THE UNIVERSITY OF CHICAGO

DEFINING THE PATHOGENESIS OF DEL(7q) IN HEMATOPOIETIC STEM CELLS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:

CANCER BIOLOGY

BY

MATTHEW RICHARD MANUEL JOTTE

CHICAGO, ILLINOIS

AUGUST 2024

For Amanda

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The McNerney lab has been a great source of scientific and personal support throughout my time in graduate school. The graduate student community is one of the lab's strongest assets. Thank you to Tanner Martinez, Raven Moten, Kasia Zawieracz, Yuqing Xue, and Weihan Liu for making graduate school far more bearable. I hope to remain a part of The Bros™ group chat and hope it continues to expand. Tanner in particular has been a great colleague – he is always willing to discuss science and talk about experiments, figures, and big ideas. Thank you for helping me troubleshoot transplants in the early years, sorting many RNA-sequencing replicates, and showing me how to make heatmaps in R Studio. Our shared experience as students in the McNerney lab and The University of Chicago MSTP, especially handling Grand Rounds catering and the 2023 admissions revisit events, and both serving as internal F30 study section review chairs, have been a strong bonding force and I look forward to returning to medical school together. I would also like to thank Raven and Yuqing for their help on the final mechanistic experiments of the thesis project. I have enjoyed watching Raven's sense of humor flourish as she progresses through graduate school. I wish our newest graduate member, Hunter Blaylock, all the best as he starts his own Ph.D. journey in the lab.

I would also like to thank Dr. Angela Stoddart, Dr. Madhavi Senagolage, Saira Khan, and Dr. Ningfei An for their guidance and mentorship in the lab. Angela in particular had a large contributing role to the direction of the thesis project, and I appreciate all the time we talked about experimental design, overarching scientific questions, and life in general. Madhavi has been an exceptional labmate, and I have enjoyed how much we can relate to each other over our five years in the lab. As with Hunter, I wish the newest

post-graduate member of the lab, Dr. Joe Cannova, best of luck. Former members of the McNerney lab, especially Dr. Molly Imgruet, Dr. Jeff Kurkewich, Bonnie Hu, and Henna Nam, have also contributed to the project and my growth as an academic scientist.

I first met Dr. Megan McNerney when I interviewed for The University of Chicago MSTP, and was immediately drawn to her science and way of thinking about problems, particularly clonal hematopoiesis. Thank you for taking me into your lab; I know I am leaving it a more confident, independent scientist. I will always admire your deft scientific writing, and I appreciate your willingness to guide me through grant proposals, conference abstracts, posters, manuscripts, and peer review. I would also like to thank my thesis committee, Dr. Alex Ruthenburg, Dr. Eileen Dolan, and Dr. Steve Kron. Dr. Kron has been an excellent thesis chair, and his consistent willingness to attend my lab meeting and work-in-progress presentations, email relevant papers, and meet to discuss data is a testament to his commitment to science and trainees. I am a better scientist for it. Dr. Andrea Piunti has also been extremely helpful discussing data, particularly for advancing the final mechanistic experiments. Thank you all for your time and mentorship.

This project would not have been possible without help from the Flow Cytometry Core, DNA Sequencing Facility, Genomics Core Facility, Animal Resources Center, and Integrated Light Microscopy Core. I am also grateful to The University of Chicago MSTP, especially Dr. Lucy Godley, and the Committee on Cancer Biology.

Both my father, Dr. Robert M. Jotte, and my grandfather, Dr. Angelo M. Alves, attained their M.D./Ph.D. degrees and are role models for showing how fulfilling and stimulating this career path can be. Finally, to Amanda, thank you for staying with me as I came to Chicago to pursue this training, and always being a source of love and support.

PREFACE

The work described in this thesis is original and was conceptualized, performed, and analyzed in the McNerney lab. I designed, performed, analyzed, and described the results for all experiments in this thesis, with assistance from Angela Stoddart with the multiplex knockout mouse cohorts, Tanner Martinez with transplanting mice and sorting cells for RNA-sequencing, Raven Moten with the immunofluorescence micrographs, Yuqing Xue with the comet tails, Molly Imgruet with the U937 dose-response curves, and Henna Nam and Bonnie Hu with the CRISPR screen data mining. Jermaine Austin performed some of the DNA isolation from blood samples of the multiplex knockout mouse cohorts. Sandeep Gurbuxani assisted with imaging and analysis of the stained spleen sections. Ningfei An and Saira Khan were helpful with promoting a lab environment conducive to productive research, including ordering and organizing reagents. The work in Chapter 1 of this thesis has been submitted for peer review and publication. Some of the information in the introduction of Chapter 1, including Figure 1, was adapted from a review I authored with Megan McNerney, which is listed as Reference 146.

Chapter 2 is an ongoing work in collaboration with Tanner Martinez. Together, we conducted and analyzed the presented experiments, and produced the figures and tables.

ABSTRACT

Loss of all or part of chromosome 7 [-7/del(7q)] is recurrent in myeloid neoplasms and associated with a poor response to chemotherapy. Chromosome 7-encoded genes that drive drug resistance and the consequences of combinatorial 7q tumor suppressor gene loss have remained unclear, the latter question being largely due to the challenges of modeling aneuploidy. Here, we use *in silico* data mining to uncover 7q genes involved in chemotherapy resistance. We establish murine models of del(7q) clonal hematopoiesis and drug resistance with multiplex CRISPR-Cas9-mediated inactivation of four genes, *Cux1*, *Ezh2*, *Kmt2c*, and *Kmt2e*. Post-genotoxic exposure, combined deficiency of *Cux1* and *Ezh2* preferentially promotes clonal myeloid expansion *in vivo*, with compounding defects in DNA damage recognition and repair. Experiments in human acute myeloid leukemia cell lines similarly illustrate central roles for *CUX1* and *EZH2* loss in cellular survival and DNA damage resolution following chemotherapy exposure. Transcriptome analysis reveals combined *Cux1* and *Ezh2* loss recapitulates -7 patient gene signatures and defective DNA damage response pathways, to a greater extent than single gene loss. This work reveals a genetic interaction between *CUX1* and *EZH2*, and sheds light on how -7/del(7q) contributes to leukemogenesis and drug resistance characteristic of these adverse-risk neoplasms. These data support the concept of 7q as a contiguous gene syndrome region, in which combined loss of multiple gene drives pathogenesis. Further, our CRISPR-based approach may serve as a framework for interrogating other recurrent aneuploid events in cancer.

CHAPTER 1

INTRODUCTION

Aneuploidy in human tissues

The human genome consists of 22 autosomes and 1 sex chromosome. Somatic cells contain two genome equivalents, with one set each originating from maternal and paternal gamete contributions. Cells possessing a full complement of 46 chromosomes are referred to as ‘euploid’, and any deviation from this number results in aneuploidy. The gain or loss of chromosomal material that arises with aneuploidy confers a dosage imbalance to all encoded genes within the affected region, resulting in excessive or insufficient gene products.[1] Hundreds to thousands of genes can be affected depending on the size of the chromosome, which has far-reaching effects on nearly every facet of cell biology, including protein homeostasis, metabolism, and proliferation.[2–4] Consequently, aneuploidy is not well-tolerated in humans, particularly during development.

Aneuploidies that arise early in embryonic development, either from chromosomal mis-segregation in a parental gamete or during one of the primary zygotic mitoses, result in an organism in which all or nearly all cells harbor an aberrant karyotype. Given the diverse cellular consequences of an unbalanced genome, aneuploidy is generally embryonic lethal in humans and is the leading cause of early miscarriages.[5] Autosomal monosomies, the loss of a chromosome, are incompatible with viability. Intriguingly, there are three autosomal trisomies, the gain of a chromosome, that can persist to birth: Trisomy 13, known as Patau syndrome; Trisomy 18, known as Edwards syndrome; and Trisomy 21, known as Down syndrome. Infants with Trisomy 13 or 18 have extremely

shortened lifespans, with only 10% of affected individuals surviving to one year of age, and display severe developmental defects including cardiac and renal malformations.[6–8] Trisomy 21 is the best-studied aneuploidy in humans. Affected individuals display a number of characteristics affecting multiple organ systems, including short stature, intellectual disability, and congenital heart defects.[9] Individuals with Trisomy 21 have a reduced life expectancy to around 60 years,[10] notably shorter than the average of approximately 80 years in the U.S. but substantially longer than the other autosomal trisomies, perhaps reflecting the fact that human chromosome 21 is the smallest of all the autosomes, comprising only 1.5-2% of the genome.[11]

Aneuploidies of the sex chromosomes are better tolerated – so long as at least one X chromosome is present, all single-chromosome deviations are compatible with survival, though many gestations still spontaneously terminate.[12,13] The most common sex chromosome aneuploidy is Klinefelter syndrome, resulting from an XXY trisomy that notably impairs testicular development and spermatogenesis.[14] An XYY karyotype, known as Jacob's syndrome, has a milder impact on primary and secondary male sexual development but still increases overall morbidity, likely due to disruptions to the psychiatric, endocrine, and neurological systems, among others.[15,16] Trisomy X, also known as Triple X syndrome, is the mildest of the sex chromosome aneuploidies; many cases go undiagnosed and the condition appears to primarily affect behavioral and psychiatric development.[17] The sole sex chromosome monosomy, known as Turner syndrome, results from an XO karyotype and confers a variety of medical and developmental problems, including short stature, infertility, and ovarian and cardiac dysgenesis.[18] The reduced phenotypic severity of sex chromosome aneuploidies may

reflect the increased propensity for dosage compensation on these chromosomes, which is less common in the autosomes.[19]

Maintenance of genome integrity is therefore critical to successful organismal development, and eukaryotic cells have evolved extensive mechanisms to safeguard chromosome copy number. The primary protective mechanism against chromosome segregation errors is the spindle assembly checkpoint. During cell division, individual chromosomes are duplicated and the resulting sister chromatids are connected to each other at the centromere and held together by cohesin complexes.[20] Microtubules originating from centrosomes at opposites poles of the cell form mitotic spindles that attach to kinetochore complexes assembled on the outer surface of the centromere, resulting in bipolar attachment and stable alignment of chromosomes to the metaphase plate via opposing spindle tension. Cleavage of the cohesin complexes by a protein called separase results in loss of sister chromatid cohesion and the initiation of chromosome segregation.[21] The spindle assembly checkpoint prevents premature exit from mitosis by halting cell cycle progression in the presence of unattached or improperly attached kinetochores.[22] Defects in the spindle assembly checkpoint greatly increase the frequency of aneuploid events, as do merotelic spindle attachments to kinetochores in which one kinetochore is attached to both poles, and premature loss of sister chromatid cohesion.[1]

When these safeguards fail in meiosis, aneuploid gametes are produced that may impair the ability of the parent organism to produce viable offspring, but ultimately have little physiologic impact on the organism itself. In contrast, segregation errors during mitosis result in aneuploid daughter cells that are retained as part of the parent tissue. In

tissues with high turnover, such as the epithelial lining of the intestine, such errors arising during replication are rapidly lost as cells are routinely shed. Further, terminally differentiated cells typically have diminished reproductive capacity which can limit the impact of a deleterious change.[23] However, errors arising in highly proliferative stem and progenitor populations will persist and be passed on to every daughter cell as the afflicted cell continues to divide. The resultant population may accumulate to detectable levels, a phenomenon known as somatic mosaicism. Somatic mosaicism arising from point mutations is common and increases with age, and has been detected in every organ yet examined.[24] Estimates of the frequency of aneuploidy in somatic tissues vary by methodology; single-cell sequencing indicates fewer than 1% of human neurons and skin fibroblasts are aneuploid.[25] Still, as only a small population of somatic cells is affected, abnormalities that would be lethal in the germline may be tolerated and, in some circumstances, may even be selected for.[26]

Aneuploidy in cancer

The consequences of aneuploidy are generally detrimental to cell physiology. The immediate alteration to gene dose that occurs with chromosome mis-segregation has effects at both the RNA and protein level: genes encoded on the aneuploid chromosome display transcriptional upregulation or downregulation proportional to copy number gain or loss, and protein levels are similarly affected.[27] Dosage compensation is minimal, particularly for genes encoding components of multi-protein complexes.[27] Globally, aneuploid cells exhibit upregulation of genes related to stress response pathways, similar to the stress response from heat shock and oxidative stress, and this ‘aneuploidy stress

response' is conserved across yeast, plant, mouse, and human cells.[28] At the protein level, the sudden increase in copy number from a trisomy can result in an increased protein load, saturating the chaperone and proteasome quality-control systems and leading to persistent protein aggregates.[29] Further, aneuploidies can disrupt ribosome stoichiometry as at least one component of the ribosome is encoded on every chromosome in the human genome except chromosome 7.[30] Impaired ribosome formation can lead to ribosome biogenesis stress and p53 activation, with subsequent cell cycle arrest and potential apoptosis initiation.[31]

Such changes generally reduce the fitness of the affected cell. Indeed, cells with deficiencies in ribosome biogenesis are eliminated in *Drosophila melanogaster* models, and murine aneuploid hematopoietic stem cells (HSCs) are outcompeted by their euploid counterparts, in part due to decreased proliferation.[32,33] These observations are at odds with the fact that aneuploid karyotypes are incredibly prevalent in human cancers, with over 90% of solid malignancies and 75% of hematopoietic malignancies deviating from the balanced complement of 46 chromosomes.[34] The high prevalence of aneuploidy in cancer suggests a pathogenic role, and aneuploidy was first proposed as a cause of tumorigenesis by Theodor Boveri in the early 20th century.[34,35] However, there is no general consensus among the scientific community. The sheer variety of chromosomal changes detected in human cancer, illustrated by the Mitelman database that currently has over 77,000 cataloged patient karyotypes,[36] as well as the observation that the degree of karyotypic complexity correlates with tumor evolution from benign to malignant,[37] have led to suggestions that aneuploidy is irrelevant to tumor initiation or an inconsequential side-effect of transformation.[38,39] Still, the recurrence of specific

chromosomal alterations in cancer, many of which are tissue-specific, suggests at least some aneuploid events contribute to the disease state.

The conflicting findings that aneuploidy can impair proliferation but is widespread in cancer, a disease of uncontrolled proliferation, have led to what is termed the 'aneuploidy paradox' and extensive efforts have been made to identify contexts in which aneuploidy can confer benefit.[40,41] One emerging theme is that aneuploidy may confer a fitness advantage under stress conditions, particularly in the presence of genotoxic agents. For example, colorectal cancer cells engineered to carry an additional copy of chromosome 13 have reduced proliferation compared to disomic parent cells in standard culture conditions, but significantly increased proliferation in both serum-free and hypoxic conditions, as well as in the presence of the 5-fluorouracil, a cytotoxic chemotherapy agent used to treat colorectal cancer.[42] Intriguingly, the trisomy 13 cells also displayed a higher proliferation rate than trisomy 7 cells in these same suboptimal conditions, notable as trisomy 13 is specifically found in colorectal cancer whereas trisomy 7 is common across other cancer types. In a different set of experiments, BRAF-mutant melanoma cells treated with an inhibitor of the spindle assembly checkpoint kinase MPS1, which induces chromosomal instability and mis-segregation during the treatment window, gradually decreased in frequency in standard culture conditions after mixing with equal numbers of untreated cells.[43] However, these same cells completely overtook the culture in the presence of vemurafenib, a targeted chemotherapy for BRAF-mutant melanoma, further suggesting that aneuploidy can confer context-specific fitness advantages. Such experiments also carry the ominous implication that aneuploidy can contribute to chemotherapy resistance.

Increasing emphasis is being placed on identifying whether the effects of an aneuploidy are mediated by a specific gene or genes encoded in the altered region. In support of this concept, many chromosomal aberrations are not whole chromosome aneuploidies and instead only affect specific segmental regions. For example, the short arm of chromosome 17 (17p) contains *TP53*, a tumor suppressor gene aptly named the 'guardian of the genome' for its myriad roles in DNA repair, cell cycle, and apoptosis. 17p deletions occur widely in cancer and invariably include *TP53*,[44–46] suggesting the deletions function as a mechanism to eliminate one copy of this critical gene. In addition, many cancers with 17p deletions harbor mutations in the remaining *TP53* allele, strengthening the premise that *TP53* is the target of 17p loss.[47] However, these deletions are large and often encompass hundreds of neighboring genes; intriguingly, acute myeloid leukemia (AML) patients with both a *TP53* mutation and 17p deletion have a worse prognosis than those with homozygous *TP53* mutations, suggesting the adjacent genes co-deleted with *TP53* in 17p loss are not innocuous passengers.[47] Indeed, shRNA knockdown of either *Trp53* or *Eif5a*, two genes encompassed by nearly all lymphoma-associated 17p deletions, in *Eμ-Myc* mouse HSCs accelerated lymphoma development, which was further enhanced by combined knockdown of both genes.[47] These experiments provide compelling evidence that multiple genes within an aneuploid segment can be important for a cancer phenotype, and, critically, that the effects of altered gene dose can be additive or synergistic.

Multiple lines of evidence have emerged that further support aneuploid regions housing multiple genes related to cancer development, and having distinct phenotypes from alterations to individual genes. Co-suppression of three genes mapping to

chromosome band 8p22, commonly deleted in a number of solid cancers including lung cancer, breast cancer, and hepatocellular carcinoma (HCC), synergistically promoted tumor growth in a mouse model of HCC.[48] Multiple regions of chromosome 3p, deletions of which are nearly ubiquitous in small and non-small cell lung cancers, have been demonstrated to contain genes relating to lung development, carcinogen sensitivity, and cell cycle progression, among other pathways.[49] Deletions of the long arm of chromosome 5 (5q), recurrent in myelodysplastic syndrome (MDS) and AML, have been localized to two regions that are minimally necessary for clinical symptoms at chromosome bands 5q31 and 5q32-33.[50–52] Heterozygous loss of the ribosomal gene *RPS14* on 5q33 recapitulates the anemia observed in del(5q) patients, and reduced expression of the nearby microRNAs miR-145 and miR-146a causes thrombocytosis characteristic of del(5q).[53,54] Combined single-copy loss of the orthologs of *EGR1*, located on 5q31, and *APC*, located on 5q23, accelerates the development of anemia and monocytosis, features of MDS, in mouse models with reduced *Trp53*.[55]

Of note, for many segmental deletions the homologous chromosome remains intact, and there are few if any mutations or epigenetic silencing of the second alleles of genes contained in the deletion region. In addition to hindering identification of candidate genes that may contribute to the disease phenotype, this also suggests that affected genes are haploinsufficient, whereby single-copy loss is sufficient to induce a pathogenic phenotype. Experimental evidence supports that many 5q genes act in a haploinsufficient manner, in contrast to Alfred Knudson's 'two-hit hypothesis' that tumor suppressor genes require bi-allelic inactivation.[56] A recent computational analysis of over 8,200 tumor-normal tissue pairs suggests that the majority of the genome is dose-sensitive, and that

cumulative haploinsufficiency and triplosensitivity of critical gene clusters driven by chromosome loss or gain, respectively, can predict the complex patterns of aneuploidy characteristic of cancer genomes.[57] Further, this analysis suggests that the classic 'two-hit' model from Knudson, exemplified by secondary *TP53* mutations in cases with del(17p), may be the exception rather than the norm. In this way, aneuploidy may serve to simultaneously alter the copy number of multiple dose-sensitive linked genes.

Though aneuploidy likely contributes to disease development, it is unlikely to be sufficient for transformation as a sole abnormality. The simplest route to transformation in human cells disrupts the pRb and p53 tumor suppressor pathways and activates the telomerase and RAS proliferation pathways, converging on multiple hallmarks of cancer.[38,58,59] There is no conclusive evidence that any one chromosomal aneuploidy can sufficiently perturb these pathways to the degree necessary for full malignant transformation, though trisomy 21 causes a transient abnormal myelopoiesis antecedent to AML, and some structural rearrangements create potent fusions that drive disease.[60,61] Still, there is increasing documentation that aneuploid events can arise early in tumorigenesis and even potentially serve as initiating events. In colorectal cancer, inactivating mutations in the *APC* (adenomatous polyposis coli) gene on chromosome 5q are considered tumor-initiating events, activating WNT signaling pathways that ultimately reduce the probability of a cell being shed into the intestinal lumen.[62–65] Accordingly, deletions of 5q containing *APC* have been observed at early stages of neoplasia and many precancerous lesions harbor additional major chromosomal abnormalities, suggesting chromosome instability occurs quite early.[62,66]

Chromosome 7 deletions in hematopoietic malignancies

Cancers of the hematopoietic system, which gives rise to all lineages of blood, are generally classified as either myeloid or lymphoid according to the branch of hematopoiesis that is disrupted. These diseases, which are thought to originate in hematopoietic stem cells,[67] tend to have a relatively low mutational burden among the spectrum of human cancers, but are notable for their high degree of chromosomal translocations.[68] It is estimated that over 50% of leukemias, and nearly all lymphomas, exhibit or are caused by translocations, which can generate novel fusion proteins or forcibly alter the expression level of an otherwise unmutated gene to drive disease development.[69] The reasons for such a high prevalence of translocations are still under active study, and at least for lymphoid disease may be related to the V(D)J somatic recombination events that occur during the early development of B and T lymphocytes.

The most prevalent malignancies of the myeloid lineage of hematopoiesis are myelodysplastic syndrome and acute myeloid leukemia. AML is the most common form of acute leukemia in adults and also has the shortest survival, accounting for over 60% of leukemic deaths.[70] Though AML has its own repertoire of recurrent translocations, chromosomal aneuploidy is also routinely observed in cytogenetic analyses, most commonly gains of chromosomes 8 and 13, 17p deletions, 5q deletions, and loss of all of part of chromosome 7 [-7/del(7q)].[71]

Chromosome 7 loss was first described in a brief report from 1964, which proposed a new disease consisting of a refractory anemia and myeloid hyperplasia associated with chromosome 7 absence that frequently developed to acute leukemia.[72] It is now appreciated that -7/del(7q) is among the most common chromosomal abnormalities

across essentially all myeloid malignancies, including both pediatric and adult disease.[73,74] Interestingly, chromosome 7 alterations in hematopoietic cancers are almost always deletions or copy-neutral loss of heterozygosity (CN-LOH); in solid cancers, chromosome 7 gains or focal amplifications are observed, thought to increase the dose of driving oncogenes including *EGFR* and *MET*.[57,75] Within the hematopoietic field, the sheer prevalence of -7/del(7q) strongly suggests these deletions contribute to pathogenesis, but the specific contributions to disease initiation and development remain unknown.

-7/del(7q) is found in 5-10% of AML and adult MDS, 40% of pediatric MDS, 40% of myeloid neoplasms arising from cancer predisposition syndromes, and 50% of therapy-related myeloid neoplasms.[76–79] Deletions are associated with higher-risk MDS, faster time to transformation to AML, and poor overall survival in AML, and are considered adverse prognostic events.[76,80] Within these diseases, -7/del(7q) often co-occurs with 5q deletions, chromosome 8 gains, and RAS pathway mutations, but also frequently occurs as an isolated cytogenetic event.[81] Intriguingly, a recent whole exome sequencing study of pediatric MDS identified chromosome 7 deletions as the sole abnormality in 13% of cases – no other karyotype alterations or coding mutations were detected.[78] Though it is possible non-coding changes were present but not detected, such data raise the question as to whether monosomy 7 is sufficient to promote MDS. Spontaneous remission of monosomy 7 in pediatric MDS with subsequent disease resolution has also been reported, albeit rarely, further suggesting -7 is critical for enabling the disease state.[82,83]

There is also accumulating evidence that -7/del(7q) can serve as an initiating event in myeloid disease. In addition to cases in which -7 is the sole detectable abnormality,[78]

analyses of clonal hierarchies in myeloid neoplasms have demonstrated that chromosome 7 deletions can be founding events.[84,85] The degree to which chromosome 7 loss impacts the fitness of an HSC when it occurs so early remains unclear. Aneuploidy in HSCs reduces proliferation rates, and though no ribosome components are encoded on chromosome 7, there are likely indirect effects on ribosome biogenesis.[30,33] For example, the *SBDS* gene on 7q11.21 encodes a highly conserved ribosome maturation factor essential for complete ribosome assembly;[86] mutations within the gene are associated with Schwachman-Bodian-Diamon syndrome, an autosomal recessive bone marrow failure syndrome for which the gene is named, and even partial reduction in *SBDS* expression impairs hematopoiesis.[87,88] Consequently, it seems likely that HSCs with founding -7/del(7q) events, and indeed any aneuploidy, would face selective pressure to acquire secondary mutations that alleviate associated stresses. Regardless, the fact remains that chromosome 7 deletions can occur early in disease development.

Whether the effects of -7 and del(7q) are equivalent remains an open question. Monosomy 7 and del(7q) are often group together clinically despite differing mechanisms of occurrence: monosomy 7 results from a chromosome segregation failure, whereas del(7q) results from chromosomal breakage. Some studies have assessed -7 separately from del(7q) and report better prognosis for del(7q),[89,90] though others have found no difference.[85,91] More genes are affected by a monosomy 7 than a del(7q) event, perhaps accounting for these observations, but the recurrence of 7q deletions implies many, if not most, of the important chromosome 7 genes reside on the long arm. Additionally, there is heterogeneity in the breakpoints for 7q deletions, and whether different deletion spans carry unique prognostic implications remains unclear.

Clonal hematopoiesis and chromosome 7 alterations

Many large-scale sequencing studies have also identified chromosome 7 alterations in clonal hematopoiesis of indeterminate potential (CHIP), the phenomenon by which otherwise healthy individuals with no history of hematopoietic malignancy harbor low-frequency variants associated with leukemia in their blood, including *DNMT3A*, *TET2*, *ASXL1*, and *TP53*.[92–96] Somatic mutations accumulate as an organism ages, with adult HSCs acquiring approximately 14 mutations per year,[97] and by chance some of these mutations occur in coding regions. If the mutation confers a fitness advantage, the cell bearing it should expand and become over-represented in the blood – the hematopoietic equivalent of somatic mosaicism. Improved deep sequencing technologies have revealed that clonal hematopoiesis is ubiquitous in adults, with one study identifying mutations in leukemia-associated genes in 95% of individuals, albeit some at extremely low variant allele frequencies (VAF).[98] The clinical threshold for CHIP is a VAF of ≥2%, and by this definition CHIP increases with age and is estimated to occur in approximately 20% of elderly individuals.[99] This observation raises the question of how to reconcile the high prevalence of CHIP with the clearly lower incidence of hematopoietic disorders. As the presence of CHIP is associated with increased risk of developing a hematologic malignancy, this discrepancy highlights the need to understand which individuals are at highest risk of progression.[100]

The presence of mutations in the absence of overt disease is not unique to hematopoiesis; the number of cell divisions that occur during ontogeny, as organ systems are generated from a single-cell zygote, is vastly higher than the number required to maintain self-renewing adult tissues.[101] Even in the hematopoietic system, which requires

continuous output across life, the number of cell divisions needed to constitute the complete blood production network and HSC pool is certainly higher than the steady-state division rate of adult HSCs, which is quite low.[102] Consequently, it is estimated that a substantial fraction of mutations occur and accumulate during development, at odds with epidemiologic data that the majority of cancers occur later in life.[103] The disconnect between cancer incidence and time-dependent mutation accumulation suggests oncogenic mutations are necessary but not sufficient for tumorigenesis.[101] This premise holds even for strong driver mutations – BCR-ABL translocations, the initiating event of chronic myeloid leukemia (CML),[104] can be detected in leukocytes of up to 30% of healthy individuals over sufficient lengths of time to suggest the alterations reside in HSCs, but the incidence of CML is far less than 30% of the general population.[105,106]

Considerable effort has therefore been dedicated toward identifying interacting factors that drive clonal expansion of mutant cells. The size of the CHIP clone and number of mutations are primary metrics that influence likelihood of progression, as are the specific gene or genes that are mutated.[99] Exposure to cytotoxic therapy, such as anti-cancer treatment for a solid malignancy, can promote expansion of clones with mutations in DNA damage response genes, including *TP53* and *PPM1D*.[107,108] Regenerative hematopoiesis after HSC transplant can promote expansion of *DNMT3A* mutations and actually decreases the VAF of some *PPM1D* mutations, demonstrating that the behavior of different CHIP clones is impacted both by the type of stress as well as the specific mutations they harbor.[107] Inflammatory signaling and the microbiome have also been shown to influence clonal expansion.[109,110] Collectively, these observations support a

model in which selective pressures from multiple sources alter the fitness landscape of the tumor microenvironment and influence the oncogenic potential of different mutations.

The incidence of chromosomal alterations in CHIP is lower than that of point mutations, but similarly increases with age to an estimated 2-3% of elderly individuals and also increases the risk of developing a hematologic malignancy.[92,93] Two recent studies examined the effects of copy number alterations in CHIP in conjunction with point mutations, in large cohorts from Memorial Sloan Kettering (MSK) and BioBank Japan.[111,112] Both groups detected chromosome 7 deletions and CN-LOH at similar levels as prior studies. Intriguingly, every 7q CN-LOH event in the MSK cohort co-localized with an *EZH2* mutation, located on 7q36.1, suggesting this alteration served to duplicate a previously acquired *EZH2* mutation.[111] This has also been reported in previous studies of chromosome 7 uniparental disomy,[80,113,114] providing supporting evidence that *EZH2* is a critical gene on 7q. However, none of the -7/del(7q) events in the MSK cohort had a second *EZH2* mutation. The BioBank Japan study found that individuals with del(7q) and 7q CN-LOH had significantly increased risk for subsequent development of hematologic malignancy, particularly myeloid disease.[112] The risks associated with chromosome 7 abnormalities were similar to those of 17p, which contains *TP53*, indicating del(7q) CHIP is a biomarker for risk of disease progression that warrants close monitoring.

Cell extrinsic factors likely have a large degree of influence on cells with -7/del(7q), evidenced by the high prevalence in therapy-related myeloid neoplasms (t-MN), high-risk malignancies arising after prior exposure to chemotherapy or radiation.[79] CHIP is also associated with increased risk of t-MN development,[95,115] suggesting a model in which a pre-existing clonal population harboring -7/del(7q) is selected for by radiation and/or

chemotherapy, and subsequently expands and transforms to t-MN. Chromosome 7 deletions are also found in hematopoietic cells of benzene-exposed workers as well as AML in elderly patients, which often resembles t-MN, supporting that -7 may be selected for in the context of environmental exposure and aging.[116,117] This echoes observations that chemotherapy can select for *TP53* and *PPM1D* mutations, and that aneuploidy can be advantageous in stress contexts.[43,107] Chromosome 7 loss may therefore allow cell survival and persistence in a suboptimal bone marrow environment. Further, -7/del(7q) may contribute to the inherent drug resistance of t-MNs as the initiating population has already undergone selection for increased fitness under genotoxic stress. The identification of specific chromosome 7 genes modulating expansion and drug resistance has therefore become the focus of intense efforts.

Chromosome 7 tumor suppressor genes

Except for *EZH2*, there are few genes on chromosome 7 that have recurrent bi-allelic inactivation in myeloid disease.[114,118] Accordingly, chromosome 7 tumor suppressor genes (TSG) are likely to act in a haploinsufficient manner, similar to del(5q) genes. The lack of 'second hit' mutations is a barrier to identifying critical TSGs, but alignment of patient 7q deletion segments has mapped three commonly deleted regions (CDRs) at the cytogenetic bands 7q22, 7q34, and 7q35-36.[80,119,120] Several candidate TSGs have since been identified in these regions on the basis of mutational status in myeloid or solid cancers, as well as germline mutations that increase susceptibility to myeloid cancers, and have been validated in animal knockout models.[121] Mutations in the 7q genes *CUX1*, *EZH2*, and *LUC7L2* have also been identified in clonal hematopoiesis

and are predicted to increase cell fitness.[100,122–124] A summary of function, cellular and hematopoietic deletion phenotypes, and clinical associations for major 7q genes is provided in **Table 1**, and individual 7q TSGs are discussed in greater detail below:

CUX1

CUX1, previously known as *CUTL1* and CCAT displacement protein (*CDP*), is a ubiquitously expressed, non-clustered homeobox transcription factor that is both evolutionarily and functionally conserved from *Drosophila melanogaster* to humans.[125] Located in the 7q22 CDR, *CUX1* is one of the few chromosome 7 genes that is recurrently mutated in cancer, with mutations identified in 2-4% of myeloid diseases and 1-5% of various solid tumors.[126,127] *CUX1* mutational patterns fit a signature representative of TSGs, with frameshift and nonsense alterations distributed throughout the coding frame[128] **(Figure 1)**. Bi-allelic mutations are rare, however, suggesting haploinsufficiency. MDS and AML patients with inactivating *CUX1* mutations have decreased survival compared to those with wild-type *CUX1*, with overall survival mirroring that of patients with -7/del(7q).[126,127] Collectively, the clinical data strongly implicate *CUX1* inactivation in myeloid disease development and support *CUX1* being a critical 7q TSG.

The cellular function of *CUX1* and role of *CUX1* in myeloid malignancies is under active exploration. Investigation of *CUX1* is complicated by the complexity of the locus: the *CUX1* gene is large, spanning 340 kilobases and 33 exons, with multiple RNA and protein isoforms.[129] Hematopoietic cells express only the full length p200 isoform,[130] which contains four DNA-binding domains consisting of three CUT-repeat domains and one homeodomain[131] **(Figure 1)**. *CUX1* is further complicated by being one of the few

mammalian genes that shares exons with a second, independent gene *CASP* (Cux1 Alternative Splice Product).[132] Exons 15-24 are unique to *CUX1* and contain the four DNA-binding domains; consequently, CASP does not have DNA-binding domains nor is it located in the nucleus **(Figure 1)**. Instead, CASP is a highly expressed Golgi-associated protein thought to be involved in vesicle transport.[133] *CASP* and *CUX1* isoforms are routinely aggregated in genomics datasets such as RNA sequencing, making it challenging to parse out independent roles of *CUX1* and *CASP*, and failure to assess antibody specificity to CUX1 and CASP can produce inconclusive or misleading results.[130]

Due in part to this complexity, and the requirement for *Cux1* during development, establishment of traditional *Cux1* knockout mice has been challenging.[134] To circumvent these issues, the McNerney lab developed inducible shRNA-based murine models of *Cux1* knockdown, reducing CUX1 protein levels to 54% ($Cux1^{mid}$) and 12% ($Cux1^{low}$) in thymocytes.[135] The $Cux1^{mid}$ shRNA targets an exon shared by all *Cux1* and *Casp* transcripts, whereas the $Cux1^{low}$ model affects CUX1-encoding transcripts only **(Figure 1)**. Ubiquitous shRNA expression in $Cux1^{mid}$ mice leads to a normocytic anemia and splenomegaly, and $Cux1^{low}$ mice develop MDS/myeloproliferative neoplasms with a lethal anemia.[135] Recently, a second group reported a *Cux1* knockout model in which exons 15-18 were excised in the hematopoietic compartment driven by *Vav1-iCre*.[136] This approach, which avoids *Casp* isoforms, removes the first two DNA-binding domains and ablates CUX1 protein expression in an allele-dependent manner in splenocytes. Similar to $Cux1^{mid}$ mice, $Cux1^{+/-}$ mice develop mild anemia and bone marrow dysplasia.[136] This phenotype is exacerbated with full *Cux1* loss; $Cux1^{-/-}$ mice develop a disease akin to $Cux1^{low}$ mice, providing compelling evidence for the pathogenesis of *CUX1* loss in

myeloid disease. These models also suggest the effects of mutations in shared exons can likely be attributed to CUX1 disruption and not CASP.

On a molecular level, CUX1 preferentially binds enhancer elements and acts as a transcriptional activator or repressor in a context-dependent manner.[135,137,138] CUX1 interacts with the BAF nucleosome remodeling complex in HSCs to open chromatin, particularly at enhancer regions, and *CUX1* deficiency in $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) results in a proliferative gene signature similar to MDS patients with -7/del(7q).[135,139–141] Recently, the McNerney lab reported *CUX1* loss also impacts the epigenetic landscape of cells, both basally and in the context of irradiation-induced DNA damage.[142] After irradiation, *CUX1*-null cells show an impaired DNA damage response with decreased di- and tri-methylation of histone 3 lysine 9 and 27 (H3K9me2/3; H3K27me2/3), marks normally associated with DNA repair.[142–144] Further, *Cux1*-deficient cells continue to proliferate after alkylating agent exposure, ultimately leading to alkylator-induced t-MN in *Cux1*-deficient mice.[142] Collectively, the data demonstrate a novel role for CUX1 in the recognition and repair of chemotherapy-induced DNA damage. Given the epidemiologic connection between alkylating agent chemotherapy and -7/del(7q) t-MNs,[145] as well as the presence of both *CUX1* mutations and -7/del(7q) in CHIP, these findings provide a mechanistic link between del(7q) and t-MN. Overall, there is overwhelming clinical, genetic, and experimental evidence for *CUX1* as a critical 7q TSG.

EZH2

EZH2 is a well-known histone-modifying enzyme encoded in the 7q36 CDR that catalyzes the mono-, di-, and tri-methylation of histone 3 lysine 27. H3K27me3 is the best-

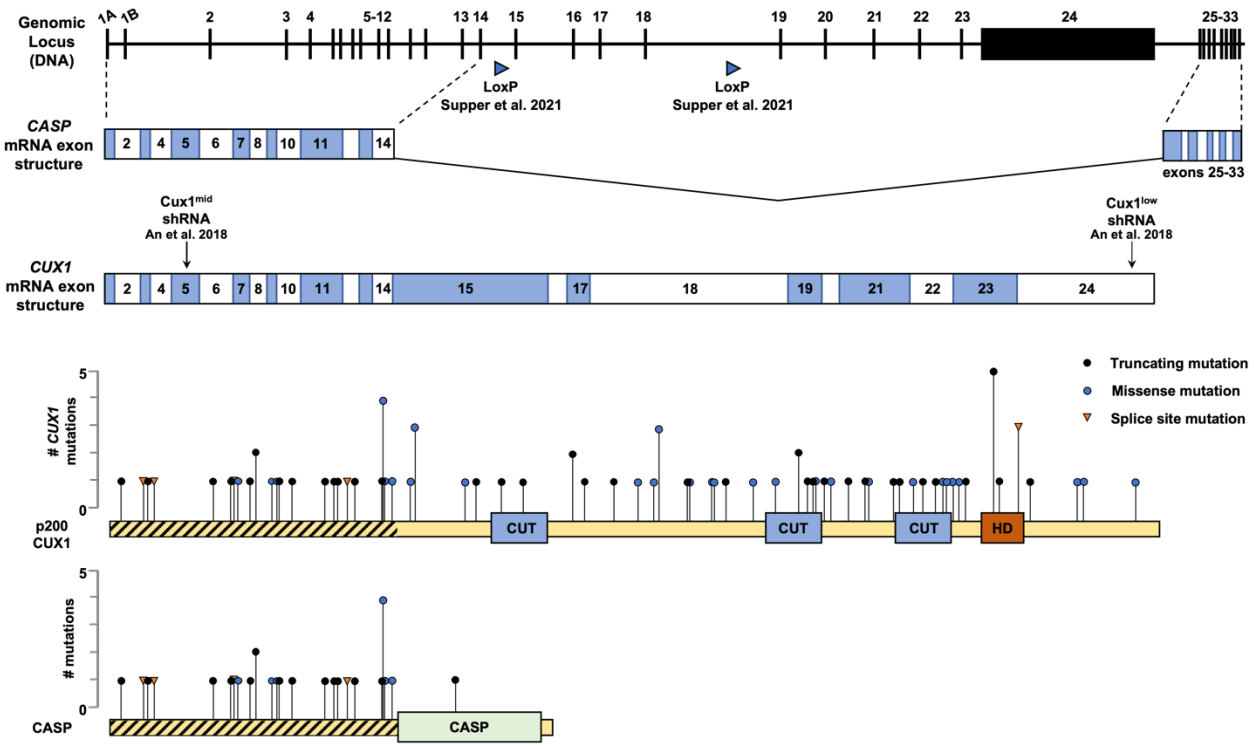

Figure 1: Structures of *CASP* and *CUX1*.

The genomic locus of *CUX1* has two alternative start sites (exons 1A and 1B) and contains 33 exons which encode two gene products, *CUX1 and CASP*. The locus organization is conserved between humans and mice. *CUX1* contains 24 exons; *CASP* is spliced from exons 1-14 and 25-33. The *CUX1* NM_181552 mRNA exon structure is shown with Cux1^{mid} and Cux1^{low} shRNA targeting locations,[135] and LoxP recombination sites are shown below the genomic locus.[136] The p200 CUX1 protein is depicted below the exon structure with the four DNA-binding domains; exon length is drawn to scale to match the protein. Overlaid is a plot of *CUX1* mutations from AACR project GENIE disease classes "Leukemia", "Myelodysplastic Syndromes", "Myeloproliferative Neoplasms", and "Myelodysplastic/Myeloproliferative Neoplasms".[128] The distribution of mutations fits a pattern representative of tumor suppressor genes.[57] A plot of CASP is shown below CUX1; there is only a single mutation within the *CASP* exons not shared by *CUX1*. Regions shared by *CUX1* and *CASP* are hatched.

Source: Jotte and McNerney 2022[146]

studied of these three marks, and is associated with negative regulation of gene expression and is an abundant component of transcriptionally silent facultative heterochromatin.[147,148] Originally identified in *Drosophila melanogaster*, EZH2 forms the core of the multi-subunit methyltransferase Polycomb Repressive Complex 2 (PRC2) along with EED, SUZ12, and RBBP4 or RBPP7.[148] The paralog of EZH2, EZH1, can also complex with other polycomb subunits to form PRC2, and there are many accessory proteins that can associate with PRC2 to form complex variants.[148,149] The polycomb proteins are well-conserved in animals and critical for transcriptional repression of developmental genes; accordingly, the core subunits of PRC2 are essential for murine embryonic development and homozygous loss of either *Ezh2*$^{-/-}$, *Eed*$^{-/-}$, or *Suz12*$^{-/-}$ is embryonic lethal around gastrulation.[150–152]

The role of EZH2 in HSCs is somewhat challenging to dissect, due in part to EZH1 compensation that maintains a degree of H3K27me3 in *Ezh2*-null cells.[153] Conflicting reports of EZH2 function and deletion phenotypes add further complexity: *Ezh2* levels decrease in aged murine HSCs, and overexpression of *Ezh2* can enhance self-renewal and repopulating potential,[154,155] but loss of PRC2 function via *Suz12* inactivation also increases HSC self-renewal and repopulation in serial transplantation.[156] *Ezh2* has been reported by multiple groups to be dispensable for adult hematopoiesis in mice,[157,158] but *EZH2* is among the most commonly mutated genes in human hematologic malignancies.[121] Even within a disease context, *Ezh2* loss has been reported to both enhance and disrupt MLL-AF9 leukemia at different stages of development.[159,160]

The spectrum of *EZH2* mutations varies in hematologic disease. In B-cell lymphomas, oncogenic point mutations in the catalytic domain result in increased

H3K27me3 that represses cell cycle checkpoint genes and impairs B-cell differentiation.[161–163] This has led to the successful development and implementation of EZH2 inhibitors for use in follicular lymphomas.[164,165] In contrast, myeloid disease is more so characterized by inactivating mutations in *EZH2*, associated with poor prognosis, though overexpression has been reported as well.[118,166,167] *EZH2* is among the only 7q genes with recurrent bi-allelic inactivating mutations, suggesting a canonical tumor suppressor role for *EZH2* in these diseases.[114,168] In cases with chromosome 7 alterations, bi-allelic mutations are most commonly observed with uniparental disomy; 'second hit' mutations are rare in -7/del(7q).[80,113,114] Animal models of *Ezh2* loss support a tumor suppressive role, as mice with hematopoietic-specific *Ezh2* knockout develop myelodysplasia and late myelodysplastic disorders.[169] Some mice develop T-cell disease, likely reflecting the prevalence of inactivating *Ezh2* mutations in a subset of T-cell leukemias.[170] Further, *Ezh2* loss accelerates disease development when co-deleted with other genetic drivers, including *Tet2* and *Runx1* in models of MDS.[171–173] Intriguingly, decreased EZH2 and H3K27me3 levels are observed in approximately half of relapsed AML, and pharmacologic and genetic reduction in EZH2 activity induces resistance to standard AML therapeutics, implicating *EZH2* in drug resistance.[174]

Mechanistically, the pathogenesis of *EZH2* loss is thought to be driven by de-repression of PRC2 target genes, particularly the *HOX* clusters of transcription factors. The *HOX* genes are well-known leukemic oncogenes that are normally repressed during hematopoietic differentiation, and inappropriate *HOXA9* and *HOXA10* expression is observed in -7/del(7q) AML and MDS.[175–177] Primary AML samples with low EZH2 levels display increased *HOXA9* and *HOXB7* expression, and *EZH2* knockdown reduces

H3K27me3 in the HOX clusters with subsequent increases in expression.[174,178] Of note, the *HOXA* cluster is located on 7p and thus partially lost by a monosomy 7; whether HOX de-repression is equivalent in patients with -7 vs del(7q) has not been characterized. EZH2 has also been reported to contribute to DNA repair, and is recruited to sites of DNA damage.[179–181] Loss of EZH2 decreases the efficiency of break repair, which may be related to the increased drug resistance phenotypes in AML with reduced EZH2 levels.[181,182] Collectively, there is strong evidence for *EZH2* as a mediator of del(7q) pathogenesis.

KMT2C and KMT2E

The mixed lineage leukemia (MLL) gene family, also known as the lysine N-methyltransferase 2 (KMT) family, comprises the core enzymatic subunits of the COMPASS H3K4 methyltransferase complexes and is frequently disrupted in leukemia.[183,184] Translocations of *KMT2A/MLL1* are numerous in both myeloid and lymphoid acute leukemia, and are well-known oncogenic drivers.[185] Genomic sequencing studies have identified recurrent mutations in *KMT2C/MLL3* and *KMT2D/MLL4* across a variety of hematologic and solid malignancies, with *KMT2C* being among the most frequently mutated genes in cancer.[121,183] *KMT2C* is encoded in the 7q36 CDR, and together with KMT2D forms the core and scaffold for the MLL3/4 COMPASS-like complex that catalyzes H3K4me1, a histone modification commonly found in enhancer regions.[186,187] Similar to the Polycomb group proteins, the KMT genes are active during development and deficiencies are lethal during embryogenesis or shortly after birth with exception of *KMT2E/MLL5*.[183,188]

Though KMT2C and KMT2D have overlapping functions, distinct roles have been demonstrated for *KMT2C* deficiency in myeloid disease. Mice expressing catalytically inactive *Kmt2c*, compatible with viability, displayed aberrant myelopoiesis with decreased lymphoid output, increased myeloid output, and splenomegaly, but fail to develop frank malignancy.[189] However, single copy loss of *Kmt2c* is sufficient to accelerate *Trp53*$^{-/-}$-driven leukemogenesis in a transplant setting, supporting a role for *Kmt2c* as a haploinsufficient myeloid TSG.[190] Further, *Km2tc*$^{+/-}$ HSCs exhibit increased self-renewal over serial transplant and have a selective advantage after exposure to the alkylating agent cyclophosphamide.[191] As alkylating agent exposure is linked to t-MN development, this finding implicates *KMT2C* in -7/del(7q) t-MNs.[145] *KMT2C* mutations are also detected in relapse samples of pediatric AML, providing additional evidence for *KMT2C* in promoting chemotherapy resistance.[192] Similar to CUX1 and EZH2, KMT2C has been reported to mediate the DNA damage response and is recruited to sites of damage; cells deficient in *KMT2C* have reduced expression of repair genes, which may provide a mechanistic underpinning for increased cell survival in the face of chemotherapy.[193,194]

KMT2E/MLL5, located in the 7q22 CDR, is classified as a KMT family member based on sequence homology.[195] However, crystallography analysis indicates a lack of key residues in the catalytic domain and the protein does not show any histone methyltransferase activity *in vitro*.[196] In contrast to other KMT family members, homozygous loss of *Kmt2e/Mll5* is not embryonic lethal and knockout animals survive to adulthood with no gross anatomical abnormalities, though animals are smaller and spermatogenesis defects renders males infertile.[197–199] *Kmt2e* loss has relatively mild hematopoietic effects, with slightly decreased erythroid output and increased white blood

cells; knockout mice are also susceptible to bacterial infection, suggesting a decrease in neutrophil function.[197–199] Intriguingly, *Kmt2e*-null cells have severely reduced engraftment and repopulating potential in transplant experiments, indicating intrinsic HSPC defects. These phenotypes were later linked to an impaired DNA damage response, as *Kmt2e*-deficient HSPCs accumulate DNA damage and reactive oxygen species which dampens their functionality.[200]

Collectively, the 7q genes *CUX1*, *EZH2*, *KMT2C*, and *KMT2E* converge on epigenetic modifications and the DNA damage response, ultimately dysregulating myelopoiesis. Compounding dysfunction of the pathways moderated by these genes may increase cell tolerance to stressful environments and enable persistence and expansion, as has been shown for many epigenetic regulators including *EZH2*.[201] Whether combinatorial loss of 7q genes acts in an additive or epistatic fashion remains an important, unanswered question.

Additional candidate 7q TSGs

-7/del(7q) likely deregulates additional cellular pathways involved in myeloid disease, including cell signaling, RNA splicing, and energy metabolism. Germline mutations in the 7q genes *SAMD9* and its paralog *SAMD9L*, encoded in 7q21, are strongly associated with monosomy 7 as well as the multi-system disorder MIRAGE syndrome.[78,202] These mutations are thought to be gain-of-function as enforced overexpression in $CD34^+$ HSPCs interferes with ribosome assembly, suppressing protein synthesis and cell growth.[203] Intriguingly, patients that develop monosomy 7 often lose the chromosome harboring the mutant allele, known as adaptation by aneuploidy, arguing

against a tumor suppressive role for *SAMD9/SAMD9L*.[202,204] However, mice deficient in *Samd9l* develop lethal myeloid dysplasia and *Samd9l*$^{+/-}$ HSPCs display hypersensitivity to cytokines, suggesting the gene may have some TSG function.[205] Mutations in the splicing factor *LUC7L2*, encoded in the 7q34 CDR, have been identified in CHIP and AML with antecedent MDS.[206] Splicing factor mutations occur in 50% of over MDS cases but are challenging to characterize due to poor overlap of alternative splicing events, thought to be the mechanism underlying these mutations.[207,208] Two recent studies independently reported an unexpected downregulation of glycolysis genes following *LUC7L2* loss, with subsequent shifting of metabolism toward oxidative phosphorylation.[209,210] Both studies identify exon skipping as a mechanism of decreased gene expression and link alternative splicing events to glucose metabolism, a novel mechanism not previously ascribed to splicing factors. Erythropoiesis is also affected by 7q gene deficiencies in addition to *CUX1*, as decreased levels of *DOCK4* or *ACHE* in CD34^{+} HSPCs reduce erythroid differentiation.[211,212]

7p genes

Though the majority of implicated chromosome 7 TSGs are located on 7q, some genes on 7p promote pathogenic phenotypes when lost. *PMS2*, encoding a mismatch repair endonuclease, is commonly mutated in Lynch syndrome, a hereditary cancer predisposition syndrome.[213] *Pms2*$^{-/-}$ murine fibroblasts are deficient in mismatch repair and also display low-level resistance to platinum agents, chemotherapeutics that are associated with t-MN development.[145,213] The transcription factor FOXK1 has also been reported to influence the DNA damage response, and *FOXK1* knockout cells display

diminished DNA repair capabilities after irradiation.[214] The aryl hydrocarbon receptor (AHR), a nuclear ligand-activated transcription factor, is implicated in the development of many cancers and is attributed both oncogenic and tumor suppressive functions.[215] AHR antagonists increase HSC expansion *ex vivo*, suggesting *AHR*-deficient HSCs may similarly have a proliferative advantage.[216] In line with this, AHR signaling is repressed in AML blasts, and administration of an AHR agonist significantly impairs leukemic growth and self-renewal.[217] The *ACTB* gene, encoding one of the six different actin proteins, is also located on 7p. A major component of the cytoskeleton, actin filaments have also been found in the nucleus and promote DNA repair, and inhibition of actin nucleation increases the number of double-strand breaks and impairs DNA end-processing.[218,219] Whether 7p genes contribute to -7/del(7q) pathogenesis remains largely unexplored.

7q as a contiguous gene syndrome region

A contiguous gene syndrome (CGS) is a genetic disorder caused by large-scale chromosomal alterations affecting copy number, leading to dosage imbalance of multiple neighboring genes.[220] Given the numerous bona fide TSGs and myeloid regulators on 7q, it is likely the 7q CDRs function as CGS regions in cancer, similar to those observed on 5q and 8p.[48,55] Several 7q TSGs have functions that converge on similar pathways, particularly on the DNA damage response. De-regulation of DNA repair is a common thread across human cancer, with numerous cancer predisposition syndromes affecting the fidelity of DNA repair including Fanconi anemia and Lynch and Bloom syndromes.[221] Chromosome 7 deletions may therefore enable cells to survive genotoxic stress via altered DNA damage recognition and repair, particularly in the development of t-MNs.

Though it seems counterintuitive that a decreased capacity for DNA repair confers a competitive advantage, the resultant genomic instability can serve as a mechanism to generate the genetic diversity that expedites the acquisition of cancer hallmarks.[59] Intriguingly, the overall mutation burden of t-MN is similar to that of *de novo* AML, though signatures from chemotherapy exposure can be detected, suggesting therapy-induced mutations are not the primary driver of t-MNs.[222,223] Loss of multiple critical 7q TSGs that collectively impinge on the DNA damage response may thus be a major factor in the etiology of t-MNs. While mechanistic studies of 7q genes have traditionally focused on individual genes, studies investigating combined gene deletions are warranted to refine the understanding of how -7/del(7q) drives malignancy.

Modeling aneuploidy

The biggest challenge to furthering understanding of aneuploidies is developing appropriate model systems. Most laboratory techniques to study aneuploidy are based on random chromosomal mis-segregation, achieved through pharmacologic or genetic disruption to mitotic spindle assembly components.[1] Cell lines of interest must be karyotyped to identify the chromosomal aberration, but this technique can be of use for studying general properties of aneuploidy. Defined aneuploid lines harboring a specific alteration can be generated through microcell-mediated transfer of single chromosomes to a host cell, but this technique is restricted to modeling trisomies and the resultant cells are often karyotypically unstable.[224] Inclusion of a selectable marker can allow stable propagation, but only a limited number of aneuploidies have been generated through this

approach.[1] Further, generating lines with multiple abnormalities is challenging and precludes studies of complex karyotypes.

Studying specific human aneuploidies in animal models is hindered by a lack of chromosomal synteny between species. The genes encoded on human chromosome 7 are located largely on chromosomes 5 and 6 in the mouse;[225] consequently, generating models of whole-chromosome aneuploidy corresponding to human sequences is infeasible. Alternative approaches include Cre-Lox recombination to remove specific chromosomal segments, which has been used to engineer two large-scale deletions of regions syntenic to the human 7q22 CDR in the mouse.[226–228] While this approach can inform whether specific regions affect hematopoiesis, variations in patient 7q deletion location and lengths make determining boundaries challenging and the rate of Cre-recombination is quite low.[226] Further, haploinsufficiency in the mouse may not fully recapitulate phenotypes of human patients – mice with heterozygous loss of a 1.5 Mb region corresponding to human 7q22, that includes *Cux1* and 27 other genes, display largely normal hematologic parameters at >1 year of age.[228] As mice with ~90% reduction in CUX1 protein display a more severe phenotype than mice with ~50% reduction, decreases past 50% may be necessary for some genes.[135] Similar observations have been made in mice deficient in the hematopoietic transcription factor *Sfpi1* (PU.1).[229]

Human inducible pluripotent stem cells (iPSCs) have also been used to study 7q deletions, either derived directly from patient samples or engineered to carry large deletions.[230,231] The use of isogenic controls from the same patient enables direct testing of the deletion region to recapitulate disease-associated phenotypes; however, iPSCs are difficult to culture and can undergo spontaneous dosage correction that restores the

missing chromosome 7 segment to the diploid state.[230] Further, engineered deletions are still subject to boundary determinations. While iPSCs can have utility for mapping disease-associated regions, further experiments are necessary to dissect the contributions of specific genes.

Combinatorial deletion of multiple genes is an appealing strategy for modeling monosomies, as has been used to interrogate 5q and 8p deletions.[48,55] These approaches circumvent synteny issues, though it becomes prohibitively complex to target more than two genes particularly in inducible or tissue-specific shRNA or Cre-based systems. Recently, CRISPR-Cas9 has been used to simultaneously target multiple loci on different chromosomes to model CHIP.[232,233] Up to six genes can be modified in a single cell with this technique, more than have been affected with short hairpin approaches.[48,232] Multiplex CRISPR-Cas9-based gene deletion may therefore be a novel means to model del(7q) that overcomes the challenges of other approaches.

Critical knowledge gaps to be addressed

Much remains to be understood about the contributions of aneuploidy to cancer development. Clinical evidence is mounting that chromosome 7 deletions and *CUX1* mutations can be early, driving events in myeloid disease. Though *CUX1* is clearly a critical 7q TSG, additional 7q genes have been identified that are mutated in cancer and yield hematopoietic phenotypes when deleted in mice.[121] As no single gene deletion fully recapitulates the MDS/AML phenotype, it is essential to innovate new models to interrogate combined gene deficiency. It remains to be discovered whether 7q gene

deficiencies are additive, epistatic, or even redundant, particularly for genes that act in the same molecular pathways.

As both chromosome 7 deletions and individual 7q gene mutations have been identified in CHIP, uncovering the mechanisms by which loss of chromosome 7 genes increase fitness in response to different environmental pressures is a critical outstanding question. Emerging data indicate that certain pressures, such as genotoxic therapy, can select for CUX1-deficient clones, and this fitness advantage likely corresponds with the inherent drug resistance of malignancies arising from these clones.[142] This premise is further supported by the prevalence of -7/del(7q) in t-MN and association with poor prognosis in many diseases, as well as the identification of 7q gene mutations in relapsed patients.[174,192] The advent of high-throughput genomics and plethora of genome-wide screens may also serve as powerful tools to identify candidate genes driving drug resistance, in tandem with patient mutational data. Collectively, this work aims to assess the feasibility of a multiplex CRISPR-Cas9-based approach to model aneuploidy in piecewise fashion, in the context of -7/del(7q). Establishing a model of del(7q) clonal hematopoiesis and drug resistance will shed light on how these deletions contribute to the pathogenesis of high-risk myeloid disease, and test the idea of 7q as a CGS region. Further, our CRISPR-based methodology may serve as a tractable framework for interrogating other recurrent aneuploid events more broadly in cancer.

Table 1: 7q genes Implicated in Myeloid Disease Based on Clinical and Experimental Data

Gene Name (location)	Cellular Function	Cellular Deletion Phenotype	Hematopoietic Murine Deletion Phenotype	Clinical Associations
SAMD9/SAMD9L **(7q21.1)**	Endosomal fusion protein; terminates surface receptor signaling[205]	Persistent cytokine receptor signaling and hypersensitivity[205]	Enhanced HSC colony-forming potential and *in vivo* reconstitution[205]	Germline activating mutations cause MIRAGE syndrome[202]
	Regulates protein synthesis[203]	Gain of function mutations disrupt ribosome assembly[203]	Late MDS in germline +/- and -/- mice[205]	Mutant allele lost through -7 via adaptation by aneuploidy[78,202,204]
ACHE **(7q22.1) CDR**	Hydrolyzes acetylcholine; linked to stress hematopoiesis[234]	Enhanced proliferation, decreased apoptosis in mouse bone marrow cultures[235]	Increased neutrophil cell number in +/- mice[236]	None reported
		Impaired erythroid differentiation in human $CD34^+$ cells[212]		
CUX1 **(7q22.1) CDR**	Homeobox transcription factor[137]	Enhanced proliferation, activation of AKT signaling[135]	$Cux1^{mid}$: monocytosis $Cux1^{low}$: monocytosis and lethal anemia[135]	Significantly decreased expression in -7/del(7q) leukemias[237]
	Recruited to DNA DSB, interacts with EHMT2 methyl-transferase[142]	Decreased apoptosis in hematopoietic progenitors[136]	Hematopoietic-specific +/- mice: increased mean RBC volume -/- mice: monocytosis and anemia[136]	Inactivating mutations found in MDS, AML, and MDS/MPN-overlap syndromes[126,127,238]
	Interacts with BAF complex in HSCs to open chromatin at enhancers[139]	Impaired DNA damage response[142]	$Cux1^{low}$ and -/- mice develop MDS/MPN-like disease[135,136]	Inactivating mutations associated with poor prognosis in MDS[126,127]
			$Cux1^{low}$ mice treated with ENU develop rapid, fatal t-MN[142]	

Table 1 continued

RASA4 **(7q22.1) CDR**	RAS GTPase-activating protein[239]	Elevated Erk phosphorylation after FcyR stimulation[239]	No overt disease; germline -/- mice have impaired macrophage phagocytosis[239]	Promoter hypermethylation in JMML[240]
KMT2E **(7q22.3) CDR**	Epigenetic regulator, capable of binding H3K4 methylation[196]	Cell cycle arrest in lung fibroblasts and colon cancer cells[241]	No overt disease; impaired neutrophil maturation, decreased RBC and hematocrit in germline -/- mice[197–199]	High expression linked to favorable outcome in cytogenetically normal AML[242]
	Reported catalytically inactive[196]	Increased ROS, impaired DNA damage response in +/- and -/- hematopoietic progenitors[200]	Mild thymocyte and splenocyte reduction, decreased RBC and hematocrit, increased RBC distribution width in germline +/- mice[197,236]	
DOCK4 **(7q31.1)**	GTPase activator[211,243]	RBC cytoskeletal defects, erythroid dysplasia in human CD34+ cells[211,243]	No reported hematopoietic phenotype in germline -/- mice[236,244]	Significantly decreased expression in MDS, linked to decreased overall survival[211,243]
MKLN1 **(7q32.3)**	Organization of F-actin networks[245,246]	Decreased retrograde intracellular transport in neurons[246]	No reported hematopoietic phenotype in germline +/- or -/- mice[246]	Linked to inherited predisposition for MPN[247]
				Mutations observed in relapsed pediatric AML[192]
TRIM24 **(7q33)**	RING-domain E3 ubiquitin ligase; nuclear receptor co-regulator[248,249]	Increased proliferation in human CD34+ cells, decreased proliferation in AML cell lines[212,250]	Germline -/- mice develop HCC but have no hematopoietic phenotype[251]	High expression in AML reported, linked to poor survival[250]
	Targets p53 for degradation[252]	Reduced DSB repair complex recruitment[253]	***conflicting evidence for TRIM24 as hematopoietic oncogene or TSG***	

Table 1 continued

	Serine and threonine nuclear kinase[254–256]	Decreased p53 activation and apoptosis in breast cancer cells[254]	No reported hematopoietic phenotype in germline -/- mice[236,257]	Low frequency missense mutations in MDS and AML[256]
HIPK2 **(7q34) CDR**	Phosphorylates p53 to activate apoptosis[254]	Decreased erythroid differentiation and expansion in human $CD34^+$ cells[230]		
		Increased cisplatin resistance in colon cancer and lung cancer cell lines[255]		
LUC7L2 **(7q34) CDR**	Splicing factor, co-localizes with U1 snRNP[113,209,210]	Altered splicing in AML and cervical cancer cell lines[209,210]	No overt disease; Increased platelet volume in germline +/- mice[236]	Heterozygous inactivating mutations observed in MDS and AML[113,206]
		Decreased expression of glycolysis genes; metabolic shift to OX-PHOS in AML and cervical cancer cell lines[209,210]	No reported hematopoietic phenotype in other germline -/- mice[258]	Decreased expression linked to reduced survival in MDS[113,206]
ATP6V0E2 **(7q36.1) CDR**	Intracellular proton pump[230]	Decreased erythroid differentiation and expansion in human $CD34^+$ cells[230]	None reported	None reported
CUL1 **(7q36.1) CDR**	E3 ubiquitin ligase; transcriptional repressor[259–261]	Increased transcription of c-MYC target genes[261]	Germline -/- is embryonic lethal[259]	Mutations observed in myeloid neoplasms[113]
			Deletion in T-cell lineage yields T-cell lymphomas[260]	

Table 1 continued

EZH2 **(7q36.1) CDR**	Catalytic component of PRC2, methylates H3K27[147]	Decreased H3K27 methylation; partial compensation by EZH1[153,169]	Conditional knockout mice observed up to 30 weeks do not develop myeloid disease[262]	Mono-allelic and bi-allelic inactivating mutations observed in myeloid disease[80,114,168]
	Gene silencing, developmental gene repression[148]		Transplant recipients of *Ezh2*$^{-/-}$ cells develop mixed disease, including T-cell lymphoma and late MDS[169,172]	Inactivating mutations linked to poor prognosis in MDS, drug resistance in AML[118,167,174]
KMT2C **(7q36.1) CDR**	H3K4 methyltransferase, core component of COMPASS complex[184]	Decreased H3K4me1 at some enhancer regions; compensation by KMT2D[187,191]	No overt disease; increased HSC number and self-renewal, splenomegaly in hematopoietic-specific and germline +/- and -/- mice[191,236]	Mutations observed in AML, possibly over-represented due to pseudogene[189,190,263]
	Recruited to DNA DSBs[193]	Decreased expression of DNA damage response genes in bladder cancer cells[194]	Mono-allelic knockout accelerates sh*Nf1*, *Trp53*$^{-/-}$-driven leukemogenesis[190]	Mutations observed in relapsed pediatric AML[192]
			Selective advantage in +/- and -/- HSCs after chemotherapy[191]	

Abbreviations: +/- heterozygous deletion; -/- homozygous deletion; AML, acute myeloid leukemia; DSB, double-strand break; HCC, hepatocellular carcinoma; HSC, hematopoietic stem cell; JMML, juvenile myelomonocytic leukemia; MDS, myelodysplastic syndrome; MPN, myeloproliferative neoplasm; OX-PHOS, oxidative phosphorylation; PRC2, Polycomb Repressive Complex 2; RBC, red blood cell; t-MN, therapy-related myeloid neoplasm.

MATERIALS AND METHODS

In silico data mining and patient deletion alignment

Nine genome-wide CRISPR knockout screens for resistance against cisplatin, doxorubicin, or etoposide were downloaded from the Biological General Repository for Interaction Datasets open repository of CRISPR screens (BioGRID ORCS).[264–273] For each screen, we identified genes with positive z-scores above background, indicating increased drug resistance after gene deletion, then narrowed this list to only chromosome 7 genes expressed in human hematopoietic stem cells.[135] We further subsetted to only genes located in 7q CDRs, then entered the final list of 39 candidates into the Molecular Signatures Database (mSigDB) to identify relevant functional pathways.[274] Patient SNP array data from the TCGA PanCancer Atlas[275] was accessed from cBioPortal (cbioportal.org/datasets);[276] data from the University of Chicago was accessed through GEO, accession number GSE42482. Data were visualized with Integrated Genome Viewer version 2.[277]

Mouse models

All animal studies were approved by the University of Chicago Institutional Animal Care and Use Committee. Mice were housed in specific pathogen-free animal care facilities at the University of Chicago, accredited by the Association for Assessment and Accreditation of Laboratory Animal Care. Mice were housed separately by sex with up to five mice per cage. Mice of both sexes were used in all experiments. Mice harboring short hairpin RNA (shRNA) targeting either *Cux1* (shCux1) or renilla luciferase as a control (shCtrl) mice were generated as previously described;[135] briefly, a second-generation

reverse tet-transactivator (M2rtTA) expressed from the endogenous *Rosa26* promoter activates expression of shCux1 or shCtrl, downstream of the *Col1a1* locus, as well as GFP. Mice were genotyped by tail clip, and mice used were either rtTA-M2$^{tg/tg}$;Col1a1$^{shCux1/WT}$ or rtTA-M2$^{tg/tg}$;Col1a1$^{shRen/WT}$ littermate controls, on a mixed C57BL/6 x 129/Sv background. The transgene was induced by keeping mice on a continuous doxycycline-containing diet (TD.12006, 1 mg/kg, Envigo). For all experiments, age-matched, sex-matched, littermate adult mice aged 6-12 weeks were used.

Cell transfection and culture

CRISPR guide RNAs were designed by intersecting the top hits from the Synthego CRISPR design tool and the sgRNA Designer by the Broad Institute. We selected guides that targeted early exons or functional domains to maximize chances of generating a nonfunctional protein. We transfected cells with Cas9:gRNA ribonucleoprotein (RNP) complexes using the NEON transfection system (ThermoFisher). Cas9:gRNA complexes were assembled by mixing guide RNA with Cas9-2NLS protein (Synthego) in a 3:1 molar ratio and incubating at room temperature for 15-30 minutes.[278,279] For multiplex transfections, we used equal amounts of guide RNA for each targeted gene while maintaining the 3:1 molar ratio with Cas9. We electroporated $2x10^5$ - $1x10^6$ cells with the RNP mixture using the conditions: 1650 volts, 35 ms, 1 pulse for human U937 cells; 1900 volts, 20 ms, 1 pulse for mouse c-kit$^+$ HSPCs. Cells were cultured in media for 2-5 days post-transfection before isolating DNA and performing Sanger sequencing of the gRNA cut site. PCR amplification was performed using Quick Taq^{TM} HS Dye master mix (Diagnocene). Sequencing was performed by the University of Chicago DNA Sequencing

Facility, and sequencing traces were aligned using the Synthego Inference of CRISPR Edits (ICE) tool to determine CRISPR editing and gene knockout efficiencies. A table of guide RNAs and primers is provided in **Table 2**.

U937 cells were grown in RPMI 1640 media supplemented with 10% fetal bovine serum (R&D systems) and 1% antibiotic/antimycotic (Gibco). Single cell clones were generated from the transfection pool and clonality was confirmed with Sanger sequencing and western blotting. Mouse HSPCs were grown on fibronectin-coated plates (StemCell Technologies) with either SFEM base media (StemCell Technologies) supplemented with 50 ng/mL murine SCF, FLT-3 ligand, and TPO[135] (PeproTech), or Ham's F12 nutrient mix (Gibco) supplemented with 10 mM HEPES, 1x Insulin-Transferrin-Selenium-Ethanolamine, 1 mg/mL polyvinyl alcohol, 1x penicillin-streptomycin-glutamine, 100 ng/mL TPO, and 10 ng/mL SCF.[280] In both cases media was supplemented with 1.5 μg/mL doxycycline to induce shRNA against *Cux1* or renilla luciferase.

Western blot for protein knockout

Protein lysate was generated by isolating up to $1x10^6$ cells, washing once with 1X PBS, and lysing in RIPA buffer with EDTA and protease/phosphatase inhibitor for one hour at 4˚C. Lysates were spun at 15,000g for 20 minutes, and the supernatant was resuspended in loading buffer with β-mercaptoethanol and boiled for 5 minutes. Equal volumes of lysate were loaded onto a 4-15% gradient polyacrylamide gel (BioRAD), run at 20 mA until separated, and transferred to a PVDF membrane for 90 minutes at 150 mA. All primary antibodies were used at 1:1000 dilution, except antibodies probing histones which were used at 1:10,000 dilution. Secondary antibodies were used at a

1:2000 dilution. A table of antibodies used is provided in **Table 3**. A second CUX1-specific antibody was used and generated as follows: Rabbit polyclonal antibodies recognizing the C-terminal amino acids 1223-1242 of CUX1 (CEPPSVGTEYSQGASPQPQH) were generated and affinity purified by the Pocono Rabbit Farm and Laboratory (Canadensis, PA).

Bone marrow transplant

HSPCs were isolated from littermate shCtrl and shCux1 donors using a MACS column after incubation with CD117 microbeads (Miltenyi Biotec). After resting in an incubator for 2-3 hours, $2x10^5$ - $1x10^6$ c-kit$^+$ HSPCs were electroporated with RNPs targeting genes of interest and cultured as described above. Two days post-electroporation, cells were collected, washed once with 1X PBS, and mixed at a 1:2 ratio with RBC-depleted helper marrow (C57BL/6, CD45.1). Recipient mice (C57BL/6, CD45.1) aged 6-10 weeks were lethally irradiated (γ-irradiation, 8.5 Gy), anesthetized with isoflurane four hours later (4% vaporized in O2), and transplanted retro-orbitally. Mice were monitored post-transplant until anesthetic wore off. Recipient mice were immediately shifted to doxycycline-containing chow to maintain shRNA induction.

Tracking gene knockout in peripheral blood and tissues

After a three-week initial engraftment period, serial blood samples were collected every four weeks from the submandibular vein and subjected to complete blood count, flow cytometry, and DNA sequencing analyses. Complete blood counts were obtained using a Hemavet 950 counter (CDC Technologies, Oxford, CT). For flow cytometry, 50 uL

of peripheral blood was incubated with a murine Fc receptor blocking agent (Invitrogen) for 10 minutes, then RBCs were lysed with 1X BD FACS Lysis buffer (BD Biosciences) for 20 minutes before washing and resuspension with a staining antibody cocktail for 30 minutes on ice (CD45.2, CD45.1, CD11b, CD3e, B220, Gr-1). For autopsy analyses, bone marrow and spleen samples were RBC-depleted with ACK lysis buffer (StemCell Technologies) for 5 minutes before Fc receptor blocking and staining with biotin-conjugated lineage markers for 20 minutes on ice. Cells were washed and resuspended with a staining antibody cocktail for 30 minutes on ice before sample acquisition (CD45.1, CD45.2, CD4, CD8, CD11b, CD3e, B220, Gr-1, CD117, Sca-1, CD34, CD150, CD48). A complete list of flow cytometry antibodies used is provided in **Table 4**. DNA was isolated from peripheral blood mononuclear cells and hematopoietic tissues at autopsy using the Wizard Genomic DNA Purification kit (Promega) according to the manufacturer instructions. PCR amplification of the gRNA cut site, Sanger sequencing, and sequencing trace alignment was performed as described above.

Inhibitors, drugs, and DNA damage treatment

Daunorubicin HCL was purchased from Selleck Chem, and etoposide was purchased from Sigma Aldrich. The EZH2 inhibitors tazemetostat (EPZ-6438) and GSK126 were purchased from Fisher Scientific. All compounds were dissolved in DMSO and added to cells in culture for the indicated times. All primary cell experiments utilized 0.01% DMSO. Cell line experiments utilized 1% DMSO. N-Ethyl-N-Nitrosourea (ENU) was purchased from Sigma Aldrich and dissolved in 10% ethanol and 90% PBS to 100 mg/mL. ENU was added to cells in culture (0.025 mg/mL) for the indicated times, or

administered intraperitoneally to mice at 100 mg/kg. For experiments with irradiation, DNA damage was induced by exposure to a ^{60}Co source (MDS Nordion). Dosage rates varied between 70.29 cGy/min and 75.33 cGy/min depending on the experiment date. Non-irradiated samples were mock irradiated, and cells were allowed to recover in an incubator for allotted time before use.

Dose-response curves and cell viability assays with daunorubicin and etoposide

For initial viability assays, $1x10^5$ U937 7q knockout or control HPRT cells were seeded per well in a 48-well plate and cultured with 3 nM daunorubicin, 50 nM etoposide, or DMSO vehicle control for 72 hours. For dose-response curves with daunorubicin and etoposide, $1x10^4$ U937 cells were seeded per well in a 96-well plate and cultured with increasing doses of either compound for 72 hours. Viability in both assays was assessed with the CellTiter-Glo 2.0 Cell Viability Assay (Promega). Equal volumes of cells and reagent were mixed in 96-well white-walled plate and imaged on a BioTek Synergy 2 analyzer. Assays were performed in at least triplicate with 2-3 technical replicates per experiment. To assess replicative capacity after culture with daunorubicin, equal numbers of viable cells were taken from vehicle and daunorubicin-treated wells, washed with warm media, and re-seeded at 1 cell per well to 96-well plates. The number of wells that formed colonies with ≥50 cells after 2 weeks was scored and normalized to vehicle plates. Apoptosis after culture with daunorubicin was measured by washing cells with Annexin V binding buffer, then staining with Annexin V-APC (Biolegend) at a 1:50 dilution for 30 minutes on ice. Cells were resuspended with 7-AAD viability solution (Biolegend) for 15 minutes before flow cytometry acquisition.

Table 2: Guide RNAs and Primers

Gene	Species	gRNA sequence	Forward primer (5' -> 3')	Reverse primer (5' -> 3')
EZH2	Human Exon 3	ACAGAAGUCAGGAUGUGCAC	AAAGATGGACACCCTGAGGTC	GGATGGCAATCGTTTCCTGT
Ezh2	Mouse Exon 7	UUAUCAGCAGGAAAUUUCCG	AACTGAGGTACACTGCAGAAGAG	AGTGCGTCATCCTGGTCATAC
KMT2C	Human Exon 3	GACACAGAUCGCUGAAGAGU	TAGTGTGCTGCCCTGCTCTA	GCGGTAGAGAATAAAGATGCACA
Kmt2c	Mouse Exon 3	AGAAAUCAAAGAACAAUCUG	GGAAGCATACTGCCATCCAGA	ACGTTGACCTCAAGGCACAG
KMT2E	Human Exon 6	GGCAACAUAUUGACUGCAUG	CCCAAACGAGGTTCGAGTCA	ACCATGCCCGGCTGTGATTC
*Kmt2e**	Mouse Exon 10	GUGAUACACACCUCUACAGG	GGGTTCAGCACCAGAAATTGA	TGCTTCAAACTGTTCTCTCAGC
Actb	Mouse Intron 5	AGGUUGCUCUGACAACCACA	CGCCTTCACCGTTCCAGTTT	GTGACAAAACTCCTGAGGCCATA
CUX1	Human Exon 4	UGCACUGAGUAAAAGAAGCA	CCCTCCTAGACCCTGAGCTT	TTCATGTGTCCTGCACTCCC
HRPT	Human Intron 2	GCAUUUCUCAGUCCUAAACA	AACATCAGCAGCTGTTCTGAGTA	TGCATAGCCAGTGCTTGAGAA
PPM1D	Human Exon 6	GAAGGCAUUGCUACGAACCA	TGCATAGATTTGTTGAGTTCTGG	TGGTTCTGGATCTTTTGAGGGT
Luc7l2	Mouse Exon 2	AACCACATACTGTTCCAGAA	CTGGGGTTCAGTGGGGATTC	TGACCCTGTGAGAACATGCC
Rasa4	Mouse Exon 6	TTAGCACCAAAAGATCGGAA	ATTCTTCCGCTCATCCACGG	TGCCACGCTCTTCAGATGAT
Samd9l	Mouse Exon 2	TGATGTGGTGACTAAAGACA	CCCTGTCTCAGCAACACGTA	ACAGTGATCATGGTCTCAGGG

* Denotes guide RNA used in multiplex-editing experiments

Table 3: Western Blotting and Immunofluorescence Antibodies

Protein	Catalog number	Supplier
CUX1	ABE217	Millipore
EZH2	3147S	Cell Signaling
KMT2C	53641	Cell Signaling
KMT2E	Sc-377182	Santa Cruz
ACTB-HRP	Sc-47778 HRP	Santa Cruz
HSC70-HRP	Sc-7298 HRP	Santa Cruz
H3K27me3	9733S	Cell Signaling
Total H3	Ab1791	Abcam
Anti-rabbit IgG-HRP	7074S	Cell Signaling
Anti-mouse IgG-HRP	7076S	Cell Signaling
H2AX pS139 (Rb)	9718S	Cell Signaling
H2AX pS139 (Ms)	05-636	Millipore
53BP1	PA5-17578	ThermoFisher
Anti-rabbit AF647	SAB4600177	Millipore
Anti-mouse AF488	SAB4600387	Millipore

Table 4: Flow Cytometry Antibodies

Target	Clone	Fluorophore(s)
CD45.1	A20	PE, BUV-395
CD45.2	104	APC, BUV737
B220	RA3-6B2	PerCP-Cy5.5, Biotin-conjugated
CD3e	145-2C11	BUV-737, Biotin-conjugated
CD11b	M1/70	APC-Cy7, Biotin-conjugated
Gr1	RB6-8C5	PE, Biotin-conjugated
CD19	1D3	Biotin-conjugated
Ter119	TER-119	Biotin-conjugated
CD5	53-7.3	Biotin-conjugated
Streptavidin		BUV-737, BUV-395
CD4	GK1.5	PE
CD8	53-6.7	APC
CD117	2B8	APC-Cy7
Sca-1	D7	PE-Cy7, APC-Cy7
CD34	RAM34	APC
CD48	HM48-1	PerCP-Cy5.5
CD150	Q38-480	PE

Colony forming assays

For primary cell colony-forming assays, $2x10^3$ cells were mixed with MethoCult M3434 containing drug or vehicle, and plated in a 6-well plate or 10-cm dish for passage 1. Each condition was plated in at least duplicates. M3434 was supplemented with 1.5 μg/mL doxycycline to induce shRNA expression, and with the EZH2 inhibitors tazemetostat or GSK126 where indicated. After seven days, total colonies were manually scored and wells were scraped, pooled, and total cell number counted. Up to $1x10^4$ cells were re-seeded for subsequent passages without additional drug or vehicle, and scored as in passage 1 after 10-12 days. Colony size was assayed by imaging culture wells in a Bio-Rad ChemiDoc XRS$^+$ imaging system or on a wide-field microscope with a 4x objective, fitting polygons to colonies in ImageJ, and measuring area. All colony areas were then normalized to the average area of shCtrl colonies to generate relative area distributions.

RNA-Sequencing

c-kit$^+$ bone marrow cells were isolated from adult shCtrl and shCux1 mice. For sequencing replicates with ENU, cells were pre-cultured on 24-well fibronectin-coated plates with doxycycline to induce shRNA and tazemetostat to inhibit EZH2 activity for four days, then media was replenished and supplemented with 0.025 mg/mL ENU for 72 hours to induce DNA damage. For sequencing replicates in the absence of ENU, cells were transfected with RNPs targeting *Ezh2*, *Kmt2c*, or an intron of *Actb* as a control and cultured on 24-well fibronectin-coated plates with doxycycline to induce shRNA for five days. In both experiments, Lineage$^-$, Sca-1$^+$, c-kit$^+$ (LSK) cells were sorted on a FACSAria

Fusion running FACSDiva (Special Order Research Product, BD Biosciences), equipped with 18 detectors and 355 nM, 405nM, 488 nM, 561 nM, and 640 nM lasers, directly into TRIzol. RNA was purified with the RNeasy Micro Kit (Qiagen) and libraries were prepared by the University of Chicago Genomics Facility. Three to four biological replicates were performed.

Barcoded libraries were sequenced with Illumina (NovaSEQ-6000) to generate 50-bp paired end reads. Technical replicate fastq files were concatenated and reads were aligned using STAR.[281] Differential gene expression analysis across all genotypes was performed using DESeq2,[282] with a negative binomial general linearized model fitting and Wald statistics used to generate p values, which were adjusted with the Benjamini-Hochberg method. Pairwise differential gene expression analyses between genotypes, for the purpose of Gene Set Enrichment Analysis (GSEA)[283], were performed using EdgeR,[284] with a negative binomial general linearized model fitting and an exact test, analogous to Fisher's exact test, used to generate p values, which were adjusted with the Benjamini-Hochberg method. The data discussed in this work are deposited in the NCBI Gene Expression Omnibus[285] and are accessible through with the GEO Series Accession number GSE266586.

γH2AX and neutral comet electrophoresis assays

c-kit$^+$ bone marrow cells were isolated from adult shCtrl and shCux1 mice, and cultured for five days with doxycycline and tazemetostat or vehicle to induce *Cux1* knockdown and EZH2 inhibition. Cells were irradiated (2 Gy) or mock irradiated, and collected at one- and four-hours post-IR. Lineage-negative cells were gated. Clonal U937

populations were irradiated (6 Gy) or mock irradiated, and collected at the time points indicated. Cells were fixed and permeabilized with the eBioscience FoxP3 Transcription Factor Staining Buffer set (ThermoFisher) and stained 1:50 with mouse anti-H2AX (pS139)-AlexaFluor-647 antibody (clone N1-430, BD Biosciences).

For single-cell comet electrophoresis assays, c-kit$^+$ bone marrow cells were isolated from adult shCtrl and shCux1 mice and cultured for five days with doxycycline and tazemetostat or vehicle to induce *Cux1* knockdown and EZH2 inhibition. Cells were irradiated (2 Gy) or mock irradiated, and embedded in 1% low melting point agarose (Invitrogen) at the indicated time points. Comet assays were performed with the Trevigen Comet Kit using the manufacturer instructions with the following modifications: Cells were embedded at 2x density, electrophoresis was run at 21 V for 30 minutes, and comets were stained with SYBR Green rather than SYBR Gold. Images of comet slides were taken on a wide-field microscope with a 20 X air objective. Images were analyzed with the OpenComet software plug-in on ImageJ.[286]

Immunofluorescence micrographs

For imaging experiments, round #1 cover slips were pre-treated with 0.1% poly-L-lysine for 12 hours and placed in a 24-well plate. $5x10^5$ U937 cells were seeded on each cover slip and allowed to settle for 12-16 hours. Irradiation was performed *in situ*. Cells were fixed by incubating with 4% paraformaldehyde (v/v) for 20 minutes at room temperature, washed three times with 1X PBS, and permeabilized with 0.1% Triton-X-100 in PBS for 15 minutes at room temperature. Cells were washed three times with 1X PBS, then blocked with 10% goat serum in PBS for 1 hour at room temperature. Primary

antibodies were diluted in 10% goat serum at the following concentrations: 53BP1 1:50, H2AX pS139 1:100, H3K27me3 1:5000. Coverslips were incubated overnight at 4˚C in primary antibodies, then washed three times with 1% goat serum in PBS. Secondary antibodies were diluted in 10% goat serum at 1:3000 for anti-rabbit AF647 and 1:1000 for anti-mouse AF488, and coverslips were incubated with secondary antibody for 2 hours at room temperature, away from light. Coverslips were washed three times with 1% goat serum in PBS, then stained for 10 minutes at room temperature with DAPI at 300 ng/mL in 1% goat serum, away from light. Coverslips were mounted on slides with ProLong Gold (Invitrogen), visualized with Zeiss Observer 7 microscope with a 63X oil objective, and pseudo-colored. 3 biological replicates were performed for each replicate and ≥50 cells were imaged per genotype per replicate.

Immunofluorescence image analysis

Foci counting was performed in ImageJ by thresholding and segmenting nuclei, then counting foci in each nucleus via a thresholding and FindMaxima routine. Foci intensity analysis was performed by segmenting the foci as above, then measuring the MFI of each focus. Foci within each nucleus were then grouped, and the mean Foci-MFI values for each nucleus were reported. H3K27me3 intensity was calculated by thresholding and segmenting nuclei, then measuring the MFI of H3K27me3 within each nucleus after background subtraction.

Histology

Tissues were fixed in 10% formalin, embedded in paraffin, sectioned at 4 μm, and stained with hematoxylin and eosin. Images were taken on a wide-field microscope with a 4 X air objective.

Data availability and statistical analysis

All statistical tests were conducted as indicated and all plots show biological replicates unless otherwise indicated. Tests were performed in GraphPad Prism version 10.2.2 or within R Studio for RNA-sequencing FDR value calculation. Significance testing was performed as described in the figure legends, with significance denoted for all plots as: ns, $p > 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Box plots show min-to-max values with a line at the mean. RNA-sequencing heatmaps were generated in R Studio using pheatmap. All R code for data generation, analysis, and plotting is available upon request.

RESULTS: DEFINING THE PATHOGENESIS OF DEL(7q) IN HEMATOPOIETIC STEM CELLS

Introduction

Aneuploid karyotypes are prevalent in human cancer, with over 90% of solid and 75% of hematopoietic malignancies deviating from the balanced complement of 46 chromosomes.[34] Large-scale gain or loss of chromosomal material alters the dosage of hundreds to thousands of genes, affecting multiple cellular processes including protein homeostasis, metabolism, and proliferation.[2–4] Despite this pervasiveness, the pathogenesis of recurrent aneuploidies remains largely unclear, in part due to a paucity of model systems. Isolated aneuploidies, particularly monosomies, are challenging to generate and often unstable in cell culture systems,[230,287] and the lack of chromosomal synteny between humans and animal species is a barrier to developing animal models of aneuploidy.[121,227,287]

Loss of all or part of chromosome 7 [-7/del(7q)] is recurrent in high-risk myeloid malignancies, including pediatric and adult myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML), and is associated with a poor prognosis.[76–78] Strikingly, -7/del(7q) occurs in up to 50% of therapy-related myeloid neoplasms (t-MN), also referred to as myeloid neoplasms post-cytotoxic therapy, high-risk malignancies which arise after prior exposure to chemotherapy or radiation.[79,288,289] Recently, chromosomal alterations have also been identified in clonal hematopoiesis, and del(7q) clonal hematopoiesis is associated with increased risk of developing myeloid disease.[94,96,111,112] As clonal hematopoiesis is associated with increased risk of t-MN development,[95,115] these observations suggest a model in which a pre-existing clonal population is selected for by

radiation and/or chemotherapy, and subsequently expands and transforms to t-MN. This may also contribute to the inherent drug resistance of t-MNs, as the initiating population has already undergone selection for increased fitness under genotoxic pressure.[121]

Decades of work have identified multiple candidate tumor suppressor genes (TSGs) within commonly deleted regions (CDR) of 7q.[74,85,121,146] Our lab identified the homeobox transcription factor *CUX1*, encoded in the 7q22 CDR, as a potent haploinsufficient TSG.[237] *Cux1* deficiency induces MDS with a lethal anemia, and *Cux1*-deficient cells are rapidly selected for after alkylating agent exposure and found t-MNs, providing multiple lines of evidence that even this single 7q gene deficiency can drive disease development.[135,136,142] However, additional 7q TSGs have hematopoietic deletion phenotypes and the effects, if any, of combinatorial 7q gene deletions on disease pathogenesis and drug resistance remain unclear.

In this study, we develop a murine model of del(7q) clonal hematopoiesis and drug resistance by simultaneously inactivating up to four 7q gene candidates (*Cux1*, *Ezh2*, *Kmt2c*, and *Kmt2e*) in hematopoietic stem cells (HSCs) using multiplex CRISPR-Cas9 editing. We observe significant myeloid expansion of clones edited for *Cux1* and *Ezh2*, which is rapidly accelerated upon chemotherapy exposure. *Cux1;Ezh2*-deficient cells also display reduced sensitivity to the AML therapeutic daunorubicin, linking these 7q genes to the drug resistance characteristic of t-MNs. Mechanistically, *Cux1;Ezh2*-deficient cells have a profoundly impaired DNA damage response (DDR) after genotoxic stress, with compounding perturbations to double-strand break recognition and repair. Overall, our data reveal a genetic interaction between *CUX1* and *EZH2* loss, and support the concept of 7q as a contiguous gene syndrome region whereby collective loss of multiple

TSGs promotes disease pathogenesis and drug resistance. Further, our CRISPR-based approach may serve as a tractable framework for interrogating recurrent aneuploid events more broadly in cancer.

In silico data mining identifies chromatin modifiers as candidate 7q genes modulating drug resistance

Previous work from our lab demonstrated a role for the 7q gene *CUX1* in chemotherapy-resistance and epigenetic-driven DNA repair.[142] To identify additional chromosome 7 genes that may modulate chemotherapy resistance, we leveraged publicly available genome-wide CRISPR-Cas9 knockout screens for chemotherapy resistance. We mined nine screens available on BioGRID for resistance to etoposide and cisplatin, chemotherapeutics which are associated with t-MN development, and anthracyclines, used in the treatment of *de novo* AML and t-MN.[121] These screens were conducted in human cell lines derived from both myeloid leukemias and solid tumors. Within each screen, we first identified any chromosome 7 gene in which targeting gRNAs were enriched in drug-treated cells compared to vehicle, then narrowed to genes expressed in $CD34^+$ HSCs. This yielded 113 genes in aggregate, which we then filtered to genes located in 7q CDRs, resulting in 39 gene candidates **(Table 5)**. Pathway analysis demonstrated significant enrichment for transcriptional co-regulators and chromatin-modifying enzymes, notably the histone modifiers *EZH2*, *KMT2C/MLL3*, and *KMT2E/MLL5* **(Figure 2A)**. Loss of *EZH2*, encoding the catalytic component of the H3K27 methyltransferase Polycomb Repressive Complex 2 (PRC2),[148,152] is associated with resistance to anthracyclines and poor prognosis in AML, validating our *in silico*

approach.[174] Heterozygosity of *KMT2C*, a component of the COMPASS complex that catalyzes H3K4 monomethylation,[183,290] promotes resistance to alkylating agents in HSCs.[191] Inactivating *KMT2C* mutations are also observed in relapsed pediatric AML, further supporting a role in chemotherapy resistance.[192] Loss of *KMT2E*, also a member of the KMT/MLL family, has been reported to impair HSC function,[197–199] and decreased expression of *KMT2E* is associated with worse outcomes in AML.[242]

KMT2C, *KMT2E*, and *EZH2* are among the most highly expressed 7q candidates in human HSPCs **(Figure 2A)**, and all three have been reported to mediate DNA damage repair;[181,182,193,194,200] therefore, we reasoned these three epigenetic regulators were appropriate candidates to test for increased drug resistance and interaction with *CUX1* loss. Despite megabase distances between these genes, they are collectively lost in over 70% of patients with -7/del(7q), based on copy-number analysis of two patient cohorts from The University of Chicago[237] and The Cancer Genome Atlas[275] **(Figure 2B)**. Collectively, the functional convergence of these genes on epigenetic regulation and the DDR provides strong rationale for investigating 7q as a contiguous gene syndrome region in which loss of multiple TSGs drives disease pathogenesis and drug resistance.

CRISPR RNP transfection achieves high-efficiency knockout in HSPCs

Using an inducible shRNA-based transgenic model of *Cux1* knockdown, our lab has reported that *Cux1*-deficient cells rapidly outcompete wild-type cells and expand following exposure to the DNA-damaging alkylating agent N-ethyl-N-nitrosourea (ENU) in mixed bone marrow chimera experiments.[142] To test whether additional 7q gene deficiencies promote expansion, either alone or in combination with shRNA-induced *Cux1*

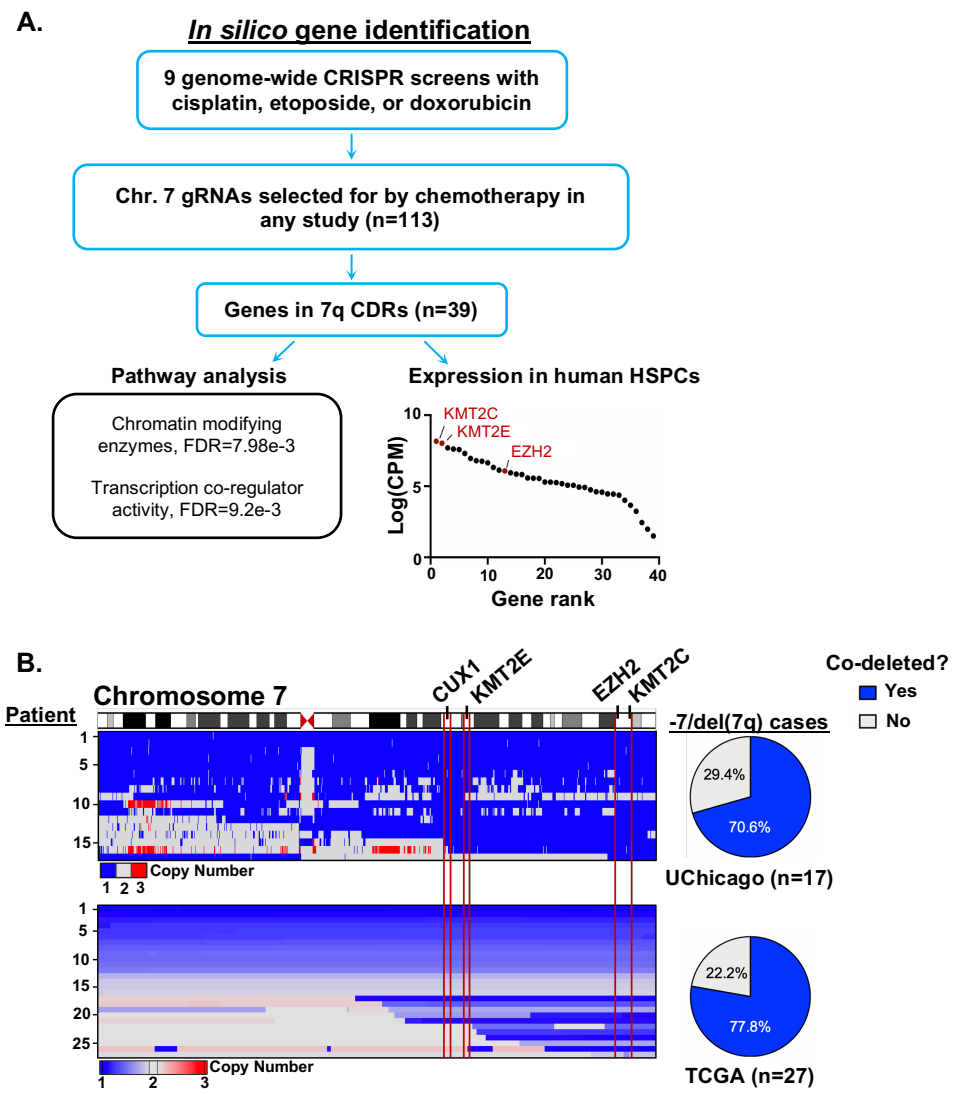

Figure 2: *In silico* data mining identifies *KMT2C*, *KMT2E*, and *EZH2* as candidates 7q TSGs modulating drug resistance.

A) Nine genome-wide screens were analyzed to identify chromosome 7 genes that promote resistance to chemotherapy when inactivated. Pathway analysis and RNA expression in human HSPCs is shown in counts per million. **B)** Copy-number analysis of chromosome 7 from two patient cohorts from The University of Chicago and TCGA. The regions encoding *CUX1*, *KMT2E*, *KMT2C*, and *EZH2* are denoted in red, and the percentage of -7/del(7q) patients with deletion of all four genes is shown.

deficiency, we used an RNP-based CRISPR-Cas9 delivery method[278,279] to knock out *Ezh2*, *Kmt2c*, or *Kmt2e* in HSPCs **(Figure 3A)**. This approach yields high levels of insertions and deletions (indels) detectable within 72 hours of transfection with consequent reductions in protein levels **(Figure 3B, C)**. We transplanted wild-type CD45.1 recipients with RNP-transfected CD45.2 HSPCs harboring shRNA against *Cux1* (shCux1) or renilla luciferase as a control (shCtrl). After an initial engraftment period, we injected recipient mice with ENU (100 mg/kg) or vehicle to model chemotherapy exposure, and tracked gene knockout in peripheral blood, defined as the sum of frameshift indel frequencies. Mice treated with ENU experienced a transient leukopenia consistent with exposure to genotoxic agents, demonstrating efficacy of ENU treatment **(Figure 3D)**. As a control for off-target effects of Cas9, we included a transplant cohort receiving HSPCs transfected with RNPs targeting an intron of *Actb*.[291] We detected gene knockout beyond sixteen weeks post-transplant, indicating this RNP transfection-based approach successfully targets long-term HSCs.[292] We observed no significant changes in the frequency of *Actb* indels after ENU exposure **(Figure 3E)**.

Combined *Cux1* and *Ezh2* deficiency promotes hematopoietic expansion after exposure to alkylating agents

In striking contrast to *Actb*, we observed significant expansion of *Ezh2* knockout in ENU-treated mice given transfected shCux1 HSPCs **(Figure 4A)**. Indeed, we detected significant increases in *Ezh2* knockout in every hematopoietic tissue in ENU-treated shCux1 recipients **(Figure 4B)**. In mice given shCtrl-transfected HSPCs there was a general increase in *Ezh2* knockout over time, but ENU administration did not affect *Ezh2*

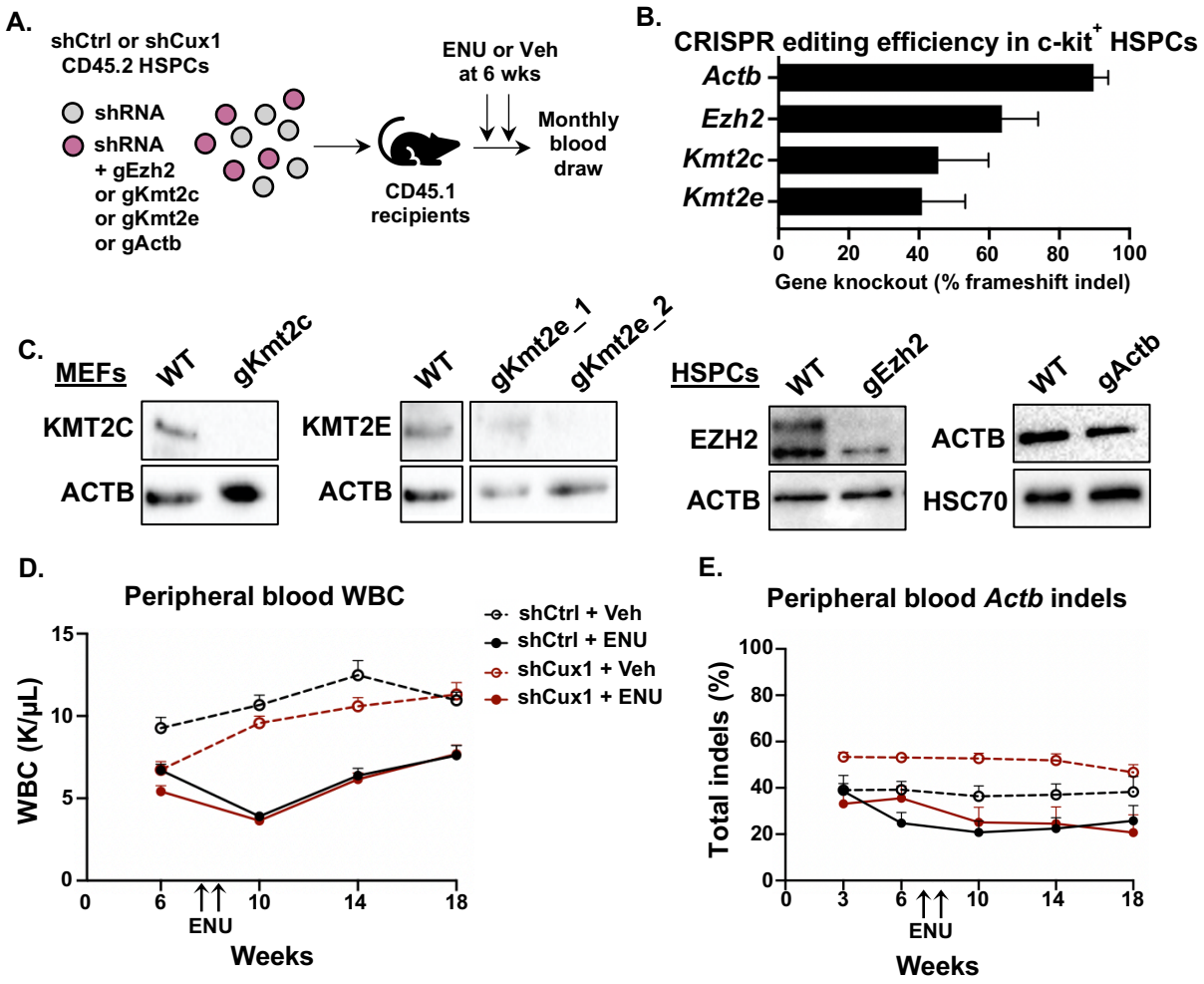

Figure 3: CRISPR RNP transfection achieves high efficiency knockout in HSCs.

A) Experimental scheme for transplant cohorts. shCtrl and shCux1 c-kit+ HSPCs were transfected with RNPs targeting Ezh2, Kmt2c, Kmt2e, or an intron of β-Actin and transplanted into wild-type recipients. Two doses of ENU (100 mg/kg) were administered intraperitoneally nine days apart at week six post-transplant. **B)** 7q gene knockout percentages, defined as the sum of frameshift indels, after RNP transfection. Total indels, including those in-frame, are plotted for *Actb* as the gRNA targets an intron. n=4 trials for *Actb*, n=15 for *Ezh2*, n=17 for *Kmt2c*, and n=14 for *Kmt2e*. Data are plotted as mean + SD. **C)** Western blot validation of 7q gRNAs in mouse embryonic fibroblasts (MEFs) or HSPCs. MEFs were used to achieve higher knockout efficiency to confirm knockdown. **D)** Total white blood cell (WBC) counts in peripheral blood for all mice receiving shCtrl or shCux1 HSPCs with a single 7q gene deficiency. Data are plotted as mean + SEM, n=28-30 mice/experimental group. **E)** Neutral *Actb* indel frequencies were serially tracked in peripheral blood mononuclear cells by Sanger sequencing before and after ENU administration. The percentage of DNA with indels is shown for vehicle- and ENU-treated recipients of shCtrl and shCux1 HSPCs. Two independent biological replicates were performed. Data are plotted as mean + SEM, n=10-15 mice/experimental group.

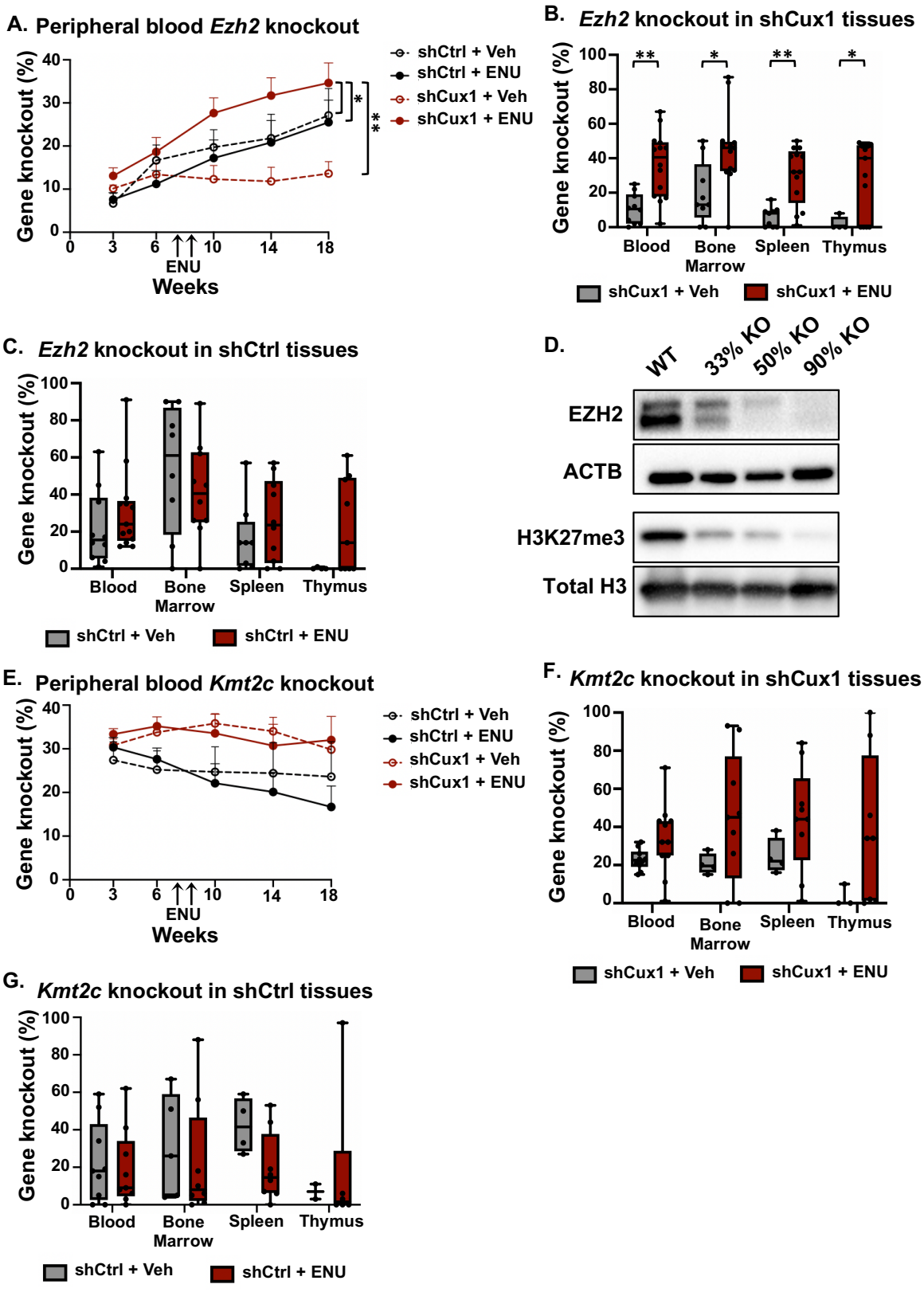

Figure 4: Combined *Cux1* and *Ezh2* loss is preferentially selected for after ENU.

Figure 4 continued

A) *Ezh2* gene knockout was serially tracked in peripheral blood mononuclear cells by Sanger sequencing before and after vehicle or ENU administration. Data are plotted as mean + SEM, and two independent biological replicates were performed with 5-7 mice per replicate per genotype, for a total of n=10-14 mice/experimental group. Significance from a two-way ANOVA is shown. * $p < 0.05$, ** $p < 0.01$. **B)** Box plots of *Ezh2* knockout in hematopoietic tissues at euthanasia of vehicle- or ENU-treated shCux1 recipient mice, 18-22 weeks post-transplant, n=10-13 mice. Significance from a two-tailed t-test is shown. * $p < 0.05$, ** $p < 0.01$. **C)** Box plots of *Ezh2* knockout in hematopoietic tissues at euthanasia of vehicle- or ENU-treated shCtrl recipient mice, 18-22 weeks post-transplant, n=8-13 mice per tissue with exception of vehicle-treated thymus samples, with n=4. **D)** Western blot of bone marrow harvested from ENU-treated mice 18-22 weeks post-transplant with a range of *Ezh2* knockout levels, confirming sustained decreases to EZH2 protein and H3K27me3. **E)** *Kmt2c* knockout was tracked as in A for vehicle- and ENU-treated recipients. Data are plotted as mean + SEM, and two independent biological replicates were performed with 4-6 mice per replicate per genotype, for a total of n=9-11 mice/experimental group. **F)** Box plots of *Kmt2c* knockout in hematopoietic tissues of shCux1 recipient mice at euthanasia as in B, n=3-11 mice. **G)** Box plots of *Kmt2c* knockout in shCtrl recipient mice as in C, n=4-9 mice per tissue with exception of vehicle-treated thymus samples, with n=2.

Ezh2 expansion in peripheral blood or other tissues **(Figure 4B, C)**. Bone marrow *Ezh2* knockout approached 90% in some ENU-treated mice, consistent with bi-allelic *Ezh2* loss **(Figure 4B, D)**. Bi-allelic *EZH2* mutations have been observed in patients with 7q loss of heterozygosity, suggesting a stronger fitness advantage from complete loss of *EZH2*.[80,111]

In contrast to *Ezh2*, we did not observe selection for *Kmt2c* knockout in the peripheral blood of either transplant group after ENU exposure **(Figure 4E)**. We did detect increased levels of *Kmt2c* knockout in other hematopoietic tissues of ENU-treated shCux1 recipients **(Figure 4F)**, in line with reports that *Kmt2c* loss promotes HSC expansion after exposure to alkylating agents,[191] but we did not detect increases in shCtrl recipients **(Figure 4G)**. Further, the expansion was highly variable within shCux1 recipients, with knockout levels near 90% in some mice and below the limit of detection in others, indicating complete or near-complete dropout of edited cells **(Figure 4F)**. Thus, while there may be some degree of genetic interaction between *Cux1* and *Kmt2c* loss after ENU exposure, our data demonstrate a stronger connection between *Cux1* and *Ezh2*.

It has been reported that *Kmt2e*-null cells fail to engraft after transplant.[197] Consistent with this, we noted very low levels of *Kmt2e* knockout in peripheral blood **(Figure 5A)**, despite detectable *Kmt2e* knockout after RNP transfection and robust engraftment of CD45.2 cells in both transplant groups **(Figure 5B)**. However, we did observe increased *Kmt2e* knockout in the bone marrow of shCux1 recipients compared to shCtrl, suggesting *Cux1* deficiency may partially rescue engraftment defects **(Figure 5C)**. Overall, using a series of transplant cohorts, we demonstrate combined deficiency of *Cux1* and *Ezh2* uniquely promotes outgrowth under genotoxic stress.

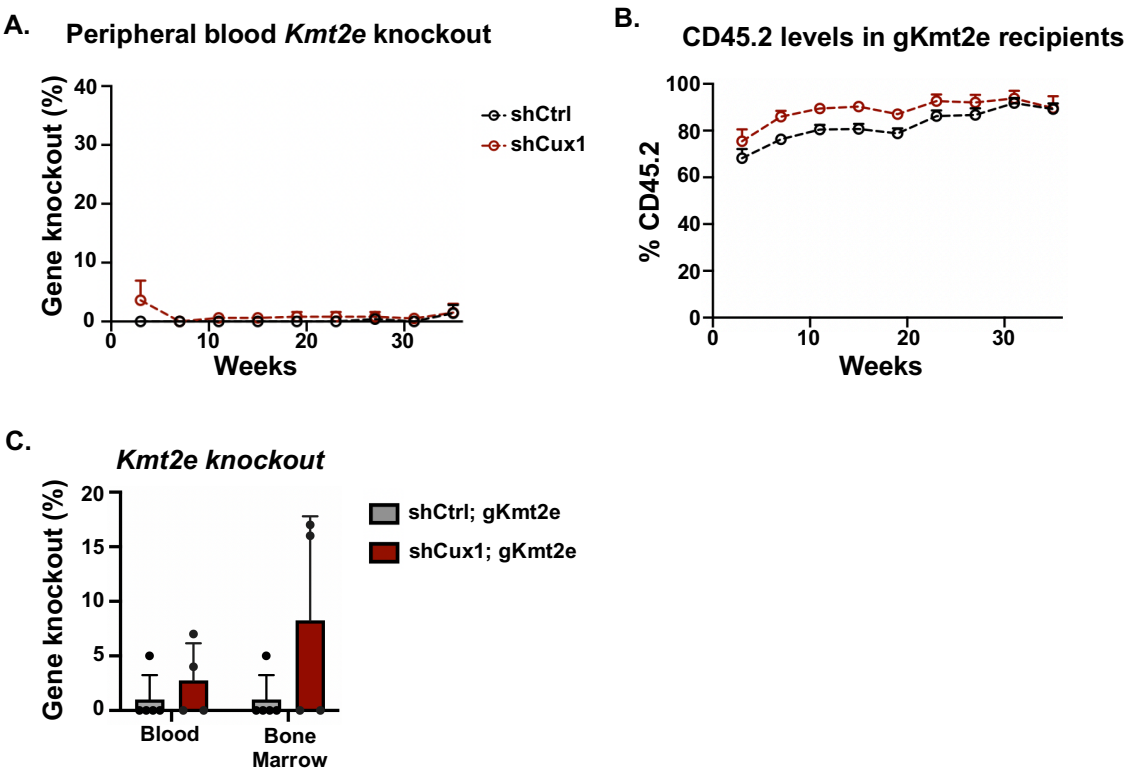

Figure 5: *Cux1* deficiency partially rescues *Kmt2e*-null engraftment failure.

A) *Kmt2e* gene knockout was serially tracked in peripheral blood mononuclear cells by Sanger sequencing over 36 weeks, n=5 mice/experimental group. **B)** Total CD45.2 positivity in peripheral blood for the same mice in A. Data are plotted as mean + SEM. **C)** Kmt2e knockout in peripheral blood and bone marrow at euthanasia. Data are plotted as mean + SEM, n=5 mice for shCtrl and n=4 mice for shCux1.

Multiplex knockout of 7q genes in HSPCs indicates that combined *Cux1* and *Ezh2* deficiency is the main driver of expansion after chemotherapy exposure

Having observed strong selection for combined *Cux1;Ezh2* deficiency, and potential selection for combined *Cux1;Kmt2c* deficiency, we wanted to test whether cells deficient in up to all four 7q genes would have an even strong selective advantage. Therefore, we expanded our RNP approach and transfected shCtrl and shCux1 CD45.2 HSPCs with a pool of RNPs targeting *Ezh2*, *Kmt2c*, and *Kmt2e* simultaneously **(Figure 6A)**. To ensure the low engraftment of *Kmt2e*-null cells was not due to off-target effects, we used a second *Kmt2e* gRNA targeting a different exon in these experiments **(Figure 2C)**. Genotyping colonies derived from individual HSPCs revealed this multiplex knockout approach is highly efficient, with 75% of colonies containing gene edits and 60% showing perturbations to all three genes **(Figure 6B)**. Further, we observed both mono-allelic and bi-allelic knockout across the population, enabling modeling of monosomy as well as recapitulating the bi-allelic *EZH2* and *KMT2C* mutations observed, albeit infrequently, in patients.[80,111,293]

We reasoned the knockout combination with the highest fitness should be selected for, indicating which 7q gene deficiencies drive expansion. We transplanted multiplex-edited shCtrl and shCux1 HSPCs into CD45.1 recipients, treated with ENU or vehicle to model chemotherapy exposure, and tracked gene knockout in peripheral blood. Input knockout levels of all three genes were comparable between shCtrl- and shCux1-modified cells for both vehicle- and ENU-treated cohorts **(Figure 6C)**. At a population level, knockout frequencies were relatively stable within the vehicle-treated mice **(Figure 7A)**, although there were some exceptions characterized below. In contrast, exposure to ENU

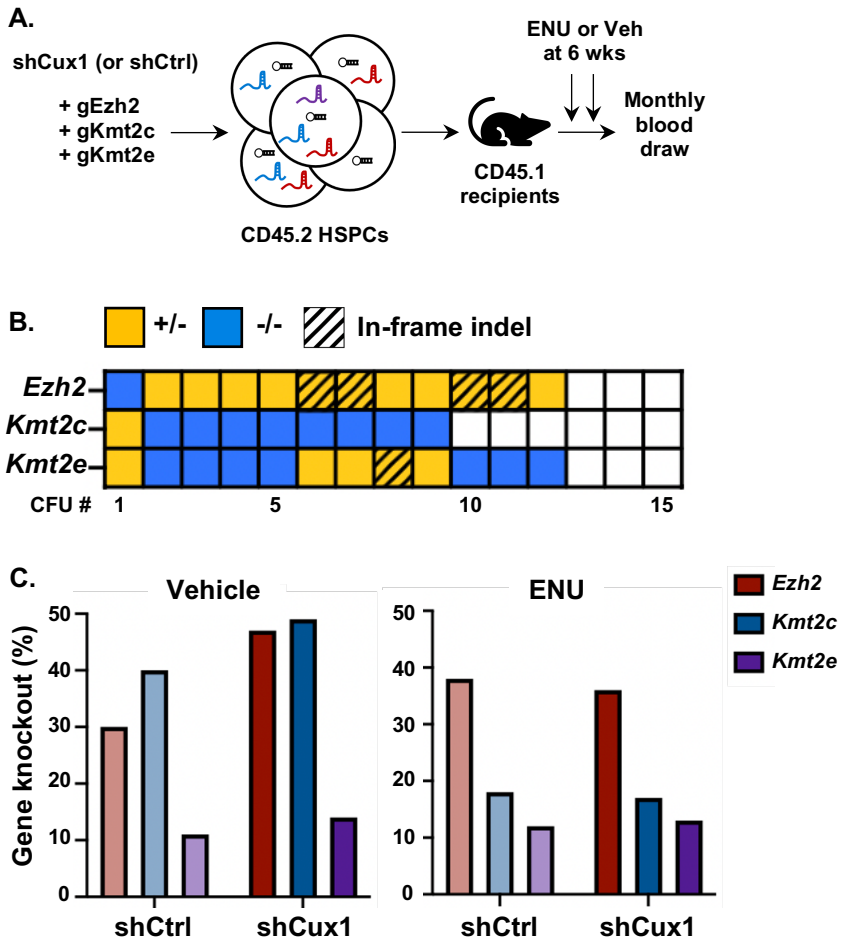

Figure 6: A pooled RNP transfection approach achieves high-efficiency multiplex gene knockout in HSCs.

A) Experimental schematic for multiplex-knockout transplant cohorts. shCtrl and shCux1 c-kit+ HSPCs were transfected with a pool of RNPs targeting *Ezh2*, *Kmt2c*, and *Kmt2e* to generate mosaic populations of cells harboring different combinations of 7q gene knockout. Two doses of ENU were administered to recipient mice (100 mg/kg) nine days apart beginning six weeks post-transplant. **B)** Transfected HSPCs were plated in MethoCult M3434, cultured for 7-10 days, and colonies derived from individual cells were genotyped. Each column represents data from one colony. Blue shading denotes homozygous knockout, yellow denotes heterozygous knockout, and hatched lines denote the second allele harbored an in-frame indel that does not cause a frameshift. 15 colonies were analyzed across 2 independent multiplex transfections. **C)** CRISPR knockout efficiencies for shCtrl and shCux1 HSPCs transfected with the pool of RNPs. 4 total sets of transfections were performed to generate the vehicle- and ENU-treated transplant cohorts.

rapidly accelerated expansion of *Ezh2* knockout in all mice, with shCux1 recipients showing significantly higher *Ezh2* knockout than shCtrl recipients over 40 weeks **(Figure 7B)**. ENU-treated shCux1 recipients also displayed a striking expansion of myeloid cells in the peripheral blood that strongly correlated with *Ezh2* knockout levels **(Figure 7C)**. Compared to shCtrl recipients, the bone marrow of ENU-treated shCux1 recipient mice displayed significantly higher levels of myeloid cells and *Ezh2* knockout **(Figure 7D)**. These data demonstrate strong selection for combined *Cux1;Ezh2* deficiency, and suggest loss of these two 7q genes directly contributes to pathogenic myeloid expansion, particularly under genotoxic pressure.

Individual mice inform the consequences of multiple 7q gene loss on clonal expansion

In recipients of multiplex-edited shCtrl HSPCs, 70% (7/10) of vehicle-treated mice showed stable knockout scores for *Ezh2*, *Kmt2c*, and *Kmt2e* for the duration of the study. In the remaining 30% (3/10), we observed a successive increase in the percent of *Ezh2*, *Kmt2c*, and *Kmt2e* knockout **(Figure 8A)**. Mouse 391 was euthanized early due to severe anal prolapse and penile infection (balanitis), but displayed a rapid increase for all three 7q genes. In mouse 392 and 458, the steady increase in knockout after 20 weeks suggests a stem cell with combinatorial loss of these epigenetic regulators may have a clonal advantage independent of *Cux1* loss. In the bone marrow of mouse 458, we observed a major clone with bi-allelic *Ezh2* and mono-allelic *Kmt2e* knockout **(Figure 8B)**. As *Kmt2e* knockout was not readily detected early, this suggests an $Ezh2^{-/-}$;$Kmt2e^{+/-}$ stem cell existed after transplant but only gained a clonal advantage through aging.

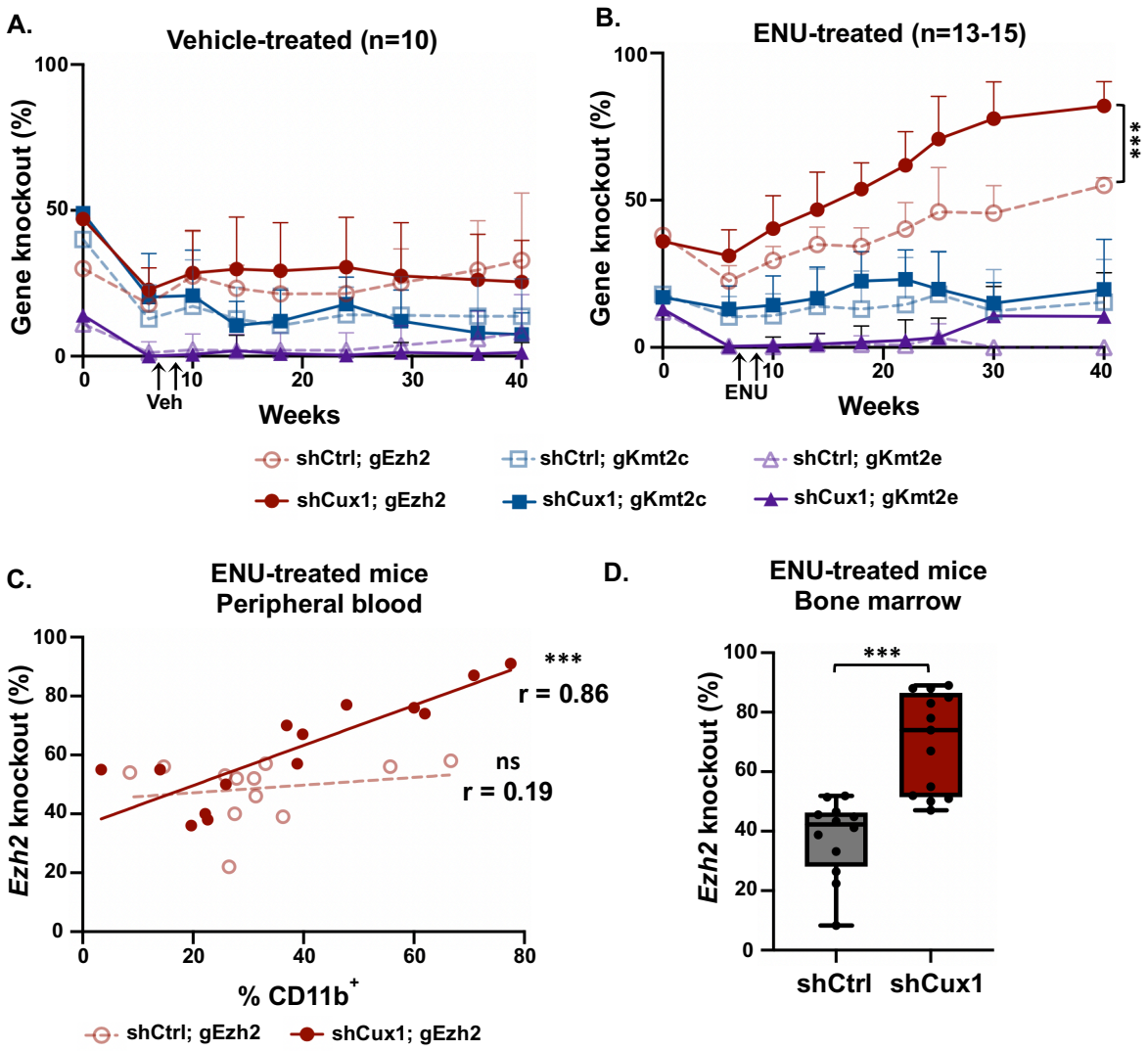

Figure 7: Multiplex knockout of 7q genes in HSCs demonstrates combined *Cux1;Ezh2* deficiency drives expansion after chemotherapy exposure.

A) Gene knockout frequencies for *Ezh2*, *Kmt2c*, and *Kmt2e* were serially tracked in peripheral blood of transplant recipients before and after vehicle administration. Data are plotted as mean + SD, total n=10 mice receiving multiplex-edited shCtrl or shCux1 cells. **B)** Gene knockout frequencies in peripheral blood as in A, for mice given ENU. Data are plotted as mean +SD, total n=13-15 mice receiving multiplex-edited shCtrl or shCux1 cells. Significance from a two-way ANOVA shown. *** $p < 0.001$. **C)** The percent of CD11b+ myeloid cells (x-axis) and *Ezh2* knockout (y-axis) in the peripheral blood of ENU-treated recipients at euthanasia. CD11b positivity was assessed by flow cytometry. Simple linear regression was performed to fit a line to the data, and significance from Pearson's r is shown. *** $p < 0.001$. **D)** Box plots showing total *Ezh2* knockout in the bone marrow of ENU-treated recipients at euthanasia. Significance from a two-tailed t-test is shown. *** $p < 0.001$.

For recipients of multiplex-edited shCux1 HSPCs, 70% (7/10) of vehicle-treated mice also displayed unchanging knockout scores. Clonal expansion in the remaining 30% (3/10) of mice was only monitored short-term clonal expansion since they developed malignancies under 24 weeks. **(Figure 8C)**. Mouse 442 contained a high percentage of $Gr1^+$ $Cd11b^+$ myeloid cells in the bone marrow and spleen (81% and 46%) and analysis of individual bone marrow-derived colonies showed that *Cux1*-deficient clones were haploinsufficient for *Ezh2*, alone and in combination with *Kmt2c* or *Kmt2e* **(Figure 8D)**. Compared to shCtrl, the shCux1 recipients displayed a concurrent increase in total WBC and decrease in RBC, along with an increase in peripheral blood neutrophils and decrease in lymphocytes **(Figure 9A, B)**, consistent with -7/del(7q) being recurrent in myeloid disease. However, the bone marrow and spleen of mouse 441 and 466 were infiltrated with >75% T cells **(Figure 9C)**, and secondary transplants confirmed that the major malignant clones in these mice were of T cell origin. Ultimately, these experiments provide proof-of-principle that combinatorial loss of multiple 7q genes can promote clonal expansion at low frequency and long latency, and *Cux1* loss may positively impact the rate of expansion.

ENU administration promoted the expansion of *Cux1;Ezh2*-deficient clones **(Figure 7B)**. Since ENU is a known inducer of thymic lymphomas,[294] most of our mice succumbed to this regardless of genotype. However, 31% (4/13) and 27% (4/15) of recipients of multiplex-edited shCtrl and shCux1 HSPCs, respectively, did not develop thymic lymphomas; all eight mice presented with progressive development of anemia and showed splenomegaly at time of euthanasia **(Figure 10A, B)**. The effacement of splenic architecture, granulocytosis, and monocytosis, was more severe in shCux1 recipient mice

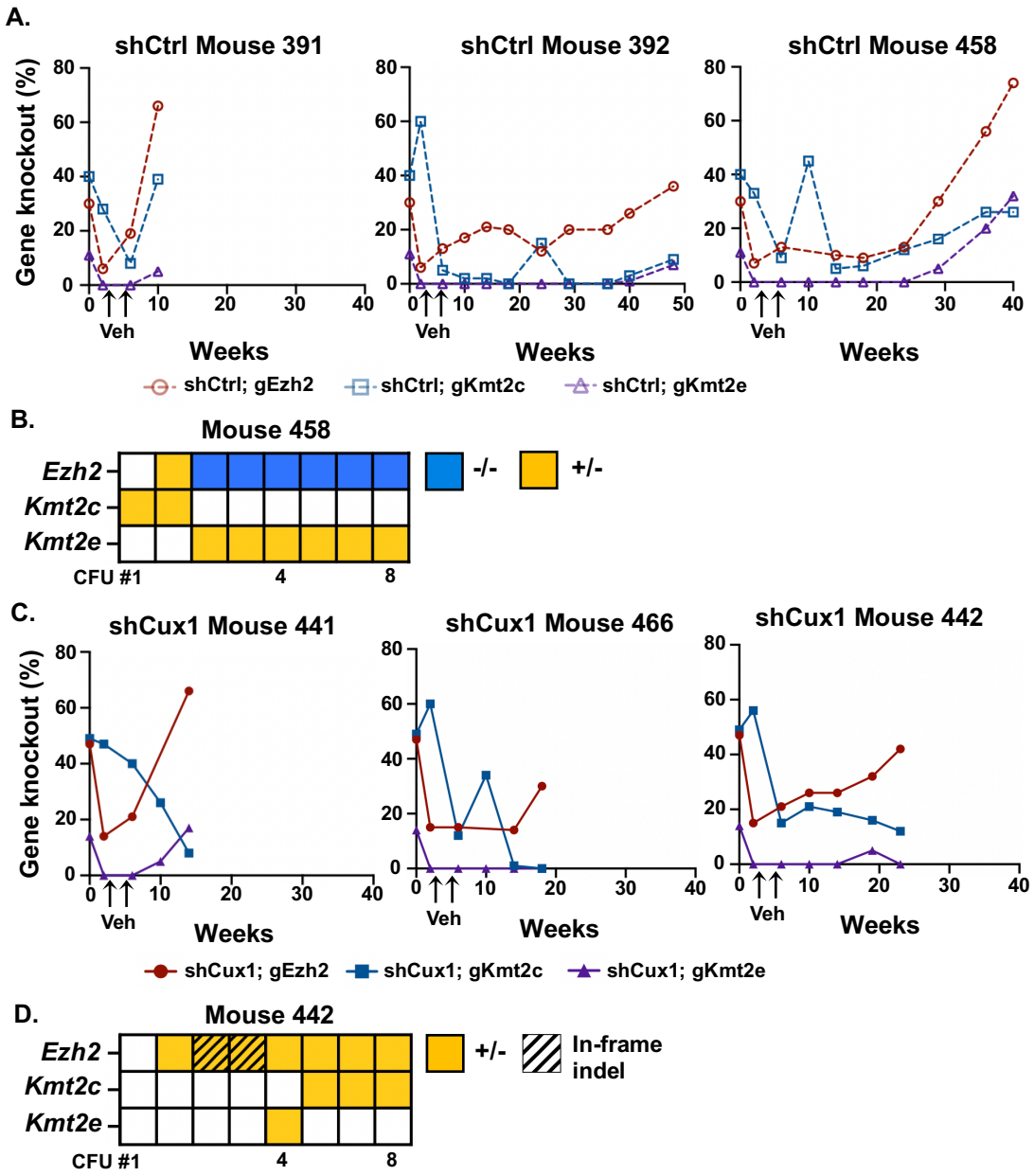

Figure 8: *Cux1* loss accelerates clonal expansion in vehicle-treated multiplex-knockout recipients.

A) Peripheral blood gene knockout frequencies for *Ezh2*, *Kmt2c*, and *Kmt2e* in three vehicle-treated recipients of multiplex-edited shCtrl HSPCs that demonstrated knockout expansion over time. **B)** Bone marrow cells from shCtrl Mouse 458 were plated in MethoCult M3434 methylcellulose, cultured for one week, and colonies derived from individual cells were genotyped. Each column represents data from one colony. Blue shading denotes homozygous knockout, and yellow shading denotes heterozygous knockout. **C)** Gene knockout frequencies as in A for three vehicle-treated recipients of multiplex-edited shCux1 HSPCs that demonstrated knockout expansion over time.

Figure 8 continued

D) Genotyping colonies as in B from bone marrow of shCux1 mouse 442. Yellow shading denotes heterozygous knockout, and hatched lines denote the second allele harbored an in-frame indel that does not cause a frameshift. Each column represents data from one colony.

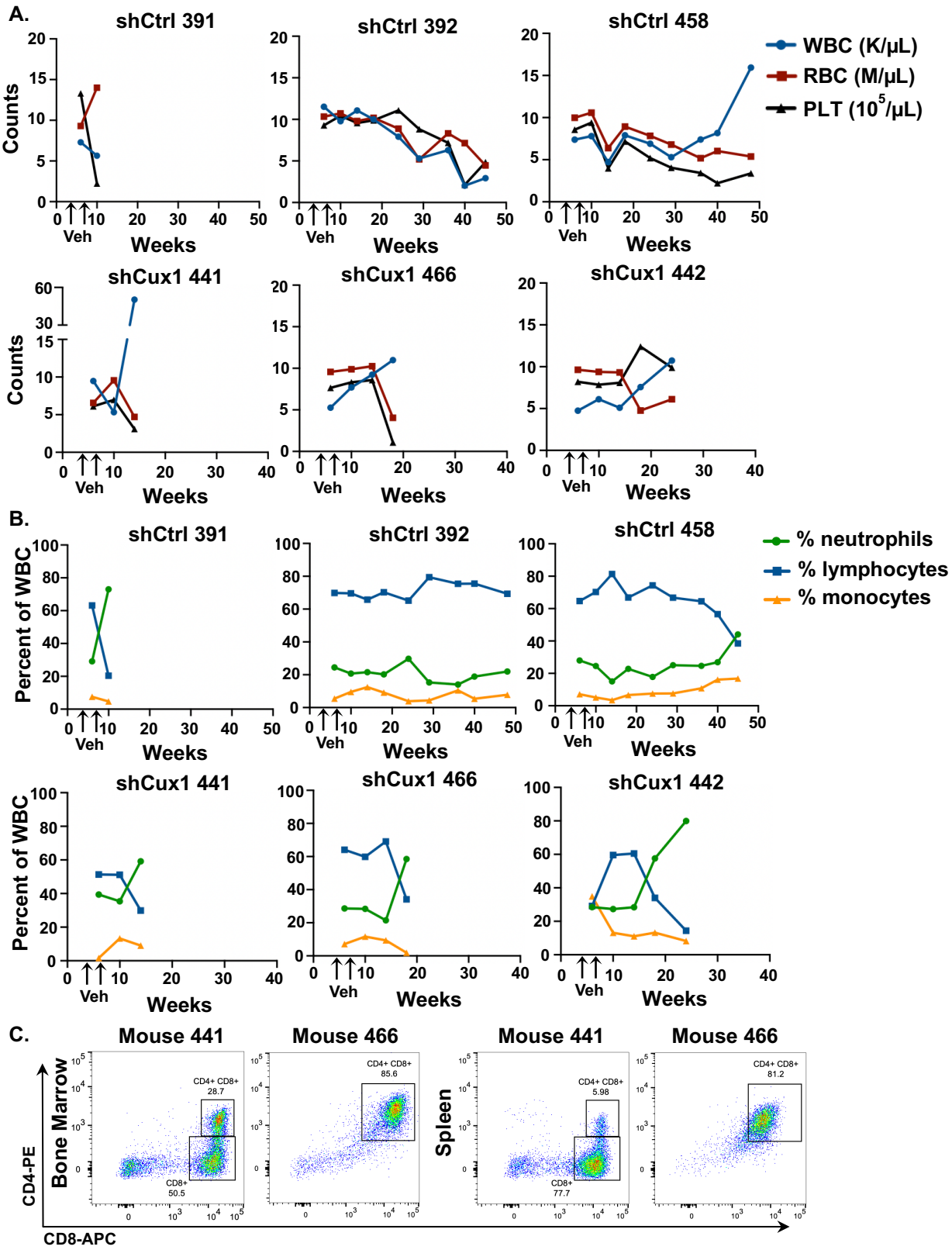

Figure 9: shCux1 multiplex-knockout recipients display increased WBCs and neutrophils and decreased RBCs and lymphocytes.

Figure 9 continued

A) Total white blood cell (WBC), red blood cell (RBC), and platelet (PLT) counts over time in peripheral blood of the vehicle-treated shCtrl recipient mice (top) and shCux1 recipient mice (bottom) with expansion of *Ezh2* knockout. Complete blood counts were obtained on a Hemavet 950 counter. **B)** A breakdown of WBC into neutrophils, lymphocytes, and monocytes is shown for the mice in A. **C)** Flow cytometry dot plots of Mouse 441 and Mouse 466 bone marrow and spleen showing T-cell expansion.

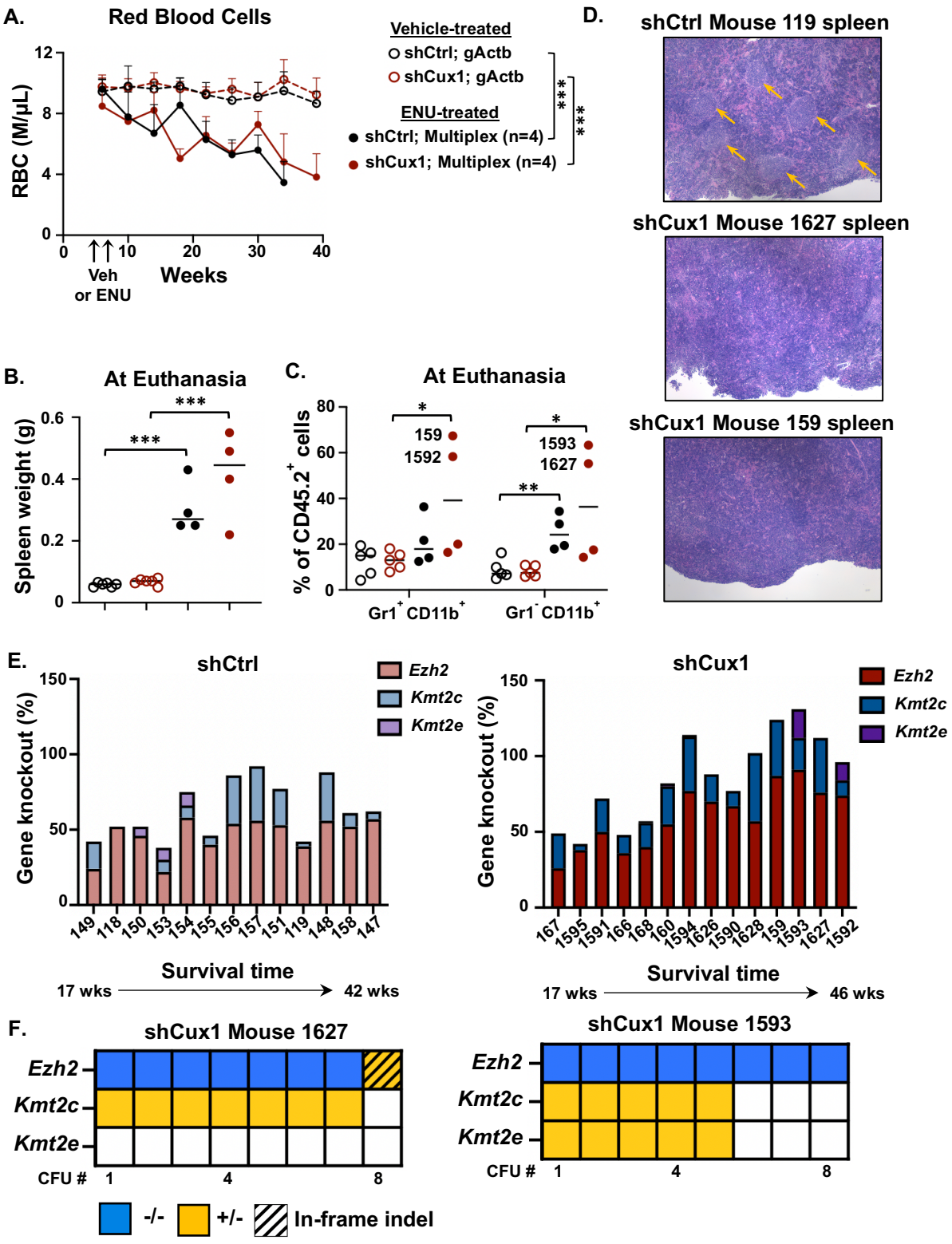

Figure 10: *Cux1* loss increases myeloid output and selects for bi-allelic *Ezh2* loss in ENU-treated multiplex-knockout recipients.

Figure 10 continued

A) Total red blood cell count over time in vehicle-treated recipients of HSPCs transfected with RNPs targeting an intron of *Actb*, and ENU-treated recipients of multiplex-edited HSPCs. Vehicle-treated *Actb* mice were used as a comparison as ENU-treated *Actb* mice died by 14 weeks. Data are plotted as mean + SD, n=5 mice/experimental group for *Actb* recipients, n=4 mice/experimental group for multiplex recipients. Significance from a two-way ANOVA is shown. *** $p < 0.001$. **B)** Spleen weight at sacrifice for the same mice as in A, with significance from a two-tailed t-test shown. *** $p < 0.001$. **C)** CD45.2+ peripheral blood granulocytes (Gr1+ CD11b+) and monocytes (Gr1- CD11b+) for the same mice as in A and B, with significance from two-tailed t-tests shown. * $p < 0.05$, ** $p < 0.01$. **D)** Hematoxylin and eosin (H&E) staining of spleen sections from ENU-treated recipients of multiplex-edited HSPCs. shCtrl Mouse 119 retains visible boundaries between white pulp and red pulp, indicated with arrows, whereas shCux1 Mouse 1627 and Mouse 1593 show dramatic effacement of splenic architecture. Images were taken at 4X magnification. **E)** Stacked bar plots of 7q gene knockout in recipients of multiplex-edited shCtrl cells (left) and shCux1 cells (right), arranged by survival time. Ezh2 knockout approaches 100%, consistent with bi-allelic gene loss, in aged recipients of shCux1 cells. **F)** Bone marrow cells from shCux1 Mouse 1627 and Mouse 1593 were plated in Methocult M3434, cultured for 7-10 days, and colonies derived from individual cells were genotyped. Each column represents data from one colony. Blue shading denotes homozygous knockout, yellow denotes heterozygous knockout, and hatched lines denote the second allele harbored an in-frame indel that does not cause a frameshift. Many colonies harbor homozygous Ezh2 knockout.

(Figure 10C, D). At euthanasia, peripheral blood *Ezh2* knockout approached 100% in shCux1 recipients (74-91%) and ranged from 39-57% in shCtrl recipients **(Figure 10E)**. Consistent with this, the majority of bone marrow clones in shCux1 recipients had bi-allelic *Ezh2* knockout **(Figure 10F)**. This suggests complete *Ezh2* inactivation may be positively selected for by *Cux1* loss and stress.

Kmt2c loss associates with but does not drive expansion of *Ezh2*- or *Cux1;Ezh2*-deficient clones

While all ENU-treated mice displayed peripheral blood expansion of *Ezh2* knockout, some mice showed simultaneous expansion of *Ezh2*- and *Kmt2c*-edited cells. An example of concurrent expansion is given in **Figure 11B**, with a comparison in **Figure 11A**. We reasoned the expanding clones in mice with significantly correlated *Ezh2* and *Kmt2c* knockout levels likely harbored inactivation of both genes **(Figure 11C)**. Genotyping individual bone marrow-derived colonies confirmed these mice displayed a dominant clone with knockout of both *Ezh2* and *Kmt2c*. In contrast, there were no colonies with both *Ezh2* and *Kmt2c* inactivation in mice with uncorrelated peripheral blood *Ezh2* and *Kmt2c* knockout **(Figure 11D, E)**.

We wanted to assess whether clones with concurrent *Ezh2* and *Kmt2c* loss expanded faster or to a greater extent after ENU exposure than those with solely *Ezh2* loss. We used maximum peripheral blood knockout as a measure of the largest clonal expansion, and compared maximum *Ezh2* and *Kmt2c* knockout values between mice with and without significantly correlated *Ezh2* and *Kmt2c* knockout trajectories. In both shCtrl and shCux1 recipients, the maximum *Kmt2c* knockout was almost two-fold higher in mice

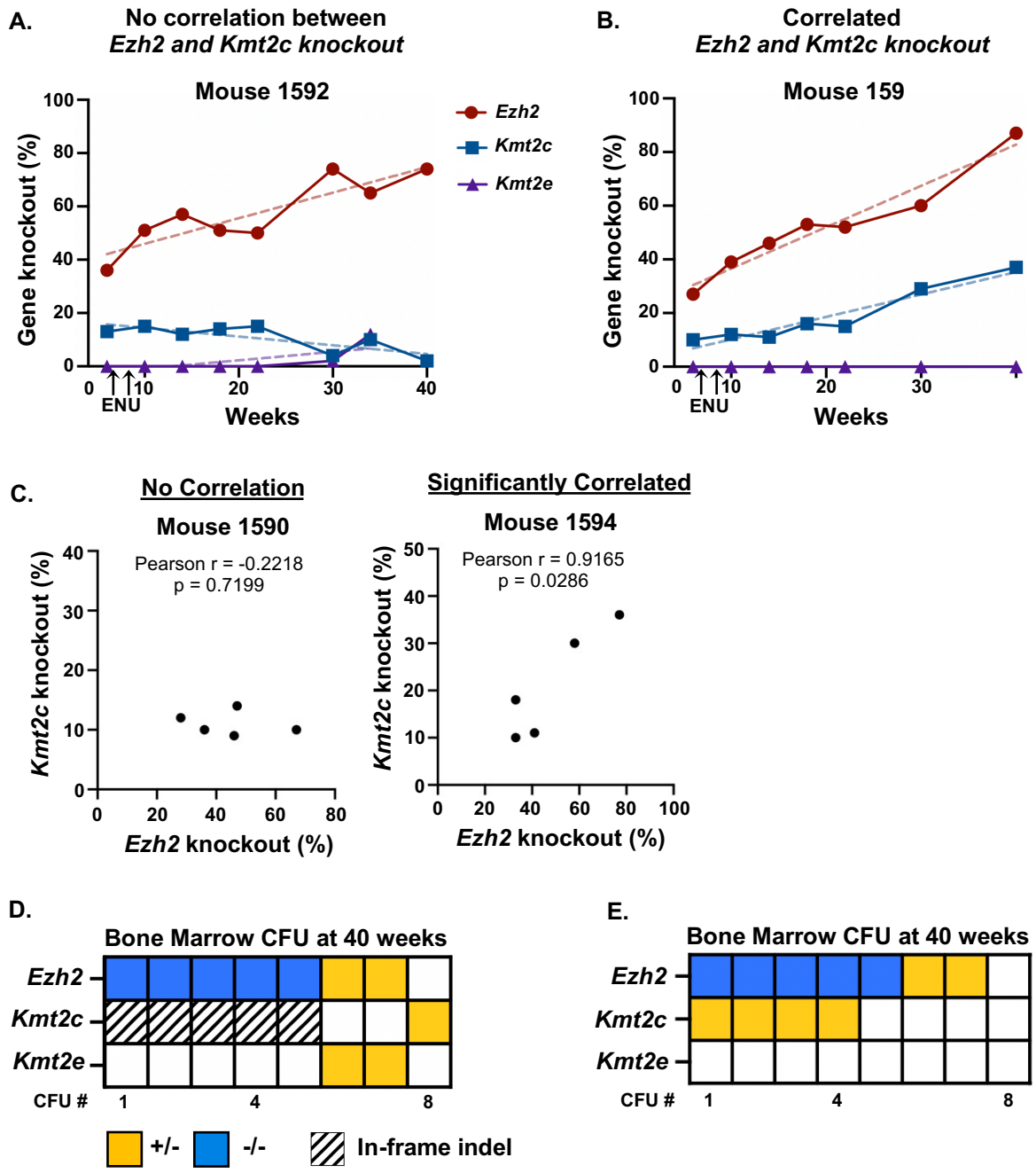

Figure 11: *Kmt2c* and *Ezh2* are co-edited in mice with significantly correlated gene knockout.

A) Peripheral blood gene knockout frequencies of an individual mouse (shCux1 Mouse 1592) in which *Ezh2* and *Kmt2c* knockout trajectories are not correlated. Simple linear regression was performed to fit a line to the data, shown in dashes. **B)** Same as in A, for an individual mouse (shCux1 Mouse 159) with significantly correlated Ezh2 and Kmt2c knockout. **C)** Example plots of correlations between Ezh2 and Kmt2c knockout in peripheral blood of ENU-treated recipients of multiplex-edited HSPCs. Gene knockouts

Figure 11 continued

were plotted against each other and Pearson's r was calculated, with $p < 0.05$ used as a benchmark for significant correlation indicating concurrent Ezh2 and Kmt2c knockout. **D)** Bone marrow cells from the mouse in A (shCux1 Mouse 1592) were plated in MethoCult M3434 methylcellulose, cultured for one week, and colonies derived from individual cells were genotyped. Each column represents data from one colony. Blue shading denotes homozygous knockout, yellow denotes heterozygous knockout, and hatched lines denote the second allele harbored an in-frame indel that does not cause a frameshift. **E)** Same as in D, for the mouse in B with no correlation between *Ezh2* and *Kmt2c* knockout (shCux1 Mouse 159).

with correlated (concurrent) vs. non-correlated *Ezh2*/*Kmt2c* knockout **(Figure 12A)**. In shCtrl recipients, *Ezh2* knockout remained at ~50% regardless of correlation with *Kmt2c* knockout, suggesting *Kmt2c* knockout is a passenger mutation in these clones and has no influence on the dynamics of *Ezh2* expansion **(Figure 12B)**. In contrast, in shCux1 recipients there was a trend (p=0.066) toward higher *Ezh2* knockout in mice with correlated *Ezh2*/*Kmt2c* levels **(Figure 12B)**. As a second metric to assess the effects of concurrent *Ezh2* and *Kmt2c* loss, we compared the slopes of *Ezh2* and *Kmt2c* knockout in mice with and without correlated expansion **(Figure 12C)**. This analysis also demonstrated a trend toward increased expansion in mice with combined *Cux1;Ezh2;Kmt2c* deficiency compared to *Cux1;Ezh2* alone (p=0.08). Collectively, our data suggest *CUX1* and *EZH2* loss influence -7/del(7q) clonal expansion and t-MN foundation. However, the notable association of *Kmt2c* knockout with expanding *Cux1;Ezh2*-deficient clones may have progressive consequences on malignancy.

***CUX1*;*EZH2* deficiency promotes resistance to the myeloid neoplasm therapeutic daunorubicin**

As t-MNs are often drug resistant, we wanted to assess whether loss of 7q genes can also promote resistance to chemotherapy agents used to treat these diseases. We generated single-cell CRISPR knockout clones in U937 human myeloid leukemia cells with heterozygous loss of each 7q gene **(Figure 13A)**, and tested for increased resistance to the anthracycline daunorubicin, a core component of induction therapy for myeloid neoplasms including MDS, AML, and t-MN.[295] As controls, we used a clone with a neutral indel in an intron of the *HPRT* gene, and a clone with a truncating *PPM1D* mutation known

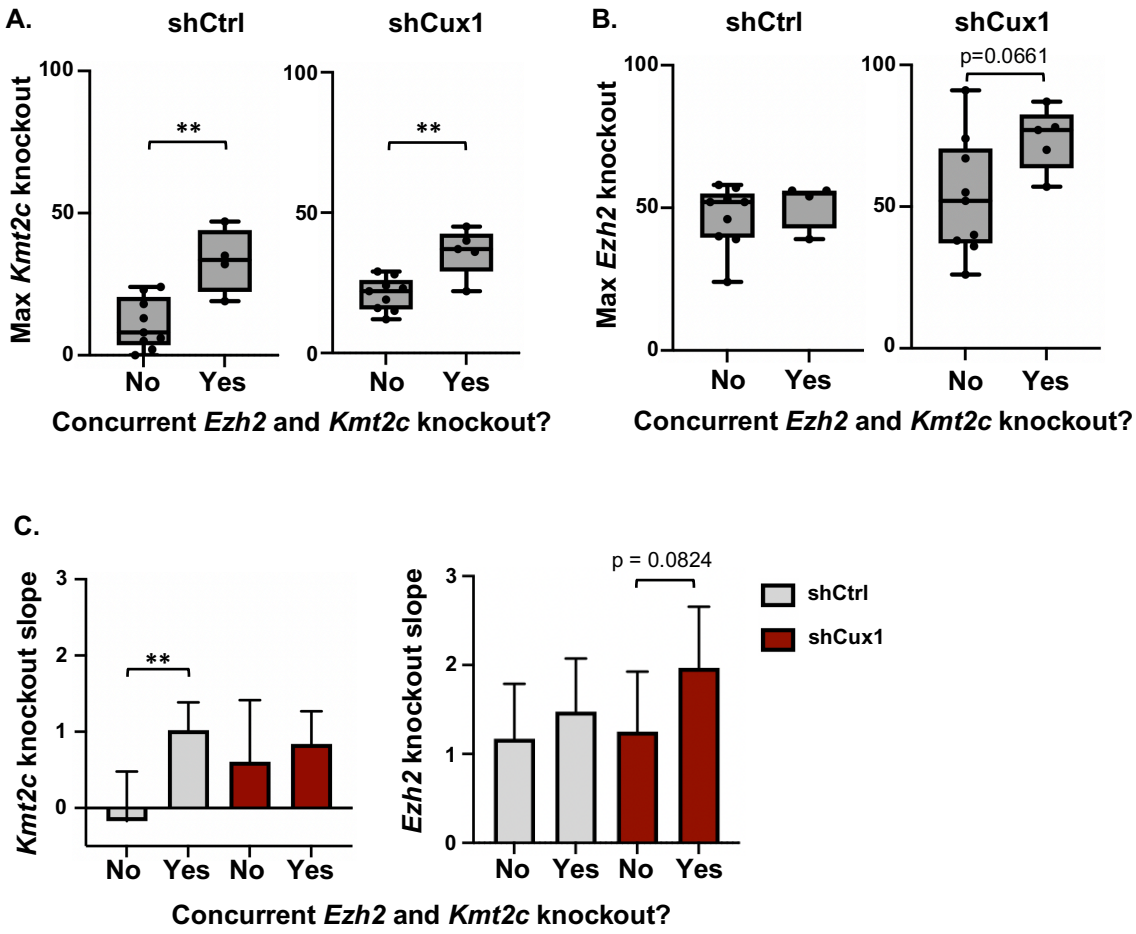

Figure 12: *Kmt2c* loss associates with, but does not enhance, expansion of *Ezh2*-deficient clones.

A) Box plots showing the maximum observed Kmt2c knockout values in peripheral blood of mice receiving multiplex-edited shCtrl (left) and shCux1 (right) HSPCs. Each plot compares the maximum observed knockout between mice with and without significantly correlated Ezh2 and Kmt2c knockout trajectories. Significance from a two-tailed t-test is shown. ** $p < 0.01$. **B)** Same as in A for maximum observed Ezh2 knockout values. A two-tailed t-test was performed. **C)** Bar plots showing the slope of gene knockout trajectories for Kmt2c (left) and Ezh2 (right) as an alternative metric to gauge the effect of combinatorial gene knockout. Each plot compares the slopes between mice with and without concurrent Kmt2c and Ezh2 knockout. Slope was calculated by performing linear regression and fitting a line to knockout values of each individual gene, as shown in **Figure 11A** and **Figure 11B**. Significance from a two-tailed t-test is shown, ** $p < 0.01$.

to promote drug resistance.[108,291] We observed significant increase in viability with even single-copy loss of *CUX1* and *EZH2*, comparable to *PPM1D*-mutant cells **(Figure 13B)**. We observed similar results when treating cells with etoposide **(Figure 13C)**, associated with the development of t-MN.

To test combined loss of *CUX1* and *EZH2* on resistance, we generated a double-heterozygous clone and established dose-response curves with daunorubicin and etoposide **(Figure 14A)**. Indeed, *CUX1*;*EZH2*$^{+/-}$ cells showed a further increased IC_{50} than either single gene deficiency for both daunorubicin and etoposide **(Figure 14B, C)**. Accordingly, *CUX1*;*EZH2*-deficient cells showed decreased apoptosis after culture with daunorubicin compared to *HPRT* cells **(Figure 15A, B)**. Even after washing out daunorubicin and re-seeding equal numbers of viable cells in new culture plates, we observed significantly more growth from *CUX1*;*EZH2*-deficient cells **(Figure 15C)**. Notably, we did not observe further increased daunorubicin IC_{50} values in either a triple heterozygote (*CUX1*;*EZH2*;*KMT2C*) or a quadruple heterozygote **(Figure 15D)**.

We next assessed the effects of *Cux1* and *Ezh2* loss on chemotherapy resistance in primary shCtrl and shCux1 HSPCs. We transfected HSPCs with RNPs targeting *Ezh2* or *Actb* as a control, and plated cells in methylcellulose with daunorubicin or ENU **(Figure 16A)**. We genotyped colonies after two passages and observed a significant increase only in the proportion of shCux1 colonies harboring *Ezh2* knockout; there was no difference in the proportion of colonies with *Actb* indels **(Figure 16B, C)**. As a parallel approach, we plated shCtrl and shCux1 HSPCs in methylcellulose with the FDA-approved EZH2 inhibitor tazemetostat, to uniformly mimic *Ezh2* loss, and daunorubicin **(Figure 17A, B)**. Across two passages, tazemetostat-treated shCux1 cells demonstrated reduced

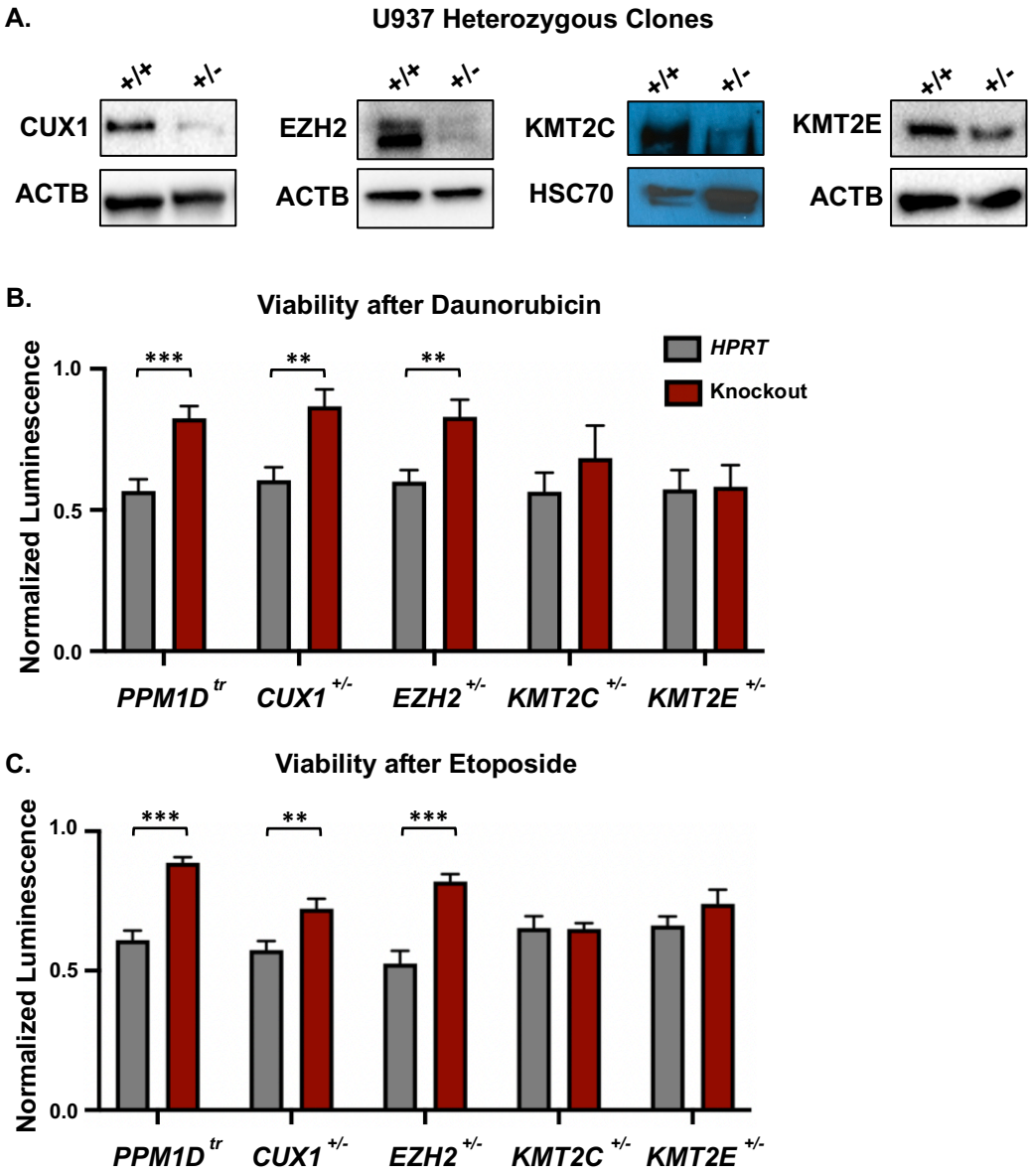

Figure 13: Single copy loss of *CUX1* or *EZH2* promotes chemotherapy resistance.

A) U937 human myeloid leukemia cells were transfected with RNPs targeting CUX1,EZH2, KMT2C, or KMT2E, and clonal lines were derived harboring heterozygous knockout. Western blot validation is shown. Data are representative of n=2 biological replicates. **B)** Heterozygous knockout clones were plated with 3 nM daunorubicin or vehicle for 72 hours, and viability was assessed via CellTiter-Glo luminescence. Cells transfected with RNPs targeting an intron of the HPRT gene were used as a wild-type control. Luminescence normalized to vehicle-treated wells is shown. Two-tailed t-tests were performed to compare normalized luminescence between 7q gene knockout cells and gHPRT control cells. Three biological replicates, each with three technical replicates, were performed. Data (n=3 biological replicates) are plotted as mean + SEM. ** $p < 0.01$,

Figure 13 continued

*** $p < 0.001$. **C)** U937 7q knockout clones were plated with 50 nM etoposide or vehicle for 72 hours, and viability was assessed with CellTiter-Glo luminescence as in B. Three biological replicates, each with three technical replicates, were performed. Data (n=3 biological replicates) are plotted as mean + SEM. ** $p < 0.01$, *** $p < 0.001$.

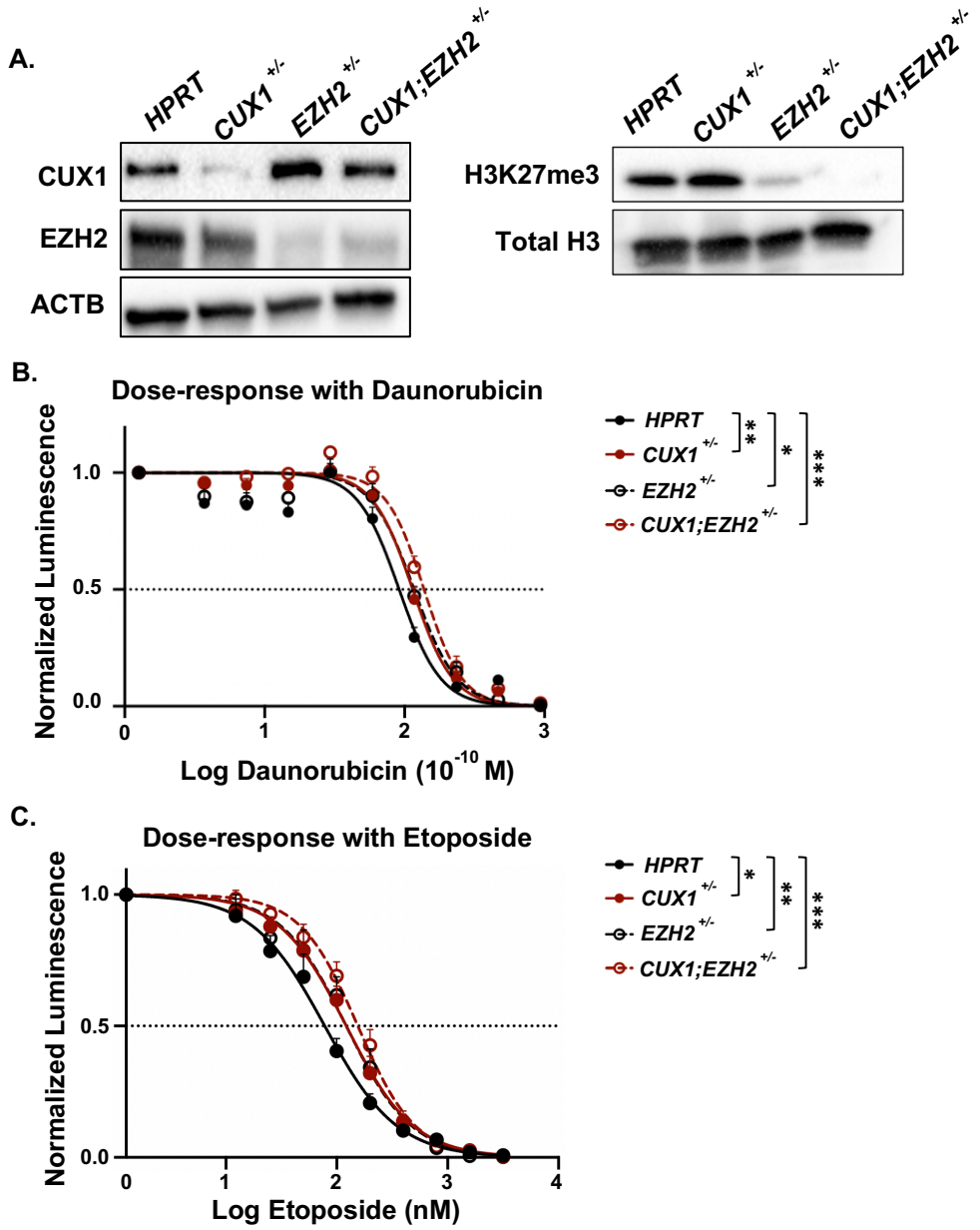

Figure 14: Combined *CUX1*;*EZH2* loss further promotes chemotherapy resistance.

A) Western blot validation of clonal CUX1;EZH2-double heterozygous knockout U937 cells. **B)** *HPRT* control, single heterozygous knockout, and double heterozygous knockout U937 cells were plated with increasing doses of daunorubicin (range 0.375 nM to 96 nM) to generate a dose-response curve. Three biological replicates were performed, each with three technical replicates. Data (n=3 biological replicates) are plotted as mean + SEM, and significance from a two-way ANOVA is shown. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **C)** *HPRT* control, single heterozygous knockout, and double heterozygous knockout U937 cells were plated with increasing doses of etoposide (range 12.5 nM to 3.2 µM) to generate a dose-response curve as in B. Three biological replicates were performed, each with three technical replicates. Data (n=3 biological replicates) are plotted as mean

Figure 14 continued

+ SEM, and significance from a two-way ANOVA is shown. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

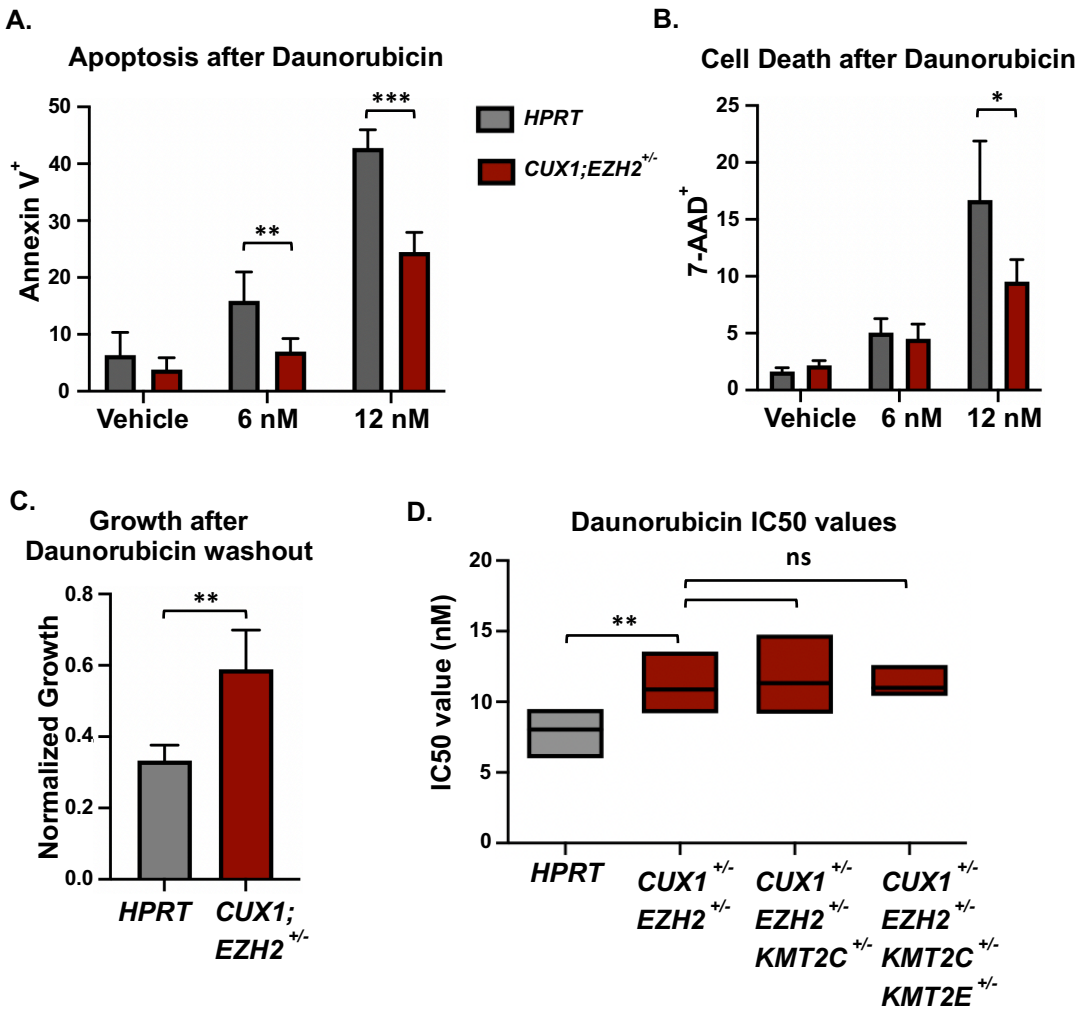

Figure 15: *CUX1*;*EZH2*-deficient cells have reduced apoptosis after daunorubicin exposure.

A) Apoptosis was measured by flow cytometry using Annexin V at two doses of daunorubicin after 72 hours of treatment. Three biological replicates were performed, each with technical duplicates. Data (n=3 biological replicates) are plotted as mean + SD, and significance from a two-tailed t-test is shown. ** $p < 0.01$, *** $p < 0.001$. **B)** Cell death was measured by flow cytometry for the nucleic acid stain 7-AAD in the same cells as in A, after culture with two doses of daunorubicin for 72 hours. Three biological replicates were performed, each with technical duplicates. Data (n=3 biological replicates) are plotted as mean + SD, and significance from a two-tailed t-test is shown. * $p < 0.05$. **C)** After 72-hour treatment with 6 nM daunorubicin, drug was washed out and equal numbers of vehicle-treated and drug-treated *HPRT* and *CUX1*;*EZH2*$^{+/-}$ cells were re-seeded in new culture plates. Cell growth in the drug-treated plates was normalized to that of vehicle-treated plates for both *HPRT* and *CUX1*;*EZH2*$^{+/-}$. Data (n=4 biological replicates) are plotted as mean + SD, and significance from a two-tailed t-test is shown. ** $p < 0.01$. **D)** Box plots of daunorubicin IC_{50} values from wild-type *HPRT*, *CUX1*;*EZH2*$^{+/-}$

Figure 15 continued

double heterozygous knockout, *CUX1*;*EZH2*;*KMT2C*$^{+/-}$ triple heterozygous knockout, and *CUX1*;*EZH2*;*KMT2C*;*KMT2E*$^{+/-}$ quadruple heterozygous knockout cells. Data from n=3 biological replicates are plotted, and significance from two-tailed t-tests is shown. ** $p < 0.01$, ns = not significant.

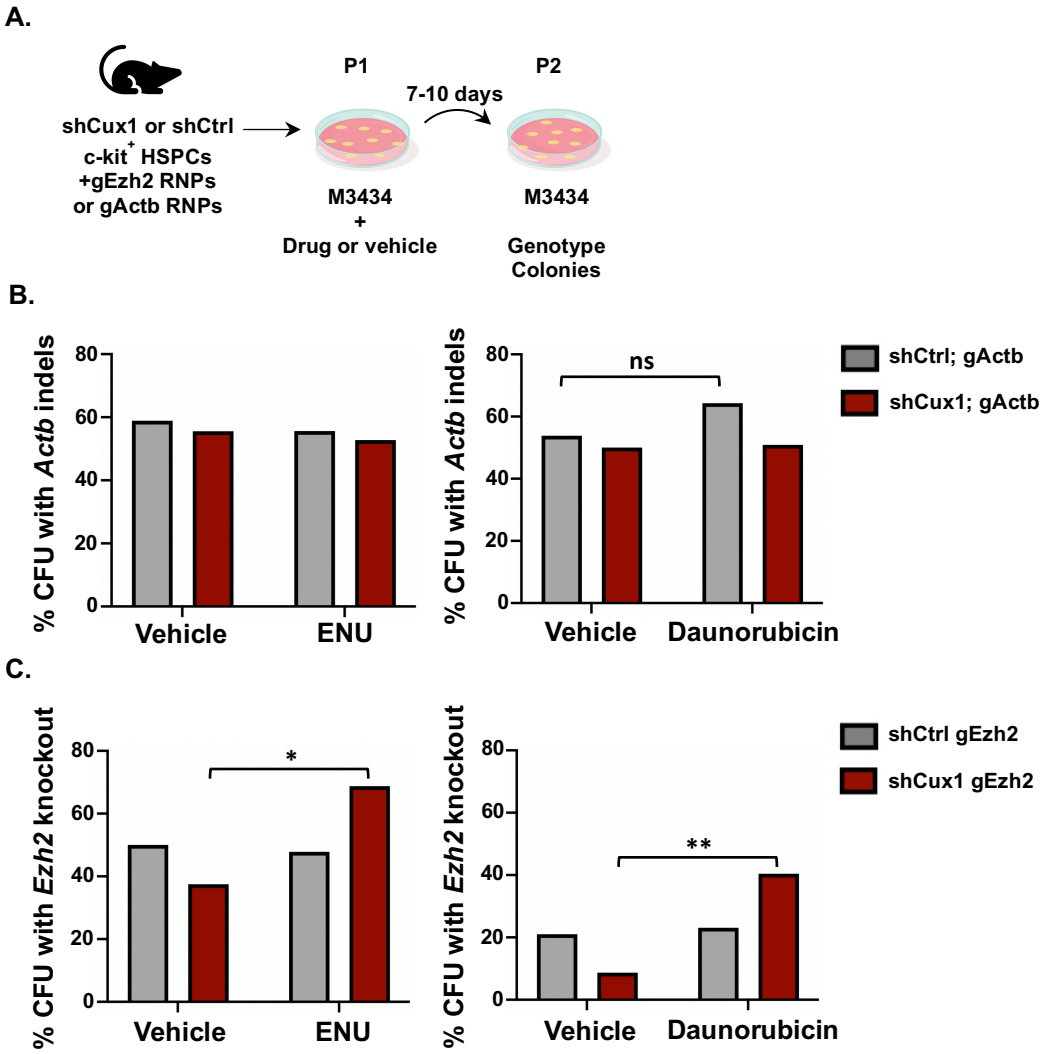

Figure 16: Combined *Cux1*;*Ezh2* deficiency promotes chemotherapy resistance in HSPCs.

A) Experimental schematic for primary cell drug resistance assays. shCtrl and shCux1 HSPCs were transfected with RNPs targeting *Ezh2* or an intron of *Actb* as a control. Cells were then plated in Methocult M3434 methylcellulose with 0.15 mg/mL ENU, 2.5 nM or 5nM daunorubicin, or vehicle. After 7-10 days, culture plates were scraped and colonies were re-seeded to fresh M3434 for a second passage. Individual colonies were genotyped at the end of P2. Results from both doses of daunorubicin are pooled. **B)** The percent of CFU with indels in *Actb* for culture plates with ENU (left) or daunorubicin (right). The number of genotyped CFU are as follows, from left to right: For ENU – 34 shCtrl vehicle, 36 shCux1 vehicle, 36 shCtrl ENU, 36 shCux1 ENU. For daunorubicin – 26 shCtrl vehicle, 28 shCux1 vehicle, 56 shCtrl daunorubicin, 55 shCux1 daunorubicin. Significance from a one-tailed z-score test for proportions is shown. ns not significant. **C)** The percent of CFU with *Ezh2* knockout for culture plates with ENU (left) or daunorubicin (right). The number of genotyped CFU are as follows, from left to right: For ENU – 22 shCtrl vehicle,

Figure 16 continued

16 shCux1 vehicle, 23 shCtrl ENU, 16 shCux1 ENU. For daunorubicin – 19 shCtrl vehicle, 23 shCux1 vehicle, 39 shCtrl daunorubicin, 47 shCux1 daunorubicin. Significance from a one-tailed z-score test for proportions is shown. * $p < 0.05$, ** $p < 0.01$.

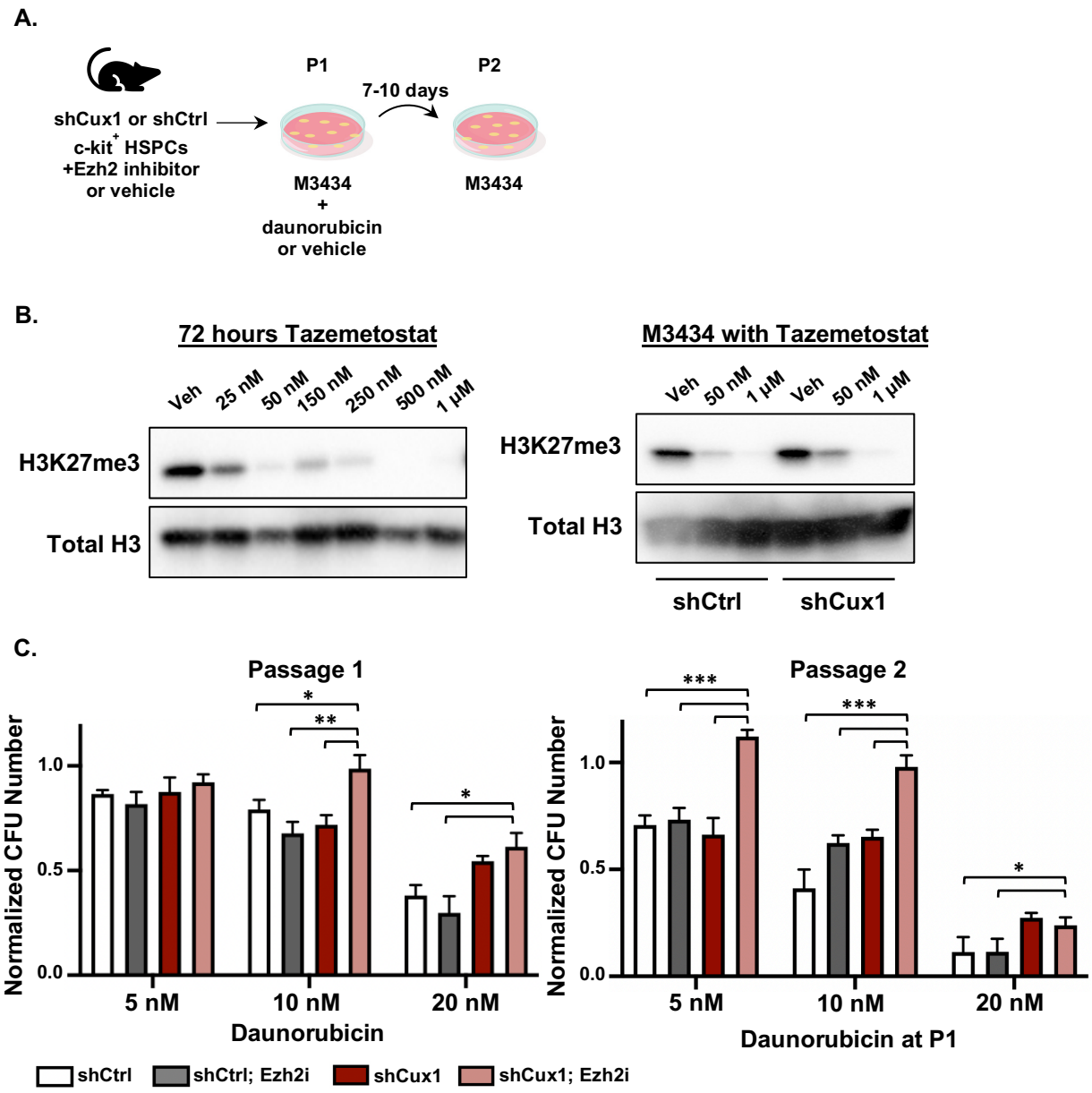

Figure 17: Pharmacologic inhibition of EZH2 promotes daunorubicin resistance on a *Cux1*-deficient background.

A) Experimental schematic for primary cell drug resistance assays with pharmacologic inhibition of EZH2. shCtrl and shCux1 HSPCs were cultured with the EZH2 inhibitor tazemetostat (1 μM) or vehicle for five days, with tazemetostat replenishment on after three days. Cells were then plated in Methocult M3434 methylcellulose ± tazemetostat and daunorubicin. After 7-10 days, colonies were counted and re-seeded to fresh M3434 for a second passage. Colonies were counted again after passage 2. Vehicle control was used for both tazemetostat and daunorubicin. **B)** Western blot showing decreased H3K27me3 in mouse HSPCs after a 72-hour culture with increasing doses of the EZH2 inhibitor tazemetostat (left). Also shown is a second blot showing sustained reduction in

Figure 17 continued

H3K27me3 after passaging HSPCs in MethoCult M3434 supplemented with tazemetostat (right). Data are representative of n=2 independent replicates. **C)** Clonogenic survival of treated shCtrl and shCux1 HSPCs after passage 1 (left) and passage 2 (right). For each genotype, the number of colonies was normalized to vehicle-treated control. Daunorubicin was present in passage 1. Three independent biological replicates with technical duplicates were performed. Data (n=3 biological replicates) are plotted as mean + SEM, and significance from a one-way ANOVA with multiple comparisons is shown. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

sensitivity to multiple doses of daunorubicin as measured by colony numbers **(Figure 17C)**.[296] Further, we recovered a significantly higher number of cells at each passage from tazemetostat-treated shCux1 wells, suggesting increased progenitor proliferation **(Figure 18A)**. We also observed decreased shCux1 sensitivity to daunorubicin with GSK126, a second EZH2 inhibitor **(Figure 18B, C)**. shCux1 colonies treated with EZH2 inhibitors were also significantly larger and denser than treated shCtrl colonies, particularly at later passages **(Figure 18D)**. Thus, combined deficiency in CUX1 and EZH2 promotes resistance to daunorubicin in primary HSPCs to a greater degree than either single gene deficiency alone. These data again demonstrate an interaction between *CUX1* and *EZH2* loss, and indicate these two genes play a driving role in both the founding of disease as well as the treatment resistance associated with -7/del(7q).

Cux1;*Ezh2*-deficient cells have an abrogated DNA damage response after genotoxic stress

To determine whether combined *Cux1* and *Ezh2* deficiency induces expression of a unique set of genes or is transcriptionally additive, we performed RNA-sequencing on shCtrl and shCux1 HSPCs transfected with gEzh2 or intronic gActb RNPs as a control **(Figure 19A)**. Differential gene expression (DE) analysis identified 1,711 DE genes (FDR < 0.05) across all four genotypes **(Table 6)**. Hierarchical clustering analysis resolved clear gene signatures for *Cux1* and *Ezh2* loss, with de-repression of PRC2 target genes in gEzh2 cells **(Figure 19B)**.[169] *Cux1* and *Ezh2* were the top two down-regulated genes, confirming the efficacy of our approach **(Figure 19C)**. *Cux1*;*Ezh2*-deficient cells displayed both gene signatures, indicating an additive transcriptional impact from combined gene

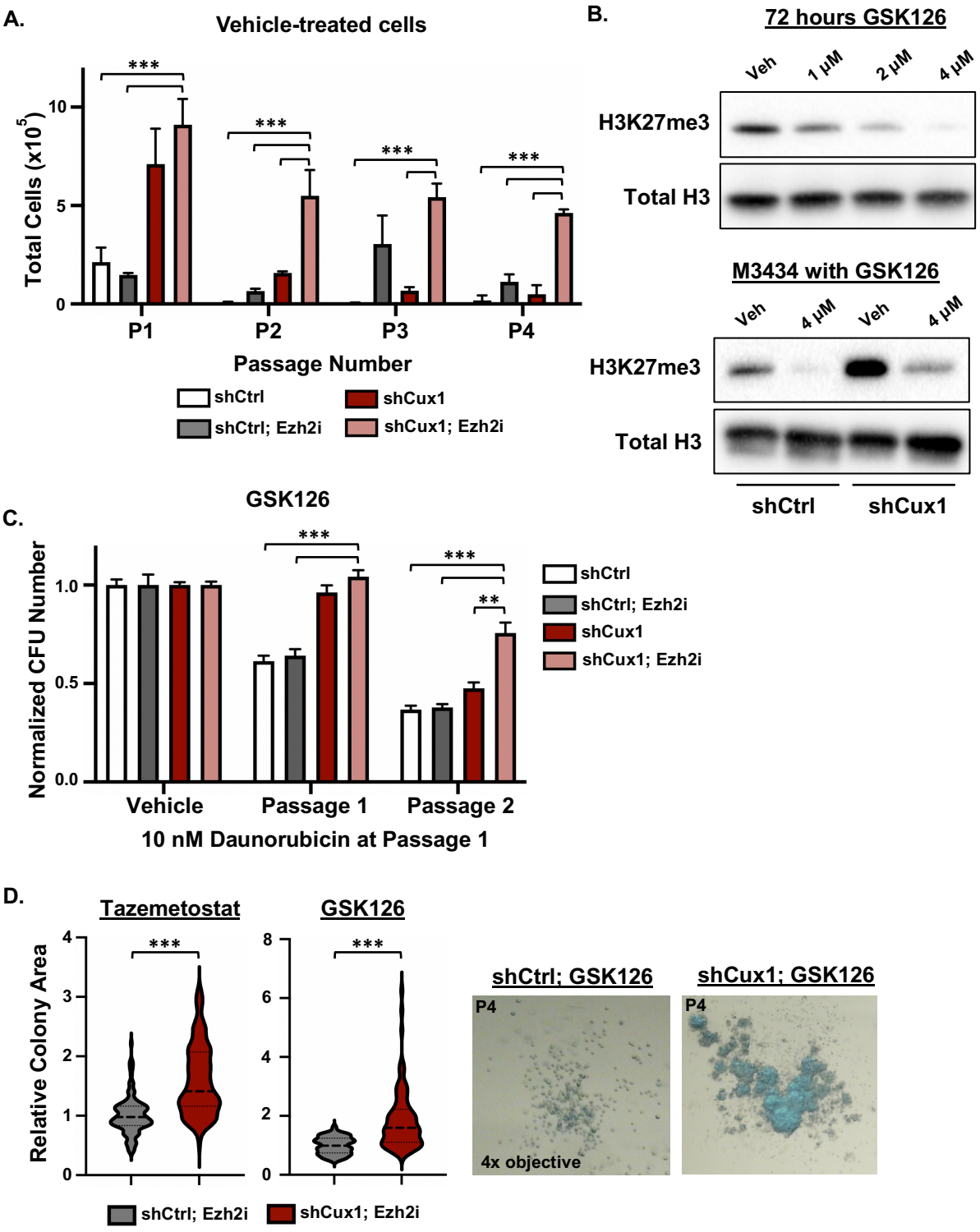

Figure 18: *Cux1*;*Ezh2*-deficient progenitors have higher output and form larger colonies.

A) Total number of cells recovered at each passage from vehicle-treated wells in **Figure**

Figure 18 continued

17C. Three independent biological replicates were performed, and one representative replicate is shown. Data are mean + SD, and significance from one-way ANOVA with multiple comparisons is shown. *** $p < 0.001$. **B)** Western blot showing decreased H3K27me3 in mouse HSPCs after a 72-hour culture with increasing doses of the EZH2 inhibitor GSK126 (top). Also shown is a second blot showing sustained reduction in H3K27me3 after passaging cells in MethoCult M3434 supplemented with GSK126 (bottom). Data are representative of n=2 biological replicates. **C)** Clonogenic survival of shCtrl and shCux1 HSPCs over two passages after treatment with GSK126 and plating in MethoCult M3434 with daunorubicin in the first passage. For each genotype, the number of colonies was normalized to vehicle-treated control. Two independent replicates with technical duplicates were performed. Data (n=2 biological replicates) are plotted as mean + SEM, and significance from one-way ANOVA with multiple comparisons is shown. ** $p < 0.01$, *** $p < 0.001$. **D)** Violin plots of relative colony area in passages 3 and 4 of shRen and shCux1 cells grown in MethoCult M3434 supplemented with either tazemetostat or GSK126. Three independent biological replicates were performed for tazemetostat, two for GSK126. Total number of points are as follows: 132, 117, 44, 97 from left to right. Representative colonies from passage 4 with GSK126 are shown to the right. Significance from a two-tailed t-test is shown. *** $p < 0.001$

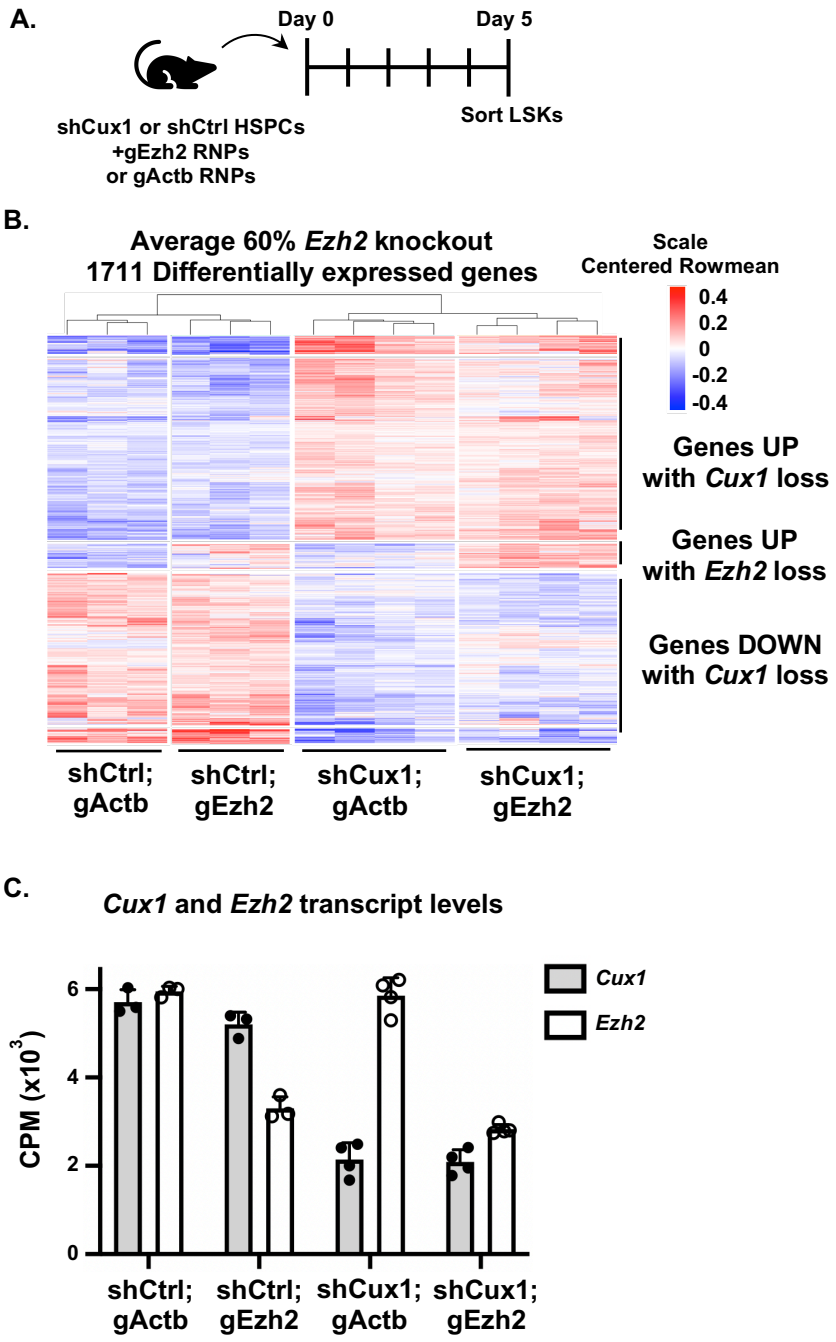

Figure 19: Combined *Cux1* and *Ezh2* loss is transcriptionally additive.

A) Experimental schematic for RNA-sequencing. HSPCs from shCtrl and shCux1 mice were transfected with gEzh2 or gActb RNPs and cultured for five days in SFEM base media with 50 ng/mL murine SCF, FLT-3, and TPO. Lin^- $Sca\text{-}1^+$ $c\text{-}kit^+$ (LSK) cells were sorted directly into TRIzol and RNA was extracted for sequencing. **B)** DE genes were calculated with DESeq2.[282] A heatmap of differentially expressed genes is shown, generated via unsupervised clustering. Clear signatures for *Cux1* loss and *Ez2h* loss are

Figure 19 continued

resolved. Genes upregulated upon *Ezh2* loss include known Polycomb targets such as *Bmpr1a1* (log_2FC=1.43, FDR=6.85e-7) and *Bmi1* (log_2FC=0.72, FDR=4.2e-38). n=3 biological replicates are plotted for shCtrl samples and n=4 for shCux1 samples. Samples had 60% *Ezh2* knockout on average across replicates. Heatmap was generated in R studio with the pheatmap package. **C)** RNA transcript counts per million for *Cux1* and *Ezh2* for RNA-seq libraries in B. *Cux1* and *Ezh2* are the top two downregulated genes, demonstrating effectiveness of the shRNA and RNP-based approaches.

deficiency. Accordingly, we noted *Cux1*;*Ezh2*-deficient cells were more strongly enriched for patient-derived del(7q) gene signatures than either single gene deficiency **(Figure 20A)**.[297] We also noted *Cux1*;*Kmt2c*-deficient cells were not enriched for the -7/del(7q) signature (q-val=0.58) despite clear enrichment for a *Kmt2c*$^{-/-}$ gene signature,[191] further supporting *Cux1* and *Ezh2* as key mediators of the del(7q) phenotype **(Figure 20B)**. *Cux1*;*Ezh2*-deficient cells were enriched for a gene signature resembling *TP53* loss, which may contribute to the selective advantage we observed after genotoxic stress **(Figure 20C)**.

Our lab has also shown that CUX1 promotes epigenetic-driven DNA repair via interaction with the histone methyltransferase EHMT2/G9a.[142] Others have shown PRC2 directly interacts with EHMT2 and also localizes to DNA breaks to deposit H3K27 methylation, an important component of the early DDR.[179–182,298] Therefore, we hypothesized that the selective advantage from *CUX1*;*EZH2* deficiency derives from further disruption to DNA damage recognition and repair. We performed a second RNA-sequencing experiment on shCtrl and shCux1 HSPCs in the presence of ENU to induce DNA damage **(Figure 21A)**. We treated cells with tazemetostat to uniformly suppress EZH2 activity prior to ENU addition. DE gene analysis identified 3,457 genes with FDR < 0.01 **(Table 7)**, and gene set enrichment analyses revealed striking differences between genotypes. The top enriched pathways in shCtrl cells after ENU exposure were related to p53 activation and the DDR **(Figure 21B)**. Pathways related to protein translation were negatively enriched, a known response to DNA damage.[299,300] Both *Cux1*- and *Ezh2*-deficient cells displayed reduced enrichment of these pathways, consistent with their role in the DDR, and cells with combined *Cux1*;*Ezh2* deficiency had an abject failure to

activate DDR pathways **(Figure 21B)**. Blunted stress responses have been observed after inactivation of epigenetic regulators, including *EZH2*,[201] suggesting these compounding DDR defects may enable persistence of *Cux1*;*Ezh2*-deficient cells following exposure to genotoxic stress.

To further explore this mechanism, we assessed the response of *Cux1*;*Ezh2*-deficient HSPCs to acute DNA damage from irradiation (IR). An early DDR step is the phosphorylation of the histone variant H2A.X to form γH2AX at sequences flanking the break point.[301] We previously showed *Cux1*-deficient cells have reduced γH2AX deposition after IR, and indeed *Cux1*;*Ezh2*-deficient cells displayed a larger reduction in γH2AX deposition **(Figure 21C)**. We similarly observed significantly decreased γH2AX deposition in *CUX1*;*EZH2*-null U937 cells **(Figure 21D, E)**. Consequently, *CUX1*;*EZH2*-deficient cells have reduced 53BP1 foci number and intensity 1-hour post-IR **(Figure 22A)**, likely due to decreased EHMT2 and H3K27me2/3, both of which are required for efficient 53BP1 recruitment **(Figure 22B)**.[142,144,302] To assess whether the decrease in DNA damage recognition results in persistent, unrepaired damage, we performed neutral comet assays on irradiated shCtrl and shCux1 HSPCs. In contrast to control cells, which resolve DNA damage within 24 hours post-IR, cells deficient in *Cux1* or *Ezh2* displayed significantly elevated levels of DNA breaks and *Cux1*;*Ezh2*-deficient HSPCs had even higher levels at 24- and 48-hours post-IR **(Figure 22C)**. Even a low-dose pulse of 0.5 Gy resulted in elevated DNA breaks in *Cux1*;*Ezh2*-deficient cells. Collectively, our data indicate CUX1 and EZH2 loss converges on DDR deregulation, which may enable mutant cells to persist through genotoxic stress and subsequently expand.

Discussion

Dissecting the pathogenesis of -7/del(7q) and other recurrent aneuploid events in cancer remains a challenge. Here we develop a model of del(7q) clonal hematopoiesis and drug resistance, taking an orthogonal approach to identify candidate 7q genes most likely to interact with *CUX1* loss by leveraging publicly available CRISPR screen data and existing literature. Our multiplex knockout approach achieves population-level editing in primary HSCs, without viral plasmid delivery or antibiotic selection. Further, this approach generates a knockout population in under one week and circumvents the lack of chromosomal synteny between humans and mice. Using this approach, we identify *CUX1* and *EZH2* as critical mediators of drug resistance and clonal expansion that converge on deregulation of the DNA damage response. This work supports the concept of 7q as a contiguous gene syndrome region, similar to those identified on 5q and 8p.[48,53–55] Further, we link the drug resistance and poor prognosis associated with -7/del(7q) to specific 7q genes.

KMT2C mutations have been identified in primary AML and relapsed pediatric AML,[192] and *Kmt2c*$^{+/-}$ HSCs expand after exposure to alkylating agents.[191] In our model, *Kmt2c* loss associates with *Cux1* deficiency but the effects on expansion are highly variable **(Figure 4F)**. Though *Kmt2c* loss trends toward promoting expansion in the setting of combined *Cux1*;*Ezh2* deficiency **(Figure 12B, C)**, suggesting a minor contribution, our data demonstrate a stronger interaction between *Cux1* and *Ezh2* loss. We also note that loss of multiple 7q genes is insufficient to fully rescue the engraftment defects from *Kmt2e* loss, though there were notable examples in which *Kmt2e* was co-mutated with *Ezh2* **(Figure 8A, C)**.[197,200] As these four 7q genes have been implicated in

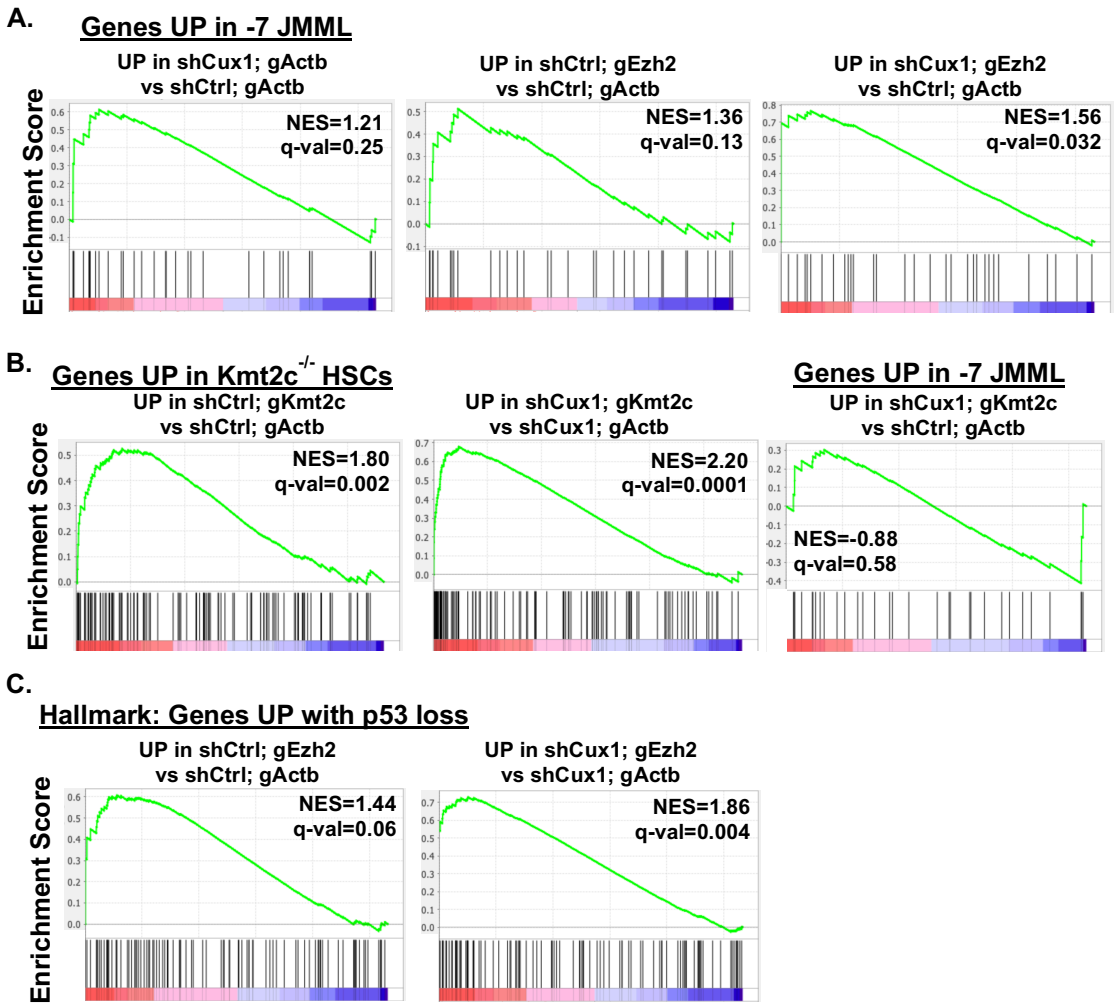

Figure 20: *Cux1*;*Ezh2*-deficient cells are enriched for gene signatures of patient-derived del(7q) disease and *TP53* loss.

A) Gene set enrichment analysis (GSEA)[283] plots showing enrichment for *Cux1* and *Ezh2* single deficiencies (top, middle) and combined *Cux1;Ezh2* deficiency (bottom) using a gene signature derived from juvenile myelomonocytic leukemia (JMML) patients with -7/del(7q).[297] Cells with combined *Cux1;Ezh2* deficiency have the highest enrichment score and achieve statistically significant enrichment (q-value = 0.037). **B)** RNA-sequencing was also performed on shCtrl and shCux1 HSPCs transfected with RNPs targeting *Kmt2c*, with an average 40% *Kmt2c* knockout. GSEA plots show enrichment for a *Kmt2c*-null HSC gene signature (left, middle), confirming efficacy of gene knockout.[191] *Cux1*;*Kmt2c*-deficient cells do not display enrichment for -7/del(7q) gene signatures (right). **C)** GSEA plots for *Ezh2* deficiency on the shCtrl background (top) and shCux1 background (bottom) showing enrichment for the P53_DN.V1_UP gene signature,[283] representing genes upregulated in NCI-60 panel of cell lines with mutated *TP53*. Combined *Cux1*;*Ezh2* deficiency has stronger enrichment and statistical significance (q-value = 0.004).

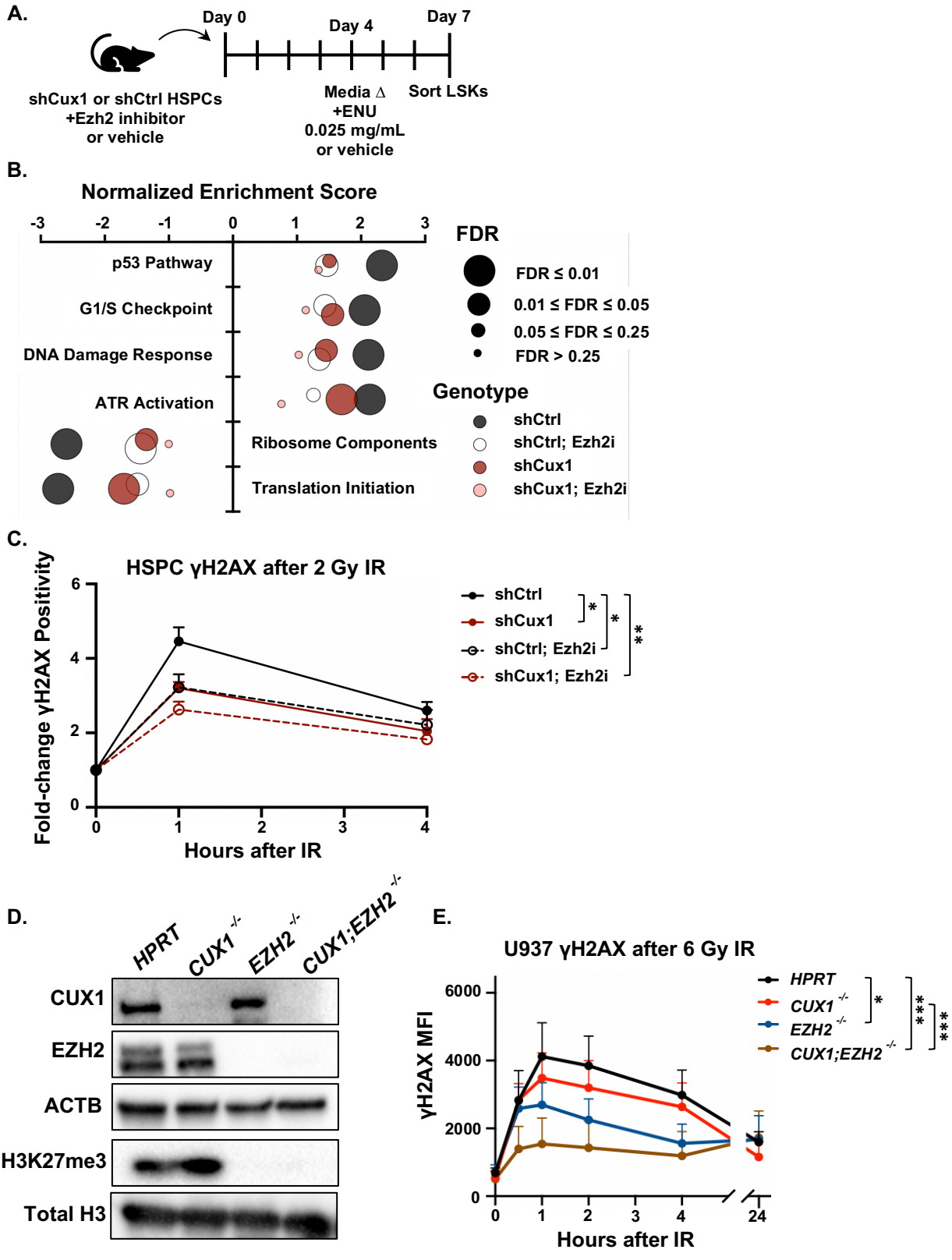

Figure 21: *Cux1*;*Ezh2*-deficient cells have a diminished DNA damage response.

Figure 21 continued

A) Experimental scheme for RNA-sequencing with ENU. shCtrl and shCux1 HSPCs were cultured with 1 µM tazemetostat or vehicle for four days to mimic *Ezh2* loss. Media was replaced with fresh tazemetostat and 0.025 mg/mL ENU or vehicle to induce DNA damage. After 72 hours of ENU treatment, Lin^- $Sca\text{-}1^+$ $c\text{-}kit^+$ (LSK) cells were sorted diretly into TRIzol. RNA was extracted and prepared for sequencing. A wild-type shCtrl group without ENU treatment was included as a control. Three biological replicates were sequenced. **B)** Differential gene expression analysis was conducted with DeSeq2 and gene rank lists were created for gene set enrichment analysis.[282,283] All genotypes were compared to shCtrl cells without ENU exposure. A multivariate plot showing the top positively and negatively enriched gene sets in shCtrl cells after ENU exposure is shown. Bubble color indicates genotype, and size denotes false discovery rate (FDR). **C)** shCtrl and shCux1 HSPCs were cultured for five days with 1 µM tazemetostat or vehicle to mimic *Ezh2* loss, with tazemetostat replenishment after three days. Cells were then irradiated to induce DNA damage (2 Gy). γH2AX deposition was followed over four hours by intracellular flow cytometry, and the fold-change of cells staining positive for γH2AX is plotted. Data (n=4 biological replicates) are plotted as mean + SEM, and significance from two-way ANOVA is shown. * $p < 0.05$, ** $p < 0.01$. **D)** Western blot validation of clonal *CUX1*$^{-/-}$, *EZH2*$^{-/-}$, and *CUX1*;*EZH2*-double homozygous knockout U937 cells. **E)** Clonal homozygous knockout U937 cells were irradiated to induce DNA damage (6 Gy), and γH2AX deposition was followed over 24 hours by intracellular flow cytometry. The total γH2AX mean fluorescence intensity (MFI) is shown; data (n=4 biological replicates) are plotted as mean + SEM, and significance from two-way ANOVA is shown. * $p < 0.05$, *** $p < 0.001$.

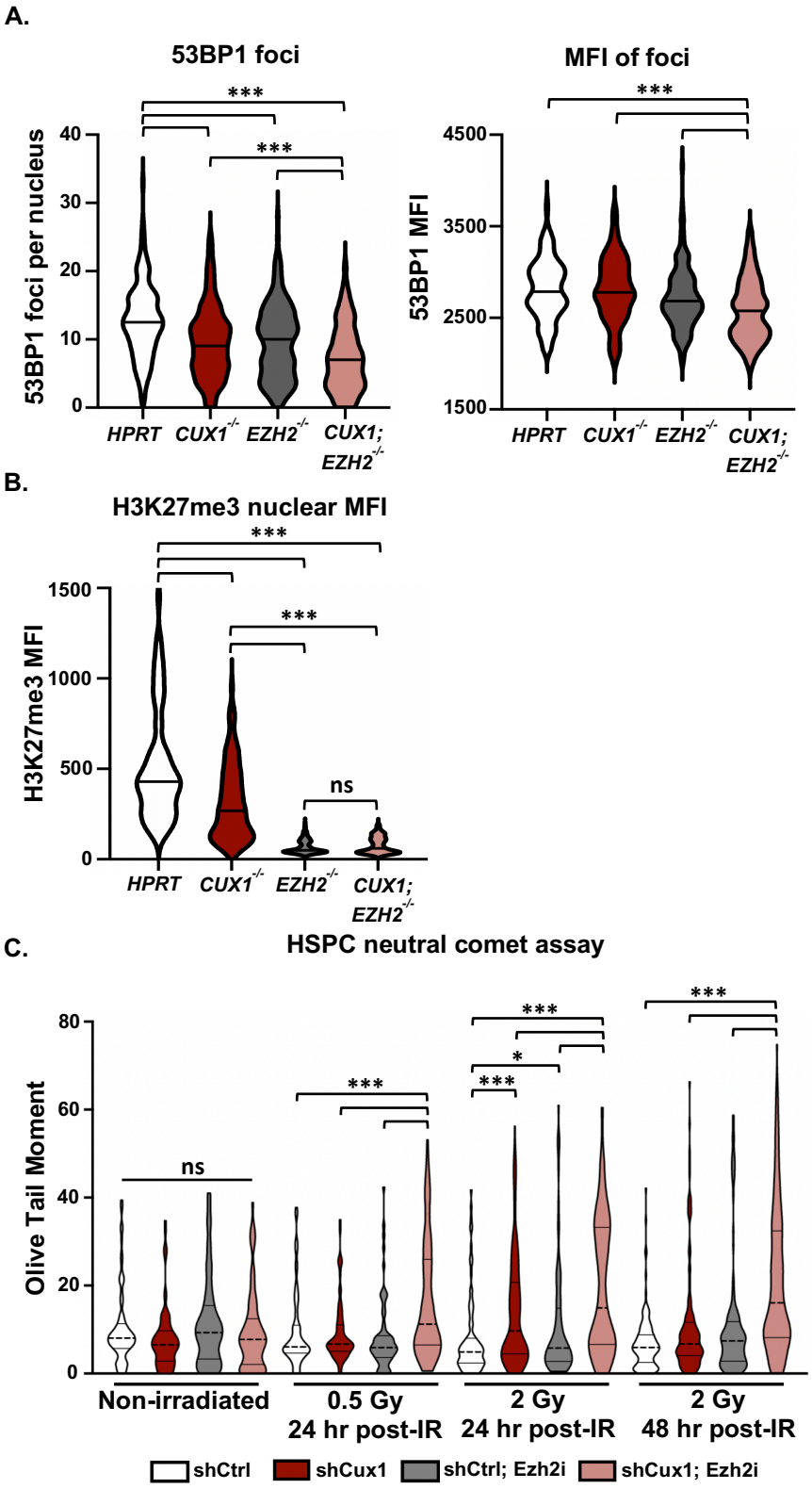

Figure 22: *CUX1*;*EZH2*-deficient cells fail to retain 53BP1 at γH2AX foci and have decreased DNA repair efficiency.

Figure 22 continued

A) U937 cells were irradiated (6 Gy) and stained for 53BP1 1-hour post-IR. The number of 53BP1 foci per nucleus (left) and the mean fluorescence intensity (MFI, right) of 53BP1 foci are plotted. Three independent biological replicates were performed with at 20-70 nuclei analyzed per replicate, and significance from a one-way ANOVA with multiple corrections is shown. ***p < 0.001. **B)** U937 cells were irradiated (6 Gy) and stained for H3K27me3 1-hour post-IR. The total nuclear MFI of H3K27me3 is plotted. Two biological replicates with 35-60 nuclei analyzed per replicate, and significance from a one-way ANOVA with multiple comparisons is shown. *** p < 0.001, ns = not significant. **C)** shCtrl and shCux1 HSPCs were cultured for five days with 1 μM tazemetostat or vehicle to mimic *Ezh2* loss, with tazemetostat replenishment after three days. Cells were then irradiated with 0.5 Gy or 2 Gy and neutral comet assays were performed. Non-irradiated cells were used as a control. Data are plotted as violin plots with quartiles, and three independent biological replicates were performed. Total number of points are as follows: 110, 129, 79, 84 (No IR); 86, 65, 72, 103 (0.5 Gy); 100, 84, 128, 133 (2 Gy, 24 hours); 120, 117, 118 174 (2 Gy, 48 hours) from left to right. Significance from Mann-Whitney U tests is shown. * p < 0.05, *** p < 0.001, ns = not significant.

DNA damage repair, there may be a degree of functional redundancy between deletion phenotypes with *Cux1* and *Ezh2* loss possessing the highest penetrance.

Patient-derived monosomy 7 cells display downregulated DNA damage checkpoint and apoptosis genes,[303] and indeed our RNA-sequencing revealed a profound failure of *Cux1*;*Ezh2*-deficient cells to activate the DDR following ENU exposure **(Figure 21B)**. Both CUX1 and EZH2 localize to sites of DNA damage,[142,179,181] and EZH2 inhibition also results in downregulation of a set of DDR genes via decreased methylation of activating pioneer transcription factors.[304] Thus, the compounding deregulation of early stages of DNA damage recognition we observed may be multifaceted, with both transcriptional and non-transcriptional origins **(Figure 21B, C, 22A, B)**.

Clonal hematopoiesis variants in DDR genes are selected for following genotoxic stress, likely due to increased survival of mutant cells.[107,108,291] Variants in epigenetic regulators are also positively selected for, likely via increased self-renewal as well as blunted responses to external stressors.[201,305] Our work positions *CUX1* and *EZH2*, which are themselves mutated in clonal hematopoiesis,[100,111,122–124] at the intersection of these pathways and offers a new perspective on the functional impact of -7/del(7q). Whether these pathways interact with other elements of -7/del(7q) biology, including altered splicing from *LUC7L2* deletions[209,210] and cytokine hypersensitivity from *SAMD9/SAMD9L* loss,[205] remains to be discovered. Future studies characterizing -7/del(7q) should be performed with combined *CUX1* and *EZH2* deficiency as a foundation.

Table 5: Candidate 7q Genes Promoting Chemotherapy Resistance When Deleted

Gene name	Resistance to drug (s)	Chr. 7 Location	CDR?	Expression (log CPM) in HSCs	Gene Rank (Expression)
ACTB	Doxorubicin Etoposide	7p22.1		12.43464999	1
KMT2C	Cisplatin	7q36.1	Yes	8.1613083	2
KMT2E	Cisplatin	7q22.3	Yes	8.02147745	3
TNRC18	Doxorubicin	7p22.1		7.8969443	4
GARS	Etoposide	7p14.3		7.69836195	5
TSC22D4	Doxorubicin	7q22.1	Yes	7.69362045	6
PDIA4	Cisplatin	7q36.1	Yes	7.6119871	7
MCM7	Etoposide	7q22.1	Yes	7.581156	8
LAT2	Cisplatin	7q11.23		7.55329255	9
FSCN1	Cisplatin	7p22.1		7.52386375	10
MDH2	Etoposide	7q11.23		7.4447735	11
AHR	Etoposide	7p21.1		7.414710915	12
CDK6	Doxorubicin Etoposide	7q21.2		7.330048215	13
SEPT7	Doxorubicin	7p14.2		7.3109188	14
TRRAP	Etoposide	7q22.1	Yes	7.30041885	15
DBNL	Cisplatin	7p13		7.0956998	16
TNPO3	Doxorubicin	7q32.1		7.01815195	17
LUC7L2	Etoposide	7q34	Yes	6.95291305	18
CLIP2	Cisplatin	7q11.23		6.94981756	19
BAZ1B	Cisplatin	7q11.23		6.948382395	20
TNS3	Etoposide	7p12.3		6.9357258	21
INTS1	Cisplatin Etoposide	7p22.3		6.88998145	22
CAPZA2	Etoposide	7q31.2		6.85388647	23
CASP2	Cisplatin	7q34	Yes	6.792406455	24
VPS41	Cisplatin	7p14.1		6.76296405	25
KCNH2	Cisplatin	7q36. 1	Yes	6.75421315	26

Table 5 continued

RAC1	Etoposide	7p22.1		6.74926465	27
NUP205	Etoposide	7q33		6.67084445	28
PMPCB	Etoposide	7q22.1	Yes	6.64917425	29
GATAD1	Etoposide	7q21.2		6.540701	30
KBTBD2	Etoposide	7p14.3		6.4849094	31
NCAPG2	Etoposide	7q36.3		6.407732	32
PLOD3	Doxorubicin	7q22.1	Yes	6.3222285	33
PSMA2	Doxorubicin	7p14.1		6.3102175	34
SUN1	Cisplatin	7q22.1		6.2747199	35
MEPCE	Etoposide	7q221	Yes	6.128333	36
C7orf50	Cisplatin	7p22.3		6.10481603	37
EZH2	Doxorubicin	7q36.1	Yes	6.0784322	38
POR	Doxorubicin	7q11.23		6.05392465	39
AP5Z1	Cisplatin	7p22.1		6.053633205	40
NUDCD3	Etoposide	7p13		6.03272645	41
SNX13	Cisplatin	7p21.1		6.0171493	42
COPS6	Cisplatin Doxorubicin Etoposide	7q22.1	Yes	5.94342525	43
POM121C	Etoposide	7q11.23		5.8728336	44
BZW2	Etoposide	7p21.1		5.857022625	45
SMARCD3	Cisplatin	7q36.1	Yes	5.850817	46
BUD31	Etoposide	7q22.1	Yes	5.818738125	47
SEC61G	Etoposide	7p11.2		5.81838385	48
ZNF12	Doxorubicin	7p22.1		5.78878275	49
POT1	Etoposide	7q31.33		5.76346205	50
ANLN	Doxorubicin	7p14.2		5.73109011	51
YKT6	Etoposide	7p13		5.63096425	52
TBRG4	Etoposide	7p13		5.5853183	53
ZNF467	Doxorubicin	7q36.1	Yes	5.57616885	54
FIS1	Cisplatin	7q22.1	Yes	5.56162995	55

Table 5 continued

TRIM56	Cisplatin	7q22.1	Yes	5.5488226	56
PNPLA8	Cisplatin	7q31.1		5.51440535	57
RFC2	Etoposide	7q11.23		5.46608145	58
CNOT4	Cisplatin	7q33		5.45210147	59
SBDS	Etoposide	7q11.21		5.37263265	60
SNX8	Cisplatin	7p22.3		5.31122035	61
SMURF1	Etoposide	7q22.1	Yes	5.2987093	62
PRKAG2	Doxorubicin	7q36.1	Yes	5.28718275	63
TAF6	Etoposide	7q22.1	Yes	5.2529728	64
MICALL2	Cisplatin	7p22.3		5.239486	65
ZNF394	Cisplatin	7q22.1	Yes	5.18028245	66
DDX56	Doxorubicin Etoposide	7p13		5.17699915	67
ZSCAN25	Cisplatin	7q22.1	Yes	5.0704956	68
MOSPD3	Cisplatin	7q22.1	Yes	5.0585665	69
SLC12A9	Cisplatin	7q22.1	Yes	4.93533145	70
CPSF4	Etoposide	7q22.1	Yes	4.90868355	71
RSBN1L	Etoposide	7q11.23		4.85171275	72
CYCS	Etoposide	7p15.3		4.83860385	73
ZNF680	Etoposide	7q11.21		4.77863885	74
ZNF746	Cisplatin	7q36.1	Yes	4.7473932	75
STK17A	Cisplatin Etoposide	7p13		4.6358003	76
GPR146	Cisplatin	7p22.3		4.6356683	77
MPLKIP	Cisplatin	7p14.1		4.619847	78
RHEB	Etoposide	7q36.1	Yes	4.612742	79
URGCP	Cisplatin	7p13		4.599517	80
CLDN15	Cisplatin	7q22.1	Yes	4.575562075	81
NUPL2	Etoposide	7p15.3		4.53726245	82
TMEM60	Etoposide	7q11.23		4.50227845	83
COA1	Etoposide	7p13		4.4827737	84

Table 5 continued

BRAF	Etoposide	7q34	Yes	4.45897065	85
MAFK	Cisplatin	7p22.3		4.454891	86
POLR2J	Cisplatin	7q22.1	Yes	4.44925215	87
RINT1	Doxorubicin	7q22.3	Yes	4.3722493	88
TECPR1	Cisplatin	7q21.3		4.33362355	89
NOM1	Etoposide	7q36.3		4.267146	90
HUS1	Etoposide	7p12.3		4.10332005	91
POP7	Cisplatin	7q22.1	Yes	4.0140661	92
PMS2	Cisplatin Doxorubicin	7p22.1		3.9973874	93
RBM48	Cisplatin Etoposide	7q21.2		3.92957895	94
PSMG3	Etoposide	7p22.3		3.69165685	95
GIMAP1	Etoposide	7q36.1	Yes	3.6611238	96
MALSU1	Cisplatin Etoposide	7p15.3		3.6219315	97
FBXL18	Cisplatin	7p22.1		3.58264835	98
MRPS17	Etoposide	7p11.2		3.3083495	99
ITGB8	Cisplatin	7p21.1		3.2735072	100
STAG3	Cisplatin	7q22.1	Yes	3.238211	101
TP53TG1	Doxorubicin	7q21.12		3.2229157	102
ABHD11	Cisplatin Etoposide	7q11.23		3.217253005	103
PSPH	Doxorubicin	7p11.2		3.217253005	104
GPR85	Cisplatin	7q31.1		2.96008415	105
CYP3A5	Cisplatin	7q22.1		2.4439418	106
INHBA	Cisplatin	7p14.1		2.38370695	107
HGF	Cisplatin	7q21.11		2.11720705	108
HOXA5	Etoposide	7p15.2		2.102264	109
GPC2	Cisplatin	7q22.1	Yes	1.97561535	110
RNF32	Etoposide	7q36.3		1.8406499	111

Table 5 continued

LRRC4	Cisplatin	7q32.1		1.7577302	112
PVRIG	Cisplatin	7q22.1	Yes	1.5035818	113

Table 6: RNA-Sequencing without ENU Reads Table

RNA-seq on Murine LSKs	PF Clusters Forward	PF Clusters Reverse	Mapped Reads
shCtrl gActb Rep 1	17,472,449	15,744,832	27,095,570
shCtrl gActb Rep 2	18,062,660	14,624,270	26,034,111
shCtrl gActb Rep 3	16,289,684	17,379,485	27,765,213
shCtrl gActb Rep 4	16,257,076	15,635,605	25,989,126
shCtrl gEzh2 Rep 1	15,267,096	17,890,252	20,691,766
shCtrl gEzh2 Rep 2	15,516,756	17,373,720	26,679,306
shCtrl gEzh2 Rep 3	15,260,316	16,613,091	26,073,646
shCtrl gEzh2 Rep 4	17,605,724	13,985,831	26,159,787
shCux1 gActb Rep 1	16,470,206	17,184,318	26,797,310
shCux1 gActb Rep 2	17,825,947	12,888,119	24,875,557
shCux1 gActb Rep 3	16,027,222	18,019,491	27,686,103
shCux1 gActb Rep 4	16,212,155	17,847,380	28,427,404
shCux1 gEzh2 Rep 1	15,914,111	17,885,653	27,045,659
shCux1 gEzh2 Rep 2	17,907,047	13,886,926	25,937,240
shCux1 gEzh2 Rep 3	15,192,098	18,331,917	27,713,406
shCux1 gEzh2 Rep 4	15,057,931	18,897,828	28,558,782
shCtrl gKmt2c Rep 1	17,045,805	15,855,620	27,502,202
shCtrl gKmt2c Rep 2	16,648,206	15,663,759	25,935,285
shCtrl gKmt2c Rep 3	18,497,129	12,496,579	25,381,402
shCtrl gKmt2c Rep 4	17,208,144	14,730,294	26,369,917
shCux1 gKmt2c Rep 1	17,077,605	15,340,216	25,935,827
shCux1 gKmt2c Rep 2	18,399,439	12,977,172	25,507,280
shCux1 gKmt2c Rep 3	16,513,825	16,651,193	27,400,238
shCux1 gKmt2c Rep 4	15,617,570	17,486,948	27,560,270

Table 7: RNA-Sequencing with ENU Reads Table

RNA-seq on Murine LSKs	PF Clusters Forward	PF Clusters Reverse	Mapped Reads
shCtrl No ENU Rep 1	16,692,265	16,872,730	27,373,104
shCtrl No ENU Rep 2	17,508,275	17,730,287	29,285,188
shCtrl No ENU Rep 3	16,142,874	16,363,872	26,846,502
shCtrl ENU Rep 1	17,366,630	17,544,655	28,722,146
shCtrl ENU Rep 2	17,232,339	17,494,213	28,459,583
shCtrl ENU Rep 3	15,594,110	15,754,377	25,431,050
shCtrl; Ezh2i ENU Rep 1	16,817,955	17,008,116	27,835,519
shCtrl Ezh2i ENU Rep 2	16,224,076	16,442,462	26,727,858
shCtrl; Ezh2i ENU Rep 3	14,882,095	15,067,996	23,888,661
shCux1 ENU Rep 1	12,531,618	12,740,901	20,527,902
shCux1 ENU Rep 2	16,306,797	16,420,167	26,989,532
shCux1 ENU Rep 3	18,370,468	18,460,912	29,274,971
shCux1; Ezh2i ENU Rep 1	14,202,050	14,430,283	23,401,007
shCux1; Ezh2i ENU Rep 2	16,866,354	17,015,619	28,209,602
shCux1; Ezh2i ENU Rep 3	17,523,695	17,670,791	28,507,889

DISCUSSION

Overview

In this study, we leveraged CRISPR-Cas9 gene editing to explore a novel method of modeling aneuploidy in the context of chromosome 7q deletions in high-risk myeloid neoplasms, targeting four candidate TSGs simultaneously in hematopoietic stem cells (*Cux1*, *Ezh2*, *Kmt2c*, *Kmt22*). We find that combined loss of the 7q genes *Cux1* and *Ezh2* promotes clonal myeloid expansion to a greater degree than either single gene loss alone, supporting the notion of 7q as a contiguous gene syndrome region harboring multiple TSGs. Intriguingly, loss of *Kmt2c* and *Kmt2e* does not have a significant impact on expansion even when combined with *Cux1*;*Ezh2* deficiency, indicating *Cux1* and *Ezh2* loss are the primary drivers of expansion within this tested gene set.

Chromosome 7 deletions are highly associated with t-MNs that arise after prior exposure to chemotherapy or radiation, and associated with poor prognosis. Therefore, we also examined the role of 7q gene loss in drug resistance, investigating alkylating agents and topoisomerase inhibitors associated with t-MN development, and anthracyclines which are a frontline therapeutic for t-MN patients. Across all tested compounds, and using both murine and human cell systems, we observed a significantly higher degree of resistance in cells with combined *Cux1*;*Ezh2* deficiency compared to single gene deficiencies and wild-type cells. Similar to our *in vivo* experiments testing clonal expansion, we noted that loss of *KMT2C* and *KMT2E* did not significantly alter resistance profiles in human AML cells, again demonstrating combined *CUX1* and *EZH2* deficiency is the driver of these phenotypes. Collectively, these observational experiments link specific 7q genes to a drug resistance phenotype. Moreover, our results suggest that

the initial foundation of a t-MN and the subsequent poor response to chemotherapies are manifestations of the same molecular mechanism, as we observed similar findings in the presence of multiple drug classes as well as irradiation.

Our mechanistic studies revealed converging roles for CUX1 and EZH2 in regulation of the DNA damage response that are conserved across murine and human cell systems. The DDR is broadly composed of a recognition phase, in which DNA double strand breaks are marked with the γH2AX histone modification, and a resolution phase, in which repair factors such as 53BP1 and ATM accumulate at γH2AX foci and regulate break repair.[301] It has been shown that both *Cux1*- and *EZH2*-deficient cells display reduced γH2AX deposition after irradiation,[142,182] and in our studies we observed an even greater reduction in γH2AX deposition in cells with combined *Cux1*;*Ezh2* loss, indicating further defects with break recognition. Consequently, *Cux1*;*Ezh2*-deficient cells fail to transcriptionally activate DDR pathways after genotoxic stress, in striking contrast to wild-type cells. The compounding defects to DNA break recognition subsequently impact DNA repair. Post-irradiation, we observed reduced 53BP1 foci number and intensity in *CUX1*;*EZH2*-deficient human leukemia cells, and increased residual DNA damage in *Cux1*;*Ezh2*-deficient hematopoietic progenitors. Our results add to a growing body of literature reporting non-transcriptional roles for CUX1 and EZH2 in DDR regulation.

This study also generated a dataset of 113 chromosome 7 gene candidates that promote drug resistance when lost **(Table 5)**. This list can serve as a resource for future studies, and highlights the utility of mining through the wealth of publicly available data from published genome-wide CRISPR-Cas9 screens. This *in silico* approach also serves

as a useful orthogonal method to identify candidate genes in addition to a traditional literature search.

Combined *CUX1*;*EZH2* loss confers a competitive advantage after genotoxic stress

As the life expectancy of the human population increases, so too will the prevalence of somatic mosaicism, CHIP, and cancer.[24,99] Age-related mutation acquisition is largely unavoidable; thus, it is paramount that efforts be made to understand the selective forces that drive expansion of mutations from a single ancestral cell to a lethal malignancy. These processes have implications beyond cancer incidence as well, as clonal hematopoiesis is also associated with increased mortality from non-hematologic causes, particularly atherosclerotic cardiovascular disease.[306–308]

Many individuals with CHIP display stable VAFs and do not progress to frank disease, suggesting external factors are a major selective force directing expansion.[309] Our *in vivo* and cell culture models demonstrate strong selection for *CUX1* and *EZH2* deficiency following exposure to genotoxic stress, and our mechanistic studies build on prior studies positioning CUX1 and EZH2 in the local DNA damage response.[142,182] Mutations in both *CUX1* and *EZH2* are observed in clonal hematopoiesis,[124] and our results are consistent with prior studies linking chemotherapy and radiation exposure to expansion of clones with mutations in DDR genes, including *PPM1D*, *TP53*, *CHEK2*, *ATM*, and *SRCAP*.[107,108,145,222,291,310–312] CUX1 and EZH2 also function as epigenetic regulators, disruption of which blunts cellular responses to stress via transcriptional numbness.[201] Thus, the increased survival and subsequent expansion we observed may reflect a combination of abrogated DDR function as well as epigenetic perturbations.

General aneuploidy has been shown to promote chemotherapy resistance,[43,313] and we extend these studies to -7/del(7q) and t-MN by directly linking two genes to these pathogenic phenotypes.

There are many additional routes by which a clonal population can gain a competitive advantage and begin expanding.[314] Some mutations affect cell intrinsic properties such as cell division; loss of *Dnmt3a*, for example, enables HSCs to self-renew through serial transplant far past when wild-type cells exhaust, functionally conferring indefinite longevity.[315,316] *Tet2* loss similarly confers HSCs with increased self-renewal, measured by increased replating capacity in serial colony-forming assays as well as increased chimerism in competitive transplants.[317] In our studies, we also observed increased replating capacity and output in shCux1 HSPCs treated with EZH2 inhibitors, suggesting increased progenitor proliferation and self-renewal; future serial transplant experiments would provide additional evidence for combined *Cux1* and *Ezh2* deficiency increasing HSC self-renewal, especially as there is conflicting literature on the role of EZH2 in self-renewal.[155,318]

It is now widely accepted that tobacco is a potent carcinogen that causes cancer, and evidence is beginning to emerge that tobacco smoke can also promote expansion of cells with *TET2* and *ASXL1* inactivation as well as the gain-of-function *JAK2*^{V617F} mutation.[319,320] Increasing attention has also been given to understanding how inflammatory stress, which normally induces HSC proliferation and can result in HSC exhaustion under chronic inflammation,[321] promotes expansion of mutant clones. As inflammatory cytokine production gradually increases with age, a phenomenon termed ‘inflammaging’, this area of study may be particularly relevant.[314,322] Mutations in *asxl1*

stimulate expression of anti-inflammatory genes in hematopoietic progenitors of zebrafish, and exposure to the pro-inflammatory cytokines TNFα, IL-1β, and IL-6 can select for mutations in *Dnmt3a*, *Cebpa*, and *Tet2*, respectively.[323–326] Inflammation from chronic infection, as well as other microbial signals, can also stimulate expansion of cells bearing mutations in *Dnmt3a* and *Tet2*.[110,327] These studies also highlight the fact that mutations in a single gene can confer a competitive advantage under multiple stressors. Whether smoking habits and inflammation similarly promote expansion of -7/del(7q) cells with is an important, unanswered question.

Targeting *CUX1*;*EZH2*-mediated expansion

As the hematopoietic field increasingly understands the selective pressures driving expansion of CHIP clones, efforts to counter the selective advantages conferred by these mutations will move to the forefront. While drugging histone methyltransferases is a growing area of study,[328] the loss-of-function *EZH2* mutations characteristic of myeloid disease represent a therapeutic challenge. The global nature of EZH2 is one such complication: PRC2, the multi-subunit complex containing EZH2, remains the only identified methyltransferase with H3K27 activity and is solely responsible for H3K27 methylation in mouse embryonic stem cells, though EHMT2/G9a can help recruit PRC2 to target loci.[298,329,330] As H3K27me3 is a major repressive histone modification, reduced PRC2 activity may contribute to pathogenesis by decreasing the threshold for transcriptional activation of oncogenes, as has been observed for NOTCH signaling in T-cell leukemia and RAS signaling in malignant peripheral nerve sheath tumors.[328,331,332] Alternatively, PRC2 loss may generally increase transcriptional noise which can

contribute to drug resistance, though bromodomain inhibitors may represent a potential strategy to combat increased transcription.[332–334] Similarly, it may be possible to target downstream effectors that become upregulated with *EZH2* loss, though it is unlikely that this will fully counter the effects of genome-wide deregulation.[328]

Loss-of-function mutations in *CUX1* pose a similar challenge. As restoration of CUX1 levels can reverse pathogenic phenotypes including splenomegaly and anemia in mice,[135,142] targeting negative regulators of *CUX1* is a potential therapeutic strategy to increase expression of the second, intact *CUX1* allele for patients with -7/del(7q) or *CUX1* mutations. Targeting downstream effectors that increase in expression following *CUX1* loss is also a potential strategy, and a finer resolution of CUX1 binding sites at enhancer and promoter regions will aid in this.[137,139]

We observed a profound failure of *Cux1*;*Ezh2*-deficient cells to active DDR signaling following genotoxic exposure. Defective regulation of DNA repair is a common theme in oncogenesis, evidenced by studies of selective forces driving CHIP as well as genetic cancer predisposition diseases including Lynch and Bloom Syndrome.[221,305] Given these shortcomings in DNA repair, it is tempting to speculate that PARP inhibitors may have therapeutic efficacy against these cells, leveraging defective DDR to induce selective synthetic lethality similar to patients with *BRCA1/2* mutations.[335] The PARP1 enzyme canonically functions in the base excision repair (BER) pathway, and PARP inhibition traps the protein on DNA, forming crosslinks that result in unrepaired DNA, replication fork collapse, and additional replication-induced DNA damage.[336] As these breaks occur during the S-phase of the cell cycle they are repaired by homology-directed repair (HDR), for which the BRCA1/2 proteins are crucial, and HDR-deficient cells are

therefore selectively vulnerable to PARP inhibition.[337] CUX1 and CUX2 have been reported to function in BER,[338–340] and our data showing decreased 53BP1 recruitment after CUX1 and/or EZH2 loss suggests a role for CUX1 and EZH2 in non-homologous end joining (NHEJ),[142] the second major repair pathway in addition to HDR. The data therefore indicate *CUX1*;*EZH2*-deficient cells may have decreased fidelity of NHEJ and an increased reliance on HDR, which would render PARP inhibitors less effective. A meta-analysis of randomized controlled trials investigating PARP inhibition identified an increased risk of subsequent MDS and AML compared to placebo,[341] and a small case-study cohort identified chromosome 7 abnormalities in these patients at approximately the same frequency as general t-MNs.[342] PARP inhibition is thus unlikely to represent a viable therapy for -7/del(7q) and may actually promote t-MN development similar to platinating and alkylating agents.[121,145]

On the surface, it seems paradoxical that cells with defective DDR have a selective advantage as broken DNA should not be faithfully segregated during cell division, resulting in massive genomic instability. However, genomic instability is itself a hallmark of human cancer, thought to generate wide genetic diversity and increase the probability of a cell reaching a mutational state with a selective advantage.[59] There are numerous examples of genomic instability in human cancers, from microsatellite instability and hyper-mutable phenotypes in colon cancer to chromothripsis, a chaotic phenomenon in which massive DNA damage and rearrangements occur within a single chromosome.[343,344] Oddly, the defective DDR in *Cux1*;*Ezh2*-deficient cells does not necessarily translate to increased mutational load: though *CUX1*-mutant myeloid neoplasms have an increased mutational burden, the increase is modest and primary

myeloid disease in general has a low overall mutation burden compared to other cancers.[68,142] Further, t-MNs do not have increased coding mutational burdens compared to *de novo* AML, despite detectable mutation signatures from chemotherapy exposure.[222,223] Our data indicate *CUX1*;*EZH2*-deficient cells have decreased apoptosis following exposure to chemotherapy. This may enable cells to survive an insult that is lethal to wild-type cells and provide time to repair DNA to a sufficient degree to enable cell division without inducing excessive mutations. Future studies testing the kinetics of cell cycle entry and the fidelity of DNA segregation would be informative in this regard.

There are also some unconventional treatment options that may one day represent promising therapeutic avenues. It has recently been shown that antibody-mediated elimination of myeloid-biased HSCs in aged mice restores features of a more youthful immune system, including increased lymphoid progenitors and an improved adaptive immune response to viral infection.[345] Applying this concept to myeloid disease may be a route to selectively eliminate the HSCs that produce leukemic progeny, though it is dependent on such HSCs displaying the targeted surface markers. Finally, it may be possible to remove the selective advantage of mutant cells entirely by artificially recapitulating the mutation in healthy cells, for example administering an EZH2 inhibitor prior to chemotherapy and/or radiation.[121] As fitness is relative, uniform EZH2 suppression could eliminate the advantage conferred from EZH2 loss to the mutant population while still maintaining susceptibility to chemotherapies that target rapidly dividing cancerous cells.

Continued modeling of chromosome 7 deletions

Chromosome 7 contains approximately 900 protein-coding genes, and many additional candidate tumor suppressors have been identified in addition to *CUX1* and *EZH2*.[146,346,347] Though *Kmt2c* and *Kmt2e* are also reported to be involved in DNA damage repair,[191,200] we did not observe strong selection for inactivating mutations under steady-state or genotoxic stress conditions. However, there are other chromosome 7 genes implicated in the DDR, and many other genes that have functions similar to those of common driver mutations in AML **(Figure 23)**.[81] Whether combined *CUX1* and *EZH2* deficiency interacts with loss of additional chromosome 7 DDR genes is an important consideration, though our data indicates there may be a degree of functional redundancy for at least *KMT2C* and *KMT2E*. Examining the DDR capacity of -7/del(7q) iPSCs with deletions spanning *CUX1* and/or *EZH2* will be informative as to whether there are additional genes that further diminish the response to DNA damage.

CUX1;*EZH2* deficiency may also interact with loss of chromosome 7 genes that function in different pathways. Identifying whether these interactions are simply additive or if there are emergent properties arising from impinging on multiple cell biology pathways will be critical to furthering our understanding of chromosome 7 biology. To begin probing this question, we extended our RNP-based transfection approach even further to target six 7q TSG candidates **(Figure 24A, B)**. Genotyping colonies derived from individual cells revealed up to five gene knockouts can be induced within a single cell, with three gene knockouts being most common **(Figure 24C)**. Similar to our three-gene multiplex transfections, we observed both mono- and bi-allelic gene knockout.

These proof-of-principle experiments highlight the power of multiplex CRISPR-Cas9 gene knockout to dissect monosomies.

Finally, analysis of chromosome 7 genes offers insight as to whether chromosome loss itself can be leveraged therapeutically. Many chromosome 7-encoded genes are essential genes, such as *BUD31* and *NAMPT*,[348,349] and homozygous loss is incompatible with cell viability. As -7/del(7q) cells are missing a copy of these critical genes, they may be particularly sensitive to further dose perturbations and offer a wide therapeutic window. This has led to classifying such genes as CYCLOPS genes (Copy-number alterations Yielding Cancer Liabilities Owing to Partial losS),[350] and indeed -7/del(7q) AML cells are selectively sensitive to NAMPT inhibition.[351]

Conclusions

Great strides have been made in defining and categorizing the recurrent features of human cancer. A more refined understanding of how these recurrent features influence pathogenesis will be critical to enable discovery of novel therapeutics as well as identify which patient populations will benefit most from treatment. Here, we leverage multiplex CRISPR-Cas9 gene inactivation to model chromosome 7q deletions in primary HSCs by simultaneously targeting multiple candidate genes, which circumvents the lack of chromosomal synteny between humans and mice. We find combined deficiency of *Cux1* and *Ezh2* selectively results in clonal myeloid expansion, particularly after genotoxic stress. We also report increased drug resistance in both human and mouse cell systems following CUX1 and EZH2 loss. We also identify compounding defects to DNA damage recognition and repair, as *CUX1*;*EZH2*-deficient cells display reduced yH2AX deposition

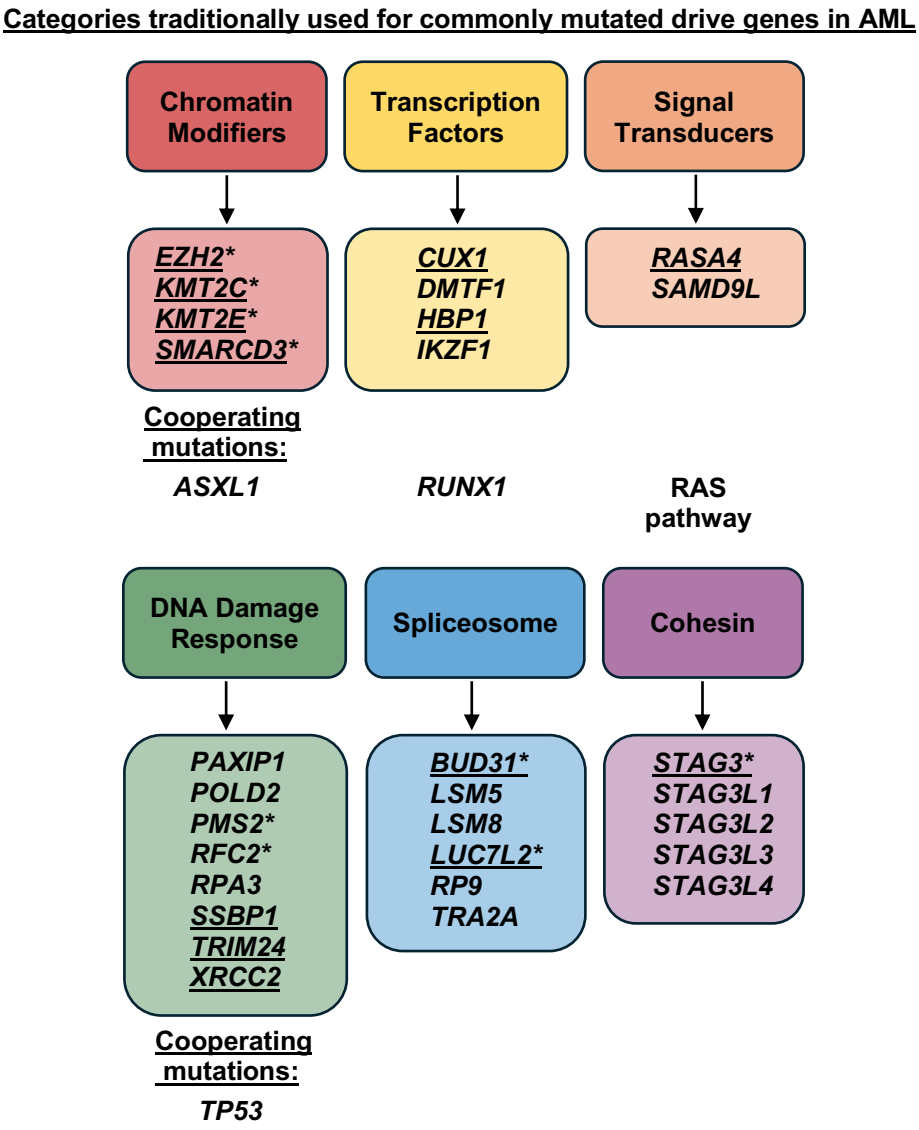

Figure 23: Functional groupings of chromosome 7 genes with co-occurring mutations.

Multiple chromosome 7 genes can be functionally classified into one of the eight categories of driver mutations that predominate in AML; some genes may function in multiple categories, but are classified according to their primary known function.[81] *Tumor Suppressors* and *NPM1* are not included, as all chromosome 7 gene candidates are thought to act as tumor suppressors in deletion events, and *NPM1* is not located on chromosome 7. Within each category, underlining indicates the gene is encoded within a CDR, and an asterisk denotes the gene was identified in data mining of the CRISPR-Cas9 screens for chemotherapy resistance **(Table 5)**. Shown beneath each category are mutations that commonly co-occur with -7/del(7q) in myeloid neoplasms.[81] Identifying whether combined *CUX1*;*EZH2* deficiency interacts with additional chromosome 7 gene deletions in the DNA damage response or other pathways, or with cooperating AML mutations, is a promising area of future study.

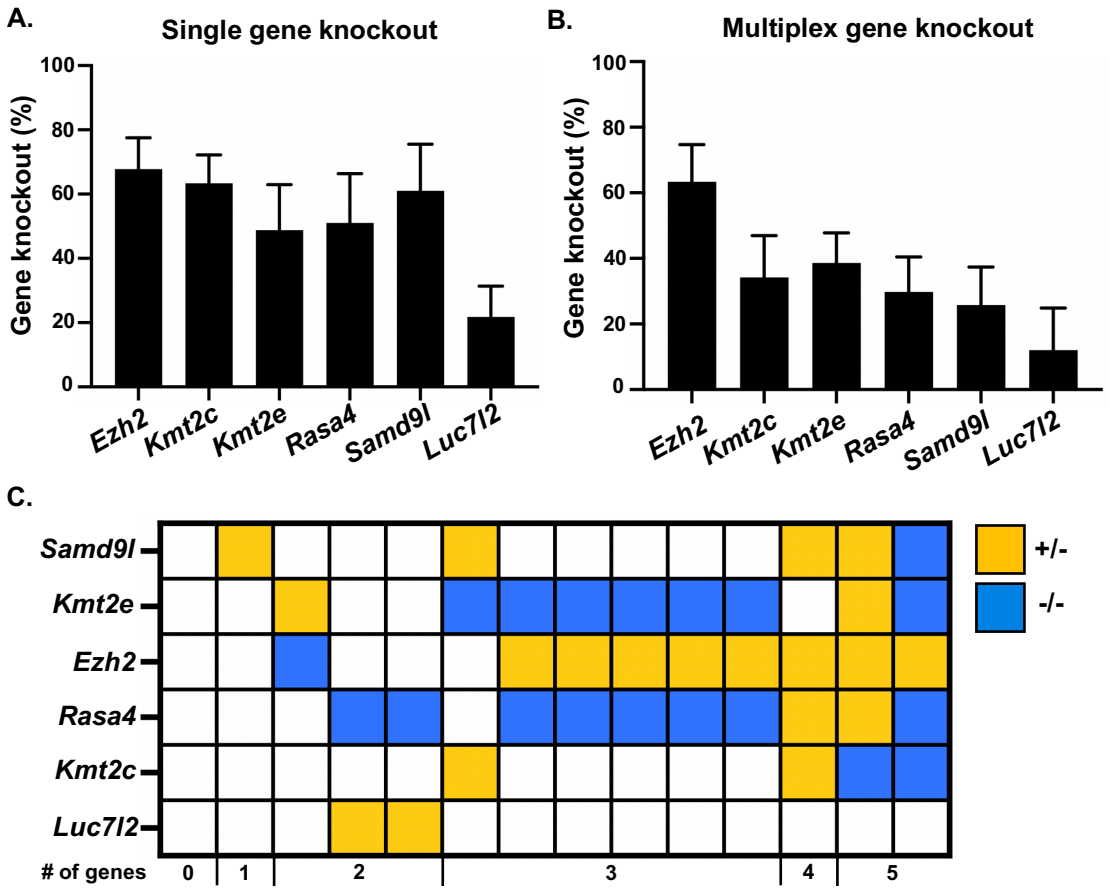

Figure 24: Continued modeling of 7q as a contiguous gene syndrome region.

A) Mouse HSPCs were isolated and transfected with RNPs targeting 7q genes. Cells were cultured for five days post-transfection, then DNA was isolated to determine gene knockout score. Data (n=5 biological replicates) is plotted as mean + SD. **B)** Mouse HSPCs were isolated and transfected with a pool of RNPs targeting six 7q genes simultaneously. Cells were cultured for five days post-transfection, and gene knockout was calculated as in A. Data (n=5 biological replicates) is plotted as mean + SD. **C)** HSPCs were plated in MethoCult M3434 for 7-10 days, then colonies derived from individual cells were genotyped. Up to fives genes were knocked out in the same cell, with three gene knockout occurring most frequently.

and 53BP1 recruitment, and increased residual DNA damage, following acute genetic insult. Our work provides additional evidence for non-transcriptional roles of CUX1 and EZH2 in the DDR, and our multiplex knockout approach is broadly applicable to other recurrent monosomies in cancer. With these studies, we conclude that 7q is a contiguous gene syndrome region that promotes cell fitness by reducing the efficacy of DNA damage recognition and repair **(Figure 25)**.

Future Directions

1. What additional stressors select for -7/del(7q) or *CUX1* mutations? Our study examined cellular response to the chemotherapeutics ENU, etoposide, and daunorubicin. ENU is well-characterized in mice but is not used to treat human disease; do additional alkylating agents such as cyclophosphamide or melphalan, or other classes of chemotherapeutics such as platinating agents or antimetabolites, similarly select for *Cux1*;*Ezh2*-deficient cells? Similarly, the interaction of inflammation and -7/del(7q) is understudied.
2. Do additional 7q genes function in the DDR that would synergize with *CUX1* and *EZH2* loss to further disrupt DNA damage recognition and repair?
3. What additional facets of cell biology are affected by del(7q), and are these disruptions additive or synergistic with *CUX1* and *EZH2* loss? Some patient deletions do not include the 7q22 CDR encoding *CUX1* **(Figure 2)**; do these patients have better outcomes, or are there additional 7q genes that synergize with *EZH2* loss?

4. Are there downstream effectors following *CUX1* and *EZH2* loss that can be therapeutically targeted? Similarly, are there any additional synthetical lethalities or CYCLOPS genes that can be targeted? As -7/del(7q) confers resistance to conventional chemotherapies, molecularly informed targeted therapies are likely necessary to improve patient outcomes.
5. What role do 7p genes have in -7 events? Most effort to identify TSGs has focused on 7q CDRs, but many genes identified in our chemotherapy resistance data mining reside on 7p **(Table 5)**.
6. How does -7/del(7q) interact with other recurrent features of myeloid neoplasms, including 5q and 17p deletions, and *RUNX1* and *ASXL1* mutations? Our lab has shown an interaction between *Cux1* deficiency and RAS signaling,[352] and whether such interactions exist with other karyotypic alterations remains unknown.

The current treatment regimens for cancer patients are sorely outdated. The induction therapy for patients with acute and therapy-related myeloid leukemias, a cycle of seven days of intravenous cytarabine followed by three days of intravenous daunorubicin, has been in use since 1973. Not only have there been no major changes in over 40 years, but there is a lifetime maximum exposure for daunorubicin due to cardiac toxicity and the data from human patients, as well as the data presented here, clearly demonstrate that these chemotherapies can select for cells with -7/del(7q). Our knowledge of the pathogenesis of -7/del(7q) has advanced to the degree that it borders on a moral quandary to continue administering the same non-specific cytotoxic therapies to t-MN patients despite clear evidence that those with -7/del(7q) often have little or no response.

While additional therapies are actively being piloted in *de novo* acute leukemia, including menin inhibitors for *NPM1*-mutant AML, there remains a dearth of options for t-MN patients. Leveraging the dose reduction in CYCLOPS genes encoded within 7q deletion segments may be a viable therapy worthy of further consideration. The use of short hairpin RNA or other nucleotide-based suppression methods would likely bear the greatest chance of success; this approach would avoid the binary dose levels established by CRISPR-Cas9 and would not require the extensive and expensive development of small molecule inhibitors. Further, the CYCLOPS gene target(s) can easily be adjusted to account for differing 7q deletion lengths. Our data suggest that the 7q region is likely a buffered system, with built-in redundancies such that effects observed with single gene deficiencies, for example *Kmt2c*, do not necessarily translate into additive or synergistic effects when combined with other gene deficiencies. A therapy that delivers multiple CYCLOPS shRNAs to hematopoietic cells of patients with -7/del(7q) disease, whether *de novo* or therapy-related, would effectively circumvent even a buffered system and should yield an appreciable therapeutic window.

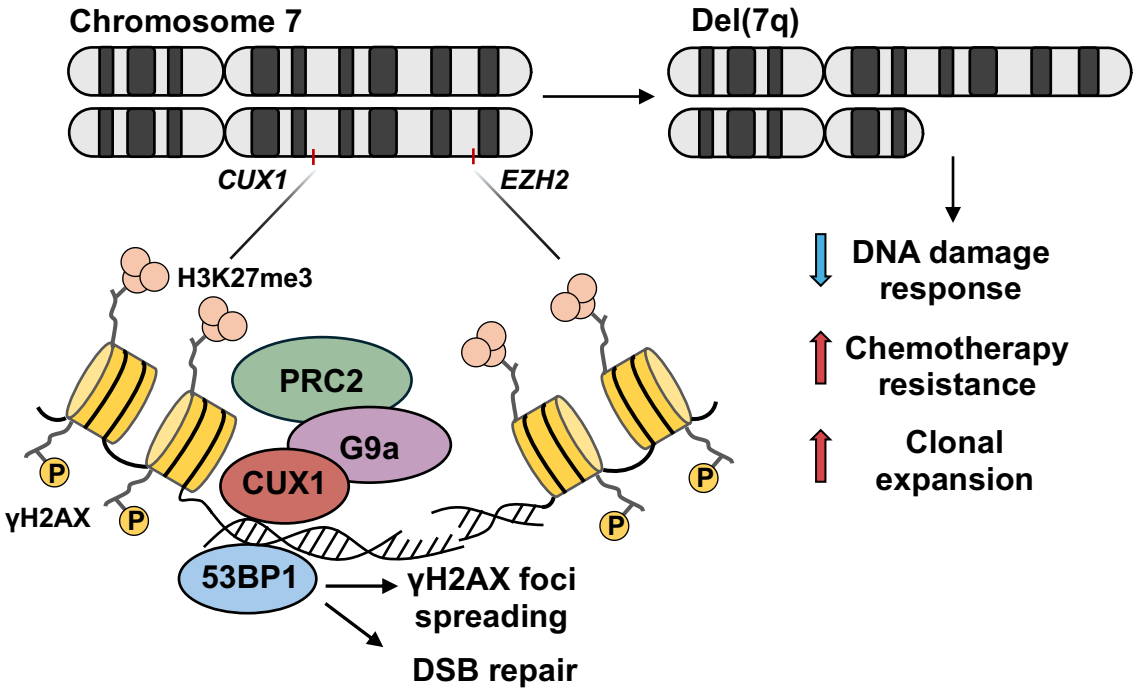

Figure 25: 7q is a contiguous gene syndrome region that promotes fidelity of DNA repair.

CUX1 and EZH2 function to promote DNA repair via recruitment of 53BP1, mediated in part by EHMT2/G9a and PRC2-catalyzed H3K27me3. -7/del(7q) events reduce the dosage of *CUX1* and *EZH2*, decreasing activation of the DNA damage response and the fidelity of DNA repair. This confers cells with increased resistance to DNA-damaging chemotherapy agents, enabling expansion.

CHAPTER 2

INTRODUCTION

Characteristics of stem cells

The identification and characterization of tissue-regenerative stem cells has been of long-standing interest to biological and medical communities. Decades of work have established a number of stem cell features that appear conserved across the various tissues of the body:

1) Stem cells typically reside in specialized niche environments with distinct anatomical locations, for example the crypts located between intestinal villi and the bulge of the hair follicle in skin.[353,354] The niche functions to protect stem cells from the external environment, but must also enable communicative signals to reach the stem cell compartment to induce proliferation as needed during tissue maintenance or injury repair.

2) Stem cells retain the ability to give rise to all lineages of their native tissue. As stem cells are thought to reside at the apex of the cell hierarchy, it follows they must be able to generate all types of mature cells. This is exemplified by the hematopoietic system, as transplant of purified HSCs into lethally irradiated recipients with ablated bone marrow will reconstitute all mature lineages of blood cells.[355] Stem cells also retain the ability to self-renew and maintain the stem cell pool; consequently, cell division can be characterized as either symmetric, in which a stem cell divides to produce a second stem cell, or asymmetric, in which the resultant daughter cell is committed to differentiation and exits the stem cell compartment.[356]

3) Stem cells are long-lived cells that exist across the organism's lifespan. Mature cells are continually lost through a variety of mechanisms, including physical forces that

slough off epithelial cells in the skin and intestine, programmed apoptosis that removes neutrophils to prevent accumulation of their histotoxic cargo,[357] and routine tissue maintenance that affects all tissues, including those with low proliferative rates such as cardiomyocytes.[358] Accordingly, the stem cell pool must persist through the organismal lifespan to replenish mature cells, either by maintaining the original stem cell population or supplementation by dedifferentiation of progenitor cells. This feature is also exemplified by the hematopoietic system, in which only HSCs are able to maintain blood production past sixteen weeks whereas restricted progenitors exhaust at earlier time points.[292]

4) Stem cells are largely quiescent, owing to the need to persist across decades of time, depending on the organism, and avoid excessive mutation accumulation from the inherent error rate of cell division. Even in highly proliferative tissues such as the hematopoietic system, which must replenish billions of mature cells daily, stem cell division is infrequent and human HSCs are estimated to divide only once per 40 weeks despite estimates of only 50,000-200,000 total HSCs.[359,360] In many regenerative tissues, the majority of cell divisions necessary to produce mature progeny are performed by transit-amplifying cells, highly proliferative progenitor cells that ultimately exhaust and differentiate, sparing the stem cell pool from replication-induced mutations.[361]

5) Stem cells have distinct biomarkers. Though all cells can be roughly distinguished based on gene expression profiles and protein repertoires, much effort has been spent identifying those that are unique to stem cells. Characterization of the CD48 and CD150 SLAM cell surface molecules of HSCs, for example, has enabled easier identification and flow cytometry-based purification of true stem cells from other progenitor populations.[362] Identification of other stem cell markers, such as the

intermediate filament Nestin in neural stem cells and the G-protein coupled receptor LGR5 in intestinal and other epithelial stem cells,[363,364] has greatly enhanced the ability to further study and characterize the functions of these stem cells.

Tools to identify and characterize stem cell function: the H2B-GFP mouse model

Accumulating evidence now suggests that extensive heterogeneity can exist even within a pool of immunophenotypically similar cells. For example, single-cell murine HSC transplant experiments have demonstrated that a surprisingly small fraction of stem cells produce balanced numbers of mature progeny, with many HSCs disproportionately generating myeloid or lymphoid output.[365,366] Lineage tracing studies, in which HSCs are labeled with unique DNA barcodes that can be tracked over time via single-cell sequencing, have been used to more clearly define transition boundaries over the differentiation process and examine the factors that influence cell fate decisions.[367,368] Collectively, these experiments highlight the differences present even within stringently-defined immunophenotypic populations, and have resulted in new models of hematopoiesis in which HSCs exist along a continuum of differentiation potential, in contrast to the binary bifurcations posited by earlier models.[369,370]

Stem cell proliferation rates were initially characterized by the retention of labeled nucleotides, for example pulsing cells with the thymidine analog 5-bromo-2-deoxyuridine (BrdU) followed by a chase period in which dividing cells dilute the BrdU and the signal diminishes.[371,372] Quiescent cells with low proliferation rates will retain the label and can be identified in flow cytometric or immunofluorescence microscopy assays. However, the cells need to be fixed and permeabilized to visualize the pulsed nucleotide, preventing

any further characterization of the labeled cells. To circumvent these barriers, the Elaine Fuchs lab developed a transgenic model in which GFP is fused to histone H2B under the control of a tetracycline-responsive regulatory element (TRE) and cytomegalovirus promoter (TRE-mCMV-H2B-GFP).[373] This model constitutively expresses *H2B-GFP* owing to the strong activity of the CMV promoter.[374] When crossed to mice harboring a tet-repressor protein, the addition of doxycycline suppresses H2B-GFP expression enabling a pulse-chase system. This system can be modified to be tissue-specific depending on the promoter or enhancer used to drive expression of the tet-repressor. This system offers two significant advantages over nucleotide labeling: first, cells do not need to divide to incorporate the label as histone H2B is a core component of the nucleosome; and second, GFP can easily be visualized with flow cytometry, enabling sorting and isolation of live label-retaining cells for downstream characterization.

The H2B-GFP model has been used extensively in a variety of tissues following its premier utilization to identify label-retaining, quiescent cells in the bulge of the hair follicle.[373] The hematopoietic field adapted the H2B-GFP model by placing the tet-repressor under the HSC-specific stem cell leukemia (SCL) promoter, and a number of landmark papers have been published with this system **(Table 8)**. The system has also been modified to be under the control of a reverse tetracycline-controlled transactivator, such that addition of doxycycline drives expression of *H2B-GFP*, enabling an inducible pulse with withdrawal of doxycycline initiating the chase period.[375,376] However, there have also been noted concerns that the system is 'leaky' and the *H2B-GFP* transgene is expressed independently of doxycycline administration.[377] Further, although some histone proteins can be quite long-lived, the half-life of H2B-GFP is approximately 4 to 6

Table 8: Landmark Hematopoiesis Studies Using the H2B-GFP Mouse

Reference	System	Findings
Tumbar et al. 2004, *Science*	K5 Tet-OFF	Skin stem cells in the niche rarely divide but can abruptly begin cycling[373]
Wilson et al. 2008, *Cell*	SCL Tet-OFF	HSCs reversibly switch from dormancy to self-renewal under stress conditions[102]
Foudi et al. 2009, *Nat. Biotech.*	Rosa26 Tet-ON	HSC quiescence is regulated by *Gfi1*, not *p21*[376]
Sugimura et al. 2012, *Cell*	SCL Tet-OFF	Non-canonical Wnt signaling maintains HSC quiescence[378]
Walter et al. 2015, *Nature*	SCL Tet-OFF	Stress-induced HSC activation accumulates DNA damage; HSCs with defects in the Fanconi Anemia pathway die after repeat activation[379]
Bernitz et al. 2016, *Cell*	hCD34 Tet-OFF	HSCs that exceed 5 divisions can no longer reconstitute irradiated recipients[380]
Cabezas-Wallscheid et al. 2017, *Cell*	SCL Tet-OFF	Label-retaining HSCs lack biosynthetic activity[381]
Kokkoliaris et al. 2020, *Blood*	SCL Tet-OFF	Cycling and dormant HSCs occupy different niches[382]
Reference	**System**	**Findings**
Challen & Goodell 2008, *PLoS One*	CMV Tet-OFF	85% of HSCs remain GFP$^+$ after 12 weeks of Dox chase; 5-FU treatment does not dilute GFP fluorescence intensity or positivity[377]
Foudi et al. 2009, *Nat. Biotech.*	Rosa26 Tet-ON	H2B-GFP is not specific to HSCs as a single parameter; short-term progenitors retain GFP even after 24 weeks of chase[376]
Morcos et al. 2020, *JEM*		The half-life of H2B-GFP is only 4-6 weeks[383]

weeks, significantly shorter than the chase duration of many studies utilizing the H2B-GFP mice **(Table 8)**.[383,384] Repeat pulse-chase experiments show label-retaining HSCs do not abruptly enter quiescence after four divisions, calling into question the results and conclusions from many high-profile studies.[380,383]

Knowledge gaps to be addressed

Both the integration site of the *H2B-GFP* transgene, as well as the number of intact copies that integrated, are currently unknown. Transgenes randomly integrate into the host genome, and the insertion site can have substantial influence; insertional mutagenesis is a well-known consequence of a transgene integrating nearby or into existing genes, disrupting their endogenous expression levels. While this has been used as a discovery tool to identify cancer-promoting genes in forward genetics screens, it has also occurred in gene therapy trials and is considered a severe adverse effect.[385–387] The integration site and orientation can also have profound effects on the transgene itself, known as the position effect.[388,389] Integration near heterochromatin can result in progressive silencing of the transgene, and, conversely, integration near strong activating regulatory elements can increase basal expression, resulting in leakiness. Given the concerns of the high background of the *H2B-GFP* transgene, identifying and characterizing the integration site will likely provide insight as to whether the genomic position influences its expression level. Conclusively defining the insertion site will also provide additional context that can aid in interpreting results obtained from this model.

MATERIALS AND METHODS

Mouse models

H2B-GFP mice were purchased from Jackson Laboratories (Strain 005104) and crossed with mice carrying a reverse tetracycline-controlled transactivator protein under control of the *Rosa26* locus (*Rosa26*^{rtTA}, Tet-ON). H2B-GFP genotyping was performed by Transnetyx. Cux1^{mCherry} reporter mice were generated using CRISPR-Cas9. Single cell fertilized ova from B6.SJL/J mice were injected with a repair construct (12.5 ng/μL) consisting of a 2.8 kb 5' homology arm, the mCherry reporter sequence, and a 1 kb 3' homology arm along with sgRNA targeting the stop codon of exon 24 of *Cux1* (5'-CCAUCGAAUGGGAGUUC-3') (25 ng/μL, IDT) and Cas9 (100 ng/μL, IDT). Repair constructs were purified from E. coli cultured for 8 hours in LB dissolved in cell culture-grade water (37˚C, 250 rpm) using the EndoFree Plasmid Maxi Kit (Qiagen). Pups were screened for homologous repair by PCR and Sanger Sequencing to confirm the entire sequence inserted without mutation. Transgenic mice were backcrossed with B6.SJL/J for one generation to generate F1 founders.

H2B$^{GFP/GFP}$; Rosa26$^{rtTA/rtTA}$ mice were crossed to Cux1$^{mCh/mCh}$ mice, and the resultant F1 progeny were crossed to attempt to generate H2B^{GFP}; Cux1$^{mCh/mCh}$; Rosa26$^{rtTA/+}$ mice.

H2B-GFP pulse-chase and flow cytometry

H2B^{GFP}; Rosa26^{rtTA} mice were given Dox chow for 30 days to load GFP, then switched to normal chow for a 150-day chase period. After the chase, bone marrow was isolated and RBCs were depleted with ACK lysis buffer. Cells were lineage depleted and

stained for the HSC markers Sca-1, c-kit, CD48, and CD150 using the same antibodies as in **Table 4**. Age-matched $H2B^{GFP}$; $Rosa26^{+}$ mice were used as controls.

Copy-number identification and targeted locus amplification

qPCR was used to identify the copy-number[390] of the *H2B-GFP* transgene by comparing the amplification cycle of GFP with that of a single-copy reference sample, the GFP downstream of the *Cux1* shRNA from shCux1 and shCtrl mice, described in Chapter 1. qPCR primers for GFP gDNA are: Forward 5' GAACCGCATCGAGCTGAA 3'; Reverse 5' TGCTTGTCGGCCATGATATAG 3'.

Targeted locus amplification (TLA) was performed to identify the integration site of the *H2B-GFP* transgene.[391] Briefly, $1x10^7$ splenocytes from $H2B\text{-}GFP^{tg/tg}$ mice, and shCux1 mice as a control, were cross-linked with 37% formaldehyde for 10 minutes at room temperature before reverse-crosslinking with 1 M glycine and cell lysis. For DNA digestion, 5% SDS and 20% Triton X-100 were added for overnight incubation at 37˚C with NlaIII restriction endonuclease (New England Biolabs). T4 ligase (ThermoFisher) was added for 2 hours at room temperature, and DNA was purified and digested with NspI restriction endonuclease (New England Biolabs). T4 ligase was added again, and the now-circularized DNA was magnetically isolated with AMPure XP beads (Beckman). TLA PCR was performed on the purified DNA using 9 pairs of inverse primers originating in GFP and directed outward **(Table 9)**, and the PCR product was magnetically purified by AMPure XP beads. Purified PCR products were submitted to the University of Chicago Genomics Core Facility for library preparation and Illumina sequencing (NovaSeq-6000) at a depth of 2 million reads.

Table 9: PCR Primers for Targeted Locus Amplification

Primer Pair	Forward primer (5' -> 3')	Reverse primer (5' -> 3')
1	TGATGGTCGAGCGCTTG	CGTGTCCGAGGGTACTAAG
2	CTCGACCAGGATGGGCA	AAGTTCATCTGCACCACCG
3	GTCAGCTTGCCGTAGGT	CAGTGCTTCAGCCGCTA
4	TTGAAGAAGATGGTGCGC	ATCGAGCTGAAGGGCAT
5	TGTCGCCCTCGAACTTC	ACAAGCTGGAGTACAACTACA
6	TGTTCTGCTGGTAGTGGT	CCAACGAGAAGCGCGAT
7	GATGTTGTGGCGGATCTTGAA	AGTCCGCCCTGAGCAAA
8	TTGATGCCGTTCTTCTGC	ACAACCACTACCTGAGCA
9	TGTTGTAATGCGCCAGG	CTAAGGCCATCACCAAGTAC

RESULTS: DETERMINING THE INTEGRATION SITE OF THE *H2B-GFP* TRANSGENE

H2B-GFP mice display leaky transgene expression

To assess the basal expression of the *H2B-GFP* transgene and identify label-retaining cells, we set up a pulse-chase experiment with mice containing one copy of GFP-tagged H2B and a reverse tetracycline-controlled transactivator (rtTA) protein driven by the *Rosa26* locus. Addition of doxycycline in the chow induces *H2B-GFP* expression. We initiated a 30-day pulse of doxycycline chow followed by a 150-day chase period **(Figure 26A)**, using age-matched controls with no rtTA. Flow cytometry analysis of HSCs, gated on the classical LSK-SLAM markers,[362] did reveal a population of GFP^+ cells; however, age-matched controls displayed equivalent GFP positivity **(Figure 26B)**, indicating aberrant transgene expression even in the absence of doxycycline to drive transcription. We note that our H2B-GFP system is a Tet-ON system, in contrast to many models referenced in **Table 8**, which utilize a Tet-OFF system. Consequently, our system may be particularly susceptible to leakiness as there is no tTA protein physically repressing transcription. Still, our results demonstrate inappropriate *H2B-GFP* expression and transgene leakiness in agreeance with previous studies.[377,383]

H2B-GFP is located on chromosome 5 of the mouse genome

Our lab has previously shown that shRNA-based reduction in *Cux1* levels alter HSPC self-renewal and proliferation, resulting in lethal myeloid disease in mice.[135,142] To further characterize how variations in endogenous CUX1 levels influence HSC function, we generated a reporter mouse in which the endogenous 3' end of the *Cux1* gene is tagged with an in-frame mCherry fluorescent protein cDNA. This enables sorting cells into

fractions with high or low levels of CUX1, termed Cux1^{bright} and Cux1^{dim}, respectively, based on mCherry fluorescence which directly reflects CUX1 levels **(Figure 27A)**. To identify differences in proliferation between Cux1^{dim} and Cux1^{bright} HSCs, we crossed homozygous Cux1$^{mCh/mCh}$ mice to homozygous H2B$^{GFP/GFP}$ mice. As the location of the *H2B-GFP* transgene is unknown, we performed a binary genotyping with primers for GFP and gel electrophoresis cannot distinguish heterozygosity from homozygosity. We therefore generated homozygous H2B$^{GFP/GFP}$ mice by crossing heterozygous founders together, then interbreeding the F1 generation until a cross yielded exclusively H2B^{GFP} pups. Homozygosity was confirmed by interbreeding the F2 animals and again obtaining exclusively H2B^{GFP} pups. After the initial P generation cross between H2B$^{GFP/GFP}$ and Cux1$^{mCh/mCh}$ mice, we interbed the resultant heterozygous F1 offspring with the goal of generating H2B^{GFP}; Cux1$^{mCh/mCh}$ mice to study GFP retention in Cux1^{dim} and Cux1^{bright} HSCs **(Figure 27B)**.

We expected Mendelian ratios of F2 offspring following independent assortment **(Figure 27C)**; however, we observed ratios of offspring that closely approximated the 1:2:1 ratio characteristic of a monohybrid cross **(Figure 27D)**, indicating strong linkage disequilibrium. As *Cux1* is located within the 5A3 chromosome band in the murine genome, these results demonstrate that the *H2B-GFP* transgene is located on murine chromosome 5. We recovered only a single recombinant offspring in 88 pups in which the H2B^{GFP} allele had segregated away from the Cux1^{mCherry} allele, resulting in a H2B$^{+/+}$; Cux1$^{mCh/+}$ genotype **(Figure 27D)**. It is possible there are additional recombinants with the genotype H2B$^{GFP/GFP}$; Cux1$^{mCh/+}$ in which the recombination event conferred homozygosity at the *H2B-GFP* locus; however, these offspring would be indistinguishable

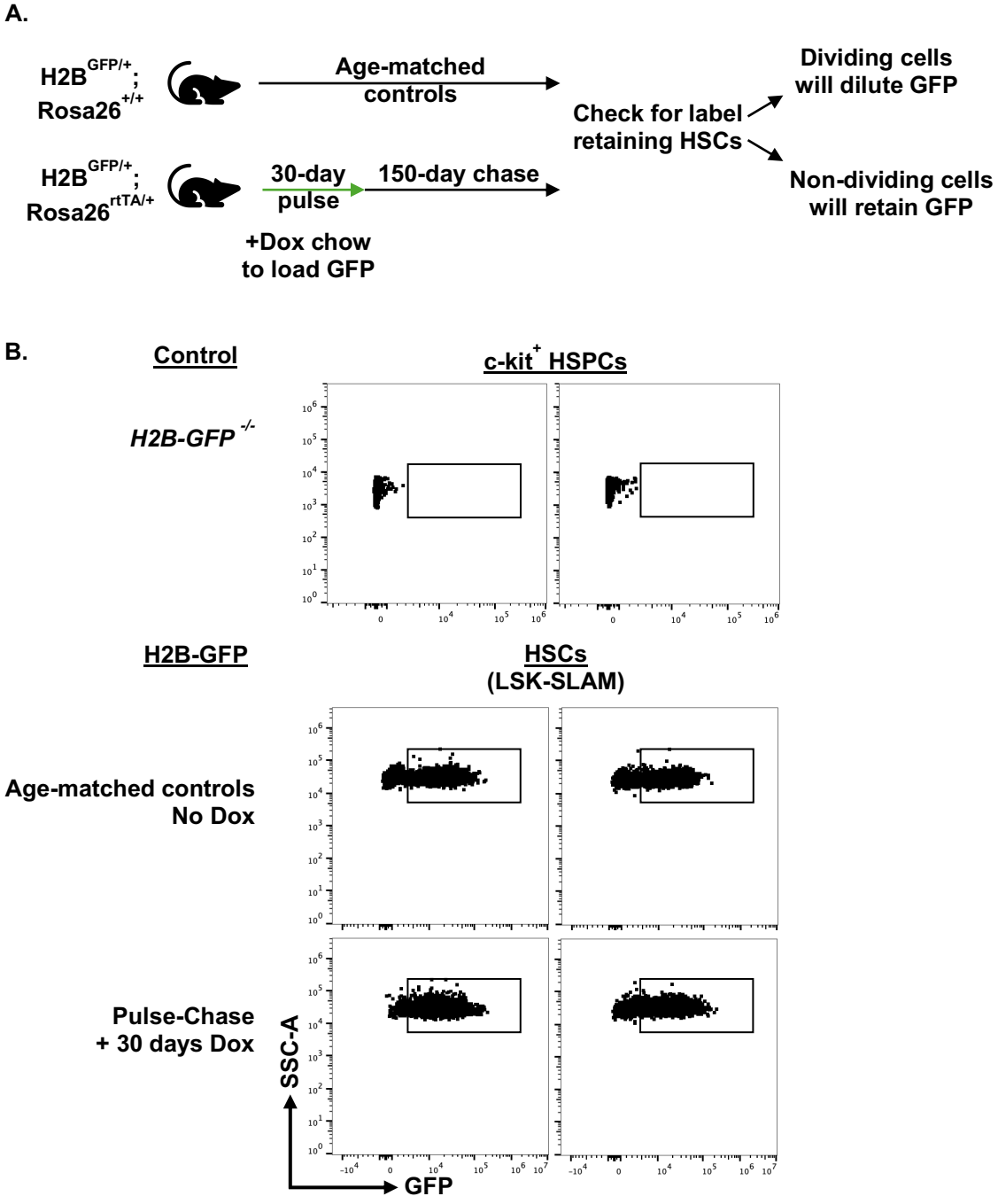

Figure 26: The *H2B-GFP* transgene is promiscuously expressed.

A) Experimental schematic for the pulse-chase designed to identify label-retaining HSCs. H2B-GFP mice with the Tet-ON rtTA were given doxycycline-containing chow for 30 days to load GFP, then switched to normal chow for 150 days to allow GFP dilution. **B)** HSCs (Lineage$^-$; Sca-1$^+$; c-kit$^+$; CD48$^-$; CD150$^+$) were gated and examined for GFP retention. Age-matched controls display high levels of GFP fluorescence despite no rtTA element. A negative control with no *H2B-GFP* transgene is shown on top.

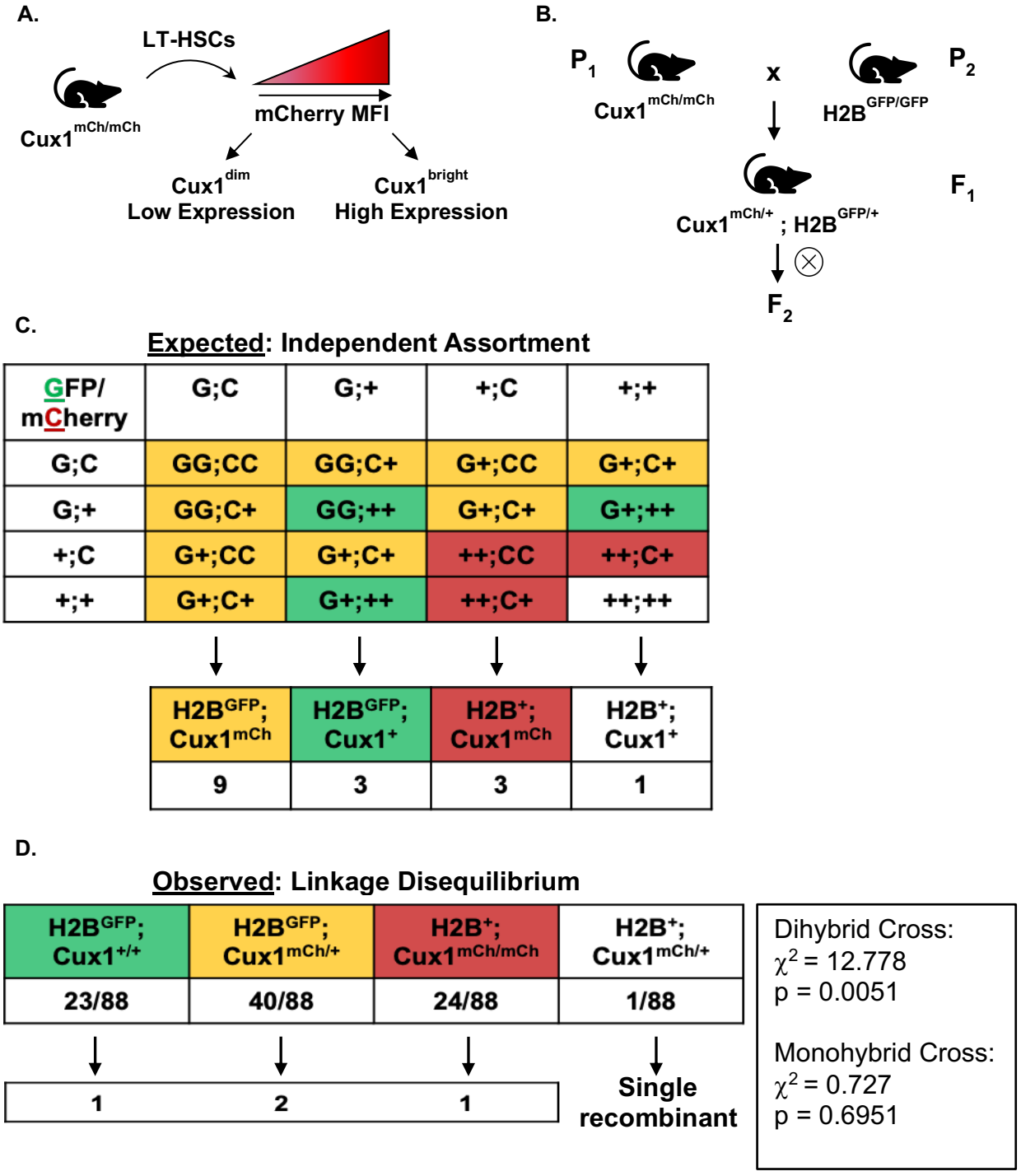

Figure 27: Linkage disequilibrium of *H2B-GFP* and *Cux1^{mCherry}*.

A) Schematic showing the distinction between Cux1^{dim} and Cux1^{bright} HSCs based on the spectrum of mCherry fluorescence, which directly reflects CUX1 protein levels. **B)** Outline of the breeding scheme to cross the *H2B-GFP* and *Cux1^{mCherry}* alleles together. Parent mice homozygous for each gene are bred together to generate F1 heterozygotes, which

Figure 27 continued

are interbred in to set up a dihybrid cross. Approximately 1 in 16 resulting offspring should be homozygous for both *H2B-GFP* and *Cux1*^{mCherry}. **C)** The expected phenotypes from a dihybrid cross that follows Mendel's law of independent assortment, yielding a 9:3:3:1 phenotypic ratio. Cells are color-coded to denote the outward phenotype. Because the location of the *H2B-GFP* transgene is unknown, zygosity cannot be determined. **D)** The observed phenotypes from the F1 dihybrid cross, which closely approximates the 1:2:1 ratio characteristic of a monohybrid cross. In 88 pups genotyped, only a single animal displayed a detectable genotype consistent with meiotic recombination between the *H2B-GFP* and *Cux1*^{mCherry} loci. Chi-square test statistic for dihybrid cross $\chi^2 = 12.778$, $p = 0.0051$. Chi-square test statistic for monohybrid cross $\chi^2 = 0.727$, $p = 0.6951$.

from H2B$^{GFP/+}$; Cux1$^{mCh/+}$ animals. Therefore, we conservatively estimate the *H2B-GFP* transgene to be located approximately 1 megabase from the *Cux1* locus.[392]

Three copies of *H2B-GFP* integrated into the mouse genome

We used qPCR to identify the number of copies of *H2B-GFP* in the murine genome. We established a standard curve using genomic DNA from shCux1 mice, which harbor a single copy of GFP.[135] We compared the amplification cycle of GFP in DNA isolated from the splenocytes of heterozygous H2B-GFP mice to the calibration curve, and determined there are likely three copies of *H2B-GFP* integrated into the mouse genome **(Figure 28)**.

To precisely identify the insertion location, we performed targeted locus amplification (TLA), a cross-linking-based technique that generates DNA libraries covering approximately 100 kb of sequence flanking an inverse primer pair.[391] We generated TLA libraries using DNA from heterozygous H2B-GFP founder mice and shCux1 mice as a positive control, since the location of *GFP* in these mice is known to be downstream of the *Col1a1* locus.[135] As fewer than one million reads can provide adequate depth and locus coverage to identify integration sites, we submitted 9 sequencing libraries using multiple inverse primer pairs originating in *GFP* **(Table 9)**. Bioinformatic analysis, including *in silico* NlaIII digestion for reads that initially fail to align, should reveal the boundary region where known genomic sequences meet *GFP*. Sanger sequencing can then be performed to identify the precise nucleotide break points, and examination of the integration region may provide insight regarding inappropriate transgene expression.

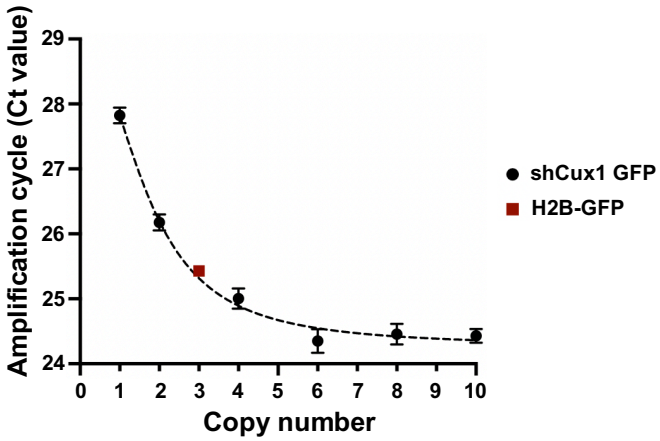

Figure 28: Three copies of *H2B-GFP* integrated into the mouse genome.

A qPCR standard amplification curve was generated by adding increasing amounts of genomic DNA isolated from shCux1 splenocytes, which harbor one copy of GFP. The amplification cycle of genomic DNA isolated from H2B-GFP splenocytes is located between two and four copies of GFP, closely approximating three copies.

DISCUSSION

Overview and Implications

Here we have initiated the process of identifying the integration site of the *H2B-GFP* transgene in the mouse genome to determine if its chromosomal location has a transcriptional influence. Our preliminary data utilizing a Tet-ON system demonstrate leakiness and inappropriate GFP expression even in mice lacking the rtTA element, which should be unable to induce transcription in the absence of doxycycline **(Figure 26)**. Though our system differs from most published models, which are Tet-OFF, the location of the *H2B-GFP* transgene remains the same in both systems and thus defining the precise integration site will still be informative.

We narrowed the integration location to murine chromosome 5 through our attempts to cross the *H2B-GFP* transgene into a *Cux1*-mCherry-tagged strain, noting that the inheritance pattern closely matched a monohybrid Mendelian cross **(Figure 27)**. Comparison of the qPCR amplification cycle of *H2B-GFP* to a calibration curve suggests three copies of the transgene integrated within chromosome 5 **(Figure 28)**. The precise boundaries of the integrate site are expected to be revealed by our TLA sequencing libraries. Presumably, the transgene integrated into a region of open chromatin near strong enhancer or promoter sequences that can drive transcription even in the presence of the tTA transcriptional repressor. Further functional studies, such as CRISPR inactivation or removal of such sequences, will gauge their influence on the *H2B-GFP* transgene. Many high-profile publications have utilized the H2B-GFP model and drawn conclusions about HSC biology and function **(Table 8)**; a better understanding of the methodological tool used in these studies may therefore suggest revising the conclusions.

REFERENCES

1. Siegel, J. J. & Amon, A. New insights into the troubles of aneuploidy. *Annu. Rev. Cell Dev. Biol.* **28**, 189–214 (2012).

2. Williams, B. R. *et al.* Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* **322**, 703–709 (2008).

3. Donnelly, N., Passerini, V., Dürrbaum, M., Stingele, S. & Storchová, Z. HSF1 deficiency and impaired HSP90-dependent protein folding are hallmarks of aneuploid human cells. *EMBO J.* **33**, 2374–2387 (2014).

4. Chunduri, N. K. & Storchová, Z. The diverse consequences of aneuploidy. *Nat. Cell Biol.* **21**, 54–62 (2019).

5. Stephenson, M. D., Awartani, K. A. & Robinson, W. P. Cytogenetic analysis of miscarriages from couples with recurrent miscarriage: a case-control study. *Hum. Reprod. Oxf. Engl.* **17**, 446–451 (2002).

6. Rasmussen, S. A., Wong, L.-Y. C., Yang, Q., May, K. M. & Friedman, J. M. Population-based analyses of mortality in trisomy 13 and trisomy 18. *Pediatrics* **111**, 777–784 (2003).

7. Moerman, P., Fryns, J. P., van der Steen, K., Kleczkowska, A. & Lauweryns, J. The pathology of trisomy 13 syndrome. A study of 12 cases. *Hum. Genet.* **80**, 349–356 (1988).

8. Cereda, A. & Carey, J. C. The trisomy 18 syndrome. *Orphanet J. Rare Dis.* **7**, 81 (2012).

9. Antonarakis, S. E. *et al.* Down syndrome. *Nat. Rev. Dis. Primer* **6**, 9 (2020).

10. Tsou, A. Y. *et al.* Medical Care of Adults With Down Syndrome: A Clinical Guideline. *JAMA* **324**, 1543–1556 (2020).

11. Watkins, P. C., Tanzi, R. E., Cheng, S. V. & Gusella, J. F. Molecular genetics of human chromosome 21. *J. Med. Genet.* **24**, 257–270 (1987).

12. Hassold, T. & Hunt, P. To err (meiotically) is human: the genesis of human aneuploidy. *Nat. Rev. Genet.* **2**, 280–291 (2001).

13. Urbach, A. & Benvenisty, N. Studying early lethality of 45,XO (Turner's syndrome) embryos using human embryonic stem cells. *PloS One* **4**, e4175 (2009).

14. Smyth, C. M. & Bremner, W. J. Klinefelter syndrome. *Arch. Intern. Med.* **158**, 1309–1314 (1998).

15. Bardsley, M. Z. *et al.* 47,XYY syndrome: clinical phenotype and timing of ascertainment. *J. Pediatr.* **163**, 1085–1094 (2013).

16. Berglund, A., Stochholm, K. & Gravholt, C. H. Morbidity in 47,XYY syndrome: a nationwide epidemiological study of hospital diagnoses and medication use. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **22**, 1542–1551 (2020).

17. Otter, M., Schrander-Stumpel, C. T. R. M. & Curfs, L. M. G. Triple X syndrome: a review of the literature. *Eur. J. Hum. Genet. EJHG* **18**, 265–271 (2010).

18. Gravholt, C. H., Viuff, M. H., Brun, S., Stochholm, K. & Andersen, N. H. Turner syndrome: mechanisms and management. *Nat. Rev. Endocrinol.* **15**, 601–614 (2019).

19. Prestel, M., Feller, C. & Becker, P. B. Dosage compensation and the global re-balancing of aneuploid genomes. *Genome Biol.* **11**, 216 (2010).

20. Brooker, A. S. & Berkowitz, K. M. The roles of cohesins in mitosis, meiosis, and human health and disease. *Methods Mol. Biol. Clifton NJ* **1170**, 229–266 (2014).

21. Kucej, M. & Zou, H. DNA-dependent cohesin cleavage by separase. *Nucl. Austin Tex* **1**, 4–7 (2010).

22. Musacchio, A. The Molecular Biology of Spindle Assembly Checkpoint Signaling Dynamics. *Curr. Biol. CB* **25**, R1002-1018 (2015).

23. Ruijtenberg, S. & van den Heuvel, S. Coordinating cell proliferation and differentiation: Antagonism between cell cycle regulators and cell type-specific gene expression. *Cell Cycle Georget. Tex* **15**, 196–212 (2016).

24. Kapadia, C. D. & Goodell, M. A. Tissue mosaicism following stem cell aging: blood as an exemplar. *Nat. Aging* (2024).

25. Knouse, K. A., Wu, J., Whittaker, C. A. & Amon, A. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13409–13414 (2014).

26. Youssoufian, H. & Pyeritz, R. E. Mechanisms and consequences of somatic mosaicism in humans. *Nat. Rev. Genet.* **3**, 748–758 (2002).

27. Pavelka, N. *et al.* Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468**, 321–325 (2010).

28. Sheltzer, J. M., Torres, E. M., Dunham, M. J. & Amon, A. Transcriptional consequences of aneuploidy. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12644–12649 (2012).

29. Oromendia, A. B., Dodgson, S. E. & Amon, A. Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* **26**, 2696–2708 (2012).

30. Kenmochi, N. *et al.* A map of 75 human ribosomal protein genes. *Genome Res.* **8**, 509–523 (1998).

31. Golomb, L., Volarevic, S. & Oren, M. p53 and ribosome biogenesis stress: the essentials. *FEBS Lett.* **588**, 2571–2579 (2014).

32. Baker, N. E. Emerging mechanisms of cell competition. *Nat. Rev. Genet.* **21**, 683–697 (2020).

33. Pfau, S. J., Silberman, R. E., Knouse, K. A. & Amon, A. Aneuploidy impairs hematopoietic stem cell fitness and is selected against in regenerating tissues in vivo. *Genes Dev.* **30**, 1395–1408 (2016).

34. Weaver, B. A. A. & Cleveland, D. W. Does aneuploidy cause cancer? *Curr. Opin. Cell Biol.* **18**, 658–667 (2006).

35. Boveri, T. *Zurfrage der entstehung maligner tumoren*. Jena, Germany: Fischer; 1914.

36. Mitelman, F., Johannson, B., & Mertens F. (Eds.) Database of Chromosome Aberrations and Gene Fusions in Cancer (2024). https://mitelmandatabase.isb-cgc.org.

37. Albertson, D. G., Collins, C., McCormick, F. & Gray, J. W. Chromosome aberrations in solid tumors. *Nat. Genet.* **34**, 369–376 (2003).

38. Hahn, W. C. *et al.* Creation of human tumour cells with defined genetic elements. *Nature* **400**, 464–468 (1999).

39. Marx, J. Debate surges over the origins of genomic defects in cancer. *Science* **297**, 544–546 (2002).

40. Weaver, B. A. & Cleveland, D. W. The aneuploidy paradox in cell growth and tumorigenesis. *Cancer Cell* **14**, 431–433 (2008).

41. Sheltzer, J. M. & Amon, A. The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends Genet. TIG* **27**, 446–453 (2011).

42. Rutledge, S. D. *et al.* Selective advantage of trisomic human cells cultured in non-standard conditions. *Sci. Rep.* **6**, 22828 (2016).

43. Lukow, D. A. *et al.* Chromosomal instability accelerates the evolution of resistance to anti-cancer therapies. *Dev. Cell* **56**, 2427-2439.e4 (2021).

44. Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *Nature* **606**, 984–991 (2022).

45. Boyd, K. D. *et al.* The clinical impact and molecular biology of del(17p) in multiple myeloma treated with conventional or thalidomide-based therapy. *Genes. Chromosomes Cancer* **50**, 765–774 (2011).

46. Teoh, P. J. *et al.* p53 haploinsufficiency and functional abnormalities in multiple myeloma. *Leukemia* **28**, 2066–2074 (2014).

47. Liu, Y. *et al.* Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. *Nature* **531**, 471–475 (2016).

48. Xue, W. *et al.* A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8212–8217 (2012).

49. Zabarovsky, E. R., Lerman, M. I. & Minna, J. D. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. *Oncogene* **21**, 6915–6935 (2002).

50. Boultwood, J. *et al.* Narrowing and genomic annotation of the commonly deleted region of the 5q- syndrome. *Blood* **99**, 4638–4641 (2002).

51. Fairman, J., Chumakov, I., Chinault, A. C., Nowell, P. C. & Nagarajan, L. Physical mapping of the minimal region of loss in 5q- chromosome. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7406–7410 (1995).

52. Ebert, B. L. Molecular dissection of the 5q deletion in myelodysplastic syndrome. *Semin. Oncol.* **38**, 621–626 (2011).

53. Ebert, B. L. *et al.* Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008).

54. Starczynowski, D. T. *et al.* Identification of miR-145 and miR-146a as mediators of the 5q- syndrome phenotype. *Nat. Med.* **16**, 49–58 (2010).

55. Stoddart, A. *et al.* Cell intrinsic and extrinsic factors synergize in mice with haploinsufficiency for Tp53, and two human del(5q) genes, Egr1 and Apc. *Blood* **123**, 228–238 (2014).

56. Knudson, A. G. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 820–823 (1971).

57. Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948–962 (2013).

58. Hahn, W. C. *et al.* Enumeration of the simian virus 40 early region elements necessary for human cell transformation. *Mol. Cell. Biol.* **22**, 2111–2123 (2002).

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

60. Bhatnagar, N., Nizery, L., Tunstall, O., Vyas, P. & Roberts, I. Transient Abnormal Myelopoiesis and AML in Down Syndrome: an Update. *Curr. Hematol. Malig. Rep.* **11**, 333–341 (2016).

61. Martinez, T. C. & McNerney, M. E. Haploinsufficient Transcription Factors in Myeloid Neoplasms. *Annu. Rev. Pathol.* **19**, 571–598 (2024).

62. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319**, 525–532 (1988).

63. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).

64. Zeuner, A., Todaro, M., Stassi, G. & De Maria, R. Colorectal cancer stem cells: from the crypt to the clinic. *Cell Stem Cell* **15**, 692–705 (2014).

65. Tariq, K. & Ghias, K. Colorectal cancer carcinogenesis: a review of mechanisms. *Cancer Biol. Med.* **13**, 120–135 (2016).

66. Shih, I. M. *et al.* Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res.* **61**, 818–822 (2001).

67. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **3**, 730–737 (1997).

68. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

69. Canoy, R. J. *et al.* Specificity of cancer-related chromosomal translocations is linked to proximity after the DNA double-strand break and subsequent selection. *NAR Cancer* **5**, zcad049 (2023).

70. Shallis, R. M., Wang, R., Davidoff, A., Ma, X. & Zeidan, A. M. Epidemiology of acute myeloid leukemia: Recent progress and enduring challenges. *Blood Rev.* **36**, 70–87 (2019).

71. Yang, J. J., Park, T. S. & Wan, T. S. K. Recurrent Cytogenetic Abnormalities in Acute Myeloid Leukemia. *Methods Mol. Biol. Clifton NJ* **1541**, 223–245 (2017).

72. Freireich, E. J. *et al*. Refractory anermia, granulocytic hyperplasia of bone marrow, and a missing chromosome in marrow cells. A new clinical syndrome? *Clin. Res.* **12**:284 (1964).

73. Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).

74. Inaba, T., Honda, H. & Matsui, H. The enigma of monosomy 7. *Blood* **131**, 2891–2898 (2018).

75. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20007–20012 (2007).

76. Papaemmanuil, E., Döhner, H. & Campbell, P. J. Genomic Classification in Acute Myeloid Leukemia. *N. Engl. J. Med.* **375**, 900–901 (2016).

77. Haase, D. Cytogenetic features in myelodysplastic syndromes. *Ann. Hematol.* **87**, 515–526 (2008).

78. Schwartz, J. R. *et al.* The genomic landscape of pediatric myelodysplastic syndromes. *Nat. Commun.* **8**, 1557 (2017).

79. Smith, S. M. *et al.* Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: the University of Chicago series. *Blood* **102**, 43–52 (2003).

80. Jerez, A. *et al.* Loss of heterozygosity in 7q myeloid disorders: clinical associations and genomic pathogenesis. *Blood* **119**, 6109–6117 (2012).

81. McNerney, M. E. *et al.* The spectrum of somatic mutations in high-risk acute myeloid leukaemia with -7/del(7q). *Br. J. Haematol.* **166**, 550–556 (2014).

82. Mantadakis, E. *et al.* Transient monosomy 7: a case series in children and review of the literature. *Cancer* **85**, 2655–2661 (1999).

83. Edington, H. J. & Lowe, E. J. Spontaneous Remission of Monosomy 7 Six Years After Diagnosis. *J. Pediatr. Hematol. Oncol.* **41**, e177–e178 (2019).

84. Dimitriou, M. *et al.* Perturbed hematopoietic stem and progenitor cell hierarchy in myelodysplastic syndromes patients with monosomy 7 as the sole cytogenetic abnormality. *Oncotarget* **7**, 72685–72698 (2016).

85. Mori, M. *et al.* Genomics of deletion 7 and 7q in myeloid neoplasm: from pathogenic culprits to potential synthetic lethal therapeutic targets. *Leukemia* **37**, 2082–2093 (2023).

86. Wong, C. C., Traynor, D., Basse, N., Kay, R. R. & Warren, A. J. Defective ribosome assembly in Shwachman-Diamond syndrome. *Blood* **118**, 4305–4312 (2011).

87. Boocock, G. R. B. *et al.* Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat. Genet.* **33**, 97–101 (2003).

88. Rawls, A. S., Gregory, A. D., Woloszynek, J. R., Liu, F. & Link, D. C. Lentiviral-mediated RNAi inhibition of Sbds in murine hematopoietic progenitors impairs their hematopoietic potential. *Blood* **110**, 2414–2422 (2007).

89. Cordoba, I. *et al.* Better prognosis for patients with del(7q) than for patients with monosomy 7 in myelodysplastic syndrome. *Cancer* **118**, 127–133 (2012).

90. Hasle, H. *et al.* Monosomy 7 and deletion 7q in children and adolescents with acute myeloid leukemia: an international retrospective study. *Blood* **109**, 4641–4647 (2007).

91. Hussain, F. T. N. *et al.* Sole abnormalities of chromosome 7 in myeloid malignancies: spectrum, histopathologic correlates, and prognostic implications. *Am. J. Hematol.* **87**, 684–686 (2012).

92. Jacobs, K. B. *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* **44**, 651–658 (2012).

93. Laurie, C. C. *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **44**, 642–650 (2012).

94. Machiela, M. J. *et al.* Characterization of large structural genetic mosaicism in human autosomes. *Am. J. Hum. Genet.* **96**, 487–497 (2015).

95. Takahashi, K. *et al.* Copy number alterations detected as clonal hematopoiesis of indeterminate potential. *Blood Adv.* **1**, 1031–1036 (2017).

96. Loh, P.-R. *et al.* Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* **559**, 350–355 (2018).

97. Osorio, F. G. *et al.* Somatic Mutations Reveal Lineage Relationships and Age-Related Mutagenesis in Human Hematopoiesis. *Cell Rep.* **25**, 2308-2316.e4 (2018).

98. Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. E. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat. Commun.* **7**, 12484 (2016).

99. Jaiswal, S. & Ebert, B. L. Clonal hematopoiesis in human aging and disease. *Science* **366**, eaan4673 (2019).

100. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).

101. DeGregori, J. Challenging the axiom: does the occurrence of oncogenic mutations truly limit cancer development with age? *Oncogene* **32**, 1869–1875 (2013).

102. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).

103. Frank, S. A. Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proc. Natl. Acad. Sci. U. S. A.* **107 Suppl 1**, 1725–1730 (2010).

104. Rowley, J. D. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–293 (1973).

105. Biernaux, C., Loos, M., Sels, A., Huez, G. & Stryckmans, P. Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals. *Blood* **86**, 3118–3122 (1995).

106. Matioli, G. T. BCR-ABL insufficiency for the transformation of human stem cells into CML. *Med. Hypotheses* **59**, 588–589 (2002).

107. Wong, T. N. *et al.* Cellular stressors contribute to the expansion of hematopoietic clones of varying leukemic potential. *Nat. Commun.* **9**, 455 (2018).

108. Hsu, J. I. *et al.* PPM1D Mutations Drive Clonal Hematopoiesis in Response to Cytotoxic Chemotherapy. *Cell Stem Cell* **23**, 700-713.e6 (2018).

109. Muto, T. *et al.* Adaptive response to inflammation contributes to sustained myelopoiesis and confers a competitive advantage in myelodysplastic syndrome HSCs. *Nat. Immunol.* **21**, 535–545 (2020).

110. Meisel, M. *et al.* Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. *Nature* **557**, 580–584 (2018).

111. Gao, T. *et al.* Interplay between chromosomal alterations and gene mutations shapes the evolutionary trajectory of clonal hematopoiesis. *Nat. Commun.* **12**, 338 (2021).

112. Saiki, R. *et al.* Combined landscape of single-nucleotide variants and copy number alterations in clonal hematopoiesis. *Nat. Med.* **27**, 1239–1249 (2021).

113. Hosono, N. *et al.* Recurrent genetic defects on chromosome 7q in myeloid neoplasms. *Leukemia* **28**, 1348–1351 (2014).

114. Ernst, T. *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.* **42**, 722–726 (2010).

115. Takahashi, K. *et al.* Preleukaemic clonal haemopoiesis and risk of therapy-related myeloid neoplasms: a case-control study. *Lancet Oncol.* **18**, 100–111 (2017).

116. Zhang, L. *et al.* Leukemia-related chromosomal loss detected in hematopoietic progenitor cells of benzene-exposed workers. *Leukemia* **26**, 2494–2498 (2012).

117. Appelbaum, F. R. *et al.* Age and acute myeloid leukemia. *Blood* **107**, 3481–3485 (2006).

118. Zhang, Q. *et al.* Mutations in EZH2 are associated with poor prognosis for patients with myeloid neoplasms. *Genes Dis.* **6**, 276–281 (2019).

119. Le Beau, M. M. *et al.* Cytogenetic and molecular delineation of a region of chromosome 7 commonly deleted in malignant myeloid diseases. *Blood* **88**, 1930–1935 (1996).

120. Döhner, K. *et al.* Molecular cytogenetic characterization of a critical region in bands 7q35-q36 commonly deleted in malignant myeloid disorders. *Blood* **92**, 4031–4035 (1998).

121. McNerney, M. E., Godley, L. A. & Le Beau, M. M. Therapy-related myeloid neoplasms: when genetics and environment collide. *Nat. Rev. Cancer* **17**, 513–527 (2017).

122. Yoshizato, T. *et al.* Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. *N. Engl. J. Med.* **373**, 35–47 (2015).

123. Zink, F. *et al.* Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* **130**, 742–752 (2017).

124. Robertson, N. A. *et al.* Longitudinal dynamics of clonal hematopoiesis identifies gene-specific fitness effects. *Nat. Med.* **28**, 1439–1446 (2022).

125. Sansregret, L. & Nepveu, A. The multiple roles of CUX1: insights from mouse models and cell-based assays. *Gene* **412**, 84–94 (2008).

126. Aly, M. *et al.* Distinct clinical and biological implications of CUX1 in myeloid neoplasms. *Blood Adv.* **3**, 2164–2178 (2019).

127. Wong, C. C. *et al.* Inactivating CUX1 mutations promote tumorigenesis. *Nat. Genet.* **46**, 33–38 (2014).

128. AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov.* **7**, 818–831 (2017).

129. Rong Zeng, W. *et al.* Exon/intron structure and alternative transcripts of the CUTL1 gene. *Gene* **241**, 75–85 (2000).

130. Krishnan, M. *et al.* Genomic studies controvert the existence of the CUX1 p75 isoform. *Sci. Rep.* **12**, 151 (2022).

131. Aufiero, B., Neufeld, E. J. & Orkin, S. H. Sequence-specific DNA binding of individual cut repeats of the human CCAAT displacement/cut homeodomain protein. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7757–7761 (1994).

132. Bürglin, T. R. & Cassata, G. Loss and gain of domains during evolution of cut superclass homeobox genes. *Int. J. Dev. Biol.* **46**, 115–123 (2002).

133. Gillingham, A. K., Pfeifer, A. C. & Munro, S. CASP, the alternatively spliced product of the gene encoding the CCAAT-displacement protein transcription factor, is a Golgi membrane protein related to giantin. *Mol. Biol. Cell* **13**, 3761–3774 (2002).

134. Tufarelli, C., Fujiwara, Y., Zappulla, D. C. & Neufeld, E. J. Hair defects and pup loss in mice with targeted deletion of the first cut repeat domain of the Cux/CDP homeoprotein gene. *Dev. Biol.* **200**, 69–81 (1998).

135. An, N. *et al.* Gene dosage effect of CUX1 in a murine model disrupts HSC homeostasis and controls the severity and mortality of MDS. *Blood* **131**, 2682–2697 (2018).

136. Supper, E. *et al.* Cut-like homeobox 1 (CUX1) tumor suppressor gene haploinsufficiency induces apoptosis evasion to sustain myeloid leukemia. *Nat. Commun.* **12**, 2482 (2021).

137. Arthur, R. K., An, N., Khan, S. & McNerney, M. E. The haploinsufficient tumor suppressor, CUX1, acts as an analog transcriptional regulator that controls target

genes through distal enhancers that loop to target promoters. *Nucleic Acids Res.* **45**, 6350–6361 (2017).

138. Lievens, P. M., Donady, J. J., Tufarelli, C. & Neufeld, E. J. Repressor activity of CCAAT displacement protein in HL-60 myeloid leukemia cells. *J. Biol. Chem.* **270**, 12745–12750 (1995).

139. Liu, W. *et al*. CUX1 regulates human hematopoietic stem cell chromatin accessibility via the BAF complex. *Cell Rep.* **43**, 114227 (2024).

140. Pellagatti, A. *et al.* Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia* **24**, 756–764 (2010).

141. Venezia, T. A. *et al.* Molecular signatures of proliferation and quiescence in hematopoietic stem cells. *PLoS Biol.* **2**, e301 (2004).

142. Imgruet, M. K. *et al.* Loss of a 7q gene, CUX1, disrupts epigenetically driven DNA repair and drives therapy-related myeloid neoplasms. *Blood* **138**, 790–805 (2021).

143. Ayrapetov, M. K., Gursoy-Yuzugullu, O., Xu, C., Xu, Y. & Price, B. D. DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9169–9174 (2014).

144. Zhang, Y. *et al.* Histone H3K27 methylation modulates the dynamics of FANCD2 on chromatin to facilitate NHEJ and genome stability. *J. Cell Sci.* **131**, jcs215525 (2018).

145. Bolton, K. L. *et al.* Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat. Genet.* **52**, 1219–1226 (2020).

146. Jotte, M. R. M. & McNerney, M. E. The significance of CUX1 and chromosome 7 in myeloid malignancies. *Curr. Opin. Hematol.* **29**, 92–102 (2022).

147. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).

148. Piunti, A. & Shilatifard, A. The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat. Rev. Mol. Cell Biol.* **22**, 326–345 (2021).

149. van Mierlo, G., Veenstra, G. J. C., Vermeulen, M. & Marks, H. The Complexity of PRC2 Subcomplexes. *Trends Cell Biol.* **29**, 660–671 (2019).

150. O'Carroll, D. *et al.* The polycomb-group gene Ezh2 is required for early mouse development. *Mol. Cell. Biol.* **21**, 4330–4336 (2001).

151. Shumacher, A., Faust, C. & Magnuson, T. Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature* **383**, 250–253 (1996).

152. Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23**, 4061–4071 (2004).

153. Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491–502 (2008).

154. Sun, D. *et al.* Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* **14**, 673–688 (2014).

155. Kamminga, L. M. *et al.* The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. *Blood* **107**, 2170–2179 (2006).

156. Majewski, I. J. *et al.* Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. *PLoS Biol.* **6**, e93 (2008).

157. Mochizuki-Kashio, M. *et al.* Dependency on the polycomb gene Ezh2 distinguishes fetal from adult hematopoietic stem cells. *Blood* **118**, 6553–6561 (2011).

158. Xie, H. *et al.* Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* **14**, 68–80 (2014).

159. Basheer, F. *et al.* Contrasting requirements during disease evolution identify EZH2 as a therapeutic target in AML. *J. Exp. Med.* **216**, 966–981 (2019).

160. Neff, T. *et al.* Polycomb repressive complex 2 is required for MLL-AF9 leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5028–5033 (2012).

161. 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).

162. Yap, D. B. *et al.* Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* **117**, 2451–2459 (2011).

163. Béguelin, W. *et al.* EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* **23**, 677–692 (2013).

164. McCabe, M. T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108–112 (2012).

165. Knutson, S. K. *et al.* Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. *Mol. Cancer Ther.* **13**, 842–854 (2014).

166. Lund, K., Adams, P. D. & Copland, M. EZH2 in normal and malignant hematopoiesis. *Leukemia* **28**, 44–49 (2014).

167. Bejar, R. *et al.* Clinical effect of point mutations in myelodysplastic syndromes. *N. Engl. J. Med.* **364**, 2496–2506 (2011).

168. Nikoloski, G. *et al.* Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat. Genet.* **42**, 665–667 (2010).

169. Mochizuki-Kashio, M. *et al.* Ezh2 loss in hematopoietic stem cells predisposes mice to develop heterogeneous malignancies in an Ezh1-dependent manner. *Blood* **126**, 1172–1183 (2015).

170. Danis, E. *et al.* Ezh2 Controls an Early Hematopoietic Program and Growth and Survival Signaling in Early T Cell Precursor Acute Lymphoblastic Leukemia. *Cell Rep.* **14**, 1953–1965 (2016).

171. Muto, T. *et al.* Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. *J. Exp. Med.* **210**, 2627–2639 (2013).

172. Sashida, G. *et al.* Ezh2 loss promotes development of myelodysplastic syndrome but attenuates its predisposition to leukaemic transformation. *Nat. Commun.* **5**, 4177 (2014).

173. Sashida, G. *et al.* The loss of Ezh2 drives the pathogenesis of myelofibrosis and sensitizes tumor-initiating cells to bromodomain inhibition. *J. Exp. Med.* **213**, 1459–1477 (2016).

174. Göllner, S. *et al.* Loss of the histone methyltransferase EZH2 induces resistance to multiple drugs in acute myeloid leukemia. *Nat. Med.* **23**, 69–78 (2017).

175. Cui, K. *et al.* Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* **4**, 80–93 (2009).

176. Thorsteinsdottir, U. *et al.* Overexpression of HOXA10 in murine hematopoietic cells perturbs both myeloid and lymphoid differentiation and leads to acute myeloid leukemia. *Mol. Cell. Biol.* **17**, 495–505 (1997).

177. Khan, S. N. *et al.* Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. *Leukemia* **27**, 1301–1309 (2013).

178. Xu, F. *et al.* Genomic loss of EZH2 leads to epigenetic modifications and overexpression of the HOX gene clusters in myelodysplastic syndrome. *Oncotarget* **7**, 8119–8130 (2016).

179. O'Hagan, H. M., Mohammad, H. P. & Baylin, S. B. Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLoS Genet.* **4**, e1000155 (2008).

180. O'Hagan, H. M. *et al.* Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell* **20**, 606–619 (2011).

181. Campbell, S., Ismail, I. H., Young, L. C., Poirier, G. G. & Hendzel, M. J. Polycomb repressive complex 2 contributes to DNA double-strand break repair. *Cell Cycle Georget. Tex* **12**, 2675–2683 (2013).

182. Lutze, J., Wolfgeher, D. & Kron, S. J. *Global Epigenetic Analysis Reveals H3K27 Methylation as a Mediator of Double Strand Break Repair*. *BioRxiv,* (2021). doi:10.1101/2021.09.20.461136.

183. Meeks, J. J. & Shilatifard, A. Multiple Roles for the MLL/COMPASS Family in the Epigenetic Regulation of Gene Expression and in Cancer. *Annu. Rev. Cancer Biol.* **1**, 425–446 (2017).

184. Miller, T. *et al.* COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12902–12907 (2001).

185. Armstrong, S. A., Golub, T. R. & Korsmeyer, S. J. MLL-rearranged leukemias: insights from gene expression profiling. *Semin. Hematol.* **40**, 268–273 (2003).

186. Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25**, 15–30 (2007).

187. Dorighi, K. M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol. Cell* **66**, 568-576.e4 (2017).

188. Lee, J.-E. *et al.* H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife* **2**, e01503 (2013).

189. Arcipowski, K. M., Bulic, M., Gurbuxani, S. & Licht, J. D. Loss of Mll3 Catalytic Function Promotes Aberrant Myelopoiesis. *PloS One* **11**, e0162515 (2016).

190. Chen, C. *et al.* MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell* **25**, 652–665 (2014).

191. Chen, R. *et al.* Kmt2c mutations enhance HSC self-renewal capacity and convey a selective advantage after chemotherapy. *Cell Rep.* **34**, 108751 (2021).

192. McNeer, N. A. *et al.* Genetic mechanisms of primary chemotherapy resistance in pediatric acute myeloid leukemia. *Leukemia* **33**, 1934–1943 (2019).

193. Chang, A. *et al.* Recruitment of KMT2C/MLL3 to DNA Damage Sites Mediates DNA Damage Responses and Regulates PARP Inhibitor Sensitivity in Cancer. *Cancer Res.* **81**, 3358–3373 (2021).

194. Rampias, T. *et al.* The lysine-specific methyltransferase KMT2C/MLL3 regulates DNA repair components in cancer. *EMBO Rep.* **20**, e46821 (2019).

195. Zhang, X., Novera, W., Zhang, Y. & Deng, L.-W. MLL5 (KMT2E): structure, function, and clinical relevance. *Cell. Mol. Life Sci. CMLS* **74**, 2333–2344 (2017).

196. Mas-Y-Mas, S. *et al.* The Human Mixed Lineage Leukemia 5 (MLL5), a Sequentially and Structurally Divergent SET Domain-Containing Protein with No Intrinsic Catalytic Activity. *PloS One* **11**, e0165139 (2016).

197. Madan, V. *et al.* Impaired function of primitive hematopoietic cells in mice lacking the Mixed-Lineage-Leukemia homolog MLL5. *Blood* **113**, 1444–1454 (2009).

198. Heuser, M. *et al.* Loss of MLL5 results in pleiotropic hematopoietic defects, reduced neutrophil immune function, and extreme sensitivity to DNA demethylation. *Blood* **113**, 1432–1443 (2009).

199. Zhang, Y. *et al.* MLL5 contributes to hematopoietic stem cell fitness and homeostasis. *Blood* **113**, 1455–1463 (2009).

200. Tasdogan, A. *et al.* DNA Damage-Induced HSPC Malfunction Depends on ROS Accumulation Downstream of IFN-1 Signaling and Bid Mobilization. *Cell Stem Cell* **19**, 752–767 (2016).

201. Loukas, I. *et al.* Selective advantage of epigenetically disrupted cancer cells via phenotypic inertia. *Cancer Cell* **41**, 70-87.e14 (2023).

202. Narumi, S. *et al.* SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. *Nat. Genet.* **48**, 792–797 (2016).

203. Thomas, M. E. *et al.* Pediatric MDS and bone marrow failure-associated germline mutations in SAMD9 and SAMD9L impair multiple pathways in primary hematopoietic cells. *Leukemia* **35**, 3232–3244 (2021).

204. Wong, J. C. *et al.* Germline SAMD9 and SAMD9L mutations are associated with extensive genetic evolution and diverse hematologic outcomes. *JCI Insight* **3**, e121086, 121086 (2018).

205. Nagamachi, A. *et al.* Haploinsufficiency of SAMD9L, an endosome fusion facilitator, causes myeloid malignancies in mice mimicking human diseases with monosomy 7. *Cancer Cell* **24**, 305–317 (2013).

206. Singh, H. *et al.* Putative RNA-splicing gene LUC7L2 on 7q34 represents a candidate gene in pathogenesis of myeloid malignancies. *Blood Cancer J.* **3**, e117 (2013).

207. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).

208. Nguyen, H. D. *et al.* Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes. *Cancer Res.* **78**, 5363–5374 (2018).

209. Daniels, N. J. *et al.* Functional analyses of human LUC7-like proteins involved in splicing regulation and myeloid neoplasms. *Cell Rep.* **35**, 108989 (2021).

210. Jourdain, A. A. *et al.* Loss of LUC7L2 and U1 snRNP subunits shifts energy metabolism from glycolysis to OXPHOS. *Mol. Cell* **81**, 1905-1919.e12 (2021).

211. Sundaravel, S. *et al.* Reduced DOCK4 expression leads to erythroid dysplasia in myelodysplastic syndromes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6359-6368 (2015).

212. Baeten, J. T., Liu, W., Preddy, I. C., Zhou, N. & McNerney, M. E. CRISPR screening in human hematopoietic stem and progenitor cells reveals an enrichment for tumor suppressor genes within chromosome 7 commonly deleted regions. *Leukemia* **36**, 1421–1425 (2022).

213. Fink, D., Nebel, S., Aebi, S., Nehme, A. & Howell, S. Loss of DNA mismatch repair due to knockout of MSH2 or PMS2 results in resistance to cisplatin and carboplatin. *Int. J. Oncol.* **11**, 539–542 (1997).

214. Tang, M. *et al.* FOXK1 Participates in DNA Damage Response by Controlling 53BP1 Function. *Cell Rep.* **32**, 108018 (2020).

215. Elson, D. J. & Kolluri, S. K. Tumor-Suppressive Functions of the Aryl Hydrocarbon Receptor (AhR) and AhR as a Therapeutic Target in Cancer. *Biology* **12**, 526 (2023).

216. Boitano, A. E. *et al.* Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science* **329**, 1345–1348 (2010).

217. Ly, M. *et al.* Diminished AHR Signaling Drives Human Acute Myeloid Leukemia Stem Cell Maintenance. *Cancer Res.* **79**, 5799–5811 (2019).

218. Belin, B. J., Lee, T. & Mullins, R. D. DNA damage induces nuclear actin filament assembly by Formin -2 and Spire-½ that promotes efficient DNA repair. [corrected]. *eLife* **4**, e07735 (2015).

219. Schrank, B. R. *et al.* Nuclear ARP2/3 drives DNA break clustering for homology-directed repair. *Nature* **559**, 61–66 (2018).

220. Schmickel, R. D. Contiguous gene syndromes: a component of recognizable syndromes. *J. Pediatr.* **109**, 231–241 (1986).

221. Sharma, R., Lewis, S. & Wlodarski, M. W. DNA Repair Syndromes and Cancer: Insights Into Genetics and Phenotype Patterns. *Front. Pediatr.* **8**, 570084 (2020).

222. Wong, T. N. *et al.* Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* **518**, 552–555 (2015).

223. Diamond, B. *et al.* Tracking the evolution of therapy-related myeloid neoplasms using chemotherapy signatures. *Blood* **141**, 2359–2371 (2023).

224. Saxon, P. J. & Stanbridge, E. J. Transfer and selective retention of single specific human chromosomes via microcell-mediated chromosome transfer. *Methods Enzymol.* **151**, 313–325 (1987).

225. Kolishovski, G. *et al.* The JAX Synteny Browser for mouse-human comparative genomics. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* **30**, 353–361 (2019).

226. Wong, J. C. Y. *et al.* Use of chromosome engineering to model a segmental deletion of chromosome band 7q22 found in myeloid malignancies. *Blood* **115**, 4524–4532 (2010).

227. Wong, J. C. *et al.* Functional evidence implicating chromosome 7q22 haploinsufficiency in myelodysplastic syndrome pathogenesis. *eLife* **4**, e07839 (2015).

228. Wong, J. C. *et al.* 5G2 mutant mice model loss of a commonly deleted segment of chromosome 7q22 in myeloid malignancies. *Leukemia* (2024).

229. Rosenbauer, F. *et al.* Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat. Genet.* **36**, 624–630 (2004).

230. Kotini, A. G. *et al.* Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. *Nat. Biotechnol.* **33**, 646–655 (2015).

231. Kotini, A. G. & Papapetrou, E. P. Engineering of targeted megabase-scale deletions in human induced pluripotent stem cells. *Exp. Hematol.* **87**, 25–32 (2020).

232. Tothova, Z. *et al.* Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. *Cell Stem Cell* **21**, 547-555.e8 (2017).

233. Labuhn, M. *et al.* Mechanisms of Progression of Myeloid Preleukemia to Transformed Myeloid Leukemia in Children with Down Syndrome. *Cancer Cell* **36**, 123-138.e10 (2019).

234. Pick, M. *et al.* Stress-induced cholinergic signaling promotes inflammation-associated thrombopoiesis. *Blood* **107**, 3397–3406 (2006).

235. Soreq, H. *et al.* Antisense oligonucleotide inhibition of acetylcholinesterase gene expression induces progenitor cell expansion and suppresses hematopoietic apoptosis ex vivo. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7907–7911 (1994).

236. Meehan, T. F. *et al.* Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. *Nat. Genet.* **49**, 1231–1238 (2017).

237. McNerney, M. E. *et al.* CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently inactivated in acute myeloid leukemia. *Blood* **121**, 975–983 (2013).

238. Lindsley, R. C. *et al.* Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. *N. Engl. J. Med.* **376**, 536–547 (2017).

239. Zhang, J., Guo, J., Dzhagalov, I. & He, Y.-W. An essential function for the calcium-promoted Ras inactivator in Fcgamma receptor-mediated phagocytosis. *Nat. Immunol.* **6**, 911–919 (2005).

240. Poetsch, A. R. *et al.* RASA4 undergoes DNA hypermethylation in resistant juvenile myelomonocytic leukemia. *Epigenetics* **9**, 1252–1260 (2014).

241. Cheng, F. *et al.* RNA interference against mixed lineage leukemia 5 resulted in cell cycle arrest. *Int. J. Biochem. Cell Biol.* **40**, 2472–2481 (2008).

242. Damm, F. *et al.* Prognostic importance of histone methyltransferase MLL5 expression in acute myeloid leukemia. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **29**, 682–689 (2011).

243. Zhou, L. *et al.* Aberrant epigenetic and genetic marks are seen in myelodysplastic leukocytes and reveal Dock4 as a candidate pathogenic gene on chromosome 7q. *J. Biol. Chem.* **286**, 25211–25223 (2011).

244. Guo, D. *et al.* Autism-like social deficit generated by Dock4 deficiency is rescued by restoration of Rac1 activity and NMDA receptor function. *Mol. Psychiatry* **26**, 1505–1519 (2021).

245. Valiyaveettil, M. *et al.* Novel role of the muskelin-RanBP9 complex as a nucleocytoplasmic mediator of cell morphology regulation. *J. Cell Biol.* **182**, 727–739 (2008).

246. Heisler, F. F. *et al.* Muskelin regulates actin filament- and microtubule-based GABA(A) receptor transport in neurons. *Neuron* **70**, 66–81 (2011).

247. Bao, E. L. *et al.* Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature* **586**, 769–775 (2020).

248. Teyssier, C., Ou, C.-Y., Khetchoumian, K., Losson, R. & Stallcup, M. R. Transcriptional intermediary factor 1alpha mediates physical interaction and functional synergy between the coactivator-associated arginine methyltransferase 1 and glucocorticoid receptor-interacting protein 1 nuclear receptor coactivators. *Mol. Endocrinol. Baltim. Md* **20**, 1276–1286 (2006).

249. Tisserand, J. *et al.* Tripartite motif 24 (Trim24/Tif1α) tumor suppressor protein is a novel negative regulator of interferon (IFN)/signal transducers and activators of transcription (STAT) signaling pathway acting through retinoic acid receptor α (Rarα) inhibition. *J. Biol. Chem.* **286**, 33369–33379 (2011).

250. Li, C., Xin, H., Shi, Y. & Mu, J. Knockdown of TRIM24 suppresses growth and induces apoptosis in acute myeloid leukemia through downregulation of Wnt/GSK-3β/β-catenin signaling. *Hum. Exp. Toxicol.* **39**, 1725–1736 (2020).

251. Khetchoumian, K. *et al.* Loss of Trim24 (Tif1alpha) gene function confers oncogenic activity to retinoic acid receptor alpha. *Nat. Genet.* **39**, 1500–1506 (2007).

252. Allton, K. *et al.* Trim24 targets endogenous p53 for degradation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11612–11616 (2009).

253. Wang, Y. *et al.* TRIM24 is critical for the cellular response to DNA double-strand breaks through regulating the recruitment of MRN complex. *Oncogene* **42**, 586–600 (2023).

254. D'Orazi, G. *et al.* Homeodomain-interacting protein kinase-2 phosphorylates p53 at Ser 46 and mediates apoptosis. *Nat. Cell Biol.* **4**, 11–19 (2002).

255. Di Stefano, V., Rinaldo, C., Sacchi, A., Soddu, S. & D'Orazi, G. Homeodomain-interacting protein kinase-2 activity and p53 phosphorylation are critical events for cisplatin-mediated apoptosis. *Exp. Cell Res.* **293**, 311–320 (2004).

256. Li, X.-L. *et al.* Mutations of the HIPK2 gene in acute myeloid leukemia and myelodysplastic syndrome impair AML1- and p53-mediated transcription. *Oncogene* **26**, 7231–7239 (2007).

257. Sjölund, J., Pelorosso, F. G., Quigley, D. A., DelRosario, R. & Balmain, A. Identification of Hipk2 as an essential regulator of white fat development. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7373–7378 (2014).

258. Li, C. *et al.* The RNA-binding protein LUC7L2 mediates MITA/STING intron retention to negatively regulate innate antiviral response. *Cell Discov.* **7**, 46 (2021).

259. Dealy, M. J. *et al.* Loss of Cul1 results in early embryonic lethality and dysregulation of cyclin E. *Nat. Genet.* **23**, 245–248 (1999).

260. Piva, R. *et al.* In vivo interference with Skp1 function leads to genetic instability and neoplastic transformation. *Mol. Cell. Biol.* **22**, 8375–8387 (2002).

261. Sweeney, M. A. *et al.* The ubiquitin ligase Cullin-1 associates with chromatin and regulates transcription of specific c-MYC target genes. *Sci. Rep.* **10**, 13942 (2020).

262. Shimizu, T. *et al.* Loss of Ezh2 synergizes with JAK2-V617F in initiating myeloproliferative neoplasms and promoting myelofibrosis. *J. Exp. Med.* **213**, 1479–1496 (2016).

263. Bowler, T. G. *et al.* Misidentification of MLL3 and other mutations in cancer due to highly homologous genomic regions. *Leuk. Lymphoma* **60**, 3132–3137 (2019).

264. Olivieri, M. *et al.* A Genetic Map of the Response to DNA Damage in Human Cells. *Cell* **182**, 481-496.e21 (2020).

265. Findlay, S. *et al.* SHLD2/FAM35A co-operates with REV7 to coordinate DNA double-strand break repair pathway choice. *EMBO J.* **37**, e100158 (2018).

266. Wijdeven, R. H. *et al.* Genome-Wide Identification and Characterization of Novel Factors Conferring Resistance to Topoisomerase II Poisons in Cancer. *Cancer Res.* **75**, 4176–4187 (2015).

267. Goodspeed, A., Jean, A. & Costello, J. C. A Whole-genome CRISPR Screen Identifies a Role of MSH2 in Cisplatin-mediated Cell Death in Muscle-invasive Bladder Cancer. *Eur. Urol.* **75**, 242–250 (2019).

268. He, Y. J. *et al.* DYNLL1 binds to MRE11 to limit DNA end resection in BRCA1-deficient cells. *Nature* **563**, 522–526 (2018).

269. Liu, C., Banister, C. E. & Buckhaults, P. J. Spindle Assembly Checkpoint Inhibition Can Resensitize p53-Null Stem Cells to Cancer Chemotherapy. *Cancer Res.* **79**, 2392–2403 (2019).

270. Huang, D. *et al.* A highly annotated database of genes associated with platinum resistance in cancer. *Oncogene* **40**, 6395–6405 (2021).

271. Awah, C. U. *et al.* Ribosomal protein S11 influences glioma response to TOP2 poisons. *Oncogene* **39**, 5068–5081 (2020).

272. 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).

273. Oughtred, R. *et al.* The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci. Publ. Protein Soc.* **30**, 187–200 (2021).

274. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinforma. Oxf. Engl.* **27**, 1739–1740 (2011).

275. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).

276. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).

277. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

278. Gundry, M. C. *et al.* Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. *Cell Rep.* **17**, 1453–1461 (2016).

279. Brunetti, L., Gundry, M. C., Kitano, A., Nakada, D. & Goodell, M. A. Highly Efficient Gene Disruption of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. *J. Vis. Exp. JoVE* 57278 (2018).

280. Wilkinson, A. C., Ishida, R., Nakauchi, H. & Yamazaki, S. Long-term ex vivo expansion of mouse hematopoietic stem cells. *Nat. Protoc.* **15**, 628–648 (2020).

281. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).

282. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

283. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

284. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26**, 139–140 (2010).

285. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

286. Gyori, B. M., Venkatachalam, G., Thiagarajan, P. S., Hsu, D. & Clement, M.-V. OpenComet: an automated tool for comet assay image analysis. *Redox Biol.* **2**, 457–465 (2014).

287. Vasudevan, A. *et al.* Aneuploidy as a promoter and suppressor of malignant growth. *Nat. Rev. Cancer* **21**, 89–103 (2021).

288. Arber, D. A. *et al.* International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood* **140**, 1200–1228 (2022).

289. Khoury, J. D. *et al.* The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia* **36**, 1703–1719 (2022).

290. Hu, D. *et al.* The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol. Cell. Biol.* **33**, 4745–4754 (2013).

291. Kahn, J. D. *et al.* PPM1D-truncating mutations confer resistance to chemotherapy and sensitivity to PPM1D inhibition in hematopoietic cells. *Blood* **132**, 1095–1105 (2018).

292. Challen, G. A., Boles, N., Lin, K. K.-Y. & Goodell, M. A. Mouse hematopoietic stem cell identification and analysis. *Cytom. Part J. Int. Soc. Anal. Cytol.* **75**, 14–24 (2009).

293. Halik, A. *et al.* Genomic Characterization of Acute Myeloid Leukemia with Aberrations of Chromosome 7: A Multinational Cohort of 523 Patients. *Blood* **142**, 63–63 (2023).

294. Kakinuma, S. *et al.* Frequent retention of heterozygosity for point mutations in p53 and Ikaros in N-ethyl-N-nitrosourea-induced mouse thymic lymphomas. *Mutat. Res.* **572**, 132–141 (2005).

295. Döhner, H. *et al.* Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood* **140**, 1345–1377 (2022).

296. Guryanova, O. A. *et al.* DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nat. Med.* **22**, 1488–1495 (2016).

297. Helsmoortel, H. H. *et al.* LIN28B overexpression defines a novel fetal-like subgroup of juvenile myelomonocytic leukemia. *Blood* **127**, 1163–1172 (2016).

298. Mozzetta, C. *et al.* The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol. Cell* **53**, 277–289 (2014).

299. Halim, V. A. *et al.* Doxorubicin-induced DNA Damage Causes Extensive Ubiquitination of Ribosomal Proteins Associated with a Decrease in Protein Translation. *Mol. Cell. Proteomics MCP* **17**, 2297–2308 (2018).

300. Silva, E. & Ideker, T. Transcriptional responses to DNA damage. *DNA Repair* **79**, 40–49 (2019).

301. Kinner, A., Wu, W., Staudt, C. & Iliakis, G. Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Res.* **36**, 5678–5694 (2008).

302. Ginjala, V. *et al.* Protein-lysine methyltransferases G9a and GLP1 promote responses to DNA damage. *Sci. Rep.* **7**, 16613 (2017).

303. Zhao, X. *et al.* Single-cell RNA-seq reveals a distinct transcriptome signature of aneuploid hematopoietic cells. *Blood* **130**, 2762–2773 (2017).

304. Liao, Y. *et al.* Inhibition of EZH2 transactivation function sensitizes solid tumors to genotoxic stress. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2105898119 (2022).

305. Challen, G. A. & Goodell, M. A. Clonal hematopoiesis: mechanisms driving dominance of stem cell clones. *Blood* **136**, 1590–1598 (2020).

306. Fuster, J. J. *et al.* Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**, 842–847 (2017).

307. Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N. Engl. J. Med.* **377**, 111–121 (2017).

308. Jaiswal, S. Clonal hematopoiesis and nonhematologic disorders. *Blood* **136**, 1606–1614 (2020).

309. Midic, D. *et al.* Prevalence and dynamics of clonal hematopoiesis caused by leukemia-associated mutations in elderly individuals without hematologic disorders. *Leukemia* **34**, 2198–2205 (2020).

310. Wong, T. N. *et al.* Rapid expansion of preexisting nonleukemic hematopoietic clones frequently follows induction therapy for de novo AML. *Blood* **127**, 893–897 (2016).

311. Coombs, C. C. *et al.* Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell* **21**, 374-382.e4 (2017).

312. Chen, C.-W. *et al.* SRCAP mutations drive clonal hematopoiesis through epigenetic and DNA repair dysregulation. *Cell Stem Cell* **30**, 1503-1519.e8 (2023).

313. Ippolito, M. R. *et al.* Gene copy-number changes and chromosomal instability induced by aneuploidy confer resistance to chemotherapy. *Dev. Cell* **56**, 2440-2454.e6 (2021).

314. Florez, M. A. *et al.* Clonal hematopoiesis: Mutation-specific adaptation to environmental change. *Cell Stem Cell* **29**, 882–904 (2022).

315. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31 (2011).

316. Jeong, M. *et al.* Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. *Cell Rep.* **23**, 1–10 (2018).

317. Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11–24 (2011).

318. 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).

319. Dawoud, A. A. Z., Tapper, W. J. & Cross, N. C. P. Clonal myelopoiesis in the UK Biobank cohort: ASXL1 mutations are strongly associated with smoking. *Leukemia* **34**, 2660–2672 (2020).

320. Ramanathan, G. *et al.* Cigarette smoke stimulates clonal expansion of Jak2V617F and Tet2-/- cells. *Front. Oncol.* **13**, 1210528 (2023).

321. Bogeska, R. *et al.* Inflammatory exposure drives long-lived impairment of hematopoietic stem cell self-renewal activity and accelerated aging. *Cell Stem Cell* **29**, 1273-1284.e8 (2022).

322. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat. Rev. Endocrinol.* **14**, 576–590 (2018).

323. Avagyan, S. *et al.* Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis. *Science* **374**, 768–772 (2021).

324. SanMiguel, J. M. *et al.* Distinct Tumor Necrosis Factor Alpha Receptors Dictate Stem Cell Fitness versus Lineage Output in Dnmt3a-Mutant Clonal Hematopoiesis. *Cancer Discov.* **12**, 2763–2773 (2022).

325. Higa, K. C. *et al.* Chronic interleukin-1 exposure triggers selection for Cebpa-knockout multipotent hematopoietic progenitors. *J. Exp. Med.* **218**, e20200560 (2021).

326. Cai, Z. *et al.* Inhibition of Inflammatory Signaling in Tet2 Mutant Preleukemic Cells Mitigates Stress-Induced Abnormalities and Clonal Hematopoiesis. *Cell Stem Cell* **23**, 833-849.e5 (2018).

327. Hormaechea-Agulla, D. *et al.* Chronic infection drives Dnmt3a-loss-of-function clonal hematopoiesis via IFNγ signaling. *Cell Stem Cell* **28**, 1428-1442.e6 (2021).

328. Richart, L. & Margueron, R. Drugging histone methyltransferases in cancer. *Curr. Opin. Chem. Biol.* **56**, 51–62 (2020).

329. Laugesen, A., Højfeldt, J. W. & Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell* **74**, 8–18 (2019).

330. Højfeldt, J. W. *et al.* Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nat. Struct. Mol. Biol.* **25**, 225–232 (2018).

331. Ntziachristos, P. *et al.* Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat. Med.* **18**, 298–301 (2012).

332. De Raedt, T. *et al.* PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies. *Nature* **514**, 247–251 (2014).

333. Farquhar, K. S. *et al.* Role of network-mediated stochasticity in mammalian drug resistance. *Nat. Commun.* **10**, 2766 (2019).

334. Piunti, A. *et al.* Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. *Nat. Med.* **23**, 493–500 (2017).

335. Faraoni, I. & Graziani, G. Role of BRCA Mutations in Cancer Treatment with Poly(ADP-ribose) Polymerase (PARP) Inhibitors. *Cancers* **10**, 487 (2018).

336. Giudice, E. *et al.* PARP Inhibitors Resistance: Mechanisms and Perspectives. *Cancers* **14**, 1420 (2022).

337. Li, H. *et al.* PARP inhibitor resistance: the underlying mechanisms and clinical implications. *Mol. Cancer* **19**, 107 (2020).

338. Pal, R. *et al.* CUX2 protein functions as an accessory factor in the repair of oxidative DNA damage. *J. Biol. Chem.* **290**, 22520–22531 (2015).

339. Ramdzan, Z. M. *et al.* RAS transformation requires CUX1-dependent repair of oxidative DNA damage. *PLoS Biol.* **12**, e1001807 (2014).

340. Ramdzan, Z. M. *et al.* The function of CUX1 in oxidative DNA damage repair is needed to prevent premature senescence of mouse embryo fibroblasts. *Oncotarget* **6**, 3613–3626 (2015).

341. Morice, P.-M. *et al.* Myelodysplastic syndrome and acute myeloid leukaemia in patients treated with PARP inhibitors: a safety meta-analysis of randomised controlled trials and a retrospective study of the WHO pharmacovigilance database. *Lancet Haematol.* **8**, e122–e134 (2021).

342. Oliveira, J. L., Greipp, P. T., Rangan, A., Jatoi, A. & Nguyen, P. L. Myeloid malignancies in cancer patients treated with poly(ADP-ribose) polymerase (PARP) inhibitors: a case series. *Blood Cancer J.* **12**, 11 (2022).

343. Nojadeh, J. N., Behrouz Sharif, S. & Sakhinia, E. Microsatellite instability in colorectal cancer. *EXCLI J.* **17**, 159–168 (2018).

344. Crasta, K. *et al.* DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**, 53–58 (2012).

345. Ross, J. B. *et al.* Depleting myeloid-biased haematopoietic stem cells rejuvenates aged immunity. *Nature* **628**, 162–170 (2024).

346. Hillier, L. W. *et al.* The DNA sequence of human chromosome 7. *Nature* **424**, 157–164 (2003).

347. Gilbert, F. Chromosome 7. *Genet. Test.* **6**, 141–161 (2002).

348. Wang, Z. *et al.* Splicing factor BUD31 promotes ovarian cancer progression through sustaining the expression of anti-apoptotic BCL2L12. *Nat. Commun.* **13**, 6246 (2022).

349. Zhang, L. Q. *et al.* Metabolic and molecular insights into an essential role of nicotinamide phosphoribosyltransferase. *Cell Death Dis.* **8**, e2705 (2017).

350. Ohashi, R. *et al.* Allele Loss and Reduced Expression of CYCLOPS Genes is a Characteristic Feature of Chromophobe Renal Cell Carcinoma. *Transl. Oncol.* **12**, 1131–1137 (2019).

351. Eldfors, S. *et al.* Monosomy 7/del(7q) cause sensitivity to inhibitors of nicotinamide phosphoribosyltransferase in acute myeloid leukemia. *Blood Adv.* **8**, 1621–1633 (2024).

352. An, N. *et al.* Oncogenic RAS promotes leukemic transformation of CUX1-deficient cells. *Oncogene* **42**, 881–893 (2023).

353. Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019).

354. Blanpain, C. & Fuchs, E. Epidermal stem cells of the skin. *Annu. Rev. Cell Dev. Biol.* **22**, 339–373 (2006).

355. Pinho, S. & Frenette, P. S. Haematopoietic stem cell activity and interactions with the niche. *Nat. Rev. Mol. Cell Biol.* **20**, 303–320 (2019).

356. Morrison, S. J. & Kimble, J. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* **441**, 1068–1074 (2006).

357. McCracken, J. M. & Allen, L.-A. H. Regulation of human neutrophil apoptosis and lifespan in health and disease. *J. Cell Death* **7**, 15–23 (2014).

358. Mehanna, R. A. *et al.* Cardiac stem cells: Current knowledge and future prospects. *World J. Stem Cells* **14**, 1–40 (2022).

359. Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P. & Abkowitz, J. L. The replication rate of human hematopoietic stem cells in vivo. *Blood* **117**, 4460–4466 (2011).

360. Lee-Six, H. *et al.* Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, 473–478 (2018).

361. Zhang, B. & Hsu, Y.-C. Emerging roles of transit-amplifying cells in tissue regeneration and cancer. *Wiley Interdiscip. Rev. Dev. Biol.* **6**, (2017).

362. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).

363. Lendahl, U., Zimmerman, L. B. & McKay, R. D. CNS stem cells express a new class of intermediate filament protein. *Cell* **60**, 585–595 (1990).

364. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).

365. Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell* **1**, 218–229 (2007).

366. Haas, S., Trumpp, A. & Milsom, M. D. Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell* **22**, 627–638 (2018).

367. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381 (2020).

368. Rodriguez-Fraticelli, A. E. & Camargo, F. Systems analysis of hematopoiesis using single-cell lineage tracing. *Curr. Opin. Hematol.* **28**, 18–27 (2021).

369. Macaulay, I. C. *et al.* Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep.* **14**, 966–977 (2016).

370. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).

371. Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61**, 1329–1337 (1990).

372. Fuchs, E. & Horsley, V. Ferreting out stem cells from their niches. *Nat. Cell Biol.* **13**, 513–518 (2011).

373. Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).

374. Radhakrishnan, P., Basma, H., Klinkebiel, D., Christman, J. & Cheng, P.-W. Cell type-specific activation of the cytomegalovirus promoter by dimethylsulfoxide and 5-aza-2'-deoxycytidine. *Int. J. Biochem. Cell Biol.* **40**, 1944–1955 (2008).

375. Das, A. T., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Curr. Gene Ther.* **16**, 156–167 (2016).

376. Foudi, A. *et al.* Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nat. Biotechnol.* **27**, 84–90 (2009).

377. Challen, G. A. & Goodell, M. A. Promiscuous expression of H2B-GFP transgene in hematopoietic stem cells. *PloS One* **3**, e2357 (2008).

378. Sugimura, R. *et al.* Noncanonical Wnt signaling maintains hematopoietic stem cells in the niche. *Cell* **150**, 351–365 (2012).

379. Walter, D. *et al.* Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature* **520**, 549–552 (2015).

380. Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H. & Moore, K. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell* **167**, 1296-1309.e10 (2016).

381. Cabezas-Wallscheid, N. *et al.* Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. *Cell* **169**, 807-823.e19 (2017).

382. Kokkaliaris, K. D. *et al.* Adult blood stem cell localization reflects the abundance of reported bone marrow niche cell types and their combinations. *Blood* **136**, 2296–2307 (2020).

383. Morcos, M. N. F. *et al.* Continuous mitotic activity of primitive hematopoietic stem cells in adult mice. *J. Exp. Med.* **217**, e20191284 (2020).

384. Mathieson, T. *et al.* Systematic analysis of protein turnover in primary cells. *Nat. Commun.* **9**, 689 (2018).

385. Ranzani, M., Annunziato, S., Adams, D. J. & Montini, E. Cancer gene discovery: exploiting insertional mutagenesis. *Mol. Cancer Res. MCR* **11**, 1141–1158 (2013).

386. Baum, C. *et al.* Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. *Mol. Ther. J. Am. Soc. Gene Ther.* **9**, 5–13 (2004).

387. Hacein-Bey-Abina, S. *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415–419 (2003).

388. Clark, A. J. *et al.* Chromosomal position effects and the modulation of transgene expression. *Reprod. Fertil. Dev.* **6**, 589–598 (1994).

389. Feng, Y. Q., Lorincz, M. C., Fiering, S., Greally, J. M. & Bouhassira, E. E. Position effects are influenced by the orientation of a transgene with respect to flanking chromatin. *Mol. Cell. Biol.* **21**, 298–309 (2001).

390. Fletcher, S. J. qPCR for quantification of transgene expression and determination of transgene copy number. *Methods Mol. Biol. Clifton NJ* **1145**, 213–237 (2014).

391. Hottentot, Q. P., van Min, M., Splinter, E. & White, S. J. Targeted Locus Amplification and Next-Generation Sequencing. *Methods Mol. Biol. Clifton NJ* **1492**, 185–196 (2017).

392. Kosambi, D. D. The Estimation of Map Distances from Recombination Values. in *D.D. Kosambi* (ed. Ramaswamy, R.) 125–130 (Springer India, New Delhi, 2016). doi:10.1007/978-81-322-3676-4_16.