Sifting through the regulatory weeds

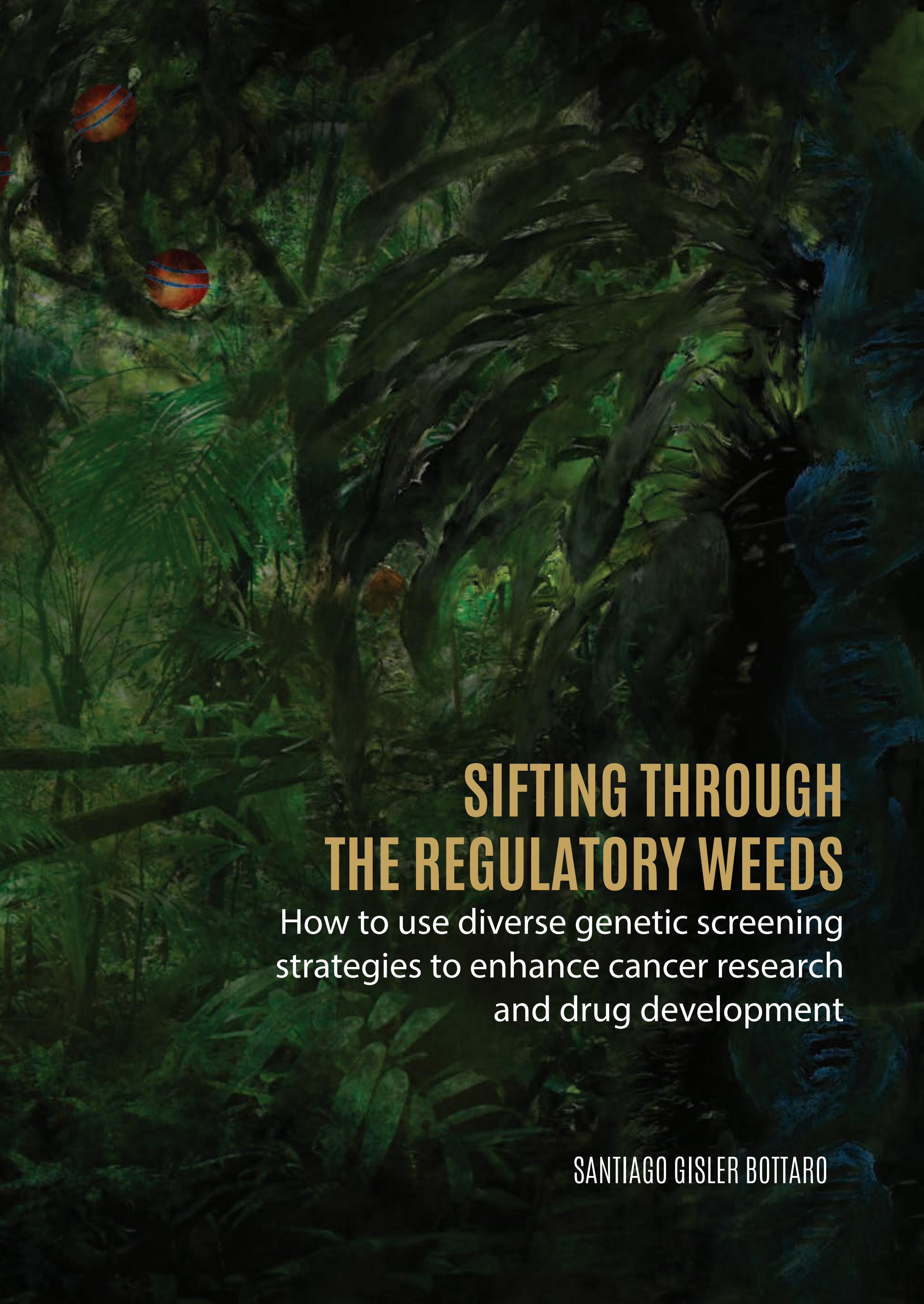
How to use diverse genetic screening strategies to enhance cancer research and drug development

Gisler, S.

Publication date
2022

Document Version
Final published version

Citation for published version (APA):
Gisler, S. (2022). *Sifting through the regulatory weeds: How to use diverse genetic screening strategies to enhance cancer research and drug development*. [Thesis, fully internal, Universiteit van Amsterdam].



SIFTING THROUGH THE REGULATORY WEEDS

How to use diverse genetic screening strategies to enhance cancer research and drug development

SANTIAGO GISLER BOTTARO

SIFTING THROUGH THE REGULATORY WEEDS

How to use diverse genetic screening strategies to enhance cancer research and drug development

Santiago Gisler Bottaro

Cover design: Anaïs Moussy
Cover, layout, and printing: Off Page, Amsterdam
ISBN: 978-94-93278-17-2

Sifting through the regulatory weeds
How to use diverse genetic screening strategies to enhance cancer research and drug development

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. G.T.M. ten Dam
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 21 september 2022, te 10.00 uur

door Santiago Gisler Bottaro
geboren te Lund

Promotiecommissie

Promotor:	prof. dr. M.M.S. van Lohuizen	AMC-UvA
Copromotor:	prof. dr. A.J.M. Berns	AMC-UvA
Overige leden:	prof. dr. P.J. Verschure	Universiteit van Amsterdam
	dr. P.M. Krawczyk	AMC-UvA
	prof. dr. L. Vermeulen	AMC-UvA
	prof. dr. R. Versteeg	AMC-UvA
	prof. dr. H.P.J. te Riele	Vrije Universiteit Amsterdam
	prof. dr. J.P. Medema	AMC-UvA

Faculteit der Geneeskunde

We shall not cease from exploration and the end of all our exploring will be to arrive
where we started and know the place for the first time.
T.S. Eliot

Man who catch fly with chopstick accomplish anything.
Mr. Nariyoshi (Keisuke) Miyagi

TABLE OF CONTENTS

Chapter 1

GENERAL INTRODUCTION I

INTRODUCTION I: UNDERSTANDING EPIGENETIC MECHANISMS AND INTERACTIONS TO ADVANCE TARGETED CANCER THERAPIES

After discovering the DNA, classical genetics assumed that DNA sequence alone dictated the phenotypes of cells and, consequently, all organisms. However, these assumptions left the wide range of phenotypic diversity within species unexplained. Similarly, the assumptions could not account for functional differences and distinct disease susceptibilities between organisms containing identical sets of genes, for example, monozygotic twins or even the different somatic cells of a single organism. The inability of genetic sequences to explain such variations called for other factors that could. One missing link to our early understanding of genotype–phenotype relationships was the epigenome and heritable epigenetic processes, which alter or regulate gene functions without changing the actual DNA sequence[1–3]. It is now widely accepted that epigenetic events regulate gene expression, cell growth, cell differentiation, and diseases, making the study of such processes essential for understanding developmental biology and cell functionality. The prevalence of epigenetic alterations in diseases makes them promising therapeutic targets, including cancer. In particular, the reversibility of epigenetic alterations, as opposed to irreversible genetic mutations, makes them very attractive drug targets.

Epigenetic events regulate cellular activity and function through processes that do not alter the genetic code. These processes include DNA methylation, histone modifications, chromatin remodeling, and non-coding RNA activity. Often, we use the analogies of "writers", "erasers", and "readers" to classify the actions of epigenetic enzymes[4]. While writers and erasers add or remove chemical components to or from the DNA or histones, readers recognize these changes and act upon them.

The following sections will introduce the reader to the functions of the most common types of epigenetic alterations and present some of the therapeutic advancements that target epigenetic regulators in cancer.

The different types of epigenetic changes

DNA methylation

DNA methylation is an essential process for developing organisms and diseases, such as cancer[5]. These events occur when DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B) add a methyl group to a cytosine residue linked to a guanine nucleoside through a phosphate (forming a CpG dinucleoside). Like a genetic off switch, DNA methylation represses gene transcription when located on gene promoters[6,7].

These heritable DNA methylation tags are relatively stable in somatic cells, showing minor global changes over time and during cell DNA replication or cell arrest[8]. Still, while primarily stable, demethylation does occur, for example, during the reversal (reprogramming) of imprinted silenced genes in embryonic stem cells and germ cells[9,10].

Apart from their stability, DNA methylation levels and occurrence are similar between individuals. However, despite these similarities, DNA-methylation levels vary between tissues and between normal and cancer cells[11–14]. While the genome of cancer cells is generally hypomethylated, tumor suppressor genes are often hypermethylated[15]. Hypermethylation in cancer cells is often observed in genes that inhibit cell proliferation and metastasis, such as genes associated with DNA damage repair or cell cycle regulation. For example, promoters of key tumor suppressor genes, such as CDKN2A, p53, and BRCA1, are highly methylated in cancer cells, compared with normal cells where the same promoters remain unmethylated[16–18]. Importantly, changes in DNA methylation, whether hypermethylation or hypomethylation, are associated with cancer development and progression. In particular, hypermethylation of CpG islands, which are regions with increased levels of CpG sites, dysregulates gene transcription and occurs in several cancer types. Studies have even shown a positive correlation between DNA methylation levels and tumor size, such as in breast cancer[19].

With the apparent involvement of DNA methylation in cancer and its reversible characteristics, it is not surprising that demethylating agents and DNMT inhibitors are attractive candidates for treating cancer. Especially two DNMT inhibitors (azanucleosides) have emerged as cancer treatments: azacytidine and decitabine. Both inhibitors are cytidine analogs with modified pyrimidine rings – replacing their carbon at position 5 for nitrogen – that incorporate into RNA (azacytidine) or DNA (azacytidine and decitabine) of quickly proliferating cells. The newly incorporated cytidine analog covalently binds the enzyme. As a result, DNMT trapping triggers the DNA damage response to degrade the DNMT enzyme and restore the hampered functionality of the DNA[20,21]. These two inhibitors have been approved by the US Food and Drug Administration (FDA) and authorized in the European Union upon recommendation by the European Medicines Agency (EMA) for the treatment of acute myeloid leukemia (AML), myelodysplastic syndrome (MDS), and chronic myelomonocytic leukemia (CMML).

Histone modifications

Instead of altering the DNA directly, epigenetic processes activate or repress genes by chemically modifying histones. In the cell nucleus, the DNA is packed into chromatin, a genetic complex consisting of units of histone octamers (eight proteins: two copies of the respective histone proteins H2A, H2B, H3, and H4) surrounded by the DNA. Specifically, an approximately 147-basepair DNA wound twice around a histone octamer creates the basic structural unit of chromatin, namely the nucleosome[22,23].

The packaging of DNA through the formation of chromatin structures organizes the DNA for proper segregation during cell division. On top of that, the dynamic properties of transferring between open (euchromatic) and closed (heterochromatic) states orchestrate fundamental biological processes that are essential for survival; transcription, replication, as well as DNA repair and genome stability[4,24,25]. These structural changes are primarily dictated by chemical modifications on the histone proteins and DNA, nucleosome remodeling, and spatial organization of chromatin. All these dynamic features of chromatin ultimately alter the accessibility of the DNA for regulatory factors[26,27]. Although complex and not fully understood, the following sections will highlight the most basic concepts of histone regulations that moderate chromatin changes and their implications in cancer.

Histone acetylation and deacetylation: acetyltransferases (HATs) and deacetylases (HDACs)

Histone acetylation and deacetylation enzymes are perhaps the most studied and best-characterized ones among epigenetic writer enzymes. These two groups of enzymes add or remove acetyl groups to or from lysine residues at the N-terminus of histones, as well as several non-histone proteins and RNA[28–32]. While lysine acetylation promotes chromatin decompaction, partly by neutralizing positive charges in the histones that attract them to the negatively charged DNA, deacetylation counteracts this neutralization process. In other words, while histone acetylation generally promotes open chromatin, which is permissive to gene transcription and DNA repair, deacetylation reverts these actions[33–35]. These two processes are orchestrated by histone acetyltransferases (HATs) and histone deacetylases (HDACs), respectively.

Human HATs are divided into two classes (types A and B) that distinguish them by their localization[36]. The type A HATs, such as p300/CBP and Gcn5, exist in the nucleus and contain a bromodomain, an approximately 110 amino-acid long protein domain that recognizes acetylated lysine residues on histones[37–39]. Upon recognition, HAT transfers an acetyl group from an acetyl-CoA co-factor to a conserved lysine residue on the histone, opening the chromatin and, consequently, permitting gene transcription. Type B HATs, on the other hand, appear in the cytoplasm and acetylate newly synthesized histones before the histones assemble into nucleosomes. Since type B HATs only acetylate newly synthesized histones without acetylated lysine residues, they lack a bromodomain.

Type A HATs acetylate lysine residues in the nucleus on histones H3 and H4, resulting in H3K9ac or H3K27ac mainly in promoter regions and enhancers of active genes[40]. Since HATs generally induce open chromatin, which promotes gene transcription, downregulation or upregulation of these enzymes or their functions are linked with cancer development and progression[41,42]. For example, imbalances of HATs can impede DNA damage repair proteins' access to damaged sites or activate genes associated with cell cycle progression and cell proliferation.

On the other side of the spectrum, HDACs usually repress gene transcription by reversing the acetylation by HATs, promoting more compact chromatin. The human genome encodes 18 different HDACs, which target different sites. In fact, HDACs are highly diverse, regulating transcription by deacetylating histones and non-histone proteins[43]. In histones, deacetylating complex formation and recruitment are promoted by adjacent methylated DNA, while histones at unmethylated DNA regions cannot recruit HDACs. Because HDACs generally repress transcription, they are implicated in tumorigenesis.

For example, recent studies have demonstrated elevated levels of different HDACs in several cancer types, including hematopoietic malignancies, radioresistant breast cancer, and the majority of squamous cell non-small cell lung cancer (NSCLC)[44–46]. Moreover, a mutagenesis screen on a panel of 180 primary tumors or cancer cell lines with mutated HDACs or HATs discovered distinct HDAC expression profiles between cancer cells and normal cells, which can be used to distinguish cancer tissues from normal tissues[47].

Recent years have seen the emergence of several HDAC inhibitors for cancer treatment. *In vitro* and *in vivo* studies of several cancer types have shown promising anticancer effects by pan-HDAC inhibitors [48]. While their mechanisms of action vary widely in terms of cancer specificity and dosage, HDAC inhibitors do induce cell cycle arrest in cancer cells through the promotion of *CDKN1A*[49–51] and cell death by regulating pro-apoptotic and anti-apoptotic genes[52–54]. Notably, several HDAC inhibitors have been approved for cancer treatment, either as single agents or in combination with other pharmaceutical drugs or radiotherapy. Particularly, these inhibitors are indicated for specific T-cell lymphomas (vorinostat, romidepsin, belinostat) or multiple myeloma (panobinostat)[55]. Apart from these, several other HDAC inhibitors are being evaluated in clinical trials to treat hematological cancers and other cancer types, including breast cancer, prostate cancer, and non-small cell lung cancer.

Histone methylation (methyltransferases)

Histone methylation occurs on the side chains of selected arginine and lysine residues. Out of these, methylation of lysine residues is the best characterized and will therefore serve as a representative example of histone methylation for our purpose. Histone methylation enzymes can add one, two, or three, methyl groups to lysine residues, resulting in mono-methylated, di-methylated, or tri-methylated products, respectively. Unlike histone acetylation, the addition of methyl groups to lysine residues (and arginine residues) does not change the electrostatic charge of the residues. In other words, adding a methyl group to the respective amino-acid residues does not result in charge-dependent conformation changes of the chromatin, such as euchromatin and heterochromatin. Instead, the specific methylated residue dictates active or repressive properties on the associated gene. For example, while methylation of H3K4, H3K36, and H3K79 often marks active genes, methylation of H3K9, H3K27, and H4K20 is often associated with gene repression[56]. Similarly, the same lysine residue may repress or activate gene expression depending on its methylation state, i.e., mono, di, or trimethylated. An example is the methylation of H3K9, which can activate genes in its monomethylated state (H3K9me1) but repress them in its trimethylated state (H3K9me3).

Histone lysine methyltransferases are highly selective of the lysine substrate. However, it is worth noting that, although being extremely specific, all but one of the histone lysine methyltransferases contain a SET (suppressor of variegation, enhancer of zeste, trithorax) domain – the exception being DOT1L, which methylates H3K79[57,58]. In essence, the SET domain methylates its substrate by transferring a methyl group from co-factor

S-adenosyl-L-methionine (SAM) to the lysine residue. This enzymatic reaction results in a methylated lysine and a byproduct of SAM (S-adenosyl-L-homocysteine; SAH).

The high substrate selectivity of lysine methyltransferases makes them especially vulnerable to aberrant alterations, usually linked to malignancies. For example, overexpression or gain-of-function of EZH2, the main enzyme of polycomb repressive complex 2 (PRC2; also including EED, SUZ12, and RbAp460), is implicated in several cancer types, including in lung[59], breast[60], and prostate cancer[61], as well as hematological malignancies[62,63]. Clinically, EZH2 overexpression or EZH2 point-mutations increasing its catalytic activity (e.g., the substitution of tyrosine 641 [Y641] of the SET domain) are also associated with poor clinical outcomes and overall survival in cancer patients with prostate cancer or B-cell lymphoma[64–66].

Identifying histone lysine demethylases shifted the original idea of a stable lysine methylation state to a more dynamic procedure[67]. These demethylases can be categorized into two classes. The first class uses flavin adenine dinucleotide (FAD) as a co-factor to demethylate monomethylated and dimethylated histone lysine residues. Lysine-specific demethylase 1A (LSD1 or KDM1A) belongs to this class of demethylases. The members of the second class, the Jumonji (JMJ) demethylases, share a JmJC domain and rely on Fe(II) ions to react with α-ketoglutarate and demethylate any of the three methylation states[68]. Ubiquitously transcribed tetratricopeptide repeat, X chromosome (UTX) and Jumanji domain containing 3 (JMJD3), which specifically demethylate H3K27 methyl marks, belong to this class of demethylases[69].

Like histone methyltransferases, histone demethylases have emerged as promising drug-target candidates. For example, LSD1 is overexpressed in many malignancies, including acute myeloid leukemia (AML), bladder and colorectal cancer, prostate cancer, breast cancer[70,71]. Similarly, several members of the JMJ family have been implicated in different cancer types, including hematological malignancies, prostate, colon, and breast cancer[68].

Later years have observed the development of several inhibitors targeting processes associated with histone methylation (including arginine methylation). For example, a handful of inhibitors of wild-type and mutant forms of EZH2 (e.g., Y641) have been approved in the US or entered clinical trials (CPI-0209, CPI-1205, tazemetostat, GSK2816126 (GSK-126), PF-06821497, and DS-3201). For example, the FDA approved SAM-competitive small molecule tazemetostat to treat metastatic or locally advanced epithelioid sarcoma[72]. H3K79 methyltransferase DOT1L has also emerged as a promising drug target for treating mixed-lineage leukemia (MLL)-rearranged leukemia, where DOT1L interacts with fusion partners. Ongoing phase 1/2 clinical trials evaluate the effect and safety of DOT1L inhibitor pinometostat to treat MLL-rearranged leukemia in combination with chemotherapy (NCT03724084) or DNMT azacytidine (NCT03701295).

Where it all comes together

This chapter has provided the reader with an overview of the different epigenetic changes involved in cancer when misregulated. Several excellent recent reviews have described such epigenetic processes and how they relate to cancer and cancer therapies more in-depth (**Table 1**).

Of course, the epigenetic processes described here do not occur as isolated events in a cell. Instead, a network of different context-dependent and collaborating or antagonizing epigenetic events establishes the respective outcomes. Epigenetic silencing and activation occur through various interdependent alterations on DNA and histones. For example, DNA methylation on transcriptionally permissive regions attracts methyl-binding domain proteins (e.g., MBD1 or MBD2), attracting other histone modifiers to the site, such as HDACs[77–79]. The process compounds the inhibitory effect on the transcription around the methylated region.

Similarly, DNA methylation and histone methylation mediated by PRC2 (H3K27me3) may anti-correlate at CpG islands in embryonic cells[80–83]. Indeed, studies suggest that PRC2 is unable to penetrate heavily CpG-methylated regions. By contrast, the removal of DNA methylation makes previously DNA-methylated regions more accessible to PRC2 and H3K27 methylation, including in promoter-distal regions. As a result, the increased accessibility dilutes PRC2 and reduces H3K27me3 at promoter sites, which may consequently increase gene activity[84–86]. Lastly, several *in vitro* and *in vivo* studies have identified specific interactions between repressive histone methylation by polycomb-group proteins and HDAC proteins that regulate transcriptional repression[87–89]. The findings suggest co-dependency between the two types of histone modifiers that can be accounted for in cancer therapy.

Table 1.

References	Title
Zhao S, Allis CD and Wang GG; 2021	The language of chromatin modification in human cancers[25]
Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, Han J and Wei X; 2019	Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials[15]
Biswas S and Rao CM; 2018	Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy[73]
Skvortsova K, Stirzaker C and Taberlay P; 2019	The DNA methylation landscape in cancer[74]
Wu D, Qiu Y, Jiao Y, Qiu Z and Liu D; 2020	Small Molecules Targeting HATs, HDACs, and BRDs in Cancer Therapy[75]
Michalak EM, Burr ML, Bannister AJ & Dawson MA ; 2019	The roles of DNA, RNA and histone methylation in ageing and cancer[76]

All these highly regulated and context-dependent events form an essential part of the human development and maintenance of cell fate and are sensitive to aberrant changes. The expression of these proteins needs to be maintained intact to ensure proper cell function. This includes "writers" like histone methyltransferases, "readers" like bromodomain-containing chromatin remodelers, or "erasers" like HDAC proteins. Their interconnected, complex pathways and interactions require new and efficient tools to reveal their exact roles and interactors in specific cancer types. A crucial advancement is large-scale next-generation sequencing, which has improved our ability to identify new interacting players in cancer. Together with chromatin-profiling sequencing advancement, such as chromatin immunoprecipitation sequencing (ChIP-seq) and assay for transposase-accessible chromatin using sequencing (ATAC-seq), functional genomics has allowed us to allocate new pieces of information to complete the epigenetic picture. Moreover, the advancements of new methods, such as CUT&RUN and CUT&Tag sequencing, which process protein–DNA associations *in situ*, have simplified and improved the analysis of post-translational modifications (PTMs). These methods are advantageous in that they have facilitated the profiling of PTMs in more physiologically relevant conditions and have enhanced the signal-to-noise ratio compared with, for example, ChIP-seq. Combining such ever-evolving techniques will allow researchers to decipher several unknowns in cancer from different perspectives; ultimately revealing the how, when, and where of cancer epigenetics.

THESIS OUTLINE

This thesis highlights the value of using different screens to investigate functional genomics in healthy and cancer cells. With a primary focus on the epigenome and epigenetic proteins, the thesis presents ideas and experiments that highlight the importance of using various large-scale genetic screens to decipher genotype–phenotype relationships. It demonstrates how these screening approaches – combined or independently – can provide valuable knowledge from different angles of cancer research, from improving biotechnological tools, such as gene-editing, to discovering novel genetic interactions that affect cancer drug sensitivity or resistance. Other tools have been overshadowed by the emergence of the diverse and highly efficient CRISPR–Cas9 tool. However, as the following chapters will demonstrate, each screening approach presents both strengths and weaknesses, and researchers should consider the available tools depending on the research question to complement their true values.

Chapter 2 gives an overview of the different genetic technologies and approaches that have laid the foundation of this thesis. The chapter describes, in particular, three mutagenesis screening tools – transposon screens, haploid genetic screens, and CRISPR screens – that can improve our understanding of the influence of the regulatory landscape in healthy and malignant cells.

Chapter 3 introduces the transposon-based screening tool Thousands of Reporters Integrated in Parallel, which was initially designed to study the effect of the regulatory landscape on transcription activity. We used the reporter system in mouse embryonic stem (mES) cells to investigate how the regulatory landscape influences the activity and properties of the gene-editing tool CRISPR–Cas9.

Chapter 4 presents how a drug resistance screen on mutagenized human haploid HAP1 cells identified resistance mechanisms to treatment with cytotoxic levels of a novel small-molecule inhibitor of BMI1 (PTC-318). The most significantly enriched mutation, *NUMA1*, was validated as a true inducer of PTC-318 resistance in both HAP1 cells and non-small cell lung cancer cell lines, providing new insight into the mitotic lethality associated with BMI1 downregulation.

Chapter 5 extrapolates the findings from Chapter 4 and combines data from two additional HAP1 screening methods, a depletion screen and a phenotype screen, to highlight a previously unknown role of chromatin remodeler WSTF as a positive regulator of H3K27me3 in cancer cells. The chapter shows how WSTF may directly or indirectly regulate an inverse correlation observed between proteins of closely related Polycomb repressive complexes PRC1 and PRC2.

Chapter 6 describes how a short-hairpin RNA (shRNA) screen with an epigenetics-focused library identified epigenetic proteins that act as a barrier to the reprogramming of somatic cells. The study highlights how TRIM28 acts as one of these reprogramming barriers, possibly by maintaining endogenous retroviruses (ERVs) silenced through H3K9me3 repression.

Chapter 7 closes with a general discussion of the technologies and results presented throughout the thesis. The chapter discusses the advantages and disadvantages of all the described technologies and how they can complement each other to extract the most information possible to enhance future research.

REFERENCES

1. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396–398 (2007).
2. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23**, 781–783 (2009).
3. Waterland, R. A. Epigenetic mechanisms and gastrointestinal development. *J. Pediatr.* **149**, 137–142 (2006).
4. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).
5. Suzuki, M. M. & Bird, A. DNA methylation landscapes: Provocative insights from epigenomics. *Nat. Rev. Genet.* **9**, 465–476 (2008).
6. Watt, F. & Molloy, P. L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* **2**, 1136–1143 (1988).
7. Ben-Hattar, J. & Jiricny, J. Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo. *Gene* **65**, 219–227 (1988).
8. Vandiver, A. R., Idrizi, A., Rizzardi, L., Feinberg, A. P. & Hansen, K. D. DNA methylation is stable during replication and cell cycle arrest. *Sci. Rep.* **5**, 17911 (2015).
9. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science (80-.).* **293**, 1089–1093 (2001).
10. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117**, 15–23 (2002).
11. Eckhardt, F. *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38**, 1378–1385 (2006).
12. Kitamura, E. *et al.* Analysis of tissue-specific differentially methylated regions (TDMs) in humans. *Genomics* **89**, 326–337 (2007).
13. Shi, M., Tsui, S. K. W., Wu, H. & Wei, Y. Pan-cancer analysis of differential DNA methylation patterns. *BMC Med. Genomics* **13**, 1–13 (2020).
14. De Bustos, C. *et al.* Tissue-specific variation in DNA methylation levels along human chromosome 1. *Epigenetics Chromatin* **2**, 7 (2009).
15. Cheng, Y. *et al.* Targeting epigenetic regulators for cancer therapy: Mechanisms and advances in clinical trials. *Signal Transduct. Target. Ther.* **4**, (2019).
16. Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene 2002 2135* **21**, 5427–5440 (2002).
17. Esteller, M. Epigenetics in Cancer. *N. Engl. J. Med.* **358**, 1148–1159 (2008).
18. Wang, L.-H., Wu, C.-F., Rajasekaran, N. & Shin, Y. K. Loss of Tumor Suppressor Gene Function in Human Cancer: An Overview. *Cell. Physiol. Biochem.* **51**, 2647–2693 (2018).
19. Christensen, B. C. *et al.* Breast cancer DNA methylation profiles are associated with tumor size and alcohol and folate intake. *PLoS Genet.* **6**, 1–10 (2010).
20. Santi, D. V., Norment, A. & Garrett, C. E. Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6993–6997 (1984).
21. Stresemann, C. & Lyko, F. Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. *International Journal of Cancer* **123**, 8–13 (2008).
22. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
23. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
24. Nair, N., Shoaib, M. & Sørensen, C. S. Chromatin Dynamics in Genome Stability: Roles in Suppressing Endogenous DNA Damage and Facilitating DNA Repair. *Int. J. Mol. Sci.* **18**, (2017).
25. Zhao, S., Allis, C. D. & Wang, G. G. The language of chromatin modification in human cancers. *Nat. Rev. Cancer 2021 217* **21**, 413–430 (2021).

26. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
27. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics* **20**, 207–220 (2019).
28. Barlev, N. A. *et al.* Acetylation of p53 Activates Transcription through Recruitment of Coactivators/Histone Acetyltransferases kinase CAK (Ko et al., 1997), and the phosphoinositol 3 kinase family members ATM (Banin et al) phosphorylate p53 in its amino terminus, leading to stablilization of p53. *Mol. Cell* **8**, 1243–1254 (2001).
29. Sharma, S. *et al.* Acetylation-Dependent Control of Global Poly(A) RNA Degradation by CBP/p300 and HDAC1/2. *Mol. Cell* **63**, 927–938 (2016).
30. Weinert, B. T. *et al.* Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. *Cell* **174**, 231 (2018).
31. Schölz, C. *et al.* Acetylation site specificities of lysine deacetylase inhibitors in human cells. *Nat. Biotechnol.* **33**, 415–425 (2015).
32. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science (80-.).* **325**, 834–840 (2009).
33. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
34. Legube, G. & Trouche, D. Regulating histone acetyltransferases and deacetylases. *EMBO Reports* **4**, 944–947 (2003).
35. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).
36. Wapenaar, H. & Dekker, F. J. Histone acetyltransferases: challenges in targeting bi-substrate enzymes. *Clin. Epigenetics 2016 81* **8**, 1–11 (2016).
37. Haynes, S. R. *et al.* The bromodomain: A conserved sequence found in human, Drosophila and yeast proteins. *Nucleic Acids Res.* **20**, 2603 (1992).
38. Jeanmougin, F., Wurtz, J. M., Le Douarin, B., Chambon, P. & Losson, R. The bromodomain revisited. *Trends in biochemical sciences* **22**, 151–153 (1997).
39. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
40. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet. 2007 393* **39**, 311–318 (2007).
41. Sheikh, B. N. & Akhtar, A. The many lives of KATs — detectors, integrators and modulators of the cellular environment. *Nat. Rev. Genet. 2018 201* **20**, 7–23 (2018).
42. Di Martile, M., Del Bufalo, D. & Trisciuoglio, D. The multifaceted role of lysine acetylation in cancer: Prognostic biomarker and therapeutic target. *Oncotarget* **7**, 55789–55810 (2016).
43. Li, Y. & Seto, E. HDACs and HDAC inhibitors in cancer development and therapy. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
44. Wang, P., Wang, Z. & Liu, J. Role of HDACs in normal and malignant hematopoiesis. *Mol. Cancer 2020 191* **19**, 1–21 (2020).
45. Sharda, A. *et al.* Elevated HDAC activity and altered histone phospho-acetylation confer acquired radio-resistant phenotype to breast cancer cells. *Clin. Epigenetics 2020 121* **12**, 1–17 (2020).
46. Bartling, B. *et al.* Comparative application of antibody and gene array for expression profiling in human squamous cell lung carcinoma. *Lung Cancer* **49**, 145–154 (2005).
47. Özdağ, H. *et al.* Differential expression of selected histone modifier genes in human solid cancers. *BMC Genomics* **7**, 90 (2006).
48. Milazzo, G. *et al.* Histone deacetylases (HDACs): Evolution, specificity, role in transcriptional complexes, and pharmacological actionability. *Genes* **11**, 556 (2020).
49. Vrana, J. A. *et al.* Induction of apoptosis in U937 human leukemia cells by suberoylanilide hydroxamic acid (SAHA) proceeds through pathways that are regulated by Bcl-2/Bcl-x(L), c-Jun, and p21(CIP1), but independent of p53. *Oncogene* **18**, 7016–7025 (1999).
50. Richon, V. M., Sandhoff, T. W., Rifkind, R. A. & Marks, P. A. Histone deacetylase inhibitor selectively induces p21WAF1

expressjon and gene-associated histone acetylation. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 10014–10019 (2000).

51. Sandor, V. *et al.* P21-dependent G1 arrest with downregulation of cyclin D1 and upregulation of cyclin E by the histone deacetylase inhibitor FR901228. *Br. J. Cancer* **83**, 817–825 (2000).
52. Minucci, S. & Pelicci, P. G. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nature Reviews Cancer* **6**, 38–51 (2006).
53. Kim, H. J. & Bae, S. C. Histone deacetylase inhibitors: Molecular mechanisms of action and clinical trials as anti-cancer drugs. *American Journal of Translational Research* **3**, 166–179 (2011).
54. Miller, C. P. *et al.* Therapeutic Strategies to Enhance the Anticancer Efficacy of Histone Deacetylase Inhibitors. *J. Biomed. Biotechnol.* **2011**, 17 (2011).
55. Eckschlager, T., Plch, J., Stiborova, M. & Hrabeta, J. Histone deacetylase inhibitors as anticancer drugs. *Int. J. Mol. Sci.* **18**, 1–25 (2017).
56. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).
57. Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12**, 1052–1058 (2002).
58. Min, J., Feng, Q., Li, Z., Zhang, Y. & Xu, R. M. Structure of the catalytic domain of human Dot1L, a non-SET domain nucleosomal histone methyltransferase. *Cell* **112**, 711–723 (2003).
59. Zhang, H. *et al.* Oncogenic deregulation of EZH2 as an opportunity for targeted therapy in lung cancer. *Cancer Discov.* **6**, 1007–1021 (2016).
60. Puppe, J. *et al.* EZH2 Is Overexpressed in BRCA1-like Breast Tumors and Predictive for Sensitivity to High-Dose Platinum-Based Chemotherapy. *Clin. Cancer Res.* **25**, 4351–4362 (2019).
61. Melling, N. *et al.* Overexpression of enhancer of zeste homolog 2 (EZH2) characterizes an aggressive subset of prostate cancers and predicts patient prognosis independently from pre- and postoperatively assessed clinicopathological parameters. *Carcinogenesis* **36**, 1333–1340 (2015).
62. Pawlyn, C. *et al.* Overexpression of EZH2 in multiple myeloma is associated with poor prognosis and dysregulation of cell cycle control. *Blood Cancer J.* **7**, (2017).
63. Li, B. & Chng, W.-J. EZH2 abnormalities in lymphoid malignancies: underlying mechanisms and therapeutic implications. *J. Hematol. Oncol. 2019 121* **12**, 1–13 (2019).
64. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629 (2002).
65. Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* **42**, 181–185 (2010).
66. Zhang, Q. *et al.* Mutations in EZH2 are associated with poor prognosis for patients with myeloid neoplasms. *Genes Dis.* **6**, 276–281 (2019).
67. Dimitrova, E., Turberfield, A. H. & Klose, R. J. "Histones and Chromatin" Review Series: Histone demethylases in chromatin biology and beyond. *EMBO Rep.* **16**, 1620 (2015).
68. Franci, G., Ciotta, A. & Altucci, L. The Jumonji family: Past, present and future of histone demethylases in cancer. *Biomolecular Concepts* **5**, 209–224 (2014).
69. Gang Xiao, Z. *et al.* The Roles of Histone Demethylase UTX and JMJD3 (KDM6B) in Cancers: Current Progress and Future Perspectives. *Curr. Med. Chem.* **23**, 3687–3696 (2016).
70. Majello, B., Gorini, F., Saccà, C. D. & Amente, S. Expanding the role of the histone lysine-specific demethylase lsd1 in cancer. *Cancers (Basel).* **11**, (2019).
71. Fang, Y., Liao, G. & Yu, B. LSD1/KDM1A inhibitors in clinical trials: advances and prospects. *J. Hematol. Oncol.* **12**, (2019).
72. FDA approves first treatment option specifically for patients with epithelioid sarcoma, a rare soft tissue cancer | FDA. Available at: https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-option-

specifically-patients-epithelioid-sarcoma-rare-soft-tissue. (Accessed: 16th May 2021)

73. Biswas, S. & Rao, C. M. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *European Journal of Pharmacology* **837**, 8–24 (2018).
74. Skvortsova, K., Stirzaker, C. & Taberlay, P. The DNA methylation landscape in cancer. *Essays Biochem.* **63**, 797–811 (2019).
75. Wu, D., Qiu, Y., Jiao, Y., Qiu, Z. & Liu, D. Small Molecules Targeting HATs, HDACs, and BRDs in Cancer Therapy. *Front. Oncol.* **0**, 2474 (2020).
76. Michalak, E. M., Burr, M. L., Bannister, A. J. & Dawson, M. A. The roles of DNA, RNA and histone methylation in ageing and cancer. *Nat. Rev. Mol. Cell Biol. 2019 2010* **20**, 573–589 (2019).
77. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).
78. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19**, 187–191 (1998).
79. Wade, P. A. *et al.* Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat. Genet.* **23**, 62–66 (1999).
80. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
81. Long, H. K. *et al.* Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife* **2013**, 1–19 (2013).
82. Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* **46**, 17–23 (2014).
83. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138 (2012).
84. Douillet, D. *et al.* Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation. *Nat. Genet.* **52**, 615–625 (2020).
85. McLaughlin, K. *et al.* DNA Methylation Directs Polycomb-Dependent 3D Genome Re-organization in Naive Pluripotency. *Cell Rep.* **29**, 1974-1985.e6 (2019).
86. Reddington, J. P. *et al.* Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* **14**, 1–17 (2013).
87. Van Der Vlag, J. & Otte, A. P. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat. Genet.* **23**, 474–478 (1999).
88. Zhang, X. *et al.* Coordinated Silencing of Myc-Mediated miR-29 by HDAC3 and EZH2 As a Therapeutic Target of Histone Modification in Aggressive B-Cell Lymphomas. *Cancer Cell* **22**, 506 (2012).
89. Fukumoto, T. *et al.* Repurposing Pan-HDAC Inhibitors for ARID1A-Mutated Ovarian Cancer. *Cell Rep.* **22**, 3393–3400 (2018).

Chapter 2

GENERAL INTRODUCTION II

INTRODUCTION II: DECIPHERING GENOTYPE–PHENOTYPE RELATIONSHIPS ASSOCIATED WITH EPIGENETICS THROUGH GENETIC MANIPULATION

To understand protein functionality and, ultimately, cellular characteristics, researchers employ different screening methods to decipher relationships between genotypes and phenotypes. Two characteristic strategies reveal these relationships in cells or animal models: genetic loss of function (LOF) and gain of function (GOF). As the names suggest, a LOF comprises any gene manipulation that reduces or inactivates the function of the gene product. In contrast, a GOF comprises manipulations that activate or enhance specific gene products. To this aim, researchers can evaluate how a specific cell or organism behaves upon specific manipulation and pinpoint unknown gene functionalities. For example:

- Does a cancer cell become more or less sensitive to a specific therapy after shutting off "gene X"?
- Do cells or organisms survive without "gene Y", and how do other gene expressions alter upon this gene loss?
- Can a cell regain functionality if "gene Z" is activated or enhanced?

To answer these questions, researchers can choose from an array of loss-of-function or gain-of-function approaches, including genetic screens, also known as mutagenesis screens. These genetic screens can, in turn, be divided into two types: reverse or forward genetic screens. A reverse genetic screen identifies the phenotype caused by specific, known mutations. On the other hand, a forward genetic screen identifies the genes responsible for cell properties by selecting phenotypes upon mutations and interrogating their genotype, for example, selecting the surviving cells in a cell population exposed to a toxic agent and mapping their mutations.

In cancer research, genetic screens can assess cancer phenotypes through either positive or negative selection of mutated cells. In a positive selection screen, cells are exposed to stringent selection to identify the genotype of the surviving, resistant cells. By contrast, in a negative selection screen, the goal is to identify the depleted mutations from a population. The latter form of screening is used when performing so-called synthetic lethality screens, which analyze cell death upon losing the expression of two or more genes. For example, the approach can assess the functionality of two genes, which independently are non-essential for cell survival but whose combined depletion causes cell lethality[1]. These strategies have therapeutic value, such as discovering drug targets in cells with inactivated tumor suppressor genes.

Prime model systems, such as the *Drosophila melanogaster* (fruit fly), served as the first models to identify genotype–phenotype relationships[2,3]. For instance, Nüsslein and Wieschaus introduced point mutations into the fruit fly's genome to identify 15 loci responsible for their larva development. Model systems like *Drosophila melanogaster*, *C. Elegans*, and *S. cerevisiae* (budding yeast) have significantly contributed to our understanding of apoptosis, cell cycle regulation, and cancer – and still are.

In recent years, thanks to the constant advancements in genetic screening and the seminal work performed on these model organisms, including the introduction of the first high throughput synthetic lethality screens, molecular biologists have been able to implement the knowledge on mammalian cells and organisms, exponentially enhancing our understanding of human genetics in healthy and pathogenic conditions.

Mutagenesis screens, epigenetics, and cancer

Since the occurrence of spontaneous mutations is rare, mutagenesis screens accelerate the rate at which researchers can assess and discover oncogenes and tumor suppressor genes. Both forward and reverse genetic screens can reveal cancer-related genes, whether associated with tumor development or suppression, in a matter of days or weeks. Hence, mutagenesis screens have offered us the proper means to study genome-wide gene interactions, regulation processes, and activities, which explain tumor properties.

Similarly, genetic screens can identify epigenetic features responsible for cancer development or suppression. Researchers can study the role of epigenetic genes or proteins in cancers *in vitro*, using different approaches. They can, for example, inhibit epigenetic protein expression or activity in mutagenized cancer cells to study mutation-specific resistance mechanisms, gene dependencies, or sensitivities. They can also incorporate genetic elements into the genome through insertional mutagenesis to study how the local chromatin affects cancer properties. The choice of specific screening tools depends on several factors, including the specific scientific question, required statistical power, and screen scale.

Three cutting-edge biological tools have emerged in the last years to screen for cancer-associated genes: transposon screens, haploid screens, and CRISPR screens. All three strategies can be used to search for genotype–phenotype interactions that define healthy and malignant cell characteristics.

Transposon screens and Thousands of Reporters Integrated in Parallel

Genome-wide transposon screens offer a versatile high-throughput alternative to identify genomic and epigenomic properties that control gene expression and cellular profiles. By randomly incorporating genetic elements, such as reporter genes, promoters, or enhancers, into the cellular genome, researchers can assess their position effects in the native chromatin environment. These genetic elements can induce loss of function or gain of function, depending on their orientations once incorporated into the genome. Compared with retroviral approaches, transposons evade cell or tissue tropism[4]. Incorporated genetic elements associated with transposon screens are easier to identify than traditional screening approaches, such as chemical or radiation strategies, thanks to the ability to implement barcodes and molecular fingerprints.

Researchers have genetically engineered recombinant transposons purposed for insertional mutagens in mammalian cells[5]. In normal mammalian conditions, most transposons are silent in the genome due to inactivating mutations or epigenetic silencing[6]. A typical engineered transposon system contains a transposon vector with inverted terminal repeat (ITR) sequences and a transposase that recognizes these repeat sequences. Upon recognition, the transposase excises the transposable element from a donor locus and moves the genetic material to specific chromosomal sites, a process sometimes referred to as a “cut-and-paste” mechanism. Researchers can investigate the phenotypic effects of specific mutations and insertional biases using specific tags (e.g., genetic barcodes) to locate the inserted transposons.

Several research groups have successfully used transposon vectors, such as piggyBac (PB) and Sleeping Beauty (SB), to identify drivers of cancer, metastatic markers, and treatment resistance[7]. The two types of transposon vectors differ in their integration bias and hopping properties, and researchers use them complementarily[8,9]. Transposon screens have become a cornerstone in genome-wide *in vitro* and *in vivo* screening, contributing to identifying genetic networks, new oncogenes, and tumor suppressor genes, including genes encoding epigenetic proteins (**Table 1**). Transposon screening carries enormous potential to expand our understanding of epigenetic activities and interactions in normal and malignant cells.

Implementing transposon screens to study epigenetic interactions in cancer

With the development of statistically powerful transposon screening tools, researchers can seamlessly decipher how the immediate epigenomic environment regulates gene activity and local interactions. These large-scale approaches help us highlight the roles of genomic or epigenomic landscapes in orchestrating gene functions and, as a result, cellular characteristics.

Our lab recently developed a high-throughput technology based on transposon mutagenesis called thousands of reporters integrated in parallel (TRIP)[10,11]. As the name suggests, TRIP integrates reporters at multiple genomic sites with distinct epigenomic features. Researchers can evaluate the position effect on such integrated genetic elements, such as reporter gene expression or promoter functionality, upon mapping the reporter sites. These types of screens present three main advantages for extracting functional genomic and epigenomic data. Firstly, TRIP transposons integrate randomly into sites previously mapped through chromatin immunoprecipitation (ChIP) data, enabling researchers to correlate loci and phenotypes with minimal genomic manipulation. Secondly, integration-specific barcoding offers the ability to study position effects in pooled populations without isolating clonal cell lines. Lastly, the design is highly flexible in that researchers can arrange sequence elements in multiple ways to study different types of processes. This flexibility can be established by including multiple promoter types to study transcription, methylated promoters to study DNA methylation or polycomb response elements (PREs) to study chromatin dynamics[10].

These experiments revealed the impact of local chromatin features on several genomic functions and elements, including gene expression, promoter activity, and enhancers. Notably, the findings confirmed that lamina-associated domains (LADs) reduced transcription and chromatin compaction partially regulates reporter activity[10]. A follow-up study extrapolated these findings by investigating so-called escaper promoters, which are maintained active despite being inside LADs. By randomly distributing promoters associated with LAD repression or escaper promoters in the genome of human K562 cells, the study could distinguish between intrinsic and extrinsic regulation by comparing the promotor activities in their new environments. Mapping the barcoded integrations revealed that both promoter sequence and the local chromatin environment influenced their sensitivity to LAD repression[12]. TRIP studies have also identified how enhancers promote distal gene activity over distances averaging approximately 20 kbp[10].

Near-random integration of reporter genes through TRIP has since paved the way to investigate how local chromatin features regulate specific genomic elements. These high-throughput studies provide sufficient data to confirm how the local environment affects gene expression, promoter activity, and enhancers[10,12,13]. The findings have also contributed to our fundamental understanding of how genomic and epigenomic environments regulate gene editing (Chapter 3 in this thesis)[14].

Transposon screens with *piggyBac* and *Sleeping Beauty* have successfully revealed mutated driver genes associated with cancer initiation in mouse models. Combined, several of these transposon screens have identified multiple genes encoding epigenetic regulators as interesting candidate driver genes of tumor initiation and maintenance in different cancer types **(Table 1)**.

For example, six independent studies have identified SWI-SNF genes, such as *ARID1A*, *ARID1B*, or *ARID5B*, implicated in different cancer types, either as driver genes or inducers of drug resisitance[15,16,24,25,28,29]. One of these studies examined gene candidates cooperating with RAS mutation (homozygous Hras^{G12V}) to induce thyroid tumorigenesis[15] and identified *ATXN7*, a component of histone acetylase complex, as well as *ARID1B* and *ARID2*, among 45 recurrent candidates. Mapping data from an *in vivo* transposon screen enabled Wu et al. to identify an Nf1-Stat3-Arid1b/β-catenin pathway responsible for driving neurofibroma, demonstrating how extensive, random mutagenesis screens can provide the necessary data to establish new pathways in cancer[24]. Another of these studies investigated drug resistance mechanisms to the chemotherapeutic drug fludarabine in chronic lymphocytic leukemia (CLL) and identified *ARID5B* among the top enriched inactivated genes[28]. The study was the first to investigate chemotherapy resistance through forward genetics in CLL, highlighting the benefit of using transposon screens to study chemotherapy sensitivity.

Apart from SWI-SNF, transposon screens have led to identifying several other genes encoding epigenetic proteins – particularly histone modifiers – being implicated in tumorigenesis or tumor maintenance. Among these, we find CpG island-associated

Table 1. Transposon mutagenesis screens identifying epigenetic hits in cancer (2015 to present)

Cancer type or cell type	**Transposon system**	**Background**	**Reference**
Thyroid cancer	Sleeping Beauty	Hras	Montero-Conde C, et al.[15]
Invasive lobular breast carcinoma	Sleeping Beauty	*Cdh*$^{fl/fl}$	Kas SM, et al.[16]
Triple negative breast cancer	Sleeping Beauty	Mutant *Pten*	Rangel R, et al.[17]
Pancreatic cancer	PiggyBac	*Kras*$^{LSL/G12D}$	Rad R, et al.[18]
B-cell acute lymphoblastic leukemia	Sleeping Beauty	Active *Stat5B*	Heltemes-Harris LM, et al.[19]
Melanoma	Sleeping Beauty	*BRAF*^{V600E}	Mann MB, et al.[20]
Colorectal cancer	Sleeping Beauty	*Tgfbr2*$^{fl/fl}$	Morris SM, et al.[21]
Gastric cancer	Sleeping Beauty	*Smad4*-KO	Takeda H, et al.[22]
Colorectal cancer	Sleeping Beauty	*Apc*^{min}, *Kras*^{G12D}, *Smad4*-KO, and *p53*^{R172H}	Takeda H, et al.[23]
Neurofibroma	Sleeping Beauty	*Nf1*$^{fl/fl}$	Wu J, et al.[24]
Prostate and skin cancer	Sleeping Beauty	*Pten*$^{SBm2/+}$; *Blm*$^{m3/m3}$	de la Rosa J, et al.[25]
Hepatocellular carcinoma	Sleeping Beauty	Mutant *Pten*	Kodama T, et al.[26]
Human colon cancer	Sleeping Beauty	hCEC-APCshRNA	Chen HJ, et al.[27]
Human chronic lymphocytic leukemia cells	PiggyBac	*APC*-KD; *Kras*^{G12D}	Pandzic T, et al.[28]
Hepatocellular carcinoma	Sleeping Beauty	*Kras*$^{G12D/+}$	Weber J, et al.[29]
B-cell lymphoma	PiggyBac	*Blm*^{m3}	Weber J, et al.[30]

demethylases and transcriptional repressors *Bmi1*, *Kdm2a,* and *Kdm6a* implicated in B-cell acute lymphoblastic leukemia or hepatocellular carcinoma[19,26]

Haploid gene-trap screens (insertional mutagenesis in haploid cell lines)

One of the main obstacles to retrieving reliable data from genetic screens in mammalian cells is their diploid (or sometimes aneuploid) nature. The ability of the homologous chromosome to genetically rescue gene changes in one allele can render the generation of complete gene silencing or homozygous mutations difficult.

Given their absence of a second set of chromosomes, haploid cells circumvent the genetic rescue effect and facilitate the generation of *bona fide* disruptions of gene functionality. The increased probability of establishing null mutations in these cell types makes them excellent tools for genetic screens probing mutation-driven phenotypic changes. Although yeast cells like fission yeast (or *Schizosaccharomyces pombe*) have long been valuable models to evaluate gene functions[31], their limited gene numbers can restrict their utility for human settings.

Haploid genetic screens to identify epigenetic interactions in cancer

Human haploid screening surged with the identification of human KBM7, a near-haploid cell line derived from a chronic myelogenous leukemia (CML) patient[32]. Except for chromosome 8 and a fraction of chromosome 15, the cells contain a haploid genome. A later attempt to establish pluripotent stem cells from KBM7 resulted in adherent HAP1 cells, which were haploid for all chromosomes except the 30-megabase fraction of chromosome 15 – a characteristic that Essletzbichler et al. later remedied using CRISPR–Cas9 excision[33,34]. Both KBM7 and HAP1 have been used successfully to identify genetic functions in various conditions, including virus entry mechanisms[35–37], T-cell mediated killing[38], and drug responses[39–41].

Although much of the research with KBM7 and HAP1 cells have focused on early signaling pathways, including viral entry, receptor signaling, or kinase pathways, studies examining the comparably slower epigenomic processes and their implication in cancer have slowly started to emerge **(Table 2)**. Although limited in numbers, these studies demonstrate how genome-wide disruption screens on human haploid cancer cells can provide extensive data on epigenetic interactions relevant to other forms of cancer. The current studies have been used to study phenotypic properties (e.g., post-translational histone modifications) or resistance to epigenetic inhibitors.

An example is heterochromatin mark H3K9me3, initially investigated through forward genetic screens in *Drosophila melanogaster*. However, while the simplicity of the fruit fly's genome has contributed to identifying heterochromatin regulators, such as Polycomb proteins[42] and heterochromatin protein 1 (HP1)[43], it does not entirely cover the more complex genome of humans.

Through genetic screens on KBM7 cells, Tchasovnikarova et al. were able to identify the human silencing hub (HUSH) complex as a previously unknown mediator of epigenetic

Table 2. Gene trap-based human haploid screens identifying epigenetic interactions in cancer

Cell line	Drug	Hits	Reference
KBM7	N/A	*SETDB1, TASOR, MPHOSPH8, PPHLN1*	Tchasovnikarova IA, et al.
HAP1	N/A	several	Brockmann M, et al.
KBM7	gemcitabine and entinostat	*DCK*	Ma YT, et al.
HAP1	PTC-318	*NUMA1*	Gisler S, et al.

repression[44]. Mapping gene-trap integrations in mutagenized KBM7 cells containing randomly incorporated GFP reporter genes served to identify mutation sites associated with "dim" (repressed) reporter expression. They showed that the loss of HUSH proteins reduced H3K9me3. Interestingly, the identified complex is absent in *Drosophila* but conserved from fish to humans, validating the reliability of human haploid cell lines to identify epigenetic proteins.

Indeed, HAP1 mutagenesis screens confirmed the links between genetic mutations and levels of specific epigenomic post-translational modifications[45]. Brockmann et al. expectedly found that gene-trap mutations in PRC2 genes *EED*, *EZH2*, and *SUZ12* were enriched in cell populations with low levels of H3K27me3. Similarly, they confirmed enrichment of EP300 mutations in populations with low histone crotonylation, a positive regulator of transcription implicated in acute kidney injury, HIV, and cancer[46]. Thus, the study highlights the connections between disease-associated mutations in genes that regulate post-transcriptional modifications and phenotype changes that do not affect cell viability or fitness. It also underscores the power of human haploid cell screens to investigate epigenetic interactions in normal conditions and diseases.

In cancer treatment and resistance, haploid genetic screens have served to identify enrichment of specific genetic mutations upon inhibition of epigenetic proteins. For example, a KBM7 screen revealed insertional enrichment of deoxycytidine kinase (DCK) upon treatment with HDAC inhibitors gemcitabine and entinostat[47]. These loss-of-function mutations were later validated through Cas9 knockouts of *DCK* in pancreatic cancer cell lines to induce resistance to the gemcitabine with or without entinostat. Through a genetic screen in HAP1 cells, our group identified that loss of mitosis gene *NUMA1* was enriched in cells treated with cytotoxic concentrations of BMI1 inhibitor PTC-318 (Chapter 4 in this thesis). Cas9-induced knockout of *NUMA1* in non-small cell lung cancer cell lines later confirmed NUMA1-associated resistance to BMI1 inhibitor PTC-318. These examples highlight the reliable outcomes of these genetic screening approaches to identify novel genetic interactions. Human haploid cells may ultimately capture novel interactions that can help distinguish genes that determine responses to targeted therapies against epigenetic proteins.

CRISPR screens to identify cancer pathways

Among the current genome-wide screening tools, clustered regularly interspaced short palindromic repeat (CRISPR) probably impacted the molecular genetics field the most. Implementing this procaryotic immune-defense system in eukaryotic cells introduced a simple and cost-effective strategy for focused gene targeting. Spearheaded by endonuclease Cas9, the CRISPR technology quickly became the go-to tool for studying genetic interactions, gene function, and gene dependencies. Researchers apply CRISPR–Cas9 to seamlessly extract large numbers of gene candidates responsible for cell survival, proliferation, or drug resistance[48].

The technology has gained much of its popularity thanks to the relative ease with which researchers can design and produce DNA-specific guide RNA (gRNA)–Cas9 complexes and the high efficacy and site-specificity of these systems. Gene editing with CRISPR–Cas9 relies on two functional components: the endonuclease Cas9 and a gRNA that recognizes the target sequence (usually in DNA). Once bound to the correct target site, the gRNA–Cas9 complex induces a double-stranded break that can result in small insertions or deletions (indels) due to cell-intrinsic errors occurring during the repair process[49]. Researchers take advantage of these sporadic indels to study gene dependencies through loss of gene function, both on a small and large scale.

Shortly after the first single-targeting CRISPR–Cas9 experiments in mammalian cells[50,51], several groups developed libraries of gRNAs for small-scale or large-scale screens[52]. Both genome-wide and focused, positive-selection and negative-selection screens have been successfully used with such generated libraries. Cas9 or gRNA can be delivered separately or combined into eukaryotic cells through lentivirus transduction to induce highly efficient loss-of-function mutations with relatively low off-target events[53]. These features and the ability to multiplex the tool enabled an early implementation of CRISPR–Cas9 as a screening tool to characterize gene essentiality[54,55] and cancer-specific drug targets in human cells[56–58].

Using genome-wide and focused sgRNA libraries to discover the involvement of epigenetic regulators in cancer and cancer treatment

For the reasons mentioned above, the emergence of CRISPR–Cas9 and its immediate adoption for functional genetic screens have accelerated the discovery rate of new genetic and epigenetic functions and interactions in cancer. These advances have also been made possible by developing focused libraries targeting epigenetic protein-encoding genes. For example, Henser-Brownhil et al. created a focused single-guide RNA (sgRNA) library targeting 450 epigenetic regulators in human cells[59]. The library enabled the investigators to saturate their screen with 3 million cells compared with approximately 200 million cells required for genome-wide screens. They targeted most of the selected genes with seven sgRNAs with the focused library, which can be compared with the two sgRNAs per targeted gene in the genome-wide library.

Both genome-wide and epigenetic-focused sgRNA libraries have contributed to identifying epigenetic mechanisms involved in cancer-associated processes. Such libraries have successfully highlighted the roles of epigenetic candidates in cancer-drug resistance, synthetic lethality, cancer dependencies, and as immunotherapeutic targets **(Table 3)**.

Several studies have extracted data of gene candidates essential for different cancer types, which could serve as therapeutic targets. In particular, many of these studies have validated possible therapeutic targets for acute myeloid leukemia, one of the most common and aggressive forms of leukemia with low survival rates[61–64,77]. The collected data presents several viable epigenetic drug candidates for acute myeloid leukemia (AML) treatment, including recurring candidates from separate studies, such as bromodomain and extraterminal (BET) family member BRD4, enriched in three of the depletion screens (mouse and human), and DOT1L. A recent study targeted exons encoding the functional domains of proteins, such as the catalytic domain of epigenetic proteins, to create in-frame variants with intact gene expression[77]. The screen assessed cancer cell dependency of 192 genes in an AML cell line, using a library of 933 sgRNAs targeting chromatin regulatory domains. The approach identified nineteen unknown dependencies and six known drug candidates for AML (Brd4, Dot1l, Ehmt1, Ehmt2, Ezh2, and Kdm1a). Interestingly, deep sequencing analysis comparing methyltransferase domains and 5′ exons of *Dot1L* and *Ezh2* confirmed that in-frame mutations in functional domains enhanced the selection approach.

Benahan et al. performed the largest screen, using genome-wide human CRISPR libraries to identify cancer gene dependencies in 30 cancer types (324 cancer cell lines)[58]. The study offers an impressive data collection of druggable cancer gene candidates, including several genes encoding epigenetic proteins.

Genome-wide CRISPR libraries have also successfully identified epigenetic candidate genes that are part of synthetic lethal interactions in cancer or confer resistance to different inhibitors. The synthetic genetic screen studies assessed sgRNA depletion in cells with mutated[65] or loss of[66] cancer-specific genes, or cells treated with known inhibitors[68] or introduced with paired sgRNA libraries[67]. In the latter study, researchers used approximately 490,000 sgRNA pairs to identify druggable synthetic lethal partners, including interactions between members of the polycomb group of proteins (EED, EZH2, and BMI1) and several candidate genes. In contrast to these depletion screens, drug resistance screens identify enriched CRISPR–Cas9-induced gene disruptions among cells treated with selected inhibitors. For example, among known genes implicated in resistance to BRAF inhibitor vemurafenib in melanoma, Shalem et al. also identified two members of the histone acetyltransferase complex STAGA (*TADA1* and *TADA2B*), which were previously unknown in the context of vemurafenib[69].

Additionally, CRISPR–Cas9 screens have been demonstrated to be a viable tool to successfully identify key epigenetic players that can serve as targets to enhance immunotherapy. Two separate *in vitro* studies recently revealed epigenetic gene

Table 3. CRISPR–Cas9 screens identifying epigenetic hits in cancer

	Cancer type	Screen type	Library	Mutation/Drug	Epigenetic hits	Reference
in vitro	Acute myeloid leukemia	Depletion	Custom	N/A	*Dot1l, Ehmt1, Ehmt2, Ezh2, Brd4, Kdm1a*	Shi J, et al.[60]
	Acute myeloid leukemia	Depletion	Murine lentiviral gRNA (version 2)	N/A	*BRD4, DOT1L,*	Tzelepis K, et al.[61]
	324 cell lines from 30 cancer types	Depletion	Human CRISPR libraries v.1.0 and v.1.1	N/A	several	Behan FM, et al.[58]
	Acute myeloid leukemia	Depletion	GeCKOv2	N/A	*Brd4, Brd9*	Yamauchi T, et al.[62]
	Acute myeloid leukemia	Depletion	GeCKO	N/A	*ENL*	Erb MA, et al.[63]
	Acute myeloid leukemia	Depletion	Custom	N/A	*SETDB1*	Cuellar TL, et al.[64]
	Colorectal cancer	Synthetic lethality	GeCKOv2	KRAS	*INO80C*	Yau EH, et al.[65]
	Hepatocellular carcinoma	Synthetic lethality	GeCKOv2	ATRX loss	*SMC5/6, ASF1A, CABIN1, HAT1*	Liang J, et al.[66]
	Chronic myeloid leukemia	Synthetic lethality	Paired single-guide RNA libraries	N/A	interactions with *HDAC8, EZH2, EED,*or *BMI1*	Han K, et al.[67]
	Triple-negative breast cancer	Synthetic lethality	H1 and H2	BET bromodomain inhibitors	*BRD2*	Shu S, et al.[68]
	Melanoma	Drug resistance	GeCKO	Vemurafenib	*TADA1, TADA2B*	Shalem O, et al.[69]
	Acute myeloid leukemia	Drug resistance	GeCKO	Quizartinib	*HDAC5*	Hou P, et al.[70]
	Chronic myeloid leukemia	Immunetherapy targets	Bassik Human CRISPR-KO	N/A	*EED, SUZ12*	Burr ML, et al.[71]
	Melanoma	Immunetherapy targets	Mouse CRISPR Brie lentiviral pooled	N/A	*Pbrm1, Arid2, Brd7*	Pan D, et al.[72]

Table 3. (continued)

	Cancer type	**Screen type**	**Library**	**Mutation/Drug**	**Epigenetic hits**	**Reference**
in vivo	Non-small cell lung cancer	Enrichment	Custom	*Kras*$^{G12D/+}$	*Utx*	Wu Q, et al.[73]
	Lung adenocarcinoma	Immunetherapy targets	epigenetic-focused sgRNA	anti-PD1	*Asf1a*	Li F, et al.[74]
	Regulatory T cells	Immunotherapy targets	Custom	N/A	*Usp22, Rnf20*	Cortez, et al.[75]
	Regulatory T cells	Immunotherapy targets	lentiCRISPRv2-Brie	N/A	*Several: e.g., Usp22, Arid1a, Brd9, Smarcd1.*	Loo, et al.[76]

candidates responsible for immune evasion associated with immunotherapy. One of the studies highlighted the role of polycomb protein in silencing MHC Class I and how inhibition of EED or EZH2 and EZH1 reverses the silencing, suggesting that polycomb repressive complex (PRC)2 inhibition may enhance immunotherapy in treating MHC-I-deficient cancer[71]. Similarly, through a genome-wide screen in mouse melanoma, Pan et al. found that inactivation of SWI-SNF genes *Pbrm1*, *Arid2*, and *Brd7* sensitized the cells to T-cell killing. Furthermore, two genome-wide loss-of-function CRISPR-screen studies in primary mouse regulatory T (Treg) cells highlighted several epigenetic regulators that either promote or disrupt the expression of Foxp3, a member of the FOX protein family and suppressor of over-reactive effector T cells[75,76]. The studies highlight gene candidates with the potential to control Treg activity, which can be implicated in immunotherapy for cancer or regulation of autoimmunity. Both studies identified members of the SAGA complex and demonstrated compromised Treg-mediated cell suppression upon *Usp22* knockout *in vivo*. Moreover, apart from the SAGA-related hits, Loo et al. found specific and opposing Foxp3-regulating roles between proteins belonging to the non-canonical BAF (ncBAF) and PBAF subcomplexes of the SWI-SNF family. Specifically, while the PBAF complex repressed Foxp3, genes encoding ncBAF complex proteins, such as Brd9, were shown to promote Foxp3. The findings, which were subsequently validated in an MC38 colorectal cancer cell-induced mouse model, raise the possibility of slowing tumor growth by indirectly downregulating Treg cells with Brd9-specific small-molecule inhibitors.

CRISPR screens have also been valuable to explore genetic and epigenetic interactions and drug candidates in cancer *in vivo*[78,79]. Compared with *in vitro* screens, *in vivo* screens have the added advantage of accounting for the interactions between cancer cells and the host microenvironment[80].

Several studies have reported *in vivo* applications of CRISPR–Cas9 to identify new mechanisms and drug targets related to cancer, including epigenetic implications. For example, an *in vivo* screen targeting epigenetic genes discovered histone chaperone Asf1a to regulate *Kras*-mutant lung adenocarcinoma sensitivity against treatment with PD-1 inhibitor[74]. A separate *in vivo* screen in a $Kras^{G12D/+}$ lung cancer mouse model identified a previously unknown role of histone demethylase Utx as a tumor suppressor in lung cancer[73]. Further validations revealed that Utx knockout enhanced the expression and activity of Ezh2 and demonstrated that pharmacological inhibition of Ezh2 suppressed the growth of Utx-deficient tumors. These studies serve to exemplify the development of CRISPR–Cas9 and its rapid application to identify the role of epigenetic proteins in cancer.

Concluding remarks

Functional genomics and mutagenesis screens have become vital to understanding biological systems and disease mechanisms advancing drug development. While the introduction of Sanger sequencing provided researchers with the means to identify genetic mutations starting in the 1970s[81,82], genetic screens were tedious, pricy, and time-consuming.

The completion of the Human Genome Project in 2003, covering nearly the complete human genome, cleared the path for large-scale genetic screens and the current discovery rate of genotype–phenotype relationships[83]. The Human Genome Project laid a foundation for these discoveries through the available genetic information (the reference genome) that researchers today apply for genetic screens. These screens include the large-scale mutagenesis approaches reviewed here, which are powerful techniques to identify genomic and epigenomic interactions that regulate healthy and malignant phenotypes. I highlight three screening approaches that can reveal functional genomic interactions underpinning cancer initiation, progression, and metastasis, as well as associated resistance mechanisms. These approaches can be used individually or combined to investigate genes that confer sensitivity or resistance to specific drugs. A recent study from our lab also used reporter-based insertional mutagenesis screens to understand how the genome and epigenome affect CRISPR–Cas9 editing[14], showing the wide applicability functional genomics brought about.

The resulting fast-paced discovery of new molecular interactions and mechanisms, as well as drug candidates, enabled by large-scale screens, also give rise to new unknowns. Faster and simpler approaches will, hence, continue to be in demand. As a result, functional genomics and large-scale mutagenesis screens will most likely continue advancing with the emergence of new screening approaches or the complementation of one screening approach with another. The following chapters will guide you through different strategies we have used to enhance molecular genetics and cancer research. To this aim, we have used variations of the screening approaches described in this chapter, either individually or combined. Hopefully, these may trigger new ideas on enhancing the search for genetic and epigenetic interactions in cancer and the development of new and more genotype-specific treatments for cancer.

REFERENCES

1. Huang, A., Garraway, L. A., Ashworth, A. & Weber, B. Synthetic lethality as an engine for cancer drug target discovery. *Nature Reviews Drug Discovery* **19**, 23–38 (2020).
2. Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in Drosophila. *Nat. 1980 2875785* **287**, 795–801 (1980).
3. Ohnishi, O. Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in Drosophila melanogaster. II. Homozygous effect of polygenic mutations. *Genetics* **87**, 529–545 (1977).
4. Friedrich, M. J. *et al.* Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nat. Protoc.* **12**, 289–309 (2017).
5. Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvák, Z. Molecular reconstruction of sleeping beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501–510 (1997).
6. Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **19**, 1419–1428 (2009).
7. DeNicola, G. M., Karreth, F. A., Adams, D. J. & Wong, C. C. The utility of transposon mutagenesis for cancer studies in the era of genome editing. *Genome Biol. 2015 161* **16**, 1–15 (2015).
8. Liang, Q., Kong, J., Stalker, J. & Bradley, A. Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. *Genesis* **47**, 404–408 (2009).
9. Ding, S. *et al.* Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473–483 (2005).
10. Akhtar, W. *et al.* Chromatin Position Effects Assayed by Thousands of Reporters Integrated in Parallel. *Cell* **154**, 914–927 (2013).
11. Akhtar, W. *et al.* Using TRIP for genome-wide position effect analysis in cultured cells. *Nat. Protoc. 2014 96* **9**, 1255–1281 (2014).
12. Leemans, C. *et al.* Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs. *Cell* **177**, 852-864.e14 (2019).
13. Brueckner, L., Van Arensbergen, J., Akhtar, W., Pagie, L. & Van Steensel, B. High-throughput assessment of context-dependent effects of chromatin proteins. *Epigenetics and Chromatin* **9**, (2016).
14. Gisler, S. *et al.* Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nat. Commun.* **10**, (2019).
15. Montero-Conde, C. *et al.* Transposon mutagenesis identifies chromatin modifiers cooperating with Ras in thyroid tumorigenesis and detects ATXN7 as a cancer gene. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4951–E4960 (2017).
16. Kas, S. M. *et al.* Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma. *Nat. Genet.* **49**, 1219–1230 (2017).
17. Rangel, R. *et al.* Transposon mutagenesis identifies genes that cooperate with mutant Pten in breast cancer progression. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E7749 (2016).
18. Rad, R. *et al.* A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nat. Genet.* **47**, 47–56 (2015).
19. Heltemes-Harris, L. M. *et al.* Sleeping Beauty transposon screen identifies signaling modules that cooperate with STAT5 activation to induce B cell acute lymphoblastic leukemia. *Oncogene* **35**, 3454 (2016).
20. Mann, M. B. *et al.* Transposon mutagenesis identifies genetic drivers of BrafV600E melanoma. *Nat. Genet.* **47**, 486 (2015).
21. Morris, S. M. *et al.* Transposon mutagenesis identifies candidate genes that cooperate with loss of Transforming Growth Factor-beta signaling in mouse intestinal neoplasms. *Int. J. cancer* **140**, 853 (2017).
22. Takeda, H. *et al.* Sleeping Beauty transposon mutagenesis identifies genes that cooperate with mutant Smad4 in gastric cancer development. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2057 (2016).

23. Takeda, H. *et al.* Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor progression. *Nat. Genet. 2014 472* **47**, 142–150 (2015).
24. Wu, J. *et al.* Insertional mutagenesis identifies a STAT3/Arid1b/β-catenin pathway driving neurofibroma initiation. *Cell Rep.* **14**, 1979 (2016).
25. Rosa, J. de la *et al.* A single-copy Sleeping Beauty transposon mutagenesis screen identifies new PTEN-cooperating tumor suppressor genes. *Nat. Genet.* **49**, 730 (2017).
26. Kodama, T. *et al.* Transposon mutagenesis identifies genes and cellular processes driving epithelial-mesenchymal transition in hepatocellular carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3384 (2016).
27. Chen, H. J. *et al.* A recellularized human colon model identifies cancer driver genes. *Nat. Biotechnol.* **34**, 845 (2016).
28. Pandzic, T. *et al.* Transposon Mutagenesis Reveals Fludarabine Resistance Mechanisms in Chronic Lymphocytic Leukemia. *Clin. Cancer Res.* **22**, 6217–6227 (2016).
29. Weber, J. *et al.* CRISPR/Cas9 somatic multiplex-mutagenesis for high-throughput functional cancer genomics in mice. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13982 (2015).
30. Weber, J. *et al.* PiggyBac transposon tools for recessive screening identify B-cell lymphoma drivers in mice. *Nat. Commun.* **10**, (2019).
31. Hoffman, C. S., Wood, V. & Fantes, P. A. An Ancient Yeast for Young Geneticists: A Primer on the Schizosaccharomyces pombe Model System. *Genetics* **201**, 403–423 (2015).
32. Kotecki, M., Reddy, P. S. & Cochran, B. H. Isolation and characterization of a near-haploid human cell line. *Exp. Cell Res.* **252**, 273–280 (1999).
33. Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* (2011). doi:10.1038/nature10348
34. Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* **24**, 2059–2065 (2014).
35. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science (80-.).* (2009). doi:10.1126/science.1178955
36. Jae, L. T. *et al.* Lassa virus entry requires a trigger-induced receptor switch. *Science (80-.).* (2014). doi:10.1126/science.1252480
37. Staring, J. *et al.* PLA2G16 represents a switch between entry and clearance of Picornaviridae. *Nature* (2017). doi:10.1038/nature21032
38. Mezzadra, R. *et al.* SLFN11 can sensitize tumor cells towards IFN-γ-mediated T cell killing. *PLoS One* (2019). doi:10.1371/journal.pone.0212053
39. Lee, C. C., Carette, J. E., Brummelkamp, T. R. & Ploegh, H. L. A Reporter Screen in a Human Haploid Cell Line Identifies CYLD as a Constitutive Inhibitor of NF-κB. *PLoS One* (2013). doi:10.1371/journal.pone.0070339
40. Planells-Cases, R. *et al.* Subunit composition of VRAC channels determines substrate specificity and cellular resistance to P t-based anti-cancer drugs . *EMBO J.* (2015). doi:10.15252/embj.201592409
41. Gisler, S., Maia, A. R. R., Chandrasekaran, G., Kopparam, J. & Lohuizen, M. van. A genome-wide enrichment screen identifies NUMA1-loss as a resistance mechanism against mitotic cell-death induced by BMI1 inhibition. *PLoS One* **15**, e0227592 (2020).
42. Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nat. 1978 2765688* **276**, 565–570 (1978).
43. James, T. C. & Elgin, S. C. Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. *Mol. Cell. Biol.* **6**, 3862 (1986).
44. Tchasovnikarova, I. A. *et al.* Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science (80-.).* **348**, 1481–1485 (2015).
45. Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal

an off-switch for GPCR signalling. *Nat. 2017 5467657* **546**, 307–311 (2017).

46. Wan, J., Liu, H., Chu, J. & Zhang, H. Functions and mechanisms of lysine crotonylation. *J. Cell. Mol. Med.* **23**, 7163–7169 (2019).
47. Ma, Y. T. *et al.* Use of a genome-wide haploid genetic screen to identify treatment predicting factors: a proof-of-principle study in pancreatic cancer. *Oncotarget* **8**, 63635–63645 (2017).
48. He, C. *et al.* CRISPR screen in cancer: status quo and future perspectives. *Am. J. Cancer Res.* **11**, 1031 (2021).
49. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, 471 (2013).
50. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819 (2013).
51. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823 (2013).
52. Addgene: CRISPR Pooled gRNA Libraries. Available at: https://www.addgene.org/crispr/libraries/. (Accessed: 18th July 2021)
53. Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34**, 631–633 (2016).
54. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science (80-.).* **343**, 80–84 (2014).
55. Wang, T. *et al.* Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **168**, 890-903.e15 (2017).
56. Housden, B. E. *et al.* Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. *Sci. Signal.* **8**, (2015).
57. Parnas, O. *et al.* A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**, 675–686 (2015).
58. Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nat. 2019 5687753* **568**, 511–516 (2019).
59. Henser-Brownhill, T., Monserrat, J. & Scaffidi, P. Generation of an arrayed CRISPR-Cas9 library targeting epigenetic regulators: from high-content screens to in vivo assays. *Epigenetics* **12**, 1065–1075 (2017).
60. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661 (2015).
61. Tzelepis, K. *et al.* A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep.* **17**, 1193 (2016).
62. Yamauchi, T. *et al.* Genome-wide CRISPR-Cas9 Screen Identifies Leukemia-Specific Dependence on a Pre-mRNA Metabolic Pathway Regulated by DCPS. *Cancer Cell* **33**, 386-400.e5 (2018).
63. Erb, M. A. *et al.* Transcription control by the ENL YEATS domain in acute leukemia. *Nature* **543**, 270 (2017).
64. Cuellar, T. L. *et al.* Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. *J. Cell Biol.* **216**, 3535 (2017).
65. Yau, E. H. *et al.* Genome-wide CRISPR screen for essential cell growth mediators in mutant KRAS colorectal cancers. *Cancer Res.* **77**, 6330 (2017).
66. Liang, J. *et al.* Genome-Wide CRISPR-Cas9 Screen Reveals Selective Vulnerability of ATRX-Mutant Cancers to WEE1 Inhibition. *Cancer Res.* **80**, 510–523 (2020).
67. Han, K. *et al.* Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* **35**, 463 (2017).
68. Shu, S. *et al.* Synthetic lethal and resistance interactions with BET bromodomain inhibitors in triple-negative breast cancer. *Mol. Cell* **78**, 1096 (2020).
69. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84 (2014).
70. Hou, P. *et al.* A genome-wide CRISPR screen identifies genes critical for resistance to FLT3 inhibitor AC220. *Cancer Res.* **77**, 4402 (2017).
71. Burr, M. L. *et al.* An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer. *Cancer Cell* **36**, 385 (2019).

72. Pan, D. *et al.* A major chromatin regulator determines resistance of tumor cells to T cell–mediated killing. *Science* **359**, 770 (2018).
73. Wu, Q. *et al.* In vivo CRISPR screening unveils histone demethylase UTX as an important epigenetic regulator in lung tumorigenesis. *Proc. Natl. Acad. Sci.* **115**, E3978–E3986 (2018).
74. Li, F. *et al.* In vivo epigenetic CRISPR screen identifies Asf1a as an immunotherapeutic target in Kras-mutant lung adenocarcinoma. *Cancer Discov.* **10**, 270 (2020).
75. Cortez, J. T. *et al.* CRISPR screen in regulatory T cells reveals modulators of Foxp3. *Nature* **582**, 416–420 (2020).
76. Loo, C. S. *et al.* A Genome-wide CRISPR Screen Reveals a Role for the Non-canonical Nucleosome-Remodeling BAF Complex in Foxp3 Expression and Regulatory T Cell Function. *Immunity* **53**, 143-157.e8 (2020).
77. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).
78. Szlachta, K. *et al.* CRISPR knockout screening identifies combinatorial drug targets in pancreatic cancer and models cellular drug response. *Nat. Commun. 2018 91* **9**, 1–13 (2018).
79. Winters, I. P., Murray, C. W. & Winslow, M. M. Towards quantitative and multiplexed in vivo functional cancer genomics. *Nat. Rev. Genet. 2018 1912* **19**, 741–755 (2018).
80. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nature Medicine* **19**, 1423–1437 (2013).
81. Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, (1975).
82. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463 (1977).
83. Abdellah, Z. *et al.* Finishing the euchromatic sequence of the human genome. *Nat. 2004 4317011* **431**, 931–945 (2004).

Chapter 3

MULTIPLEXED CAS9 TARGETING REVEALS GENOMIC LOCATION EFFECTS AND gRNA-BASED STAGGERED BREAKS INFLUENCING MUTATION EFFICIENCY

Santiago Gisler[1*], Joana P. Gonçalves[2,3*], Waseem Akhtar[1*], Johann de Jong[2],
Alexey V. Pindyurin[4,5], Lodewyk F. A. Wessels[2,3], Maarten van Lohuizen[1]

[1] Division of Molecular Genetics, Oncode and The Netherlands Cancer Institute, Amsterdam, The Netherlands
[2] Department of Intelligent Systems, Delft University of Technology, Delft, The Netherlands
[3] Division of Molecular Carcinogenesis, Oncode and The Netherlands Cancer Institute, Amsterdam, The Netherlands
[4] Institute of Molecular and Cellular Biology, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
[5] Division of Gene Regulation, Oncode and The Netherlands Cancer Institute, Amsterdam, The Netherlands
* These authors contributed equally to this work

Adapted from Nat Commun. 2019 Apr 8;10(1):1598

ABSTRACT

Understanding the impact of guide RNA (gRNA) and genomic locus on CRISPR–Cas9 activity is crucial to designing effective gene editing assays. However, it is challenging to profile Cas9 activity in the endogenous cellular environment. Here we leverage our TRIP technology to integrate ~ 1k barcoded reporter genes in the genomes of mouse embryonic stem cells. We target the integrated reporters (IRs) using RNA-guided Cas9 and characterize induced mutations by sequencing. We report that gRNA-sequence and IR locus explain most variation in mutation efficiency. Predominant insertions of a gRNA-specific nucleotide are consistent with the template-dependent repair of staggered DNA ends with 1-bp 5′ overhangs. We confirm that such staggered ends are induced by Cas9 in mouse pre-B cells. To explain observed insertions, we propose a model generating primarily blunt and occasionally staggered DNA ends. Mutation patterns indicate that gRNA-sequence controls the fraction of staggered ends, which could be used to optimize Cas9-based insertion efficiency.

INTRODUCTION

Genome engineering has seen considerable progress in recent years, nurtured by the emergence of precision editing tools based on the bacterial clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9. The CRISPR–Cas9 system complexes the endonuclease enzyme Cas9 with a guide RNA to induce double-strand breaks (DSBs) at a specific DNA locus[1–5]. For target DNA recognition and binding, CRISPR–Cas9 requires the presence of a short, conserved sequence known as a protospacer-adjacent motif (PAM). The PAM consists of nucleotides NGG and is located downstream of the target sequence[6,7].

Cas9-induced DSBs activate the cellular DNA damage response, mainly through non-homologous end-joining (NHEJ) or homology-directed repair (HDR)[8]. NHEJ is the most common DNA repair pathway[9]. In NHEJ, DNA ends are processed independently without a template prior to ligation, often producing mutations at the break site. HDR relies on sequence homology for repair and therefore depends on the availability of a donor DNA template, which can be acquired from the sister chromatid in S-phase[8]. In genome engineering applications, error-prone repair of Cas9-induced DSBs can be exploited to disrupt the target sequence and generate gene knockouts. Exogenous genetic material can also be integrated into host DNA by providing repair templates with custom oligonucleotides flanked by homology arms.

The CRISPR–Cas9 technology is used extensively for gene editing in vitro and in vivo, yet most factors controlling its nuclease activity are poorly understood. While effects of guide RNA on Cas9 nuclease efficiency and target specificity have been extensively characterized[10–14], the influence of target sequence on induced mutation patterns remains unclear. Little is also known on the impact of genomic and epigenomic context at the target locus[15]. Early studies found that chromatin accessibility or DNA methylation affects the binding of catalytically inactive Cas9 (dCas9)[16–18]. Others showed that Cas9 binding and cleavage are sensitive to chromatin changes induced by nucleosome occupancy[19,20] or administration of doxycycline[21]. Most literature suggests that genomic context influences Cas9 binding and cleavage, but effects on editing efficiency are less well understood. Several studies have observed weak correlations between epigenomic context and Cas9-induced mutation frequency at endogenous targets[17,22–24]. In particular, two of these studies showed that the significant effect of epigenomic context on Cas9 binding did not necessarily result in a detectable effect on Cas9-induced mutation frequency[17,22].

Here, we characterize Cas9-induced mutations in the genomes of mouse embryonic stem (mES) cells. We aim to survey many loci for sufficient statistical power with minimal disruption of the native environment. However, it is not trivial to scale up the number of endogenous Cas9 targets. Compromising on guide RNA specificity increases off-target effects. Targeting repetitive sequences creates challenges for alignment and mutation calling, and results might not generalize to other kinds of sequences. Both approaches can generate a large number of cleavage events per cell, eventually leading to genomic instability and unreliable Cas9 activity profiling as a result. While multiple guides could instead be used to expand Cas9 targeting, this would also introduce target heterogeneity. Alternatively, we integrate thousands of barcoded target sequences throughout the genomes of a population of mES cells using our TRIP technology[25,26]. In this way, we multiplex Cas9 cleavage while keeping the number of targets per cell under control. The use of TRIP reporters further enables sequence-independent analysis of effects across the targeted loci. We investigate the usefulness of these hybrid exogenous-endogenous loci to profile Cas9 activity and assess the impact of guide RNA sequence and targeted locus on induced mutation frequency and patterns.

RESULTS

We profiled CRISPR–Cas9-induced mutations across the genomes of mES cells. First, we used TRIP[25,26] to embed barcoded reporter genes randomly throughout the host DNA **(Fig 1A)**. We established a clonal TRIP cell line containing 36 PGK-driven integrated reporters (IRs) per cell and a multi-promoter TRIP pool with ~1k IRs distributed heterogeneously across cells **(Fig 1B)**. We designed three single-guide RNAs (sgRNAs) targeting sites near the 3’-end of the IR gene body, cloned them into Cas9–sgRNA plasmids, and used them in independent assays **(Fig 1C)**. After selecting Cas9–sgRNA-carrying cells, we amplified and sequenced IR target regions to characterize induced lesions. As a proof-of-concept, we performed Cas9 disruption assays using sgRNA1-3 in the TRIP cell line **(Fig 1C)**. By disruption, we refer to cleavage without the use of exogenous DNA. We also did disruption assays in TRIP pools to study mutations at a large number of loci. In addition, we performed editing involving the knock-in of a 21-nt single-stranded oligodeoxynucleotide (ssODN) to characterize template-dependent insertions. We used both sgRNA2/3 for disruption in TRIP pools (-ssODN), and only sgRNA2 for editing (+ssODN) since the proximity between the sgRNA3 target and the IR barcode prevented the design of proper homology arms. Finally, we analyzed all 36 IRs in the cell line, and also the 1359 IRs with at least 30 reads in all pool assays.

Variation in Cas9-induced mutation frequency across IR loci

We first analyzed Cas9-induced mutation frequencies at the targeted loci, and the effects of different factors on those frequencies. Mutation frequency was determined per IR as the fraction of reads carrying a mutation amongst all reads mapped to that specific IR. Overall, Cas9-targeted IR sequences showed high mutation frequencies genome-wide in both cell lines and pools. Cell line averages were ~50% for sgRNA1/2 and ~65% for sgRNA3, while pools reached ~30% for sgRNA1/2 and ~60% for sgRNA3 **(Figs 2A, 2B)**. We examined how Cas9-induced mutation frequencies varied with sgRNA, IR locus, ssODN, and promoter.

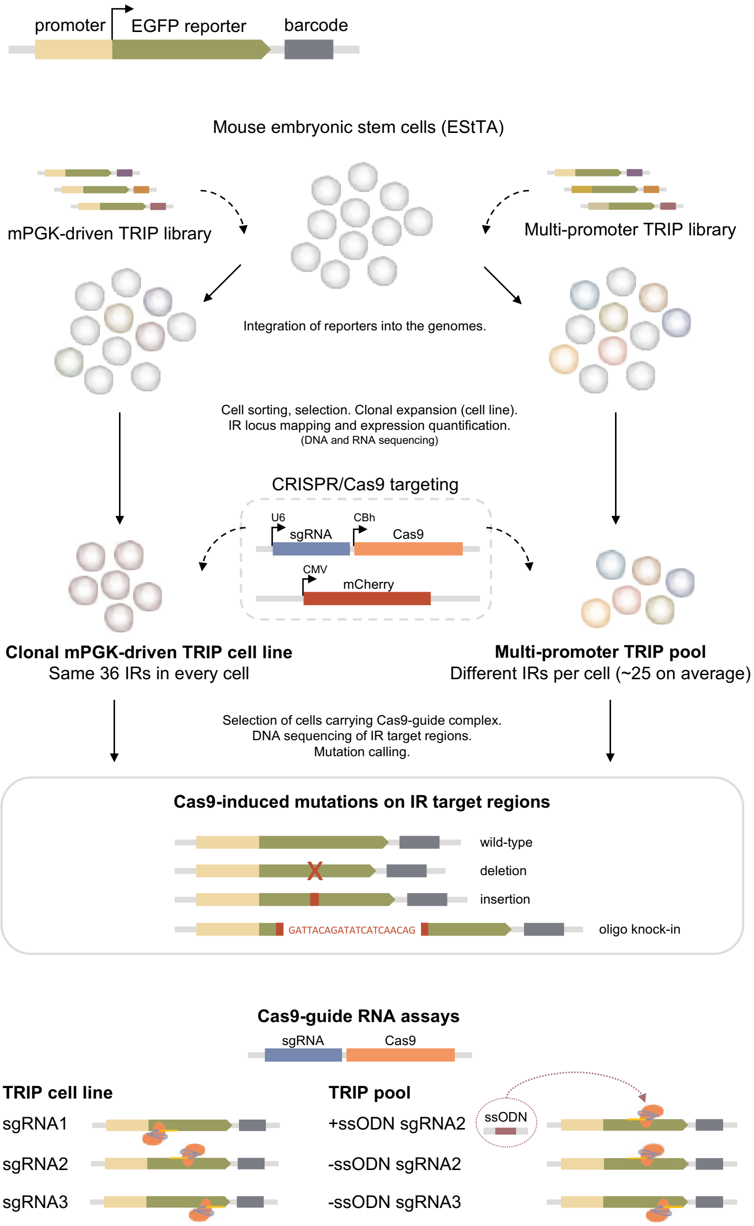


Figure 1. Overview of CRISPR-Cas9 assays in TRIP cell line and pools. (A) Barcoded TRIP reporter construct. (B) Clonal PGK-driven TRIP cell line with 36 IRs (left), and TRIP pool containing ~ 1k IRs with various promoters (right) - CMV, cMyc, Hoxb1, Nanog, Oct4, p53, PGK. Genomic location and expression of IRs were determined by DNA and RNA sequencing prior to Cas9 targeting of IR regions using different guides. Targeted DNA sequencing of IR regions was further used to characterize mutations arising from the repair of Cas9-induced DSBs. (C) Cas9-guide RNA combinations used in independent assays. TRIP cell line was targeted using Cas9 complexes with sgRNA1, sgRNA2, or sgRNA3 (left). In TRIP pool assays, Cas9 was complexed with sgRNA2 or sgRNA3 (right). Knock-in of a single-stranded oligodeoxynucleotide (ssODN) was performed with sgRNA2.

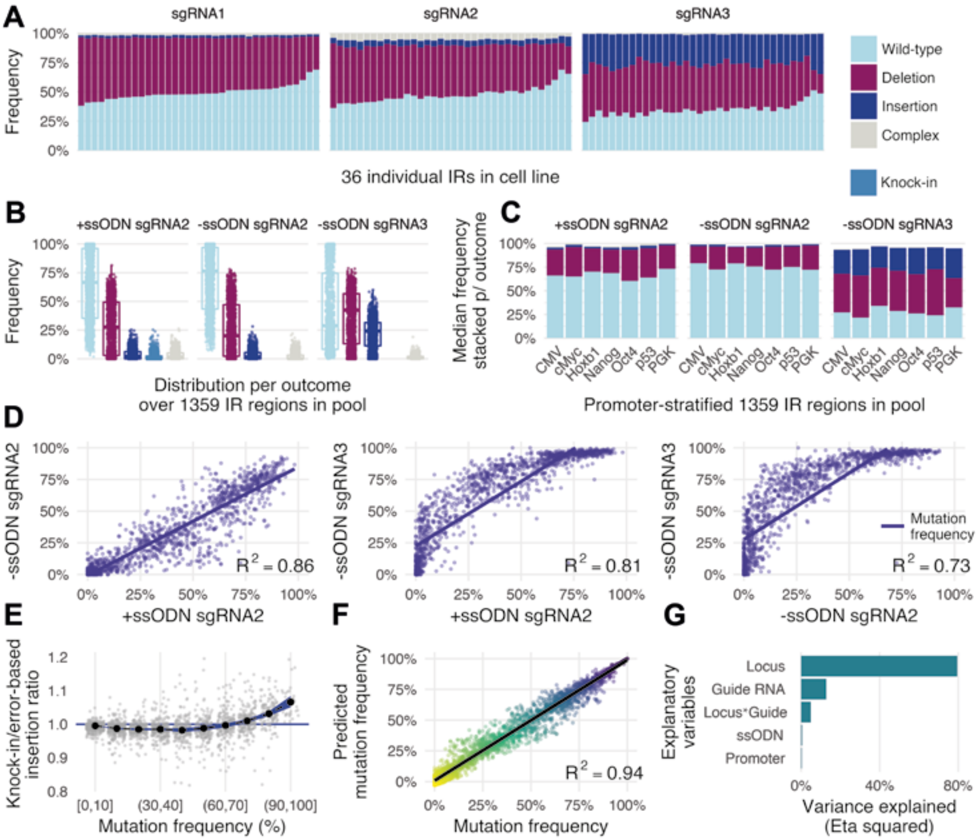


Figure 2. Contribution of IR locus, guide RNA, promoter, and ssODN to Cas9-induced mutation frequency. (A) Frequency per outcome in cell line Cas9 assays, showing effects of IR locus and guide RNA. Each bar represents one of 36 IRs in the cell line, and each colored band denotes the fraction of reads exhibiting a particular outcome among all reads mapped to such IR (vertical axis). Outcomes: wild-type in light blue, deletion in red, insertion in dark blue, and complex mix of mutations in beige. (B) Frequency per outcome in TRIP pool assays for guide RNA and ssODN inclusion combinations. Dots denote frequency (vertical axis) per outcome (color) for 1359 IRs with at least 30 reads in all assays. Boxes show the median, first and third quartiles of the frequency distributions; whiskers extend to 1.5 times the inter-quartile range from the top and bottom of the box. (C) Frequency per outcome in TRIP pools, stratified by promoter. Each bar denotes the subset of IRs associated with a given promoter; colored bands denote median frequency per outcome. (D) Correlation of IR mutation frequency across TRIP pool assays. Each dot indicates mutation frequency of a given IR in two different experiments (horizontal and vertical axes). Linear regression lines and corresponding R2 values denote correlations. (E) Ratio between knock-in and error-based insertions (vertical axis) with respect to binned IR mutation frequency (horizontal axis). Grey dots indicate ratios for individual IRs, black dots denote mean ratios within bins, blue ribbon shows 0.95 confidence interval around the mean. (F) Goodness-of-fit of linear regression model predicting mutation frequency based on IR locus, guide RNA, ssODN, promoter, and (locus, guide) interaction term. (G) Effect size or variance explained by variables in the regression model. Plotted are eta squared values for multi-way ANOVA tests based on type II sum of squares.

Guide RNA sequence-driven variation in mutation frequency

In line with previous reports, different guide RNAs led to systematic variation in IR mutation frequencies[1,13,27]. Guide sgRNA3 was most efficient in the cell line with average 1.30-fold and 1.26-fold increases in mutation frequency relative to sgRNA1/2 (effect sizes 15% ± 1% for sgRNA3 vs. sgRNA1 and 13% ± 1% for sgRNA3 vs. sgRNA2, both with $p \approx 2.91 \times 10^{-11}$, two-tailed Wilcoxon rank sum tests, **Fig 2A**). The difference was largely due to insertions, showing 12.5-fold and 5.26-fold increases with sgRNA3 compared to sgRNA1/2. Deletion frequency was more comparable across guides, varying by 1.23-fold and 1.10-fold between sgRNA3 and sgRNA1/2. Guide sgRNA3 was also the most efficient in TRIP pools, promoting an average 1.96-fold increase in mutation frequency (effect size 30% ± 1% sgRNA3-ssODN vs. sgRNA2-ssODN, $p < 2.20 \times 10^{-16}$, two-tailed Wilcoxon signed rank test; **Figs 2B, 2C**). Again, insertions increased by 6.76-fold, while deletions varied by 1.45-fold. Overall, we found that sgRNA1-3 resulted in different levels of mutation efficiency, mostly contributed by insertions.

Locus-associated variation in mutation frequency

We questioned whether certain IR loci would be more prone to mutations than others. We saw that IR mutation frequency correlated strongly across cell line assays using the three sgRNAs, with $R^2 > 0.85$ and F-test $p < 3.13 \times 10^{-16}$ **(Fig 2A)**, and across TRIP pool assays with $R^2 > 0.72$ and F-test $p < 2.16 \times 10^{-16}$ **(Fig 2D)**. Correlations were lower for insertions ($R^2 \leq 0.30$) than deletions ($R^2 \geq 0.72$), likely due to the scarcity of insertion events. The highly reproducible mutation frequencies revealed consistent locus-specific susceptibility to Cas9-induced mutations **(Figs 2A, 2D)**.

Knock-in and error-based insertion frequency

We examined the frequencies of error-based and knock-in insertions. Most insertions induced by Cas9–sgRNA2 in the TRIP pool resulted from errors in the endogenous repair of Cas9-induced DSBs, with average frequencies per IR of 3.78% and 3.24% in disruption and editing experiments, respectively **(Fig 2B)**. Knock-ins occurred only in editing assays with the integration of the designed 21-nt ssODN at the break site by HDR. Knock-in efficiency was low, as expected[15], with an average of 1.74% per IR **(Fig 2B, left plot)**. However, knock-ins were more frequent than error-based insertions at IRs with a total mutation frequency larger than 70% **(Fig 2E)**.

Effects on mutation frequency

To quantify the effect of the above factors on IR mutation frequency, we modeled mutation frequencies in the TRIP pool as a linear function of IR locus, guide RNA, ssODN inclusion, promoter, and an interaction term for the joint (non-additive) contribution of locus and guide **(Figs 2F, 2G)**. The linear regression model yielded a goodness-of-fit of $R^2 \approx 0.98$ **(Fig 2F)**. Using multi-way ANOVA tests, we determined the effect size of each factor in the model **(Fig 2G)**. IR locus explained ~79.5% of the variation in mutation frequency, while guide RNA sequence was responsible for ~12.7%, and locus-guide interaction accounted for ~4.9%. ssODN and promoter had negligible effect, with less than 1% together. These results confirmed that IR locus and guide RNA are major determinants of mutation frequency. Note that IR locus encapsulates a variety of factors that make a locus unique, including genomic context or the interaction between an IR and the host DNA.

Association between genomic context and mutation frequency

We sought to analyze the contribution of genomic context at IR loci to Cas9-induced mutation frequencies. Specifically, we examined the relationship between the mutation frequency for IRs in TRIP pools **(Fig 3A)** and transcriptional, genomic and epigenomic (TGE) features **(Figs 3B–D)**. Transcriptional features included IR expression in our TRIP cells, and gene expression and transcription factor binding in wild-type mES cells[28]. Genomic features comprised metrics such as GC content and gene proximity in mES cells[28]. Epigenomic features included chromatin density from Hi-C assays and chromatin immunoprecipitation (ChIP) data for numerous histone modifications, DNaseI hypersensitivity, and Lamin-B1 in mES cells[28].

We quantified TGE features within a region of 2 kb surrounding each IR locus and calculated their correlation with IR mutation frequency. Mutation frequency correlated weakly with TGE features (Pearson's $r = [-0.123, 0.151]$, **(Figs 3B–D)**, but insertions showed stronger association than deletions (effect size 0.021 ± 0.007, $p \approx 1.91 \times 10^{-9}$, two-tailed Wilcoxon signed rank test sgRNA2 ± ssODN and sgRNA3-ssODN insertions vs. sgRNA2 ± ssODN and sgRNA3-ssODN deletions). Correlations were also higher for sgRNA3 compared with sgRNA2, more so for insertions than deletions (effect size 0.022 ± 0.008, $p \approx 2.67 \times 10^{-7}$ for insertions, 0.016 ± 0.009, $p \approx 0.0001$ for deletions, two-tailed Wilcoxon signed rank test sgRNA3-ssODN vs. sgRNA2-ssODN). We investigated whether changes induced by TRIP IRs could have disrupted the correlation between IR mutation frequency and expression-based features. For this, we analyzed IR expression, quantified as IR barcode counts in cDNA normalized by IR barcode counts in genomic DNA obtained by high-throughput sequencing of a 168 bp region spanning the Cas9 target sites in TRIP mES cells. The correlation with mutation frequency was similarly weak for IR expression than for expression in wild-type mES cells, and we therefore excluded a potential impact of TRIP on these results **(Fig 3D, Supplementary Figure 3)**. Also reassuring was the fact that IR expression correlated strongly with wild-type TGE features (Pearson's $r = [-0.603, 0.648]$, **Fig 3D**). Additionally, we looked into translocations upon Cas9 targeting as an eventual source of variation in mutation frequency and found no evidence that they played a role in this regard **(Supplementary Tables 1 and 2)**.

Expression of IRs and endogenous genes showed the highest association with mutation frequency amongst six categories of TGE features **(Fig 3C)**. Moreover, features related to the RNA PolII complex involved in genome-wide transcription initiation produced

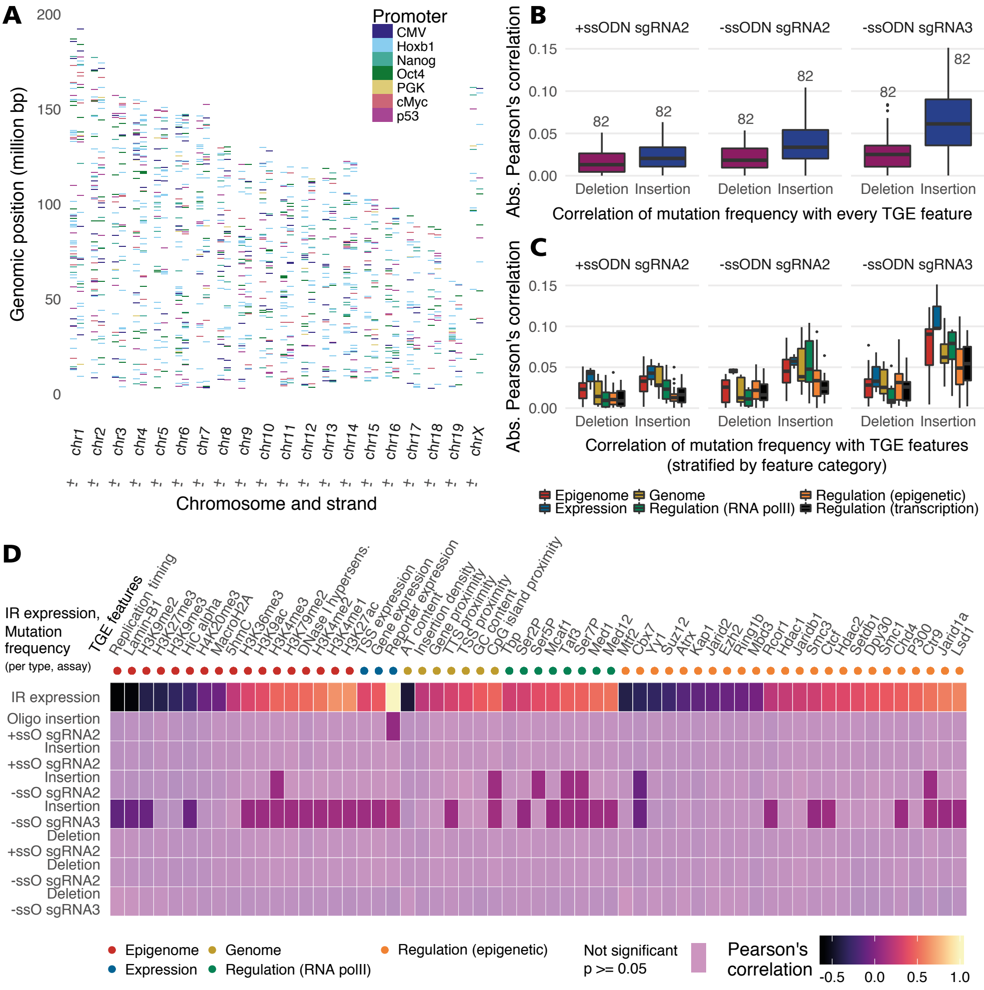


Figure 3. Correlation of TGE features with Cas9-induced IR mutation frequency in the TRIP pool. (A) Genomic location of the 1359 IRs with at least 30 mapped reads in all TRIP pool Cas9 assays. Each tick denotes the location of an IR on the chromosome, colored according to the associated promoter. (B) Correlation of TGE features with IR mutation frequency per guide RNA. Boxplots show the distribution of absolute Pearson's correlations between deletion (red) or insertion (blue) frequency and each of 82 distinct TGE features across IRs. Boxes show the median, first, and third quartiles of the frequency distributions; whiskers extend to 1.5 times the inter-quartile range from the top and bottom of the box. (C) Correlation of IR mutation frequency with TGE features stratified per category. Boxplots show the distribution of absolute Pearson's correlations between deletion or insertion frequency and each of 82 TGE features stratified into six categories (color-coded according to legend). (D) Correlation between IR expression or IR mutation frequency and TGE features. Heatmap shows the Pearson's correlation between IR expression or IR mutation frequency (deletion or insertion) in the different TRIP pool assays (rows), and individual TGE features from a subset of 62 (columns), including all except transcriptional regulators without known epigenetic activity. Cells are gradient-colored based on correlation values, and color intensity denotes significance of adjusted p-value. Colored circles at the top indicate TGE feature categories.

larger effects than transcription factors. Individually, phosphorylated RNA PolII (Ser2P, Ser5P, Ser7P) and influencers of transcription initiation such as Ctr9, Taf3, and CpG islands correlated positively with insertion frequency **(Fig 3D)**. We saw similar effects for H3K4 methylation and other histone modifications (H3K9ac, H3K27ac, H3K36me3, H3K79me2) associated with actively transcribed regions[29]. Correlating negatively with insertion frequency were known indicators of transcription inhibition, namely: lamina-associated domains (Lamin-B1), heterochromatin mark H3K9me2, and the Polycomb Repressive Complex 1 (PRC1) member Cbx7, which promotes trimethylation of H3 at Lys-9 (H3K9me3)[30].

Characterization of Cas9-induced mutation sizes and patterns

In addition to mutation frequencies, we analyzed mutation sizes and patterns. We found that Cas9 targeting yielded mostly small mutations. On average, 73.4% of all deletions in the cell line were smaller than 10-bp **(Fig 4A)**. The most common sizes were {1,2}-bp with sgRNA2 and {2,4,5}-bp with sgRNA3, each accounting for approximately 12 to 14% of all deletions. In contrast, sgRNA1 led to a large number of 3-bp deletions (41.8%, **Fig 4A**), the majority of which denoted the loss of triplet CGG, likely at positions 1–3 or 4–6 upstream of the PAM (40.2%, **Fig 4B**). Single-nucleotide insertions accounted for > 78% of all insertions with sgRNA1/2 and > 95% with sgRNA3 **(Fig 4A)**. We found similar size and pattern distributions in experiments using varying Cas9 concentrations **(Supplementary Figure 6).**

We observed that deleted regions neighbored the expected Cas9 target site between nucleotides 3|4 upstream of the PAM with sgRNA1/2 **(Fig 4C)**. This indicates that resection might preferentially occur on one rather than both DNA ends at the break site. With sgRNA3, deletions often neighbored nucleotides 4|5 instead. In addition, each sgRNA led to predictable 1-bp insertions, namely G (63.7%) for sgRNA1, G (73.9%) for sgRNA2, and T (97.9%) for sgRNA3 **(Fig 4D)**. Deletion borders and 1-bp insertions were therefore highly consistent and guide-specific.

Insight into Cas9 cleavage based on insertion patterns

We sought to understand the observed preference for specific nucleotide insertions under current Cas9 cleavage models **(Figs 4D, 5A)**. Cas9 is thought to primarily induce blunt-ended DSBs[3,31]. Besides direct re-ligation, blunt ends may be processed, leading to a deletion or, crucially, template-independent addition of nucleotides[9]. However, evidence from the seminal work on CRISPR–Cas9[3,31] and recent simulations on Cas9-domain conformation[32] suggests that Cas9 may also generate staggered DSBs. Specifically, Cas9 domains RuvC and HNH could cleave between nucleotides 3|4 on the target DNA (tDNA) and 4|5 on the non-target DNA (ntDNA) upstream of the PAM[32]. The resulting 5’ overhangs could trigger polymerase-based fill-in at position 4, producing sgRNA-specific insertions consistent with our data **(Figs 4D, 5A)**.

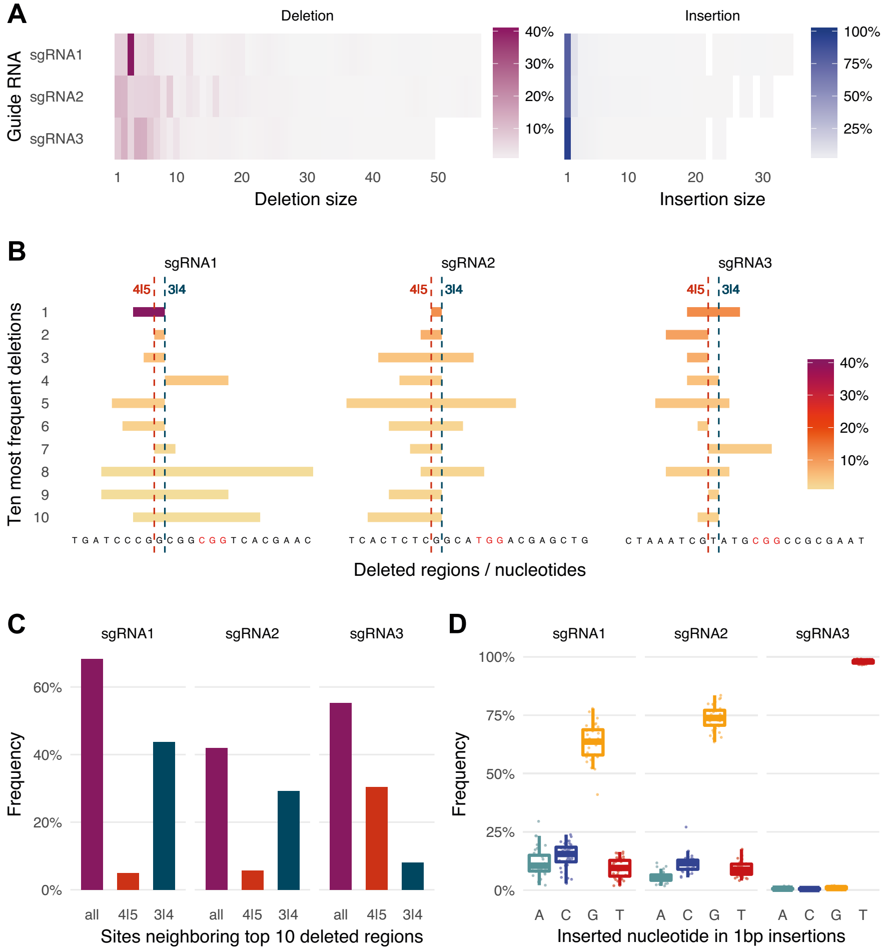


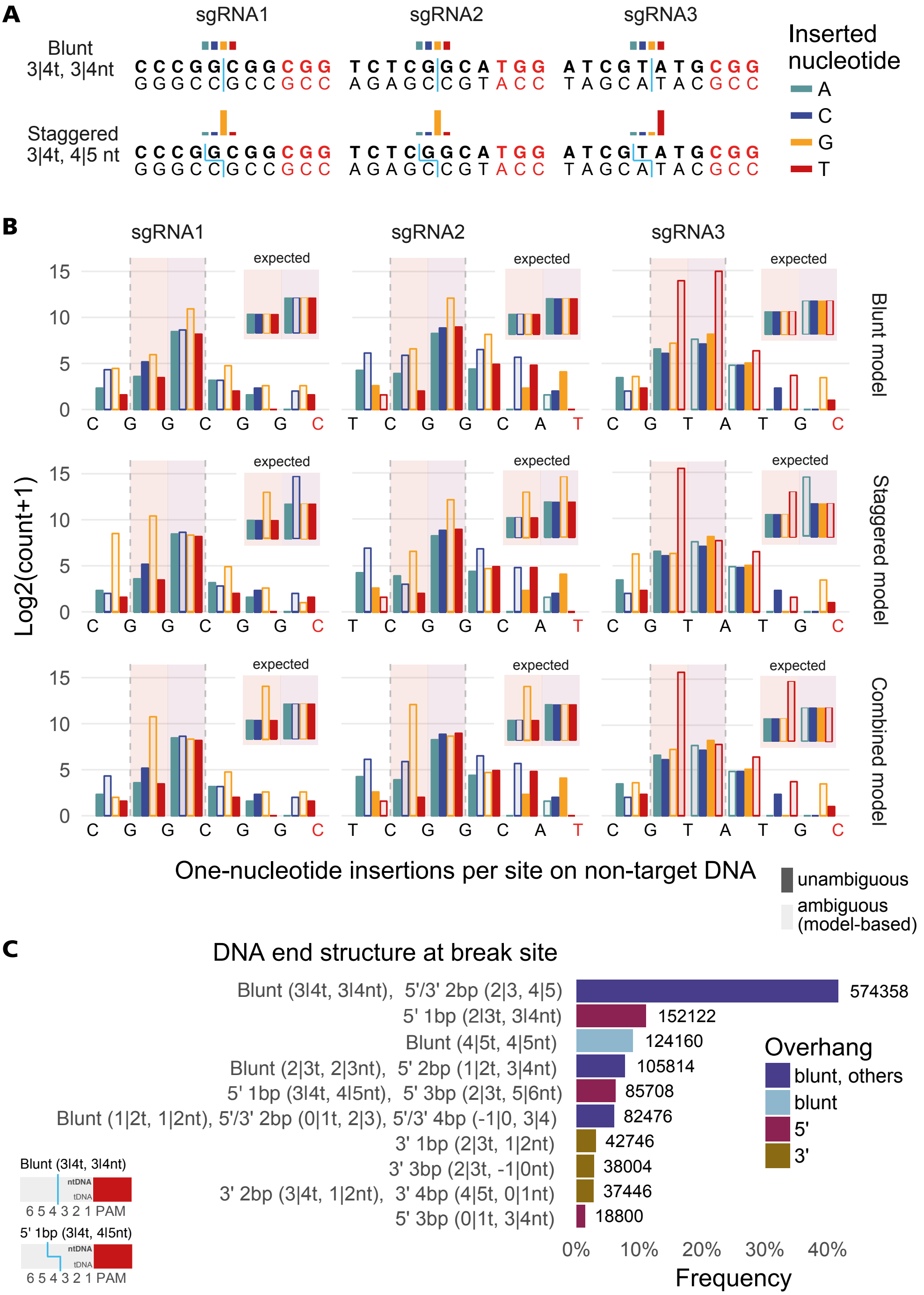
Figure 4. Mutation patterns induced by Cas9 in the 36-integration TRIP cell line. (A) Observed deletion and insertion sizes. Heatmaps show the overall frequency (color gradient) of deletions (red, left) and insertions (blue, right) per size (horizontal axis) for each guide RNA (vertical axis) in the TRIP cell line. (B) Deletion patterns and positions. Shown for each guide RNA are the ten most frequent deletion patterns with respect to the non-target DNA, from top to bottom in decreasing order of frequency. Each horizontal bar indicates the position of a deletion pattern and corresponding non-target DNA sequence lost (at the bottom), colored according to frequency. Expected 3|4 and alternative 4|5 break sites are indicated by two vertical dashed lines. (C) Frequency of sites neighboring the ten most frequently deleted regions for each guide RNA, as shown in Figure 3B. Three vertical bars indicate the proportion of all such deletions regardless of neighboring site (all, red), the subset of those deletions neighboring the expected break site (3|4, green), or the subset of those deletions neighboring the alternative break site (4|5, orange). For deletions with ambiguous positions, we weighted the frequencies by the ratio of positions meeting the criteria. We observed similar trends using all data. (D) Frequency of each nucleotide in 1-bp insertions. For each guide RNA, boxplots show the frequency (vertical axis) of insertions of each nucleotide (horizontal axis and color) across the 36 IRs (dots). Boxes show the median first and third quartiles of the frequency distributions; whiskers extend to 1.5 times the inter-quartile range from the top and bottom of the box.

Plausible Cas9 DNA cleavage and repair models

We assessed three different models of Cas9 cleavage and DNA repair **(Fig 5A)**. First, a blunt model producing blunt-ended DSBs primarily at 3|4[3,31], and leading to insertions of a random nucleotide[9]. Second, a staggered model cleaving at 3|4 (tDNA) and 4|5 (ntDNA)[32] and inducing the replication of nucleotide 4. Third, a combination of both blunt and staggered models. Each model determined an expected distribution of insertion counts across sites in the target sequence. In order to compare model distributions against observed insertions, we also addressed uncertainties caused by the addition of nucleotides identical to neighbors in the target sequence (e.g., ambiguity in insertions A**A** and **A**A). We achieved this by redistributing the counts of ambiguous insertions based on the counts of unambiguous insertions and rules derived from the models **(see Methods, Fig 5B)**. Figure 5B shows observed counts of unambiguous insertions (filled bars) and model-corrected counts of ambiguous insertions (unfilled bars) obtained for a range of possible break sites on the non-target DNA in our cell line Cas9 assays using sgRNA1-3. Each row depicts the redistribution of ambiguous counts for each assay according to a different model, and the vertical shaded areas highlight the main break sites of interest located 3|4 (right) and 4|5 (left) nucleotides upstream of the PAM. For each plot, we also include an insert labeled "expected", with an illustration of the expected model-based count distribution for the two sites 3|4 and 4|5.

Figure 5. One-nucleotide insertion patterns and DNA end structures at the break site. (A) Illustration of blunt and staggered DNA ends at the break site, and expected distribution of 1-bp insertions of the four nucleotides following DNA repair. Double-stranded sequences including PAM and 8-bp upstream, with bottom and top denoting target and non-target DNA. Blue straight and staggered lines through the sequences indicate blunt and staggered DNA ends. Colored bars on top sketch the expected distribution of 1-bp insertions upon DNA repair. Blunt model: blunt-ends primarily at 3|4 upstream of the PAM (straight line), resulting in template-independent insertion and thus similar frequencies of the four nucleotides (uniform distribution, similar-height colored bars). Staggered model: staggered ends mostly with termini at 3|4 (tDNA) and 4|5 (ntDNA) upstream of the PAM (staggered line), with template-dependent fill-in resulting in a skewed distribution with most insertions of the DNA base identical to nucleotide 4 (unequal-height colored bars). (B) Unambiguous insertion counts (filled bars) and ambiguous insertion counts (empty bars) redistributed according to blunt, staggered, and combined models. Shown are insertion counts (vertical axis) of each nucleotide (color) per site on the ntDNA (horizontal axis). Vertical shaded areas indicate the 3|4 and 4|5 sites upstream of the PAM. Unambiguous counts are directly determined from the data (filled bars), whereas ambiguous counts are redistributed over windows of ambiguous sites (empty bars) based on: (i) relative proportions of unambiguous counts, and (ii) likelihood of each nucleotide insertion according to the cleavage model. (C) Re-analysis of DNA ends generated by Cas9 targeting of a region on chromosome 6 in mouse pre-B cells deficient in DNA Ligase IV and arrested in the G1 phase. Bar length denotes relative frequency, shown for the ten most frequent DNA end structures accounting for ~ 91% of all unique patterns in the data. Absolute frequencies are displayed. Multiple DNA end structures associated with the same sequence are grouped with a single bar and label. Bars are colored by type of structure. The bottom left figure shows an illustration of two DNA structures: blunt (3|4t, 3|4nt) and 5´ 1-bp overhang (3|4t, 4|5nt).

▸


A
sgRNA1
sgRNA2
sgRNA3
Blunt
3|4t, 3|4nt
C C C G G | C G G C G G
G G G C C | G C C G C C
T C T C G | G C A T G G
A G A G C | C G T A C C
A T C G T | A T G C G G
T A G C A | T A C G C C
Staggered
3|4t, 4|5 nt
Inserted nucleotide
A
C
G
T
B
sgRNA1
sgRNA2
sgRNA3
expected
Log2(count+1)
0
5
10
15
C G G C G G C
T C G G C A T
C G T A T G C
Blunt model
Staggered model
Combined model
One-nucleotide insertions per site on non-target DNA
unambiguous
ambiguous (model-based)
C
DNA end structure at break site
Blunt (3|4t, 3|4nt), 5'/3' 2bp (2|3, 4|5) 574358
5' 1bp (2|3t, 3|4nt) 152122
Blunt (4|5t, 4|5nt) 124160
Blunt (2|3t, 2|3nt), 5' 2bp (1|2t, 3|4nt) 105814
5' 1bp (3|4t, 4|5nt), 5' 3bp (2|3t, 5|6nt) 85708
Blunt (1|2t, 1|2nt), 5'/3' 2bp (0|1t, 2|3), 5'/3' 4bp (-1|0, 3|4) 82476
3' 1bp (2|3t, 1|2nt) 42746
3' 3bp (2|3t, -1|0nt) 38004
3' 2bp (3|4t, 1|2nt), 3' 4bp (4|5t, 0|1nt) 37446
5' 3bp (0|1t, 3|4nt) 18800
0%
10%
20%
30%
40%
Frequency
Overhang
blunt, others
blunt
5'
3'
Blunt (3|4t, 3|4nt)
ntDNA
tDNA
6 5 4 3 2 1 PAM
5' 1bp (3|4t, 4|5nt)
ntDNA
tDNA
6 5 4 3 2 1 PAM

Blunt DNA cleavage-repair model

Based on the blunt model, we expected to see template-independent insertions and therefore similar insertion frequencies for every nucleotide at the break site (equal-height bars for sites 3|4 and 4|5 in “expected” inserts, top row plots of **Fig 5B**). When we redistributed the observed ambiguous counts in a nucleotide-unbiased manner, following the frequency of unambiguous insertions at each site, we did not obtain the expected nucleotide-unbiased insertion profiles. Specifically, redistributed counts showed significantly more insertions of G (sgRNA1), G (sgRNA2), and T (sgRNA3) at sites 3|4 and 4|5 (top row in **Fig 5B**, note the log-transformed scale). We further note that the nucleotide imbalance would be unavoidable for the most targeted sites, 3|4 and 4|5, given the substantial frequency gap observed between the most inserted and the remaining nucleotides.

Staggered DNA cleavage-repair model

Alternatively, the staggered model determined template-dependent insertions following a 1-nt 5’ overhang on the opposite strand. According to this preference, we expected to see significantly more insertions of the nucleotide identical to the DNA base immediately downstream of the break site (“expected” inserts in middle row plots of **Fig 5B**). For the staggered model, we redistributed ambiguous insertions in our observed counts based on both the frequency of unambiguous insertions and the 1-nt 5’ overhang template. The results obtained for this model showed the expected template-based insertion pattern at break site 4|5 for all sgRNAs (“expected” inserts *vs*. main plots in middle row of **Fig 5B**). However, the results did not follow the expected pattern at break site 3|4 for sgRNA1 and sgRNA3. Specifically, results showed similar frequencies of the four nucleotides, typical of template-independent insertions in a blunt model, while nucleotides C (sgRNA1) and A (sgRNA3) should have been significantly more frequent than others based on the staggered model (main plots *vs*. “expected” inserts in middle row of **Fig 5B**).

Combined blunt and staggered DNA cleavage-repair model

Finally, the combined blunt and staggered model provided the best fit: (i) most insertions occurred at sites 3|4 and 4|5, the expected primary targets of blunt and staggered cleavage on the non-target DNA (bottom row of **Fig 5B**); (ii) the four nucleotides were similarly likely at site 3|4, consistent with template-independent insertions at blunt DNA ends; and (iii) the most inserted nucleotide at site 4|5 matched the downstream neighbor, as expected upon fill-in of 1-nt 5’ overhangs. A combination of the blunt model with an alternative staggered model inducing 1-nt 3’ overhangs at break site 3|4 (ntDNA) could possibly fit as well, although we found no reference to such a model in the literature. We note that our data could include re-cleavage events leading to a biased selection for mutations. Specifically, accurately ligated DNA ends can be re-cleaved, while mutations become nearly fixed in the population since Cas9 is less likely to recognize and cleave at a mutated site. Over time, this could skew the ratio between mutated and wild-type sequences and influence blunt vs. staggered patterns. For this reason, we limited the scope of our analysis to show that both patterns occur, without quantifying how frequent each one is.

Staggered DNA ends in Cas9-targeted mouse pre-B cells

We investigated whether 1-nt 5′ overhangs would be generated by Cas9. For this purpose, we collected independent data generated by hairpin capture and sequencing of DNA end structures at Cas9 DSB sites (HCoDES)[33]. Specifically, we re-analyzed DNA ends induced by Cas9 targeting a region on chromosome 6 in mouse pre-B cells arrested in G1-phase and LigaseIV-deficient[33]. Many DNA end structures could not be uniquely mapped **(Fig 5C)**. We saw prevalent blunt ends at 3|4, although these could not be discerned from certain 2-nt 5′/3′ overhangs. Specifically, blunt ends at 3|4 accounted for ~20% when aligning sequences to the reference as-is, as performed in the original study[33]. We also determined a proportion of ~41.6% upon masking the noise caused by incomplete bisulfite conversions. Shorter 1-nt overhangs were favored. Additionally, we saw a preference for 5′ overhangs, particularly 1-nt 5′ at 2|3 (tDNA) 3|4 (ntDNA) accounting for ~11.0%, and at 3|4 (tDNA) 4|5 (ntDNA) confounded with a 3-nt 5′ overhang (~6.2%).

DISCUSSION

As the CRISPR–Cas9 system is widely used for gene editing, understanding Cas9 activity across the genome is crucial to identify cleavage and mutation patterns enabling new applications or improvements. Using a combination of the CRISPR–Cas9 and TRIP technologies with high-throughput DNA sequencing, we characterized mutations at ~1k loci throughout the genomes of mES cells. We showed high reproducibility in a single-promoter TRIP cell line with 36 IRs and pools of cells with thousands of heterogeneous multi-promoter TRIP IRs.

In line with previous studies[13], we saw sgRNA-dependent variation in Cas9-induced mutation frequency. Specifically, sgRNA3 led to larger mutation frequencies than sgRNA1/2. We reasoned that the high GC-content of sgRNA1/2 (75%, 70%) relative to sgRNA3 (45%) could explain the lower efficiency of sgRNA1/2, based on reports associating low and high GC-content with reduced Cas9 activity[12]. Overall, sgRNA1/2 produced comparable mutation frequencies, possibly given the high similarity, including a 19-bp reverse complement overlap. Genomic location was the largest contributor to variation in mutation frequency. However, mutation frequency correlated weakly with TGE features of wild-type cells and reporter expression in TRIP cells. Importantly, reporter expression and mutation frequency were quantified in TRIP mES cells bearing identical integrations and epigenetic landscape. This result suggests that the effects of genomic context on mutation frequency may be influenced by other factors than those surveyed here. Although studies using doxycycline-controlled chromatin states on engineered targets previously showed effects on mutation efficiency[19,21], an approach targeting endogenous sites in human cells revealed only modest effects with effect sizes that were largely dependent on guide RNA and possibly other factors[22]. We saw similar results in our study, with generally higher correlations between TGE features and mutation frequency using sgRNA3. We note that, due to the exogenous manipulation of chromatin states, those studies could produce larger contrasts between permissive and non-permissive states than those seen in the context of our work. Our findings were also consistent with two other studies assessing the association between expression or epigenome and mutation frequency at endogenous target sites[17,24]. One study reported weak correlations overall across developmental stages in zebrafish, which were slightly higher for expression than epigenetics[24]. The other study reported that native chromatin accessibility and DNA methylation were predictive of Cas9 binding, but Cas9 binding was uncorrelated with mutation frequency in mES cells[17]. Together, these findings suggest that genomic context likely influences Cas9 binding and cleavage, but that the association with mutation frequency can be further modulated by additional factors such as guide RNA sequence. It is possible that the correlation could be disrupted by stochasticity in the outcome of DNA repair, especially as a result of potential re-cleavage events. We also reason that dynamic reorganization of the regulatory landscape during the S-phase of the cell cycle could influence estimates of Cas9-induced mutation frequency but not IR expression, which is mainly produced in G1-phase. In particular, mES cells are known to exhibit unusually short G1 and long S-phases[34]. In S-phase, the DNA is unpacked to allow for replication, enabling unperturbed Cas9 binding to otherwise inaccessible loci. This could mean that most Cas9 cleavage in our assays occurred in S-phase under widespread DNA accessibility, masking differences between permissive and non-permissive domains, and hence lowering the association of mutation frequency with TGE features. Nevertheless, insertions were more dependent on the regulatory landscape than deletions, particularly concerning influencers of transcriptional activity genome-wide such as PolII complex, histone marks or Lamin-B1. We noted that deletions varied in size and pattern, while insertions comprised mostly sgRNA-specific single-nucleotides. Specifically, deletions could be generated by a variety of DNA processing events on blunt or staggered DNA ends, either aimed at ligation by NHEJ throughout the cell cycle[35] or generation of 3' overhangs in preparation for HDR in S and G2 phases[36]. In contrast, insertions likely arose by gap-filling of specific overhangs by a DNA polymerase. We hypothesize that the deterministic nature of insertions could preserve the association with TGE features better than the range of processes involved in deletions, possibly yielding varied dependencies on the regulatory landscape.

We observed small Cas9-induced lesions with all sgRNAs. The most common deletion was loss of trinucleotide CGG, accounting for ~40% of all deletions with sgRNA1. We attributed this event to microhomology-mediated end-joining (MMEJ), which could rely on the repetition of CGG in sgRNA1 to align the DNA ends, resulting in the loss of triplet CGG[37]. Regions deleted with sgRNA3 neighbored nucleotides 4|5 (ntDNA) rather than the expected target site 3|4, seen with sgRNA1/2, suggesting alternative Cas9 targeting between 4|5 (ntDNA). In addition, single-nucleotide insertions showed a high prevalence of a specific nucleotide per sgRNA, consistent with independent studies targeting the HPRT, AAVS1, and TREX1 genes in HCT116 cells[38], the GFP, NDC1, and LBR genes in K562 cells[39], as well as numerous other regions in HCT116, HEK293, and K562 human cells[38–40]. We determined that the frequently inserted DNA base matched nucleotide 4 upstream of the PAM, which hinted at eventual template-dependent repair of staggered DNA ends with termini at 3|4 and 4|5 on opposite strands. Different Cas9 cleavage and DNA repair models were analyzed, seeking to explain observed insertion patterns[3,31,32]. We propose a Cas9 cleavage model inducing primarily blunt and occasionally staggered DNA ends **(Fig 6)**. Most blunt ends are ligated, possibly upon resection, resulting in either wild-type or deletion. On occasion, blunt ends may lead to an insertion via template-independent addition of a random nucleotide (e.g., by Pol μ)[9]. We showed that insertions mainly derived from DNA ends with 1-nt 5′ overhangs, specifically with termini at positions 3|4 (tDNA) and 4|5 (ntDNA)[32]. We noted that 1-nt 3′ overhangs could also explain the observed insertions. However, re-analysis of independent data on DNA end structures generated by Cas9 targeting in mouse DNA Ligase IV-deficient pre-B cells[33] revealed a prevalence of blunt ends and preference for 1-nt 5′ over 3′ overhangs, in accordance with our hypothesis. We cannot exclude the possibility that the overhangs would arise by minimal resection of blunt ends rather than directly by Cas9 cleavage. Regardless of the process, we presented evidence that staggered DNA ends are generated and likely responsible for most insertions. The insertion patterns we observed in mES cells, mostly in S-phase, were corroborated by independent Cas9 targeting in human cells HEK293, HCT116, and K562[38–40], as well as in mouse pre-B cells arrested in G1-phase[33], using different plasmid systems. The consistency of these results indicates that our findings might generalize beyond the scope of our assays.

Here, for the first time, Cas9-induced mutation patterns were characterized at thousands of target sequences embedded throughout the genomes of mES cells. We laid out the likely combination of mechanisms of Cas9 cleavage and DNA repair underlying predictable 1-nt insertion patterns reported in the literature[38–40]. In particular, we revealed that Cas9 occasionally generates breaks with staggered DNA ends. These could be leveraged to increase knock-in efficiency and control the orientation of inserts into host DNA by homology-dependent or independent insertion[41]. Furthermore, our data indicated that guide RNA sequence determines the frequency of staggered ends, ultimately influencing insertion and overall mutation efficiency. This result suggests that guide RNAs could be designed to maximize insertion frequency. Extensive testing of a wider range of guides will be needed to build models that predict insertion efficiency

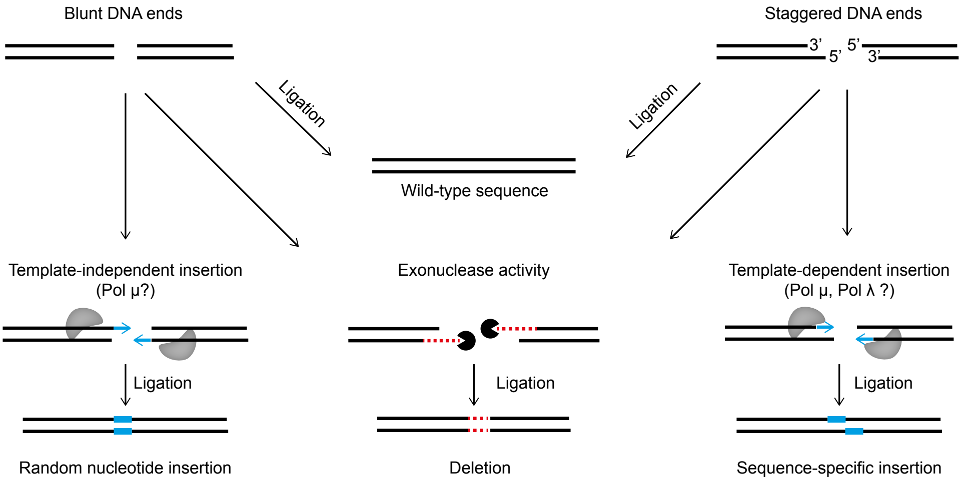


Figure 6. Illustration of DNA repair outcomes after Cas9-induced double-strand break. Both blunt and staggered ends can be directly ligated back into wild-type sequence or generate a deletion through resection by nuclease activity prior to ligation. Blunt ends can also result in an insertion by template-independent addition of a random nucleotide, possibly established by Pol μ. Staggered ends lead primarily to template-dependent insertions, possibly established by polymerases such as Pol μ or Pol λ.

based on relevant guide RNA features. Our findings could have important implications for the optimization of Cas9-mediated knock-in, which remains a major challenge in genome editing.

In addition, we introduced the TRIP technology as a tool to multiplex RNA-guided Cas9 targeting to regions in reporter genes integrated genome-wide. This CRISPR-on-TRIP approach allowed us to seamlessly isolate target sequence from genomic location and perform sequence-independent analysis of variation in Cas9 mutation efficiency and patterns in different genomic contexts. We found that genomic location is a key determinant of mutation frequency, which, however, correlated modestly with genomic and epigenomic context. We reasoned that guide RNA and stochasticity in the outcome of DNA repair, together with potential re-cleavage events and the reprogramming of the epigenetic landscape during the cell cycle, could contribute to this result. Further investigation will be needed to clarify the impact of the genomic context. Finally, we demonstrated that CRISPR-on-TRIP is a promising tool to profile Cas9 activity at a large number of target sequences scattered throughout the genome and can be combined with other assays to study the influence of a variety of processes on Cas9 activity and induced mutation patterns **(Fig 7)**.

Figure 7. Potential applications of CRISPR-on-TRIP. RNA-guided Cas9 targeting of regions within integrated TRIP reporters (CRISPR-on-TRIP) can be combined with other assays to investigate the effects of various processes on Cas9-induced mutation frequency and patterns.

MATERIALS AND METHODS

Construction of TRIP plasmid libraries

The monoclonal TRIP cell line used in this work was established in our previous TRIP study[25]. The piggyBac-based template vector pPTK-Gal4-mPGK-Puro-IRES-eGFP-sNRP-pA carrying the reporter unit in the TRIP cell line library comprised the following elements:

piggyBac 5’-TR, 14 Gal4 binding sites, mPGK promoter, puromycin resistance (PuroR) coding sequence, encephalomyocarditis virus internal ribosome entry site (IRES), EGFP coding sequence, PstI site (used to clone barcodes) + DpnII site (used to map IRs), human soluble neuropilin-1 (sNRP-1) polyA signal, and piggyBac 3’-TR (67-bp). Barcoded inserts were generated through amplification of 5 ng template vector pPTK-Gal4-mPGK-Puro-IRES-eGFP-sNRP-pA (GenBank KC710227), using primers PB-barcode-long-7 (5’-GTGACACCTGCAGGATCA$(N)_{16}$CTCGAGTTGTGGCCGGCCCTTGTGACTG-3’, where $(N)_{16}$ denotes a random 16-nt long reporter barcode) and PB-barcode-short-7 (5’-GACATAACGCGTATACTAGATTAACCCT-3’). After PCR purification, the PCR product was digested with restriction enzymes PstI and MluI (underlined). In parallel, the pPTK-Gal4-mPGK-Puro-IRES-eGFP-sNRP-pA vector was digested with the same restriction enzymes and then dephosphorylated. The digested PCR product was next ligated with the prepared vector using 10U of T4 DNA ligase (Promega). The resulting ligation product was transformed into electrocompetent *E.coli* cells and the plasmid DNA (TRIP plasmid library) was isolated using Genopure plasmid maxi kit (Roche).

For the multi-promoter TRIP pool established in this study we first generated seven libraries, each containing reporter constructs with one of seven different promoters: CMV, cMyc, Hoxb1, Nanog, Oct4, p53 and PGK. These were included to assess effects of promoters with different characteristics, such as strength[42], exogeneity, housekeeping status, retinoic acid-inducibility[43–46], and TATA-less status[45,47–50]. The piggyBac-based plasmid vectors carrying the reporter unit comprised the following elements: piggyBac 5’-TR (314 bp long), promoter of interest, EGFP coding sequence, DpnII site (used to map IRs), 5-bp promoter index unique to the promoter, KpnI site (used to clone barcodes), sNRP-1 polyA signal and piggyBac 3’-TR (242-bp). Barcoded inserts were generated by amplification of 5 ng PB template vector using primers Kpn-RandomBC-1 (5’-AAAAGGTACC$(N)_{18}$GAGTTGTGGCCGGCCCTTGTGACTG-3’, with $(N)_{18}$ denoting a random 18-nt long reporter barcode) and BssH2-A (5’-AAAAGCGCGCATACTAGATTAACCCTAGAAAGATAATCATATTG-3’). After PCR purification, the barcoded inserts were digested with restriction enzymes KpnI and BssHII (underlined). In parallel, the plasmid vectors were digested with restriction enzymes KpnI and MluI (the latter generates sticky ends compatible with those made by BssHII) and subsequently dephosphorylated. Ligation of the digested barcode insert into the digested plasmid vectors, and electrotransformation of bacterial cells were performed as described above. The seven promoter-specific plasmid libraries were mixed together in the following molar ratios: 2:2:4:2:2:2:1 to obtain the multi-promoter TRIP library.

Cell culture and TRIP library transfection

Mouse embryonic stem (mES) cells EBRTcH3 expressing the tetracycline-controlled transactivator (tTA) from the endogenous ROSA26 promoter (EStTA)[51] were cultured in 60% BRL cell-conditioned medium in the presence of 10% fetal calf serum (FCS), leukemia inhibitory factor, MEK inhibitor PD0325901, and GSK-3 inhibitor CHIR99021[52]. The EBRTcH3 ES cells were provided by the te Riele lab (Netherlands Cancer Institute), which had received them from Dr. Masui (International Research Center of Japan). The EBRTcH3 ES cells were originally derived from E14tg2a ES cells[53] by Masui and colleagues[51]. Culture dishes were coated with 0.15% gelatin and incubated at 37°C for one hour before plating.

For the cell line, six million mES cells were plated and incubated for 4 h at 37°C. The cells were then transfected with 22.5 μg of the mPGK TRIP plasmid library together with 2.5 μg of mouse codon-optimized version of PB transposase (mPB) plasmid[54] using Lipofectamine 2000 (Invitrogen) and incubated at 37°C. After 48 h of incubation, the transfected cells were sorted through fluorescence-activated cell sorting (FACS), selecting single cells with “medium” levels of EGFP expression, which were used to establish stable TRIP cell lines. The cell line EStTA-PB-B-18 used in this study carries 36 IRs.

For the TRIP pool, nine million mES cells were plated and incubated for four hours at 37°C. The cells were then transfected with 15 μg of the multi-promoter TRIP plasmid library mixture and 5 μg of PB transposase plasmid mPB-L3-ERT2.TatRRR-mCherry plasmid using Lipofectamine 2000 (Invitrogen) and incubated at 37°C. After 24 h, the cells were FACS-sorted to select for a desired mCherry signal from the PB transposase plasmid[26]. The sorted cells were resuspended in mES medium with 1 μM of tamoxifen to activate the PB transposase. After 24 h, the cells were washed and resuspended in mES medium without tamoxifen, in which they were cultured for 5 days, refreshing the medium every other day. Biological replicate pools were established by subculturing several aliquots of cells, which were grown for another week. Technical replicates were established by splitting each pool into two and growing each half separately for another week. The multi-promoter TRIP pool used in this study was grown from approximately 500 cells containing an average of ~25 IRs per cell.

Determination of genomic location and expression of IRs

Barcoded IR regions were extracted by inverse PCR followed by high-throughput DNA sequencing. The resulting reads were subject to quality control, including the filtering of aberrant barcodes arising from errors induced by PCR or sequencing. After pre-processing, the regions were aligned against the mouse genome assembly mm9 in order to map IR loci. Expression of IRs was determined by sequencing of reverse transcribed RNA (cDNA). Barcode abundance was also estimated by sequencing genomic DNA (gDNA), and used to normalize IR expression. These procedures were performed according to the TRIP protocol[26].

Association of TRIP integrations with genome-wide features

In addition to IR expression, we assessed endogenous gene expression[25] and a range of regulatory elements measured genome-wide in mES cells. These data were previously collected from public sources and preprocessed aiming to maximize comparability[28]. Most features were obtained from ChIP-seq data on epigenetic modifications such as DNase I hypersensitivity, histone marks, and binding of transcriptional and epigenetic regulators. For ChIP-based features, a score was defined as the log2-transformed ratio between signal and control. Genomic features, including GC content and gene proximity, among others, were derived from the mouse reference genome mm9 and Ensembl genes (release 66). Specifically, proximity measures were defined as the negative log2-transformed distance (+1) to the nearest concerned genomic feature (e.g., gene, transcription start site). Chromatin compaction was estimated from Hi-C data as the rate of decay in contact probability α between two loci with increasing genomic distance, locally approximated in 400 kb windows by a power-law function with scaling exponent α. Finally, the association between every IR and feature pair was computed as the mean normalized score of the genomic or regulatory feature over a region of 2 kb surrounding the integration site.

sgRNA and ssODN design

All oligonucleotides used in this study were purchased from Integrated DNA Technologies. We designed three sgRNAs targeting different regions within the EGFP reporter gene, using the CRISPR design tool http://crispr.mit.edu/ **(Supplementary Table 3)**. We considered three primary criteria. First, purposed sgRNA-targeting of a sequence within the EGFP gene body in close proximity to the barcode, enabling reliable amplification of both barcode and target site. Second, reliability according to the CRISPR design tool, with no or minimal reported off-target sites. Third, EGFP recognition in both sense and anti-sense orientation, to exclude orientation-dependent effects. In addition, we designed a 141-bp single-stranded oligodeoxyribonucleotide (ssODN) template for HDR-based knock-in in with sgRNA2. Our ssODN comprised a 21-bp sequence for knock-in with ~60-bp homology arms at each side. Following the recommendations in https://www.addgene.org/crispr/zhang/faq/ **(Supplementary Table 3)**, we avoided overlap between the ssODN homology arms and the barcode located downstream of the EGFP reporter gene. We also designed the junction of the homology arms < 10 bp from the Cas9 target site. Lastly, we designed the ssODN template with 58 and 62 nucleotide-long homology arms, in agreement with the recommended 50–80 range.

Cloning and transfection of sgRNA-guided CRISPR–Cas9

We used human codon-optimized SpCas9 and chimeric guide RNA expression plasmid pX330-U6-Chimeric_BB-CBh-hSpCas9 to complex Cas9 with different sgRNAs (1–3). In addition, we used mCherry as a fluorescent marker for visualization and sorting of Cas9–sgRNA transfected mES cells. All Cas9 targeting assays experiments were performed in triplicate. For cell line experiments, we co-transfected the Cas9–sgRNAs and mPB-L3-ERT2.TatRRR-mCherry plasmids. Ten million mES cells were first seeded on a 10-cm dish. Four hours later, the cells were transfected with 13.5 μg of Cas9–sgRNA and 1.5 μg of mPB-L3-ERT2.TatRRR-mCherry plasmids using 45 μl of Lipofectamine 2000 (Invitrogen).

For TRIP pool experiments, we cloned mCherry-expressing Cas9–sgRNA plasmids. We first digested the Cas9–sgRNA plasmids with 10 units NotI (Roche) and 10 units SbfI (NEB). We further PCR-amplified CMV-driven tatRRR-mCherry from mPB-L3-ERT2.TatRRR-mCherry using primers Fragment.FOR (10 µM) and Fragment.REV (10 µM) **(Supplementary Table 3)**. The digested Cas9–sgRNA plasmid and PCR-amplified tatRRR-mCherry fragment were then cloned using Gibson Assembly Master Mix (NEB). To avoid reduced transfection efficiency of mCherry-expressing Cas9–sgRNA plasmid and ssODN with Lipofectamine 2000, we followed the protocol from Nucleofector™ Kit for Mouse Embryonic Stem Cells (Lonza) instead. Five million mES cells per condition were trypsinized, spun down and resuspended in 90 µl of Mouse ES Cell Nucleofector™ solution. They were thereafter transfected with a total of 6.25 µg of DNA (3.75 µg mCherry-expressing Cas9–sgRNA and 2.5 µg ssODN) using the program A-024 of the Nucleofector™ Kit for Mouse Embryonic Stem Cells (Lonza). Transfected cells were resuspended in 500 µl of pre-warmed culture medium and plated in gelatin-coated 10 cm dishes. Cells were sorted 24 h after transfection. For disruption assays, we transfected five million mES cells with 8 µg of mCherry-expressing Cas9–sgRNA. For editing, five million cells were transfected with 4 µg of mCherry-expressing Cas9–sgRNA together with 1 µg of designed ssODN containing the 21-bp sequence for knock-in with ~ 60-bp homology arms. We decided to use sgRNA2 in editing assays, given that the short distance between the region targeted by sgRNA and the IR barcode prevented the design of proper ssODN homology arms **(Supplementary Figure 8)**. Additionally, sgRNA2 was preferred over sgRNA1 due to its larger mutation efficiency. In all experiments, transfected mCherry-positive mES cells were sorted by flow cytometry (MoFlo), collected in conditioned media containing 20% FCS, and subsequently spun down plated in conditioned media containing 10% FCS. Finally, Cas9–sgRNA-transfected mES cells were expanded for five days before isolating genomic DNA.

DNA isolation and preparation of samples for sequencing

After incubation of Cas9–sgRNA-transfected mES cells, DNA was isolated using DNeasy Blood & Tissue Kit (Qiagen) and prepared for sequencing. We PCR amplified cell line DNA IR regions of 414 bp surrounding the sgRNA target sites using 10 µM of PB-cDNA-forward-1-BC primers, containing different index sequences for multiple reactions, and 10 µM of PB-cDNA-Reverse-5 primer **(Supplementary Table 4)**. The PCR product was sent for Illumina MiSeq. In TRIP pool experiments, we tagged amplicons using 16-nucleotide Unique IDentifiers (UIDs) to be able to detect multiple readings of the same DNA molecules **(Supplementary Table 5)**[53]. We performed all Cas9 targeting assays in triplicate. Barcoded IR regions were PCR amplified using Phusion® High-Fidelity DNA Polymerase (NEB) following the protocol in Supplementary Table 6. We used Exonuclease-I (20 units) (Enzymatics) for cleavage of single-stranded DNA. Purified sequences were sent for Illumina HiSeq. Replicates were assessed for consistency of IR mutation frequencies and merged for subsequent analyses.

Validation of promoter–barcode index association

For identification purposes, each promoter was originally associated with a unique 5-nucleotide index located 44 base pairs downstream of the IR **(Supplementary Table 2)**. We performed Sanger sequencing on the DNA extracted for five randomly selected IRs in the TRIP pool to confirm the absence of recombination and the correct association between barcodes and corresponding promoters. We designed primers for nested PCR, including two IR-specific primers, PB-Valid.3-Out primer-1 and Inner-1, and locus-specific genome spanning primers, GEMP_mvla_outer and GEMP_mval_inner **(Supplementary Table 7)**. PCR amplification was performed in two steps using the primer combinations PB-Valid.3-Out primer-1/GEMP_mval_outer for PCR1 and Inner-1/GEMP_mval_inner for PCR2. The resulting products were diluted in water (1:5) and prepared for Sanger sequencing using 10 μM of primers together with Big Dye terminator version 3.1.

Identification and characterization of Cas9-induced lesions

We parsed the DNA reads obtained from Cas9-targeted TRIP cells in order to map barcodes and UIDs[25] and extract the sequences of interest containing the 20-nucleotide region targeted by Cas9. These were aligned against the wild-type sequence to identify and characterize Cas9-induced lesions. We used semi-global alignment with the following weights: match + 2, mismatch −2, gap opening penalty −5, gap extension penalty −0.5, and initial score 30. For TRIP cell line experiments with sgRNA1, sgRNA2, and sgRNA3, respectively, 96%, 95%, and 91% of the reads could be parsed, whereas TRIP pools yielded between 86% and 89% of successfully parsed reads. Quality control and filtering of aberrant barcodes were further performed as described in previous work[25].

Analysis of mutation frequencies, sizes and patterns

We analyzed sequenced IR regions with a read coverage of at least 30 in all Cas9 assays on TRIP cells. The average read coverage per IR was significantly higher in the TRIP cell line than in the pool **(Supplementary Figure 9)**. However, the cell line contained only 36 barcoded regions, whereas the pool offered ~ 1k IRs. As a result, we used TRIP pool data primarily to assess overall trends of variation in IR mutation frequency throughout the genome and high-resolution TRIP cell line data to identify mutation sizes and patterns.

Relative frequencies of each lesion type were calculated per IR as the ratio between the number of reads exhibiting such lesion and all reads for the given IR **(Figs 2A–C)**. We plotted frequencies per lesion type for each of 36 IRs in the TRIP cell line **(Fig 2A)**, as well as distributions **(Fig 2B)** and stacked medians stratified by promoter **(Fig 2C)** considering all 1359 IRs shared by TRIP pool assays. To assess variation in IR mutation frequency with genomic locus in the TRIP pool, we fitted linear regression models describing the relationship between mutation frequencies in each pair of assays and determined corresponding R^2 and F-statistic p-values **(Fig 2D)**. Additionally, we fitted a global linear regression model to IR mutation frequencies in all TRIP pool assays taking into account genomic locus, guide RNA, knock-in, promoter, and a significant interaction term between genomic locus and guide RNA **(Fig 2F)**. Based on this model, we determined η-squared values denoting the effect size or variance explained by each independent variable **(Fig 2G)**. For this, we relied on multi-way ANOVA tests with partial sum of squares preserving the principle of marginality (type II sum of squares). The ratio between knock-in and other insertions was calculated per IR and plotted against binned total IR mutation frequency together with the mean and confidence interval of 0.95 obtained by non-parametric bootstrapping **(Fig 2D)**. Association between IR total mutation frequency and regulatory element scores was determined using Pearson's correlation, along with the corresponding t-test *p*-values corrected for multiple testing using the Benjamini-Hochberg procedure **(Figs 3B–D)**. We calculated and plotted overall deletion and insertion frequencies per size in the TRIP cell line **(Fig 4A)**. In addition, we determined the frequencies of individual deletion patterns **(Fig 4B)** and the distributions of insertions of the four DNA bases in predominant one-nucleotide insertions **(Fig 4D)**. Deletions with ambiguous positions could not be distinguished. We took this ambiguity into account when calculating the frequency of deletions neighboring the expected or alternative break site. Specifically, we weighed down the frequency according to the ratio of ambiguous positions neighboring the corresponding sites **(Fig 4C)**. All differences between distributions reported in the manuscript were determined using non-parametric two-tailed Wilcoxon tests. All given confidence intervals are calculated for confidence level 0.95.

Analysis of one-nucleotide insertion patterns

In order to assess different models[3,31,32] of Cas9 cleavage, we had to first resolve ambiguity in insertion sites. Ambiguity arises due to insertions of DNA bases next to matching nucleotides in the target sequence. For instance, insertions of an A immediately up or downstream of a wild-type A cannot be distinguished, as they result in identical mutated sequences (**A**A and A**A)**. Since these results are confounded, both positions represent ambiguous insertion sites for the added base A. These observations are also true for longer and more complex sequences. Additionally, the deterministic nature of variant callers means that an insertion is called where the change is identified, thus once the repetition is detected. When processing sequences 5′ to 3′, insertions above would always be identified as occurring downstream of the wild-type A (A**A)**. We redistributed accumulated insertion counts throughout ambiguous sites based on unambiguous insertions and the Cas9 cleavage and DNA repair models as follows **(Fig 5B)**. The wild-type non-target sequence was processed to identify ambiguous DNA base(s) at each site, essentially those matching the 5′ and 3′ neighbors in the wild-type sequence. Individual ambiguous sites were extended to maximal windows of consecutive ambiguous sites for the same DNA base. We performed the redistribution of ambiguous DNA base insertions within these windows. For blunt ends with template-independent insertions, without bias towards specific DNA bases, ambiguous counts were distributed to follow the relative proportions of unambiguous insertions at the sites within the window. For staggered ends repaired by template-based insertions, we classified sites within a window as likely or unlikely based on the expected insertion pattern. Considering a 1-bp 5′ overhang model, ambiguous sites with downstream nucleotide identical to the inserted DNA base would be considered likely whereas those with a different downstream nucleotide would be unlikely given the conformation of the DNA ends. We re-calculated the counts for ambiguous unlikely sites as the average of the counts observed for unambiguous DNA bases at those same sites. Likely sites received the remaining counts, distributed according to the relative proportions of unambiguous DNA bases. For the combined blunt and staggered models targeting primarily between nucleotides 3–4 and 4–5 upstream of the PAM (non-target DNA), we applied the above strategies for blunt and staggered models respectively to these sites. Since the ratios between blunt and staggered were not known for the remaining sites, we used the simpler estimate based on relative proportions of unambiguous DNA base insertions. Finally, we compared the estimates against expected distributions **(Figs 5A, 5B)**.

Re-analysis of Cas9-induced DNA end structure data

We downloaded targeted sequencing data of DNA ends induced by Cas9 targeting using the SRA-toolkit, after selection using the NCBI SRA Run Selector, specifically runs SRR1617071 to SRR1617082 in project PRJNA264361. The assays involved Cas9 cleavage guided to a region on chromosome 6 in mouse Ligase IV-deficient pre-B cells arrested in G1 phase[33]. Upon Cas9 targeting, top and bottom DNA ends were ligated into a hairpin and subject to bisulfite treatment to facilitate amplification for high-throughput DNA sequencing. We aligned the sequences in each file against the wild-type hairpin sequence, assuming blunt ends between nucleotides 3–4 upstream of the PAM. For this purpose, we used BLASTn with the following options -max_target_seqs 500000 -max_hsps 1 -num_threads 30 -outfmt “6 qacc sacc qstart qend sstart send evalue bitscore length pident nident mismatch positive gaps ppos btop”. We filtered low-quality sequences and primers by keeping alignments at least 50-bp long containing at most 5 gaps. We also filtered sequences whose alignment did not span a region of at least 50-bp around the expected target site. We then used both the starting position of each alignment in the reference sequence and the BLASTn trace-back operations (BTOP) to isolate the pattern of changes relative to wild-type within a region of 30-bp around the expected target site. In parallel, we generated mutation patterns for all possible DNA end conformations within a target region of 30-bp around the expected target site. Finally, we matched the generated mutation patterns against the patterns extracted from the HCoDES data to identify the corresponding DNA end structures. Note that many of the HCoDES sequences contained substitutions at random positions, primarily TC, possibly also due to failed bisulfite conversion. Since substitutions did not show consistent patterns associated with particular DNA end conformations and had no impact on the ranking of top observed patterns, we decided to ignore them so that they would not confound otherwise valid matching patterns.

ACKNOWLEDGMENTS

We thank the NKI Genomics Core Facility and the NKI Flow Cytometry Facility for their technical support. We also thank Anton Berns, Heinz Jacobs, Hein te Riele, Tim Harmsen, Eva Brinkman and Bas van Steensel for helpful discussions and/or critical reading of the manuscript. This research was supported by Netherlands Organization for Scientific Research (NWO) grant 823.02.007 to W.A. and M.v.L.. A.V.P. was supported by European Research Council (ERC) Advanced Grant 293662 to Bas van Steensel.

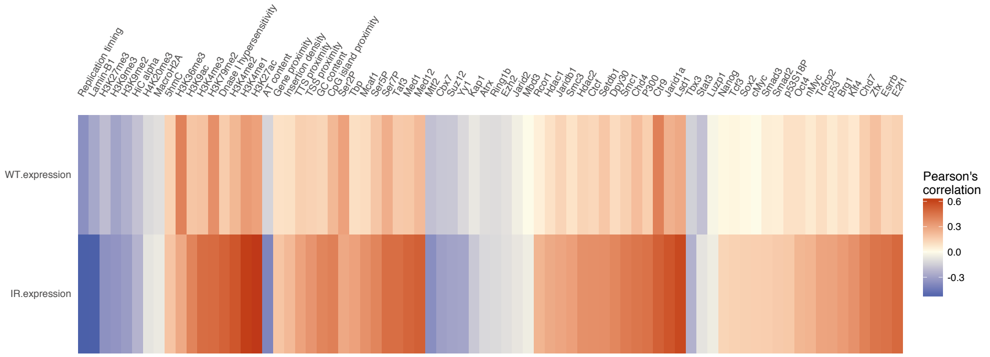
REFERENCES

1. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823 (2013).
2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819 (2013).
3. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337**, 816–821 (2012).
4. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science (80-.).* **348**, 1477–1481 (2015).
5. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (80-.).* **343**, (2014).
6. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62 (2014).
7. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569 (2014).
8. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, 471 (2013).
9. Lieber, M. R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annu. Rev. Biochem.* **79**, 181 (2010).
10. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9&ndash;mediated gene inactivation. *Nature Biotechnology* 1–8 (2014).
11. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun. 2018 91* **9**, 1–15 (2018).
12. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
13. Xu, H. *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147 (2015).
14. Chari, R., Mali, P., Moosburner, M. & Church, G. M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods* **12**, 823 (2015).
15. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262–1278 (2014).
16. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 677–683 (2014).
17. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670 (2014).
18. Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350**, 823–826 (2015).
19. Horlbeck, M. A. *et al.* Nucleosomes impede cas9 access to DNA in vivo and in vitro. *Elife* **5**, (2016).
20. Yarrington, R. M., Verma, S., Schwartz, S., Trautman, J. K. & Carroll, D. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proc. Natl. Acad. Sci.* **115**, 9351–9358 (2018).
21. Chen, X. *et al.* Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res.* **44**, 6482 (2016).
22. Daer, R. M., Cutts, J. P., Brafman, D. A. & Haynes, K. A. The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *ACS Synth. Biol.* **6**, 428 (2017).
23. Jensen, K. T. *et al.* Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett.* **591**, 1892–1901 (2017).
24. Uusi-Mäkelä, M. I. E. *et al.* Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). *PLoS One* **13**, (2018).
25. Akhtar, W. *et al.* Chromatin Position Effects Assayed by Thousands of Reporters Integrated in Parallel. *Cell* **154**, 914–927 (2013).
26. Akhtar, W. *et al.* Using TRIP for genome-wide position effect analysis in cultured cells. *Nat. Protoc. 2014 96* **9**, 1255–1281 (2014).

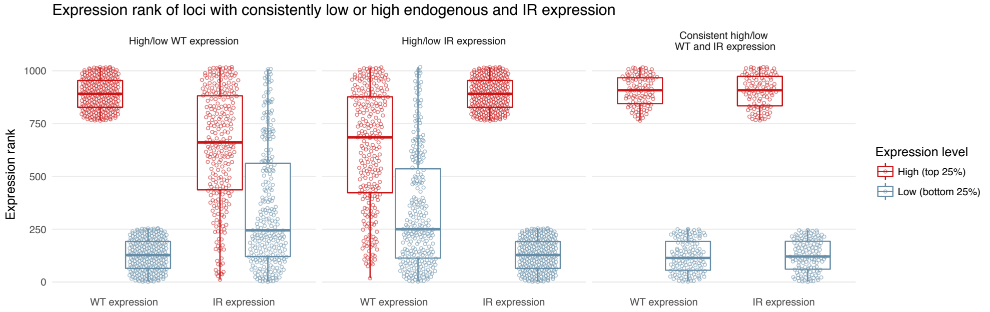
27. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827 (2013).
28. de Jong, J. *et al.* Chromatin Landscapes of Retroviral and Transposon Integration Profiles. *PLoS Genet.* **10**, 1004250 (2014).
29. Liang, G. *et al.* Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7357 (2004).
30. Li, Q. *et al.* Polycomb CBX7 Directly Controls Trimethylation of Histone H3 at Lysine 9 at the p16 Locus. *PLoS One* **5**, (2010).
31. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2579 (2012).
32. Zuo, Z. & Liu, J. Cas9-catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations. *Sci. Rep.* **5**, (2016).
33. Dorsett, Y. *et al.* HCoDES reveals chromosomal DNA end structures with single nucleotide resolution. *Mol. Cell* **56**, 808 (2014).
34. Stead, E. *et al.* Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* **21**, 8320–8333 (2002).
35. Shibata, A. *et al.* DNA Double-Strand Break Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during Homologous Recombination. *Mol. Cell* **65**, 671 (2017).
36. Symington, L. S. End Resection at Double-Strand Breaks: Mechanism and Regulation. *Cold Spring Harb. Perspect. Biol.* **6**, (2014).
37. McVey, M. & Lee, S. E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* **24**, 529 (2008).
38. Liao, S., Tammaro, M. & Yan, H. Enriching CRISPR-Cas9 targeted cells by co-targeting the HPRT gene. *Nucleic Acids Res.* **43**, (2015).
39. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
40. van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **63**, 633–646 (2016).
41. He, X. *et al.* Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.* **44**, e85 (2016).
42. Masui, S. *et al.* An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. *Nucleic Acids Res.* **33**, e43 (2005).
43. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519 (2008).
44. Boshart, M. *et al.* A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* **41**, 521–530 (1985).
45. Studer, M., Pöpperl, H., Marshall, H., Kuroiwa, A. & Krumlauf, R. Role of a conserved retinoic acid response element in rhombomere restriction of Hoxb-1. *Science* **265**, 1728–1732 (1994).
46. Ogura, T. & Evans, R. M. A retinoic acid-triggered cascade of HOXB1 gene activation. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 387 (1995).
47. Okazawa, H. *et al.* The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. *EMBO J.* **10**, 2997 (1991).
48. Kuroda, T. *et al.* Octamer and Sox Elements Are Required for Transcriptional cis Regulation of Nanog Gene Expression. *Mol. Cell. Biol.* **25**, 2475 (2005).
49. Tamaru, M. *et al.* Selective activation of testis-specific genes in cultured rat spermatogenic cells. *Biochim. Biophys. Acta* **1049**, 331–338 (1990).
50. Mcburney, M. W. *et al.* The mouse Pgk-1 gene promoter contains an upstream activator sequence. *Nucleic Acids Res.* **19**, 5755 (1991).
51. Wu, D. Y. & Yao, Z. Isolation and characterization of the murine Nanog gene promoter. *Cell Res.* **15**, 317–324 (2005).

52. Bernard, O., Cory, S., Gerondakis, S., Webb, E. & Adams, J. M. Sequence of the murine and human cellular myc oncogenes and two modes of myc transcription resulting from chromosome translocation in B lymphoid tumours. *EMBO J.* **2**, 2375 (1983).
53. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9530–9535 (2011).
54. Cadiñanos, J. & Bradley, A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* **35**, e87 (2007).

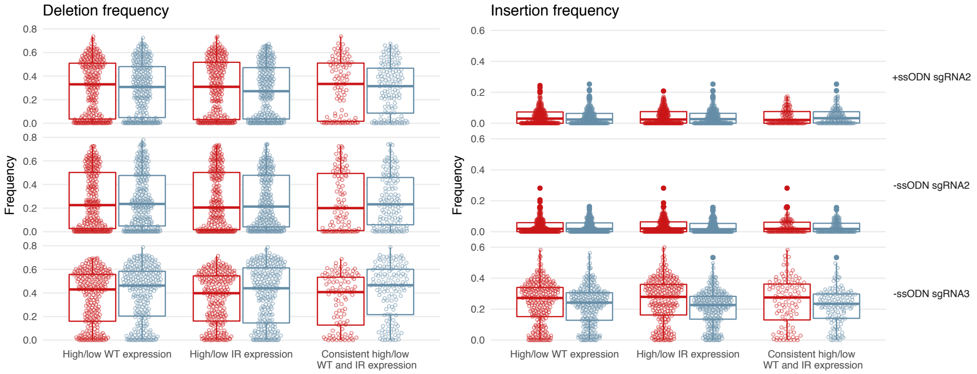
SUPPLEMENTARY MATERIAL



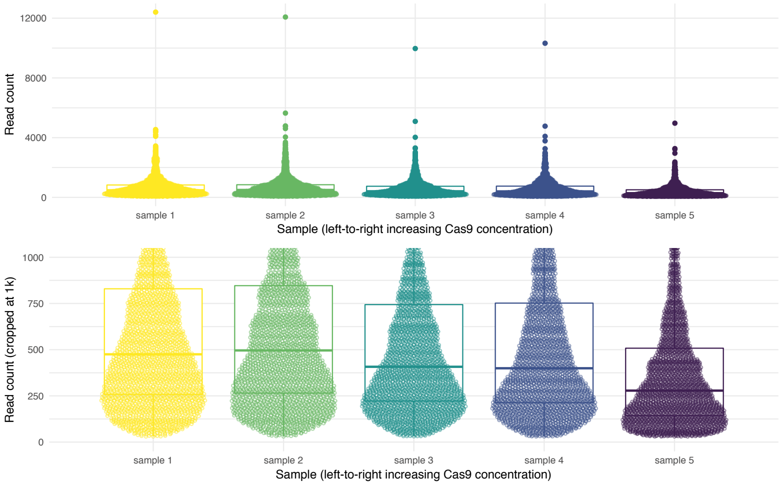
Supplementary Figure 1. Correlation of expression levels at IR loci in mES TRIP cells (IR expression) and at the corresponding endogenous loci in wild-type mES cells (WT expression), with TGE features of wild-type mES cells. Note: "IR expression" is the same as "IR expression" or "Reporter expression" in Figure 3D. "WT expression" is the same as "Gene expression" in Figure 3D.



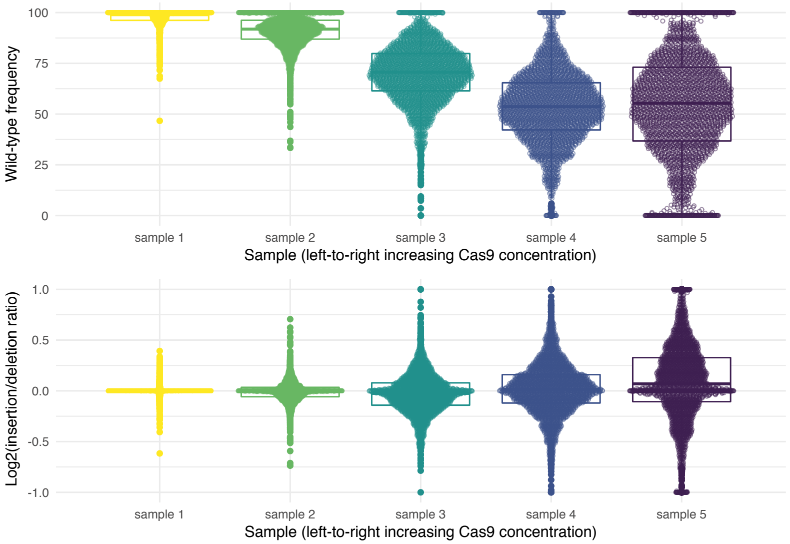
Supplementary Figure 2. Distributions of WT and IR expression levels for groups of loci with (left) high/low WT expression, (middle) high/low IR expression, and (right) consistent high/low WT and IR expression. Each circle corresponds to a distinct IR locus. Red and blue denote groups of loci with high (top 25%) and low (bottom 25%) expression, respectively.



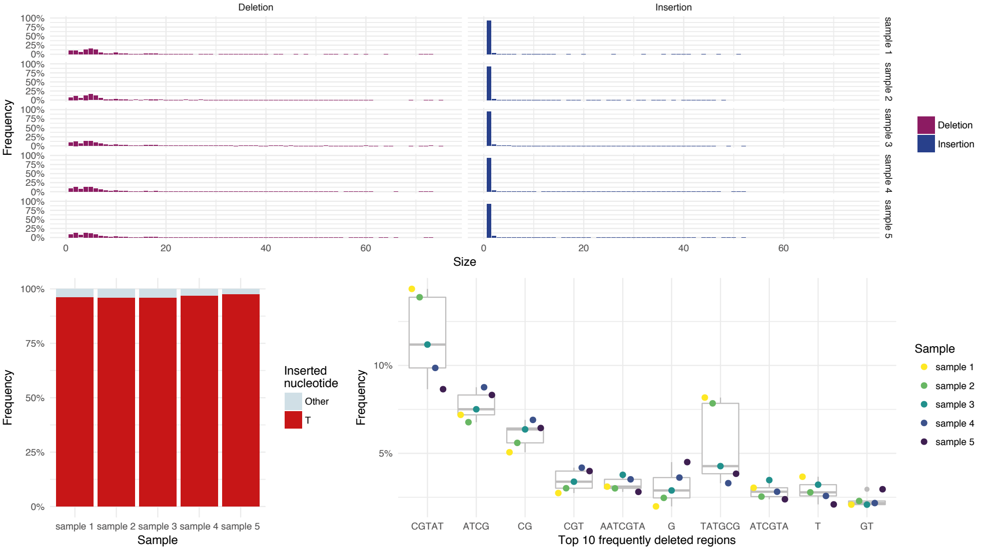
Supplementary Figure 3. Distribution of Cas9-induced mutation frequencies (deletion and insertion) obtained in the three different TRIP pool assays for the same loci groups in Supplementary Figure 2.



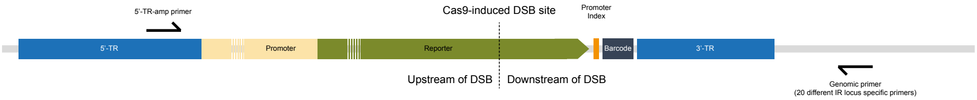
Supplementary Figure 4. Distributions of read counts per IR for five PGK-driven TRIP pools sorted by Cas9 concentration (samples 1 to 5, with increasing Cas9 concentration). Top: entire distribution. Bottom: plot cropped at 1000 read count.



Supplementary Figure 5. Frequency of wild-type reads (top) and log2 ratio between insertion and deletion frequencies (bottom).



Supplementary Figure 6. Mutation sizes and patterns in five PGK-driven TRIP pools with different Cas9 concentrations. Top left: distribution of deletion sizes in samples 1-5. Top right: distribution of insertion sizes in samples 1-5. Bottom left: frequency of the most inserted nucleotide versus the rest in samples 1-5. Bottom right: frequency of the ten most deleted regions in samples 1-5; the deleted regions are displayed left to right in decreasing order of their corresponding frequency in the cell line (Figure 4B).



Supplementary Figure 7. Reporter construct and location of amplification primers for assessment of IR-promoter missassociation.

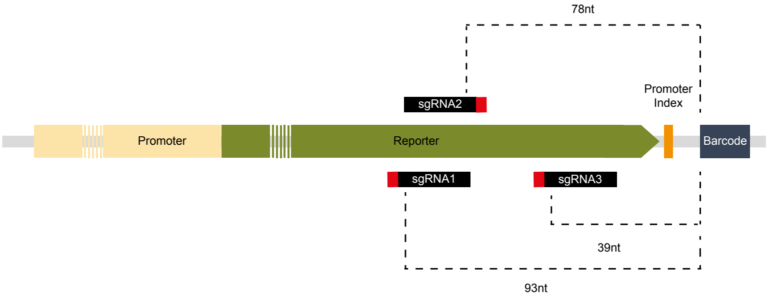
Supplementary Table 1. Validation of promoter sequences for 20 randomly chosen IRs. Column "Genomic primer" denotes one of 20 primers targeting a genomic region downstream of the IR barcode, according to Supplementary Table 4. Columns "Chrom" to "Barcode" specify IR-related information, namely the chromosome, location within the chromosome, promoter, strand, and barcode. Column "Amplification product" specifies whether we were successful in obtaining an amplification product. Column "Sequence" indicates whether we could sequence the amplification product (NA when no amplification product was available). Column "Valid sequence" indicates "match" if the sequence corresponds to the intended IR locus, and "no match" otherwise (NA if no sequence was available). Column "Promoter match" indicates "match" if the promoter for a valid sequenced IR locus corresponds to the expected promoter, "no match" otherwise (NA if no valid sequence could be obtained).

Genomic primer	Chromosome	Base	Promoter	Strand	Amplification product	Sequence	Valid sequence (match IR downstream DSB)	Promoter match
1	4	43734244	Oct4	+	no	NA	NA	NA
2	9	13630461	Hoxb1	-	yes	yes	match	match
3	20	146955246	Hoxb1	+	yes	no	NA	NA
4	6	37604358	CMV	-	yes	yes	no match	NA
5	14	96909472	Hoxb1	-	no	NA	NA	NA
6	9	115555263	Hoxb1	+	yes	yes	match	match
7	11	69347077	CMV	-	yes	yes	match	match
8	1	183806468	Hoxb1	-	yes	yes	no match	NA
9	12	4925222	Oct4	+	no	NA	NA	NA
10	6	6066083	Hoxb1	-	no	NA	NA	NA
11	6	26952426	CMV	-	yes	yes	match	match
12	2	171545278	Oct4	-	no	NA	NA	NA
13	14	101980700	Oct4	+	NA*	NA	NA	NA
14	4	108911304	PGK	-	yes	yes	no match	NA
15	6	15721098	Oct4	-	yes	yes	no match	NA
16	11	62252522	Oct4	+	no	NA	NA	NA
17	1	183049789	CMV	-	no	NA	NA	NA
18	6	86415855	Hoxb1	+	no	NA	NA	NA
19	16	97566570	cMyc	-	no	NA	NA	NA
20	9	22558808	Hoxb1	-	yes	yes	match	match

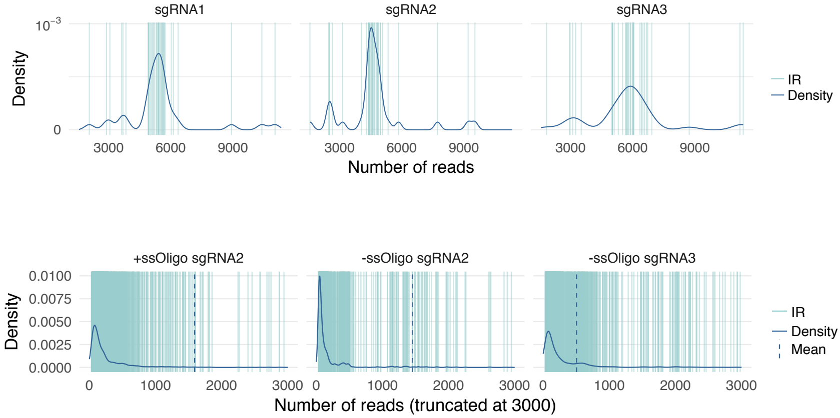
* No reliable primer could be designed

Supplementary Table 2. Primers used to investigate the frequency of translocations upon Cas9 targeting. Listed are the sequences of the 20 IR-specific genomic primers, plus the sequence of the 5′-TR primer.

Primer name	IR locus (chr. & base)		IR promoter	Promoter index	IR barcode	IR strand	Primer sequence (5′-3′)
Genomic 1	4	43734244	Oct4	GAGCG	TCTACATTTACGCATCGC	+	GTCAGAGGATATTAGAAAAGGCAGGAAGATAG
Genomic 2	9	13630461	Hoxb1	CGTCT	TAAGCAACTGATGCCGAC	-	CTCTTATATACATGGACTACGGTGGAGTAAAG
Genomic 3	20	146955246	Hoxb1	CGTCT	GATGTGTTTTGATGCCAC	+	GAGGTTAATTCTGGTAAGGGCTCAAAAGATG
Genomic 4	6	37604358	CMV	CAGCT	TCACCAGATTTCCGTTGG	-	CTACTCATTATTATTAGACCCCACCCCATCTG
Genomic 5	14	96909472	Hoxb1	CGTCT	TGACGCTTAAGCTACGAC	-	GTCTCAGGAACATTGGGAATTATACTTGACTG
Genomic 6	9	115555263	Hoxb1	CGTCT	TTCCCAACCTAAGGTTCC	+	TGAAAGATAATAAAGAACGGGCCTTGGTTG
Genomic 7	11	69347077	CMV	CAGCT	ATGCACAATGAGCGAAAC	-	GAGAGTGTAGATTAGTCAATTCAGGATCTG
Genomic 8	1	183806468	Hoxb1	CGTCT	TGTCTCCTAACGTGGTGT	-	ATATAAAGTTAGGACCTCTGTTCACTCCAGTC
Genomic 9	12	4925222	Oct4	GAGCG	CGATTTCAAATGTCCTTT	+	CGACACGTTTCTTTCTACATACTCCCTTTGTA
Genomic 10	6	6066083	Hoxb1	CGTCT	ACCGCGCTAAAAACATAT	-	GGATTTCTAATTCTATCCGTCCCTAACAAG
Genomic 11	6	26952426	CMV	CAGCT	CCTTTCGCGATAGCTCTG	-	TCTTCAGAAAATAGTCATACCCTCCTACACTG
Genomic 12	2	171545278	Oct4	GAGCG	AAAACTAGAGGAGAATAC	-	CTACAGTAGAGAGGCAGATACATGTTATTAGC
Genomic 13	14	101980700	Oct4	GAGCG	CTAATCGAGTAACATCAC	+	no sequence-specific primer
Genomic 14	4	108911304	PGK	TGTCT	CTTCCACCCAGCTACGGA	-	GATAGAACGTATCAATGTGGCTAAAACTCCCA
Genomic 15	6	15721098	Oct4	GAGCG	ACGAAATCGGGCCAGATC	-	TTATTTCATCTTCAATGGCACCTCATCTGC
Genomic 16	11	62252522	Oct4	GAGCG	AGGCCAGTGTATTTACGT	+	GCTCATCCTGCTCACCTGTTTACTATCTTAT
Genomic 17	1	183049789	CMV	CAGCT	AATGCCTCTAGCAGTAGG	-	GAGTTCCCAGAATAAACAGTCCAAGTCTTAAC
Genomic 18	6	86415855	Hoxb1	CGTCT	GATAAGGTAGCCAGAGAT	+	GATCTCTTCAAAACACACACACATTCCTCTG
Genomic 19	16	97566570	cMyc	ATACA	AGGGACTCCAGAAAGAAC	-	CCTTTCTCTGCCAGCTTTTAAACGTAATGAG
Genomic 20	9	22558808	Hoxb1	CGTCT	TTGCCCGACATCCCATTG	-	CTCTTTTAGCATTACACTCACGATCTGCAAAC
5′-TR-amp							GCGGTAAGTGTCACTGATTTTGAACTATAACG



Supplementary Figure 8. Distance between the IR barcode and the sites targeted by sgRNA1-3 within the IR gene body.



Supplementary Figure 9. Read coverage per IR in TRIP cell line and TRIP pool experiments. The average read coverage per IR was significantly larger in the TRIP cell line (top) than in the TRIP pool (bottom).

Supplementary Table 3. Sequences of the guide RNAs and the custom oligo for insertion by HDR used in the RNA-guided Cas9 targeting experiments of TRIP cell lines and pools. mCherry amplifications primers.

gRNA/oligo for HDR	Target sequence	Oligonucleotide sequence
gRNA1	GCCGAGAGTGATCCCGGCGG	CACCGCCGAGAGTGATCCCGGCGG AAACCCGCCGGGATCACTCTCGGC
gRNA2	CGCCGGGATCACTCTCGGCA	CACCgCGCCGGGATCACTCTCGGCA AAACTGCCGAGAGTGATCCCGGCGc
gRNA3	AGTGTCACCTAAATCGTATG	CACCgAGTGTCACCTAAATCGTATG AAACCATACGATTTAGGTGACACTc
	Sequence	
Custom single-stranded oligonucleotide (ssODN)	tgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgcGATTACAGATATCATCAACAGcggcatggacgagctgtacaagtaagaattcgcggccgcatacgatttaggtgacactgcag	
mCherry amplification primers	**Primer sequence**	
Fragment.FOR	TGCTGGGGAGCGGCCGCTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA	
Fragment.REV	CGCATCAGGCGCCCCTGCAGGCCAGCTGGTTCTTTCCGCCTCAGAAG	

Supplementary Table 4. Sequences of the primers used in preparation of the TRIP cell line and TRIP cell pool samples for sequencing following RNA-guided Cas9 targeting.

Targeted cells	Primer name	Illumina sequencing adaptor	Index sequence	Complementary sequence
TRIP cell line (MiSeq)	PB-cDNA-forward-1-BC-(1-4)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	NNNNNNNNNN	GTCACAAGGGCCGGCCACAA
TRIP cell line (MiSeq)	PB-cDNA-Reverse-5	ACGGCAACATCCTGGGGCACAAG		
TRIP cell pool (HiSeq)	PB-cDNA-Reverse-3	CGACAACCACTACCTGAGCACCCA		
TRIP cell pool (HiSeq)	Thio-Solexa-Ad-SE-1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTATTCCGAT*C*T		
TRIP cell pool (HiSeq)	PB-cDNA-Reverse-2a	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTGATCACATGGTCCTGCTGGAGTT		

Supplementary Table 5. Sequences of the UID-specific primers used in preparation of the TRIP cell pool samples for sequencing following RNA-guided Cas9 targeting.

Primer name	Illumina sequencing primer	Index sequence	UID 16-nt	Annealing part of primer
PB-cDNA-f-UID-16n-1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	AACGTGAT	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-2	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	AAACATCG	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-3	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	ATGCCTAA	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-4	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	AGTGGTCA	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-5	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	ACCACTGT	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-6	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	ACATTGGC	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-7	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CAGATCTG	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-8	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CATCAAGT	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-9	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CGCTGATC	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-10	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	ACAAGCTA	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-11	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	CTGTAGCC	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA
PB-cDNA-f-UID-16n-12	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	AGTACAAG	NNNNNNNNNNNNNNNN	GTCACAAGGGCCGGCCACAA

Supplementary Table 6. PCR amplification protocol parameters.

PCR amplification steps	Temperature	Time	
Assembly of reaction	98˚C	3:00	1X
	98˚C	0:15	2X
Reagents:	65˚C	0:30	
» PB-cDNA-Reverse-3 10 µM	72˚C	1:00	
» PB-cDNA-f-UID-16n-x 10 µM	4˚C	Forever	1X
Exonuclease treatment	37˚C	1:00:00	1X
Reagent:	98˚C	5:00	1X
» Exonuclease-I (20 U/µl) 3 µl	4˚C	Forever	1X
Final amplification	98˚C	0:15	24X
Reagents:	65˚C	0:30	
» Thio-Solexa-Ad-SE-1 10 µM	72˚C	1:00	
» PB-cDNA-Reverse-2a 10 µM	72˚C	5:00	1X

Supplementary Table 7. List of five IRs from the TRIP pool that were randomly selected to validate the association between barcode and promoter index prior to Cas9 targeting. Sequences provided as a Source Data file.

Tested IR	Chromosome	Base	Strand	Barcode	Promoter	Index
1	4	43734244	+	TCTACATTTACGCATCGC	Oct4	GAGCG
2	9	13630461	-	TAAGCAACTGATGCCGAC	Hoxb1	CGTCT
3	20	146955246	+	GATGTGTTTTGATGCCAC	Hoxb1	CGTCT
4	6	37604358	-	TCACCAGATTTCCGTTGG	CMV	CAGCT
5	14	96909472	-	TGACGCTTAAGCTACGAC	Hoxb1	CGTCT

General reporter primers (5'->3')	Primer sequence
PB-Valid.3-Out primer-1	TCACATGGTCCTGCTGGAGTT
Inner-1	TCCTGCTGGAGTTCGTGACC

IR locus-specific genome spanning primers	
GEMP_mval_1_outer	TAAACCAGTCACTTTCTTGCACAGC
GEMP_mval_1_inner	TAAACCAGTCACTTTCTTGCACAGC
GEMP_mval_2_outer	ACTCCCAAATGCTTGCTCCAACTCT
GEMP_mval_2_inner	TCTATGCCCTCTTCCTCATCAGCTGT
GEMP_mval_3_outer	CAATGGGCACACACATGGGTTCAAG
GEMP_mval_3_inner	ATATGGAGAGAGAAGATGGAGTTCC
GEMP_mval_4_outer	AGGCAAGGGGTTTCATCACACACAG
GEMP_mval_4_inner	AGGAGAAGACTGGACATGATGGTGC
GEMP_mval_5_outer	TCGTGAAACTGGAAAGGCAAATGGT
GEMP_mval_5_inner	GCCACAAATGATATGGGCTGGAGGA

Sequencing primer (5'->3')	
PB-Valid.Gen.Seq-1	CCGGGATCACTCTCGGCA

Chapter 4

A GENOME-WIDE ENRICHMENT SCREEN IDENTIFIES NUMA1-LOSS AS A RESISTANCE MECHANISM AGAINST MITOTIC CELL DEATH INDUCED BY BMI1 INHIBITION

Santiago Gisler[1], Ana Rita R. Maia[2], Gayathri Chandrasekaran[1],
Jawahar Kopparam[1], Maarten van Lohuizen[1]

[1] Division of Molecular Genetics, Oncode and The Netherlands Cancer Institute,
Amsterdam, The Netherlands
[2] Division of Cell Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

Adapted from PLoS One. 2020 Apr 28;15(4):e0227592

ABSTRACT

BMI1 is a core protein of the polycomb repressive complex 1 (PRC1) that is overexpressed in several cancer types, making it a promising target for cancer therapies. However, the underlying mechanisms and interactions associated with BMI1-induced tumorigenesis are often context-dependent and complex. Here, we performed a drug resistance screen on mutagenized human haploid HAP1 cells treated with BMI1 inhibitor PTC-318 to find new genetic and mechanistic features associated with BMI1-dependent cancer cell proliferation. Our screen identified NUMA1-mutations as the most significant inducer of PTC-318 cell death resistance. Independent validations on NUMA1-proficient HAP1 and non-small cell lung cancer cell lines exposed to BMI1 inhibition by PTC-318 or *BMI1* knockdown resulted in cell death following mitotic arrest. Interestingly, cells with CRISPR–Cas9 derived *NUMA1* knockout also showed a mitotic arrest phenotype following BMI1 inhibition but, contrary to cells with wild-type NUMA1, these cells were resistant to BMI1-dependent cell death. The current study brings new insights to BMI1 inhibition-induced mitotic lethality in cancer cells and presents a previously unknown role of NUMA1 in this process.

INTRODUCTION

The chromatin-modifying Polycomb-group proteins are critical epigenetic transcriptional repressors controlling cell fate decisions, such as self-renewal and differentiation of stem cells, as well as tumorigenesis, primarily through the repression of downstream genes[1–3]. B lymphoma Mo-MLV insertion region 1 homolog (BMI1), an essential protein of the polycomb repressive complex 1 (PRC1), was first identified as an oncogene, inducing lymphomas in mice by cooperating with c-MYC[4,5]. The protein is often expressed in stem cells, and several reports have implicated its overexpression in cancer stem cell maintenance and the progression of different types of cancers[6–8]. By contrast, regulation of BMI1 with inhibitors or short hairpin RNAs (shRNAs) results in cellular senescence or apoptosis of several types of cancer cells[9–13] and sensitizes tumor cells to cytotoxic agents or radiation[14,15]. Because of this, BMI1 is an attractive target for future clinical therapies of different cancers.

BMI1 overexpression is a well-established inducer of cancer cell proliferation and resistance to cancer drug treatments of various cancer cell lines[16–18], highlighting the potential of specific BMI1 inhibitors. However, although BMI1 inhibition results in growth reduction and cell death of different cancer cell lines, the underlying mechanisms are often context-dependent and uncertain[11,13,19]. As a result, little is known about the genetic interactions and variations involved in BMI1 inhibition-derived lethality or the subsequent resistance.

In the present study, we performed a genome-wide screen for gene disruptions that could result in resistance to pharmacological inhibition of BMI1 by exposing mutagenized human haploid HAP1 cells to low concentrations of the BMI1-inhibitor AB057609107 (PTC-318). PTC-318 is a new inhibitor of BMI1, developed by PTC Therapeutics, USA, designed to regulate BMI1 expression post-transcriptionally. We show that reducing BMI1 levels by shRNA or inhibition with the small molecule inhibitor PTC-318 significantly reduced cell viability of different cancer cell lines. Through our genetic screen, we unveiled new genes that rescue cell death induced by BMI1 inhibitors and selected NUMA1 for follow-up studies. We identified a novel mitotic mechanism underlying BMI1-associated lethality using CRISPR–Cas9 derived knockouts of NUMA1 in both HAP1 cells and non-small lung cancer (NSLC) cell lines. Our results highlight a mechanism relying on mitotic arrest upon loss of BMI1 and a new genetic resistance mechanism. These observations add further knowledge to the complex and context-dependent involvement of BMI1 in cancer. Our findings contribute to a better understanding of BMI1-associated pharmacological cancer treatment strategies.

RESULTS

Loss of BMI1 induces cell death in cancer cell lines

In order to test BMI1-dependent lethality in HAP1 cells, we transduced doxycycline-responsive plasmid systems into wild-type HAP1 cells to generate an inducible expression of shRNAs targeting *BMI1* (shBMI1) or control shRNA (shRandom). We assessed knockdown efficiency upon doxycycline treatment by comparing RNA and protein levels at different time points. As expected, doxycycline-induced shBMI1 decreased BMI1 expression, both at RNA (avg knockdown, 93%) and protein level after 48 hours of doxycycline treatment **(Figs 1A, 1B)**. We next evaluated HAP1 cell viability upon shBMI1 induction through colony formation, cell count, and cell viability assays. shBMI1-induction reduced HAP1 cell proliferation after doxycycline exposure compared with shRandom-induced HAP1 cells, indicating that *BMI1* knockdown is lethal in HAP1 cells **(Figs 1C, 1D, and Supplementary Figure 1)**.

In order to use a pharmacological setting, and hence more clinically relevant, we inhibited BMI1 expression post-transcriptionally with the small molecule AB057609107 (PTC-318). PTC-318 was identified by performing a high-throughput compound screening through gene expression modulation by small molecules (GEMS) technology. A luciferase open reading frame flanked by the untranslated region (UTR) of BMI1 was used as a reporter to screen for small molecules inhibiting *BMI1* at the transcriptional level.

To avoid inhibitor-derived side-effects, we screened for the lowest concentrations of PTC-318, leading to cell death. PTC-318 titrations showed that treatment with low nanomolar concentrations reduced BMI1 protein expression compared with dimethyl sulfoxide (DMSO) treated HAP1 control cells **(Fig 1E)**. Moreover, 20nM and 40nM PTC-318 treatment-induced cell death as measured by colony formation and cell proliferation assays **(Figs 1F, 1G, and Supplementary Figure 1)**. These results suggest that similar to BMI1 knockdown, the pharmacological inhibition of *BMI1* with nanomolar concentrations of PTC-318 induces lethality in HAP1 cells.

To confirm that PTC-318-induced lethality was caused by BMI1 regulation and exclude potential off-target toxicity, we expressed BMI without its UTR in HAP1 cells. These cells showed increased BMI1 protein levels compared with wild-type HAP1 cells. Importantly, upon treatment with PTC-318, cells ectopically expressing BMI1 showed increased resilience to PTC-318 treatment compared with wild-type HAP1 cells **(Figs 2A, 2B)**, suggesting that BMI1 transcriptional inhibition is responsible for PTC-318 cytotoxicity.

NUMA1-integrations cause resistance to PTC-318-derived cell death in HAP1 screen

Given the specificity and potent cytotoxicity of PTC-318, we performed an insertional mutagenesis drug resistance screen in HAP1 cells to find new BMI1-associated interactions and resistance mechanisms. Due to their haploid genome, HAP1 cells are particularly amenable to mutagenesis as only a single allele requires inactivation to obtain a loss-of-function. This approach has been successfully used to investigate various biological processes, including virus entry[20–22], T cell-mediated killing[23], and drug response[24,25] in both haploid (HAP1) and near-haploid (KBM7) human cells.

In essence, wild-type HAP1 cells were transduced with a retroviral GFP gene trap that randomly integrates into the genome. The gene trap most commonly integrates into

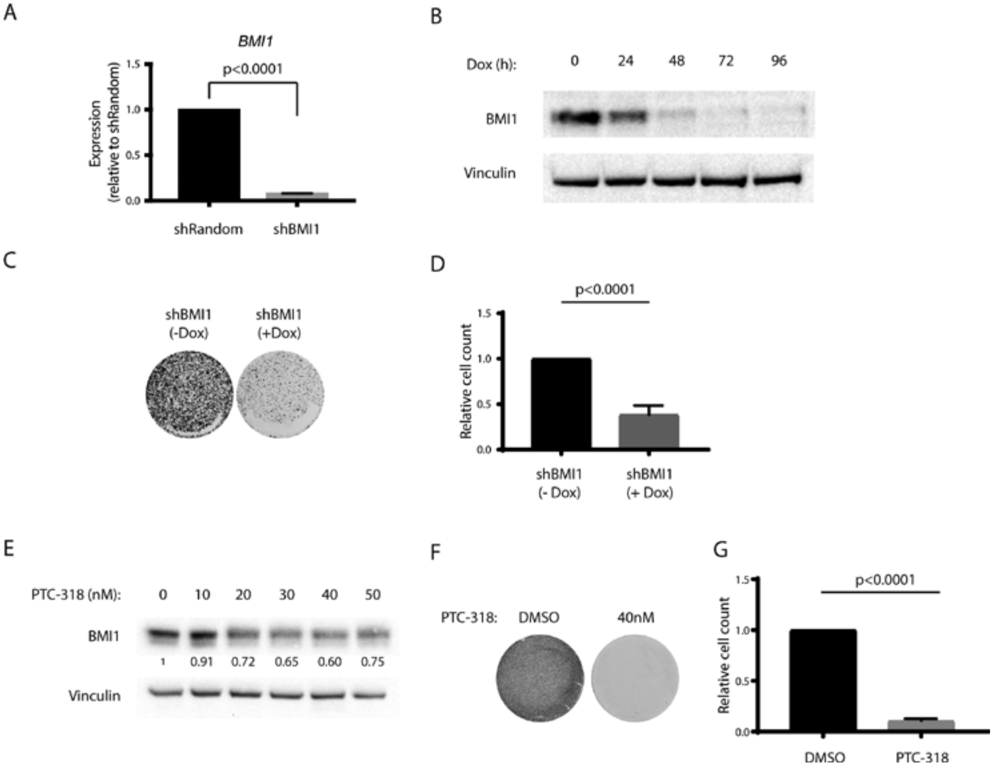


Figure 1. Genetic knockdown and pharmacological inhibition of BMI1 induces cell lethality. (A) Relative mRNA expression levels of BMI1 HAP1 cells transduced with inducible shRNA constructs FH1t-Random (shRandom) or FH1t-BMI1 (shBMI1) 96 hours after doxycycline treatment. (B) Protein expression levels of BMI1 in HAP1 cells transduced with shBMI1 0–96 hours after doxycycline treatment. Vinculin was used as a loading control. (C) Cell survival upon BMI1 knockdown was confirmed in HAP1 cells transduced with shBMI1 through a colony formation assay one week after plating with or without doxycycline treatment or (D) through relative cell counts 96 hours after plating with or without doxycycline. (E) Protein expression levels of HAP1 cells treated with different concentrations of PTC-318. Vinculin was used as a loading control. (F) Cell survival of HAP1 cells treated with PTC-318 through colony formation assay one week after treatment with DMSO (0.1%) or PTC-318 (40 nM) and (G) relative cell counts 48 hours after treatment with DMSO (0.1%) or PTC-318 (40 nM). Error bars represent SD. Student's t-test was performed for statistical testing.

introns and can land in two orientations: sense or antisense. Owing to the unidirectional design of the gene trap, sense-orientation integrations are likely to produce truncated and impaired transcription products **(Fig 2C)**.

We subjected mutagenized HAP1 cells to a stringent selection with 40 nM of PTC-318 for two weeks. Next, the genomic DNA of surviving HAP1 cells was isolated, and the insertion sites were subsequently amplified and sequenced. Gene trap integrations were then assigned to genes and counted as inactivating integrations if inserted in a sense-orientation with the transcriptional orientation in an intron or regardless of orientation

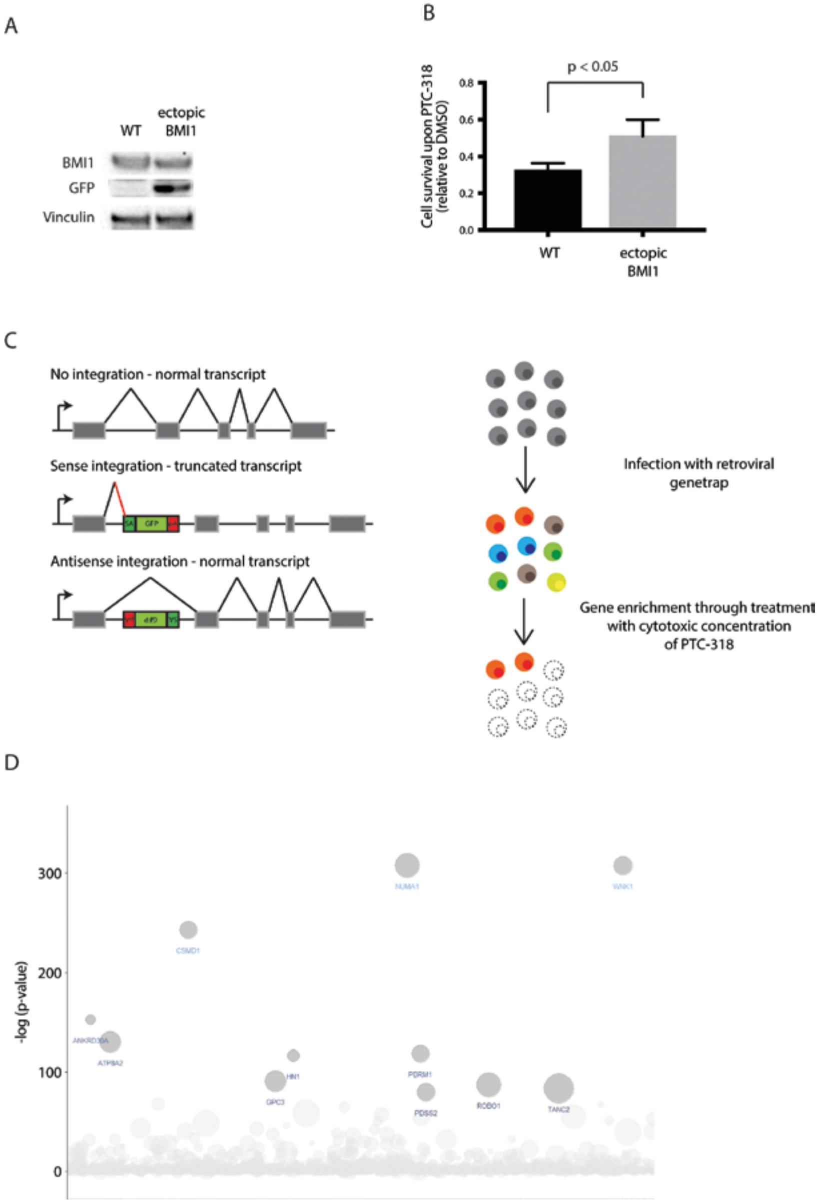


Figure 2. A haploid genetic screen shows enrichment of NUMA1 mutations upon BMI1 inhibition with PTC-318. (A) BMI1 and GFP protein expression of wildtype HAP1 cells and HAP1 cells ectopically expressing BMI1 in the absence of 3'UTR through the transduced FUGW-BMI1 construct. (B) HAP1 cell survival relative to corresponding DMSO-control 48 hours after treatment with 20 nM PTC-318. Error bars represent SD. Student's t-test was performed for statistical testing. (C) Schematic overview showing insertional mutagenesis of HAP1 cells using a retroviral gene trap (left) and enrichment screening process (right). (D) Bubble plot depicting genes enriched for unique gene-trap insertions in HAP1 cells treated with 40 nM PTC-318. The y-axis shows the significance, and the x-axis shows the genes for which the gene-trap insertions were mapped in alphabetical order. Size of the bubble corresponds to the number of unique inactivating gene-trap insertions.

in an exon. To identify enriched genes in the PTC-318 treated population, we compared the number of disruptive integrations per gene in our experimental dataset with a non-treated control dataset using a one-sided Fisher's exact test. We applied a stringent cut-off and considered all genes with an FDR-corrected p-value lower than 10^{-5} as enriched for inactivating integrations.

Interestingly, we found several reported antagonists of Polycomb repressors among the most significant hits required for cell survival, including two genes (*PBRM1* and *ARID2*) encoding essential PBAF subgroup proteins of the nucleosome remodeling complex SWI/SNF **(Fig 2D and Supplementary Table 1)**. Strikingly, the most significantly enriched gene in the HAP1 population, *NUMA1* ($p<0.0001$), had not previously been associated with *BMI1*. *NUMA1* encodes a 238 kDa nuclear mitotic apparatus protein, which is mainly reported to be involved in dynein-associated tethering of spindle pole microtubules during mitosis, recruiting dynein to the cell cortex[26–28]. NUMA1 has also been implicated in a conserved mechanism establishing asymmetric cell division associated with differentiation. The features associated with NUMA1 and its tumor-suppressing and cell-differentiating properties[29–32] suggested it was an attractive BMI1-antagonizing candidate, especially considering the role of BMI1 in maintaining cell stemness in healthy and cancer cells. To understand this new genetic association, we decided to focus on *NUMA1*-loss as a possible resistance-inducing mechanism upon BMI1 inhibition.

Knockout of *NUMA1* in HAP1 cells induces resistance to cell death upon BMI1 inhibition

As a validation of the screen results, we checked whether the loss of *NUMA1* expression would result in resistance to the lethality induced by BMI1 inhibition. We generated clonal *NUMA1* knockout HAP1 cell lines (NUMA1-KO) by targeting the first exon of *NUMA1* with CRISPR–Cas9 and selected two clones for validation to discard any phenotype changes resulting from clonality **(Figs 3A, 3D)**. We found that the two NUMA1-KO clones (NUMA1-KO 1 and NUMA1-KO 2) were significantly more resistant to the toxicity of shRNA-induced knockdown of BMI1 or PTC-318 treatment compared with wild-type HAP1 (hereafter named NUMA1-WT) cells **(Figs 3B, 3E)**. We found a more than 2-fold higher cell viability in NUMA1-KO cells compared with NUMA1-WT upon *BMI1* knockdown and inhibition, respectively, following quantification of Alamar blue-stained cells **(Figs 3C, 3F)**. Interestingly, we also observed an increase in NUMA1 protein levels upon PTC-318 treatment, which could represent an acute response to BMI1 inhibition.

In order to further validate that the NUMA1-KO resistance was specific to BMI1 inhibition, we attempted to establish BMI1 knockout (BMI1-KO) HAP1 clones. We transfected NUMA1-WT or NUMA1-KO HAP1 cells with CRISPR–Cas9 targeting *BMI1*, hypothesizing that NUMA1-WT cells would survive the BMI1-KO to a lesser extent compared with NUMA1-KO cells. While we were able to establish ten BMI1-KO clones in

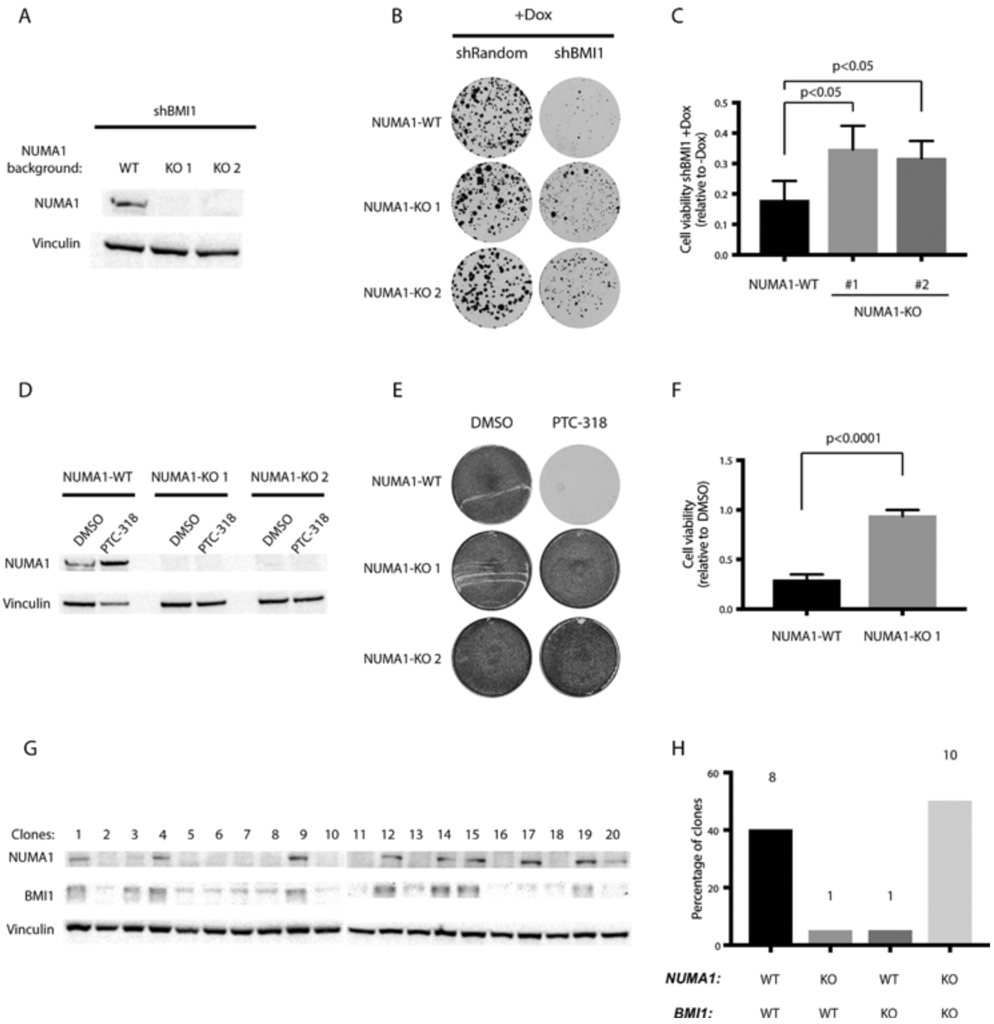


Figure 3. Cas9-induced NUMA1 knockout rescues HAP1 cytotoxicity caused by BMI1 inhibition. (A) Protein expression of NUMA1 in populations of HAP1 cells transduced with a doxycycline-inducible shBMI1 construct in NUMA1-WT or NUMA1-KO backgrounds. (B and C) Survival of the HAP1 NUMA1-WT and NUMA1-KO clones transduced with the inducible shBMI1 construct and maintained with or without doxycycline (B) colony formation and (C) cell viability. (D) Expression of NUMA1 protein in HAP1 NUMA1-WT and NUMA1-KO clones treated either with 0.1% DMSO or 40 nM of PTC-318. The survival of these cells was analyzed by (E) colony formation assay and (F) cell viability. (G) Protein expression of NUMA1 and BMI1 in NUMA1-WT clones, transfected with CRISPR–Cas9 targeting BMI1, single-cell sorted, and expanded (left). (H) Percentages of each genetic background based on the respective protein expression (right). Error bars represent SD. Student's t-test was performed for statistical testing.

a NUMA1-KO background, we retrieved only one NUMA1-WT clone with reduced BMI1 expression **(Figs 3G, 3H)**.

These results indicated that loss of NUMA1 results in resistance to the toxicity resulting from loss of BMI1, whether induced by shRNA, inhibitor, or CRISPR–Cas9. It also further confirmed the BMI1-specific effect of PTC-318 and that the NUMA1-derived resistance is specific to BMI1-associated lethality.

BMI1-loss induces mitotic arrest and cell death

Since NUMA1 is important for proper mitotic progression, we assessed the cell cycle profiles of HAP1 cells treated with a lower concentration of PTC-318 (20nM). We chose the lower concentration in order to maintain target engagement and limit the loss of cell viability.

Both genetic knockdown and pharmacological inhibition of *BMI1* resulted in an increased number of cells arrested in mitosis, as shown by the increase in Histone H3 phosphorylation (pH3). Compared with shRandom or DMSO controls, BMI1 knockdown and PTC-318 treatment resulted in a 2-fold and 8-fold increase in pH3 positive cell populations, respectively **(Fig 4A and Supplementary Figure 3)**.

Having confirmed mitotic arrest induced by BMI1 inhibition, we next followed the mitotic progression of HAP1 by live-cell imaging upon doxycycline treatment at different time-points (24, 48, and 72 hours) or 20 nM PTC-318 treatment **(Supplementary Figure 2)**. We found a modest albeit significantly increased time spent in mitosis in all doxycycline-induced shBMI1 HAP1 cells (41, 45, and 53 minutes, respectively) compared with control cells (34 minutes) **(Fig 4B and Supplementary Figure 4)**.

In contrast to the modest time-increases in mitosis observed in doxycycline-induced shBMI1 cells, analysis of NUMA1-WT HAP1 cells treated with PTC-318 showed a more prolonged mitotic arrest compared with DMSO-treated cells. While the mitotic duration of DMSO-treated cells was 31 minutes, PTC-318 treated NUMA1-WT cells spent 131 minutes in mitosis. In general, we observed that BMI1 inhibition with PTC-318 resulted in a robust mitotic arrest in both NUMA1-WT and NUMA1-KO cell lines. However, the arrest was longer in NUMA1-WT than in NUMA1-KO cell lines (131 minutes and 92 minutes, respectively).

After observing mitotic arrest induced by BMI1 inhibition both through shBMI1 and PTC-318, we examined the fates of these cells by live-cell imaging. While most untreated cells underwent chromosome segregation and completed cell division, we observed an enrichment of mitotic cell death upon BMI1 inhibition. This effect was more pronounced in PTC-318-treated cells than in the shBMI1 population **(Fig 4C)**. In agreement with the rescued cell lethality observed upon NUMA1 loss in the gene-trap screen and the subsequent validations, NUMA1-KO cells rescued the mitotic cell death induced by BMI1 inhibition. It has been described that the balance between two competing networks dictates cell fate. While the CyclinB1/cyclin-dependent kinase 1 (CDK1) complex is essential for mitotic progression, opposing signals, for example, the apoptosis-regulating protein B-cell lymphoma 2 (BCL-2), can induce mitotic cell death. The fate of a cell in mitotic

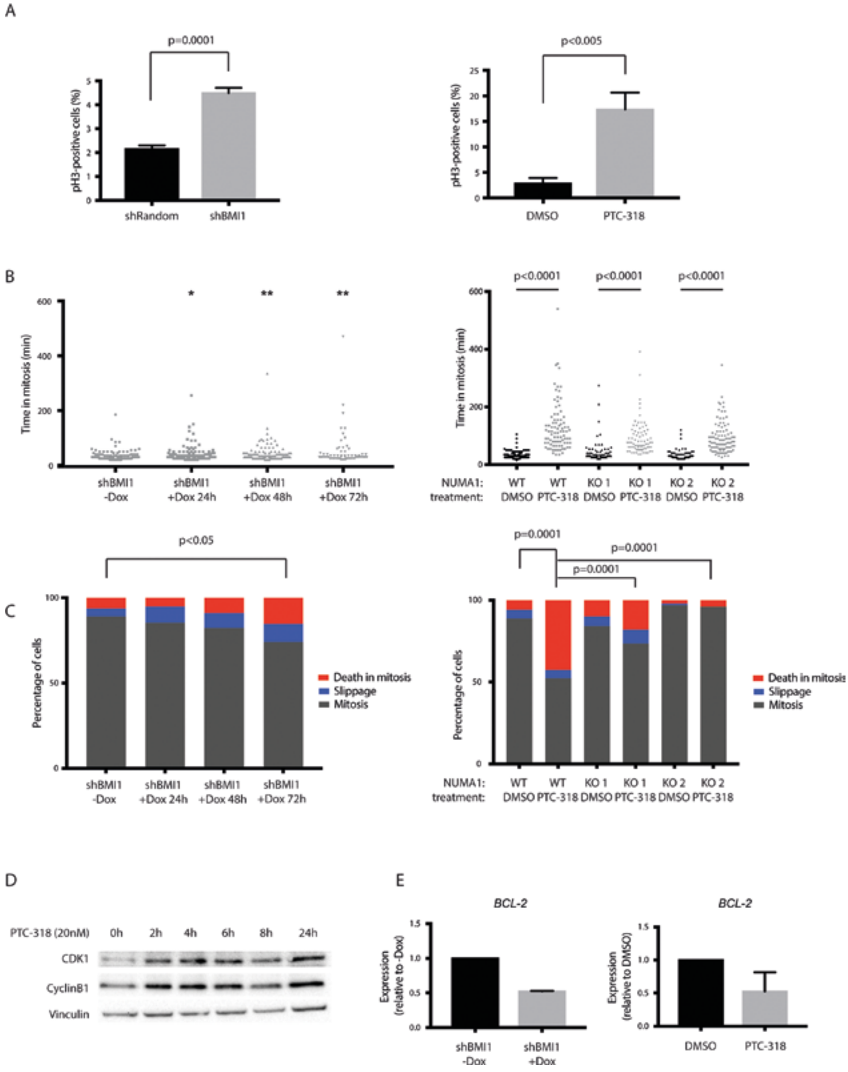


Figure 4. BMI1 inhibition induces mitotic cell death in NUMA1-WT cells but not in NUMA1-KO. (A) Quantification of phospho-histone H3 populations of HAP1 cells transduced with the inducible shRandom or shBMI1 constructs where cells were treated with doxycycline for 96 hours before measurement (left graph), and cells treated for 48 h with 0.1% DMSO or 40 nM PTC-318 (right graph). (B) Scatter dot plot representation of time in mitosis (from nuclear envelope breakdown to anaphase onset) of HAP1 cells transduced with the inducible shBMI1 construct and treated with doxycycline for 0-96 hours (left graph), or NUMA1-WT and NUMA1-KO cells treated with 20 nM PTC-318 (right). Bars represent the mean and standard deviation (SD). (C) Analysis of the mitotic fate of cells analyzed in (B). The graphs depict the percentage of cells in mitosis (dark grey), mitotic slippage (blue), and death in mitosis (red). (D) Protein expression of CDK1 and CyclinB1 in HAP1 cells at different time points after treatment with 40 nM of PTC-318. (E) Relative expression levels of BCL-2 in HAP1 transduced with inducible shBMI1 construct (with or without doxycycline for 96 hours; left) or treated with 0.1% DMSO or 20 nM PTC-318 for 4 hours (right). Error bars represent SD. Student's t-test was performed for statistical testing (*p<0.05 and **p<0.01).

arrest is thus linked to the antagonizing levels and timing of these two independent networks: if cell death signals increase, CyclinB1 levels decrease[33].

Thus, we measured the expression levels of CyclinB1/CDK1 and BCL-2 in BMI1-inhibited HAP1 cells. Consistent with the observed mitotic arrest, both CDK1 and CyclinB1 protein levels were elevated at early time-points after PTC-318 treatment (between 2 and 4 hours), and remained high up to 24 hours after treatment **(Fig 4D)**. In agreement with these results, live-cell imaging data showed that the onset of mitotic arrest was around 3–4 hours after treatment **(Supplementary Figure 4)**. In order to quantify cell death signals induced by BMI1 inhibition, we determined RNA levels of the anti-apoptotic marker *BCL-2* and observed a significant decrease after PTC-318 treatment compared with DMSO-treated control cells **(Fig 4E)**. Similar to the PTC-318 treatment, the induction of shBMI1 in the HAP1 cells significantly decreased *BCL2* expression.

Together, these results suggest that BMI1 inhibition induces mitotic arrest in HAP1 cells, ultimately causing cell death in the NUMA1-WT background but not in the NUMA1-KO population.

Inhibition of BMI1 induces mitotic cell death in NSCLC cell lines

We asked whether toxicity associated with BMI1-inhibition could be relevant to other cancer cell types and, hence, rule out that our findings were limited to haploid cells. Previous studies have demonstrated that BMI1 is overexpressed in non-small cell lung cancer (NSCLC) and that its expression is associated with NSCLC progression[34–36]. We, therefore, tested whether NSCLC would be sensitive to BMI1 inhibition. We treated NSCLC cell lines h520 and SK-MES-1 with PTC-318 and found that, as in HAP1 cells, low nanomolar concentrations of PTC-318 resulted in a significant reduction in cell survival of both lines (IC50 h520: 1.9 nM) (IC50 SK-MES-1: 7.8 nM) compared with untreated controls **(Figs 5A, 5B)**. Both NSCLC cell lines showed a significant accumulation of pH3 mitotic marker upon PTC-318 treatment, further suggesting that our findings in the HAP1 population can be generalized to other cancer types **(Figs 5C, 5D)**.

Finally, we validated the essential function of BMI1 in NSCLC as we did for HAP1 cells. We transfected the two NSCLC cell lines with CRISPR–Cas9 targeting *BMI1* and sorted single-cell clones. While we were unable to grow any single clones from the transfected SK-MES-1 population, we retrieved six h520 clones after single-cell sorting. Five of these clones showed regular expression of BMI1, and one showed reduced BMI1 expression **(Fig 5F)**. Interestingly, the clone with the lowest BMI1 expression also showed reduced levels of NUMA1 compared with the other h520 cell clones. These observations align with our main finding that BMI1 inhibition is tolerated only in the context of NUMA1 knockout or at least NUMA1 reduction, while it is lethal in NUMA1-proficient cells.

Altogether, we were able to show mitotic arrest and cytotoxicity caused by BMI1 loss in two commonly used NSCLC in vitro models, indicating that the observed phenotypes related to BMI1 inhibition extend to non-haploid cancer cells as well.

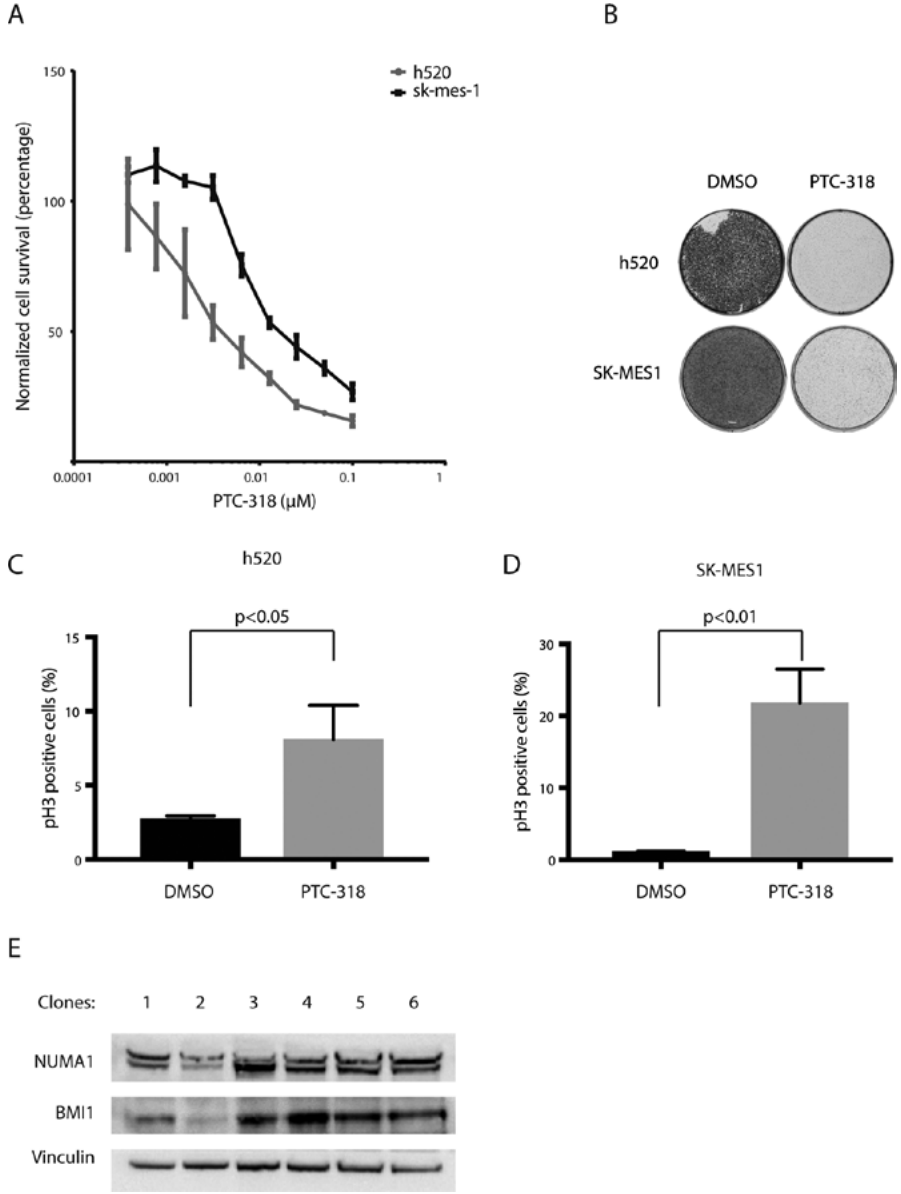


Figure 5. Non-small cell lung cancer cell lines are sensitive to BMI1 inhibition. (A) IC50 measurements of NSCLC cell lines h520, SK-MES-1, and SWI-1573. The x-axis shows the PTC-318 concentration in μM, and the y-axis the percentage of surviving cells relative to non-treated control counterparts. (B) Colony formation of h520 and SK-MES-1 cells treated with either 0.1% DMSO or 40 nM of PTC-318. (C) Percentages of h520 phospho-histone H3 populations after treatment with either 0.1% DMSO or 40 nM PTC-318 determined by flow cytometry quantification. (D) Percentages of h520 phospho-histone H3 populations after treatment with either 0.1% DMSO or 80 nM PTC-318 determined by flow cytometry quantification. (E) Protein expression of NUMA1 and BMI1 in NUMA1-WT clones of an h520 population transfected with Cas9 targeting BMI1, single-cell sorted, and expanded. Error bars represent SD. Student's t-test was performed for statistical testing.

DISCUSSION

Loss of BMI1 has been shown to induce growth arrest and cell death in cancer cells both *in vitro* and *in vivo*[37–39]. However, although BMI1 inhibition leads to impaired cell proliferation in different systems, the exact mechanisms are often context-dependent. Also, while BMI1 expression can cause resistance against different cancer treatment strategies, including chemotherapy or radiation[15–18], little is known about possible resistance-inducing responses to cell death resulting from BMI1 inhibition.

Encouraged by the therapeutic potential of BMI1 inhibition, we assessed the effect of PTC-318, a new and previously unpublished small molecule inhibitor of BMI1, on human cancer cells and evaluated its underlying mechanisms of action. We found that chemical inhibition of BMI1 efficiently caused cell death in human haploid HAP1 cells–a cell line originally derived from chronic myeloid leukemia cells[40] –and in non-small cell lung cancer (NSCLC) cell lines.

A PTC-318 resistance screen on HAP1 cells allowed us to address genetic alterations underlying resistance to BMI1 inhibition, thus expanding our understanding of previously unknown BMI1 mechanisms. Reassuringly, our screen presented significant hits, including genes of the SWI/SNF complex, one of the most studied ATP-dependent chromatin remodeling complexes, known to regulate and antagonize the function of polycomb repressive complexes[41]. Among these hits, we found PBAF-specific genes PBRM1 (BAF180) and ARID2 (BAF200) from the Polycomb-antagonistic SWI/SNF complex. While the finding of the two independent components of the PBAF chromatin-remodeling complex might be expected based on the known dose-dependent antagonistic relationship between Polycomb repressors and SWI/SNF remodelers[41], Nuclear Mitotic Apparatus proteins (*NUMA1*) was an unexpected candidate as the most significant hit in our HAP1 screen. Here, we confirmed that the loss of NUMA1 established by CRISPR–Cas9 could rescue the lethality induced by shRNA-established BMI1 knockdown and PTC-318 inhibition.

Using different genetic approaches, including genetic interference through shRNAs and ectopic BMI1 expression, we could confirm that PTC-318-derived cell death was on-target and caused by BMI1 inhibition. It is worth noticing that knockdown of *BMI1* by shRNA resulted in significant but weaker phenotypic changes in our studied cell lines compared with PTC-318 treatment. These relatively weaker phenotypes might be caused by partial inhibition of BMI1 by the shRNAs compared with PTC-318 treatment. However, we cannot entirely exclude that PTC-318 may have some additional effects beyond BMI1 inhibition.

Apart from the similarities observed between shRNA and small molecule-induced BMI1 inhibition, we also demonstrated that our attempts to establish *BMI1* knockout lines were unsuccessful in NUMA1-WT backgrounds but not in NUMA1-KO backgrounds, confirming BMI1-dependency in multiple cell lines. Interestingly, the only clone surviving BMI1 knockout in the NSCLC cell line h520 expressed lower levels of NUMA1, further underscoring the genetic interactions between NUMA1 and BMI1.

NUMA1 and BMI1 are involved in several shared cellular processes that could suggest possible direct or indirect mechanistic relationships between the two proteins. For example, just as BMI1, NUMA1 is involved in DNA damage repair and homologous recombination[42], cell differentiation[30–32], higher-order chromatin organization[43], and cell cycle progression. The cell cycle properties of NUMA1, particularly in mitosis, were especially interesting to us since previous studies have implicated the EZH2 protein of polycomb repressive complex 2 (PRC2) as a regulator of cell cycle checkpoints, including cyclin-dependent kinase inhibitor p21[44]. According to these findings, loss of EZH2 in cancer cells abrogated cell cycle arrest in G1 and G2/M upon treatment with DNA-damaging agents Adriamycin or etoposide. Importantly, several studies have reported the involvement of BMI1 in mitosis or G2/M phase[45,46]. Wei et al. analyzed the DNA damage repair mechanisms in breast cancer MCF7 cells upon ectopic expression or inhibition of BMI1. They found that while ectopic BMI1 expression resulted in reduced G2/M arrest upon etoposide-treatment, knockdown of *BMI1* had the opposite effect.

In light of our screening results and the reported role of NUMA1 in the cell cycle, we further validated that inhibition of BMI1 in NUMA1 proficient cells resulted in a mitotic arrest followed by cell death in mitosis, a phenotype that could be rescued by inactivation of *NUMA1*. The mitotic arrest observed in our study upon inhibition of BMI1 coincided with increases in CDK1 and Cyclin-B1 protein levels. The findings further support previous studies demonstrating a CDK1-induced switch from mitotic arrest to apoptosis[47]. According to their results, this apoptotic switch is derived from the phosphorylation of Bcl-2/Bax family proteins upon interfering with microtubule formation. In our study, both genetic and drug-induced BMI1 regulation decreased RNA expression of *BCL-2*, a known protein with anti-apoptotic functions[48], suggesting deregulation of both mitosis and cancer-associated survival mechanisms upon BMI1 inhibition. Still, our results from the cell cycle profiling indicate that knockout of NUMA1, while rescuing the lethal phenotype induced by BMI1 inhibition, does not rescue mitotic arrest. The latter observation suggests that NUMA1 regulates the balance between Cyclin B1/CDK1 activity and BCL-2 levels, thereby regulating mitotic fate[49] upon impaired BMI1 expression.

In conclusion, our results investigating the effects of PTC-318 on cancer cell lines indicate that inhibition of the oncogene BMI1 induces cancer cell death, making it a potential approach to treat certain types of cancers–either as a single agent or in combination with other inhibitors. The emergence and testing of new BMI1 inhibitors for the treatment of cancer suggest that BMI1 is a relevant target for cancer therapy. In our study, we unveiled how the inhibition of BMI1 induces cell death through mitotic arrest that was successfully rescued by the depletion of NUMA1 expression. In our experimental set-up, *NUMA1*-mutations were strongly associated with cancer cell survival and may have an essential role – and hence serve as a potential biomarker – in resistance mechanisms in treatments associated with BMI1 inhibition. Although further studies are needed to establish a complete picture of the pathways and mechanisms linked with

BMI1 overexpression and inhibition in cancer, our findings highlight an essential and novel mechanism of BMI1 in cancer cells, which could contribute to the development of effective cancer therapies.

MATERIALS AND METHODS

Cell culture

HAP1 cells have been described previously[50]. They were maintained in IMDM + GlutaMAX (Gibco) supplemented with 10% fetal calf serum (FCS; Sigma-Aldrich) and penicillin–streptomycin (Gibco). NSCLC cell lines h520 and SK-MES-1 (provided by the A. Berns laboratory at the Netherlands Cancer Institute, Amsterdam, the Netherlands) were maintained in DMEM/F12 + GlutaMAX (Gibco) supplemented with 10% FCS and penicillin–streptomycin (Gibco).

Haploid genetic screen

Procedures for the generation of gene-trap retrovirus and HAP1 mutagenesis have been described previously[50]. To select HAP1 variants resistant to BMI1 inhibitor PTC-318, approximatively 10^8 mutagenized HAP1 cells (>90% haploid) were seeded in fourteen T175 cell culture flasks. The cells were exposed to 40 nM of PTC-318 24 hours after seeding and incubated for fourteen days. Surviving HAP1 clones were trypsinized and washed before amplification and analysis of integration sites, as described in[22,51]. In brief, as a first step, insertion sites were amplified in a linear amplification reaction using a biotinylated primer. Products were captured on streptavidin-coated magnetic beads, washed, and subjected to single-stranded DNA linker ligation followed by a second PCR to finalize the products for Illumina sequencing. Subsequently, as described for enrichment mutagenesis screens in Staring et al., 2017, sequence reads were aligned to the human genome (HG19) and uniquely aligning insertion sites were mapped to the genomic coordinates (RefSeq) of non-overlapping protein-encoding gene regions. Gene trap integrations in sense in introns, or regardless of orientation in exons, were considered disruptive. To select genes enriched for mutations after PTC-318 selection, the number of disruptive integrations in each gene was compared to those retrieved in an unselected HAP1 population[22] using a one-sided Fisher's exact test and corrected for multiple testing (Benjamini and Hochberg FDR).

Generation of knockdown, overexpression, and knockout cell lines

For BMI1 knockdown experiments, we used doxycycline-inducible FH1-tUTG-RNAi vectors[52], as described previously[53]. For ectopic expression of BMI1, we used the BMI1-overexpression vector with FUGW vector backbone (FUGW-BMI1; Addgene 21577). Cells in 10-cm plates were transduced using 10 µg of FH1-tUTG-RNAi vector or FUGW-BMI1, 3.5 µg VSV-G, 2.5 µg REV, and 5 µg pRRE in $CaCl_2$ (2.5M). GFP positive cells were sorted by flow cytometry (MoFlo), collected in respective media containing 20% FCS, and subsequently

spun down and plated in media containing 10% FCS. shRNA targeting sequences: shBMI1, 5’-GGAGGAGGTGAAGTATAAA’; shRandom, 5’-ATTCTTACGAAACCCTTAG-3’.
For Cas9-induced knockout, we used SpCas9 and chimeric guide RNA (gRNA) expression plasmid pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene 42230) encoding gRNAs targeting NUMA1 (NuMA-KO) or BMI1 (BMI1-KO). Cells were transfected in 6-well plates with 1.2-1.6 µg of Cas9–gRNA construct together with 10% 1.5 µg of mPB-L3-ERT2.TatRRR-mCherry plasmid, following the Lipofectamine 2000 protocol. mCherry-positive cells were single-cell sorted by flow cytometry (MoFlo) into 96-well plates 48 hours after transfection and incubated for two weeks before expansion. Gene mutations were validated by Sanger sequencing and western blot analysis. gRNA targeting sequences were 5’-GACACTCCACGCCACCCGGG-3’ for NuMA-KO and 5’-AACGTGTATTGTTCGTTACC-3’ for BMI1-KO.

Western blot analysis

Whole-cell extracts pelleted and prepared in RIPA buffer (50 mM Tris, pH 8.0, 50 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, and 0.1% SDS) containing protease inhibitor cocktail (Complete; Roche) and phosphate inhibitors (10 mM Na fluoride final concentration, 1 mM sodium orthovanadate final concentration, and 1 mM NaPPi final concentration). Equal amounts of protein, as determined by a Bio-Rad Protein Assay Dye Reagent on Nanodrop 2000c, were resolved on NuPage-Novex 4–12% Bis-Tris gels (Invitrogen) and transferred onto nitrocellulose membranes (0.2 m; Whatman). Membranes were blocked in phosphate-buffered saline (PBS) with 0.1% Tween-20 (PBST) and 5% BSA for 1 h, incubated with primary antibodies in PBST 1% BSA overnight at 4°C, and incubated with secondary antibodies coupled to HRP for 45 min in PBST 1% BSA at room temperature. Membranes were imaged on a BioRad ChemiDoc XRS+. The following antibodies were used for western blot analyses: anti-BMI1 D20B7 (Cell Signalling), anti-GFP (Abcam, ab6556), anti-NuMA (Thermo Fisher, PA5-22285), anti-CDK1 (Bethyl, A303-664A), Cyclin B1 (GNS1; Santa Cruz, sc-245).

Gene expression

Total RNA was isolated with the ReliaPrep™ RNA Cell Miniprep System (Promega) according to the manufacturer’s instructions. RNA quantity and quality were assessed using a Nanodrop 2000c (Thermo Scientific). Primers details are available upon request.

Clonogenic assays

To test PTC-318 lethality, 400,000 HAP1 cells/well or 100,000 NSCLC cells/well were seeded in 2 ml/well in 6-well plates and incubated overnight at 37°C. Dimethyl sulfoxide (DMSO) diluted PTC-318 (80 µM) was further diluted with the corresponding medium to obtain the desired concentrations. For control cells, DMSO was diluted to match the DMSO concentrations of PTC-318 dilutions (below 0.1%). PTC-318 or DMSO dilutions were distributed to the cells (2 ml/well) and incubated for one week at 37°C. Cells were next washed with PBS and incubated with 0.5 ml of 0.1% crystal violet (50% methanol) for 10 min at room temperature. Imaging of the wells was performed with GelCount (Oxford Optronix).

FH1t-BMI1-induced lethality was assessed by plating 100,000 HAP1 cells/well in 6-well plates in cell medium with or without doxycycline. Doxycycline treatment was performed every day until the end of the experiment (96 hours). For time-point experiments, all evaluated time-points were plated at 0 hours and assessed 96 hours after plating to maintain similar confluency. The start of doxycycline treatment depended on conditions tested (96 h doxycycline: 0 h after seeding, 72 h: 24 hours after seeding, 48 h: 48 hours after seeding, and 24 h: 72 hours after seeding). Cells were washed with phosphate-buffered saline (PBS) and incubated with 0.5 ml crystal violet for 10 min at room temperature. Imaging of the wells was performed with a GelCount.

Cell viability assay

To assess cell viability upon PTC-318 treatment, 40,000 HAP1 cells/well or 20,000 HAP1 cells/well were plated in 24-well plates (500 µl/well) and incubated for 24 hours at 37°C. Next, the cells were treated with respective concentrations of PTC-318 (as described above) and incubated for 48 hours. Alamar blue (10X; Invitrogen) was diluted 1:10 in the corresponding cell medium and distributed to cells. Negative control cells received 3 µl of 20% sodium dodecyl sulfate (SDS). Cell viability was determined by measuring the relative fluorescence 4-6 hours after addition of Alamar blue using the HP D300 Digital Dispenser.

To assess viability upon BMI1-KD, 2,000 cells were plated in 24-well plates with or without doxycycline (500 µl/well) and incubated for 24 hours at 37°C. The doxycycline was refreshed every day for one week before adding the Alamar blue as previously described.

Flow cytometry analysis

For PTC-318, 200,000 cells were plated, treated with PTC-318, and incubated for 48 hours before collection. For shBMI1, 100,000 cells were plated and incubated for 96 hours. Doxycycline was refreshed every day. After corresponding incubation times, equal amounts of cells (500,000-2,000,000) were collected after trypsinization and thereafter centrifuged, fixed with 70% ethanol, and stored at -20°C until staining. For staining, cells were centrifuged for 3 minutes at 1,500 rpm and then washed in PBS. Cells were centrifuged again at 1,500 for 3 minutes, and pellets were resuspended in 0.25% Triton X-100 (Sigma), transferred to 1.5 ml tubes, and incubated on ice for 15 minutes for permeabilization. Cells were centrifuged at 5,000 g for three minutes, resuspended in PBS containing 1% BSA and 1:100 mouse anti-phospho-Histone H3 (Cell signaling, 9701), and incubated for 1 hour at room temperature. After washing, cells were resuspended in PBS containing 1% BSA and 1:300 Alexa 647 conjugated donkey anti-mouse (Thermofisher, A21238) and incubated for 30 min at room temperature and in the dark. Cells were after that washed, resuspended in 150 µl PBS containing 10 µg/ml DNase-free RNase A, and incubated for 30 min at 37°C. Prepared cells were stored at -20°C or 4°C until analysis on a flow cytometer (LSR II). Propidium iodide (1.5 µg) was added to the samples right before assessment.

Live cell imaging

Cells were plated on LabTek dishes, and two hours before imaging, the media was changed to Leibovitz L15 CO2-independent cell culture medium (Gibco). Mitotic progression was followed by adding SiR-DNA (Spirochrome) and drugs to the cells, that were imaged as previously described[54]. In the case of shRNA-mediated knockdown of BMI1 (shBMI1), doxycycline was added to the cells 24, 48, or 72 hours prior to SiR-DNA addition. Cells were imaged every 15 minutes in a heated chamber of 37 °C using a 20x NA 0.95 air objective on an IX71 microscope (Olympus), controlled by SoftWoRx 6.0 software (Applied Precision).

ACKNOWLEDGMENTS

We thank the NKI Genomics Core Facility and the NKI Flow Cytometry Facility for their technical support. We also want to thank Vincent Blomen, Thijn Brummelkamp, Lorenzo Bombardelli, Anke Sparmann, and Bas Tolhuis for the valuable help, discussions, and/or constructive criticism of the manuscript.

REFERENCES

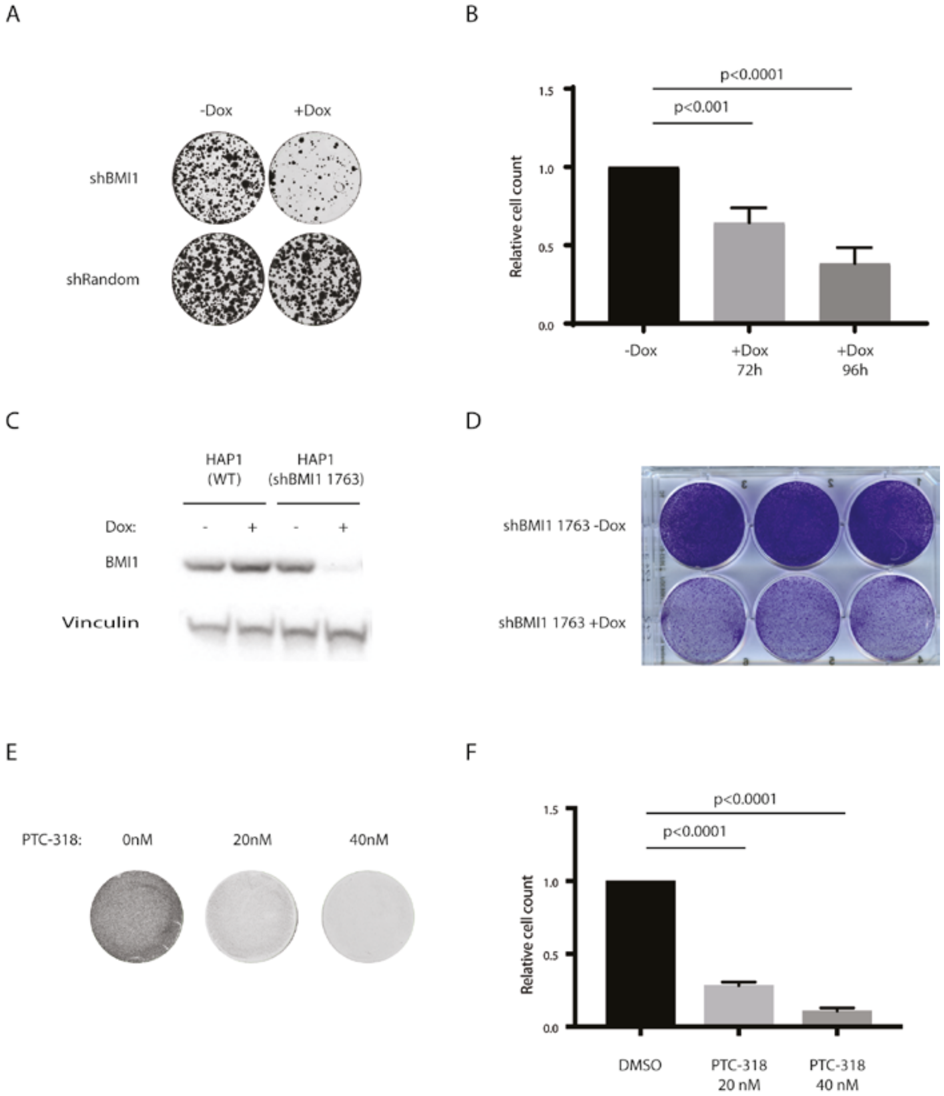
1. Pietersen, A. M. *et al.* Bmi1 Regulates Stem Cells and Proliferation and Differentiation of Committed Cells in Mammary Epithelium. *Curr. Biol.* (2008). doi:10.1016/j.cub.2008.06.070
2. Godlewski, J. *et al.* Targeting of the Bmi-1 oncogene/stem cell renewal factor by MicroRNA-128 inhibits glioma proliferation and self-renewal. *Cancer Res.* (2008). doi:10.1158/0008-5472.CAN-08-2629
3. Jacobs, J. J. L. *et al.* Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. *Genes Dev.* (1999). doi:10.1101/gad.13.20.2678
4. Haupt, Y., Alexander, W. S., Barri, G., Peter Klinken, S. & Adams, J. M. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in Eμ-myc transgenic mice. *Cell* (1991). doi:10.1016/0092-8674(91)90383-A
5. van Lohuizen, M. *et al.* Identification of cooperating oncogenes in Eμ-myc transgenic mice by provirus tagging. *Cell* (1991). doi:10.1016/0092-8674(91)90382-9
6. Biehs, B. *et al.* BMI1 represses Ink4a/Arf and Hox genes to regulate stem cells in the rodent incisor. *Nat. Cell Biol.* (2013). doi:10.1038/ncb2766
7. Lukacs, R. U., Memarzadeh, S., Wu, H. & Witte, O. N. Bmi-1 is a crucial regulator of prostate stem cell self-renewal and malignant transformation. *Cell Stem Cell* (2010). doi:10.1016/j.stem.2010.11.013
8. Abdouh, M. *et al.* BMI1 sustains human glioblastoma multiforme stem cell renewal. *J. Neurosci.* (2009). doi:10.1523/JNEUROSCI.0968-09.2009
9. Kojima, K. *et al.* Downregulation of BMI-1 By the Small Molecule PTC596 Induces Mitochondrial Apoptosis in Mantle Cell Lymphoma. *Blood* **130**, 1263 (2017).
10. Wang, Q. *et al.* Pharmacological inhibition of Bmi1 by PTC-209 impaired tumor growth in head neck squamous cell carcinoma. *Cancer Cell Int.* (2017). doi:10.1186/s12935-017-0481-z
11. Kreso, A. *et al.* Self-renewal as a therapeutic target in human colorectal cancer. *Nat. Med.* (2014). doi:10.1038/nm.3418
12. Paranjape, A. N. *et al.* Bmi1 regulates self-renewal and epithelial to mesenchymal transition in breast cancer cells through Nanog. *BMC Cancer* (2014). doi:10.1186/1471-2407-14-785
13. Maeda, A. *et al.* Targeting of BMI-1 expression by the novel small molecule PTC596 in mantle cell lymphoma. *Oncotarget* (2018). doi:10.18632/oncotarget.25558
14. Yin, T. *et al.* Bmi1 inhibition enhances the sensitivity of pancreatic cancer cells to gemcitabine. *Oncotarget* (2016). doi:10.18632/oncotarget.9293
15. Griffith, J. *et al.* Silencing BMI1 radiosensitizes human breast cancer cells by inducing DNA damage and autophagy. *Oncol. Rep.* (2017). doi:10.3892/or.2017.5478
16. Ojo, D. *et al.* Polycomb complex protein BMI1 confers resistance to tamoxifen in estrogen receptor positive breast cancer. *Cancer Lett.* (2018). doi:10.1016/j.canlet.2018.03.048
17. Wu, Z. *et al.* Overexpression of BMI-1 promotes cell growth and resistance to cisplatin treatment in osteosarcoma. *PLoS One* (2011). doi:10.1371/journal.pone.0014648
18. Yin, T. *et al.* Bmi-1 promotes the chemoresistance, invasion and tumorigenesis of pancreatic cancer cells. *Chemotherapy* (2012). doi:10.1159/000334103
19. Bhattacharya, R. *et al.* MiR-15a and MiR-16 control Bmi-1 expression in ovarian cancer. *Cancer Res.* (2009). doi:10.1158/0008-5472.CAN-09-2552
20. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science (80-.).* (2009). doi:10.1126/science.1178955
21. Jae, L. T. *et al.* Lassa virus entry requires a trigger-induced receptor switch. *Science (80-.).* (2014). doi:10.1126/science.1252480
22. Staring, J. *et al.* PLA2G16 represents a switch between entry and clearance of Picornaviridae. *Nature* (2017). doi:10.1038/nature21032
23. Mezzadra, R. *et al.* SLFN11 can sensitize tumor cells towards IFN-γ-mediated T cell killing. *PLoS One* (2019). doi:10.1371/journal.pone.0212053

24. Lee, C. C., Carette, J. E., Brummelkamp, T. R. & Ploegh, H. L. A Reporter Screen in a Human Haploid Cell Line Identifies CYLD as a Constitutive Inhibitor of NF-κB. *PLoS One* (2013). doi:10.1371/journal.pone.0070339
25. Planells-Cases, R. *et al.* Subunit composition of VRAC channels determines substrate specificity and cellular resistance to P t-based anti-cancer drugs . *EMBO J.* (2015). doi:10.15252/embj.201592409
26. Compton, D. A. & Luo, C. Mutation of the predicted p34cdc2 phosphorylation sites in NuMA impair the assembly of the mitotic spindle and block mitosis. *J. Cell Sci.* (1995).
27. Gehmlich, K., Haren, L. & Merdes, A. Cyclin B degradation leads ot NuMA release from dynein/dynactin and from spindle poles. *EMBO Rep.* (2004). doi:10.1038/sj.embor.7400046
28. Radulescu, A. E. & Cleveland, D. W. NuMA after 30 years: The matrix revisited. *Trends in Cell Biology* (2010). doi:10.1016/j.tcb.2010.01.003
29. Ohata, H. *et al.* NuMA Is Required for the Selective Induction of p53 Target Genes. *Mol. Cell. Biol.* (2013). doi:10.1128/mcb.01221-12
30. Siller, K. H. & Doe, C. Q. Spindle orientation during asymmetric cell division. *Nature Cell Biology* (2009). doi:10.1038/ncb0409-365
31. Du, Q., Stukenberg, P. T. & Macara, I. G. A mammalian partner of inscuteable binds NuMA and regulates mitotic spindle organization. *Nat. Cell Biol.* (2001). doi:10.1038/ncb1201-1069
32. Lechler, T. & Fuchs, E. Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature* (2005). doi:10.1038/nature03922
33. Gascoigne, K. E. & Taylor, S. S. How do anti-mitotic drugs kill cancer cells? *Journal of Cell Science* (2009). doi:10.1242/jcs.039719
34. Zhang, X., Tian, T., Sun, W., Liu, C. & Fang, X. Bmi-1 overexpression as an efficient prognostic marker in patients with nonsmall cell lung cancer. *Medicine (United States)* (2017). doi:10.1097/MD.0000000000007346
35. Xiong, D. *et al.* Bmi-1 expression modulates non-small cell lung cancer progression. *Cancer Biol. Ther.* (2015). doi:10.1080/15384047.2015.1026472
36. Vrzalikova, K. *et al.* Prognostic value of Bmi-1 oncoprotein expression in NSCLC patients: A tissue microarray study. *J. Cancer Res. Clin. Oncol.* (2008). doi:10.1007/s00432-008-0361-y
37. Jacobs, J. L., Kieboom, K., Marino, S., DePinho, R. A. & Van Lohuizen, M. The oncogene and Polycombgroup gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature* (1999). doi:10.1038/16476
38. Bruggeman, S. W. M. *et al.* Ink4a and Arf differentially affect cell proliferation and neural stem cell self-renewal in Bmi1-deficient mice. *Genes Dev.* (2005). doi:10.1101/gad.1299305
39. Liu, L., Andrews, L. G. & Tollefsbol, T. O. Loss of the human polycomb group protein BMI1 promotes cancer-specific cell death. *Oncogene* (2006). doi:10.1038/sj.onc.1209454
40. Carette, J. E. *et al.* Generation of iPSCs from cultured human malignant cells. *Blood* (2010). doi:10.1182/blood-2009-07-231845
41. Wilson, B. G. *et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* (2010). doi:10.1016/j.ccr.2010.09.006
42. Vidi, P. A. *et al.* NuMA promotes homologous recombination repair by regulating the accumulation of the ISWI ATPase SNF2h at DNA breaks. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gku296
43. Abad, P. C. *et al.* NuMA influences higher order chromatin organization in human mammary epithelium. *Mol. Biol. Cell* (2007). doi:10.1091/mbc.E06-06-0551
44. Wu, Z. *et al.* Polycomb protein EZH2 regulates cancer cell fate decision in response to DNA damage. *Cell Death Differ.* (2011). doi:10.1038/cdd.2011.48
45. Wei, F. *et al.* BMI1 attenuates etoposide-induced G2/M checkpoints via reducing ATM activation. *Oncogene* (2015). doi:10.1038/onc.2014.235
46. Xu, C. R. *et al.* Bmi1 functions as an oncogene independent of Ink4a/Arf repression in hepatic

carcinogenesis. *Mol. Cancer Res.* (2009). doi:10.1158/1541-7786.MCR-09-0333

47. Zhou, L., Cai, X., Han, X., Xu, N. & Chang, D. C. CDK1 switches mitotic arrest to apoptosis by phosphorylating Bcl-2/Bax family proteins during treatment with microtubule interfering agents. *Cell Biol. Int.* (2014). doi:10.1002/cbin.10259
48. Yin, X. M., Oltvai, Z. N. & Korsmeyer, S. J. BH1 and BH2 domains of Bcl-2 are required for inhibition of apoptosis and heterodimerization with Bax. *Nature* (1994). doi:10.1038/369321a0
49. Ban, K. H. *et al.* The END Network Couples Spindle Pole Assembly to Inhibition of the Anaphase-Promoting Complex/ Cyclosome in Early Mitosis. *Dev. Cell* (2007). doi:10.1016/j.devcel.2007.04.017
50. Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* (2011). doi:10.1038/nature10348
51. Jae, L. T. *et al.* Deciphering the glycosylome of dystroglycanopathies using haploid screens for Lassa virus entry. *Science (80-.).* (2013). doi:10.1126/science.1233675
52. Herold, M. J., Van Den Brandt, J., Seibler, J. & Reichardt, H. M. Inducible and reversible gene silencing by stable integration of an shRNA-encoding lentivirus in transgenic rats. *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0806213105
53. Michalak, E. M. *et al.* Polycomb group gene Ezh2 regulates mammary gland morphogenesis and maintains the luminal progenitor pool. *Stem Cells* (2013). doi:10.1002/stem.1437
54. Maia, A. R. R. *et al.* Inhibition of the spindle assembly checkpoint kinase TTK enhances the efficacy of docetaxel in a triple-negative breast cancer model. *Ann. Oncol.* (2015). doi:10.1093/annonc/mdv293

SUPPLEMENTARY MATERIAL



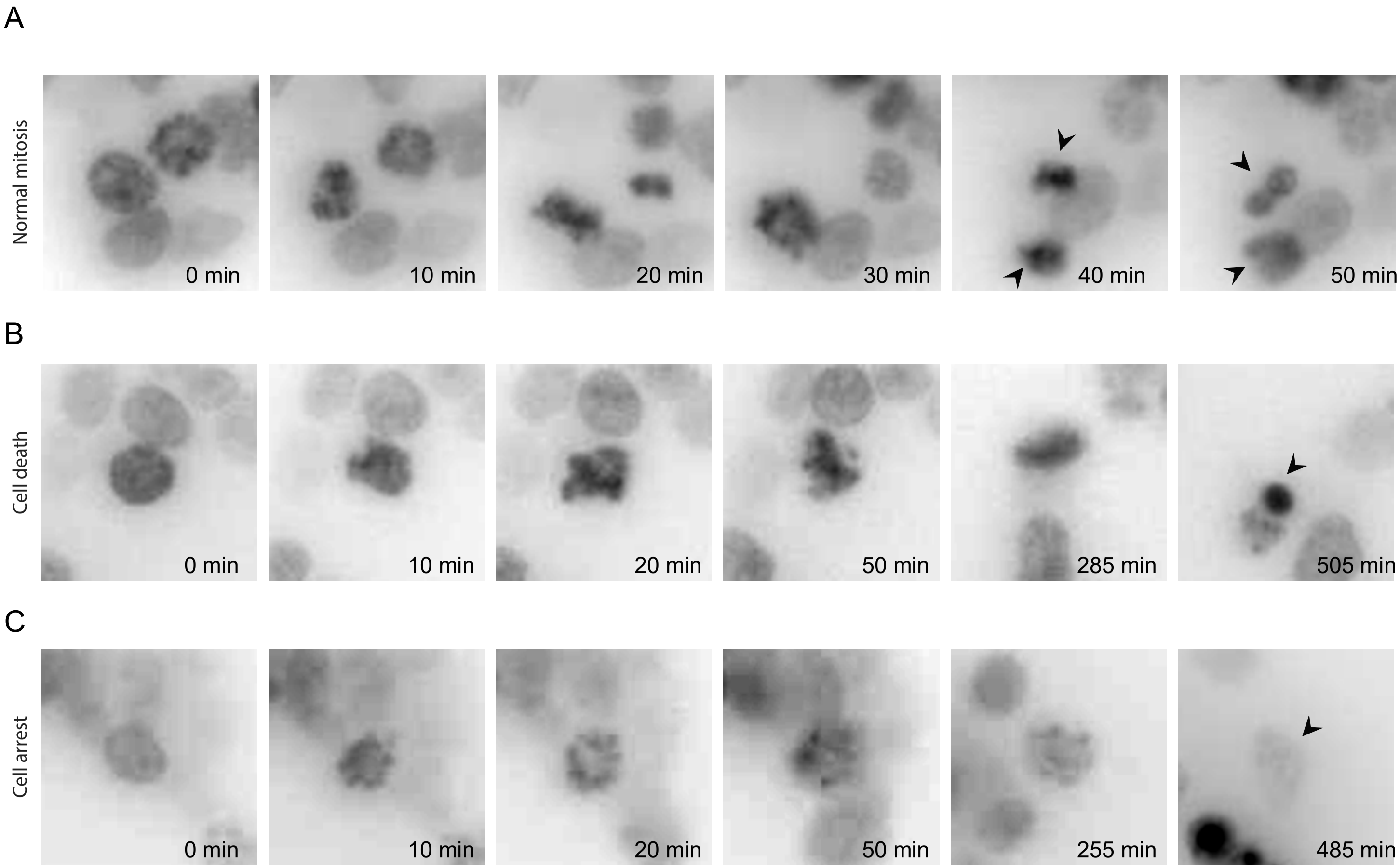
Supplementary Figure 1. Inhibition of BMI1 results in reduced HAP1 cell viability. (A) Cell survival upon BMI1 knockdown was confirmed in HAP1 cells transduced with shBMI1 through a colony formation assay one week after plating HAP1 cells transduced with either shRandom or shBMI1 and treated with doxycycline or (B) through relative cell counts of HAP1 cells transduced with shBMI1 treated with doxycycline for 72, and 96 hours after plating or without doxycycline (-Dox). (C) Cell survival of HAP1 cells treated with PTC-318 through colony formation assay one week after treatment with DMSO (0.1%) or PTC-318 (20 or 40 nM) and (D) relative cell counts 48 hours after treatment with DMSO (0.1%) or PTC-318 (20 or 40 nM). Error bars represent SD. Student's t-test was performed for statistical testing.

Supplementary Table 1. List of the 100 most significant enriched genes after HAP1 screen exposing the cells with 40 nM PTC-318.

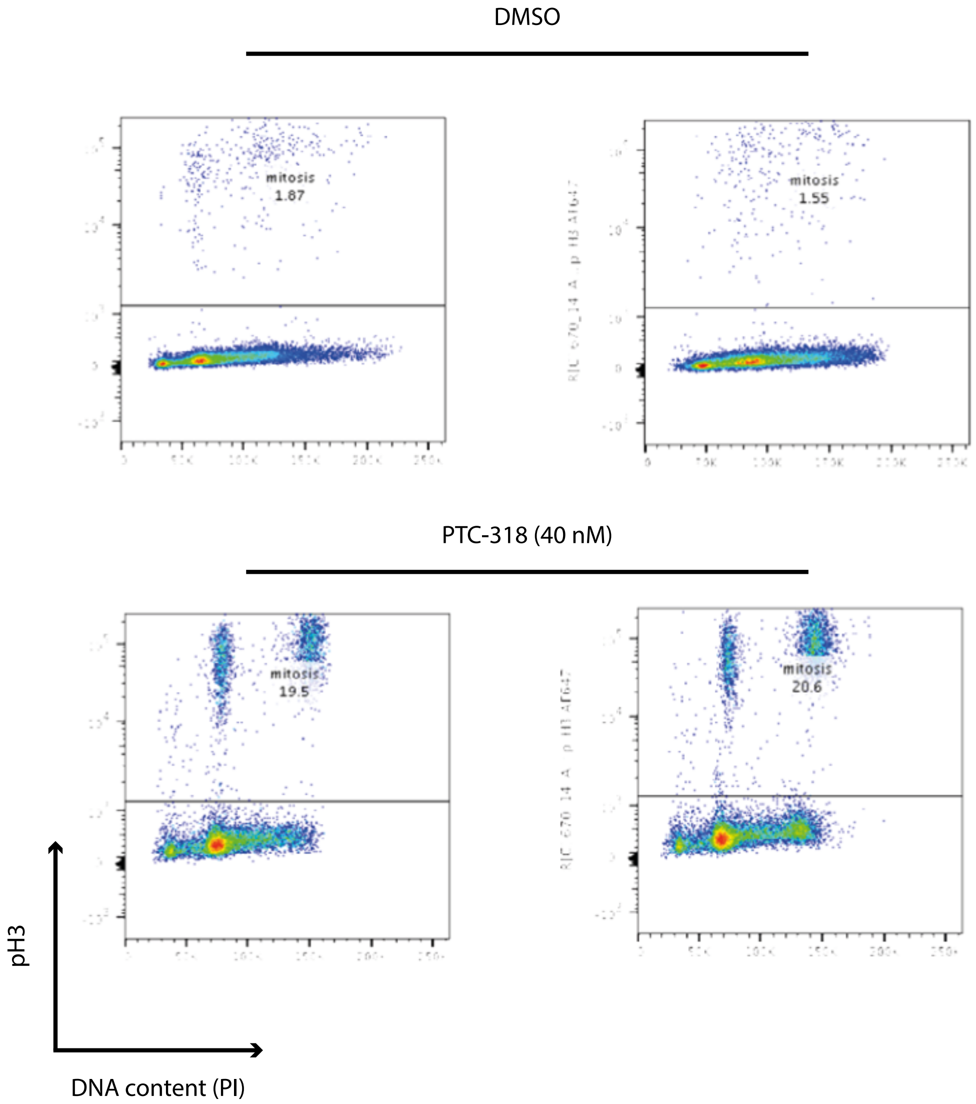
#	Gene	Number of integrations	Total number of integrations - integrations in gene	FDR-corrected Pvalue	#	Gene	Number of integrations	Total number of integrations - integrations in gene	FDR-corrected Pvalue
1	NUMA1	570	2163758	1,61E-304	51	MPPED2	1331	2162997	3,79E-25
2	WNK1	246	2164082	1,61E-304	52	GRIA3	832	2163496	2,74E-24
3	CSMD1	502	2163826	6,32E-240	53	CDKAL1	2910	2161418	2,83E-24
4	ANKRD30A	18	2164310	6,63E-150	54	MAP1B	1308	2163020	3,35E-24
5	ATP8A2	1973	2162355	4,44E-127	55	BST2	35	2164293	1,80E-23
6	PBRM1	1101	2163227	1,70E-115	56	DCUN1D3	290	2164038	8,11E-23
7	HN1	203	2164125	1,21E-113	57	AGPS	1082	2163246	1,61E-22
8	GPC3	2417	2161911	4,03E-88	58	FRMD6	1003	2163325	9,75E-22
9	ROBO1	3646	2160682	1,54E-84	59	PCYT1B	925	2163403	3,77E-21
10	TANC2	6235	2158093	5,88E-81	60	CD36	434	2163894	7,62E-21
11	PDSS2	1665	2162663	7,21E-77	61	ADGRG4	701	2163627	1,84E-20
12	CDH12	535	2163793	1,22E-65	62	SLIT2	2685	2161643	2,30E-20
13	TUBB4B	6	2164322	6,79E-65	63	TOX	1986	2162342	2,83E-20
14	LRPPRC	285	2164043	1,07E-62	64	ARHGEF6	319	2164009	4,60E-20
15	IL1RAPL1	5470	2158858	2,92E-56	65	CELF2	354	2163974	1,30E-19
16	MYBPC1	1024	2163304	2,05E-55	66	BMF	867	2163461	2,03E-19
17	TMC1	4381	2159947	1,99E-52	67	FAM124A	433	2163895	3,98E-19
18	PCDH9	1729	2162599	7,34E-52	68	ADGRV1	2783	2161545	5,33E-19
19	ZSWIM6	842	2163486	1,06E-50	69	MAGED1	287	2164041	6,17E-19
20	UNC13C	171	2164157	3,46E-49	70	MACROD2	1370	2162958	9,53E-19
21	MTF2	203	2164125	7,47E-49	71	RPL10	45	2164283	1,24E-18
22	DLG2	5625	2158703	6,46E-46	72	CRAMP1	173	2164155	2,78E-18
23	HCFC1	28	2164300	3,25E-44	73	LRRC4C	841	2163487	3,68E-18
24	CADM2	2853	2161475	2,36E-41	74	PDS5A	350	2163978	4,27E-18
25	FOXO3	994	2163334	7,05E-41	75	LDB2	1717	2162611	4,60E-18

Supplementary Table 1. (continued)

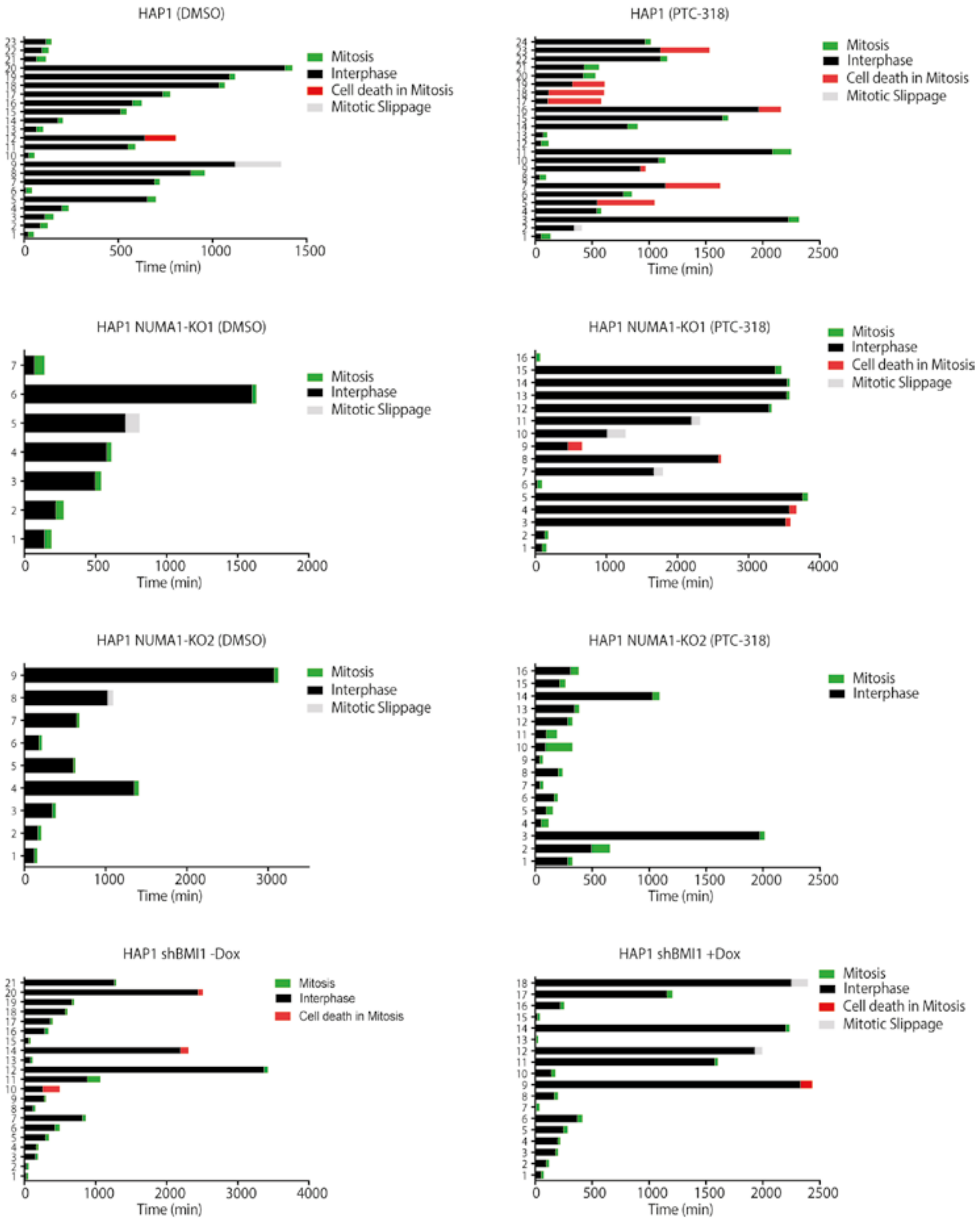
#	Gene	Number of integrations	Total number of integrations - integrations in gene	FDR-corrected Pvalue	#	Gene	Number of integrations	Total number of integrations - integrations in gene	FDR-corrected Pvalue
26	MRPS28	484	2163844	3,89E-40	76	ARID2	270	2164058	5,53E-18
27	CRADD	1678	2162650	1,70E-38	77	CDC25A	252	2164076	8,22E-18
28	ZNRF2	1784	2162544	6,16E-38	78	ATP8A1	981	2163347	1,17E-17
29	ZBTB20	5892	2158436	9,05E-38	79	MAP4	936	2163392	1,19E-17
30	LGALSL	377	2163951	1,27E-36	80	GTF2I	438	2163890	1,37E-17
31	GNAT3	726	2163602	1,51E-35	81	HNRNPD	192	2164136	1,96E-17
32	LRFN5	40	2164288	2,25E-35	82	XKR3	28	2164300	2,11E-17
33	NAV3	1338	2162990	6,04E-35	83	SHOX2	50	2164278	2,30E-17
34	MAP4K3	1736	2162592	4,17E-34	84	HDAC9	4481	2159847	2,56E-17
35	CYP7B1	525	2163803	1,59E-31	85	CNTNAP5	179	2164149	9,04E-17
36	BAZ1B	541	2163787	2,68E-30	86	SKIDA1	743	2163585	1,10E-16
37	SGF29	128	2164200	3,90E-30	87	CNTNAP4	455	2163873	1,42E-16
38	CNOT2	87	2164241	4,16E-30	88	EYS	1420	2162908	1,82E-16
39	ANGPT1	878	2163450	4,70E-30	89	RBMS3	1075	2163253	2,44E-16
40	POU6F2	3142	2161186	6,17E-30	90	ADAMTS5	153	2164175	3,02E-16
41	LIN28B	470	2163858	2,33E-29	91	DYM	3197	2161131	4,98E-16
42	NONO	111	2164217	3,13E-28	92	FOXM1	140	2164188	5,13E-16
43	MGAT4C	916	2163412	3,56E-28	93	CDK1	20	2164308	5,87E-16
44	MAD1L1	447	2163881	5,48E-28	94	SGCZ	574	2163754	5,87E-16
45	NLK	1627	2162701	5,59E-28	95	SERF2	506	2163822	1,27E-15
46	ISL1	9	2164319	5,93E-27	96	RAB30	312	2164016	2,62E-15
47	CCDC91	1656	2162672	1,56E-26	97	ADAMTSL1	1797	2162531	3,63E-15
48	CHN1	2443	2161885	3,48E-26	98	ABCA13	1883	2162445	3,63E-15
49	NARS2	896	2163432	6,42E-26	99	KY	393	2163935	3,80E-15
50	FBXL7	1842	2162486	9,38E-26	100	TTC28	4413	2159915	4,71E-15



Supplementary Figure 2. Mitotic arrest of HAP1 cells upon treatment with 40 nM PTC-318. Flow cytometry plot comparing the proportion of cells in mitosis between HAP1 cells treated with DMSO and PTC-318. Technical replicates.



Supplementary Figure 3. Representative picture of live-cell imaging using SiR-DNA. (A) Normal mitotic progression, (B) death in mitosis, and (C) slippage.



Supplementary Figure 4. Live-cell imaging of HAP1 clones upon BMI1 inhibition. Quantification of live-cell imaging data showing the times of individual cells. The upper three rows show cells treated with either DMSO (0.1%) or PTC-318 (20 nM), while the lower row shows HAP1 cells transduced with shBMI1 untreated (-Dox) or treated (+Dox) with doxycycline. The respective y-axes depict the individual clones.

Chapter 5

BAZ1B IS A POSSIBLE REGULATOR OF H3K27me3 IN AN INVERSE CORRELATION BETWEEN PRC1 AND PRC2

Santiago Gisler[1], Jawahar Kopparam[1],
Ferhat Alkan[2], Maarten van Lohuizen[1]

[1] Division of Molecular Genetics, Oncode and The Netherlands Cancer Institute,
Amsterdam, The Netherlands
[2] Division of Oncogenomics, Oncode and The Netherlands Cancer Institute,
Amsterdam, The Netherlands

ABSTRACT

Polycomb Group (PcG) proteins are well-known regulators of cell fate in development and repressors of tumor suppressor genes in cancer. In the canonical PcG pathway, polycomb repressive complex (PRC)2 proteins catalyze the trimethylation of lysine residue 27 of histone H3 (H3K27me3), which is recognized by PRC1, ultimately triggering condensation of the surrounding chromatin and gene repression. However, previous studies have demonstrated that specific PcG proteins can act as tumor suppressors in some cancers, suggesting that the canonical functions may not apply in certain contexts. Here, we present preliminary data suggesting an inverse correlation between PRC1 and PRC2 in three cell lines. We present results from different types of large-scale screens in human haploid HAP1 cells presenting recurrent enrichment of *BAZ1B* disruptions in different experimental settings, supporting the observed inverse correlation. The findings suggest that the product of *BAZ1B*, Williams syndrome transcription factor (WSTF), may directly or indirectly regulate the negative feedback between PRC1 and PRC2 in specific cancer cells, providing us with new information about their distinct, context-dependent functions.

INTRODUCTION

While being critical for controlling cell fate decisions during development, Polycomb group (PcG) proteins are found elevated in several cancer types and associated with poor cancer prognosis[1]. The canonical gene repressive pathway of the PcG proteins involves the two complexes: polycomb repressive complex (PRC) 1, which contains PCGF, CBX, PHC, and RING1, and PRC2, with the core subunits EED, SUZ12, and the histone methyltransferase enzymes EZH1/2. Canonically, PcG-mediated gene repression starts with the recruitment of PRC2 to the promoter of a target gene. Once bound, the catalytic subunits of PRC2, EZH2 (or EZH1) methylates lysine residue 27 on histone H3 (H3K27), generating H3K27me3, which facilitates transcriptional repression[2–5]. As the prototypical repressive mark of PRC2, H3K27me3 becomes the recognition site for PRC1. In essence, the chromodomain-containing CBX family of proteins recognizes and binds H3K27me3, which induces transcriptional silencing through E3 ubiquitin-protein ligase RING1-mediated histone H2A ubiquitination (H2AK119Ub) and condensation of the chromatin[6–10]. Transcriptional silencing by these PcG complexes is associated with repression of tumor-suppressor genes, among other the loci CDKN2A and CDKN2B that encode p14 (ARF), p15 (INK4B), and p16 (INK4A)[11–13].

In contrast to the canonical PcG-mediated gene silencing, an increasing number of studies have discovered functions of non-canonical PRC1 configurations that act independently of CBX proteins, PRC2, and H3K27me3. The identification of additional non-canonical PRC1 complexes has added complexity to the PcG-mediated gene regulation. For example, the configuration of some variant PRC1 complexes includes the protein RYBP, which is mutually exclusive with CBX proteins and has different chromatin activity and gene-target selectivity compared with the canonical PRC1 complex. These non-canonical complexes mediate H2A ubiquitination independently from H3K27me3 and, therefore, are also independent of PRC2 proteins, such as EED or EZH2. Recent findings have modeled a pathway in which these variant PRC1 complexes can recognize unmethylated CpG islands and ubiquitinate proximal H2AK119, which, consequently, promotes H3K27me3 through the recruitment of PRC2 to the target site[14,15].

B lymphoma Mo-MLV insertion region 1 homolog (BMI1) was initially identified as an oncogene that cooperates with c-MYC[16,17]. Overexpression of BMI1 is associated with the maintenance and progression of several cancer types, and BMI1 inhibition results in growth reduction and cancer cell death[18–20]. Similarly, overexpression of PRC2 proteins or mutations in their genes (in particular *EZH2*) have been implicated in cell proliferation, cell invasion, and metastasis[21–24]. However, although the PcG protein complexes have traditionally been associated mainly with gene silencing and oncogenic functions, recent studies also highlight their involvement in transcriptional activation and tumor suppression[25–28]. For example, BMI1 overexpression is associated with increased overall survival in patients with breast cancer, melanoma, and endometrial carcinomas[29–32]. Other non-canonical PRC1 complexes, such as the Kdm2–PRC1.1 complex, can act as transcriptional activators found on promoters of active genes or as tumor suppressors, depending on the context[33,34]. Furthermore, Eed functions as a context-dependent tumor suppressor in mouse KRAS-driven non-small cell lung cancer[35].

With these opposing functions in mind, as well as the range of subunit combinations PRCs can adopt, it becomes evident that PcG-protein functions, such as transcriptional regulation, depends on the cell type and genetic context. Consequently, the different contexts may dictate the balance between negative and positive regulation induced by PcG complexes and their associated components. Although being increasingly attractive for clinical applications, the development of PcG-mediated targeted therapies for cancer patients requires a better understanding of the contextual behaviors of PcG proteins to establish a proper anti-cancer effect.

Genetic approaches have had an enormous impact on the discovery of cancer drug targets and have propelled the identification of gene interactions that affect their efficacy. Here, we compared the results of three different genome-wide screen approaches in mutagenized human haploid HAP1 cells to evaluate the regulation of PcG protein expression. The pilot study includes data from an enrichment screen performed on cells treated with BMI1 inhibitor AB057609107 (PTC-318), a synthetic lethality (depletion) screen in EED knockout cells, and data from a phenotype screen in cells with low or high levels of H3K27me3. By combining these data, we were able to identify common hits with potential involvement in a negative feedback regulation between PRC1 and PRC2. We demonstrate how *BAZ1B*, encoding Williams syndrome transcription factor (WSTF), is enriched in all the screens, suggesting its role as a tumor-suppressor upon BMI1 inhibition and positive regulator of H3K27me3. Although preliminary and correlational, we show how combining different haploid genetic screen approaches can extract large-scale, valuable data for future validation and to decipher context-dependent functions of PcG proteins, as well as other epigenetic regulators.

RESULTS

Downregulation of PRC2 increases BMI1 levels and decreases levels of tumor-suppressors *CDKN2A* and *CDKN1A*

We had previously observed a recurring inverse correlation between PRC1 protein BMI1 and PRC2 protein EZH2 in specific cellular contexts. Specifically, we had found that inhibition of EZH2 with EZH2-selective methyltransferase inhibitor GSK2816126A (GSK-126) was not lethal to specific cell lines and that it resulted in increased levels of BMI1 protein. To validate these observations, we investigated expression profiles upon inhibition of specific PcG proteins in three cancer cell lines, namely non-small cell lung cancer (NSCLC) cell line h520, mouse small cell lung cancer (SCLC) cell line C2C, and the chronic myeloid leukemia (CML)-derived haploid cell line HAP1. As previously observed in our lab, as well as in other studies[36,37], low micromolar concentrations of GSK126 reduced the levels of H3K27me3 **(Figs 1 A and B)**. Interestingly, we also observed that the same GSK-126 concentrations also resulted in increased levels of BMI1.

We decided to use a HAP1 cell line depleted for embryonic ectoderm development (EED) to identify synthetic lethal partners of EED that can explain this inverse correlation. EED is a scaffolding protein and crucial for proper PRC2 complex formation. By knocking out EED, we could evaluate the properties of the PRC2 complex and how it relates to BMI1 levels, potentially with a stronger phenotype compared with EZH2 inhibition since ablation of EED removes PRC2 from chromatin while EZH2 inhibition affects its catalytic activity. To understand the inverse correlation between PRC1 and PRC2, we examined the gene expression and protein profiles of HAP1 cells upon loss of EED (ΔEED). The ΔEED cell line was established using CRISPR–Cas9 targeting the fourth exon of the gene. Knockout of *EED* resulted in reduced levels of EZH2 and its catalytic mark H3K27me3, as well as increased levels of the active enhancer mark H3K27Ac **(Fig 1C)**. The same cell line also showed reduced *EED* transcript levels by approximately half of the wild-type population, possibly resulting from nonsense-mediated decay from aberrant *EED* transcription[38] **(Fig 1D)**. However, compared with wild-type HAP1 cells, ΔEED cells had a more than 4-fold increased expression of *BMI1*, similar to the observed BMI1 upregulation in the lung cancer cell lines upon GSK126 inhibition.

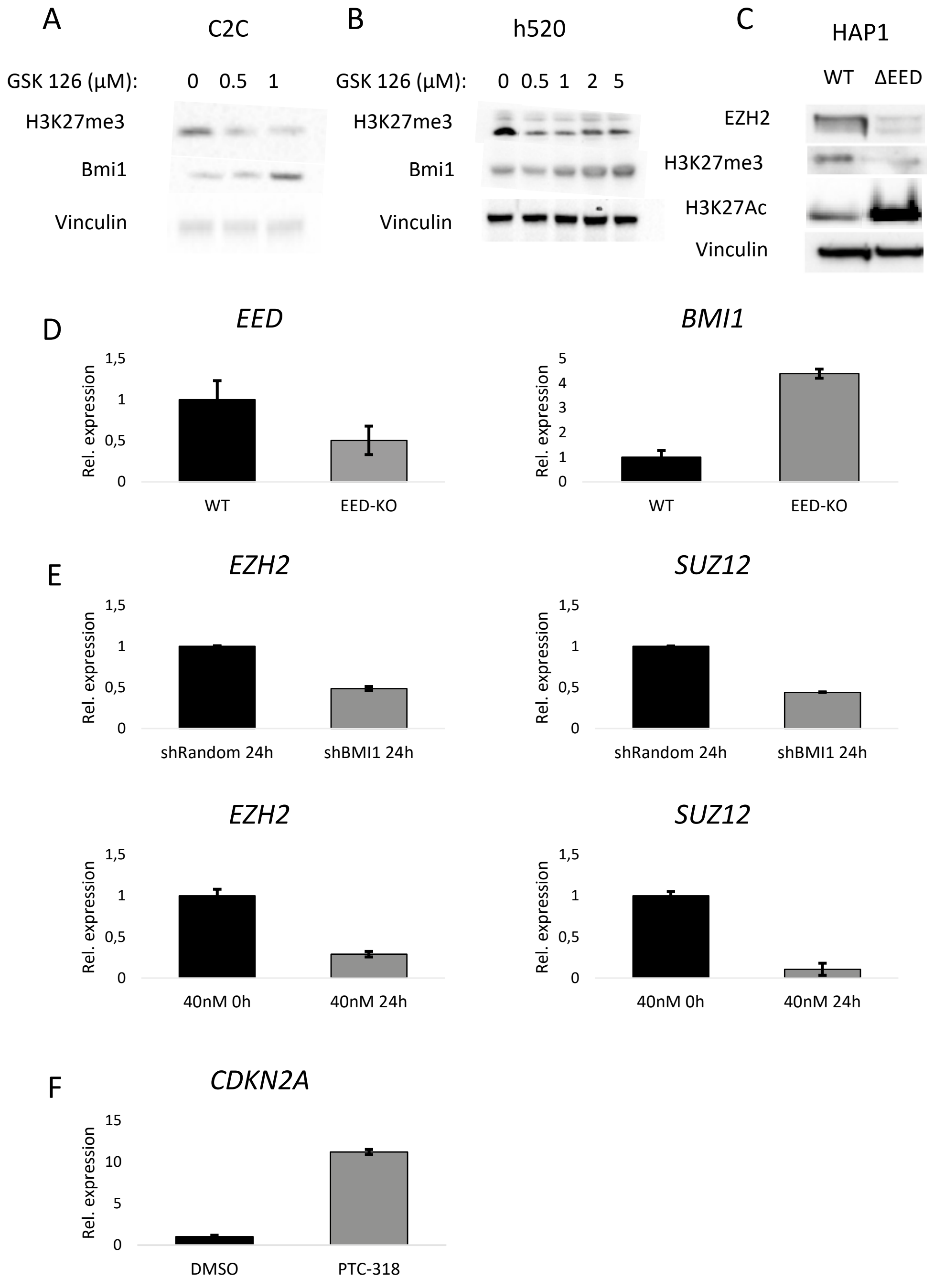
Interested to know if the negative correlation between the two PcG complexes is also present upon BMI1 inhibition, we performed transcript quantification of *EZH2*, *EED,* and *SUZ12* upon modulation of *BMI1*. To this end, we implemented two separate inhibition approaches: BMI1 knockdown with short-hairpin RNA (shRNA) targeting *BMI1* or treatment with a previously published small-molecule inhibitor of BMI1 (PTC-318)[39]. Both *BMI1* knockdown and inhibition of BMI1 with PTC-318 reduced the mRNA expression of the PRC2 proteins **(Fig 1E)**, suggesting that PRC2 expression levels depend on BMI1 expression. We also observed that inhibition of BMI1 derepressed the levels of *CDKN2A* in accordance with its role as a repressor of p16 **(Fig 1F)**.

Altogether, we found that downregulation of EZH2 or EED increased the expression levels of BMI1, indicating an inverse correlation between the two PRC complexes and suggesting that PRC2, in the specific context, could potentially function as a tumor suppressor. Importantly, downregulation of BMI1 did not increase the levels of PRC2 protein but, instead, resulted in a reduction of the respective transcripts, possibly reflecting a positive feedback mechanism between BMI1 and PRC2.

Genetic screens confirm an inverse correlation between PcG complexes with possible links to *BAZ1B*

Our expression analyses upon inhibition of PRC1 or PRC2 proteins suggested that PRC2 proteins negatively regulate BMI1 but not the other way around. We wanted to better understand this relationship by implementing different haploid genetic screen analyses and potentially identifying common gene candidates associated with loss of BMI1 or EED.

Because of their haploid nature, HAP1 cells are especially sensitive to gene mutations, as they only require gene inactivation of one allele to establish complete gene knockout. Both enrichment screens (e.g., to analyze drug resistance) and depletion screens (e.g.,



to analyze synthetic lethality) have been successfully applied on human haploid cell lines to evaluate drug response and various biological processes[39–46]. Briefly, HAP1 genetic screens are prepared by transducing the cells with retroviral GFP-reporter gene traps that integrate randomly into their genome. The gene traps are designed with a splice

◀ **Figure 1.** Inhibition of the expression of PRC2 proteins upregulates BMI1 expression but not the other way around. (A) Levels of H3K27me3 and BMI1 in C2C cells and (B) h520 cells upon treatment with increasing concentrations of GSK126. Vinculin was used as a loading control. (C) Expression levels of EZH2 and H3K27me3 and H3K27Ac in wild-type and EED knockout (ΔEED) HAP1 cells. Vinculin was used as a loading control. (D) Relative mRNA expression levels of *EED* and *BMI1* in wild-type and ΔEED HAP1 cells. (E) Relative mRNA expression levels of *EZH2* and *SUZ12* in wild-type HAP1 cells either treated with 40 nM of PTC-318 for 48 hours or transduced with doxycycline-inducible shRNA constructs FH1t-Random (shRandom) or FH1t-BMI1 (shBMI1) upon 96 hours of doxycycline treatment. (F) Relative mRNA expression of *CDKN2A* in HAP1 cells treated with PTC-318. Error bars represent SD (n = 3). A two-tailed Student's t-test was performed for statistical testing.

acceptor and poly-A tail on the respective ends of the reporter. Because the gene traps are unidirectional with a high probability of landing in introns, sense-oriented gene traps most likely truncate and impair transcription, whereas antisense-oriented gene traps leave the gene intact (unless they integrate into exons).

While we have shown that BMI1 inhibition induced cell lethality in our previously assessed panel of cell lines **(Fig 2B and Chapter 4)**, inhibition of PRC2 proteins did not markedly affect their survival. As a result, we performed two types of haploid screens: an enrichment screen with PTC-318 and a depletion screen with ΔEED **(Figs 2C, 2D, Supplementary Figure 1, and Supplementary Table 1)**. As described in Chapter 4 in the enrichment screen, we subjected HAP1 cells to 40 nM of PTC-318 to screen for gene disruptions associated with resistance to BMI1 inhibition. By contrast, we used the ΔEED HAP1 cell line in the synthetic lethality screen to identify the mutation fitness of specific candidates upon loss of PRC2 function.

Both screens identified a wide panel of gene disruptions significantly associated with either resistance to PTC-318 or synthetic lethality with ΔEED, respectively. Interestingly, we found one common gene enriched in both screens, namely Bromodomain adjacent to zinc finger domain, 1B (*BAZ1B*). *BAZ1B* encodes Williams syndrome transcription factor (WSTF), a central protein of the WHICH complex, a nucleosome mobilizer and transcriptional regulator through chromatin remodeling[47].

The data suggested that disruption of *BAZ1B* both induces resistance to HAP1 cell lethality induced by BMI1 inhibition with PTC-318 and mediates cell death in ΔEED HAP1 cells. Together, the extracted data support our observed inverse correlation between core proteins of PRC1 and PRC2.

Phenotype screening analysis reveals *BAZ1* as a positive regulator of H3K27me3

Until this point, we had performed expression analyses, followed by an enrichment screen of PTC-318 treated HAP1 cells and a depletion screen of ΔEED HAP1 cells to evaluate PcG interactions. Through these approaches, we identified an inverse correlation between BMI1 and PRC2, indicating that downregulation of the individual PRC2 proteins EED and EZH2 upregulates BMI1 – but not the other way around. Intrigued by the identification of

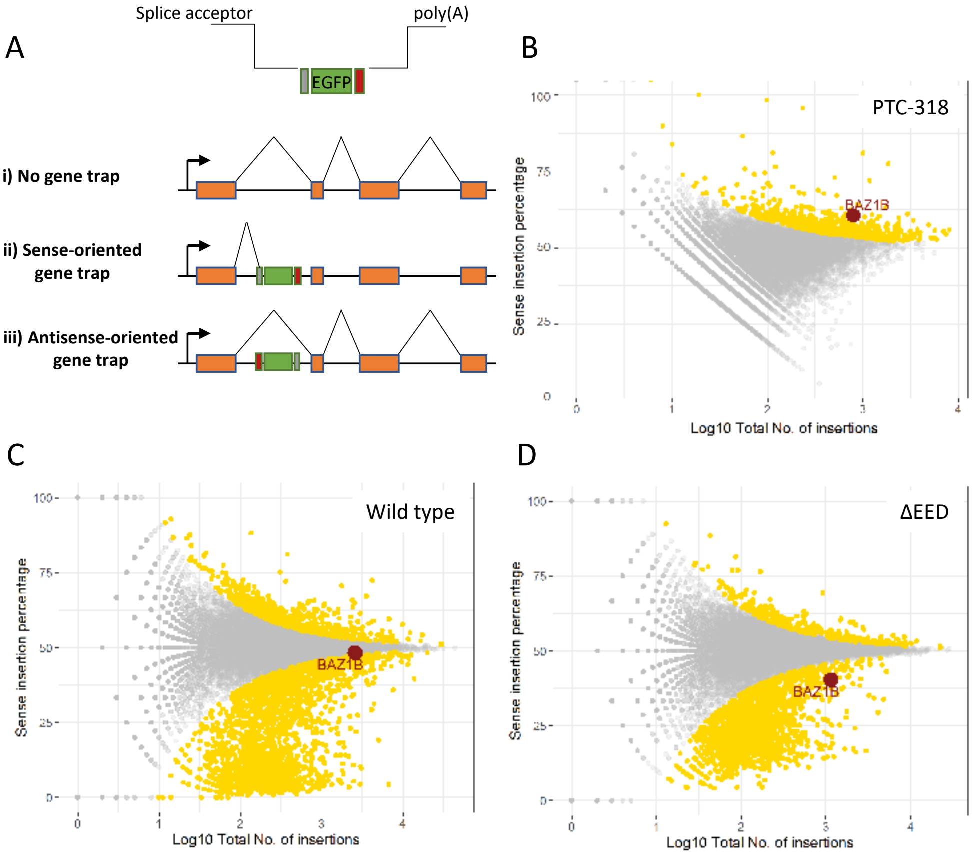


Figure 2. Inactivating gene-trap integrations in BAZ1 are enriched in BMI1-inhibited HAP1 cells but depleted in ΔEED HAP1 cells. (A) A schematic figure representing the gene trap (top) and the different gene-expression outcomes upon its sense- or antisense-oriented integrations into introns (i, ii, and iii). (B) Data from an enrichment screen of HAP1 cells treated with 40 nM of PTC-318. The y-axis shows the percentage of insertions integrating with a sense orientation, and the x-axis shows the total number of mutations assigned to the gene. Data in yellow represent significantly enriched genes (two-sided Fisher's exact test, false discovery rate-corrected $p<0.05$). (C) Representative plots of synthetic screens for wild-type (C) and ΔEED clone (D). Each dot represents an individual gene. The y-axis shows the percentage of insertions integrating with a sense orientation, and the x-axis shows the total number of mutations assigned to the gene. Data in yellow represent significantly enriched genes (two-sided Fisher's exact test, false discovery rate-corrected $p<0.05$). Enrichment screen plot of HAP1 cells treated with 40 nM PTC-318.

BAZ1B as a potential mediator of the observed inverse relationship between members of PRC1 and PRC2, we were encouraged to investigate further and validate the association between BAZ1B and H3K27me3.

To this aim, we researched previously published open-access datasets of phenotype genetic screens on gene-trap mutagenized HAP1 cells based on fluorescence-activated cell sorting (FACS) of specific protein phenotypes[48]. In these screens, 10^8 gene-trap mutagenized HAP1 cells[40,49,50] are fixed and permeabilized. The cells are then exposed to fluorescent antibodies that target the phenotype marker of interest and sorted by FACS to select populations with low or high levels of the specific marker **(Fig 3A)**.

We analyzed gene-disruption data from protein phenotype studies of H3K27me3 levels. Two distinct populations of HAP1 cells had been isolated based on their H3K27me3 levels: H3K27me3 low and H3K27me3 high. Subsequently, the high complexity gene-trap library used in the screen enabled us to extract unique genomic mutations associated with the respective H3K27me3 phenotype. Unsurprisingly, we found significant numbers of disruptive mutations in PRC2 genes *EED, EZH2, and SUZ12* in the H3K27me3 low population **(Fig 3B)**. The results confirm that disruption of any of these PRC2 proteins negatively affects H3K27me3 and, consequently, their essential roles as key mediators of gene silencing through H3K27me3. Interestingly, in the same H3K27me3 low population data, we also observed a significant enrichment of mutagenized *BAZ1B* ($p = 3.13e\text{-}26$). Similar to the PRC2 proteins, the data further suggest that the proper expression of *BAZ1B* positively regulates H3K27me3, either directly or indirectly.

Surprisingly, we also retrieved the enrichment of two components of the PBAF complex, *PBRM1* ($p = 8.14e\text{-}23$) and *ARID2* ($p = 2.58e\text{-}5$), among the negative regulators of H3K27me3 (H3K27 high population). The PBAF subcomplex belongs to the ATP-dependent nucleosome remodeling complex SWI/SNF, a reported antagonist of PcG proteins[51]. According to the protein phenotype screen results, disruption of *PBRM1* or *ARID2* is associated with high H3K27me3 levels, suggesting that they are negative regulators of the methylation mark and possibly antagonists of PRC2. Interestingly, we had previously identified enrichment of the same two gene disruptions in our HAP1 enrichment screen with PTC-318 **(Fig 2B)**[39].

In summary, while verifying positive and negative regulators of H3K27me3 from existing datasets of a protein phenotype screen on mutagenized HAP1 cells, we identified *BAZ1B* as a positive regulator of H3K27me3. These observations further support the synthetic lethality association between *EED* and *BAZ1B* and indicate that *BAZ1B* and its product WSTF may regulate H3K27me3. The screening results also revealed two enriched disruptions in SWI/SNF genes *PBRM1* and *ARID2*, suggesting that these may be involved in regulating PRC2-dependent H3K27me3 and the BMI1 pathway.

DISCUSSION

Inspired by previous observations in our lab, this pilot study extracted information from three different HAP1 screening approaches to identify potential gene candidates associated with the negative correlation between PRC1 and PRC2 protein expression. We found that inhibition or disruption of the PRC2 proteins in selected cancer cell lines increased expression of BMI1. By contrast, inhibiting *BMI1* expression with short-hairpin RNA or PTC-318 reduced levels of selected PRC2 genes. Our analyses of high-complex library data confirmed this inverse correlation between the two PcG complexes, which are

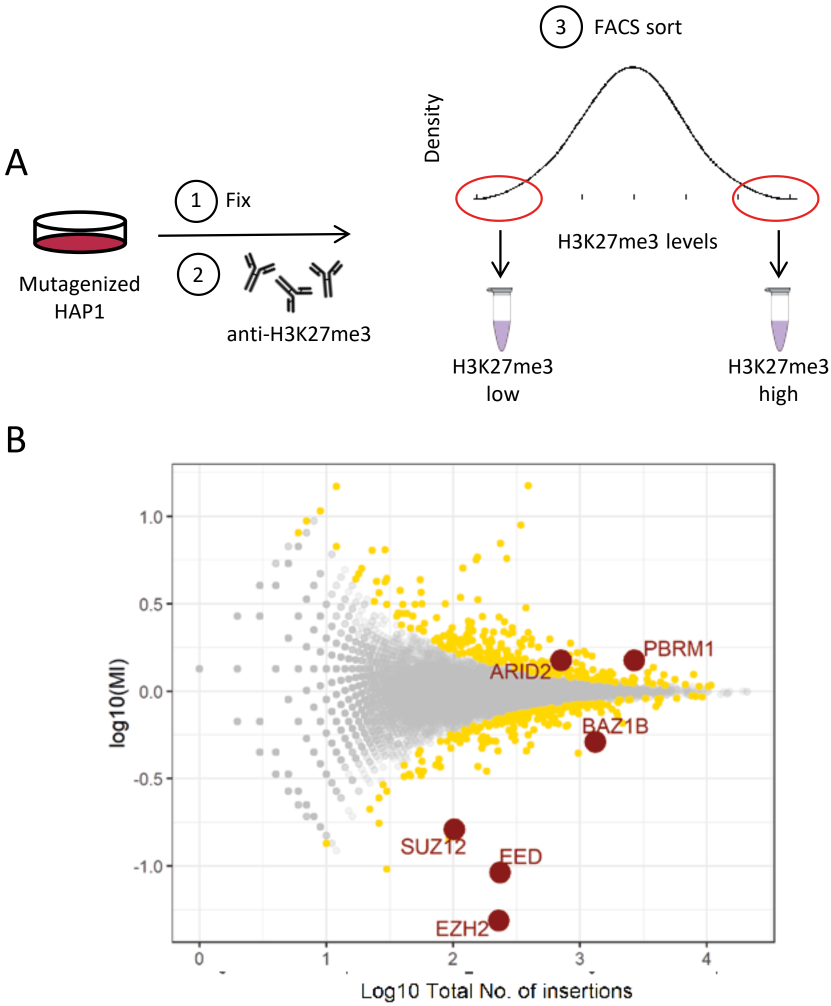


Figure 3. BAZ1B disruption reduces H3K27me3 levels according to "Fix cell genetics". (A) A schematic figure representing the set-up of phenotype genetic screen. (B) Plot of HAP1 cells after sorting cell populations according to their levels of H3K27me3 (H3K27me3 low and H3K27me3 high). Each dot on the plot represents a gene. The y-axis shows the ratio between inserted gene-traps (sense-orientation) in the respective cell populations, plotted as mutation index (MI). The x-axis shows the total number of sense-oriented gene-trap insertions mapped for a specific gene.

traditionally described as part of the same gene-repressive pathway. Interestingly, early observations in *Drosophila melanogaster* also demonstrate that mutations of *Polycomb* derepress, not only homeotic genes but also other PcG genes along the polytene chromosomes of the salivary glands[52], indicating that regulation within PcG is conserved.

Gathering data from the different HAP1 screens helped us identify *BAZ1B* as a potential tumor suppressor that positively regulates H3K27me3 and induces resistance to BMI1 inhibition when disrupted. Importantly, we observed recurrent *BAZ1B* disruptions in all three screens, which further supports its possible involvement in the inverse correlation observed between *BMI1* and PRC2 genes, notably *EED*.

These preliminary results may highlight context-dependent functions of the two PcG complexes that are distinct from the canonical PcG pathway in which PRC2-induced methylation of H3K27me3 at target promoters recruits PRC1 to stabilize repression[8,53–55]. It is becoming increasingly accepted that PRC2 proteins – especially EZH2 – besides their oncogenic functions also function as tumor suppressors[56]. For example, genotypic analyses of patients with myelodysplastic syndromes/myeloproliferative neoplasms (MDS/MPN) showed that a proportion of the patients had mutations or deletions of EZH2-encompassing regions on chromosome 7 or 7q[57]. Apart from these findings, several studies have identified inactivating mutations in the catalytic SET domain of EZH2 in various hematological malignancies, such as chronic myelomonocytic leukemia (CMML)[58]. Furthermore, a study on mouse KRAS-driven non-small cell lung cancer showed that *Eed* functions as a tumor suppressor in a *Kras* and *Trp53* mutated background[35]. Together, these studies support our preliminary findings, suggesting that PRC2 proteins may function as tumor-suppressors in our assessed cell lines. We have previously demonstrated that selected lung cancer and leukemia cell lines are sensitive to BMI1 inhibition[39] and can employ a tumor-suppressive function of PRC2 proteins[35]. It is, therefore, possible that PRC2 adapts its tumor-suppressive function in specific cell contexts by negatively regulating BMI1 expression, which appears to be essential for both HAP1 cells and other cancer cell lines. Further studies would be able to validate these findings in cancer cell lines where an inverse correlation between PRC1 and PRC2 is observed, such as the ones presented here.

Deciphering these context-dependent tumor-suppressive functions may help the development of future therapeutic strategies targeting epigenetic proteins, such as PcG proteins. Comprehensive screening strategies, such as the diverse HAP1 screens described here, may serve to identify non-canonical PcG functions and the interacting genes implicated in these processes. Our results from separate screens also retrieved recurring gene candidates, one of them being *BAZ1B*. The gene product of *BAZ1B*, WSTF, is a multifaceted nuclear protein and an essential component of the chromatin remodeling complexes WICH (WSTF-ISWI chromatin remodeling complex), B-WICH, and WNAC (WSTF including the nucleosome assembly complex). WSTF is involved in a wide range of cellular processes, including regulating replication, transcription, and DNA damage repair[47,59]. In DNA replication, WSTF is implicated in maintaining chromatin organization, preventing heterochromatin spreading[60]. Interestingly, previous findings have shown that knockout of *BAZ1B* in human cells induces inappropriate chromatin changes throughout the nucleus[61]. Although our current knowledge of the interactions and dynamics of WSTF remains limited, the observations suggest that loss of *BAZ1B* function disrupts the regulation of several genes. The findings make *BAZ1B* a viable candidate as a potential participant in context-dependent, non-canonical PcG functions.

Furthermore, the fact that we identified enrichment of the mutagenized SWI/SNF genes *PBRM1* (encoding BAF180) and *ARID2* (encoding BAF200) in two of our screens further emphasizes the complexity of these regulatory complexes. SWI/SNF complexes are known antagonists of PcG proteins[51], and both *PBRM1* and *ARID2* were enriched in the PTC-318 resistance screen and the protein phenotype screen assessing negative regulators of H3K27me3. Their simultaneous appearance in both these screens may increase the validity of the hits since they are two of the three proteins specific to PBAF – the rest of the proteins associated with PBAF appear in other SWI/SNF subcomplexes. Moreover, both SWI/SNF and ISWI, the latter in which BAZ1B is a subunit, are ATP-dependent chromatin remodelers but with distinct nucleosome displacement strategies[47]. While SWI/SNF can both activate and repress genes, ISWI mainly represses target genes[47,62]. This difference might explain why we identify *BAZ1B* and the PBAF genes *PBRM1* and *ARID2* disrupted in HAP1 cells resistant to PTC-318 but find them on opposite sides of the H3K27me3 phenotype screen.

Importantly, recent studies have demonstrated that the interplay between PcG and SWI/SNF is highly dynamic and context dependent[63–65]. One study assessed SWI/SNF and PcG-mediated gene regulation in mouse embryonic stem cells and found that PRC1 and PRC2 redistribute in the genome upon degradation of Brg1, the ATPase subunit in BAF and PBAF SWI/SNF complexes[63]. The group showed that core subunits of PRC1 and PRC2 redistribute from genomic domains highly occupied by PcG, for example, *Hox* clusters, to domains with low PcG occupancy. Moreover, reduced Brg1 levels also coincided with depleted PRC1 and PRC2 and transcriptional derepression at the *Bmi1* locus. Overall, these findings emphasize the complex dynamic processes and interactions associated with these epigenetic complexes. They show how SWI/SNF complexes, such as BAF, antagonize PcG repression through ATP-dependent eviction[66] and support repressed states on the genome, depending on the context. The findings also provide insight into the dynamic PcG regulation, which constantly competes with SWI/SNF and other chromatin modifiers in a dosage-dependent manner, and how changes in this balance can disrupt regulation between PRC complexes (e.g., PcG-mediated repression of the *Bmi1* locus).

Although limited and speculative, our current understanding of these pathways, including the interactions explained above between PcG, WSTF, and SWI/SNF proteins, leads us to suggest a preliminary, high-level model in which PRC2 negatively regulates BMI1 **(Fig 4)**. Based on the enrichment screen, we have observed that downregulation of BMI1 results in apoptosis of HAP1 cells and other cancer cell lines, perhaps through the derepression of tumor-suppressor genes, such as the INK4A/ARF locus. By contrast, overexpression of BMI1 has also been associated with apoptosis through the regulation of the anti-apoptotic protein Survivin. Therefore, it is possible that the role of PRC2, in this context, is to maintain a balance of BMI1 that maintains cell survival. Moreover, the cell lethality associated with BMI1 inhibition can be rescued through the loss of *BAZ1B* function, possibly induced by "inappropriate" rearrangements of the chromatin and decompaction of oncogenes. Our findings from the synthetic lethality screen and

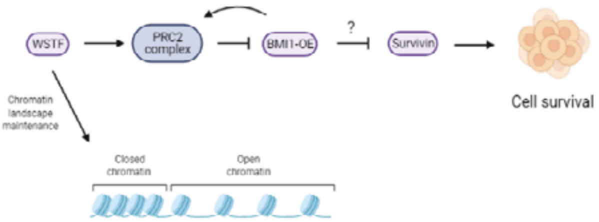


Figure 4. Representation of WSTF- and PRC2-associated regulation of BMI1 overexpression and cell survival. Synergism between PRC2 and WSTF may regulate *BMI1* expression to establish the "just-right" levels of BMI1 and maintain cell survival. Like BMI1 inhibition, BMI1 overexpression (BMI1-OE) can result in cell lethality by dysregulation of anti-apoptotic proteins, e.g., Survivin.

H3K27me3 phenotype screen suggest that WSTF is a positive regulator of H3K27me3 and, perhaps, acts synergistically with PRC2 to regulate target genes, including *BMI1*. Finally, similar to the loss of *BAZ1B*, loss of *PBRM1* or *ARID2* may redistribute relevant epigenetic modifiers and alter global gene expression, including processes involving PcG or WSTF. However, considering the complexity of these processes and the limited knowledge, the current data do not enable us to specify an exact pathway, which may vary depending on the context.

Overall, it is important to note that the current findings are preliminary and require extensive genotypic and phenotypic validation to consolidate the proposed conclusions. It will be interesting to validate these findings in a panel of cancer cell lines by establishing *BAZ1B*-null mutants and evaluating the possible synthetic lethal interactions between WSTF and EED. Furthermore, future studies *in vitro* and *in vivo* could analyze cancer-associated phenotypes or expression profiles to explain the reciprocal regulation of PRC1 and PRC2, the role of WSTF and other recurring screen hits, and a possible context-dependent tumor-suppressive role by PRC2. These may be able to investigate how deregulation of PcG and relevant chromatin remodelers affect the levels of other PgG proteins, both in terms of genome localization and redistribution and in terms of global expression. Nevertheless, our analysis of three different HAP1 screening methods achieved to, at least partly, confirm and explain the observed inverse correlation between PRC1 and PRC2 proteins.

MATERIALS AND METHODS

Cell culture

HAP1 cells (provided by the T. Brummelkamp laboratory at the Netherlands Cancer Institute, Amsterdam, the Netherlands) have been described previously[67]. The cells were maintained in IMDM + GlutaMAX (Gibco) supplemented with 10% fetal calf serum (FCS; Sigma-Aldrich) and penicillin–streptomycin (Gibco). NSCLC cell line h520 and SCLC cell line C2C (provided by the A. Berns laboratory at the Netherlands Cancer Institute, Amsterdam, the Netherlands) were maintained in DMEM/F12 + GlutaMAX (Gibco) supplemented with 10% FCS and penicillin–streptomycin (Gibco).

Haploid genetic screen

Procedures for the generation of gene-trap retrovirus and HAP1 mutagenesis have been described previously[42,67]. To select HAP1 variants resistant to BMI1 inhibitor PTC-318, approximately 10^8 mutagenized HAP1 cells (>90% haploid) were seeded in fourteen T175 cell culture flasks. The cells were exposed to 40 nM of PTC-318 24 hours after seeding and incubated for fourteen days. Surviving HAP1 clones were trypsinized and washed before amplification and analysis of integration sites, as described in[42,68]. In brief, as a first step, insertion sites were amplified in a linear amplification reaction using a biotinylated primer. Products were captured on streptavidin-coated magnetic beads, washed, and subjected to single-stranded DNA linker ligation followed by a second PCR to finalize the products for Illumina sequencing.

For synthetic lethality screens, gene trap retrovirus was produced as previously described in Jae et al., 2013[68]. Approximately 40 million ΔEED HAP1 cells were mutagenized by transduction with concentrated gene trap virus in the presence of 8 μg/ml protamine sulfate in a T175 flask for at least two consecutive days. The mutagenized cells were passaged for an additional 10-12 days following the last infection, after which cells were collected by trypsinizing. Cells were fixed using Fix Buffer I (BD biosciences). To minimize a confounding effect from diploid cells, with potentially heterozygous mutations, DAPI-stained haploid fixed cells were sorted based on G1 of haploid cells using Moflo flow cytometry. Thirty million sorted cells were lysed overnight at 56°C to allow for de-crosslinking, followed by genomic DNA isolation using a DNA Mini Kit (Qiagen).

Generation of shRNA knockdown and EED knockout cell lines

For BMI1 knockdown experiments, we used doxycycline-inducible FH1-tUTG-RNAi vectors[69], as described previously[70]. shRNA targeting sequences: shBMI1 (GATTGGATCGGAAAGTAAAC), shEZH2 (ACCATTTCCTCAATGTTTCCAG), shSUZ12 (GGATGTAAGTTGTCCAATA). For Cas9-induced knockout, we used SpCas9 and chimeric guide RNA (gRNA) expression plasmid pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene 42230) encoding a gRNA targeting *EED* (ΔEED). Cells were transfected in 6-well plates with 1.2–1.6 μg of Cas9–gRNA construct and 10% of mPB-L3-ERT2.TatRRR-mCherry plasmid, following the Lipofectamine 2000 protocol. mCherry-positive cells were single-cell sorted by flow cytometry (MoFlo) into 96-well plates 48 hours after transfection and incubated for two weeks before expansion. Gene mutations were validated by Sanger sequencing and western blot analysis. gRNA targeting sequence was: 5′-GAGGGAAGTGTCGACTGCGC-3′.

Gene expression

Total RNA was isolated with the ReliaPrep™ RNA Cell Miniprep System (Promega) according to the manufacturer's instructions. RNA quantity and quality were assessed using a Nanodrop 2000c (Thermo Scientific). Primer sets used (Integrated DNA technologies; IDT) had the following sequences: BMI1 forward CTGATGCTGCCAATGGCTCTA; BMI1 reverse GTGCATCACAGTCATTGCTGCT; EED forward 5'-ATGCTGTCAGTATTGAGAGTGGC-3'; EED reverse 5'-GAGGCTGTTCACACATTTGAAAG-3'; EZH2 forward 5'-CGCGGGACTAGG GAGTGTTCAGT-3'; EZH2 reverse 5'-AGTACATTATAGGCACCGAGGCGA-3'; SUZ12 forward 5'-CCGAGCACTGTGGTTGAGTA-3'; SUZ12 reverse 5'-AACTGCATCTGATGGTGGTG-3'.

Protein expression

Whole-cell extracts were pelleted and prepared in RIPA buffer (50 mM Tris, pH 8.0, 50 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, and 0.1% SDS) containing protease inhibitor cocktail (Complete; Roche) and phosphate inhibitors (10 mM Na fluoride final concentration, 1mM sodium orthovanadate final concentration, and 1 mM NaPPi final concentration). Equal amounts of protein, as determined by a Bio-Rad Protein Assay Dye Reagent on Nanodrop 2000c, were resolved on NuPage-Novex 4–12% Bis-Tris gels (Invitrogen) and transferred onto nitrocellulose membranes (0.2 m; Whatman). Membranes were blocked in phosphate-buffered saline (PBS) with 0.1% Tween-20 (PBST) and 5% BSA for 1 h, incubated with primary antibodies in PBST 1% BSA overnight at 4°C, and incubated with secondary antibodies coupled to HRP for 45 min in PBST 1% BSA at room temperature. Membranes were imaged on a BioRad ChemiDoc XRS+. The following antibodies were used for western blot analyses: anti-BMI1 D20B7 (Cell Signalling, 6964T), anti-EZH2 (Merck Millipore, 07-689), anti-H3K27me3 (Abcam, ab6002), anti-H3K27Ac (Active Motif, 39133), Vinculin (Sigma-Aldrich, V9131).

Statistical analyses

As described for enrichment mutagenesis screens before[42], sequence reads were aligned to the human genome (HG19) using bowtie, and uniquely aligned insertion sites were mapped to the genomic coordinates (RefSeq) of non-overlapping protein-encoding gene regions. Gene trap integrations in a sense orientation in introns, or regardless of orientation in exons, were considered disruptive. To select genes enriched for mutations after PTC-318 selection, the number of disruptive integrations in each gene was compared with those retrieved in an unselected HAP1 population (available at the SRA (SRP018361: accession SRX223544) using a one-sided Fisher's exact test. P values were false discovery rate (FDR)-corrected (Benjamini—Hochberg). The dot plot was established with R and Bioconductor packages.

Mapping of the insertion sites was achieved by aligning the deep sequencing reads to the human genome (hg19) using bowtie (Langmead et al., 2009), allowing for a single mismatch. The 65bp reads from the HiSeq 2500 were cropped to 50bp, similar to the output from the HiSeq 2000 runs. Next, the unique aligned reads were allocated to Refseq gene coordinates using Bedtools (Quinlan and Hall, 2010). Overlapping gene regions on opposite strands were omitted for further analysis since orientation bias is not readily interpretable in those regions. In contrast, names of genes with overlapping regions on the same strand were concatenated. All datasets were normalized to the aggregated insertion counts of four wild-type HAP1 datasets; see Blomen et al., 2015[50] for normalization details. A binomial test for the distribution of sense and antisense orientation insertions was performed for each replicate. A gene was considered synthetic lethal with EED when it showed increased depletion of sense insertions as compared to 4 independent WT datasets that were previously published ((Blomen et al., 2015), NCBI SRA accession no. SRP058962). To this end, a total of 16 two-sided Fisher's exact tests were performed (2 independent datasets per genotype against all 4 WT datasets) per gene. A gene was considered a synthetic lethal interactor when it passed all Fisher's tests with a P-value cut-off of 0.05 and an effect size of at least 20%.

The data from the H3K27me3 phenotype screen was available in Brockmann et al. 2017[48] (Supplementary Table 5: Histone H3(K27) trimethylation.xlsx). The dot plot with the log10(MI) on the y-axis and the genes distributed over the x-axis in alphabetical order were established with R and Bioconductor packages. For each gene, a mutation index (MI) was calculated corresponding to the ratio of the number of disruptive integrations per gene in both populations normalized by the number of total integrations in each channel (see Brockmann et al. 2017). For genes without a single insertion site in only one of the channels, a value of 1 was assigned so as not be omitted from the plots.

ACKNOWLEDGMENTS

We thank the NKI Genomics Core Facility and the NKI Flow Cytometry Facility for their technical support. We also want to thank Vincent Blomen, Joppe Nieuwenhuis, Thijn R. Brummelkamp, and Andrej Alendar for the valuable help, discussions, and/or constructive criticism of the manuscript.

REFERENCES

1. Sparmann, A. & Van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nature Reviews Cancer* **6**, 846–856 (2006).
2. Laugesen, A., Højfeldt, J. W. & Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Molecular Cell* **74**, 8–18 (2019).
3. Simon, J. A. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Molecular Cell* **49**, 808–824 (2013).
4. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nat. 2011 4697330* **469**, 343–349 (2011).
5. Glancy, E., Ciferri, C. & Bracken, A. P. Structural basis for PRC2 engagement with chromatin. *Current Opinion in Structural Biology* **67**, 135–144 (2021).
6. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a polycomb group protein complex. *Science (80-.).* **306**, 1574–1577 (2004).
7. Gao, Z. *et al.* PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell* **45**, 344–356 (2012).
8. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870 (2003).
9. Min, J., Zhang, Y. & Xu, R. M. Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823–1828 (2003).
10. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
11. Bruggeman, S. W. M. *et al.* Ink4a and Arf differentially affect cell proliferation and neural stem cell self-renewal in Bmi1-deficient mice. *Genes Dev.* (2005). doi:10.1101/gad.1299305
12. Bracken, A. P. *et al.* The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev.* **21**, 525–530 (2007).
13. Gil, J. & Peters, G. Regulation of the INK4b-ARF-INK4a tumour suppressor locus: All for one or one for all. *Nature Reviews Molecular Cell Biology* **7**, 667–677 (2006).
14. Sugishita, H. *et al.* Variant PCGF1-PRC1 links PRC2 recruitment with differentiation-associated transcriptional inactivation at target genes. *Nat. Commun. 2021 121* **12**, 1–12 (2021).
15. Blackledge, N. P. *et al.* PRC1 Catalytic Activity Is Central to Polycomb System Function In Brief. (2020). doi:10.1016/j.molcel.2019.12.001
16. Haupt, Y., Alexander, W. S., Barri, G., Peter Klinken, S. & Adams, J. M. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in Eμ-myc transgenic mice. *Cell* (1991). doi:10.1016/0092-8674(91)90383-A
17. van Lohuizen, M. *et al.* Identification of cooperating oncogenes in Eμ-myc transgenic mice by provirus tagging. *Cell* (1991). doi:10.1016/0092-8674(91)90382-9
18. Ojo, D. *et al.* Polycomb complex protein BMI1 confers resistance to tamoxifen in estrogen receptor positive breast cancer. *Cancer Lett.* (2018). doi:10.1016/j.canlet.2018.03.048
19. Wu, Z. *et al.* Overexpression of BMI-1 promotes cell growth and resistance to cisplatin treatment in osteosarcoma. *PLoS One* (2011). doi:10.1371/journal.pone.0014648
20. Yin, T. *et al.* Bmi-1 promotes the chemoresistance, invasion and tumorigenesis of pancreatic cancer cells. *Chemotherapy* (2012). doi:10.1159/000334103
21. Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11606–11611 (2003).
22. Bracken, A. P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* **22**, 5323–5335 (2003).

23. Pawlyn, C. *et al.* Overexpression of EZH2 in multiple myeloma is associated with poor prognosis and dysregulation of cell cycle control. *Blood Cancer J.* **7**, (2017).
24. Nienstedt, J. C. *et al.* EZH2 overexpression in head and neck cancer is related to lymph node metastasis. *J. Oral Pathol. Med.* **47**, 240–245 (2018).
25. Score, J. *et al.* Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. *Blood* **119**, 1208–1213 (2012).
26. Khan, S. N. *et al.* Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. *Leukemia* **27**, 1301–1309 (2013).
27. Ernst, T. *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.* **42**, 722–726 (2010).
28. Jiao, L. *et al.* A partially disordered region connects gene repression and activation functions of EZH2. *Proc. Natl. Acad. Sci.* **117**, 16992–17002 (2020).
29. Bachmann, I. M., Puntervoll, H. E., Otte, A. P. & Akslen, L. A. Loss of BMI-1 expression is associated with clinical progress of malignant melanoma. *Mod. Pathol. 2008 215* **21**, 583–590 (2008).
30. Engelsen, I. B. *et al.* Low BMI-1 expression is associated with an activated BMI-1-driven signature, vascular invasion, and hormone receptor loss in endometrial carcinoma. *Br. J. Cancer 2008 9810* **98**, 1662–1669 (2008).
31. Pietersen, A. M. *et al.* EZH2 and BMI1 inversely correlate with prognosis and TP53 mutation in breast cancer. *Breast Cancer Res. 2008 106* **10**, 1–12 (2008).
32. Choi, Y. J. *et al.* Expression of Bmi-1 protein in tumor tissues is associated with favorable prognosis in breast cancer patients. *Breast Cancer Res. Treat.* **113**, 83–93 (2009).
33. Shinoda, D. *et al.* Insufficiency of non-canonical PRC1 synergizes with JAK2V617F in the development of myelofibrosis. *Leuk. 2021* 1–12 (2021). doi:10.1038/s41375-021-01402-2
34. Cohen, I., Zhao, D., Koseki, H., Zheng, D. & Correspondence, E. E. PRC1 Fine-tunes Gene Repression and Activation to Safeguard Skin Development and Stem Cell Specification. *Cell Stem Cell* **22**, 726–739 (2018).
35. Serresi, M. *et al.* Ezh2 inhibition in Kras-driven lung cancer amplifies inflammation and associated vulnerabilities. *J. Exp. Med.* **215**, 3115–3135 (2018).
36. Zhang, Q., Dong, P., Liu, X., Sakuragi, N. & Guo, S. W. Enhancer of Zeste homolog 2 (EZH2) induces epithelial-mesenchymal transition in endometriosis. *Sci. Rep.* **7**, (2017).
37. Ito, T., Teo, Y. V., Evans, S. A., Neretti, N. & Sedivy, J. M. Regulation of Cellular Senescence by Polycomb Chromatin Modifiers through Distinct DNA Damage- and Histone Methylation-Dependent Pathways. *Cell Rep.* **22**, 3480–3492 (2018).
38. Morse, D. E. & Yanofsky, C. Polarity and the degradation of mRNA. *Nature* **224**, 329–331 (1969).
39. Gisler, S., Maia, A. R. R., Chandrasekaran, G. & van Lohuizen, M. A genome-wide enrichment screen identifies NUMA1-loss as a resistance mechanism against mitotic cell-death induced by BMI1 inhibition. *bioRxiv* (2019). doi:10.1101/2019.12.24.887851
40. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science (80-.).* (2009). doi:10.1126/science.1178955
41. Jae, L. T. *et al.* Lassa virus entry requires a trigger-induced receptor switch. *Science (80-.).* (2014). doi:10.1126/science.1252480
42. Staring, J. *et al.* PLA2G16 represents a switch between entry and clearance of Picornaviridae. *Nature* (2017). doi:10.1038/nature21032
43. Mezzadra, R. *et al.* SLFN11 can sensitize tumor cells towards IFN-γ-mediated T cell killing. *PLoS One* (2019). doi:10.1371/journal.pone.0212053
44. Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* **46**, 17–23 (2014).
45. Lee, C. C., Carette, J. E., Brummelkamp, T. R. & Ploegh, H. L. A Reporter Screen in a Human Haploid Cell Line Identifies

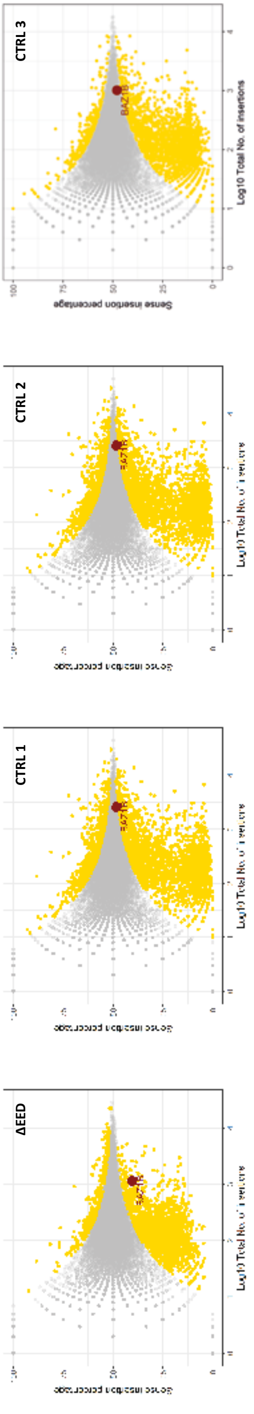
CYLD as a Constitutive Inhibitor of NF-κB. *PLoS One* (2013). doi:10.1371/journal.pone.0070339

46. Planells-Cases, R. *et al.* Subunit composition of VRAC channels determines substrate specificity and cellular resistance to P t-based anti-cancer drugs . *EMBO J.* (2015). doi:10.15252/embj.201592409
47. Barnett, C. & Krebs, J. E. WSTF does it all: A multifunctional protein in transcription, repair, and replication. *Biochemistry and Cell Biology* **89**, 12–23 (2011).
48. Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling. *Nat. 2017 5467657* **546**, 307–311 (2017).
49. Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* (2011). doi:10.1038/nature10348
50. Blomen, V. A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Sci* **350**, 1092–1096 (2015).
51. Wilson, B. G. *et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* (2010). doi:10.1016/j.ccr.2010.09.006
52. Zink, B. & Paro, R. In vivo binding pattern of a trans-regulator of homoeotic genes in Drosophila melanogaster. *Nat. 1989 3376206* **337**, 468–471 (1989).
53. Czermin, B. *et al.* Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).
54. Müller, J. *et al.* Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
55. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in polycomb-group silencing. *Science (80-.).* **298**, 1039–1043 (2002).
56. Koppens, M. & van Lohuizen, M. Context-dependent actions of Polycomb repressors in cancer. *Oncogene* 1–12 (2015).
57. Nikoloski, G. *et al.* Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat. Genet.* **42**, 665–667 (2010).
58. Jankowska, A. M. *et al.* Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX,,EZH2,and DNMT3A. *Blood* **118**, 3932–3941 (2011).
59. Xiao, A. *et al.* WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. *Nature* **457**, 57–62 (2009).
60. Bozhenok, L., Wade, P. A. & Varga-Weisz, P. WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *EMBO J.* **21**, 2231–2241 (2002).
61. Culver-Cochran, A. E. & Chadwick, B. P. Loss of WSTF results in spontaneous fluctuations of heterochromatin formation and resolution, combined with substantial changes to gene expression. *BMC Genomics* **14**, 1–18 (2013).
62. Dirscherl, S. S. & Krebs, J. E. Functional diversity of ISWI complexes. in *Biochemistry and Cell Biology* **82**, 482–489 (Biochem Cell Biol, 2004).
63. Weber, C. M. *et al.* mSWI/SNF promotes Polycomb repression both directly and through genome-wide redistribution. *Nat. Struct. Mol. Biol.* **28**, 501 (2021).
64. Shu, J. *et al.* Genome-wide occupancy of Arabidopsis SWI/SNF chromatin remodeler SPLAYED provides insights into its interplay with its close homolog BRAHMA and Polycomb proteins. *Plant J.* **106**, 200–213 (2021).
65. Menon, D. U., Shibata, Y., Mu, W. & Magnuson, T. Mammalian SWI/SNF collaborates with a polycomb-associated protein to regulate male germline transcription in the mouse. *Development* **146**, (2019).
66. Kadoch, C. *et al.* Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat. Genet.* **49**, 213–222 (2017).
67. Carette, J. E. *et al.* Global gene disruption in human cells to assign genes to phenotypes. *Nat. Biotechnol.* **29**, 542 (2011).
68. Jae, L. T. *et al.* Deciphering the glycosylome of dystroglycanopathies using haploid screens for Lassa virus entry. *Science (80-.).* (2013). doi:10.1126/science.1233675
69. Herold, M. J., Van Den Brandt, J., Seibler, J. & Reichardt, H. M. Inducible and reversible

gene silencing by stable integration of an shRNA-encoding lentivirus in transgenic rats. *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0806213105

70. Michalak, E. M. *et al.* Polycomb group gene Ezh2 regulates mammary gland morphogenesis and maintains the luminal progenitor pool. *Stem Cells* (2013). doi:10.1002/stem.1437.

SUPPLEMENTARY MATERIAL



Supplementary Figure 1. Plots of synthetic screens for ΔEED clone and biological replicates of wild-type controls (CTRL 1–3).

Supplementary Table 1. Top 50 gene hits from the enrichment screen with PTC-318 (left) and depletion screen with ΔEED (right).

	Gene	FDR-corrected Pvalue		Gene	FDR-corrected Pvalue
1	NUMA1	1,61E-304	1	KDM1A;MIR3115	6,53E-14
2	WNK1	1,61E-304	2	BAZ1B	9,60E-11
3	CSMD1	6,32E-240	3	RCOR1	3,92E-10
4	ANKRD30A	6,63E-150	4	RBM27	2,07E-7
5	ATP8A2	4,44E-127	5	CRAMP1L	2,82E-7
6	PBRM1	1,70E-115	6	PTBP1;MIR4745	6,78E-6
7	HN1	1,21E-113	7	UGDH	7,47E-6
8	GPC3	4,03E-88	8	ARHGAP18	9,39E-6
9	ROBO1	1,54E-84	9	OTUD5	1,18E-5
10	TANC2	5,88E-81	10	CHD1	1,64E-5
11	PDSS2	7,21E-77	11	LPAR1	2,37E-5
12	CDH12	1,22E-65	12	DDX5;MIR3064;MIR5047	4,01E-5
13	TUBB4B	6,79E-65	13	SELE	0,000
14	LRPPRC	1,07E-62	14	NDUFB2	0,000
15	IL1RAPL1	2,92E-56	15	AUNIP	0,000
16	MYBPC1	2,05E-55	16	LRRIQ4	0,001
17	TMC1	1,99E-52	17	SYNJ2BP-COX16;COX16;SYNJ2BP	0,001
18	PCDH9	7,34E-52	18	NDUFAF1	0,001
19	ZSWIM6	1,06E-50	19	FAM72B	0,001
20	UNC13C	3,46E-49	20	CCDC90A	0,001
21	MTF2	7,47E-49	21	UBE2G2	0,001
22	DLG2	6,46E-46	22	CHCHD2	0,002
23	HCFC1	3,25E-44	23	ARHGEF19	0,002
24	CADM2	2,36E-41	24	PIKFYVE	0,002
25	FOXO3	7,05E-41	25	ZNF37BP	0,002
26	MRPS28	3,89E-40	26	ALX1	0,002
27	CRADD	1,70E-38	27	ESCO1	0,002
28	ZNRF2	6,16E-38	28	MRPL28	0,002
29	ZBTB20	9,05E-38	29	LACTB	0,002
30	LGALSL	1,27E-36	30	ZNF800	0,002
31	GNAT3	1,51E-35	31	LOC100505989	0,002
32	LRFN5	2,25E-35	32	SGOL2	0,002
33	NAV3	6,04E-35	33	TRMT112	0,003
34	MAP4K3	4,17E-34	34	LSR	0,003
35	CYP7B1	1,59E-31	35	IL20RA	0,003
36	BAZ1B	2,68E-30	36	TRAF4	0,003
37	SGF29	3,90E-30	37	GNPTAB	0,003
38	CNOT2	4,16E-30	38	MPHOSPH8	0,003
39	ANGPT1	4,70E-30	39	MRPL14	0,003
40	POU6F2	6,17E-30	40	LAP3	0,003
41	LIN28B	2,33E-29	41	BTLA	0,003

Supplementary Table 1. (continued).

	Gene	**FDR-corrected Pvalue**		**Gene**	**FDR-corrected Pvalue**
42	NONO	3,13E-28	42	STK19	0,004
43	MGAT4C	3,56E-28	43	TEK	0,004
44	MAD1L1	5,48E-28	44	NOL3	0,004
45	NLK	5,59E-28	45	NDUFS8;MIR4691	0,004
46	ISL1	5,93E-27	46	DDX31	0,004
47	CCDC91	1,56E-26	47	CDH2	0,004
48	CHN1	3,48E-26	48	ABCF2	0,004
49	NARS2	6,42E-26	49	FIZ1	0,005
50	FBXL7	9,38E-26	50	APITD1;APITD1-CORT;CORT	0,005

Chapter 6

TRIM28 IS AN EPIGENETIC BARRIER TO INDUCED PLURIPOTENT STEM CELL REPROGRAMMING

Denise Catherine Miles[1], Nienke Alexandra de Vries[1],
Santiago Gisler[1], Cor Lieftink[2],
Waseem Akhtar[1], Ewa Gogola[3],
Inka Pawlitzky[1], Danielle Hulsman[1],
Ellen Tanger[1], Martijn Koppens[1],
Roderick Leonardus Beijersbergen[2], Maarten van Lohuizen[1, 4]

[1] Division of Molecular Genetics, Oncode and The Netherlands Cancer Institute, Amsterdam, The Netherlands
[2] Division of Molecular Carcinogenesis, NKI Robotics and Screening Center, the Netherlands Cancer Institute, Amsterdam, The Netherlands
[3] Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands;
[4] Cancer Genomics Centre (CGC.nl), Amsterdam, The Netherlands

Adapted from Stem Cells. 2017 Jan;35(1):147-157

ABSTRACT

Since the discovery of induced pluripotent stem cells, there has been intense interest in understanding the mechanisms that allow a somatic cell to be reprogrammed back to a pluripotent state. Several groups have studied the alterations in gene expression that occur as somatic cells modify their genome to that of an embryonic stem cell. Underpinning many of the gene expression changes are modifications to the epigenetic profile of the associated chromatin. We have used a large-scale shRNA screen to identify epigenetic modifiers that act as barriers to reprogramming. We have uncovered an important role for TRIM28 in cells resisting transition between somatic and pluripotent states. TRIM28 achieves this by maintaining the H3K9me3 repressed state and keeping endogenous retroviruses (ERVs) silenced. We propose that knockdown of TRIM28 during reprogramming results in more plastic H3K9me3 domains, dysregulation of genes nearby H3K9me3 marks, and up-regulation of ERVs, thus facilitating the transition through reprogramming.

INTRODUCTION

Through the ectopic expression of *Oct4*, *Sox2*, *Klf4*, and *c-Myc* (OSKM), a differentiated cell can be reverted into a stem cell state[1]. Since the discovery of induced pluripotent stem (iPS) cells, the mechanisms that allow a differentiated cell to become reprogrammed into a stem cell state have been active areas of research. Of particular interest are the alterations to the epigenetic state of the chromatin during reprogramming. Epigenetic modifications can alter the chromatin structure in several ways, including DNA methylation and discrete alterations of histone modifications. Both of these modifications dictate whether chromatin will be in an active (euchromatin) or repressed (heterochromatin) state. Changes in H3K9me2/3, H3K27me3, and H3K4me3 have been demonstrated to influence the transcription of particular important subsets of genes during the reprogramming process[2–4]. It has been further demonstrated that specific epigenetic modifications pose a barrier to differentiated cells obtaining pluripotent potential. Early studies demonstrated that the addition of histone deacetylase (HDAC) inhibitors valproic acid or butyrate increased the efficiency of reprogramming, indicating that histone acetyl-transferases play an important role in maintaining somatic cell identity[5–7].

More recently, several groups using various reprogramming systems have identified multiple epigenetic modifiers, which function to either block or enhance reprogramming. These studies targeted either a single gene or a handful of genes representing a subset of histone modifiers[3,8–12]. As each of these studies utilized different systems to both initiate and evaluate reprogramming, there is the potential for inconsistencies to arise[13]. The most effective way to overcome these potential complications is to use a large-scale screen targeting epigenetic modifiers in a single reprogramming system. Using such an approach, it has recently been discovered that *Caf1* and *Ube2i* function to maintain somatic cell identity during reprogramming[14]. Though this particular screen contained many more epigenetic modifiers than previous attempts, there is still the potential for specific genes and/or shRNAs to be underrepresented.

Recently it has also been shown that transcriptional control of transposable elements during reprogramming is important for the downstream differentiation of functional human iPS cell lines[15]. Transposable elements comprise 45%–55% of the genome. Expression of RNA transposons (retroviruses) is repressed by the binding of the epigenetic modifier Tripartite motif-containing 28 (TRIM28) to specific zinc finger proteins. This complex then recruits the methyltransferase SETDB1 to establish and maintain the repressive epigenetic mark H3K9me2/3. TRIM28 also allows the recruitment of HP1-gamma to these elements and thereby initiates the formation of heterochromatin, adding an additional layer to their transcriptional repression[16]. *Trim28* knockout in embryonic stem (ES) cells and neuronal progenitor cells results in transcription of specific endogenous retroviruses (ERVs), which are usually repressed[17,18]. Furthermore, it has also been observed that during the various stages of reprogramming, particular transposable elements increase their expression in both mouse and human reprogramming systems[19].

Using a large-scale shRNA screen targeting approximately 670 epigenetic modifiers during the reprogramming of immortalized primary mouse embryonic fibroblasts (MEFs), we have identified, for the first time, *Trim28* as a major epigenetic barrier to reprogramming. Consistent with previous reports, we also identify Setdb1 as an additional epigenetic barrier. We have found that Trim28 represses the expression of ERVs during the reprogramming process. Upon *Trim28* knockdown during reprogramming, specific ERVs and the genes surrounding these regions switch from being repressed to being highly transcribed. These upregulated regions are also located in proximity to H3K9me3 marks indicating that *Trim28* safeguards the differentiated state of somatic cells by maintaining the repression of these regions. Thus *Trim28* underpins a strong epigenetic barrier during cellular reprogramming.

RESULTS

An shRNA screen targeting epigenetic modifiers reveals *Trim28* and *Setdb1* as barriers to cell reprogramming

In order to investigate the importance of epigenetic modifiers during cell reprogramming in an unbiased manner, we performed a pooled shRNA screen with an epigenome library containing 3,003 shRNAs targeting around 670 unique genes **(Fig 1A)**. In order to obtain a sufficient number of cells for the pooled shRNA screen, primary *Oct4*-GFP reporter MEFs were immortalized by knockdown of *p53*[20,21] and subsequently infected with a dox inducible OSKM construct along with the reverse tetracycline transactivator M2rtTA[22,23]. These *Oct4*-GFP p53KD MEFs were then infected with the epigenome shRNA library. After 48 hours of puromycin selection, the cells were switched to ES cell conditions on a layer of feeder MEFs. Dox was added to the medium for 5 days to induce reprogramming. To ensure the *Oct4*-GFP reporter could be maintained without the induction of OSKM, cells were grown for an additional 2 days without dox. On day 7 cells were sorted into GFP positive and GFP negative populations. From these populations, genomic DNA was extracted, shRNAs recovered by PCR, and their relative abundance was determined using deep sequencing. Analysis of the relative abundance of the shRNAs in the GFP positive population revealed that three shRNAs were significantly ($padj \leq 0.1$) enriched with a $\log_2$ fold change of more than 2.5 compared to the GFP negative sample **(Fig 1B)**. Of the three shRNAs, two targeted *Trim28* and one targeted *Setdb1*. TRIM28 is known to interact with SETDB1 to establish the repressive H3K9me3 histone mark suggesting that H3K9me3 is a major barrier to reprogramming[24]. *Setdb1* has also previously been demonstrated to be a barrier to cell reprogramming[10].

Trim28 knockdown enhances reprogramming

To validate that the three shRNAs identified in the screen target their respective genes, we analyzed the RNA expression of *Trim28* and *Setdb1* in *Oct4*-GFP MEFs. The shRNAs targeting *Trim28* reduced its expression by 85% (KD1) and 87% (KD2) when compared

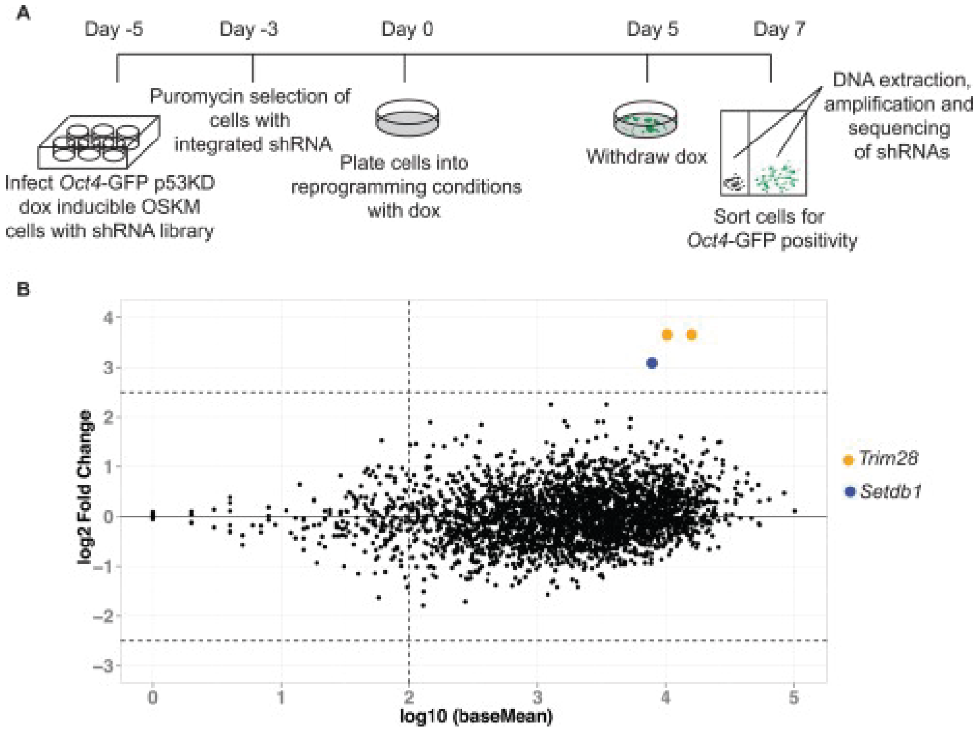


Figure 1. shRNA screening identifies *Trim28* as a barrier to cell reprogramming. (A) Diagram outlining the procedure used for the shRNA screen. (B) MA plot of shRNAs and their abundance in the *Oct4*-GFP positive population versus the *Oct4*-GFP negative population.

to the scrambled shRNA control **(Fig 2A)**. Setdb1 expression was reduced by 75% when compared to the scrambled shRNA control **(Fig 2A)**. These results were further confirmed at the translational level by western blot analysis where knockdown of *Trim28* or *Setdb1* in *Oct4*-GFP MEFs resulted in a marked decrease in their respective protein levels **(Fig 2B)**.

To validate that *Trim28* knockdown could enhance reprogramming, we knocked down *Trim28* and *Setdb1* individually and in combination in p53KD *Oct4*-GFP MEFs, reflecting the original screening conditions, and in wild type *Oct4*-GFP MEFs **(Figs 2C, 2D; Supplementary Figure 1C)**. Knockdown of *Trim28* enhanced reprogramming 17-fold (KD1, 49%) and 21-fold (KD2, 59.4%) in p53KD *Oct4*-GFP MEFs, compared to a scrambled shRNA control (2.8%), as quantified by the percentage of *Oct4*-GFP positive cells **(Fig 2C)**. As previously demonstrated, *Setdb1* knockdown also enhanced the reprogramming efficiency 20-fold (57%) **(Fig 2C)**[10]. A small but significant increase in the number of *Oct4*-GFP positive cells could be observed in p53 KD MEFs when knocking down *Trim28* KD1 and *Setdb1* together (69%, $p = .0045$) in p53 KD MEFs compared to knockdown of *Trim28* KD1 alone. Although not significant, an increase was also observed with *Trim28* KD2 and *Setdb1* together compared to *Trim28* KD2 alone. Importantly, although at

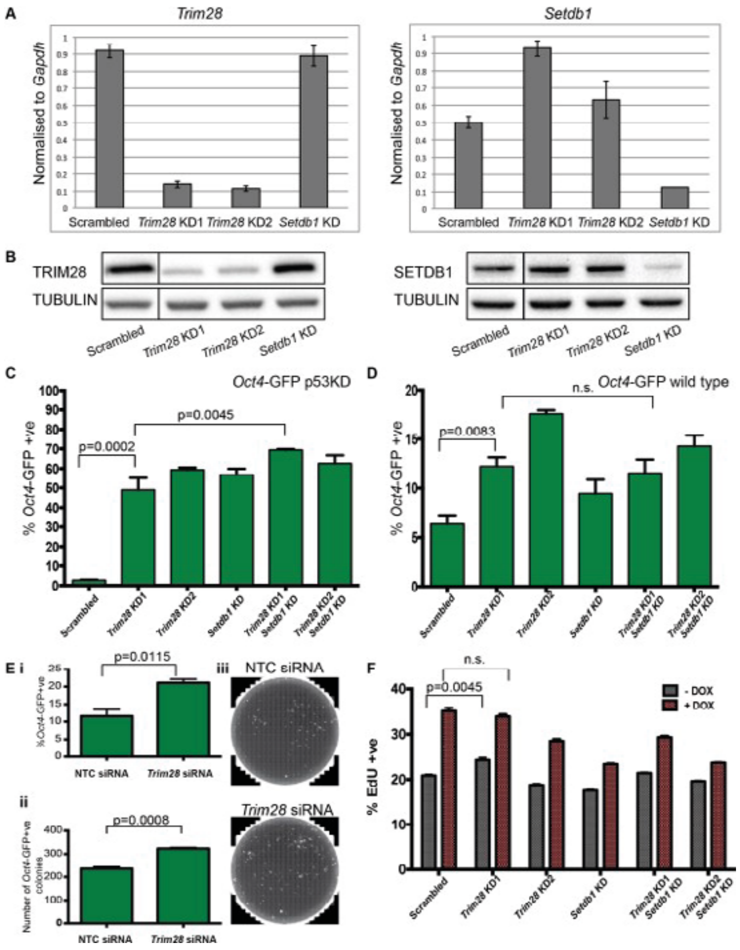


Figure 2. Knockdown of *Trim28* increases reprogramming. (A) qRT-PCR analysis of *Trim28* and *Setdb1* in *Oct4*-GFP MEFs after 7 days of reprogramming containing either a scrambled shRNA or shRNAs targeting *Trim28* or *Setdb1*. Gene expression was normalized to *Gapdh*. *n* = 3 and error bars = SEM. (B) Western blot analysis of TRIM28, SETDB1, and TUBULIN in *Oct4*-GFP MEFs after 4 days of shRNA knockdown. (C) Quantification of *Oct4*-GFP positive (+ve) cells after 7 days of reprogramming with p53KD (2 days – dox) or (D) wild type (7 days + dox), *n* = 3 and error bars = SEM). Student's *t*-test was used to determine significance, n.s. = not significant (E) Quantification of wild type *Oct4*-GFP cells after 7 days of reprogramming using either flow cytometry (i) or colony counting (ii) (7 days + dox) *n* = 3 and error bars = SEM. Student's *t*-test was used to determine significance. (iii) Representative images of wells used for colony counting (F) Quantification of EdU positive cells after 4 days in embryonic stem cell conditions with and without dox induction of OSKM. *n* = 3 and error bars = SEM. Student's *t*-test was used to determine significance, n.s. = not significant. Abbreviation: SEM, standard error of the mean.

a lower fold increase, similar results were observed in wild-type *Oct4*-GFP MEFs **(Fig 2D)**. Knockdown of *Trim28* significantly enhanced reprogramming 2-fold (KD1, 12%) and 3-fold (KD2, 18%) compared to a scrambled shRNA control (6%) in wild type *Oct4*-GFP MEFs. However, there was no significant additional increase in the percentage of *Oct4*-GFP positive cells when *Trim28* KD1 and *Setdb1* were combined (12%, p = .6984) compared to *Trim28* KD1 alone **(Fig 2D)**. Crucially, knockdown of *Trim28* and *Setdb1* without dox did not stimulate the expression of the *Oct4*-GFP reporter transgene, indicating that knockdown of *Trim28* or *Setdb1* does not interfere with its expression **(Supplementary Figure 1A)**. We also observed no increased expression of the OSKM lentiviral vector when *Trim28* or *Setdb1* was depleted during reprogramming compared to a scramble control **(Supplementary Figure 1B)**.

To analyze whether the temporary knockdown of *Trim28* was sufficient to increase the efficiency of reprogramming, we also used siRNA to knock down *Trim28* during the first few days of reprogramming in wild-type *Oct4*-GFP. Knockdown of *Trim28* compared to a non-targeting control (NTC) siRNA was confirmed using qRT-PCR prior to the induction of reprogramming **(Supplementary Figure 1D)**. In two separate experiments, cells were analyzed using either flow cytometry or colony counting after 7 days of reprogramming **(Supplementary Figure 1E)**. It was observed that when *Trim28* was knocked down, albeit temporarily, there was a significant increase in reprogramming efficiency of 1.8-fold and 1.4-fold, respectively, compared to NTC siRNA **(Fig 2E)**. These results support and validate the veracity of the initial screening system and reinforce TRIM28 as a novel barrier to cell reprogramming. To further characterize how TRIM28 functions as a barrier to reprogramming, we used wild-type *Oct4*-GFP with shRNA *Trim28* knockdown for the remainder of the study.

Loss of *Trim28* enhances reprogramming independent of cell proliferation

One known way to enhance the efficiency of cell reprogramming is to increase cell proliferation. We used EdU to ascertain whether knockdown of *Trim28* or *Setdb1* would increase the proliferation of wild-type *Oct4*-GFP MEFs **(Fig 2F)**. The knockdown of *Trim28* KD1 led to a small but significant increase (4%) in EdU positive cells relative to the scrambled control without dox (p = .0045). However, this was not observed with *Trim28* KD2. Furthermore, when dox was added and reprogramming induced, no significant change in proliferation was observed when comparing *Trim28* KD1 to the scrambled control (p = .2048). *Trim28* KD2 and *Setdb1* KD both showed a decrease in the percentage of EdU positive cells when compared to the scrambled control under the presence or absence of dox. These results indicate that TRIM28 does not influence cell proliferation during reprogramming.

Trim28 knockdown regulates the expression of genes located in repressive chromatin regions during reprogramming

To understand how *Trim28* KD increases cell reprogramming, we performed RNA sequencing. We bulk harvested RNA from wild-type *Oct4*-GFP MEFs after 7 days of OSKM induction, containing either scrambled, *Trim28* KD1 alone or *Trim28* KD1 combined with *Setdb1*KD. We also included wild-type *Oct4*-GFP MEFs containing scrambled shRNA without OSKM induction (nonreprogramming) as an additional control. Using DESeq we constructed a heatmap showing the Euclidean distances between the samples using the variance stabilizing transformation of the data **(Fig 3A)**. This analysis showed that the scrambled no dox (nonreprogramming) samples clustered away from the cells with the induction of OSKM (reprogramming). To validate our reprogramming system, we constructed an MA plot to look for differences in gene expression between reprogramming and non-reprogramming samples. As expected, in the samples that had OSKM induced, we observed an up-regulation of several genes involved in the pluripotency network, including *Nanog* (Log_2 Fold Change (FC) = 4.5), *Lin28b* (Log_2 FC 5.6), *Esrrb* (Log_2 FC 6.3), *Sall4* (Log_2 FC 8.9), and *Fgf4* (Log_2 FC 7), *Oct4* (lentiviral and endogenous, Log_2 FC 4.6), *Sox2* (lentiviral and endogenous, Log_2 FC 3.9) **(Supplementary Figures 2A, 2B)**.

In order to elucidate the impact of *Trim28*KD on gene expression during reprogramming, we compared the expression of *Trim28*KD to the scrambled control after 7 days of OSKM induction. Surprisingly, we observed that when *Trim28* is knocked down during reprogramming, only 143 genes had an increase in expression with a Log_2 fold change greater than 2 **(Figs 3B, 3C, Supplementary Table 1)**. Analysis of this list of genes using gene ontology (GO) revealed that the most statistically significant GO term was "unclassified," and no other statistically significant GO terms were observed. This result indicates that *Trim28* does not regulate a specific pathway during reprogramming.

Considering the up-regulation of 143 genes in *Trim28*KD cells and the knowledge that TRIM28 is known to modulate H3K9me3, we reasoned that these genes may be located nearby chromatin areas carrying H3K9me3 modifications. Using publicly available ChIP-Seq data for MEFs and ES cells, we searched for genes that lie within 1kb of H3K9me3 peaks[25,26]. We intersected this list with our list of 143 genes and found that 19 genes (MEFs, $p = 1.7e-4$) and 77 (ES cells, $p = 1.6e-14$) of the 143 genes were located within 1kb of H3K9me3 marks **(Fig 3D)**.

Further to this we analyzed the log_2 fold change in gene expression (*Trim28*KD vs. scrambled), which were located either far (>80 kb) or near (<40 kb) H3K9me3 peaks from both MEFs and ES cells **(Fig 3E)**. Using the Wilcoxon test, we observed a statically significant increase in expression in genes located near H3K9me3 peaks in *Trim28*KD samples. Indicating that when *Trim28* is knocked down during reprogramming, genes nearby H3K9me3 peaks increase their expression more than those located far away.

To uncover whether loss of Trim28 had an overall effect on global H3K9me3 levels, we used immunofluorescence to detect H3K9me3 in wild-type *Oct4*-GFP MEFs

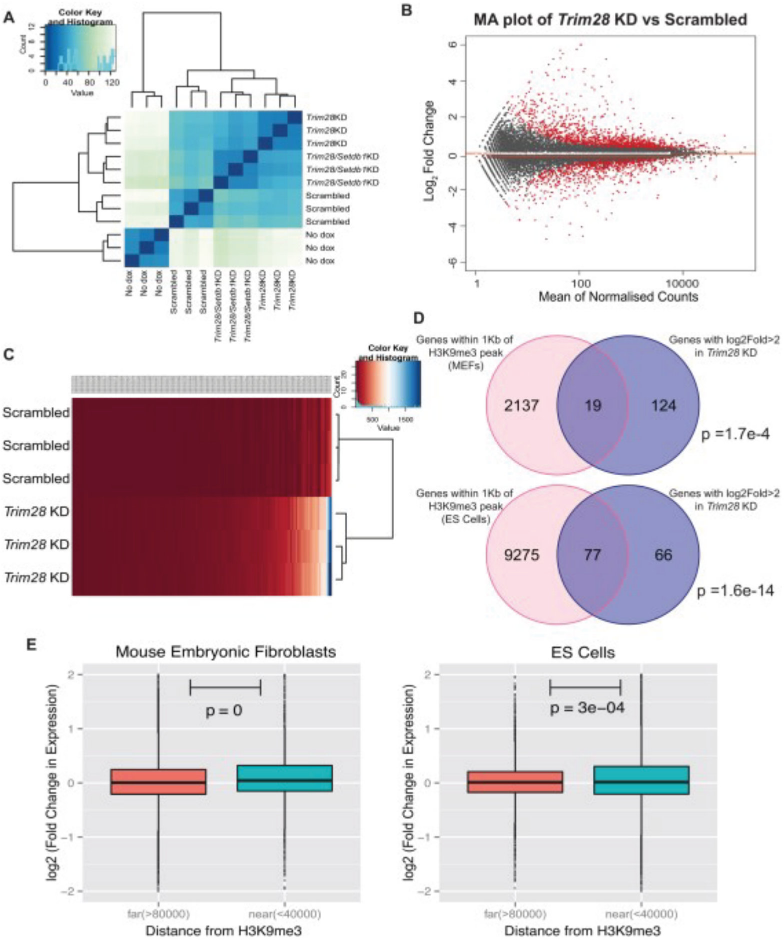


Figure 3. Knockdown of *Trim28* during reprogramming alters expression of genes nearby H3K9me3 peaks. (A) A heatmap of RNA Seq samples showing the Euclidean distances between the samples using the variance stabilizing transformation of the data (created using DeSeq). (B) mean difference (M) vs. average expression (A) plot of differential gene expression in *Trim28*KD versus scrambled cells at day 7 of reprogramming. (C) Heatmap of normalized counts of the 143 genes significantly (padj < 0.1) upregulated genes with a $\log_2$ fold change >2, in *Trim28*KD cells versus scrambled cells. Also, see Supporting Information Table 3. (D) Venn diagram showing the chromatin context of the 143 genes upregulated with a $\log_2$ fold change >2, in *Trim28*KD cells versus scrambled cells. H3K9me3 peak data obtained for MEFs and ES cells. Hypergeometric testing was used to determine significance. (E) Boxplots showing the relative change in gene expression (between *Trim28*KD and scrambled samples) situated near (<40 kb) or far (>80 kb) from H3K9me3 peaks using H3K9me3 peak data for MEFs and ES cells. Wilcoxon testing was used to determine significance. Abbreviations: MEFs, mouse embryonic fibroblasts; ES, embryonic stem cell.

with or without *Trim28*KD. Though the knockdown of *Trim28* was detected using this method **(Supplementary Figure 3A)** we did not observe a complete loss of H3K9me3, HP1 gamma foci, nor of 5 methylcytosine **(Supplementary Figures 3A-3C)**. These results indicate that knockdown of *Trim28* alone is not sufficient to erase global H3K9me3 modifications in MEFs.

It has previously been shown that knockdown of *Setdb1* can lead to a trend of decreased H3K9me3 levels near the *Nanog* locus; however, the transcription of *Nanog* was not analyzed in this study[10]. We, therefore, performed qRT-PCR to see if knockdown of *Setdb1* or *Trim28* influenced the expression of *Nanog*. Using this technique, we did not observe a significant change in *Nanog* expression upon *Trim28*KD; however, we did observe a statistically significant decrease with *Setdb1*KD **(Supplementary Figure 3D)**.

Trim28 **knockdown upregulates ERVs during reprogramming**

It is well established that one of the key functions of TRIM28 is to repress transposable elements and particular ERVs in mouse embryonic stem and neuronal progenitor cells[17,18]. Previously, it was demonstrated, by qRT-PCR analysis, that specific ERVs increase their expression during reprogramming[19]. We hypothesized that knockdown of *Trim28* may alter the expression of ERVs during reprogramming. Using RepEnrich analysis, we analyzed our RNA Seq data expression changes of ERVs during reprogramming[27]. Indeed, when we compared the scrambled control with OSKM induced (reprogramming) to the scrambled control without OSKM induced (non-reprogramming) at day 7, we observed alterations in ERV expression, confirming previous data **(Supplementary Table 2)**.

Interestingly, when we compared *Trim28*KD to the scrambled control during reprogramming, we detected a significant increase in the expression of 76 ERVs **(Fig 4A, Supplementary Table 3)**. Intriguingly, the vast majority of repetitive elements that increased their expression belonged to the long terminal repeat (LTR) class. Some of these LTR class repeats are also located nearby genes, which were upregulated in *Trim28*KD reprogramming cells and located within 1 kb of H3K9me3 peaks **(Supplementary Figure 4B)**. Breaking down this list further into ERV class and family, we observed that 50 of the 76 upregulated ERVs belong to the LTR/ERVK family. The next most common family, which was upregulated, was the LTR/ERV1 family (16 out of 76). Previously, it was shown TRIM28 knockout affects the expression of the intracisternal A-particle (IAP) and MMERVK10c elements in mouse ES cells[18]. In line with this observation, we observed that all IAP and the MMERVK10c elements increased their expression with TRIM28 knockdown during reprogramming. This result was also confirmed using qRT-PCR **(Fig 4B)**. A single LINE/L1 repeat was upregulated in *Trim28*KD cells. When we assayed LINE1 elements using qRT-PCR, we observed a modest increase in its expression. Together, the RepEnrich analysis and qRT-PCR data indicate that TRIM28 is crucial to suppress the activation of specific repeat elements during reprogramming. It is especially important in maintaining repression of IAP and MMERVK10c repetitive elements.

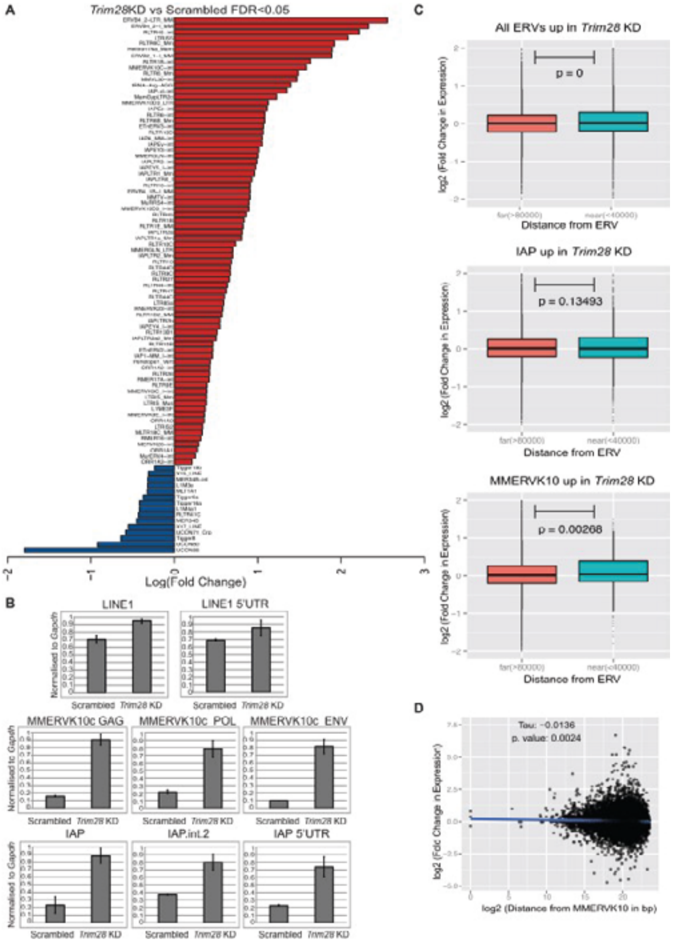


Figure 4. Trim28 controls the expression of repetitive elements during reprogramming. (A) RepEnrich analysis of repetitive elements in reprogramming *Trim28*KD versus scrambled after 7 days of reprogramming. Also see Supporting Information Table 5. (B) qRT-PCR analysis of repetitive elements LINE1, MMERVK10c, and IAP, normalized to *Gapdh* $n = 3$ and error bars = standard error of the mean. (C) Boxplots showing the relative change in gene expression (*Trim28*KD vs. scrambled) situated near (<40 kb) or far (>80 kb) from all upregulated repetitive elements, upregulated IAP elements, and upregulated MMERVK10 elements. Wilcoxon testing was used to determine significance. (D) Scatter plot showing the relative change in gene expression (*Trim28*KD vs. scrambled) against the distance from MMERVK10 elements. Regression line is shown in blue. Kendall method was used to calculate association and subsequent *p*-value.

To analyze whether this increase of expression of ERVs had an impact on gene expression, we used a similar approach to Friedli et al. and compared the $\log_2$ fold change in gene expression between *Trim28*KD and scrambled samples and related this change in expression to distance from the upregulated ERVs. When we grouped all upregulated repetitive elements (>300bp) together, we observed an increase in the overall fold change of genes located nearby these elements when *Trim28* is knocked down during reprogramming **(Fig 4C)**. Surprisingly, when we specifically looked at IAP (>300bp) elements, this difference was no longer statistically significant. However, when we analyzed the MMERVK10 elements (>300bp) we once again saw an increase in the expression of genes located near these elements compared to those far away. Furthermore, a statistically significant ($p = .0024$) negative correlation ($Tau = -0.0136$) was observed between changes in gene expression between *Trim28*KD and scrambled control and distance from the MMERV10 elements **(Fig 4D)**. Together these results indicate that the genes located nearby the repetitive elements which are upregulated with *Trim28*KD also upregulate their expression. However, the influence of derepression of repetitive elements containing chromatin regions on gene expression may be larger from some repetitive elements than others.

DISCUSSION

The maintenance of a somatic cell state is controlled by both transcription factors and the epigenetic modification of the chromatin[4,28,29]. Cells undergoing reprogramming to reach an iPS cell state are able to reset these properties, but the process is relatively inefficient. Through the use of a large-scale shRNA screen targeting a comprehensive list of epigenetic modifiers, we have discovered that Trim28 is a novel barrier to cell reprogramming. We also identified a previously known epigenetic barrier to reprogramming *Setdb1*, thereby validating our screening approach[10]. In a coordinated fashion, TRIM28 is known to bind SETDB1 to allow the maintenance of the repressive H3K9me3 epigenetic mark. This work also complements the recent study, which showed that CAF-1, a protein which is also known to interact with SETDB1 and is frequently co-bound to TRIM28 chromatin binding sites, plays an important role in the accessibility of chromatin to transcription factors during reprogramming[14]. We have shown that *Trim28* knockdown in immortalized MEFs results in a substantial increase in the percentage of cells that undergo successful reprogramming as determined using the *Oct4*-GFP reporter. This increase in reprogramming was also confirmed in the wild-type MEFs, albeit less pronounced when compared to immortalized MEFs. One possibility for this reduced efficiency is that it is well established that TRIM28 knockout ES cells do not survive under typical ES culture conditions due to the uncontrolled expression of ERVs[17]. Therefore, it is highly likely that many wild-type MEFs with high levels of *Trim28* knockdown undergo apoptosis when reprogrammed into iPS cells, whereas cells compounded with p53 KD can survive Trim28 knockdown when reprogrammed.

Setdb1 has previously been identified as a barrier to cell reprogramming[10]. As *Setdb1* was identified and validated in our study, this proves the veracity of our screening method. A number of other epigenetic modifiers, which have been reported to be barriers to cell reprogramming, were not seen to enhance reprogramming in our screen. In particular, we did not detect *Suv39h1/2*, *Dot1L*, or *Mbd3*[8–10]. This is most likely due to the fact that these previous screens use different systems of reprogramming, such as partially reprogrammed cells, secondary reprogramming MEFs or human fibroblasts. Supporting this hypothesis, it has recently been shown that different reprogramming systems have different roadblocks to reprogramming[30]. There are also limitations for the screening system used in this study, such as under-representation of specific shRNAs and off-target effects. However, supporting a role for *Trim28* as a barrier to cell reprogramming, it was recently demonstrated that knockdown of *Caf-1* and *Ube2i*, which areknown to interact with TRIM28 either directly or through *Setdb1*, enhanced reprogramming[14,31]. Though this recent screen did not identify *Trim28* as a target, this was likely due to poor representation of the shRNAs targeting *Trim28*, with only one shRNA out of nine represented consistently in the screen. Additionally, using the scoring system provided in the analysis, this single hairpin is ranked similarly to some hairpins targeting CAF-1 as enriched in *Oct4*-GFP positive cells[14]. Likewise in this study two shRNAs targeting *Chaf1a* (*Caf-1*) were seen to be enriched in the GFP positive population, with a $\log_2$ fold change greater than 1; however, they were not statistically significant. Overall these screening results support the notion that the H3K9me3 barrier is the most critical epigenetic barrier to cell reprogramming.

It is well established that an increase in reprogramming efficiency can occur when cells have an increased proliferation rate[32,33]. When we analyzed whether inhibition of *Trim28* had an effect on cell proliferation, we did not observe a change in proliferation rate, indicating that this was unlikely to be the cause of enhanced reprogramming. Importantly, we also did not observe an increase in the expression of the OSKM lentiviral vector, indicating TRIM28 does not influence the expression of this construct during reprogramming. Another method to increase reprogramming is to alter the EMT pathway. Onder et al. have previously shown that inhibition of the epigenetic modifier *Dot1L*, which controls the H3K79 methylation, increases the efficiency of reprogramming of human fibroblasts by modulating the EMT phase of reprogramming. Our RNA Seq analysis of gene expression did not reveal any consistent changes in the expression of this set of EMT genes with *Trim28* knockdown, indicating that alterations in the EMT are unlikely to be the cause of *Trim28*KD enhanced reprogramming. Interestingly, we did not observe an increase in the expression of other genes typically associated with a statistically significant GO term. Though it was previously demonstrated that knockdown of *Setdb1* led to a decrease in H3K9me3 near the *Nanog* locus[10], we did not observe an increase in *Nanog* expression with *Setdb1* knockdown during reprogramming.

We interrogated our RNA Seq data further to better understand how knockdown of *Trim28* increases cell reprogramming. Surprisingly, only a small subset of genes was upregulated upon *Trim28* knockdown when compared to the scrambled shRNA control cells. With the most significant GO term being unclassified and knowing that Trim28 functions to maintain repressive chromatin, we reasoned that these gene expression changes were likely to be related to chromatin location rather than a specific gene or signaling network. By intersecting our list of 143 upregulated genes with genes that are located nearby H3K9me3 marks in either MEFs or ES cells we demonstrated that a significant proportion of these genes are indeed located in these regions. We also determined that genes are significantly more likely to be upregulated in *Trim28*KD samples during reprogramming if located nearby these H3K9me3 regions. Together, this indicates that knockdown of *Trim28* during reprogramming leads to a derepressed chromatin state.

The role of Trim28 has been studied extensively in ES cells and neuronal progenitor cells[17,18,34]. It is well established that TRIM28 controls and maintains the repressive chromatin environment around ERVs. Additionally, it has been demonstrated that the control of ERVs occurs during both mouse and human cell reprogramming[19]. Friedli et al. have shown that during the reprogramming process, specific repetitive elements increase their expression to levels similar to that observed in ES cells. Recently it was demonstrated that precise repression of human repetitive elements is also required for the differentiation of iPSCs[15]. Continued expression of *KLF4* led to the maintained expression of the HERVK retroviruses, which caused defects in iPSC differentiation. Once KLF4 expression was attenuated, the iPSCs were able to differentiate. Our data show that knockdown of *Trim28* leads to an increase in the expression of repetitive elements during reprogramming, in particular, the IAP and MMERVK10c elements. We also show that genes located nearby MMERVK10 elements in *Trim28* knockdown samples are more likely to be upregulated than genes located far away from these elements. This is potentially due to loss of H3K9me3 maintenance at these sites, with the expression of these elements indicating a more de-condensed and open chromatin state leading to increased reprogramming potential **(Fig 5)**. However, as ERVs must be tightly controlled during reprogramming and differentiation, temporary knockdown of *Trim28* is a more suitable approach for the production of iPS cell lines.

A recent publication showed that ERVs fluctuate in their expression during the development of mouse and human embryos[35]. Moreover, the endogenous retrovirus MuERVL has also been associated with the pluripotent state of cells at the earliest stages of embryo development[36], indicating that expression of ERVs may be indicative of a cell's pluripotent potential. It has also been demonstrated that during somatic cell nuclear transfer (SCNT), which is another method used to reprogram a somatic cell into an ES cell-like state, reprogramming resistant regions (RRRs) are present at H3K9me3 enriched sites throughout the genome[37]. This work further demonstrated that ectopic expression of *Kdm4d* during SCNT enhanced the efficiency of reprogramming through the depletion of H3K9me3 marks at RRRs. Though it would be interesting to compare whether the genes we identified as upregulated with *Trim28*KD during iPS cell reprogramming are located within these RRRs, the chromosomal locations for the RRRs are, unfortunately, not publicly available. It is of note that a role for Trim28 in controlling ERVs is only seen when cells are not fully differentiated. This observation is further observed in our study where *Trim28*KD has an important role during the reprogramming of MEFs, which is in direct contrast to its

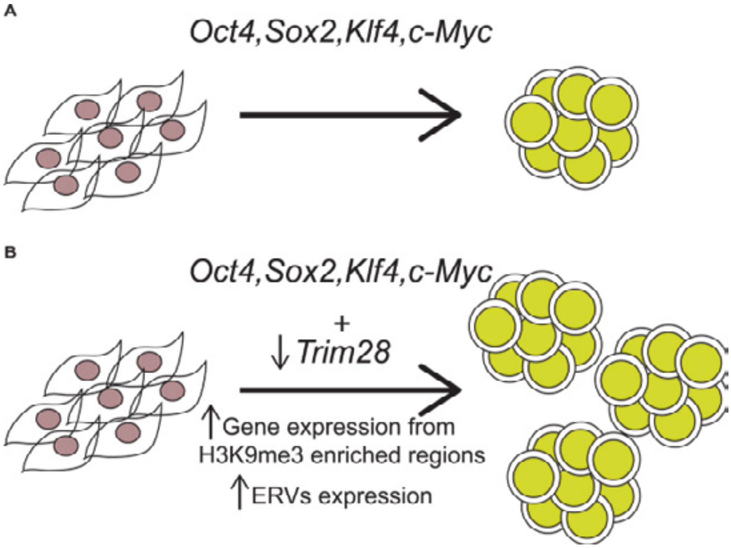


Figure 5. Model of TRIM28 function during reprogramming. (A) Ectopic expression of *Oct4*, *Sox2*, *Klf4,* and *c-Myc* (OSKM) induces the reprogramming of somatic cells into induced pluripotent stem cells at low efficiency. (B) Ectopic expression of OSKM along with knockdown of TRIM28 increases the expression of genes nearby H3K9me3 and induces the expression of endogenous retroviruses. This decondensed chromatin state increases the reprogramming efficiency.

lack of a function in differentiated MEFs[34]. Hence, in future work, it would be of interest to functionally validate a role for TRIM28 in other in vitro reprogramming methods.

CONCLUSION

In conjunction with previous work, this study now provides a better understanding of how alterations to the epigenetic landscape contribute to both maintaining the somatic cell state and resetting the pluripotent potential of a cell. From our large-scale screen, we have confirmed H3K9me3 as the major epigenetic barrier to cell reprogramming with a key function of TRIM28 in maintaining this barrier. Knockdown of *Trim28* during reprogramming results in increased expression of genes nearby H3K9me3 peaks and upregulation of specific ERVs, indicating a de-condensed and active chromatin state that facilitates the transition through reprogramming.

MATERIALS AND METHODS

Animals

Oct4-green fluorescent protein (GFP) transgenic mice[20] were housed and handled according to national regulations. The NKI animal ethics committee approved all animal experiments.

Reprogramming

For the shRNA screen, *Oct4*-GFP MEFs were expanded and infected with an shRNA targeting *p53*[21]. For validation of the screen primary, *Oct4*-GFP MEFs were collected at E13.5. *Oct4*-GFP MEFs were infected with TetO-FUW-OSKM and FUW-M2rtTA constructs (a gift from Rudolf Jaenisch Addgene #20321 and Addgene #20342, respectively)[22,23]. *Oct4*-GFP MEFs were plated onto feeders 4 days after infection in ES cell media (DMEM, 15% fetal bovine serum, LIF, NEAA, beta-mercaptoethanol, L-glutamine, and penicillin-streptomycin) with doxycycline (dox) at a final concentration of 1 µg/ml. shRNAs used for validation were MISSION pLKO.1-puro non-mammalian shRNA Control Plasmid DNA SHC002 (Sigma), pLKO targeting *Trim28* TRCN0000071366 (KD1), TRCN0000071363 (KD2), pLKO targeting *Setdb1* TRCN0000092975. siRNAs used were SMARTpool: ON-TARGETplus *Trim28* siRNA and ON-TARGETplus non-targeting Control Pool (Dharmacon).

Pooled shRNA Screen

Immortalized *Oct4*-GFP MEFs containing dox inducible OSKM were transduced with the epigenome shRNA library containing 3003 shRNAs targeting approximately 670 genes at an MOI = 0.3, to allow for a single integration per cell. The number of cells in all steps of the screens was sufficient to maintain a 300-fold complexity of the epigenome library. The transduced cells were treated with puromycin (8 µg/ml) for cells with integrated lentiviruses. After 48 hours, the selected cells were plated onto 10-cm plates into reprogramming conditions (see reprogramming) for 5 days in the presence of dox followed by an additional 2 days without dox. The reprogramming cultures were harvested and resuspended as a single-cell suspension in 1% fetal calf serum in phosphate-buffered saline and used for cell sorting. The screen was performed in six biological replicates.

Flow Cytometry

For the pooled shRNA screen *Oct4*-GFP positive and negative cells were sorted based on GFP expression using a Becton Dickinson FacsAria Sorter. Further analysis of GFP and proliferation was done using the Becton Dickinson FACS Calibur analyzer.

Amplification of shRNAs, Sequencing and Analysis

Cell pellets were treated with RNAse A (Sigma-Aldrich, R6513-50MG, St. Louis, http://www.sigmaaldrich.com). DNA was extracted using a DNAeasy Blood& Tissue kit (Qiagen, 69506, Hilden, Germany, http://www1.qiagen.com). For 300-fold coverage of the shRNA library, 6 µg of DNA was taken as an equivalent of 1 million cells per sample. shRNAs were amplified as previously described[38] using the Phusion High-Fidelity polymerase chain reaction (PCR) Kit (NEB, E0553L). After PCR amplification, the PCR product was cleaned using a MinElute PCR Purification Kit (Qiagen, 28004) according to the manufacturer's instructions. Samples were sequenced using an Illumina HiSeq2000 platform according to standard procedures. The shRNA sequences were extracted from the sequencing reads and aligned to the epigenome shRNA library. The counts of the matched reads were used for analysis with DESeq2 v1.6.3 using pairwise analysis. The shRNAs with a base mean of ≥100, a $\log_2$ fold change ≥2.5, and an adjusted p-value of ≤ .1 were considered as hits and used for validation and follow-up. Normalized read counts for each shRNA and replicate is provided in Supplementary Table 4.

qRT-PCR

RNA was extracted using standard Trizol (Invitrogen) procedure. One microgram of RNA was DNAse treated (Promega) and reverse transcribed using SuperScript II (Invitrogen) according to the manufacturer's instructions. Real-time PCR was performed using PowerSYBR Green PCR Master Mix 200 reactions AB (Life Technologies, 4368702, Rockville, MD, http://www.lifetech.com) and primers listed in Supplementary Table 5. Samples were run on an Applied Biosystems One Step Plus. Samples were assayed in biological triplicate and technical duplicate. Samples were normalized to murine *Gapdh*.

RNA Seq

Sequencing was performed using the Illumina HiSeq2000 platforms according to standard procedures. Data analysis was performed using R (http://www.r-project.org) and Bioconductor. In particular, DeSeq was used for the analysis of differential gene expression in RNA Seq samples. RNA Seq data is available at the Gene Expression Omnibus website http://www.ncbi.nlm.nih.gov/geo/under the accession number GSE80550. RepEnrich was used to analyze repetitive elements, with the exclusion of simple repeats[27]. For analysis of ChIP Data we used the H3K9me3 data set from ENCODE Data Coordination Center Accession: ENCSR857MYS (ES cells) as previously analyzed[25]. For MEF, H3K9me3 data was obtained from[26]. Raw data for GSM1303761 (IgG ChIP-seq) and GSM1303762 (H3K9me3 ChIP-seq) were downloaded from NCBI Sequence Read Archive and aligned to mm9 genome using bwa v0.7.5 with default parameters. H3K9me3 peaks were called with the peak calling algorithm: MACS2 (in broad mode)[39]. Peaks identified by MACS algorithms were used for further analysis.

Western Blot

Western blot was performed according to standard procedures. Protein samples were run on NuPAGE Novex Bis-Tris 4%–12% Protein Gels (Invitrogen). The following antibodies were used: KAP1 (Abcam ab22553, 1:1,000, Cambridge, U.K., http://www.abcam.com, Tubulin (Sigma T9026, 1:10,000), Setdb1 (Thermo Fisher Scientific MA5-15722, 1:1,000). Antibodies were detected using goat anti-mouse and goat anti-rabbit secondary antibodies conjugated to HRP (Life Technologies 62-6520, 656120).

Immunofluorescence

Immunofluorescence was performed as previously described[40]. Cells were grown on coverslips coated with 0.1% gelatin and fixed with 4% paraformaldehyde. An additional antigen retrieval step of 4 N HCl treatment for 10 minutes was included for 5 mC detection. The following antibodies were used: TRIM28/KAP1 (Abcam ab22553, 1:500), H3K9me3 (Abcam ab8898, 1:1,000), HP1-gamma (MAB3450, 1:1,000, https://www.emdmillipore.com), anti-5-methylcytosine (33D3) (Eurogentec BI-MECY-0100, 1:500, Seraing, Belgium, https://secure.eurogentec.com). Antibodies were detected using Alexa Fluor (Invitrogen) goat anti-rabbit 488 and goat anti-mouse 568 (1:1,000).

Cell Proliferation

Cells were plated and grown with and without dox for 4 days. EdU was added at a final concentration of 10 μM for 2 hours. Cells were harvested using standard procedures and assayed using the Click-iT EdU AlexaFluor 647 Flow Cytometry Assay Kit (Life Technologies, C-10424) according to the manufacturer's instructions.

ACKNOWLEDGMENTS

We thank the NKI flow facility for their assistance with cell sorting. We thank the sequencing facility, in particular Ron Kerkhoven and Iris de Rink, for the processing and analysis of sequencing data. We thank the NKI screening facility, in particular Ben Morris, for their help with the shRNA screen. We thank Ekaterina Nevedomskaya for assistance with ChIP-Seq data analysis. We thank Joost Girbnau for the Oct4-GFP MEFs. We thank Prashanth Kumar for his expert advice on shRNA screening. We thank Niek Wit for providing us with the p53 knockdown construct. We acknowledge the ENCODE Consortium and the Hardison laboratory for generating the H3K9me3 dataset. D.C.M. was supported by a National Health and Medical Research Council Early Career Fellowship (no. 1052195).

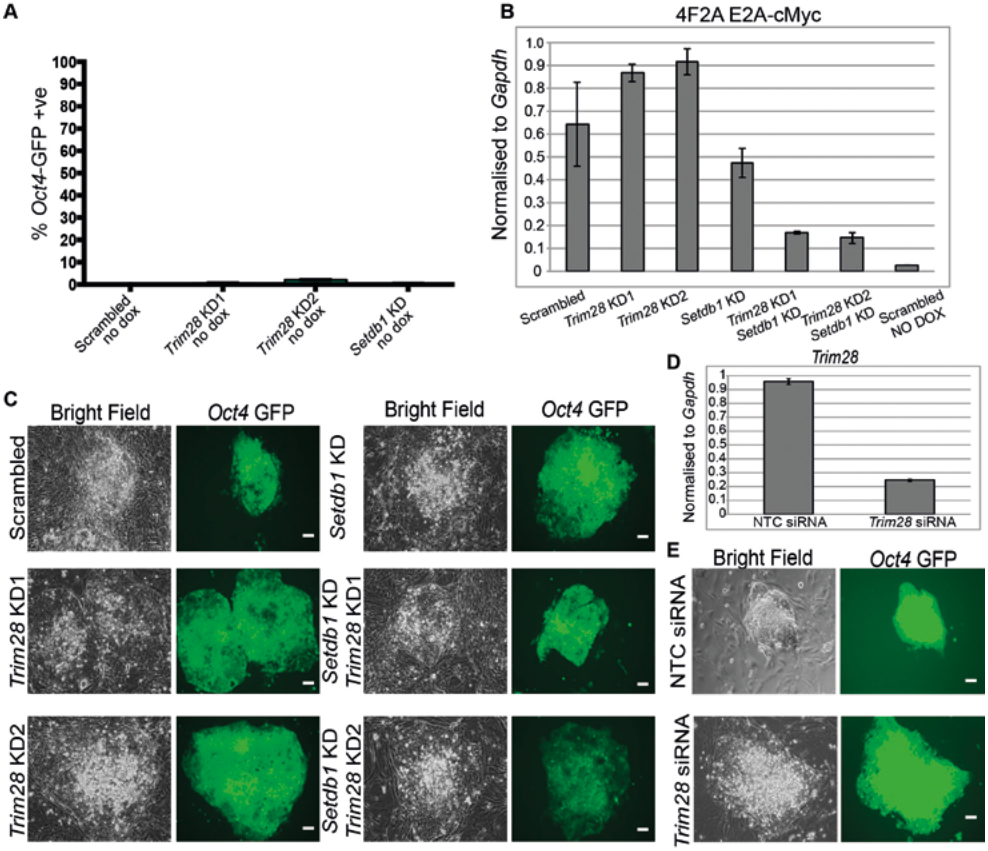
REFERENCES

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
2. Koche, R. P. *et al.* Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8**, 96–105 (2011).
3. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
4. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012).
5. Mali, P. *et al.* Butyrate Greatly Enhances Derivation of Human Induced Pluripotent Stem Cells by Promoting Epigenetic Remodeling and the Expression of Pluripotency-Associated Genes. *Stem Cells* **28**, 713–720 (2010).
6. Liang, G., Taranova, O., Xia, K. & Zhang, Y. Butyrate promotes induced pluripotent stem cell generation. *J. Biol. Chem.* **285**, 25516–25521 (2010).
7. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.* **26**, 795–797 (2008).
8. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483**, 598–602 (2012).
9. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).
10. Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45**, 34–42 (2013).
11. Wang, T. *et al.* The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* **9**, 575–587 (2011).
12. Liang, G., He, J. & Zhang, Y. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nat. Cell Biol.* **14**, 457–466 (2012).
13. Cusulin, C. *et al.* Precursor States of Brain Tumor Initiating Cell Lines Are Predictive of Survival in Xenografts and Associated with Glioblastoma Subtypes. *Stem cell reports* **5**, 1–9 (2015).
14. Cheloufi, S. *et al.* The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218–224 (2015).
15. Ohnuki, M. *et al.* Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12426–12431 (2014).
16. Sripathy, S. P., Stevens, J. & Schultz, D. C. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol. Cell. Biol.* **26**, 8623–8638 (2006).
17. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
18. Fasching, L. *et al.* TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep.* **10**, 20–28 (2015).
19. Friedli, M. *et al.* Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res.* **24**, 1251–1259 (2014).
20. Yoshimizu, T. *et al.* Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev. Growth Differ.* **41**, 675–684 (1999).
21. Dirac, A. M. G. & Bernards, R. Reversal of senescence in mouse fibroblasts through lentiviral suppression of p53. *J. Biol. Chem.* **278**, 11731–11734 (2003).
22. Carey, B. W. *et al.* Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 157–162 (2009).
23. Hockemeyer, D. *et al.* A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* **3**, 346–353 (2008).
24. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. SETDB1:

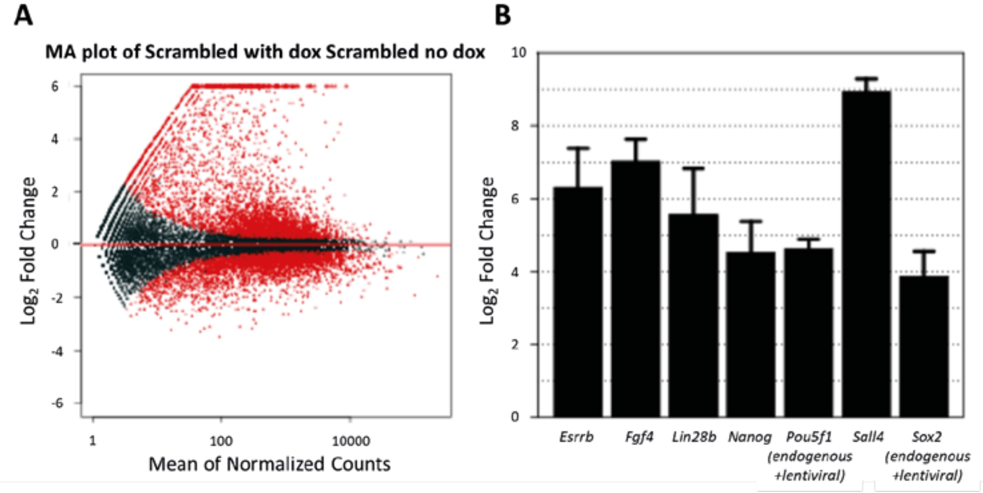
a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932 (2002).

25. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
26. Pedersen, M. T. *et al.* The demethylase JMJD2C localizes to H3K4me3-positive transcription start sites and is dispensable for embryonic development. *Mol. Cell. Biol.* **34**, 1031–1045 (2014).
27. Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M. & Neretti, N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* **15**, (2014).
28. Li, R. *et al.* A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. *Cell Stem Cell* **7**, 51–63 (2010).
29. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
30. Chantzoura, E. *et al.* Reprogramming Roadblocks Are System Dependent. *Stem Cell Reports* **5**, 350–364 (2015).
31. Yang, B. X. *et al.* Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163**, 230–245 (2015).
32. Li, H. *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460**, 1136–1139 (2009).
33. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**, 1149–1153 (2009).
34. Rowe, H. M. *et al.* TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res.* **23**, 452–461 (2013).
35. Göke, J. *et al.* Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16**, 135–141 (2015).
36. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nat. 2012 4877405* **487**, 57–63 (2012).
37. Matoba, S. *et al.* Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. *Cell* **159**, 884–895 (2014).
38. Prahallad, A. *et al.* Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **483**, 100–104 (2012).
39. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, (2008).
40. deVries, N. A. *et al.* Prolonged Ezh2 Depletion in Glioblastoma Causes a Robust Switch in Cell Fate Resulting in Tumor Progression. *Cell Rep.* **10**, 383–397 (2015).

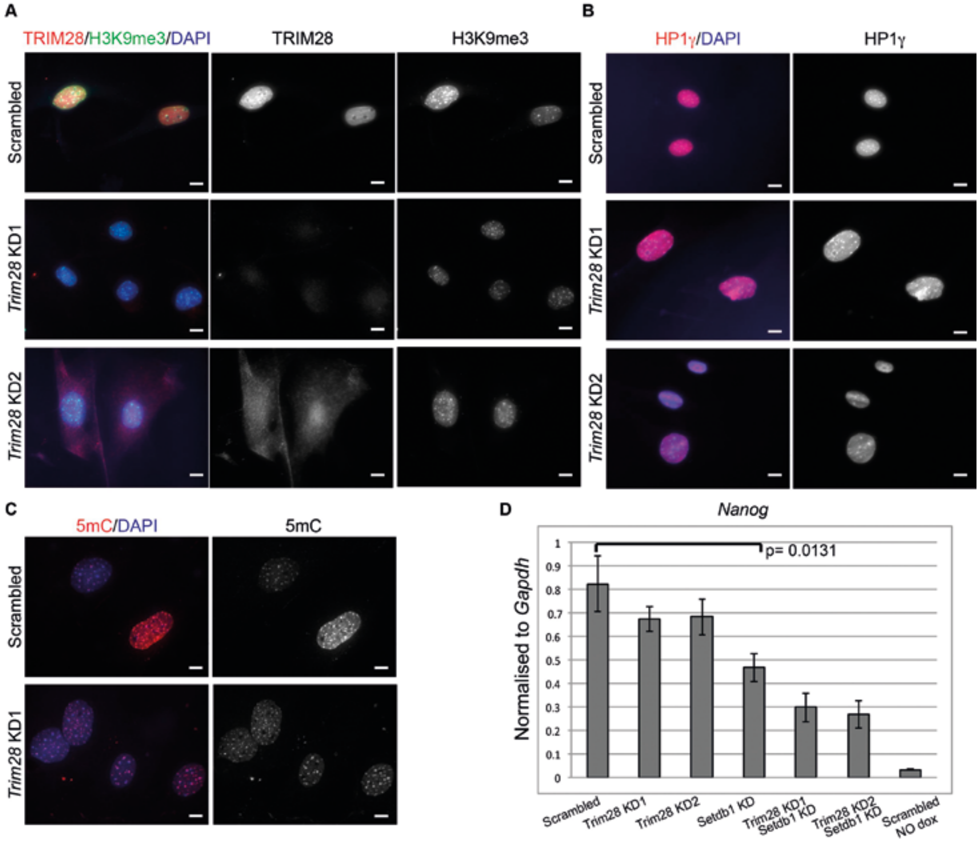
SUPPLEMENTARY MATERIAL



Supplementary Figure 1. Expression profiles of p53KD *Oct4*-GFP MEFs upon individual or combined knockdown of *Trim28* and *Setdb1*.



Supplementary Figure 2. Differences in gene expression between reprogramming and non-reprogramming samples.



Supplementary Figure 3. Levels of H3K9me3, HP1 gamma foci, 5 methylcytosine (5mC), or Nanog in *Oct4*-GFP MEFs with or without *Trim28*KD.

Supplementary Table 1. Genes with a log2Fold Change < 2 increase in expression during reprogramming with *Trim28*KD.

Ensembl ID	**log2FoldChange**	**pval**	**padj**
ENSMUSG00000064224	6,689	1,57E-50	2,24E-47
ENSMUSG00000054169	5,750	1,63E-67	3,28E-64
ENSMUSG00000079357	5,699	3,17E-45	4,26E-42
ENSMUSG00000034918	5,511	6,59E-43	6,93E-40
ENSMUSG00000052861	5,263	1,13E-27	5,83E-25
ENSMUSG00000034533	5,209	2,70E-42	2,72E-39
ENSMUSG00000097398	4,912	1,73E-25	7,89E-23
ENSMUSG00000097569	4,715	3,41E-44	3,93E-41
ENSMUSG00000099452	4,688	1,85E-43	2,04E-40
ENSMUSG00000033826	4,641	5,32E-55	8,06E-52
ENSMUSG00000058400	4,596	5,88E-30	3,31E-27
ENSMUSG00000013338	4,514	1,13E-38	9,78E-36
ENSMUSG00000085234	4,497	3,33E-11	3,01E-9
ENSMUSG00000099474	4,466	5,86E-21	1,71E-18
ENSMUSG00000032649	4,336	2,86E-18	6,71E-16
ENSMUSG00000056071	4,325	6,48E-17	1,33E-14
ENSMUSG00000071042	4,279	1,56E-44	1,88E-41
ENSMUSG00000074489	4,267	3,29E-24	1,33E-21
ENSMUSG00000074895	4,248	1,60E-32	1,07E-29
ENSMUSG00000056293	4,225	7,38E-33	5,25E-30
ENSMUSG00000015787	4,063	1,44E-23	5,53E-21
ENSMUSG00000053063	3,965	4,88E-10	3,66E-8
ENSMUSG00000021708	3,951	1,12E-130	2,71E-126
ENSMUSG00000096528	3,902	3,17E-6	0,000
ENSMUSG00000033847	3,808	2,54E-18	6,03E-16
ENSMUSG00000024365	3,694	3,27E-104	3,96E-100
ENSMUSG00000029134	3,649	4,61E-21	1,38E-18
ENSMUSG00000045009	3,586	4,30E-10	3,26E-8
ENSMUSG00000096035	3,515	4,87E-81	1,96E-77
ENSMUSG00000094626	3,407	3,01E-7	1,31E-5
ENSMUSG00000046856	3,331	1,08E-12	1,21E-10
ENSMUSG00000035910	3,310	1,97E-6	7,03E-5
ENSMUSG00000022582	3,296	1,46E-15	2,58E-13
ENSMUSG00000041608	3,295	1,06E-24	4,50E-22
ENSMUSG00000100029	3,278	4,48E-11	3,90E-9
ENSMUSG00000024575	3,247	1,21E-10	1,00E-8
ENSMUSG00000047592	3,181	6,79E-26	3,16E-23
ENSMUSG00000051726	3,147	5,34E-83	2,59E-79
ENSMUSG00000100398	3,142	0,001	0,017
ENSMUSG00000075604	3,137	7,79E-5	0,002
ENSMUSG00000075023	3,132	8,70E-6	0,000
ENSMUSG00000029925	3,027	2,91E-10	2,25E-8

Supplementary Table 1. (continued).

Ensembl ID	log2FoldChange	pval	padj
ENSMUSG00000039313	3,026	3,77E-13	4,70E-11
ENSMUSG00000063651	2,982	6,07E-79	1,84E-75
ENSMUSG00000040164	2,944	3,40E-15	5,63E-13
ENSMUSG00000038132	2,936	2,85E-10	2,21E-8
ENSMUSG00000024313	2,934	8,71E-5	0,002
ENSMUSG00000078907	2,929	4,25E-7	1,77E-5
ENSMUSG00000018126	2,915	1,07E-7	5,05E-6
ENSMUSG00000020953	2,903	4,81E-69	1,16E-65
ENSMUSG00000086563	2,884	1,52E-9	1,02E-7
ENSMUSG00000031766	2,879	0,000	0,002
ENSMUSG00000079262	2,851	8,21E-8	3,97E-6
ENSMUSG00000089667	2,848	0,005	0,051
ENSMUSG00000028148	2,839	9,79E-80	3,39E-76
ENSMUSG00000101029	2,822	2,70E-5	0,001
ENSMUSG00000069456	2,794	7,98E-9	4,70E-7
ENSMUSG00000055403	2,776	0,008	0,084
ENSMUSG00000059625	2,772	0,000	0,007
ENSMUSG00000096141	2,752	1,90E-6	6,80E-5
ENSMUSG00000038670	2,732	1,02E-5	0,000
ENSMUSG00000020431	2,723	3,20E-77	8,60E-74
ENSMUSG00000092471	2,722	2,52E-6	8,65E-5
ENSMUSG00000055827	2,710	3,41E-6	0,000
ENSMUSG00000044117	2,705	1,27E-18	3,04E-16
ENSMUSG00000034634	2,692	1,90E-21	5,75E-19
ENSMUSG00000015484	2,684	0,004	0,045
ENSMUSG00000036962	2,669	0,000	0,005
ENSMUSG00000047842	2,655	3,86E-18	8,91E-16
ENSMUSG00000075224	2,642	2,69E-7	1,19E-5
ENSMUSG00000020332	2,620	0,002	0,031
ENSMUSG00000040412	2,609	0,004	0,042
ENSMUSG00000017723	2,607	1,09E-58	1,88E-55
ENSMUSG00000024731	2,587	8,41E-11	7,08E-9
ENSMUSG00000028294	2,578	1,05E-6	4,02E-5
ENSMUSG00000021750	2,576	9,51E-16	1,74E-13
ENSMUSG00000021214	2,524	7,55E-15	1,18E-12
ENSMUSG00000092591	2,511	0,000	0,005
ENSMUSG00000073514	2,503	0,002	0,022
ENSMUSG00000020475	2,482	1,65E-10	1,34E-8
ENSMUSG00000021238	2,479	3,62E-98	2,19E-94
ENSMUSG00000092340	2,461	3,60E-5	0,001
ENSMUSG00000070570	2,455	2,11E-5	0,001
ENSMUSG00000038932	2,447	6,65E-69	1,46E-65
ENSMUSG00000038242	2,441	1,83E-5	0,001

Supplementary Table 1. (continued).

Ensembl ID	log2FoldChange	pval	padj
ENSMUSG00000034452	2,434	0,001	0,010
ENSMUSG00000001663	2,426	1,11E-16	2,21E-14
ENSMUSG00000028332	2,395	2,47E-5	0,001
ENSMUSG00000087703	2,392	0,000	0,006
ENSMUSG00000027612	2,387	3,74E-9	2,33E-7
ENSMUSG00000030935	2,372	1,36E-16	2,66E-14
ENSMUSG00000021367	2,370	4,42E-23	1,53E-20
ENSMUSG00000052087	2,369	1,68E-7	7,68E-6
ENSMUSG00000004540	2,361	5,18E-5	0,001
ENSMUSG00000092200	2,360	0,007	0,069
ENSMUSG00000044951	2,357	0,000	0,003
ENSMUSG00000087677	2,352	0,001	0,014
ENSMUSG00000045201	2,350	1,64E-6	5,99E-5
ENSMUSG00000031698	2,323	0,000	0,007
ENSMUSG00000020542	2,307	7,63E-5	0,002
ENSMUSG00000021071	2,304	1,35E-10	1,11E-8
ENSMUSG00000039543	2,303	8,86E-7	3,44E-5
ENSMUSG00000073000	2,293	0,002	0,027
ENSMUSG00000024209	2,289	5,77E-13	6,92E-11
ENSMUSG00000048349	2,285	7,80E-7	3,08E-5
ENSMUSG00000043727	2,280	3,67E-5	0,001
ENSMUSG00000044667	2,274	2,49E-8	1,32E-6
ENSMUSG00000044835	2,269	3,99E-11	3,50E-9
ENSMUSG00000056296	2,247	0,001	0,018
ENSMUSG00000037157	2,220	0,000	0,005
ENSMUSG00000100783	2,206	0,004	0,047
ENSMUSG00000028978	2,203	2,79E-7	1,22E-5
ENSMUSG00000076863	2,202	8,65E-5	0,002
ENSMUSG00000092354	2,198	0,000	0,006
ENSMUSG00000018924	2,181	0,000	0,005
ENSMUSG00000094498	2,176	0,009	0,089
ENSMUSG00000000627	2,172	7,85E-19	1,94E-16
ENSMUSG00000023247	2,169	0,000	0,004
ENSMUSG00000031489	2,164	2,30E-34	1,74E-31
ENSMUSG00000100127	2,163	0,000	0,007
ENSMUSG00000051228	2,162	9,65E-5	0,002
ENSMUSG00000079465	2,158	0,002	0,029
ENSMUSG00000021749	2,148	1,45E-15	2,58E-13
ENSMUSG00000032202	2,141	3,00E-38	2,50E-35
ENSMUSG00000032845	2,138	4,60E-5	0,001
ENSMUSG00000031022	2,128	0,002	0,031
ENSMUSG00000042453	2,125	1,36E-9	9,25E-8
ENSMUSG00000030468	2,115	3,37E-19	8,96E-17

Supplementary Table 1. (continued).

Ensembl ID	log2FoldChange	pval	padj
ENSMUSG00000055748	2,107	0,005	0,056
ENSMUSG00000073399	2,099	5,38E-13	6,47E-11
ENSMUSG00000030263	2,079	1,55E-22	5,14E-20
ENSMUSG00000049593	2,074	0,006	0,060
ENSMUSG00000026114	2,065	3,67E-22	1,15E-19
ENSMUSG00000000248	2,047	5,74E-14	8,43E-12
ENSMUSG00000024803	2,043	8,42E-31	5,10E-28
ENSMUSG00000000197	2,038	1,39E-10	1,14E-8
ENSMUSG00000024008	2,034	3,83E-19	9,97E-17
ENSMUSG00000075270	2,029	0,000	0,005
ENSMUSG00000037440	2,027	6,63E-15	1,05E-12
ENSMUSG00000027887	2,021	4,18E-5	0,001
ENSMUSG00000048521	2,019	1,32E-19	3,55E-17
ENSMUSG00000026012	2,008	9,13E-13	1,03E-10
ENSMUSG00000031549	2,003	3,49E-14	5,28E-12

Supplementary Table 2. List of primers.

Primer	Sequence
Mouse_Gapdh_L1	AGGTCGGTGTGAACGGATTTG
Mouse_Gapdh_R1	TGTAGACCATGTAGTTGAGGTCA
Mouse_Nanog_qPCR_L1	TTGCTTACAAGGGTCTGCTACT
Mouse_Nanog_qPCR_R1	ACTGGTAGAAGAATCAGGGCT
Mouse_Setdb1_L1	TCATTAACACAGGCCCTGAA
Mouse_Setdb1_R1	GGCAGGCACATTTGGATTTA
Mouse_Trim28_L1	TGGTCAATGATGCCCAGA
Mouse_Trim28_R1	TTGGTCATGGTCCAGTGCT
4F2A E2A-cMyc_L1 [22]	GGCTGGAGATGTTGAGAGCAA
4F2A E2A-cMyc_R1 [22]	AAAGGAAATCCAGTGGCGC
Mouse_IAP_L1 [40]	GCACCCTCAAAGCCTATCTTA
Mouse_IAP_R1[40]	TCCCTTGGTCAGTCTGGATTT
Mouse_LINE1_5’UTR_L1[40]	GGAGTCTGCGTTCTGATGA
Mouse_LINE1_5’UTR_R1[40]	GGCGAAAGGCAAACGTAAGA
Mouse_IAP.int.2_L1 [40]	AGCAGGTGAAGCCACTG
Mouse_IAP.int.2_R1 [40]	CTTGCCACACTTAGAGC
Mouse_Line1_L1 [17]	TTTGGGACACAATGAAAGCA
Mouse_Line1_R2 ([17]	CTGCCGTCTACTCCTCTTGG
Mouse_IAP 5’UTR_L1 [17]	CGGGTCGCGGTAATAAAGGT
Mouse_IAP 5’UTR_R1 [17]	ACTCTCGTTCCCCAGCTGAA
ERVK10CLTR_L1 [18]	GTGTGAGACACGCCTCTCCT
ERVK10CLTR_R1 [18]	GGGAGAGCTTGATTGCAGAG

Supplementary Table 2. List of primers.

Primer	Sequence
ERVK10CGAG_L1 [18]	TCAGGATCATGCTCAACAGC
ERVK10CGAG_R1 [18]	TGGCATTGTGAGCCAATCTA
ERVK10CPOL_L1 [18]	GCCACCAGAGACATGGTTTT
ERVK10CPOL_R1 [18]	CGGGCTTCTTTTCTTGTGAG
ERVK10CENV_L1 [18]	TATCGCCTCAGGGTTAATGC
ERVK10CENV_R1 [18]	TGGATGCCACACAACTCATT

Supplementary Table 3. Repeat elements that alter expression during reprogramming.

Repeat	log Fold Change Scrambled_vs_ Scrambled_no_DOX	FDR Scrambled_vs_ Scrambled_no_DOX
MurSAT1	6,559	8,21E-8
MLT1H2-int	4,255	6,79E-14
Tigger18a	2,845	2,88E-78
MER136	2,500	1,96E-5
tRNA-Lys-AAG	2,469	9,98E-7
MER91C	2,266	0,000
MER96	2,053	6,67E-6
Helitron2Na_Mam	1,952	0,035
X11_DNA	1,869	3,39E-13
RLTR19D	1,831	0,002
LTR58	1,773	0,013
RLTR46A	1,725	0,002
UCON88	1,698	0,042
Charlie17	1,489	1,77E-14
MMETn-int	1,440	9,17E-30
B2_Mm1a	1,377	2,51E-21
UCON84	1,368	0,009
ETnERV-int	1,354	9,17E-30
X26_DNA	1,326	1,09E-22
RLTR9C	1,263	7,24E-6
B2_Mm1t	1,260	4,68E-19
MER91B	1,214	4,36E-8
ERVB4_1C-LTR_Mm	1,203	1,06E-25
ORR1A1-int	1,129	3,64E-26
X6a_DNA	1,095	4,14E-6
ERVB4_2-LTR_MM	1,089	3,81E-6
LTR103b_Mam	1,069	0,032
RLTR22_Mus	1,037	1,25E-24
RLTR43C	0,986	1,33E-14
RMER12B	0,975	2,81E-17

Supplementary Table 3. (continued)

Repeat	log Fold Change Scrambled_vs_ Scrambled_no_DOX	FDR Scrambled_vs_ Scrambled_no_DOX
HAL1ME	0,971	1,52E-12
RLTR44C	0,970	1,76E-15
RLTR44-int	0,951	4,66E-15
LTR108a_Mam	0,900	0,000
UCON73	0,866	0,048
RMER17A2	0,803	3,65E-9
ERVB7_1-LTR_MM	0,787	5,03E-14
B2_Mm2	0,774	1,12E-11
RLTR1F_Mm	0,754	0,008
RMER16	0,719	2,60E-9
LTR65	0,708	0,021
ORR1A0-int	0,692	4,59E-12
MARE10	0,661	0,000
RLTRETN_Mm	0,654	1,67E-9
U1	0,642	0,003
MERVL-int	0,622	7,60E-15
RLTR13G	0,606	5,41E-7
MLT1B-int	0,597	0,021
ERVL-B4-int	0,585	1,07E-14
ERVL-int	0,582	1,86E-6
Charlie19a	0,572	0,033
MER34B-int	0,558	1,46E-6
RLTR44D	0,549	0,020
ORR1A0	0,541	8,82E-12
RMER16A2	0,535	0,000
MT2C_Mm	0,518	6,92E-16
U14	0,501	0,039
RLTR1A2_MM	0,499	5,87E-5
RLTR13B2	0,479	0,025
RLTR26	0,470	0,000
LTRIS2	0,449	1,10E-8
MLT1A1	0,448	4,02E-10
L1MEg	0,444	2,83E-7
RLTR1B	0,433	4,57E-6
ORR1A1	0,427	1,18E-9
LTR37-int	0,425	0,007
IAPA_MM-int	0,407	0,021
IAPEz-int	0,394	0,019
RLTR13E	0,388	3,73E-6
MMTV-int	0,388	0,044
MLT1E1	0,386	0,030
RLTR45	0,379	0,001

Supplementary Table 3. (continued)

Repeat	log Fold Change Scrambled_vs_ Scrambled_no_DOX	FDR Scrambled_vs_ Scrambled_no_DOX
Charlie24	0,376	0,008
MERVL_2A-int	0,371	2,94E-7
IAPLTR3-int	0,347	0,021
ERVB4_2-I_MM	0,347	0,011
IAPEy-int	0,343	0,022
RLTR1B-int	0,329	5,30E-5
LTR16A	0,328	0,033
ORR1A2-int	0,327	7,83E-7
RMER6A	0,324	1,98E-5
RMER19B	0,320	0,001
ORR1C2-int	0,319	0,000
LTRIS3	0,317	0,036
MT2_Mm	0,312	0,007
MERV1_I-int	0,301	0,004
ORR1A3-int	0,282	0,000
MuRRS4-int	0,279	0,000
MTB-int	0,270	6,61E-5
IAPEY3-int	0,266	0,021
RMER19A	0,256	0,021
ORR1A4	0,254	0,000
RLTR19-int	0,248	0,015
MLTR25A	0,242	0,018
ORR1A2	0,235	0,000
RMER4B	0,225	0,000
ORR1C1	0,217	0,003
RodERV21-int	0,206	0,048
MTC-int	0,200	0,003
ORR1A3	0,175	0,021
L2b	-0,157	0,036
RLTR21	-0,157	0,022
MIR	-0,160	0,046
MER58A	-0,165	0,050
MamRTE1	-0,174	0,021
L1MA6	-0,189	0,021
MTE2b-int	-0,203	0,015
MTE-int	-0,205	0,002
L1ME4b	-0,217	0,006
L3b	-0,218	0,021
L1MC2	-0,227	0,011
B1_Mur1	-0,238	0,018
Charlie1	-0,241	0,023
A-rich	-0,255	0,037

Supplementary Table 3. (continued)

Repeat	log Fold Change Scrambled_vs_ Scrambled_no_DOX	FDR Scrambled_vs_ Scrambled_no_DOX
MuLV-int	-0,269	8,94E-5
MER31A	-0,271	0,036
MLT1L	-0,271	0,029
L1ME3A	-0,272	0,001
RLTR11B	-0,277	0,002
RLTR6-int	-0,291	0,001
MLT1H	-0,300	0,007
Kanga2_a	-0,301	0,048
MER131	-0,305	0,036
L2d	-0,329	0,000
MMVL30-int	-0,362	0,002
MER102b	-0,393	0,006
Tigger15a	-0,399	0,001
RLTR4_MM-int	-0,404	1,52E-9
U2	-0,406	0,008
RLTR4_Mm	-0,409	1,21E-7
MER74A	-0,450	0,000
RLTR6_Mm	-0,451	5,69E-6
Charlie30b	-0,469	5,01E-5
Charlie15a	-0,482	0,002
X32_DNA	-0,486	5,87E-5
UCON49	-0,500	0,036
X15_LINE	-0,512	4,56E-7
RLTR13D2	-0,512	0,000
MER58D	-0,530	0,023
L1ME4a	-0,548	6,05E-9
Charlie29b	-0,548	0,030
RLTR6C_Mm	-0,554	1,68E-8
CR1-3_Croc	-0,559	0,021
X7A_LINE	-0,564	0,003
MamRep605	-0,570	1,04E-6
hAT-5_Mam	-0,603	0,010
L1M2a	-0,608	0,012
X9b_DNA	-0,647	0,012
MER117	-0,691	4,08E-8
MER70B	-0,725	0,014
Tigger14a	-0,744	3,82E-6
MADE2	-0,748	5,02E-5
X33a_DNA	-0,750	1,96E-7
MLT1H-int	-0,758	3,69E-14
tRNA-Ala-GCA	-0,781	4,72E-14
LTR55	-0,786	0,006

Supplementary Table 3. (continued)

Repeat	log Fold Change Scrambled_vs_ Scrambled_no_DOX	FDR Scrambled_vs_ Scrambled_no_DOX
Charlie16a	-0,788	1,26E-7
UCON6	-0,807	1,52E-9
ERVB4_1B-LTR_MM	-0,809	1,34E-14
MamGypLTR1b	-0,870	6,73E-6
L1M2b	-0,873	2,01E-6
X4b_DNA	-0,942	1,78E-27
U4	-0,972	0,000
UCON80	-1,024	0,000
MamGypLTR3a	-1,051	5,27E-11
MER101B	-1,075	0,000
LTR107_Mam	-1,154	0,037
MamGypLTR1c	-1,223	4,62E-13
Penelope1_Vert	-1,272	1,16E-61
MER91A	-1,295	3,94E-16
UCON20	-1,509	0,008
MER135	-1,658	1,68E-8
X24_DNA	-1,818	0,049
MER97b	-1,819	4,97E-11
MLT1I-int	-1,905	0,029
Eulor6D	-2,904	0,038

Supplementary Table 4. Repeat elements that alter expression during reprogramming with *Trim28*KD. Normalized read counts for each shRNA and replicate from the sequencing data, aligned to the epigenome shRNA library.

Repeat	log Fold Change Trim28_vs_Scrambled	FDR.Trim28_vs_Scrambled
ERVB4_2-LTR_MM	2,555	3,25E-73
ERVB4_2-I_MM	2,329	8,77E-121
RLTR45-int	2,217	2,69E-185
LTRIS5	2,083	1,69E-17
RLTR6C_Mm	1,919	4,90E-105
Helitron1Na_Mam	1,893	0,017
ERVB2_1-I_MM	1,887	9,33E-12
RLTR1B-int	1,632	4,17E-119
MMERVK10C-int	1,582	2,99E-114
RLTR6_Mm	1,485	8,55E-63
MMVL30-int	1,471	3,53E-49
tRNA-Arg-AGG	1,395	0,012
IAP-d-int	1,349	3,68E-53
MamGypLTR2c	1,228	0,032
MMERVK10D3_LTR	1,126	0,001

Supplementary Table 4. (continued).

Repeat	**log Fold Change Trim28_vs_Scrambled**	**FDR.Trim28_vs_Scrambled**
IAPEz-int	1,112	1,61E-16
RLTR6-int	1,085	1,09E-49
RLTR6B_Mm	1,083	1,79E-46
ETnERV3-int	1,068	5,59E-39
RLTR10D	1,065	3,97E-15
IAPA_MM-int	1,054	1,31E-15
IAPEy-int	1,052	1,69E-20
IAPEY3-int	1,012	1,13E-30
MMERGLN-int	1,001	7,72E-37
IAPLTR3-int	0,991	8,11E-17
IAPEY5_I-int	0,977	2,53E-18
IAPLTR1_Mm	0,962	6,98E-18
IAPLTR4_I	0,930	1,25E-14
RLTR10-int	0,911	1,37E-24
ERVB4_1B-I_MM	0,909	6,12E-25
MMTV-int	0,907	2,07E-10
MuRRS4-int	0,896	8,69E-44
MMERVK10D3_I-int	0,895	3,75E-23
RLTR45	0,862	3,45E-21
RLTR1B	0,836	1,18E-23
RLTR1E_MM	0,824	1,59E-10
IAPLTR2a	0,814	1,76E-13
IAPLTR1a_Mm	0,806	1,88E-12
RLTR10C	0,735	2,95E-17
MMERGLN_LTR	0,701	1,80E-15
IAPLTR2_Mm	0,700	6,83E-11
RLTR10	0,685	1,86E-10
RLTR44D	0,672	0,001
RLTR9C	0,672	0,015
RLTR27	0,659	3,97E-5
RLTR44-int	0,636	1,27E-7
RLTR47	0,624	0,030
RLTR44C	0,601	3,91E-7
LTR85a	0,593	0,013
RNERVK23-int	0,582	2,41E-6
RLTR1D2_MM	0,573	0,008
IAPLTR2b	0,548	2,88E-8
IAPEY4_I-int	0,547	6,11E-17
RLTR13B1	0,514	1,90E-13
IAPLTR2a2_Mm	0,502	4,71E-9
RLTR10B	0,464	0,012
ETnERV2-int	0,463	7,45E-10
IAP1-MM_I-int	0,458	3,35E-8

Supplementary Table 4. (continued).

Repeat	log Fold Change Trim28_vs_Scrambled	FDR.Trim28_vs_Scrambled
Penelope1_Vert	0,432	2,28E-7
ORR1A0-int	0,425	8,84E-5
RLTR26	0,424	0,001
RMER17A-int	0,419	0,002
RLTR9E	0,392	0,001
MMERVK9C_I-int	0,389	5,93E-6
LTRIS_Mm	0,385	0,004
LTRIS_Mus	0,384	0,001
L1ME3F	0,370	0,008
MMERVK9E_I-int	0,361	1,25E-6
ORR1A0	0,358	1,47E-5
LTRIS2	0,340	1,57E-5
MLTR18C_MM	0,335	0,032
RMER16-int	0,318	3,50E-6
MERVK26-int	0,290	0,011
ORR1A1	0,276	0,000
MurERV4-int	0,251	0,001
ORR1A3-int	0,212	0,012
Tigger19b	-0,236	0,013
X15_LINE	-0,311	0,013
MER34B-int	-0,318	0,030
L1M3e	-0,320	0,046
MLT1A1	-0,325	1,58E-5
Tigger6a	-0,372	0,002
Tigger16a	-0,419	0,038
L1M4a1	-0,419	0,024
RLTR41C	-0,429	0,014
MER34B	-0,452	0,007
X17_LINE	-0,550	3,13E-5
UCON71_Crp	-0,582	0,001
Tigger8	-0,641	0,001
UCON80	-0,919	0,012
UCON88	-1,796	0,039

Chapter 7

GENERAL DISCUSSION

GENERAL DISCUSSION

This thesis highlights how pioneering screening tools can serve as powerful approaches to resolve previously unexplored genetic or epigenetic processes and interactions. The findings can help improve our understanding of epigenetic interactions and functions in normal or cancer cells. They also provide new insights into epigenetic functions that may regulate targeted therapies, gene editing, or reprogramming for therapeutic development. Whether used individually or combined, these approaches are powerful assets to unravel complex and often context-dependent epigenetic processes from different angles, ultimately, to propel cancer research. From a biotechnological standpoint, the transposon-based screening approach Thousands of Reporters Integrated in Parallel (TRIP) identified how the epigenomic context regulates Cas9 gene editing efficiency (Chapter 3). In addition, a focused shRNA library targeting epigenetics-associated genes deciphered how epigenetic alterations regulate cellular reprogramming, which can help our understanding of normal development and cancer and possibly aid the development of stem-cell therapies (Chapter 6). From a clinical standpoint, different haploid genetic screens revealed molecular interactions that may affect the sensitivity or resistance of cancer cells to small-molecule inhibitors of epigenetic proteins.

The impact of genomic and epigenomic context on gene editing

Precision gene editing with clustered regularly interspaced short palindromic repeats (CRISPR)–associated protein 9 (Cas9) has rapidly emerged as the primary gene-editing tool in fundamental and clinical research[1–3]. From monogenic loss-of-function studies to large-scale mutagenesis screens, CRISPR–Cas9 has offered researchers a simple and reliable tool to identify previously unknown interactions in any cell type. Notably, the successful implementation of CRISPR–Cas9-based gene editing in human cells has paved the way for currently ongoing clinical trials evaluating the technology's safety and efficacy in different types of diseases, including cancer[4,5]

However, despite these rapid advancements, little is known about its activity profile, especially in the context of the epigenomic or genomic landscape of a cell[6]. Chapter 3 describes how targeting thousands of barcoded reporter sequences integrated throughout pools of mouse embryonic stem (mES) cell genomes with gRNA-directed Cas9 characterized the effects of the epigenomic or genomic context on gene editing. Using the TRIP technology, we found that the gRNA sequence and the specific loci of the integrated reporters explain most of the observed variations in Cas9 mutations. The sequencing data from the screen also enabled us to assess the types of small insertions and deletions (indels) established primarily by the non-homologous end joining (NHEJ) pathway upon Cas9 cleavage. Analyses of the sequencing data and published results from mouse pre-B cells7 helped us establish a model in which Cas9 primarily cleaves DNA in a blunt fashion and occasionally creates staggered DNA ends. The high-throughput, nucleotide-resolution data from the reporter-based screen enabled us to detect mutation patterns that would have been missed at a small scale or outside of the cellular context.

Multiplexing the target sequence of gRNA–Cas9 through TRIP provides enough statistical power and minimal disruption of the endogenous genomic environment to capture actual Cas9-mutation frequencies in cell lines. The approach also circumvents confounding effects, such as heterogeneous outcomes when targeting multiple genomic sites or mutation calling upon targeting repetitive elements. Also, since the TRIP technology assesses living cells, it is ideal for studying Cas9 cleavage and the resulting DNA damage repair (DDR)-mediated mutations. By contrast, targeting synthesized, nucleosome-occupied sequences or doxycycline-induced heterochromatin *in vitro*[8–10] limits the evaluation of indels. Instead, it restricts the outcomes to quantifying cleavage efficacy or assessing binding properties of catalytically dead Cas9 (dCas9) rather than mutation frequencies.

Furthermore, exogenous manipulation of chromatin states may create a binary (permissive versus non-permissive) representation of the genome that discounts cell-endogenous factors that may influence Cas9 activity and mutation events. Two earlier studies[11,12] supported our observed effect of endogenous factors, suggesting that while genomic or epigenomic contexts affect Cas9 activity, other factors, such as target sequence or cell cycle profiles, likely contribute to the mutation frequency. Indeed, mES cell lines devote a longer time in the S phase than does, for example, mouse embryonic fibroblast (MEF) cells[13]. These differences may affect the choice of DDR pathways – with the template-dependent homology-directed repair being more pronounced in S phase than the error-prone NHEJ repair – which, ultimately, influences the mutation frequency. Since DNA is unpacked to a larger proportion in the S phase, our results derived from mES cell lines may explain the weak relationship we observed between the genomic or epigenomic context and Cas9 activity.

Nevertheless, our approach revealed several known and previously unknown features associated with Cas9 that can be implemented to improve Cas9 gene editing for future applications. Firstly, using the TRIP approach, we confirmed that Cas9 activity is dependent on gRNA sequence, with the GC-content being a prominent influencer. Our observations suggested that lower GC content correlates with higher mutation efficiency, in line with previous reports[14]. Secondly, using the TRIP technology to multiplex genome-wide sgRNA–Cas9 targets also opened up new possibilities to assess the influence of transcriptional, genomic, and epigenomic (TGE) features on mutation frequency in living cells. While we found that genomic location significantly influenced mutation frequency, the TGE features we evaluated were weak predictors of mutation frequency. As described above, the findings suggest that other factors not interrogated in this study might regulate Cas9-induced mutation frequency. Thirdly, we also found that insertions at the cleavage sites were more dependent on transcriptional features than deletions. In particular, insertion frequency correlated with features associated with regulating transcription, including phosphorylation of RNA PolII, transcription initiation, several histone marks, or Lamin-B1. Fourthly, using a 21-nucleotide long single-stranded oligodeoxynucleotide (ssODN) as a template for homology-directed repair, we observed low knock-in efficiency, in line with previous reports.

Interestingly, the proportion of ssODN knock-in increased with higher total mutation frequencies compared with insertion frequency from erroneous double-strand break repair. Lastly, our approach gave us insight into the possible cleavage patterns of Cas9 and the subsequent repair processes. While deletion sizes varied widely (mostly between 1 and 10 nucleotides), the vast majority of insertions consisted of a single gRNA-specific nucleotide. The observations made us hypothesize that insertions possibly were derived from staggered cuts rather than blunt cuts and that the predictable, gRNA-specific insertion arose from the filling of the resulting overhang. This deterministic nature of insertions could also explain why insertion frequency depends more on TGE features than deletion frequency.

We analyzed different cleavage models and proposed a model in which Cas9 primarily creates blunt and occasionally staggered DNA ends with 5′ overhangs. Our model was further supported by confirming evidence from independent studies on different cell lines[15–17] and G1-arrested pre-B cells[7] exposed to Cas9. We attributed these template-dependent insertions to DSB repair by either Pol μ or Pol λ, two polymerases that seem more flexible than other polymerases during template-dependent DNA synthesis[18]. However, more focused research on the different NHEJ pathways is needed to confirm the specific repair pathways involved in the staggered-end repair process.

Nevertheless, our findings consistently suggested that gRNA sequence determines the frequency of staggered cuts, which can be exploited for more precise gene editing, e.g., potentiating single-nucleotide insertions, improving knock-in efficiency, and dictating insert orientation during knock-in. In addition, the most common deletion pattern observed accounted for approximately 40% of all deletions and consisted of a specific trinucleotide sequence in the targeted reporter sequences upon exposure to one of the three sgRNAs, namely gRNA1. Since the specific trinucleotide (CGG) was repeated in the target sequence of gRNA1, we attributed the event to microhomology-directed end joining (MMEJ), an alternative NHEJ pathway that relies on microhomologous sequences to align the DNA ends at the DNA double-strand break[19]. As a result, the pathway can delete the microhomology at the cleavage site, a procedure that has been confirmed in a more recent study of Cas9-induced repair mechanisms[20]. Similar to the gRNA-dependent staggered cuts, the findings demonstrate that one can manipulate indel patterns and create specific small NHEJ- or MMEJ-derived "scars" in the target sequence depending on the gRNA sequence of choice. They can guide gRNA design to avoid in-frame mutations and potentiate frameshift mutations. However, as for the suggested template-dependent insertions, future research on these pathways will be necessary to confirm the extent to which NHEJ or MMEJ is involved in the respective repair of Cas9-induced editing.

We extracted extensive data to improve gene editing with CRISPR–Cas9 (or other CRISPR–Cas) technologies using the TRIP approach. Our results highlight the importance of proper gRNA design and can guide future research to enhance Cas9 efficiency through, for instance, target-sequence selection. Further studies will be able to specify TGE features that influence Cas9 efficiency and confirm whether specific cell types (for example, with shorter S phase) or cutting patterns might be more dependent on TGE features than others. We could envision that specific sequences, which are more prone to cause staggered cuts, can enhance knock-in efficiency and template orientation. Implementing this knowledge when designing CRISPR–Cas9 and gRNAs would prove valuable to improve the otherwise low efficiency of ssODN knock-ins with CRISPR–Cas9.

Identifying novel gene functions associated with targeted PcG treatment using haploid genetic screens

Haploid genetic screens have provided an excellent approach to discovering mechanisms of action associated with drug candidates. Whether using inhibitors, gene editing, or RNAi-induced gene regulation, haploid genetic screens are excellent choices to extract high-throughput data revealing genetic features responsible for drug responses, such as drug sensitivity or resistance. The haploid nature of the cell lines lends itself to efficiently creating large libraries of genome-wide null mutations that can be used to investigate gene properties upon treatment with inhibitors of interest. Chapters 4 and 5 shift the focus from investigating genomic and epigenomic features regulating gene editing to identifying genes and gene products that affect targeted therapies with Polycomb Group (PcG) inhibitors. Of note, the choice of drug targets can, of course, be extended to other drug classes[21,22].

Chapter 4 describes how a straightforward enrichment screen revealed mutations responsible for cell resistance to cytotoxic levels of AB057609107 (PTC-318), a novel small-molecule inhibitor designed to regulate BMI1 post-transcriptionally. These random integration screens provide a powerful starting point when investigating drug candidates, such as BMI1 inhibitors, that induce growth arrest and cell death[23–25] but where the contributing genetic interactions are largely unknown[26–28]. They can also be combined with other screening types to compare gene enrichment between two distinct but related pathways (Chapter 5; discussed below).

Chapter 4 validated the most significantly enriched gene disruption, Nuclear Mitotic Apparatus proteins (*NUMA1*), and its functions associated with PTC-318 resistance. To our knowledge, although BMI1 and NUMA1 share several common cellular processes, NUMA1 has previously not been directly associated with BMI1 in cancer. The functions of NUMA1 include DNA damage repair and homologous recombination[29], cell differentiation[30–32], higher-order chromatin organization[33], and cell cycle progression. Specifically, NUMA1 binds to microtubules to regulate mitotic cell division through spindle pole formation and control of proper chromosome segregation[34]. Similarly, both PRC1 and PRC2 have been implicated in cell cycle arrest in cancer cells, showing, for example, that aberrant expression of BMI1 or EZH2 reduces arrest at different parts of the cell cycle[35–37].

In line with these findings, our study showed that HAP1 cells treated with PTC-318 arrested and died in mitosis. We further showed that this direct relationship between BMI1 expression, mitotic arrest, and mitotic cell death is rescued by NUMA1 loss. Moreover, the PTC-318-induced mitotic phenotype coincided with increasing levels of CDK1 and Cyclin-B1, two proteins implicated in the cell cycle as markers of late G2/S cell-cycle phases. They also coincided with reduced expression levels of *BCL*-2, which encodes the anti-apoptotic protein BCL-2. These observations align with the known inverse relationship reported between the CyclinB1/cyclin-dependent kinase 1 (CDK1) complex and apoptotic signals[35,38]. The two independent networks maintain a balance in cell fate, where increasing levels of one of these two are associated with the decreasing levels of the other.

Interestingly, it has been demonstrated that CDK1-induced phosphorylation at serine 128 of pro-apoptotic protein Bad abrogates anti-apoptotic Bcl-2[39,40]. Ser128 phosphorylation has also been shown to promote cell death in mitosis upon treatment with an inhibitor of microtubule formation noncodazole[41]. Moreover, recent studies on PTC596, a small-molecule inhibitor that reduces levels of BMI1, have shown that it inhibits microtubule formation, which was demonstrated to induce mitotic arrest and apoptosis in pancreatic ductal adenocarcinoma cell lines[42,43]. It is, therefore, possible that BMI1 inhibition causes mitotic arrest through microtubule destabilization, which, in the presence of NUMA1, induces apoptosis through the upregulation of the CyclinB1/ CDK1 complex and, consequently, downregulation of anti-apoptotic BCL-2. Taken together, our observations suggest a role of BMI1 in cell cycle progression and a relationship between BMI1 and NUMA1 in mitotic regulation.

The fact that we picked up gene profiles seemingly unrelated to PcG further underscores the advantages of unbiased large-scale screening approaches. By contrast, similar screens with focused libraries (e.g., RNAi or gRNA libraries) limit the findings to the realm of specific libraries, e.g., associated with epigenetics or kinases. To this aim, haploid genetic screens are less biased and can present unexpected hits that would otherwise be missed in focused screens. Importantly, Cas9-induced knockout of NUMA1 in HAP1 cells and selected non-small cell lung cancer cell lines confirmed the involvement of NUMA1 in PTC-318 resistance. It confirmed that the findings are not limited to the haploid cell line but can be extended to other cancer cell lines, indicating their value in discovering therapeutic candidates.

Similarly, Chapter 5 shows how combining different types of HAP1 screens helps unveil gene candidates involved in inversely correlating phenotypes between related protein complexes. A depletion screen on HAP1 cells provided an extensive list of genes whose disruptions in combination with *EED* knockout lead to cancer cell death. Epigenetic regulators, such as the PcG, control the transcriptional activity of many genes. Because of this regulatory coverage, it is expected that disruption of core epigenetic proteins, such as EED, will be associated with a wide range of synthetic lethal partners. However, comparing the most significant data from a synthetic lethality HAP1 screen with data from other related HAP1 screens can reveal common candidates involved in an inverse correlation between otherwise closely related proteins or protein complexes. These comparison screens narrow down the data of gene candidates, specifying the findings to more specific purposes, e.g., comparing opposite functions between these related proteins. Our lab had previously observed that downregulation of PRC2 proteins EZH2 or EED with inhibitors or shRNAs resulted in an upregulation of BMI1 expression. Interestingly, transcriptional regulation between PcG proteins seems to be conserved between *Drosophila melanogaster* and human cells. Early studies on *Drosophila melanogaster* demonstrated that inactivating mutations of *Polycomb* derepressed several PcG targets along the polytene chromosomes, including homeotic genes and other PcG genes. PRC2 may, therefore, be a negative regulator of BMI1 in human cancer cells. Regulation of *BMI1* may be essential to maintain the "just-right" (Goldilocks) levels of BMI1 and, hence, cell fitness. Indeed, previous studies have shown that overexpression of Bmi1 in neural stem cells causes apoptosis through the repression of anti-apoptotic protein Survivin[44,45], further supporting a regulatory function of PRC2 on *BMI1*.

Together with data from our enrichment screen with PTC-318 (discussed above) and open-access data from an H3K27me3 phenotype screen[46], we identified the enrichment of a recurring gene candidate, *BAZ1B*, suggesting its potential involvement in the inverse relationship between PcG complexes. Although preliminary, our findings suggest that WSTF, the gene product of *BAZ1B*, is a positive regulator of H3K27me3 and potentially participates in a context-dependent, non-canonical PcG pathway. Although the interactions of WSTF are incompletely defined, considering previous findings indicating that loss of *BAZ1B* function redistributes the chromatin landscape in the cell nucleus[47], it is also possible that *BAZ1B* dysregulation disrupts global chromatin and gene regulation in our settings. For example, loss of WSTF could cause global chromatin rearrangement that activates the transcription of oncogenes or repress the transcription of tumor-suppressor genes. These changes could, for example, be responsible for rescuing cell death induced by BMI1 inhibition while also being synthetic lethal with disruption or inhibition of PRC2 genes. While incomplete, the findings provide a starting point to unravel the complexity associated with the interactions between these epigenetic pathways, which would have been difficult to retrieve with individual screens alone. Combining data from several large-scale screening approaches ultimately enabled us to isolate the recurring gene candidate.

Four reassuring observations confirmed the reliability of HAP1 screening approaches. Firstly, two of the screens identified *PBRM1* (encoding BAF180) and *ARID2* (encoding BAF200), two indispensable components of PBAF, one of the subcomplexes of the SWI/SNF complex and known antagonists of PcG[48]. Secondly, Cas9-induced knockouts of selected hits from the enrichment screen confirmed their association in the resistance to PTC-318 cytotoxicity. Thirdly, we reproduced the observed inverse correlation between PRC2 proteins and BMI1 with shRNAs. Lastly and importantly, we could demonstrate our findings in independent cancer cell lines, confirming that our observations were not specific to HAP1 cells.

Especially identifying SWI/SNF proteins in two different screens (Chapters 4 and 5) indicate an intricate and context-dependent relationship between PcG and SWI/SNF. Recent studies have demonstrated how SWI/SNF proteins, apart from being antagonists of PcG through its eviction[49], also can collaborate with PcG proteins[50–52]. The delicate balance between these two systems – both in healthy and malignant cells – is evident from a study by Weber et al. showing that degradation of Brg1, the ATPase subunit of SWI/SNF complexes, redistributes PRC1 and PRC2 from domains that PcG highly occupies to domains with lower PcG occupancy. This redistribution coincides with the derepression of primary PcG targets, e.g., the *Bmi1* locus. These findings show an interdependency between the different chromatin-modifying complexes that, depending on the genotypic or phenotypic context, can change the regulatory landscape in a dosage-dependent manner. These context-dependent interactions may explain the inverse relationship we identified between PRC1 protein BMI1 and the two core PRC2 proteins, EZH2 and EED. They may further corroborate the role of WSTF, a subcomponent of chromatin-remodeling complexes WICH, B-WICH, and WINAC, in the proposed inverse relationship between BMI1 and PRC2.

Comparing RNA interference screens with mutagenesis screens

The recent successes of gene-disruptive technologies – particularly with CRISPR–Cas9 – rapidly replaced a more than a decade-long implementation of RNA interference (RNAi) in eukaryotic cells and organisms. Chapters 3–5 highlight the high efficacy and reliability (reproducibility) gene-disruptive technologies can offer functional genomics and research in general. The chapters describe how knockout by CRISPR–Cas9, retroviral gene trapping, or insertional mutagenesis by transposon-derived TRIP uncovered several novel genotype–phenotype relationships on a large scale. The data were seamlessly retrieved through a combination of positive selection screens, negative selection screens, and phenotype screens in HAP1 cells, as well as transcriptional activity screens in mES cells. Moreover, the scientific evidence indicates that CRISPR–Cas9 outperforms RNAi in efficacy and frequency of off-target effects[53,54], which begs the question of whether RNAi screens have become obsolete.

Chapter 6 introduces a screen in mouse embryonic fibroblasts (MEFs) with an epigenome shRNA library targeting around 670 unique genes, which identified key regulators acting as a barrier to reprogramming somatic cells. The study highlights how TRIM28 acts as one of these reprogramming barriers, possibly by maintaining endogenous retroviruses (ERVs) silenced through H3K9me3 repression. Knockdown of *Trim28* resulted in increased H3K9me3 plasticity, upregulated ERVs – specifically IAP and MMERVK10c – and increased reprogramming.

The chapter presents a prime example of an experimental setting where RNAi silencing may be preferred over gene-disruptive mutagenesis approaches. Reprogramming somatic cells into pluripotent stem cells can help researchers better understand developmental processes, epigenetics, and diseases like cancer. For instance, cellular reprogramming can enable researchers to create specific and rare cell types (such as cancer cells) from somatic cells and study processes associated with the corresponding phenotype (such as epigenetic changes). The bidirectional transitions between somatic cells and induced pluripotent stem cells require transient modifications in the gene expression. To this aim, RNAi provides temporary repression of specific genes. This can be compared with gene disruptions, which establish permanent mutations in the genome and, thus, a "scar" that may impede further reliable assessments of the cells.

RNAi, such as shRNAs, may also be preferred over knockout technologies in drug discovery. One of the advantages of gene mutagenesis screens is their established null-mutation, which can mitigate the background noise of the partial transcription sometimes associated with RNAi. However, knockdown with shRNAs might better represent the transient and sometimes partial loss of protein activity induced by therapeutic drugs. Furthermore, knockout of essential genes might impede attempts to evaluate genotype–phenotype properties in cells due to cell lethality. In these cases, partial gene regulation through knockdown strategies is advantageous and may serve as a better tool for studying cancer phenotypes.

Although not the case for all knockout screen approaches (e.g., CRISPR screens), the mutagenesis methods described in Chapters 3–5 rely on cell-specific properties. For example, haploid genetic screens rely on observations from specific cell lines derived from chronic myeloid leukemia, namely KBM7 and HAP1[55,56]. The cell-type-specific approaches may limit the findings to the particular cell line and may, hence, not apply to other cancer types of interest. Extending the findings to additional cell types requires additional validation steps in different cell lines to confirm bona fide gene interactions. Similarly, since TRIP cell lines can be tedious to create[57], researchers may rely on already established cell lines to study transcriptional activity and its downstream effects. In Chapter 3, mES cells served to study the effect of TGEs on Cas9 activity. However, as mentioned in the discussion above, the extended time that mES cells spend in the S phase may have been responsible for the weak relationship between genomic or epigenomic context and Cas9 activity. By contrast, shRNA libraries are applicable in any cell line of interest and may, therefore, be a more sensible approach compared with cell-type-specific mutagenesis screen approaches when comparing phenotypes between different cell types. Note that gRNA libraries share the same advantageous properties as shRNA libraries in the same context[58].

Apart from these advantages, RNAi has limitations, which is why other functional genomic tools have surpassed it. While other types of libraries, for example, gRNA libraries, are associated with off-target effects, RNAi has been demonstrated to exhibit sequence-independent off-target effects that trigger an interferon effect[59]. Although it is important to note that off-target effects of earlier CRISPR–Cas9 systems may be more prevalent than initially thought[60], the advancements of CRISPR gene-knockout approaches have contributed to highly efficient and specific gene-editing systems with low false-positive backgrounds. These attributes contrast the confounding drawbacks of RNAi screens, namely incomplete gene silencing and increased probability of off-target effects. Similarly, in contrast to gene-disruptive screens, RNAi screens are arguably less suitable for negative selection screens as the technology is associated with off-target effects and presents high background noise.

Finally, shortly after implementing CRISPR–Cas9 in mammalian cells, a catalytically dead mutant form of Cas9 (dCas9) was found to block gene transcription in both bacteria and mammalian cells[61,62]. The process, known as CRISPR interference (CRISPRi), was further enhanced to repress mammalian gene transcription by coupling the dCas9 with repressor domains, such as the Krüppel-associated box (KRAB) domain, which can induce heterochromatin. This implementation also led to the development of the CRISPR activation (CRISPRa) approach, where dCas9 is coupled to a transcriptional effector (e.g., VP64 or the VPR approach[63]) to activate gene transcription. Both the CRISPRi and CRISPRa systems can manipulate gene transcription without editing gene sequences. This property entices a comparison between RNAi and, in particular, CRISPRi systems for screening purposes. Although CRISPRi is still in its early development compared with the more mature RNAi systems, there are advantages and disadvantages associated with both systems. The RNAi-induced repression depends on the cells' endogenous RNA-silencing machinery, enabling the introduction of a single element (i.e., siRNA or shRNA). While this is an advantage over CRISPRi, which may be inefficient to introduce into specific cells, the dependency of RNAi on cells' endogenous RNAi machinery can result in non-specific toxicity[64]. On the other hand, CRISPRi offers higher specificity compared with RNAi[62], but its effector domains may knock down adjacent genes of bidirectional promoters. However, next-generation libraries constantly enhance the efficacy and specificity of CRISPRi (and CRISPRa), ultimately outperforming first-generation libraries and yielding comparable results as libraries with catalytically active Cas9[65].

Outlook

The thesis has outlined several genetic approaches that can be used independently to study functional genomics. Whether being the integrational bias and cell specificity of insertional mutagenesis approaches, such as haploid or TRIP screens, or the confounding off-target effects of shRNA or CRISPR library screens, all the described tools are associated with their respective shortcomings. However, the same techniques can work complementarily to uncover the genetic wiring behind specific phenotypes. This work has demonstrated how the different tools uniquely contribute to either of three phases of cancer research. Our unbiased CRISPR-on-TRIP strategy in Chapter 3 was optimal to decipher how TGEs affect the activity of CRISPR–Cas9, which can significantly improve the technology for future research. Further, our haploid genetic screens, described in Chapters 4 and 5, enabled us to validate drugs and novel genetic interactions in cancer cells. Finally, Chapter 6 described how transient silencing with RNAi on reprogramming cells is optimal for revealing key genes that affect reprogramming, ultimately contributing to enhanced research on developmental processes, epigenetics, and cancer.

The rapid advancements of CRISPR seem to overshadow most other genetic technologies. However, it is imperative not to discredit alternative technologies as their respective functions may, depending on assays, be superior to those of CRISPR. For example, in CRISPR screens, phenotype differences are identified by amplifying and quantifying the DNA encoding the respective gRNAs, similar to phenotypic selection in RNAi screens[66]. This process is relatively simple and less prone to sequencing errors due to biased PCR amplification than insertional mutagenesis screens in haploid cells[67]. However, because gRNA amplification is an indirect quantification of phenotypes, the measurement may be confounded by off-target effects[68,69]. Hence, insertional mutagenesis screens, such as TRIP-like or haploid genetic screens, may occasionally be more favorable for unbiased, genome-wide identification of genotype- phenotype interactions. It is also important to note that CRISPR screens depend on error-prone DNA repair to induce indels and frameshift mutations. By contrast, insertional mutagenesis screens with haploid cells require only one gene integration to cause a null mutation, providing less background signal than expected with CRISPR–Cas9. However, the simplicity of haploid cell lines also presents one major disadvantage: the data extracted from haploid genetic screens may be specific to the haploid cell line. Translating the findings to other cancer cells of interest requires further validation. Nevertheless, CRISPR–Cas9 and (as shown in Chapters 4 and 5) haploid cells are reliable and powerful tools to extract data from positive and negative selection screens.

Furthermore, as already discussed, transient gene silencing using RNAi screening approaches may be preferable for reprogramming assays since they keep the gene intact for further differentiation of the cells and their applications. Similarly, partial silencing may be preferable over null mutations in such assays or in identifying new drug targets – since it better mimics the effects of therapeutic drugs.

These examples highlight the benefits of widening the focus of screening tools to obtain a complete picture of functional genomics in the relevant research fields. Rather than replacing one another, these technologies can complement and overlap to extract the most reliable genotype–phenotype results. It will be interesting to see these technologies develop and become more robust, for example, by establishing TRIP libraries in additional cell types or haploidy in cell lines unrelated to chronic myeloid leukemia. As seen in Chapter 3, different technologies can also be combined. However, the combination of tools is not limited to TRIP and CRISPR, but haploid cell lines can also be combined with CRISPR libraries for focused targeting[22]. For example, exposing haploid cells to barcoded transposons has been valuable for identifying drug sensitivity in cancer cells[70]. Complementing retrovirus integrations with transposon integrations may help account for integration biases specific to the two technologies[71]. Alternatively, chemical mutagenesis could be used to study environmental effects on genotypes and phenotypes, for example, in haploid cell lines[72].

REFERENCES

1. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337**, 816–821 (2012).
2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819 (2013).
3. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823 (2013).
4. Stadtmauer, E. A. *et al.* CRISPR-engineered T cells in patients with refractory cancer. *Science (80-.).* **367**, (2020).
5. Lu, Y. *et al.* Safety and feasibility of CRISPR-edited T cells in patients with refractory non-small-cell lung cancer. *Nat. Med. 2020 265* **26**, 732–740 (2020).
6. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262–1278 (2014).
7. Dorsett, Y. *et al.* HCoDES reveals chromosomal DNA end structures with single nucleotide resolution. *Mol. Cell* **56**, 808 (2014).
8. Horlbeck, M. A. *et al.* Nucleosomes impede cas9 access to DNA in vivo and in vitro. *Elife* **5**, (2016).
9. Chen, X. *et al.* Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res.* **44**, 6482 (2016).
10. Yarrington, R. M., Verma, S., Schwartz, S., Trautman, J. K. & Carroll, D. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proc. Natl. Acad. Sci.* **115**, 9351–9358 (2018).
11. Uusi-Mäkelä, M. I. E. *et al.* Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). *PLoS One* **13**, (2018).
12. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol. 2014 327* **32**, 670–676 (2014).
13. Stead, E. *et al.* Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* **21**, 8320–8333 (2002).
14. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science (80-.).* **343**, 80–84 (2014).
15. Liao, S., Tammaro, M. & Yan, H. Enriching CRISPR-Cas9 targeted cells by co-targeting the HPRT gene. *Nucleic Acids Res.* **43**, (2015).
16. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
17. van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **63**, 633–646 (2016).
18. Lieber, M. R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annu. Rev. Biochem.* **79**, 181 (2010).
19. McVey, M. & Lee, S. E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* **24**, 529 (2008).
20. Fu, Y. W. *et al.* Dynamics and competition of CRISPR–Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. *Nucleic Acids Res.* **49**, 969–985 (2021).
21. Gerhards, N. M. *et al.* Haploid genetic screens identify genetic vulnerabilities to microtubule-targeting agents. *Mol. Oncol.* **12**, 953–971 (2018).
22. Lau, M.-T. *et al.* Systematic functional identification of cancer multi-drug resistance genes. *Genome Biol.* **21**, (2020).
23. Jacobs, J. L., Kieboom, K., Marino, S., DePinho, R. A. & Van Lohuizen, M. The oncogene and Polycombgroup gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature* (1999). doi:10.1038/16476
24. Bruggeman, S. W. M. *et al.* Ink4a and Arf differentially affect cell proliferation and neural stem cell self-renewal in Bmi1-deficient mice. *Genes Dev.* (2005). doi:10.1101/gad.1299305
25. Liu, L., Andrews, L. G. & Tollefsbol, T. O. Loss of the human polycomb group protein BMI1 promotes cancer-

specific cell death. *Oncogene* (2006). doi:10.1038/sj.onc.1209454

26. Kreso, A. *et al.* Self-renewal as a therapeutic target in human colorectal cancer. *Nat. Med.* (2014). doi:10.1038/nm.3418
27. Maeda, A. *et al.* Targeting of BMI-1 expression by the novel small molecule PTC596 in mantle cell lymphoma. *Oncotarget* (2018). doi:10.18632/oncotarget.25558
28. Bhattacharya, R. *et al.* MiR-15a and MiR-16 control Bmi-1 expression in ovarian cancer. *Cancer Res.* (2009). doi:10.1158/0008-5472.CAN-09-2552
29. Vidi, P. A. *et al.* NuMA promotes homologous recombination repair by regulating the accumulation of the ISWI ATPase SNF2h at DNA breaks. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gku296
30. Siller, K. H. & Doe, C. Q. Spindle orientation during asymmetric cell division. *Nature Cell Biology* (2009). doi:10.1038/ncb0409-365
31. Du, Q., Stukenberg, P. T. & Macara, I. G. A mammalian partner of inscuteable binds NuMA and regulates mitotic spindle organization. *Nat. Cell Biol.* (2001). doi:10.1038/ncb1201-1069
32. Lechler, T. & Fuchs, E. Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature* (2005). doi:10.1038/nature03922
33. Abad, P. C. *et al.* NuMA influences higher order chromatin organization in human mammary epithelium. *Mol. Biol. Cell* (2007). doi:10.1091/mbc.E06-06-0551
34. Radulescu, A. E. & Cleveland, D. W. NuMA after 30 years: The matrix revisited. *Trends in Cell Biology* (2010). doi:10.1016/j.tcb.2010.01.003
35. Wu, Z. *et al.* Polycomb protein EZH2 regulates cancer cell fate decision in response to DNA damage. *Cell Death Differ.* (2011). doi:10.1038/cdd.2011.48
36. Wei, F. *et al.* BMI1 attenuates etoposide-induced G2/M checkpoints via reducing ATM activation. *Oncogene* (2015). doi:10.1038/onc.2014.235
37. Xu, C. R. *et al.* Bmi1 functions as an oncogene independent of Ink4a/Arf repression in hepatic carcinogenesis. *Mol. Cancer Res.* (2009). doi:10.1158/1541-7786.MCR-09-0333
38. Gascoigne, K. E. & Taylor, S. S. How do anti-mitotic drugs kill cancer cells? *Journal of Cell Science* (2009). doi:10.1242/jcs.039719
39. Yang, E. *et al.* Bad, a heterodimeric partner for Bcl-xL and Bcl-2, displaces bax and promotes cell death. *Cell* **80**, 285–291 (1995).
40. Zhang, J., Liu, J., Yu, C. & Lin, A. BAD Ser128 Is Not Phosphorylated by c-Jun NH2-Terminal Kinase for Promoting Apoptosis. *Cancer Res.* **65**, 8372–8378 (2005).
41. Zhou, L., Cai, X., Han, X., Xu, N. & Chang, D. C. CDK1 switches mitotic arrest to apoptosis by phosphorylating Bcl-2/Bax family proteins during treatment with microtubule interfering agents. *Cell Biol. Int.* (2014). doi:10.1002/cbin.10259
42. Eberle-Singh, J. A. *et al.* Effective Delivery of a Microtubule Polymerization Inhibitor Synergizes with Standard Regimens in Models of Pancreatic Ductal Adenocarcinoma. *Clin. Cancer Res.* **25**, 5548–5560 (2019).
43. Shapiro, G. I. *et al.* Pharmacokinetics and Safety of PTC596, a Novel Tubulin-Binding Agent, in Subjects With Advanced Solid Tumors. *Clin. Pharmacol. drug Dev.* **10**, 940–949 (2021).
44. Yadirgi, G. *et al.* Conditional activation of Bmi1 expression regulates self-renewal, apoptosis, and differentiation of neural stem/progenitor cells in vitro and in vivo. *Stem Cells* **29**, 700–712 (2011).
45. Acquati, S. *et al.* Epigenetic Regulation of Survivin by Bmi1 Is Cell Type Specific During Corticogenesis and in Gliomas. *Stem Cells* **31**, 190–202 (2013).
46. Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling. *Nat. 2017 5467657* **546**, 307–311 (2017).
47. Culver-Cochran, A. E. & Chadwick, B. P. Loss of WSTF results in spontaneous fluctuations of heterochromatin formation and resolution, combined with substantial changes to gene expression. *BMC Genomics* **14**, 1–18 (2013).
48. Wilson, B. G. *et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* (2010). doi:10.1016/j.ccr.2010.09.006

49. Kadoch, C. *et al.* Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat. Genet.* **49**, 213–222 (2017).
50. Weber, C. M. *et al.* mSWI/SNF promotes Polycomb repression both directly and through genome-wide redistribution. *Nat. Struct. Mol. Biol.* **28**, 501 (2021).
51. Shu, J. *et al.* Genome-wide occupancy of Arabidopsis SWI/SNF chromatin remodeler SPLAYED provides insights into its interplay with its close homolog BRAHMA and Polycomb proteins. *Plant J.* **106**, 200–213 (2021).
52. Menon, D. U., Shibata, Y., Mu, W. & Magnuson, T. Mammalian SWI/SNF collaborates with a polycomb-associated protein to regulate male germline transcription in the mouse. *Development* **146**, (2019).
53. Smith, I. *et al.* Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map. *PLOS Biol.* **15**, e2003213 (2017).
54. Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34**, 631–633 (2016).
55. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science (80-.).* (2009). doi:10.1126/science.1178955
56. Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* (2011). doi:10.1038/nature10348
57. Akhtar, W. *et al.* Using TRIP for genome-wide position effect analysis in cultured cells. *Nat. Protoc. 2014 96* **9**, 1255–1281 (2014).
58. Addgene: CRISPR Pooled gRNA Libraries. Available at: https://www.addgene.org/crispr/libraries/. (Accessed: 18th July 2021)
59. Bridge, A. J., Pebernard, S., Ducraux, A., Nicoulaz, A. L. & Iggo, R. Induction of an interferon response by RNAi vectors in mammalian cells. *Nat. Genet.* **34**, 263–264 (2003).
60. Schaefer, K. A. *et al.* Unexpected mutations after CRISPR-Cas9 editing in vivo. *Nature Methods* **14**, 547–548 (2017).
61. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
62. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442 (2013).
63. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
64. Grimm, D. The dose can make the poison: Lessons learned from adverse in vivo toxicities caused by RNAi overexpression. *Silence* **2**, (2011).
65. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun. 2018 91* **9**, 1–15 (2018).
66. Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc. 2017 124* **12**, 828–863 (2017).
67. Carette, J. E. *et al.* Global gene disruption in human cells to assign genes to phenotypes. *Nat. Biotechnol.* **29**, 542 (2011).
68. Fu, Y. *et al.* High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822 (2013).
69. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132 (2014).
70. Pettitt, S. J. *et al.* Genome-wide barcoded transposon screen for cancer drug sensitivity in haploid mouse embryonic stem cells. *Sci. Data 2017 41* **4**, 1–8 (2017).
71. Jong, J. de, Wessels, L. F. A., Lohuizen, M. van, Ridder, J. de & Akhtar, W. Applications of DNA integrating elements: Facing the bias bully. *Mob. Genet. Elements* **4**, 1 (2014).
72. Forment, J. V. *et al.* Genome-wide genetic screening with chemically-mutagenized haploid embryonic stem cells. *Nat. Chem. Biol.* **13**, 12 (2017).

Addendum

ENGLISH SUMMARY

NEDERLANDSE SAMENVATTING

LIST OF PUBLICATIONS

AUTHOR CONTRIBUTIONS

PORTFOLIO

ACKNOWLEDGMENTS

ENGLISH SUMMARY

Epigenetic processes alter and regulate gene expression without changing the actual DNA sequence. These heritable epigenetic processes include methylating the DNA, modifying the histone proteins that stabilize the chromosome structure, restructuring chromatin, and activating non-coding RNAs. Such reversible changes to the DNA affect the expression of proteins in both healthy and cancer cells and, hence, the cells' traits (phenotype). The reversibility of epigenetic changes and their impact on gene expression, cell growth, and cell identity make epigenetic proteins desirable drug targets.

The development of loss-of-function and gain-of-function genetic screens have provided us with powerful tools to extract comprehensive data on gene properties that can help us understand the impact of genetic and epigenetic changes on cancer properties and cancer therapies. By using these tools, we can seamlessly gather data that explain the properties that facilitate or hinder cancer therapies or other disease-mitigating technologies.

Chapter 3 analyzes how the target sequence of CRISPR–Cas9 and the regulatory landscape of the DNA affect the activity and mutation patterns of CRISPR–Cas9-mediated gene editing. By targeting the same sequence integrated into several regions of the DNA of mouse embryonic stem cells, mapping, and quantifying the resulting mutations, the activity of Cas9 could be measured. The chapter shows that the target sequence of the gRNA–Cas9 complex and its localization in the genome explained most of the differences in mutation efficiency. It also shows that small nucleotide insertions (mainly composed of one nucleotide) after Cas9 cuts correlated with specific transcriptional, genomic, and epigenomic features. Among these features, phosphorylation of RNA PolII, transcription initiation, several histone marks, or Lamin-B1 were the most prominent. Analyses of published data, sequence modeling, and the fact that these small nucleotide insertions were highly predictable based on adjacent nucleotides at the cutting site indicated that Cas9, apart from blunt cuts, also produces staggered cuts that can be "filled in" by a DNA polymerase.

In **Chapter 4**, a drug resistance screen on mutagenized human cancer cells (called HAP1) found that disruption of the gene encoding Nuclear mitotic apparatus 1, *NUMA1*, mainly implicated in mitosis, rendered cells resistant to cytotoxic levels of PTC-318, a small-molecule inhibitor of Polycomb-group (PcG) protein BMI1. Functional analyses revealed that NUMA1-expressing HAP1 cells arrested and died in mitosis when treated with PTC-318. These cell traits coincided with increased expression levels of two proteins implicated in the cell cycle, CDK1 and Cyclin-B1, which are markers of late G2/S cell-cycle phases. They also coincided with reduced expression of *BCL*-2, a gene implicated in preventing a type of cell death known as apoptosis. By contrast, loss of *NUMA1* expression protected the PTC-318-treated cells from mitotic arrest and cell death.

Chapter 5 presents combined data from three different screening approaches with haploid cells, HAP1, that confirm an inverse relationship between polycomb repressive complexes PRC1 and PRC2. Canonically, the two complexes collaborate to repress specific genes in an organism's development and in diseases, such as cancer. However, according to the study, downregulation of PRC2 proteins EZH2 or EED results in an upregulation of PRC1 protein BMI1. Findings from the enrichment screen presented in Chapter 4, a new depletion screen, and published data from a phenotype HAP1 screen served to identify chromatin remodeler WSTF as a positive regulator of H3K27me3 in cancer cells which is enriched in BMI1-inhibited cells. The combined data suggest that WSTF, as well as chromatin remodelers BAF180 (*PBRM1*) and BAF200 (*ARID2*) may directly or indirectly regulate the inverse correlation observed between PRC1 and PRC2.

Chapter 6 moves away from mutagenesis screens to describe how transient gene repression can be used to identify genes of epigenetic proteins involved in reprogramming. Specifically, the chapter describes how a short-hairpin RNA (shRNA) screen with an epigenetics-focused library identified epigenetic proteins that act as a barrier to reprogramming somatic mouse cells. The study highlights how TRIM28 acts as one of these reprogramming barriers, possibly by maintaining endogenous retroviruses (ERVs) silenced through H3K9me3 repression.

Finally, **Chapter 7** discusses the results from each of the previous chapters to highlight their contribution in identifying unknown genetic and epigenetic interactions in cancer cells. The discussion highlights how the they impact future technologies, such as CRISPR–Cas, targeted cancer therapies, or reprogramming and proposes future directions that can facilitate cancer research. Overall, this thesis outlines the power of using different types of screening tools (CRISPR–Cas9, transposons, or RNA interference), both individually or combined, to understand different aspects of cancer cells and therapies.

NEDERLANDSE SAMENVATTING

Epigenetische processen veranderen en reguleren genfuncties zonder de eigenlijke DNA-sequentie te veranderen. Deze overervende epigenetische processen omvatten methylatie van het DNA, modificatie van histon-eiwitten die de structuur van chromosomen stabiliseren, herstructurering van chromatine en de activering van niet-coderende RNA's. Dergelijke omkeerbare veranderingen in het DNA beïnvloeden de expressie van eiwitten in zowel gezonde als kankercellen en dus ook de eigenschappen van de cellen (fenotype). De omkeerbaarheid van epigenetische veranderingen, die genexpressie, celgroei en cel identiteit kunnen beïnvloeden, maakt epigenetische eiwitten zeer aantrekkelijke doelwitten voor geneesmiddelen.

De ontwikkeling van genetische screenings voor het bestuderen van verlies, dan wel verwerving van genfunctie, heeft ons krachtige hulpmiddelen gegeven om gedetailleerde informatie over geneigenschappen te verkrijgen, die ons kunnen helpen om de gevolgen van genetische en epigenetische veranderingen op de eigenschappen van kankers en kankertherapieën te begrijpen. Door deze tools te gebruiken kunnen we naadloos gegevens verzamelen die de eigenschappen verklaren die kankertherapieën en andere behandelingen faciliteren, dan wel belemmeren.

Hoofdstuk 3 analyseert hoe de doelsequentie van CRISPR–Cas9 en het regulerende landschap van het DNA de activiteit en mutatiepatronen van CRISPR–Cas9-gemedieerde genbewerking beïnvloeden. Door dezelfde doelsequentie te targeten, die geïntegreerd is in verschillende regio's van het DNA van embryonale stamcellen van muizen, en de resulterende mutaties te volgen en te tellen, kon de activiteit van Cas9 worden gemeten. Het hoofdstuk laat zien dat de doelsequentie van het gRNA-Cas9-complex, en de lokalisatie ervan in het genoom, de meeste verschillen in mutatie-efficiëntie kan verklaren. Het laat ook zien dat kleine nucleotide-inserties (meestal samengesteld uit één nucleotide) na een Cas9-knip correleerden met specifieke transcriptionele, genomische en epigenomische kenmerken. Van deze kenmerken waren fosforylering van RNA PolII, transcriptie-initiatie, verschillende histonmarkeringen, of Lamin-B1 de meest prominente. Analyses van gepubliceerde data, sequentiemodellering en het feit dat deze kleine nucleotide-inserties zeer goed te voorspellen waren op basis van aangrenzende nucleotiden op de plaats van de knip, gaven aan dat Cas9, afgezien van stompe eindes, ook overhangende eindes produceert die kunnen worden "ingevuld" door een DNA-polymerase.

In **Hoofdstuk 4** wordt beschreven hoe – in een geneesmiddelresistentiescreening op gemutageniseerde menselijke kankercellen (genaamd HAP1) – verstoring van het gen dat codeert voor Nuclear mitotic apparatus 1, *NUMA1*, voornamelijk betrokken bij mitose, cellen resistent maakt tegen cytotoxische niveaus van PTC-318, een *small molecule*-remmer van *Polycomb-group* (PcG) eiwit BMI1. Functionele analyses onthulden dat HAP1-cellen die NUMA1 tot expressie brengen, stopten met delen en doodgingen tijdens mitose, wanneer behandeld met PTC-318. Deze celkenmerken correleerden met toenemende expressieniveaus van twee eiwitten die betrokken zijn bij de celcyclus, CDK1 en Cyclin-B1, markers van de late G2/S-celcyclusfasen. Ze correleerden ook met verminderde expressie van BCL-2, een gen dat betrokken is bij het voorkomen van een type celdood dat bekend staat als apoptose. Daarentegen beschermde het verlies van NUMA1-expressie de met PTC-318 behandelde cellen tegen mitotische stop en celdood.

Hoofdstuk 5 beschrijft gecombineerde data van drie verschillende types van screenings met haploïde cellen, HAP1, die een omgekeerde relatie bevestigen tussen PRC1 en PRC2, twee polycomb-repressieve complexen. Deze complexen werken samen om specifieke genen te onderdrukken tijdens de ontwikkeling van een organisme, als ook in ziektes zoals kanker. Echter, in deze studie resulteert de neerwaartse regulatie van PRC2 eiwitten EZH2 of EED tot opwaartse regulatie van PRC1 eiwit BMI1. Resultaten van de verrijkingsscreening omschreven in Hoofdstuk 4, een nieuwe depletiescreening, en gepubliceerde gegevens van een HAP1 fenotype-screening, hielpen om chromatine-remodeler WSTF te identificeren als een positieve regulator van H3K27me3 in kankercellen (welke verrijkt is in BMI1-geremde cellen). De data tezamen suggereren dat WSTF, alsmede chromatinemodelers BAF180 (*PBRM1*) en BAF200 (*ARID2*), direct of indirect de inverse correlatie zouden reguleren tussen PRC1 en PRC2.

Hoofdstuk 6 beschrijft hoe tijdelijke remming van genen gebruikt van worden om genen te identificeren van epigenetische eiwitten die betrokken zijn bij herprogrammering. Het hoofdstuk beschrijft hoe een *short-hairpin* RNA (shRNA) screening met een bibliotheek gefocust op epigenetica epigenetische eiwitten identificeert die fungeren als een barrière voor de herprogrammering van somatische muiscellen. De studie laat zien hoe TRIM28 als een van deze herprogrammeringsbarrières dient, mogelijk door endogene retrovirussen (ERV's) te onderdrukken door H3K9me3-repressie.

Ten slotte worden in **Hoofdstuk 7** de resultaten van elk van de voorgaande hoofdstukken besproken om hun bijdrage te benadrukken bij het identificeren van onbekende genetische en epigenetische interacties in kankercellen. De discussie bespreekt de impact die deze zouden kunnen hebben op toekomstige technologieën, zoals CRISPR-Cas, gerichte kankertherapieën of herprogrammering, en suggereert toekomstige richtingen die kankeronderzoek kunnen vergemakkelijken. Dit proefschrift schetst de kracht van het gebruik van verschillende soorten screeningtools (CRISPR-Cas9, transposons of RNA-interferentie), zowel afzonderlijk als gecombineerd, om verschillende aspecten van kankercellen en therapieën te begrijpen.

LIST OF PUBLICATIONS

Gisler S, Maia ARR, Chandrasekaran G, Kopparam J, van Lohuizen, M. A genome-wide enrichment screen identifies NUMA1-loss as a resistance mechanism against mitotic cell-death induced by BMI1 inhibition. PLoS One 15, e0227592 (2020).

Gisler S, Gonçalves JP, Akhtar W, de Jong J, Pindyurin AV, Wessels LFA, van Lohuizen M. Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. Nat. Commun. 10, (2019).

Miles DC, de Vries NA, **Gisler S**, Lieftink C, Akhtar W, Gogola E, Pawlitzky I, Hulsman D, Tanger E, Koppens M, Beijersbergen RL, van Lohuizen M. TRIM28 is an Epigenetic Barrier to Induced Pluripotent Stem Cell Reprogramming. Stem Cells 35, (2017).

Norddahl GL, Wahlestedt M, **Gisler S**, Sigvardsson M, Bryder D. Reduced repression of cytokine signaling ameliorates age-induced decline in hematopoietic stem cell function. Aging Cell 11, 1128–1131 (2012).

AUTHOR CONTRIBUTIONS

Chapter 3: Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency

Authors: Santiago Gisler, Joana P Gonçalves, Waseem Akhtar, Johann de Jong, Alexey V Pindyurin, Lodewyk FA Wessels, and Maarten van Lohuizen

Contributions: SG, JPG, WA, AVP, LFAW and MvL designed the research. SG, WA and AVP performed the experimental work. WA and JdJ preprocessed the TRIP and epigenomic data. JPG designed and performed the subsequent data analysis, and made most figures. SG, JPG and WA interpreted the results. SG and JPG wrote the manuscript. LFAW and MvL revised the manuscript and supervised the analytical and experimental research.

Chapter 4: Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency

Authors: Santiago Gisler, Ana Rita R Maia, Gayathri Chandrasekaran, Jawahar Kopparam, Maarten van Lohuizen

Contributions: SG and MvL conceived and supervised the study. SG and ARRM performed the experiments assisted by JK. ARRM curated the imaging data and created the corresponding gigures, GC curated the sequencing data and created the corresponding figures, SG, ARRM, and GC analyzed the data. SG wrote the manuscript and created the remaining figures with input from ARRM, GC, JK, and MvL.

Chapter 5: *BAZ1B* is a possible regulator of H3K27me3 in an inverse correlation between PRC1 and PRC2

Authors: Santiago Gisler, Jawahar Kopparam, Ferhat Alkan, Maarten van Lohuizen

Contributions: SG and MvL conceived and supervised the study. SG and JK performed the experiments. FA analyzed the screening data and created the corresponding figures. SG wrote the manuscript and created the remaining figures with input from JK and MvL.

Chapter 6: TRIM28 is an epigenetic barrier to induced pluripotent stem cell reprogramming

Authors: Denise Catherine Miles, Nienke Alexandra de Vries, Santiago Gisler, Cor Lieftink, Waseem Akhtar, Ewa Gogola, Inka Pawlitzky, Danielle Hulsman, Ellen Tanger, Martijn Koppens, Roderick Leonardus Beijersbergen, Maarten van Lohuizen

Contributions: DCM and MvL conceived and supervised the study. DCM, NdV, and SG collected, assembled, analyzed, and interpreted data. CL and WA analyzed and interpreted data. IP conceptualized and designed experiments and collected and assembled data. EG, DH, ET, and MK collected and assembled data. RLB conceptualized and designed experiments and analyzed and interpreted data. DCM wrote the manuscript and created the figures with input from NdV, SG, and MK. DCM and MvL gave the final approval of manuscript.

PORTFOLIO

Name PhD student:	Santiago Gisler Bottaro
PhD period:	2011 – 2022
Name PhD supervisors:	Prof. dr. Maarten van Lohuizen and Prof. dr. Anton Berns

	Year
Courses	
Laboratory Animals Course for Article 9	2011
Stralingsdeskundigheidscursus	2012
Epigenetics and non-coding RNA: from mechanism to disease	2015
Writing and presenting in biomedicine	2015
Introduction to R and statistics	2015
ImageJ/Fiji: Image processing & analyses	2017
Retreats	
OOA PhD student retreats	2012-2014
Seminars and workshops	
Division research meetings (oral presentations)	2012-2018
NKI research club meetings (oral presentations)	2013-2017
NKI CRISPR mini symposium (oral presentation)	2014
NKI staff meeting: Molecular genetics (oral presentation)	2017
Meetings and conferences	
NIRM Consortium, Rotterdam (poster)	2013
NIRM Consortium, Rotterdam (talk)	2014
IMB Conference: DNA Repair & Genome Stability in a Chromatin Environment, Mainz (poster)	2015
Cold Spring Harbor Conference: Genome Engineering – The CRISPR-Cas Revolution, New York (poster)	2017
Student supervision	
Lucia Selfa Aspiroz: "*A possible novel interaction between BMI1 and PBAF proteins PBRM1 and ARID2*"	2014

ACKNOWLEDGMENTS

This thesis would not have been possible without the help, support, and friendship of people in my life. In one way or another, you have been part of developing my personality and, by extension, the resulting thesis.

First, thank you, Maarten, for giving me the opportunity to join your lab and introducing me to the team and exciting projects. You allowed me to follow up on my ideas and collaborate with other groups inside and outside the institute. I also appreciate your guidance ahead of presentations and your helpful involvement in reviewing this thesis.

Other special thanks to my co-supervisor, Ton, for your support and discussions about my projects and the thesis preparation afterward. Thanks also to the rest of my PhD committee, Hein and Jos, for your helpful advice.

I also want to convey a big thank you to my supervisors during my MSc project in Madrid, Erwin and Latifa. You and the rest of the department at CNIO activated the researcher in me. You showed me how fun research can be (which includes the hundreds of [non-precast gel] western blots produced in such a short time). Rest assured, you scarred me and I could not stop talking about my time in your department until long after leaving the lab.

Special thanks go out to my former lab members. Thanks, Martijn, my office mate and fellow PhD student for some years. Although we never shared any projects, discussing science, trips, and The Expendables 1–3 with you was always a pleasure. Maybe we'll get together and revisit these topics one day (the fourth movie seems to be out). Next, Waseem, an excellent colleague and one of the most essential pillars in my development as a researcher. You were one of the most curious, self-driven, and loving characters I've met. You wholeheartedly supported and believed in me and enjoyed sharing your extensive repertoire of scientific expertise and were genuinely open to discussing and learning from less experienced researchers like me. So, thanks to the mastermind behind my first paper as a graduate student. I'm happy to have had you as a colleague and – most importantly – a great friend; you will always be missed. Thanks also to the last setup of the MvL lab before I left: Jawa, Gayathree, Gaurav, Jitendra, and Danielle. You introduced me to the real India during our lunches. I took mental notes and feel now I know what to expect and how to orient myself in your chaotic cities. Thanks also for your help with the experiments, data analyses, and insightful scientific discussions during the years we worked together.

To the old MvL people that were part of the lab when I arrived. Thanks to the Italian band, Cesare, Michela, Gaetano, and Betty, for introducing me to the institute and research. I also enjoyed our lively discussions about cultures and how nobody really gets Italians. Of course, I still think you can get too sensitive about some aspects of your culture but let's discuss the topic over a pineapple pizza and cappuccino. Thanks to the almost-completely-Dutch-lunch-crew: Nienke, Ellen, Joep, Yme, Pauline, and Denise.

You were all involved in my formation, helping out whenever I needed. Finally, thanks to all the former members who helped in the projects in some way or another: Anke and Bas (for helping me ignite the haploid projects), Maxime, and Lucia (my first official student).

Then we have the H4 people. You were many, but a specific group comes to mind. Lorenzone, my almost-flat-mate, we moved to Amsterdam more or less at the same time, and I feel we knew each other about 2 weeks before we even met. Perhaps we've enjoyed too many beers and fogs together. Still, you never hesitate to help out when things get serious, and I know I can count on you even when the shit hits the fan. Thanks! Johann, the Friday lab member, was probably my first connection with a true computational scientist. Thanks for dropping by once a week to collaborate on projects, sharing your knowledge on informatics, and, most importantly, contributing to making Fridays a bit more laid back. Andrej, you were always busy with other stuff and never had time for lunch or drinks ('next time I'll join, man, I swear' never happened). Still, it was always great chatting with you in the hallway. You may break lunch promises, but you've also supported me tremendously, especially during the last rush of my thesis writing. I appreciate your help, my friend, and I wish you the best in your future projects. Apart from these, I also want to give big thanks to Giusi, Paul, Ivo, Colin, Hilda, John H, Nathalie, Jan-Paul, Huub, Remco, Gerjon, Margriet, Metello, Monique, Elvira.

I was fortunate to get to know NKIers from other floors through collaborations and Friday borrels. Thanks to Thijn and his lab members, Vincent, Lucas, Joppe, and Jaquelin, for introducing me to the haploidy and Hap screening world. You guided me through an exciting and, at times, frustrating haploid adventure. Thanks, Rita, for your friendship and collaboration on the haploid project. I could not have finalized the project without your impressive imaging skills; SiR-DNA magic. Thanks, Lodewyk, for collaborating on the CRISPR/TRIP project and for being helpful. Joana and Ferro, we've shared many fun moments outside the institute walls, but we also had science-related fun. Thanks for all your help and friendship. Liam and Rob, thanks for letting an outsider like me join your lab lunches; they were nice escapes from the stress in the lab. Keep doing what you do with that team, which looks like a dream lab from the outside. The sporty NKI nerds: Kong and Lorenzo; it is always fun to train and race together with a tiny competitive edge. Margarida, great to have had the pleasure of discussing CRISPR with you. Thanks for all your help once I left the NKI. A group of beer-loving nerds helped me wrap up the week and realize life was not only western blots, PCRs, and flow cytometry. Great discussions and fun at NKI's Friday borrels and afterward with Jacobien, Sedef, Judith, Flo, Joanna, Rui, Anna, Philip, Joao, Gozde, Deepani, Arne, Liqin, Ahmed, Eric, Lisanne, Chelsea, Marco.

To the staff at the flow cytometry facility, Anita, Frank, and Martijn, thank you for your patience and help with my cells. Thank you Lenny, Marjolijn, and Bram at the microscopy facility for your guidance, support, and overall positive and friendly attitude. Desiree, thank you for your help during my brief experience with radiology; always nice to talk with you! And, thanks to Ron and the rest of the sequencing facility for your efficient work and flexibility.

Life is not only research. Sports have been a critical part of my successes in science and this thesis. All work and no play make me a dull boy. First, thanks to the Carlson Gracie team in Amsterdam and Flexa for teaching me resilience by constantly smashing me, bruising me, messing up my neck, and finally leaving me with a dopamine/adrenaline-filled smile. That's what true teammates do for each other in the gentle art of catharsis. Gracias también a Alessandro Nagaishi y los excelentes personajes de CheckMat Medellín que me mantuvieron físicamente y psíquicamente sano durante nuestro tiempo en Medellín. Finally, thanks to the handful of ATACers I got to know, with whom I swam/biked/ran over the years. I never paid a membership, but you never whined when I ate your ATAC burgers.

I also have a mental group of close friends from the Netherlands that I love. Renato & Quirine, Petrit & Ale, Arnaud, Michael & Jitske, Santi(ago) II, Manos, Álvaro, thank you all for being part of my inner circle, sharing good and crap moments with me, and listening to my frustrating PhD stories. I'll share a few words about each of you when we meet. Petrit, an extra thanks for being my paranymph. It's always great to have you around, pička, whether for dumb hangouts or a bit more sophisticated ones (with emphasis on "a bit"), including our never-ending philosophical and political debates. Thanks to the rest of my friends around the Netherlands: Phil & Fred, Lidra & Alex, Fred & Celine, Branca & Vladimir, Rion & Barbara, Davide, Lauriane, Naomi & Craig. You all kept me sane and active in one way or another.

Stort tack till mitt svenska gäng, kärnan i min umgängeskrets, älskar er med. Tack Sidde, Stoffe & Victoria, Jonny & Emma, Mea & Anna, Reggie, Baroj, Sara, MK, Andy, Apat & Malin. Jag kan lovprisa er alla och kanske gör så nån gång i dimman. Ni utgör en otroligt viktig del av mitt liv och har alltid stöttat mig på ett eller annat sätt; i med och motgång. Och så vänner som David, Vittorio och M S-ÅRK Jönsson; jag har många bra minnen med er och det är skönt att veta att vi fortfarande kan ha kontakt med varandra lite då och då. Gracias tambien a nuestros queridos amigos en Medellín, Juan Guillermo & familia y Walter.

Merci à tous mes amis français, Antoine, Claire, Sushi, Vincent, Laura, Kevin Aha, Julie, Mawine, Bruno, Mika, Pasca, Kevin, Arnaud, Marine, Elo, Gabriel, PePe, TeTe, BeuBeu. Vos distractions ont été une partie importante dans mon voyage. A bientôt dans le sud !

Merci beaucoup aussi à ma famille Française : papa Joel, mama Sofie, Romain, Flore, Remi. Cela a commencé avec la langue des signes, les langues brisées et un Atlas, mais, d'une manière ou d'une autre, nous parvenons toujours à nous comprendre et à apprécier notre compagnie. Vous avez été d'un soutien incroyable pendant ce voyage ! Avec cette thèse à l'écart, je vais me concentrer davantage sur la réparation de mon français cassé.

A mi familia, muchas gracias por todo el apoyo y amor que me han dado y que me siguen dando; los quiero mucho (tqm). Mamá, sé que no es fácil acordarse de todos los proyectos o temas con cual trabajo. Sin embargo, me impresiona tu esfuerzo para acordarte de las palabras claves en orden de poder explicarlos a tus amig@s. Admiro tu fuerza, energía y dedicación que siempre demostraste y que me ha inspirado a nunca darme por vencido. Gracias por apoyarme y soportarme en tiempos difíciles y buenos. También me hubiese encantado poder disfrutar de estos momentos con mi padre. Tus historias, teorías y optimismo en la vida fueron fundamentales para mi manera de solucionar problemas, dirigirme en la vida y formar ideas (por males o bienes). Te extraño. Mis queridos hermanos y mejores amigos, Ramiro y Mariana, siempre me alentaron y sufrieron conmigo, como hinchando la Celeste en 2010 (como ven, el sufrimiento forma parte). Ramiro, gracias por introducirme a la vida científica; no hubiese conseguido esto sin tu ayuda y apoyo incondicional. Mariana, gracias por empujarme a pensar independientemente y encontrar mi propio camino; siempre lo tengo en cuenta. Gracias a mis sobrinos queridos, Linn, Nina, Julián, Ida, Elvis y Sötis. Especialmente los tres primeros, los quiero mucho. Y a mis queridos cuñados, Cilla y Robert, tack und danke schön for the good times and great moments we've spent together.

Le meilleur pour la fin. Ce tour de montagnes russes aurait été complètement différent si la personne la plus importante de ma vie n'avait pas été a mes cotés avec son sourire réconfortant. Je te serai toujours reconnaissant, Anaïs, d'être la partenaire la plus attentionnée et la meilleure que l'on puisse souhaiter. Tu es restée à mes côtés pendant tous ces rebondissements et m'a motivé a ne jamais abandonner. J'imagine que cela n'a pas toujours été facile, mais j'espère que l'espace de mon cerveau libéré compensera les moments de stress. Je t'aime, Banana !