa-mir-1282	ATCTTGTCCTTTTTGCCCTC	6.384710635	-0.6246414	1.75E-06
GAREML	GTCTCGCTCACTGACGCCCT	3.67511633	1.5857976	1.29E-05
PRSS12	TTAGGATCAGTACGACTTCG	5.350928324	-0.84063794	5.71E-05
TFRC	AAATTCATATGTCCCTCGTG	3.858111503	-1.399218281	0.000145293
RAB40AL	CGAGTCCCCGTACAGTCACC	4.919655611	0.885386444	0.001066239
MBNL3	GTTTGCCGAGAATTTCAGCG	5.233135502	-0.788984584	0.001339261
SSTR3	CATCATCTACACGGCCGCAC	4.535977697	-1.04300008	0.001477864
INSR	CTCGAATCAGACGTAACTTC	5.167170575	-0.775294686	0.002098435
TFRC	ATCACTATAGATCCATTCAC	3.14757658	-1.519148668	0.003371022
S100A9	AGTTCATCATGCTGATGGCG	3.910008361	-1.260519052	0.006206328
KIAA0586	TGCAGTGCAAGTGTTGCCTT	6.173766828	0.498897759	0.010953438
OR10A7	ACTGGTTTGATTTGGGCCGC	5.405005935	0.659335369	0.01194204
CCDC121	TAATCAGTGCCTAAATAGAC	3.320353378	1.322273665	0.015657074
DPM1	AAGGGATGTTGCTGAACAGT	6.056043745	0.505506814	0.016482377
OR6K2	ACGAGCCATCGTCATGATTC	4.880666398	-0.777606037	0.017623418
VPS35	TTCTGATGAACTGCACTACT	6.389024319	-0.436198632	0.019000421
DYNLRB2	TATTCCCATCCGAACAACCT	2.982972736	-1.445812805	0.021529824
HELZ	TTCTCTTGGCCAGTACTCCA	4.436042034	0.928913167	0.024660709
MRPL17	TGGCCGCGTATTTCGCCGTA	3.397806191	-1.236371676	0.028483685
hsa-mir-744	ACACTGGGTTGGGCAAGGTG	4.998521345	0.722628994	0.028483685
PORCN	ACAACTTCCACCATTGACCG	4.224101739	-0.973516992	0.02883915
CALB2	GCTGACGGCGTCCCAGTTCC	4.493247393	-0.902452894	0.0303282
SLC11A1	TGCACCCACCGTAGTTATCG	5.525226175	0.588575137	0.0303282
ZNF599	GACATCAGCCATATAACGAC	3.472269125	-1.198191467	0.03170255
KLRC4	GACAAGACATATCACTGCAA	2.702993312	1.527742069	0.03170255
S100A5	CGACCAGGAGATCGACTTCA	2.653513495	-1.550641607	0.033097595
ACAD11	CCACTCCGATAGCTGTTGCA	3.994527488	1.030232014	0.034273863
TAS1R2	CCAAGTGGCGCAGCTCCGTG	4.841551831	0.756509473	0.039867965
hsa-mir-4256	CTCCACAGCCACCCCAGAAA	4.800876233	-0.76697731	0.040997317
hsa-mir-556	AGTCTAGAAGGCCTTGTTGA	4.095853105	-0.992464434	0.042087835
CDHR1	TCTCACGTATACACCCTGAA	4.592322015	-0.816135032	0.042087835
ST3GAL3	CTCTGAAAAGTTTGCGTACT	5.335739639	0.633974831	0.042087835
NUP133	TGCTCTGTCACCTATTACTA	4.357229235	0.903202778	0.042087835
MCF2L	CTTGTCCGGAATCTCGCTGA	2.705926065	-1.531449377	0.042444081
TNFSF13	TGAGCAGAGTTCCGATGCCC	4.258813611	0.96419125	0.046570865
HCST	CCAGTCCACCATGATCCATC	4.395890847	-0.855262212	0.061763042
C10orf91	CCATGTGTACTGAGTTATCC	5.143388604	0.642502949	0.06229548

IGFBP3	CCTCGCTCAGTGCGCACGTC	5.462688348	0.566081621	0.06819082
SGSM1	CAAGAACTTCCCGCCGGCTG	2.911651163	-1.339536981	0.070031076
SPZ1	ACCATTGCCTTATTCGAAAT	5.540737323	-0.540371262	0.078123532
hsa-mir-3939	GTACGTGCACGCCCACATAC	3.749343359	-1.02177037	0.078586249
ADORA2A	CTCCTCGGTGTACATCACGG	3.976961256	0.979458353	0.078586249
ARHGEF37	CCGGGAGCTCATCGACACTG	3.37862938	-1.152083139	0.087609174
CYB5B	AGGCAGTATGTCCGGTTCAA	3.828956905	-0.99747245	0.087609174
KRT33A	TCTCATTCAGGACCTGGTTC	5.180954879	0.605941966	0.087609174
hsa-mir-95	TGAAATGCGTTACATTCAAC	3.946006538	0.961081547	0.087609174
HAUS6	TCCGAAAACATTGCTGTGAA	4.536056687	0.778372455	0.090121871
CCL5	CTGAGACTCACACGACTGCT	3.806439211	0.990029007	0.093158453
KCNJ11	GCGCAAGCGCAGCGTGCCCA	3.488684506	-1.06444047	0.09666145
C8orf47	CTCCAGCAGCGCCCTCAACA	4.141356353	-0.896413918	0.09666145
VEPH1	CATGCAAAAGCCGTAGTAGA	4.402190493	0.816705843	0.09666145
TSPAN9	GGTGAACGAGAACGCCAAGA	3.452309344	-1.076671254	0.097869826
RAB3IP	TTCCGCACAATTTAAACGTC	3.91813209	-0.95306167	0.097869826
hsa-mir-592	TATGCGATGATGTGTTGTGA	5.056933517	0.62622438	0.097869826

Table S6: TBI/NTBI secondary screen candidates ranked by FDR.
Candidate selection based on individual sgRNA sequences showing differential abundance between TBI and NTBI with FDR < 0.05.

Gene	sgRNA Sequence	log2 CPM	Log2 FC	FDR
TFRC	AAATTCATATGTCCCTCGTG	8.212277	-2.26064	1.82E-12
TFRC	CTATACGCCACATAACCCCC	7.599824	-3.06445	3.66E-12
TFRC	AATTGGTGTGTTGATATACA	7.375847	-2.86754	3.66E-12
TFRC	CAGGAACCGAGTCTCCAGTG	6.526095	-3.55255	9.97E-11
CCDC115	AGTTCCTCACAGTCTACGTC	7.483923	-2.10763	7.54E-09
TFRC	TAGATCAGCATTCTCTAACT	7.246904	-1.68877	2.24E-06
SPATA31A5	ACCTCTGCAGTTGTGAAAGC	7.64459	-1.4424	8.81E-06
AQP3	CAGCACACACACGATAAGGG	8.842352	-0.98073	5.2E-05
CRISPLD2	AGGAACAACTTGTGTTACCG	7.834527	-1.07933	0.00055
TFRC	ATCACTATAGATCCATTCAC	8.401425	-0.89896	0.00065
CCDC115	GGCTTCGACCCCAGTCAATG	7.035921	-1.28584	0.001099
SLC11A2	ATGAGAACGCCACCCCACAG	7.759858	-0.91556	0.007099
ACSL3	TATCTAAAGTATCACATCCA	7.741815	-0.9315	0.007752
KATNB1	GATGGAGCCCGACTGAGAGC	8.201856	-0.75637	0.020699
CREBZF	TAAGAGAAGACGACGCCTTC	8.301445	-0.71354	0.029206
RBBP4	CACTACGACAGTGAGAAAGG	4.793556	1.958295	0.037358
LGR6	CCCATAGGGACAGCACTGGT	8.488924	-0.61763	0.041563
CTDNEP1	CCCCACCACCACACACCTTG	8.594243	-0.61479	0.041563
RNF219	GATCCTTTAACCTTGGTGCA	8.911796	-0.58812	0.041563
TNS1	CATCACGAAGCTCCATGCCA	8.677762	0.659217	0.041563
ZNF236	GCATGTGCGGTCGCACACCG	6.896929	1.004073	0.041563
PRRT2	CGTGGACGGGGCCCAGCGTC	8.62089	-0.64975	0.045387

Table S7: Hemin genome-wide screen candidate genes ranked by FDR.
Candidate selection based on individual sgRNA sequences showing differential abundance in hemin compared to vehicle control with FDR < 0.1.

Gene	sgRNA sequences	log2 CPM	log2 FC	FDR
TMEM199	TCCAAGGTAAGTGCAGACGA	3.485457547	3.885669352	1.52E-06
AP2A1	ACGCGCCGTCCACTCGCCCA	6.178576697	2.773139604	1.52E-06
CCDC115	AGTTCCTCACAGTCTACGTC	5.642584013	2.731677598	1.52E-06
AAGAB	AGCAGACATCAATCTATGTG	5.413224937	2.554627615	1.52E-06
WDR7	ATCGTAACAGGATGTCACGA	4.387871251	2.498488221	1.52E-06
ATP6AP2	ACGTCTGGGATCCGCTCTCC	4.50201545	2.384837202	4.05E-06
DENND1A	AGTGTCTGTCCATCTCAGCG	4.130783462	2.25760731	2.31E-05
DENND1A	CCCCACTTACCACGCTGAGA	5.748401946	2.07604429	2.31E-05
ATP6AP1	CGCCTCGAAGTCCACAGCAA	4.19176983	2.092681059	2.32E-05
FAM122A	CGGTCTGCACCTCCCGTTCG	5.513786856	1.90152007	4.74E-05
WDR7	GTGACATCCTGTTACGATCG	4.14679289	2.389880744	6.14E-05
AP2B1	AGTGATTGCTGCTATGACCG	3.378013733	2.004884815	0.000543347
ATP6AP1	TACCTACTTAGATCCCGCCC	5.198547002	1.79706117	0.000543347
AAGAB	GCCTTGATAGTGTCTCCTCA	5.985506112	1.645386903	0.000543347
CKMT1A	CATGCCCACAGTCTTGATGA	5.400362819	-1.732806339	0.000543347
SNX9	AATGAACTGACGGTTAATGA	4.058876978	1.57952582	0.000818992
DENND1A	TGAGCACGAATGTGAAGTTC	6.737742591	1.5584616	0.000885043
AP2M1	CTATAAGATGTGTGACGTGA	4.057831641	1.609686544	0.000894222
CCDC115	GCCCATCGCGTAGCGAGCCT	3.951593914	1.792015187	0.001044937
RNASEK	CCATTCCGCTGTGTTGATTG	4.2265317	1.439612437	0.001167383
C15orf57	TAAAGTAGCCGTCAGCACAG	2.905155792	2.233903086	0.001400542
VPRBP	GAAACTGAAGAGTATCGTAA	3.165729455	-1.953087232	0.001402667
TSC1	ATGAAAGAGTGCGTACACAC	3.345548129	-1.859308634	0.002208839
TMEM199	TCGCTCGCCCGCAAGCAAAG	3.100954521	2.266286271	0.002742126
GTF2H1	GACCGTTACAGCCATCAGTT	6.022149898	1.709079342	0.002742126
EXOC7	AGATCCAGACCTACCAGCGC	3.892136029	1.510078367	0.003975311
VMA21	GGAATGAAGGCTCACGACAG	3.20825759	1.680322195	0.004235467
LHPP	CCAACGAGTCGCAGAAGTCC	4.047801006	-1.434266609	0.00486888
POR	ACATGCCTCGCATCCCGTAG	3.889943309	-1.472683738	0.005255845
RAB35	CTTGAAATCCACTCCGATCG	3.309482997	1.603669693	0.005434045
IREB2	AACAAGACCCGTATTGAGTA	3.657014471	1.84896766	0.005520487
RPA1	AGGTACTTACGATGACTTGG	4.191262539	-1.83604289	0.005520487
RUNX1	CACTTCGACCGACAAACCTG	3.831334005	1.389873852	0.00556469
SNX9	TCTTACTTCAACGTAGTCTG	6.320095279	1.275110211	0.005568886
SMARCD2	AGCCTGTTACGACATCGATG	4.86765189	1.145811491	0.006207996
ABCF2	ATGATCAGTACGTGAAGACG	3.860116688	-1.526544949	0.006272786
AGFG1	TGCTTCGACTGCGACCAGCG	4.6268894	1.159559179	0.006430786
INTS12	AACTTACCTACAAACAACGC	3.340861358	1.525924529	0.006877585
DHX16	CTGTCGAACCCGCTCAGCAA	3.570611959	-2.264344115	0.006974726
NMUR1	CCACATAGCGTTCCACGCTC	5.832105917	1.388524416	0.008725645
FCAMR	CCACACGGTCACGATAGCGA	4.192837251	-1.195804691	0.008725645
UMODL1	CTATGAAGTGATCAGCGTCC	4.757447313	-1.235610264	0.012049171
YBEY	TGACAGGTGACGGCCACCCA	2.969763628	1.733761636	0.012727038
KPNA5	TATGCTTGAAAGTCCTATAC	5.153825213	-1.062816193	0.014554794
ARID1A	CCCCTCAATGACCTCCAGTA	3.992765446	1.216507129	0.015022663
HMOX2	CTGTCCCTTCCGAACAGCTA	5.053651184	-1.054138509	0.015259063
METTL17	GCCCGTCTTTGCAACTCCTC	3.534722621	1.375179323	0.016746878
MMACHC	GACCGTATCGCCCTACTCGA	3.38862252	1.411544259	0.017113312

CNKSR2	ATGGTCTCCGAGTCAAGTAG	4.935017841	1.230621653	0.017113312
NONO	CTGGACAATATGCCACTCCG	3.26303902	1.467087979	0.017601646
UBR4	CCGGAACCAACTTCAGTCAG	3.233494331	1.425477491	0.017601646
LYZL6	TGCGCAAAAAGGATTGTGTC	4.885909782	-1.042457751	0.022573476
CBFB	AGGCGTTCTGGAAGCGTGCC	3.623750611	1.283661452	0.024422913
RABIF	CTCATGGGAAACTCGTTCCA	4.615494203	0.998786383	0.024422913
hsa-mir-154	GGGATTTGGTACTGAAAAAT	2.741068576	1.902499282	0.024818381
CARM1	CCCGTACTCACGGCTGTAGA	4.340381663	1.00138925	0.026669635
GDI2	GCACTCTCTCCTCCGTAGTA	3.544497881	1.322161413	0.028445228
ATP6V0C	GACATGTCCGAGTCCAAGAG	0.980088517	3.074658263	0.029483695
PICALM	TGGTCTATCTCTGACCAAAG	4.777730906	0.991913728	0.029483695
UROD	CAGGCAGGCCGTTACTTACC	3.22847148	1.33241932	0.030528055
ARL8B	CACCGCGATGACATTGACGA	3.666995831	-1.249226453	0.031424597
GGT5	GGCATCGCTCGCATAAGTCC	5.023241337	-1.13573987	0.033660188
YWHAG	GCGCTACGACGACATGGCCG	4.461438995	-0.966818518	0.033878992
RPL37A	CTAGACTCACTTGTACGTCC	3.668688191	-1.463901355	0.035434741
C9orf41	TTAGCACCAGTATGCATGAG	5.271341212	0.97499492	0.036524924
AP2M1	ATCAGCTTTATCCCGCCAGA	2.9824349	1.4081107	0.036757609
PICALM	TGATATACCAGACCTTTCAC	4.887564801	0.947061611	0.036757609
ADAMTS10	CCAGACGCACACCATCGACA	3.907019115	-1.069343952	0.038576549
CPSF6	TGGGAGTGCTATTGAGACAC	3.668750922	1.189019889	0.039855228
ARL8B	CACCTTCGTCAATGTCATCG	5.01600137	-0.984042508	0.039936988
CBFB	GCCGACTTACGATTTCCGAG	2.414653691	1.825682516	0.041103908
SNAP29	CAGGAGCTCGCCCGTCAGCG	1.70034944	2.596340836	0.044423811
CNOT4	GTCTCGCAGTCCTGATGCGA	3.967831224	1.02465801	0.044423811
hsa-mir-544b	TAGTCTCAGAGCTACAGAAA	4.076630611	-1.009878042	0.044423811
CCDC115	CTACGCGATGGGCGCCAAGT	3.005515912	1.421648767	0.044526695
PPBP	AAGGGGTGGTATCAAGTCTG	4.922195492	-1.182298435	0.044526695
COMMD7	AGTGGAAAAGCACCTCTGTC	3.318192896	-1.359086379	0.048484866
NCBP1	TGATTACACAGATGATCCCG	2.159995933	2.378338556	0.050759198
SPANXN3	GAAGGCGTAGACTTATCTGA	3.255272741	1.257505395	0.050759198
RER1	TGTGCGATGGGTCGTGACAC	4.787253568	-0.967773384	0.050759198
GDI2	TCACTGACATTATACCTGAC	4.618106067	0.931708545	0.054272324
AP2A1	GCGAGGCCTTGTACAGTCGA	4.870442979	0.90585115	0.054272324
YBEY	GCAGTAAGATCGAGATTGTA	4.164772331	0.976900896	0.055423798
CHP1	CTCTTCGTCCCGCAGTAACG	4.767693855	-0.998194368	0.056570509
ABCB10	TACCATCATACTGATGCCTA	4.988866663	-1.002292871	0.057321079
EXOC8	ATTCGTCAGCTTCGCATCGA	5.297252044	0.888400523	0.058169578
ZDHHC5	AAGGATACGTGACAGCCGTG	3.916335641	-1.053465885	0.058169578
IMPAD1	GAACGAGACACAACGATCCT	4.609963425	-0.884158096	0.059597832
HTR6	CCCGTACAGCGCGTTCAGCA	4.328868607	0.927010046	0.060033882
DNAJA1	CAGCTCGTTGAAGCACTGTG	3.773868835	-1.06365463	0.065275973
RAB35	CCACCGCTTGACGTTGACAA	5.217374651	0.892473191	0.066551433
GEMIN7	TGCGGAGGTAACGCTCCCGA	3.764092362	1.025514453	0.067806306
CAMK2B	CGATGTTGGAATGCTTCAGA	3.238092348	1.16666194	0.068093874
BAG6	AGCACCTCGACATACTCCGC	4.042643927	1.040931409	0.068093874
SCAF4	CGCCGTCAACGCCTTCAACC	5.064005447	0.851273463	0.068093874
SOCS1	AGTAGAATCCGCAGGCGTCC	1.857008626	-2.015631253	0.068093874
SLC5A10	CCGGCCACCACGTCATGTCC	4.247833123	-0.966650449	0.070912878
CHKA	CCGGGATGAACTGCTCCAGT	1.307244205	-2.576805057	0.070912878
MSI1	CGGACTCGCCGCACGACCCC	4.64668736	-1.061826988	0.072817196
QPRT	CGCCTCGCACCGCTACGACC	5.232036664	-0.849401194	0.073202825
CYP26A1	CGCGGCCGCTCACGCAGTAC	3.770320205	-1.271018477	0.073202825

SRPK1	CCGCTCACCTTTCCGCTCCA	3.730700476	1.091815252	0.075801638
MYLK3	GAGTGACGACAATGACCACG	4.435832562	-0.862746639	0.077109659
FTSJ2	GCTTGAAGGCGCTTCGACAC	2.788006035	1.435769972	0.077298679
PTMS	AGGATGAAGCGGATCCCAAA	2.457091417	1.763145551	0.080084333
KMT2B	TCTGCGCTCCGATCCCAGCG	3.96475079	0.949059141	0.080084333
HNRNPD	TCTTTGTAGTGTCCCAGCTA	4.386535541	0.85332485	0.080084333
CLDN3	GCCGTGTACTTCTTCTCGCG	3.670219	-1.79700519	0.083899059
SCAF8	TAGCATACCTTGTCATCCCC	4.471693279	0.861483742	0.090768853
FOXD4L1	GAAGGGTGCCGGGGGTCGCC	1.496154455	2.206059139	0.091699995
DNAJB6	CTTTCCAAGATATCGGAAAC	3.398147245	1.205694286	0.091699995
DCLK1	ACCGGAGAAGTTCCGTTACC	5.035839638	0.866233888	0.091699995
hsa-mir-142	TTTATGGATGAGTGTACTGT	2.844170094	-1.310595088	0.091699995
YLPM1	GGCTGACCATCTACCACCTC	3.01049759	1.248722396	0.094179828
FAT2	TGACGGTATTAGCAGTGGAC	4.167758193	-0.9528327	0.094179828
HARBI1	TAGAGGGACACTAATGACCG	2.897543798	-1.421398844	0.094179828
RUNX1	GATGAGCGAGGCGTTGCCGC	3.461147184	1.058969444	0.094620242
RABEP2	AGAGGTGAGCGAGAGCACGA	3.556015808	-1.032751204	0.094620242
MRPL33	GCCCCACTTACAGAAGACCG	1.181797485	2.213662075	0.097579953

Table S8: Top 20 enrichment hits in the hemin screen based on MAGeCK analysis.

Gene Name	Used sgRNAs	FDR	Rank	Good sgRNAs	Log FC
DENND1A	3	0.002475	1	3	2.0651
CCDC115	3	0.002475	2	3	1.8144
WDR7	3	0.007426	4	3	2.3838
AP2A1	3	0.007426	3	3	1.2945
AGFG1	3	0.007426	5	3	0.96908
SNRPA	3	0.007426	6	3	1.061
TMEM199	3	0.007779	7	3	2.343
ATP6AP1	3	0.009282	8	2	1.8184
AAGAB	3	0.013366	10	2	1.6489
ATP6V1A	3	0.013366	9	3	1.1353
hsa-mir-1202	4	0.022952	11	4	0.95097
SNX9	3	0.038366	12	2	1.2645
VMA21	3	0.039985	13	2	1.6686
IPO9	3	0.055516	14	3	0.75449
FTSJ2	3	0.061716	15	3	0.94048
CBFB	3	0.090347	19	2	1.291
UROD	3	0.090347	20	2	1.3218
AP2M1	3	0.090347	16	3	1.4031
CNOT4	3	0.090347	18	3	1.0155
MOCS3	3	0.090347	17	3	0.94457

Table S9: Gene ontology (GO) analysis of enrichment hits identified in the hemin screen.

Pathway ID	Biological Process	No. of Genes	FDR
GO.0072583	clathrin-mediated endocytosis	4	0.00131
Pathway ID	**Cellular Component**	**No. of Genes**	**FDR**
GO.0030135	coated vesicle	8	4.62E-05
GO.0030136	clathrin-coated vesicle	7	4.62E-05
GO.0005905	coated pit	5	0.000358
GO.0030122	AP-2 adaptor complex	3	0.000358
GO.0030128	clathrin coat of endocytic vesicle	3	0.000457
GO.0005654	nucleoplasm	22	0.001
GO.0030120	vesicle coat	4	0.001
GO.0030132	clathrin coat of coated pit	3	0.001
GO.0031981	nuclear lumen	24	0.001
GO.0044428	nuclear part	25	0.001
GO.0045334	clathrin-coated endocytic vesicle	4	0.001
GO.0031410	cytoplasmic vesicle	13	0.00133
GO.0043231	intracellular membrane-bounded organelle	45	0.00139
GO.0030662	coated vesicle membrane	5	0.00156
GO.0030659	cytoplasmic vesicle membrane	8	0.00176
GO.0016023	cytoplasmic membrane-bounded vesicle	12	0.0021
GO.0030666	endocytic vesicle membrane	5	0.00272
GO.0070013	intracellular organelle lumen	25	0.00294
GO.0030139	endocytic vesicle	6	0.00296
GO.0030665	clathrin-coated vesicle membrane	4	0.00296
GO.0097458	neuron part	12	0.00302
GO.0044446	intracellular organelle part	36	0.00343
GO.0048475	coated membrane	4	0.00343
GO.0043229	intracellular organelle	45	0.00534
GO.0042382	paraspeckles	2	0.00626
GO.0030669	clathrin-coated endocytic vesicle membrane	3	0.00721
GO.0043227	membrane-bounded organelle	45	0.0105
GO.0043226	organelle	46	0.0159
GO.0042995	cell projection	13	0.021
GO.0044456	synapse part	7	0.0247
GO.0016514	SWI/SNF complex	2	0.031
GO.0000145	exocyst	2	0.0349
GO.0097481	neuronal postsynaptic density	3	0.0351

Table S10: ATO primary screen candidate genes obtained by individual sgRNA analysis.

Gene	sgRNA sequence	log2CPM	log2FC	FDR
KEAP1	AGCGTGCCCCGTAACCGCAT	5.498070295	2.206646081	7.9461E-28
EEFSEC	AGGTGCACCGATTATACCCG	4.76457748	2.447600338	2.85607E-23
KEAP1	GCAATGAACACCATCCGAAG	4.302124154	2.454744139	4.69759E-21
SEPHS2	CAATGCCAACGATCCACGCT	3.485022246	2.525060503	2.09322E-13
PSTK	AAACTGATCAGACACTCCGA	3.62658477	2.455067634	2.67542E-13
KRT73	GCTGAGGAGCGTGCGCGAAG	4.978811478	-4.862250309	6.68982E-12
ARID1B	GGAAGCAACCAGTCTCGATC	4.172487244	1.88339974	6.68695E-11
SLC6A12	CACCTCGGGCATCCATGACC	5.253565253	1.831354871	1.72085E-08
DLGAP5	TACGAACGAAATAGACACTT	5.00151697	-3.25056815	1.22553E-07
SET	GAGCAGCACCATGTCGGCGC	3.447588594	1.892617536	1.79028E-07
SECISBP2	TCAGACTTGTCCTCAACCGA	4.610540392	1.389909252	1.62816E-06
GNB2L1	TCCATAGTTGGTCTCATCCC	4.001243797	-2.100384517	2.30129E-06
TRABD	GAAGCCTCTGACGGCACAAC	4.083520131	-1.994811298	4.39806E-06
STMN1	ACTGGAACGTTTGCGAGAGA	3.696609718	-1.556977138	1.4338E-05
TECR	CAGGTCCCGGAAGTACAGTG	4.160829634	1.328864916	1.48671E-05
ZNF844	CTTCTTTGCGTATACACAAA	4.386705642	-1.298136636	0.000129899
DCLRE1A	CCAGTACATGCCAGATCATA	5.269096733	-1.97860754	0.000132348
CCL3L1	TGCCGTCCTCCTCTGCACCA	5.629889778	1.030557481	0.000132348
CCL3L3	TGCCGTCCTCCTCTGCACCA	5.629889778	1.030557481	0.000132348
POLR2J	CGAGTCGTTCTTGCTCTTCG	4.18738889	1.1841672	0.000252559
POLR2J2	CGAGTCGTTCTTGCTCTTCG	4.18738889	1.1841672	0.000252559
POLR2J3	CGAGTCGTTCTTGCTCTTCG	4.18738889	1.1841672	0.000252559
DYNC1LI1	AGTACTAGTAGTTTGCACAA	4.199569114	-1.183357921	0.000386268
SEPHS2	GACTGACGCGGCCGGACGTG	3.518296449	1.445307822	0.000574407
STK25	TCATCGACCGCTATAAGCGC	5.016526469	-1.276681037	0.000583291
ABCC1	CAAGTTCGTGAATGACACGA	2.982268927	-1.685501283	0.000622028
RBM5	TTGACAGCTAAAGACGCGTA	3.572158635	1.425619616	0.000712475
COX8C	CTCAACTCACCGCGGCAGAC	2.975120396	2.346778087	0.000725805
TXNDC17	TTCAGCCCCTCTCGTACGAC	4.53570388	1.009456872	0.000911489
UBE2H	TAGCATCGAGAGTAAACATG	3.728652736	-1.260702287	0.001509903
ST3GAL3	CTCTGAAAAGTTTGCGTACT	5.443564212	1.239760161	0.001550167
RRAGC	CATTCCACAACTGCCGACCT	4.253569232	1.042808792	0.002487092
TACR1	GGCTGGTTGCACGAACTGAT	3.670882593	-1.612444435	0.003329384
GDF9	GAATTTGTTGGGACGTGCCA	4.525708186	-1.130769355	0.003393369
STAG2	AGTCCCACATGCTATCCACA	2.624478359	1.700466057	0.00360915
CLEC3B	GAGATCACCGCGCAACCCGA	3.703543867	1.849140511	0.006144094
SRGAP2B	CCGTGGAGTCAGGGGAGCGT	4.201463984	1.297418483	0.006144094
SRGAP2C	CCGTGGAGTCAGGGGAGCGT	4.201463984	1.297418483	0.006144094
KRTAP10-11	ACCTGCACACGGGGTGGCAG	2.802611839	1.589091601	0.007255271
MTF1	CCTTCGTGTGCACTCGCACG	4.998343535	0.843493441	0.008850061
FADS3	CCTCGACGGATGACTCCCCC	4.876405402	0.838456642	0.011659098
PPRC1	TCAGCTCTGATCGAGTCATG	2.995025916	1.43470298	0.013201715
DPH6	TAGCTACATGTATCAGACAG	3.648736086	1.150454773	0.013291321
RETSAT	AGATGTAGCTGAGTACTGCC	4.9818765	-0.785074893	0.02091083
TRIP13	AAGTTGTCATCACATAATCG	2.596847263	1.530451875	0.02511263
CEACAM4	CCCAGTCACGATGCCAGCGA	4.657990492	0.881365296	0.02511263
VSTM2L	CGCGTCATCGACTTCAGCGA	4.08830214	1.031047611	0.0294714
CASC4	GAACAACATATCGTATCAGA	3.290221429	-1.283719118	0.032999795
MACROD1	GTGGACGCCATCGTCAACGC	4.656324831	0.844489029	0.033286069
IMPAD1	TCGAAAGTACGTCACTACTA	4.737744743	0.798103737	0.034848741

C16orf59	TTGCAATTGTCGCTGTGCGC	4.098828287	0.951119196	0.035360988
DR1	CACCAGCTCTCGAGCATCGT	4.269414644	0.877619888	0.039177222
LYPLA2	GCATCGTGGCGTTGAGCTGC	4.79477996	-0.772537365	0.042931425
ELF4	CGGGACCTTGTGCATGACGC	5.318359182	-0.845143374	0.048367996
LSMD1	ACACCACATCGTTTCCATTG	5.079166657	0.718980293	0.051131236
MRPL3	TTGCCAGGTTCCTTTAGAGT	2.740870837	1.374115465	0.051658893
PCDHA13	GACGCCCTCGACGAAGCCTG	3.458841484	1.217736264	0.051658893
PIGS	CTTGCACACTGCCCGATGAC	5.167066718	0.787484799	0.051658893
TSPAN5	ACATCGCTCTCGACTTGCAT	2.76173937	1.351626539	0.052687387
BRD8	TCTTGCTTGACCGTCATTTC	3.074021665	1.226128061	0.052687387
PPP4R2	TCACATTGTTTCTCCAGTCT	4.794162054	0.754980897	0.055053133
RHOXF2	TCTTTCTCGTCGTCGACTGC	4.870325465	-0.752360678	0.055448034
SEPSECS	CAGCCAGCTTTATAATGTCC	2.724565612	1.36670024	0.055931249
MEP1B	AGGGTGACATCAGACTTGAT	2.761836936	-1.433460595	0.055931249
MRGBP	CAGCGCCTGCATGTCGTACA	3.000099982	1.335909714	0.055931249
SRM	GTCAGTGATGATCACGTCGA	3.69544681	1.008592279	0.055931249
ATG2A	CACTGCACAGTGCGCGTGTC	3.914339003	-1.067579158	0.055931249
PCDH20	TGGGGCGTCTACATCGTCCC	4.348696532	0.910624832	0.055931249
QPRT	CGCCTCGCACCGCTACGACC	5.631937393	-0.729840852	0.055931249
BCLAF1	ACCTAGAAGATCTATATGAC	3.624311389	1.054791571	0.059988781
KIF2A	GATCCAGATCGGGATTTACG	4.456834377	-0.851230761	0.059988781
PADI1	GCTACTTACCGACGCCAGTG	3.344567589	-1.299299463	0.060704614
HSPA1B	CCCTGAGCCCCGAAGCCGCC	4.20110798	-0.884140607	0.060704614
VPS39	GGAGGTGTTACGGATGTGTG	3.724106559	0.989869257	0.060850375
FLCN	CGCCCGTTACCAGGCAAAGG	2.534616747	1.461821818	0.061161311
ACSL3	CGAGTGGATGATAGCTGCAC	3.02001565	-1.21766288	0.061161311
hsa-mir-4697	CAGGAGTCACTGACATGAAG	3.591013082	-1.025852553	0.061161311
RCE1	CCTCCGTAGGGACCATCCCG	3.992696211	0.93910948	0.061161311
SCRN1	ACCAGACCATCCTTAGCACG	4.127428034	0.867226939	0.061161311
C2orf82	CATCGTGATCGCCGCCCTGC	4.328616527	0.818753412	0.061161311
CREB3L2	TACTTACATGAGGAGCCGTG	4.444098099	-0.827080846	0.061161311
NDUFA9	TGCTGTGTTGTACTACTCGT	5.422012369	0.841232214	0.061161311
DLX6	AGGGTAGTGCGCGCCTGCCA	3.608129155	1.014010253	0.061209409
PLD6	CGGAGGACGACGAGTACGTG	3.281959153	1.145801232	0.061965073
FLOT1	CCTCCGCCTGCATAATTAGT	3.47148492	-1.057297244	0.06315919
TUBA4A	GCTGTCGGTATGGGCCATTT	3.592500149	-1.018899395	0.063815158
TOR1AIP2	AACACAAACATGGCCGACAG	3.048066504	-1.219673586	0.076539871
CREBZF	TAAGAGAAGACGACGCCTTC	4.21049139	-0.836144585	0.076539871
PRRT2	ACCGCCTAAGTTGATGACGC	4.547260365	-0.761539341	0.076539871
SLC22A18AS	CGACAGTTCTGCCGGATGAG	2.67160115	-1.319195714	0.083677625
hsa-mir-15a	TGTGCTGCTACTTTACTCCA	5.072476684	-0.686116046	0.085791659
CTDNEP1	ATCTATTTCAGGTGAGCCAG	2.89479787	1.25016555	0.087396916
DHPS	CGCCTTCTAAGATGATCGCC	2.503895862	1.370746895	0.090592729
SPATA5L1	GATACTGCTTTGTTACGACC	2.943655555	1.190914352	0.090592729
HLA-DRB1	CACGTCGCTGTCGAAGCGCA	3.246757098	-1.084886854	0.090592729
CXCR7	CCGTTTCCTTACCTCCGGGC	3.284490015	1.079350155	0.090592729
hsa-mir-6806	TTCAGCATCACATGGAACCT	3.321244138	1.066513559	0.090592729
EED	CCACGTAAGATTGCAACAAC	4.164355237	0.843759826	0.090592729
COMMD2	ATTGCAGTGGTCGCCGAGTT	4.202069523	0.849371086	0.090592729
RIPK1	GCTCGGGCGCCATGTAGTAG	4.337078348	-0.9248091	0.090592729
ANKDD1A	TCGACGCCATCGAGCAACAG	2.592544905	1.340352044	0.093764074
ATPBD4	CTCTATCGCCGAACCATAAG	3.655901784	0.996877334	0.096999534
DPH6	CTCTATCGCCGAACCATAAG	3.655901784	0.996877334	0.096999534

PSTK	CCTGTCGAAGCAATTTCCAT	3.910423766	0.879834306	0.096999534
NSMCE2	TTGTTCGCATGATTGAGTCC	3.566434086	0.986406269	0.097286862
FAM194B	CCATGCTTCTTCGTGATGCA	4.323354514	-0.787945465	0.099271844

Table S11: ATO primary screen candidate genes obtained by the analysis method based on summing all sgRNAs targeting each gene.

Gene	log2 CPM	log2 FC	FDR
KEAP1	6.195856497	2.053495734	6.87E-55
SEPHS2	4.808106915	1.772087851	2.06E-19
EEFSEC	5.602030416	1.252563233	7.97E-14
PSTK	4.804957084	1.494678508	1.77E-13
KRT73	5.42718556	-2.526887668	1.26E-11
ARID1B	4.487341569	1.42269958	1.99E-09
TXNDC17	5.458282973	0.902147195	1.00E-06
SLC6A12	6.186614468	0.918074847	2.37E-06
DCLRE1A	6.234476692	-1.058982463	1.34E-05
DLGAP5	6.159995354	-1.082764819	6.38E-05
ST3GAL3	5.920229979	0.956789344	8.67E-05
RBM5	5.236625999	0.829328757	0.000133696
DPH6	6.157267391	0.663406135	0.000133696
DYNC1LI1	5.061426224	-0.85160534	0.00014902
STAG2	3.92967649	1.137190618	0.000386437
GNB2L1	4.738147437	-1.081489558	0.000926518
SET	5.210768394	0.704160534	0.004095575
CABIN1	3.906544142	0.995937382	0.005752644
KDSR	5.763682695	0.599614752	0.005752644
QPRT	6.036316076	-0.600115149	0.005752644
ELF4	5.982243205	-0.637951733	0.006800045
SECISBP2	6.062896506	0.552038614	0.00781699
CENPE	4.664911111	0.879990739	0.008736886
POLR2J2	5.277756852	0.656831688	0.008897833
PAPOLA	3.927710734	0.947643363	0.010290839
POLR2J3	4.910510552	0.723040896	0.010290839
LSMD1	5.622455405	0.58995689	0.010290839
PIGS	5.555300657	0.648779701	0.010566121
PPAN	5.765030476	0.583153221	0.010566121
PPAN-P2RY11	5.765030476	0.583153221	0.010566121
TACR1	4.671427879	-0.825622753	0.010917316
SRM	4.982579194	0.691270293	0.012245039
PIH1D1	6.119418262	0.535739704	0.012245039
POLR2J	4.650357067	0.768996486	0.012329936
PAK1IP1	4.78250441	0.76657683	0.012329936
PEAR1	5.346187373	-0.625968303	0.013818931
RSL24D1	3.408118321	1.104632434	0.014884953
KCNA4	6.042248186	0.527510508	0.014884953
XRN2	4.745181725	0.757500941	0.01661252
DR1	5.760858476	0.542497207	0.01662935
ATPBD4	5.952957787	0.533967158	0.016735829
C15orf41	6.207056397	0.50595243	0.018853213
CCNL2	5.86682919	0.522180395	0.022876601
TSPAN5	2.904333517	1.090087654	0.023009671
GTF3A	5.163502877	0.663291342	0.023140924
MTF1	5.563624608	0.557935502	0.023179038
KIF2A	5.879581139	-0.51991088	0.023185686
ATF1	6.024256181	0.49528334	0.02585737
PSMC3	5.069390991	0.639799806	0.026706032

MACROD1	5.543679483	0.56010723	0.027380377
CDK9	5.635978527	0.536606728	0.029847077
TGM4	5.886854554	-0.504027837	0.029847077
GFI1B	4.659071034	0.734205091	0.030268646
KIF20B	3.757480422	-0.887851205	0.030414513
TECR	5.485207494	0.577488707	0.030414513
LEMD2	6.137730697	-0.57191459	0.030414513
IPO5	5.200425377	0.597840292	0.032155338
MTPN	5.785119817	-0.513623522	0.032155338
SP7	5.130901139	-0.599908894	0.036862519
PPP4R2	5.598753828	0.538430805	0.040463211
FLCN	5.01269312	0.622508647	0.04281953
PRRT2	5.81084329	-0.490335803	0.044670625
DHODH	4.599212432	0.704299614	0.046819764
ZNF432	6.662664011	0.434292646	0.047806644
RREB1	5.785005605	0.491804492	0.047952299
RRAGC	5.064358261	0.594472059	0.048382316
MRGBP	4.609343167	0.695040736	0.049281497
AQP3	5.096388883	0.59532396	0.049281497
VSTM2L	5.011539905	0.625726362	0.04953084
NDUFA9	6.501584008	0.485166551	0.051076807
DPH5	6.055418589	0.460675645	0.053177467
STIL	5.444547473	0.533829523	0.058027014
IMPAD1	6.187842353	0.445992857	0.058525513
RBBP4	4.447767976	0.704861729	0.060506339
PPIH	5.495304062	0.520142624	0.061649974
ATXN2L	5.696709467	0.49066072	0.061649974
RETSAT	5.399314909	-0.533254932	0.066134949
EIF2AK1	5.580615297	0.503732433	0.066134949
FAM46A	4.9126114	-0.606244585	0.070678704
CHM	5.325240628	0.530302905	0.077054823
CNOT2	5.787405248	-0.470956673	0.077054823
UBE2H	5.380255692	-0.526540476	0.083250528
RCE1	4.343816432	0.734539237	0.083759058
ABCC1	4.760675209	-0.62422275	0.083759058
SLC30A1	5.36154085	0.529138248	0.083759058
NCAPD3	5.091091612	-0.562010935	0.085891161
SBSN	5.660025676	0.501094236	0.087815282
TUBA1B	4.249141528	-0.709041154	0.09015983
PCDH20	4.975933116	0.618780924	0.09015983
PINX1	5.760386731	0.463859469	0.09015983
MEIS2	6.162270573	0.433295943	0.09015983
PKMYT1	3.274578997	-0.877261252	0.098595976
ZZZ3	5.271383632	0.523602726	0.098595976
HIP1	5.586848467	0.483953261	0.098595976
RBBP6	5.622966671	0.477994887	0.098595976
FLOT1	5.94036317	-0.440095948	0.098595976
TOR1A	6.209670059	0.427406246	0.098595976

Table S12: Top 10 hits (FDR<0.1) of the ATO primary screen based on MAGeCK analysis.

Gene	Used sgRNAs	FDR	Rank	Good sgRNAs
PSTK	3	0.00165	1	3
SEPHS2	3	0.00165	2	3
KEAP1	3	0.00165	3	3
DPH6	3	0.003713	4	3
PIH1D1	3	0.010891	5	3
STAG2	3	0.015677	6	3
hsa-mir-663a	4	0.028996	7	3
GFI1B	3	0.069926	8	3
TXNDC17	3	0.076458	9	3
PAK1IP1	3	0.083663	10	3

Table S13: ATO secondary screen candidates ranked by FDR.
Candidate selection based on individual sgRNA sequences showing differential abundance in ATO relative to vehicle control with FDR < 0.001.

Gene	sgRNA sequence	logCPM	logFC	FDR
KEAP1	AGTACGACTGCGAACAGCGA	9.9284922	2.85282	8.61E-118
KEAP1	TCGTAGCCCCCATGAAGCAC	10.087803	2.803095	4.13E-117
KEAP1	ACAACCCCATGACCAATCAG	11.20124	2.771692	1.01E-114
KEAP1	AGCGTGCCCCGTAACCGCAT	10.289278	2.674672	1.18E-113
KEAP1	AAGTGCCACTCATCCCGCTC	10.953638	2.439859	2.30E-106
KEAP1	CAGCACCGTTCATGACGTGG	11.038812	2.99108	2.99E-104
KEAP1	CATTCGGGTCACCTCGCTCC	10.401677	2.681653	5.75E-102
EEFSEC	AGGTGCACCGATTATACCCG	9.7886859	1.752534	4.34E-80
EEFSEC	CGGGTATAATCGGTGCACCT	9.7886859	1.752534	4.34E-80
SEPSECS	ATGACGACGAGCAACCAGTG	9.2306148	1.753533	6.68E-73
TXNDC17	CCTCACCCTGCACGCAGTCG	10.108006	1.341048	2.48E-61
ABCC1	AGTACACGGAAAGCTTGACC	7.8188746	-2.37009	5.95E-59
SEPHS2	CGAAAAGGTAACGCCACTCA	8.9053585	1.48679	7.95E-58
FLCN	TGATGATGCTGTACCAGCGC	8.8082305	1.53124	1.67E-56
ABCC1	CAAGTTCGTGAATGACACGA	8.2672621	-1.80797	3.03E-54
SECISBP2	GGAGTCCAAGGGATTGTGCG	8.6653688	1.569443	1.33E-53
TXNDC17	CAGAGAACAACATTTCCACC	9.6504118	1.206332	2.17E-53
SEPHS2	CAATGCCAACGATCCACGCT	9.3347974	1.273544	2.28E-53
TXNDC17	GGGCTGAAGCACATTAGTGA	9.311995	1.270435	5.61E-53
ABCC1	AACCTGACAGCATCGAGCGA	8.0371058	-1.8802	1.34E-51
PSTK	ATGCTGGACTCGGAATTGTG	9.1953384	1.349271	1.02E-50
PSTK	ATCATACGCGACAACACCGA	8.6786825	1.469814	2.31E-49
RRAGC	TGACTTACAGGATGACTACA	8.7746493	1.361725	7.37E-49
EEFSEC	GGGGATGCAGACCCAGTCAG	8.7039932	1.415736	1.55E-48
GFI1B	CCTTGTTGCACTTCACACAG	9.3261309	1.31422	3.15E-48
ABCC1	TCTGCTTCGTCAGTGGCATG	7.6653956	-1.93459	4.27E-46
FLCN	CGCCCGTTACCAGGCAAAGG	8.4297478	1.484598	1.90E-45
TXNDC17	CTGAACCAGTCGTACGAGAG	9.2901466	1.250227	4.91E-45
PSTK	CTCTTGAGACCTGTTTACAG	7.8315095	1.747477	4.34E-44
GFI1B	CTACCACCAGCCCCGTGTGC	9.5198517	1.059481	1.34E-43
SEPSECS	CAGCCAGCTTTATAATGTCC	8.9773282	1.14951	4.90E-43
RRAGC	AGTCCCATGAGCAGAATCCT	8.7492916	1.240372	1.07E-42
TXNDC17	TTCAGCCCCTCTCGTACGAC	9.1803149	1.103122	1.40E-42
FLCN	TGGACCAAGGTATCCTCGGT	9.247931	1.115471	1.48E-42
TXNDC17	AGCTGAACCAGTCGTACGAG	9.0084002	1.151361	5.37E-41
TXNDC17	GTGCCTACACTACTTAAGTA	8.6745311	1.250068	4.91E-40
RRAGC	CATTCCACAACTGCCGACCT	9.0438417	1.066597	6.79E-40
FLCN	TCACGCCATTCCTACACCAG	10.151996	0.920112	6.87E-40
SEPHS2	ATTTGGTTGGCATACTACAG	8.7131547	1.191108	1.36E-39
PSTK	TTGTCTGAGATACAATCCTT	8.9900152	1.379049	3.79E-39
SLC30A1	CCAGCACGTCCGACAGCATG	9.3005593	1.010869	1.03E-38
SEPSECS	GCTGAGTGACAGCTTTGTCA	8.7798524	1.137831	1.36E-38
RRAGC	GATATACGTCATTGACGCAC	8.9550757	1.140151	1.74E-38
SEPHS2	ACAGCAACCTGGGTTCCTAA	8.4452176	1.706766	1.84E-37
FLCN	CGACGCTGGCCCCCTCTGCG	8.7311113	1.139491	3.49E-37
GFI1B	CATTGTGCTGTCCCGACCCC	9.8922561	0.923207	7.94E-37
GFI1B	GGGGTCGGGACAGCACAATG	9.8922561	0.923207	7.94E-37
EEFSEC	GCATGCTCACCCGCTCCACA	8.7422926	1.10144	4.71E-36

ABCC1	AAAATGTGATTGGCCCCAAG	7.5872333	-1.63337	6.48E-36
SECISBP2	CCCCCCAAGTCCAACTGCAC	8.8401262	1.046284	6.64E-36
ABCC1	CCAGCCGAAAGAGAGTTCCA	7.5047865	-1.70239	1.16E-34
SLC30A1	CTGGATCCGAGCCGAGGTAA	9.5023139	0.935991	3.95E-34
NAA38	GCGCGCCCGACAGCAGCTAG	8.7019196	1.04421	6.69E-33
SECISBP2	TCAGACTTGTCCTCAACCGA	8.7639701	1.002663	3.56E-32
GFI1B	GAAGCTAGGCTTGTAGAATG	8.0710087	1.219278	2.13E-31
PSTK	AAACTGATCAGACACTCCGA	7.8408698	1.330606	6.86E-31
NAA38	ACACCACATCGTTTCCATTG	9.3650955	0.827749	8.71E-31
PIH1D1	GTTGACAGCTACGTCGTAGG	8.3069283	1.13755	9.46E-31
MRGBP	CAGCGCCTGCATGTCGTACA	7.9890807	1.227932	1.40E-29
MTF1	CCTTCGTGTGCACTCGCACG	8.7653325	0.937646	1.85E-29
FLCN	TGTACTCTCTGGCAACACAG	7.8900014	1.353029	2.37E-29
VPS39	TCTGATTGACTACCTGACAC	8.2145449	1.121412	4.48E-29
PSTK	TCCGAGTCCAGCATGTGCTT	8.9640339	0.853391	8.45E-29
PSTK	TGTTAAGAGGTAGCAAGACT	8.7277042	0.932584	1.22E-27
GFI1B	GCGGTTCATCTTCCTGCACA	9.6113869	0.771494	2.45E-27
TXNDC17	CGGGTTCTAAGGACGCCGGG	7.9987288	1.134209	6.51E-27
SLC30A1	GATCCGAGCCGAGGTAATGG	8.5017274	0.993341	7.60E-27
SEPSECS	GCTCGCATGAGCACCTCATA	9.0394854	0.773447	2.96E-25
MTF1	CCACGTGCGAGTGCACACGA	8.4340248	0.952746	1.33E-24
SLC30A1	GCTGGACAACTTAACATGCG	8.7146165	0.834107	5.30E-24
DYNC1LI1	GTAGCACGTTCTTCCCCGCA	8.6038662	-0.84898	1.04E-22
EEFSEC	AAAAGATACCGTATTTCCGG	7.1241036	1.357499	2.25E-22
SET	GCTCAACTCCAACCACGACG	8.0116238	1.003742	2.29E-22
SET	ACATTTGTCAACCATCCACA	9.0598201	0.724939	1.03E-21
SET	GCTCTCACCTGAGGTCTCGT	8.7550012	0.760855	2.32E-21
SECISBP2	CAGTTCAGGAAAGTCCAGTG	9.8458991	0.613733	2.51E-21
TECR	CAGGTCCCGGAAGTACAGTG	8.5439149	0.829251	2.88E-21
EED	TTGTGAATGACATTCATACA	8.7930802	0.727963	2.13E-20
DYNC1LI1	AGTACTAGTAGTTTGCACAA	7.5693671	-1.07445	1.30E-19
SET	GAGCAGCACCATGTCGGCGC	8.381443	0.828753	1.47E-19
RREB1	ATCTACCTGGCGAATGTGCA	8.3903723	0.832302	1.87E-19
SET	AGATTTCCATTTGATTTCGG	8.5967659	0.761356	3.64E-19
NAA38	GCATTCGCATGACAGATGGA	6.9644824	1.327452	6.48E-19
ABCC1	TAGAAGTAGCCCTGCCAGTC	8.6352759	-0.72976	2.58E-18
SEPHS2	AGCTATGCGCCCCATCATGT	8.6644011	0.721771	2.93E-18
RREB1	CAGCACAACACAGACACTGG	7.7480434	0.962886	5.25E-18
DPH6	TAGCTACATGTATCAGACAG	7.9799336	0.850035	4.99E-17
UBE2H	GCTTGTATTCTTCTGGTCGG	8.3406519	-0.76472	2.44E-16
UBE2H	GCTGCAGCGTCACCATTGAG	8.8418251	-0.62011	3.51E-16
RRAGC	TCAAAAAAGTCCATTTGCCC	9.5724259	0.532051	3.62E-16
NCAPD3	CAGTGCAGACTTCCTAACGT	7.5676416	-0.9582	4.28E-16
NUP133	CTCACAGAGCAACAGTCGAG	7.7597347	0.892124	7.78E-16
PPP4R2	TCACATTGTTTCTCCAGTCT	9.1024617	0.571548	1.33E-15
SLC30A1	TGTCAGAGAACCTGACCATA	8.9784641	0.565781	5.15E-15
RREB1	AGTCCGTGAATCCTAAGTCG	8.0799738	0.7616	5.66E-15
VPS39	GTTCGGATCTCAACATATCG	8.7966805	0.611273	5.66E-15
RRAGC	TGCTTATGACAAAGAATCTA	8.3707446	0.701354	1.67E-14
GFI1B	GGTTTGGGAATAGAGTGCTA	9.0955963	0.55207	1.93E-14
EED	AAGAGAATGATCCATACCAC	6.3809094	1.322177	2.09E-14
EED	ACAAACACGCCAAATGCACC	8.5871897	0.637551	3.31E-14
TRIP13	CGAGTCGCCAACGGTCCACG	9.0055082	0.544231	3.97E-14

PINX1	CTTGGAATCGTCATTACTCC	8.8729019	0.558916	4.91E-14
PIH1D1	TAAAGACCAACTCCTCGGAA	7.148122	0.997612	5.55E-14
C15orf41	GAGGTTTCTACAGGAACACG	8.3267177	0.711341	7.55E-14
MTF1	GCACATTCGAACTCATACAG	6.8231136	1.146872	9.78E-14
PINX1	ATGGGCAGGAAACCACAGGT	7.5908229	0.869709	1.08E-13
TRIP13	TACCGTGGAGCAGCACCACC	9.3306708	0.500858	1.30E-13
NUP133	GGGACAGACTCAGCCCACCG	7.934797	0.768465	1.88E-13
KEAP1	CCTGGAGGATCATACCAAGC	9.1512068	1.632728	3.95E-13
VPS39	GGAGGTGTTACGGATGTGTG	7.5861804	0.829772	5.54E-13
NDE1	TCCACGCCCATTGCTCACCG	7.67052	-0.8056	5.70E-13
EED	CCACGTAAGATTGCAACAAC	8.784348	0.55508	6.25E-13
DR1	AGAGACTCTTCCTAATGTCC	6.9475631	1.041509	9.45E-13
SLC30A1	CGCGTTAAGAGCACCCGCCC	9.1931702	0.489701	1.16E-12
SLC30A1	GGGCGGGTGCTCTTAACGCG	9.1931702	0.489701	1.16E-12
UBE2H	CAAATTCATTAAGTCCTCCC	8.9325652	-0.51118	1.73E-12
MRGBP	AGAGGCACACCTCCACCTCG	7.5219577	0.829381	1.80E-12
RCE1	CCCATGCACGGTGCTAACAT	9.3090744	0.471514	2.61E-12
C16orf59	TTGCAATTGTCGCTGTGCGC	8.1460557	0.649	6.46E-12
DPH6	CTCTATCGCCGAACCATAAG	7.2335091	0.88445	7.38E-12
DPH6	CTTATGGTTCGGCGATAGAG	7.2335091	0.88445	7.38E-12
MRGBP	AGTTCAGCCAGAACATCGGG	6.132796	1.247659	8.35E-12
KDSR	GCATCCAAGTTTGCCATAAG	7.296564	0.861032	1.14E-11
DR1	CAACGATGCTCGAGAGCTGG	7.4833885	0.80046	1.30E-11
DEPDC5	GCTACATCAGTGAAGATACC	8.8037682	-0.52158	1.63E-11
DR1	CACCAGCTCTCGAGCATCGT	8.0960622	0.646807	1.78E-11
AQP3	TACAACAACCCCGTCCCCCG	9.0432916	0.470785	3.33E-11
TECR	ACTCACGGGTCTTAGTGAAG	8.8709832	0.488964	3.53E-11
DPH6	TGACTATCAGCGTATTCGAG	8.0602191	0.647616	3.72E-11
RREB1	CCTGGTGCGAAACAAACCTG	7.0751757	0.90311	3.87E-11
EED	ATGGCTCGTATTGCTATCAT	8.938448	0.475295	3.92E-11
TECR	TCTTCCTAACAGAGTACGCG	7.4871473	0.769366	8.63E-11
NDE1	TAGGGCAGAAAATACGCAAG	6.9787754	-0.9048	1.12E-10
TECR	TGTACTCACTGGGGTCCAGG	7.7838362	0.680612	2.13E-10
C15orf41	TCTGCTGAGCATCTTCTCCC	8.7424638	0.503109	2.28E-10
RBM5	AGCAAGACCATCATGCTGCG	8.5223694	0.554269	3.54E-10
NCAPD3	CTTATCTCCTTCTCCATGAA	8.5559925	-0.52966	4.43E-10
NAA38	CGGAGAGCGCGAGGACTCGG	8.1148384	0.593918	5.48E-10
PAPOLA	TGAGACATTGAAACCCTTTG	8.9965212	0.44279	5.48E-10
EEFSEC	CAGCAGCTTAGGGTCAAACT	8.8256278	0.466848	5.67E-10
NDE1	TGCAGCACTCTGAAGGCTAC	8.6896933	-0.50285	6.48E-10
PINX1	CCATCTTCTCTAGCATCCGC	7.8276376	0.672486	6.69E-10
DPH6	TCAGGAAAAAGAAGAAGTAG	7.3669795	0.768775	9.60E-10
VPS35	ATTCCAAATGAAGAGATGCC	7.4867151	0.722932	1.10E-09
NPRL2	GGCATACTTGTTGTGTAGTG	9.0954434	-0.4261	1.10E-09
TECR	TACGTCCAGTCGGCATACAG	8.2190284	0.563799	1.35E-09
NCAPD3	CAATGTTTCCCATAGTACAC	7.4150192	-0.73382	1.41E-09
UBE2H	TAGCATCGAGAGTAAACATG	8.2223944	-0.56119	1.66E-09
ZNF236	TTCGCTCGTGATCCCTCATG	8.8692822	0.441931	3.07E-09
SRM	ACTCGTCTCTCTCCGTGCAC	8.3409136	0.532291	3.63E-09
DPH5	GCCTCCATAATGAATGCTGT	6.6887337	0.908276	3.72E-09
PPP4R2	GTGAATTGTTAACAGATCCA	7.1634463	0.767992	4.60E-09
PINX1	TAGGGAAGGATCTGTCATCT	10.104703	0.369038	5.01E-09
MTF1	AATGCACTTCCACAACACAA	7.3460987	0.710034	1.03E-08

PAPOLA	GTAATGCCATAGTGCTTCTG	8.8109392	0.429496	1.50E-08
NPRL2	TCCAGCAAGATGGTCATGAT	9.0537667	-0.4009	1.64E-08
SECISBP2	CTCTTACCTGAGACTACCAC	8.1318073	0.536088	1.69E-08
DPH5	GTTCTAAGAGCAACAAAGCT	6.496626	0.928686	1.86E-08
RBM5	TGCTCAGTGGTCATCCACCC	9.169172	0.383667	2.57E-08
MTPN	GGAGAAGAGCTTTGATTGCC	7.3918528	-0.67845	2.57E-08
NDE1	TGGCGATGACCTACAAACAG	7.396494	-0.67559	2.75E-08
MRGBP	TTCCATTCCCGAATCCAGAG	6.9547585	0.77414	4.09E-08
IPO5	ACCATCATTGCTAACATCTG	8.9359233	0.397503	4.09E-08
C15orf41	TGGTAGGCGGCACAGACACT	8.1368933	0.520977	4.52E-08
MTPN	CAATCTGCTGCATAATGAAG	7.2649672	-0.70717	4.69E-08
BRD8	ATAAGTACCTATATCTCTCC	7.1567786	0.717138	4.86E-08
NCAPD3	AACATAGCGTATTCCCCACA	8.4302443	-0.47984	6.23E-08
RRAGC	TTTCTGCACCACCTTACTAA	6.9724708	0.752399	7.01E-08
CNOT2	AGTCTTCTATGTTTCCACAT	8.0320885	-0.53489	8.32E-08
IRF3	GAGGTGACAGCCTTCTACCG	7.8766706	0.556472	9.50E-08
DHPS	GATGCCCATTCTGGACCAGA	7.4956505	0.61776	1.55E-07
NCAPD3	GTTGACTGTTACCAGTGCGT	7.378441	-0.64528	1.82E-07
NAA38	CTCTCCGTCCGAATCCTGCG	8.4135233	0.456843	3.32E-07
CNOT2	ATGATGAACCACTCCCAGGT	7.7268778	-0.55588	4.49E-07
VPS35	CAACATAGTCCACACGATCA	7.7576083	0.550062	4.67E-07
PAPOLA	TAAGATGGTACCTATCACTG	7.9332127	0.511218	7.66E-07
UBE2H	ACGCTTCATCAATGTTGGGA	8.1856259	-0.46413	8.36E-07
EED	GACGCTGTCAGTATAGAAAG	8.5137879	0.428534	9.53E-07
CDKN1A	AGTCGAAGTTCCATCGCTCA	7.7534987	0.53908	9.58E-07
PPP4R2	ACTTCAGCTCCTGAGCCAAG	7.9631112	0.494718	1.10E-06
SEPSECS	CTGGACATTATAAAGCTGGC	7.6778778	0.547263	1.18E-06
PIH1D1	TCCTCCCGCCGACGTGACCG	8.3652038	0.436365	1.63E-06
RBM5	TTGACAGCTAAAGACGCGTA	7.9657051	0.486126	1.89E-06
SLC11A1	TGCACCCACCGTAGTTATCG	9.6247222	0.320757	1.90E-06
SESN3	AGCTGCTAGCACATCGACCT	8.875587	0.357793	1.91E-06
CDK2	AGAAGCATTACCTTGATGAG	7.6336439	0.541528	2.29E-06
CNOT2	GACAGCTGACCTGATGTAGA	7.6128249	-0.54253	2.37E-06
C16orf59	AAGCAGCTTCCCAGAACTCG	6.8687051	0.690113	2.52E-06
C15orf41	CAATTGAAAGTTATTACCAG	7.9236381	0.480301	3.49E-06
TRIP13	AAGTTGTCATCACATAATCG	8.2973863	0.432514	3.49E-06
MEIS2	TCATCATCGTCACCTGTACC	9.1548949	0.325663	3.57E-06
CTDNEP1	ATCTATTTCAGGTGAGCCAG	9.0973857	0.327984	3.86E-06
MTPN	GGCCCTGAAAAACGGAGACT	7.5135895	-0.5464	4.26E-06
NUP133	TGGTGTGGGAACATTCGTGT	7.9403501	0.478335	4.27E-06
IRF3	AGAAGGGTTGCGTTTAGCAG	7.2956683	0.586306	4.31E-06
SPATA31A5	ATGCTTGGCTTGAGGATCCG	8.4189318	0.410245	4.90E-06
MTPN	GTCAACCGGACACTAGAAGG	7.3145983	-0.57513	5.83E-06
POLR2J	TCCTAGGATCACCATTAACA	5.9852313	0.867153	6.44E-06
SEPSECS	CCTGGCTCCAGAAACCTGTC	8.1924563	0.429247	6.67E-06
ANKDD1A	TGTGCTGGCGTTCATAATGG	8.5031243	0.394603	8.26E-06
KATNB1	CTGAAGGATGGCCACCCCTG	8.4051342	0.400676	9.44E-06
RBM5	TACCCCTAGCGATCATTCTT	8.3998222	0.408989	1.08E-05
RBBP6	TGAGACACAACAATTCATCT	8.3133605	0.407069	1.27E-05
CNOT2	GAATGTTAGTAACCCGACCT	7.4803577	-0.53033	1.47E-05
NPRL2	CTCATTGCAGACCTCTAGTG	8.8343578	-0.33116	1.82E-05
PPP4R2	CATGACAAAGAAACTGATCC	8.0864986	0.416655	2.15E-05
NUP133	TGCTCTGTCACCTATTACTA	7.6616112	0.4883	2.23E-05

EED	TTTAACTGGCACAGTAAAGA	6.7573365	0.65591	2.27E-05
AQP3	CAAGCTGCCCATCTACACCC	8.5356005	0.368842	2.69E-05
AQP3	TGCCCCGGCTGAGCACAACC	8.0375096	0.419678	3.10E-05
PAPOLA	ACGAGAAATCAGTGAAAGCA	8.127386	0.406798	4.07E-05
DEPDC5	GGATCAGTATATTGGCCGTG	8.979135	-0.30228	4.58E-05
C15orf41	TGCTTGATGCAGTCCACTAG	7.5983154	0.481821	4.63E-05
BPGM	GGATCGCCTCTTGGTCACCC	8.8953114	-0.30647	4.73E-05
CNOT2	TATGCTTGGCGAGCTTCTGT	7.1402697	-0.55079	5.49E-05
CCNL2	TCCGCTTCAGCATGTGTCAA	8.9143149	0.296296	8.54E-05
TNFSF13	GCTGCAGAGCCTCAGGAGAG	7.8963202	0.418365	9.36E-05
DHPS	CTTCTGGGCCCAGTAATACA	8.1049874	0.388676	0.000102
ERCC5	AAATAATGTTCTTCTGCGCT	9.173373	0.287071	0.000103
PPP4R2	CAGGCCACTTAATCGACCAA	8.9801544	0.287029	0.000108
DPH5	TACAGTTAGGACTGAGGTGT	5.3177551	0.92567	0.000109
HOXC9	CCCCGTTACGATGTCGGCGA	8.6169413	0.333761	0.000114
DEPDC5	TGTCCCCATAGATGTCGCAG	9.036679	-0.28458	0.00013
KDSR	GCTATCGAGTGCTATAAACA	6.7465177	0.602309	0.000132
TAS1R2	AGGCCTGCCCGAATAACGAG	8.2620452	-0.36589	0.000143
C15orf41	TGCTTCCGAAGTATGATGTT	8.8035425	0.301769	0.000144
MRGBP	GTGCCTCTTCCACGCCATGC	8.222281	0.368726	0.000166
ELF4	ACATGAACTATGAGACAATG	8.1753668	0.367144	0.000176
DHPS	TAAGTGCGGGACTAAACACA	3.8576526	1.333295	0.000181
CTDNEP1	ATAGCAGAAGCATTCTTAAG	8.5832666	0.333719	0.000189
IPO5	GCTCTAGATCGAATGGCTTG	8.5701863	0.339636	0.000196
MTPN	GTGCGACAAGGAGTTCATGT	7.6796435	-0.43162	0.000197
PRRT2	ACCGCCTAAGTTGATGACGC	8.8927223	-0.28481	0.000208
RBM5	ATTACTATGATCCGACAACA	7.8285737	0.407765	0.00022
CTDNEP1	CCCCACCACCACACACCTTG	8.7638194	0.300376	0.00022
IPO5	GCAAGATCAGGAACACCCGT	8.1614958	0.361865	0.00025
RREB1	CTTCTCGATATCCTTGCGGG	7.7817047	0.412332	0.00025
PPRC1	CCCAAGCTTACCCTTGGACA	6.3020525	0.655597	0.000291
CCNL2	CTGGAACCGCACGAAGACGT	7.873057	0.403328	0.000291
APOOL	TTGTAGACATGGCGGCCATC	8.9740725	-0.2756	0.000299
CHMP4B	GAATGCCAACACCAACACCG	7.0080155	0.515678	0.000364
PKMYT1	CCCCGGAATGACACCCGCCG	7.220759	-0.4862	0.000372
MTPN	GGATGAGGTGAAAGACTATG	7.3515821	-0.46583	0.000381
ERICH6B	TACTGAAGTAGAACTAGATG	8.9802607	0.266071	0.000408
UVSSA	GGATCAGAAACTTTCGAAGT	8.1477635	-0.34579	0.000504
MEIS2	CTTACCAAATTGTCCAGCTC	7.5506164	0.42651	0.000524
RCE1	CTCCAGAGCAGCACGCACAG	9.197853	0.251215	0.000526
HELQ	GTTGACAGCAAAGCTGAGAA	8.0234604	-0.35963	0.000532
RETSAT	AAAAGCCTCCTTTCATGTAG	8.390651	-0.32637	0.000609
NPRL2	AGAGAGCAGCTTCGTGTCCA	9.1223872	-0.25298	0.000635
ATF1	AAGTATCTGCTGTCCATCAG	9.0144103	0.259688	0.000638
TNFSF13	CTGCCCCCCAACTCAACCAG	9.1850038	0.251167	0.000656
VPS39	CCAGCACACAGAGACCGGTG	8.2866373	0.328139	0.000661
CCNL2	CTTCGTGAAGCACTCCATGG	7.1814912	0.474003	0.000699
SOWAHA	CCACTTCGACGGTCTTGCCG	8.12279	-0.3403	0.000699
MTF1	ACGTTTAACTGTGAATCTGA	7.809736	0.383439	0.000722
SRGAP2B	TCGATTTGTACAAGTCAGCG	8.5403108	-0.30965	0.00076
DEPDC5	TCTTAAGGAAGATTTACAGA	7.3765565	-0.43515	0.000859
CDK9	GCTGACTGATGAGGGCGAGT	6.6589189	0.552872	0.000952

Table S14: Top enrichment hits (FDR <0.05) in the ATO secondary screen based on MAGeCK analysis.

Gene	Used sgRNAs	FDR	Rank	Good sgRNAs	Log FC
KEAP1	8	0.000354	1	8	3.6209
TXNDC17	8	0.000354	2	8	1.3756
PSTK	7	0.000354	3	7	1.5591
GFI1B	7	0.000354	4	7	1.0784
SLC30A1	7	0.000354	5	7	0.96709
FLCN	7	0.000354	6	7	1.2972
EED	7	0.000354	7	7	0.67214
RRAGC	8	0.000354	8	8	1.0128
EEFSEC	7	0.000354	9	6	1.6044
C15orf41	7	0.000354	10	7	0.55323
SET	8	0.000354	11	7	0.83751
SEPHS2	7	0.000354	12	6	1.404
SEPSECS	8	0.000354	13	7	0.74617
DPH6	7	0.000354	14	6	0.84199
NAA38	8	0.000928	15	8	0.70017
MTF1	8	0.000928	16	7	0.59462
RREB1	7	0.001375	17	6	0.92188
AQP3	6	0.001375	18	6	0.40152
TECR	7	0.001824	19	5	0.59738
SECISBP2	8	0.003713	20	6	0.67625
PINX1	7	0.004479	21	6	0.45232
CCNL2	7	0.006976	22	7	0.37641
PPRC1	8	0.008394	23	7	0.373
PPP4R2	8	0.012079	24	6	0.42572
RBM5	7	0.012079	25	6	0.41979
PAPOLA	7	0.018469	26	6	0.42426
MRGBP	8	0.019252	27	6	0.58367
NUP133	8	0.020686	28	6	0.46963
SRM	7	0.025435	29	7	0.18889
IPO5	7	0.038167	30	6	0.27192
CTDNEP1	8	0.038167	31	6	0.2788
CABIN1	7	0.049041	32	6	0.33024

Table S15: Top depletion hits (FDR <0.05) in the ATO secondary screen based on MAGeCK analysis.

Gene	Used sgRNAs	FDR	Rank	Good sgRNAs	Log FC
ABCC1	8	0.000619	1	8	-2.1472
MTPN	7	0.000619	2	7	-0.69635
NCAPD3	7	0.000619	3	6	-0.65907
DEPDC5	7	0.000619	4	7	-0.44416
UBE2H	8	0.000619	5	7	-0.58585
NPRL2	6	0.000619	6	6	-0.34835
CNOT2	7	0.000619	7	7	-0.60749
NDE1	8	0.000619	8	7	-0.65595
BPGM	7	0.00275	9	7	-0.23163
CD84	8	0.012376	10	8	-0.12878
KIF20B	7	0.020252	11	7	-0.23933
QPRT	8	0.026815	12	8	-0.15952
FOXM1	6	0.027799	13	5	-0.20636
SH3TC1	8	0.029349	14	8	-0.20469
UGT2A3	8	0.034653	15	7	-0.28277
DYNC1LI1	7	0.035582	16	6	-0.36164

Table S16: Functional classification of validated genes involved in ATO toxicity.

Pathway	Gene	Validating sgRNAs	Knockout Phenotype
Selenocysteine Metabolism	PSTK	7	Resistance
	EEFSEC	6	Resistance
	SEPSECS	6	Resistance
	SEPHS2	5	Resistance
	SECISBP2	5	Resistance
Oxidative Stress	KEAP1	8	Resistance
	TXNDC17	8	Resistance
Diphthalmide Biosynthesis	DPH6	5	Resistance
	DPH5	3	Resistance
Histone Modification	GFI1B	7	Resistance
	FLCN	6	Resistance
	SET	5	Resistance
	UBE2H	5	Sensitivity
	MRGBP	5	Resistance
	PIH1D1	3	Resistance
	DR1	3	Resistance
Transport	SLC30A1	7	Resistance
	ABCC1	7	Sensitivity
	MTF1	5	Resistance
	AQP3	3	Resistance
RNA Processing	RRAGC	7	Resistance
	PPP4R2	5	Resistance
	RBM5	5	Resistance
	PAPOLA	4	Resistance
	CCNL2	3	Resistance
Miscellaneous	C15orf41	6	Resistance
	MTPN	6	Sensitivity
	EED	6	Resistance
	CNOT2	5	Sensitivity
	NCAPD3	5	Sensitivity
	RREB1	5	Resistance
	TECR	5	Resistance
	VPS39	4	Resistance
	PINX1	4	Resistance
	IPO5	3	Resistance
	CTDNEP1	3	Resistance
	DHPS	3	Resistance
	KDSR	2	Resistance
	MEIS2	2	Resistance
	RCE1	2	Resistance
	C16orf59	2	Resistance
	DYNC1LI1	2	Sensitivity
	TRIP13	2	Resistance

Table S17: Acetaldehyde primary screen candidate genes ranked by FDR.
Candidate selection based on individual sgRNA sequences with log2 count per million (CPM) > 2.5 showing differential abundance in acetaldehyde relative to untreated control with FDR < 0.1.

Gene	sgRNA sequence	log2CPM	log2FC	FDR
C10orf91	TTCCCCGCTTCAGGCTTCGA	4.715521822	-2.651912829	1.82E-23
SPATA31A5	CTCCTTGCTGCCTTCACCTG	7.094653559	-1.112377751	9.32E-07
EML4	GTCATCATAGTTGTCAACAT	4.289403276	-1.550650502	4.56E-06
OVCA2	TGTCTCACCGAAGTCTGATC	3.913842481	-1.524465716	5.19E-05
ERCC8	CAGTGGTATCCTCATGACAC	4.751137662	-1.39154304	5.19E-05
SYNE1	GATATCCCGAGGACACCGGG	3.725530733	-1.368817324	0.005547657
APOOL	TTGTAGACATGGCGGCCATC	4.694454477	1.009099034	0.00976383
HPSE	CGTCACCATTGACGCCAACC	6.273532571	1.061416113	0.011651179
SNAPC5	CTTACATCATGTGACTGTTC	3.25452683	-1.440586607	0.015855838
UVSSA	AAAGCGTGCCATTGACCTGA	4.81653696	-0.93800656	0.015855838
hsa-mir-4502	TGTAATCTTTTTGCTGATGG	5.303877674	0.941117777	0.015855838
ZNF181	GCCATATGTGATCACGTTAT	4.306319351	1.069683081	0.015855838
NIPBL	TACGCCCCACAAAGCCCTGC	2.840966583	1.660083516	0.015855838
hsa-mir-548av	AACTGCAGTTTTGCCATTAA	6.459566527	-0.783523524	0.018104775
CDK2	TGTTCGTACTTACACCCATG	4.024402051	1.104451881	0.020858877
RHOBTB1	ACAAGCGTGTTTGACCAGTT	3.489698074	-1.374699179	0.023111681
SH3TC1	ACATCTTCTTCGACGGGGCC	5.171907015	-0.906561613	0.02630088
XIAP	CCATCACTAACTAGAAGAAT	4.47304379	0.975464707	0.027258657
SESTD1	AACAGTAGTCGTAATGCTAC	4.302632674	1.053343946	0.02887497
DCUN1D1	CAGCTGTACAATAGATACAA	2.716635983	-1.588662558	0.031943527
ARHGEF6	TACGTCACACGAGTTGAAGA	2.876769838	-1.519581047	0.03237278
CDKN1A	CCATTAGCGCATCACAGTCG	4.13208618	1.047875999	0.03237278
SSNA1	CCTCTCGCAACGAGTTCGAC	5.566603618	0.770638595	0.039967681
RHOV	GCGCGCGGGGTACCCATTGC	3.037432117	-1.52414793	0.052260391
ORAI1	CGGCGAAGACGATAAAGATC	3.180629044	-1.412568648	0.052773132
HELQ	ACCAATGCTACCAAGTGATC	4.596867414	-0.923077191	0.052773132
TRAM1L1	CGAGCCTTCCATATAAGAGT	2.569945869	1.649792798	0.052773132
KATNB1	ATGACTGCCGCGTCAACCTG	4.816170937	0.84204845	0.055075837
ANXA8L2	ACTTACCCAGGCTTTCCACC	3.158634085	-1.290210539	0.062239097
BPGM	GGATCGCCTCTTGGTCACCC	3.589793717	-1.22588119	0.065693536
FBXL20	TTACCGTAACAGGAGTTCTT	5.565260784	0.829030883	0.065693536
hsa-mir-4440	CACTGCACTTGCTTGGTGAG	3.885917967	1.05078871	0.075065485
NTF3	TACCGTGGCAAAAGTAACCA	5.065851306	-0.775249459	0.077397912
PALB2	TGCCCGATGGACGAGCCTCC	3.936908314	-1.029820712	0.084180762
RSRC2	CAAGATCTAGGTCTCGTGAA	3.194756894	1.224704644	0.084180762
MIS12	CGCCACAAACGTGCATGCTT	3.401345968	1.21615504	0.08527368
CD84	GAGTCAGAATTGTTGCTGAC	4.862781729	-0.905644253	0.088670728
HIRA	TACCTAAGTGATTGTCCATC	4.132822759	0.968753125	0.094050209
ZNF236	TTCGCTCGTGATCCCTCATG	4.507375738	0.838411896	0.094548004

Table S18: Acetaldehyde secondary screen candidates ranked by FDR.
Candidate selection based on individual sgRNA sequences showing differential abundance in acetaldehyde relative to untreated control with FDR < 0.05. Additional candidates with FDRs between 0.05 and 0.2 are highlighted.

Gene	sgRNA sequence	log2CPM	log2FC	FDR
OVCA2	GCCGAGCTCGTGTGCCTCAG	8.046624537	-1.157108238	1.07E-10
OVCA2	GACACCAAGAGGATAAACCG	7.158844413	-1.26928468	5.21E-09
OVCA2	TGTCTCACCGAAGTCTGATC	8.145570747	-0.892537728	7.54E-09
HELQ	TGCTGGAATAGATACTATTG	7.622997719	-0.881622174	3.73E-08
HELQ	GGAGTTGCCTATCACCACAG	7.290085494	-1.002562453	6.99E-07
HELQ	TCCTGATCACTTGGTAGCAT	8.035895961	-0.695629563	6.99E-07
OVCA2	TTCCAATGCGGAGAAAACGT	6.79425597	-0.953213802	6.99E-07
OVCA2	CGGGGCTTCCGTGAGAAGAC	7.082896022	-0.870287941	1.49E-06
HELQ	GTTGACAGCAAAGCTGAGAA	7.806332951	-0.655537411	1.72E-06
OVCA2	GGGCTTCCGTGAGAAGACCG	6.265461438	-1.065906067	1.98E-06
HELQ	TGAAGTATATCATCCAATCA	8.357406538	-0.514507	1.05E-05
OVCA2	GAGGGCGCCAGATCAGACTT	7.09078245	-0.840154002	5.53E-05
PPP4R2	CATGACAAAGAAACTGATCC	7.500701398	-0.547150433	0.000259819
ERCC8	GCCAAGATATAGTCATAACG	8.218613933	-0.410175088	0.00035922
ERCC5	TTAATGGCTGAAAGAGTCCG	8.828892732	-0.334715995	0.000556072
PPP4R2	TCACATTGTTTCTCCAGTCT	8.528171372	-0.350102576	0.000569644
HELQ	ACCAATGCTACCAAGTGATC	8.343288647	-0.419828906	0.000699568
ERCC8	CAGTGGTATCCTCATGACAC	7.945572797	-0.377891985	0.000797574
RREB1	AGTCCGTGAATCCTAAGTCG	8.151079573	0.425329918	0.000877949
OVCA2	CAACTGGCCAGCCAATTTCC	8.321641702	-0.340876337	0.001456098
ERCC8	TGTAAAGCAGTGTGTTCCAT	8.775961331	-0.311834132	0.004158919
KEAP1	AGTACGACTGCGAACAGCGA	8.005919258	-0.521423355	0.004395591
NANS	GAGATCGGCCAGAACCACCA	6.504548953	-0.520151258	0.021651217
RAB40AL	CTGGTGGGCGACAGGGACGT	8.923830742	0.32464674	0.027747298
NANS	TATGTGACGTTCCAACACCT	8.007839997	-0.267683328	0.036401637
HELQ	CTTATCTCTTACCTTCGAGC	6.257301768	-0.480664874	0.045995708
PPP4R2	CAGGCCACTTAATCGACCAA	8.676211407	-0.225368825	0.069313514
ZNF671	GAACTCGCCTGTGCTCAATC	8.819471945	0.206796893	0.070997476
PINX1	TAGGGAAGGATCTGTCATCT	10.10064454	0.28070039	0.070997476
KEAP1	CATTCGGGTCACCTCGCTCC	7.832030335	0.261322345	0.070997476
FBXO40	AACCTCCGGCTTAATGGCAA	9.28521364	-0.255373421	0.080809167
PALB2	GCCCCTCAGCTGTGAGGAGA	6.2107472	-0.431805615	0.114117565
RHOBTB1	GAGCCAGGCGCCCGTTAGCA	7.418245906	0.308143187	0.115843487
SRM	CCCCGCCGAAAGTCTCTTCA	8.494907737	0.212946544	0.115843487
UVSSA	AATTGAATCCTGCTTGACGG	8.325020516	-0.211390942	0.115852232
PPP4R2	ACTTCAGCTCCTGAGCCAAG	7.392559312	-0.309546104	0.138614989
PALB2	TGCCCGATGGACGAGCCTCC	6.567316477	0.599901779	0.145886779
NCAPD3	CAATGTTTCCCATAGTACAC	7.671611239	-0.272893383	0.145886779
KIF20B	AAATAGCTTCAATACTAGAG	7.579400152	0.257828429	0.167401155
FOXM1	AACTCATCTTTCGAAGCCAC	8.883536926	-0.164355538	0.167401155
TUBA1C	TTGCCTGTGATGAGTTGCTC	8.094514146	-0.22005734	0.176860732
ANKRD32	GCTTATCAGTTCTAACAAGG	6.07831791	-0.437849189	0.176860732
ACSL3	GTGGTGAAGAGTAACCAATG	7.444505141	-0.311991066	0.186105986
UVSSA	AAAGCGTGCCATTGACCTGA	8.580059514	-0.193312579	0.186494016
STEAP3	AGCCACAAAGATGACCTCCG	8.782615393	-0.174944064	0.188189293
ERCC8	ACCTGCTACCAAACAGTGCT	8.095079135	-0.238213369	0.194703495