NOVEL APPROACHES FOR *IN VITRO* AND *IN VIVO* MANIPULATION OF THE MOUSE GENOME

by

Rui Gao

A Dissertation

Presented to the Faculty of the Louis V. Gerstner, Jr.

Graduate School of Biomedical Sciences,

Memorial Sloan-Kettering Cancer Center

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

New York, NY

December, 2021

_____	_____
Andrea Ventura, MD, PhD Dissertation Mentor	Date

Dedication

This thesis is dedicated to my mother, Wan Qiong, and my partner, Michael; their unwavering support, from near and far, has carried me through my PhD journey.

Abstract

Cancer is an umbrella term for a body of distinct but related diseases that are characterized by changes in the genome that result in the uncontrolled growth and proliferation of cells. In the last decade, large-scale next-generation sequencing efforts have illuminated the genetic landscape of multiple types of cancers, revealing the chromosomal rearrangements that underlie gene expression changes that drive tumorigenesis. The wealth of data generated by next-generation sequencing efforts must be complemented by experimental studies to not only identify which genes are *bona fide* players in cancer, but also to better understand the mechanisms by which various genomic alterations lead to the perturbation of normal gene expression and function.

The discovery and development of CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR associated protein 9) gene editing technology has been a boon to the field of cancer research. Initially discovered as an immune mechanism in bacteria, it has been harnessed to precisely, efficiently, and easily manipulate the genome, which has greatly facilitated the ability to model and characterize genomic alterations found in human cancers. Importantly, CRISPR/Cas9 has also allowed both germline and somatically edited mouse models to be generated within accelerated timelines.

In this thesis, we employ CRISPR/Cas9-based methods to model different types of chromosomal alterations that are found in cancers. The bulk of this thesis

centers on using CRISPR/Cas9-based methods to model the formation of circular extrachromosomal DNAs (ecDNAs) harboring oncogenes, which are found in multiple cancer types and thought to be a mechanism for oncogene amplification. We designed a strategy that exploits Cre-Lox mediated circularization and use it to target the *Myc* and *Egfr* oncogenes both *in vitro* in human and mouse cell lines and *in vivo* by engineering novel germline mouse models. We show that our system is able to induce ecDNA formation and can allow us to observe the behavior of ecDNAs over time and in response to selective pressure. Using our novel germline mouse models, we can conditionally induce the formation of *Myc*- and *Egfr*-ecDNAs and anticipate that we will be able to use these models to address the question of whether oncogene-carrying ecDNAs have transformative potential in a spatiotemporally regulated manner.

I also investigate the effects of deleting a CTCF-binding site within a minimally deleted region in B-cell chronic lymphocytic leukemia (B-CLL) which putatively acts as a boundary between functional gene neighborhoods to prevent crosstalk between enhancers and target genes located in different topologically-associating domains (TADs). By generating an allelic series of isogenic human cell lines with or without deletion of the CTCF-binding site and interrogating the resulting gene expression profiles by RNA-sequencing, I test the hypothesis that removal of the CTCF-binding site can lead to local gene expression changes, while at the same time probing for gene expression changes that may result from the deletion of other elements within the minimally deleted region. While I did not find

local gene expression changes linked to the deletion of the CTCF-binding site, I identify changes in gene candidates that would support further investigation of other elements within the region.

Finally, I explore alternative methods of delivering CRISPR/Cas9 components *in vivo,* in recognition of the fact that some organs, in particular the gastrointestinal tract, are inaccessible for somatic gene editing through conventional viral delivery of CRISPR. I show that an engineered strain of *E. coli* can deliver CRISPR/Cas9 to mammalian cells and induce gene editing at desired loci *in vitro*, and further carry out *in vivo* experiments to test the feasibility of bacteria-mediated delivery to the gastrointestinal tract of mice.

Biographical Sketch

Rui Gao was born on May 4, 1991, in Hubei, China. At age two, she and her parents moved to Singapore, where they later became naturalized citizens. She attended the Swiss Cottage Primary School in Singapore until age ten, when she and her parents moved to Beijing, China. There, she attended the International School of Beijing from the fourth grade onwards until she graduated from high school in 2010 with an International Baccalaureate Diploma. Following high school, she decided to pursue her joint interest in chemistry and biology with a biochemistry degree from Imperial College, London. During her time at Imperial, she gained her first research experience working in the lab of Dr. Chris Bakal at the Institute of Cancer Research, investigating novel regulators and interactors of mitotic spindle assembly using high-throughput, combinatorial RNAi screens in *Drosophila* cells. After graduating with a Bachelor of Science degree in Biochemistry in 2014, she enrolled in the Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences at Memorial Sloan Kettering in 2015 and joined the lab of Dr. Andrea Ventura in 2017 to model genomic changes observed in human cancers in mice using CRISPR/Cas9 technology.

Acknowledgements

I extend my sincerest gratitude to Dr. Andrea Ventura, my PhD adviser, for opening the doors of his laboratory to me at the start of my PhD, when I was briefly without a research home. He took a chance on me, and since then his support has never yielded. His belief in me has enabled me to rise to challenges I never would have otherwise had the courage to set for myself, and this has played an instrumental role in my growth as a scientist and as a person in the last several years. In addition, Andrea's generosity of spirit has made it possible to weather the storms of research. His office door has always been open, and he has always made time to discuss latest results or to offer support and encouragement, even if it meant momentarily putting aside his own work (or lunch!). His genuine care for everyone in his lab is palpable. In short, I could not have asked for a better mentor, and am proud to have trained under him.

I would also like to thank my committee members, Drs. Maria Jasin and Robert Benezra, for their advice and support over the years. Their scientific insights into my projects have greatly strengthened them and I am honored to have been able to discuss my research with such accomplished scientists. I have especially benefited from more frequent scientific discussions with Dr. Benezra during our joint lab meetings. I am also grateful to Dr. Omar Abdel-Wahab for agreeing to chair my defense committee, and for readily connecting me with postdocs in his lab when I was seeking technical help during my CLL project. I

would also like to thank Dr. Luke Dow for agreeing to be my external examiner, and for lending his time to discuss my bacteria project, and even attending one of my committee meetings to further contribute his insight.

In addition, I would like to thank all current and former members of the Ventura lab, who have made the lab such a wonderful place to work. During their time as former postdocs, Drs. Ram Kannan, Turgut Dogruluk, and Joana Vidigal lent me their time and expertise when I faced research challenges. I would particularly like to thank Ram for his advice and encouragement on career-related matters. I have also benefited greatly from the expertise of current postdocs in the lab, Drs. Minsi Zhang and Gaspare La Rocca. It has been a privilege to have learned from such generous and excellent scientists. In particular, I have been inspired by Minsi's breadth of knowledge and work ethic and am very grateful to have learned from her, both in terms of technical skills and through scientific discussions. I have no doubt that she will one day be an excellent mentor in her own lab. A special thank you goes to Dr. Davide Pradella, the newest postdoc in the lab. Though we have only had the chance to collaborate for a short amount of time, it has been truly a joy to work alongside him; his passion, thoughtfulness, and scientific rigor has already driven our project forward and I know will continue to do so. My thanks also goes to Dr. Xiaoyi Li, a recent PhD graduate from the lab, who has shared advice on experiments over the years, and who contributed her computational know-how by performing genome alignment of the RNA-sequencing data for my CLL project. I would like to acknowledge Chiara Mastroleo, Kevin Chen,

and Paul Ogrodowski for their help with experiments, maintaining mouse colonies, and for keeping the lab in good shape over the years. I also extend my thanks to Sue Woube, the lab's administrative assistant, for her help with ordering reagents and scheduling meetings.

I am tremendously grateful to the collaborators with whom I've had the pleasure of working. The technical expertise of Dr. Gouri Nanjangud and Marta Lisi from the Molecular Cytogenetics Core have allowed us to make great inroads in the ecDNA project, and I must especially thank them for their generosity with regards to their time, reagents, and resources. I am also grateful for our collaboration with the Mouse Genetics Core, in particular Drs. Yas Furuta and Joe Giacalone, whose technical assistance and support have been invaluable in generating our mouse strains. Members of the Flow Cytometry Core have also provided key technical assistance and guidance over the years. In sum, I have been fortunate to have had access to such top-notch facilities at MSKCC.

Next, I would like to thank every administrative member of the Gerstner Sloan Kettering Graduate School. Their commitment to each graduate student under their care has truly made GSK feel like a family. Deans Dr. Ken Marians and Dr. Michael Overholtzer, Associate Dean Dr. Linda Burnley, and Assistant Dean Dr. Thomas Magaldi have exemplified true leadership and support with all facets of students' lives, and registrar David McDonagh, program coordinator Stacy De La Cruz, and admissions coordinator Julie Masen have helped students navigate their PhD journeys. I have personally benefited from their assistance and support

over the years and am very grateful. I would also like to thank my fellow GSK classmates, past and present, for being such a terrific group of individuals; it was always a pleasure to interact with them, in and outside of research.

Outside of research, my life has been filled with people who have made my time in New York something on which I'll look back fondly. This PhD journey would have been impossible without my friends. I am so lucky to have found some of my closest ones within the graduate school. Space will not permit me to adequately thank everyone, but I would like to acknowledge (Dr.) Amy Shyu, Miguel de Jesus, and Mollie Chipman for their friendship and in some cases, cat-sitting duties! To Paige Arnold, Gemma Regan-Mochrie, and (Dr.) Emily Kansler: there are not enough words to express my love and gratitude for you three. Our friendship has buoyed me not only through my PhD, but also through life. No matter where our lives take us, I know that we'll still be gathering in old age, catching up with each other and chatting away, albeit in rocking chairs.

My family has persistently supported me from abroad for the last decade, and their unconditional love has sustained me from afar. I am eternally grateful that they have given me free reign to pursue my goals so far from home, but I know that they are never too far away. 瑞儿想你们。

Last but not least, my endless gratitude goes to my partner, Dr. Michael Hankinson, who has been, and continues to be, my bedrock. His unwaning love and support is my north star. I look forward to building our shared life together, with Kenzie by our side. 我爱你！

Table of Contents

LIST OF TABLES xv

LIST OF FIGURES xvi

LIST OF ABBREVIATIONS xix

1 CHAPTER 1: INTRODUCTION 1

1.1 CHROMOSOMAL STRUCTURAL CHANGES ARE IMPORTANT DRIVERS OF CANCER 1

1.1.1 COMPLEXITY AND DIVERSITY OF CANCER-ASSOCIATED CHROMOSOMAL STRUCTURAL CHANGES 3

1.2 CHROMOSOMAL STRUCTURAL INSTABILITY (CSI) CAN BE CAUSED BY IMPROPER REPAIR OF DNA DOUBLE STRAND BREAKS (DSBS) 15

1.2.1 MAJOR SOURCES OF PATHOLOGICAL DSBS 17

1.2.2 PARADOXICAL ROLES OF DNA DAMAGE REPAIR PATHWAYS IN CANCER FORMATION AND DEVELOPMENT 18

1.2.3 NON-HOMOLOGOUS END JOINING (NHEJ) 18

1.2.4 HOMOLOGOUS RECOMBINATION (HR) 23

1.2.5 ALTERNATIVE DSB REPAIR PATHWAYS 27

1.3 HISTORICAL PERSPECTIVE ON MODELING CHROMOSOMAL REARRANGEMENTS IN MOUSE 29

1.3.1 THE USE OF TRANSGENIC ANIMALS TO STUDY CHROMOSOMAL REARRANGEMENTS 31

1.3.2 HOMOLOGOUS RECOMBINATION-MEDIATED TARGETING STRATEGIES 33

1.3.3 CRE-LOXP-MEDIATED CHROMOSOMAL ENGINEERING 33

1.4 CRISPR-BASED METHODS OF MODELING CHROMOSOMAL REARRANGEMENTS 35

1.4.1 THE DISCOVERY AND ADAPTATION OF BACTERIAL CRISPR SYSTEMS HAS REVOLUTIONIZED THE FIELD OF GENETIC RESEARCH 35

1.4.2 THE PROMISE OF CRISPR FOR *IN VIVO* MODELING OF CHROMOSOMAL REARRANGEMENTS 37

1.4.3 LIMITATIONS OF CRISPR 40

1.5 OBJECTIVES OF THIS THESIS 44

2 CHAPTER 2: MATERIALS AND METHODS 46

2.1 CELL CULTURE AND DRUG TREATMENT. 46

2.2 CLONING AND PLASMID CONSTRUCTION. 46

2.3 CASSETTE TARGETING. 47

2.4 ADENOVIRAL TRANSDUCTION OF CRE RECOMBINASE. 49

2.5 GFP EXPRESSION ANALYSIS AND ISOLATION OF GFP-POSITIVE CELLS. 50

2.6 FLUORESCENCE IN SITU HYBRIDIZATION (FISH) ANALYSIS. 50

2.7 SOUTHERN BLOTTING. 52

2.8 QUANTITATIVE PCR. 52
2.9 ZYGOTE ELECTROPORATION. 53
2.10 GENERATION OF NOVEL FLOXED MOUSE STRAINS. 54
2.11 STATISTICAL ANALYSIS AND QUANTIFICATION. 54

3 CHAPTER 3: *IN VITRO* MODELING OF ECDNA 55

3.1 INTRODUCTION 55
3.2 RESULTS 60
3.2.1 A SPLIT-GFP STRATEGY FOR MODELING CIRCULARIZATION 60
3.2.2 GENERATION AND TESTING OF CASSETTES 62
3.2.3 GENERATING DOUBLE-TARGETED CLONES FLANKING THE *EGFR* LOCUS 64
3.2.4 CRE-INDUCED RECOMBINATION LEADS TO RECONSTITUTION OF THE GFP REPORTER 67
3.2.5 GFP-POSITIVE CELLS LOSE GFP EXPRESSION OVER TIME AND INCREASE GFP EXPRESSION IN RESPONSE TO SELECTIVE PRESSURE 71
3.2.6 CRE TREATMENT LEADS TO DIFFERENT KARYOTYPE OUTCOMES IN DOUBLE-TARGETED HCT116 CLONES 74
3.2.7 SEQUENTIAL SORTING OF GFP-POSITIVE D9 CELLS SELECTS FOR TANDEM DUPLICATION 80
3.2.8 CRE-MEDIATED RECOMBINATION LEADS TO FORMATION AND MAINTENANCE OF ECDNAS IN MURINE *P53-/-* ADULT NEURAL STEM CELLS 80
3.3 DISCUSSION 86

4 CHAPTER 4: *IN VIVO* MODELING OF ECDNAS 91

4.1 INTRODUCTION 91
4.2 RESULTS 93
4.2.1 MODELING *EGFR*- AND *MYC*-ECDNA IN VIVO 93
4.2.2 GENERATION OF NOVEL MOUSE MODELS VIA DIRECT ZYGOTE ELECTROPORATION OF CRISPR COMPONENTS 95
4.2.3 GENERATION OF DOUBLE CASSETTE-TARGETED MOUSE EMBRYONIC STEM CELLS BY SEQUENTIAL TARGETING 99
4.2.4 CRE-INDUCED RECOMBINATION LEADS TO RECONSTITUTION OF THE GFP REPORTER IN MESCS 99
4.2.5 CONFIRMATION OF ES CELL TARGETING BY SOUTHERN BLOT ANALYSIS 100
4.2.6 RECOMBINED MESCS SHOW MOSAIC EXPRESSION OF THE RECONSTITUTED REPORTER AND DISPLAY HETEROGENEOUS KARYOTYPES 103
4.3 DISCUSSION 108

5 CHAPTER 5: DISCUSSION 115

5.1 CONCLUSIONS **115**
5.2 FUTURE DIRECTIONS **118**
5.2.1 *IN VITRO* AND *IN VIVO* MODELING OF ECDNA 118
5.2.2 *IN VIVO* SOMATIC ENGINEERING BY GENETICALLY ENGINEERED *E. COLI* 131
5.2.3 TILING ARRAY TO ELUCIDATE GENETIC ELEMENT WITH FUNCTIONAL LINK TO *IL7R* IN CLL 133

APPENDIX I: EXPLORING THE USE OF ENGINEERED BACTERIA AS AN ALTERNATIVE DELIVERY VECTOR FOR MAMMALIAN SOMATIC GENE EDITING **134**

APPENDIX II: DISSECTING THE ONCOGENIC ROLE OF DEL13Q14 IN CHRONIC LYMPHOCYTIC LEUKEMIA **157**

6 REFERENCES **180**

List of Tables

Table 1. Examples of commonly deleted tumor suppressor genes 7

Table 2. crRNAs used in Chapters 3 and 4 of this study 48

Table 3. Primers used for genotyping PCR and Sanger sequencing to confirm cassette insertion in human and mouse cell lines 49

Table 4. Primers used for detecting Cre-mediated recombination resulting from split-GFP strategy in human and mouse cell lines 50

Table 5. qPCR primers used in Chapter 3 53

Table 6. Primers used in Chapter 4 to identify double-targeted N0 and F1 mice 54

List of Figures

Figure 1.1. Cancer genomes harbor a diverse and complex array of chromosomal rearrangements 14

Figure 1.2. Different DSB repair pathways in mammalian cells 30

Figure 1.3. Examples of oncogenic chromosomal rearrangements that have been modeled using CRISPR somatic gene editing *in vivo*. 42

Figure 3.1. Overview of the design of the cassettes and expected outcomes upon recombination 63

Figure 3.2. Generation and testing of cassettes for split-GFP strategy 65

Figure 3.3. Generation and verification of HCT116 double-targeted clones 68

Figure 3.4. Cre recombinase induces unification and expression of the reporter in double-targeted clones 72

Figure 3.5. Recombined cells lose reporter expression when passaged under neutral selection and increase reporter expression in response to selective pressure 75

Figure 3.6. Cre-mediated recombination can result in different recombination outcomes in our double-targeted cells 77

Figure 3.7. Cre-treatment results in excision or tandem duplication of the EGFR locus in double-targeted clones 79

Figure 3.8. Sequential sorting of GFP-positive D9 cells selects for tandem duplication 81

Figure 3.9. Murine *p53-/-* aNSCs are able to form and maintain ecDNAs upon induction by Cre recombinase 84

Figure 4.1. Overview of the targeted murine *Egfr* and *Myc* loci in our germline mouse models 96

Figure 4.2. Generation of loxP-targeted germline mouse models using zygote electroporation 98

Figure 4.3. Generation and verification of and cassette-targeted mESCs 101

Figure 4.4. Confirmation of mESC targeting by Southern blot analysis 104

Figure 4.5. mESC clones show mosaic GFP expression after Cre-mediated recombination and are biased against retention of the circle 107

Figure 4.6. Schematic illustrating proposed mouse experiments to test the oncogenic potential of ecDNAs *in vivo* 113

App I Figure 1. Infection of CMT93 with transformed *BM4570* leads to reporter expression in cells 140

App I Figure 2. The ability of *BM4570* to transfer reporter plasmid to cells is specific to the strain 142

App I Figure 3. Delivery of CRISPR components by BM4570 leads to gene editing *in vitro*. 145

App I Figure 4. *In vivo* administration of BM4570 does not lead to effective infection of intestinal epithelial cells. 148

App II Figure 1. Genomic and 3D organization of the del13q14 locus in humans and mice 161

App II Figure 2. Generation of an allelic series of human cell lines harboring various deletions in chr13q14 165

App II Figure 3. Bioinformatic analysis on RNA-seq data from clones harboring chr13q14 deletions is suggestive of an important genetic element between *miR-15a~16-1* and the CBS 168

App II Figure 4. *IL7R* and other genes are negatively correlated with *DLEU2* expression 170

App II Figure 5. Investigating a potential regulatory relationship between *DLEU2* and *IL7R* 172

App II Figure 6. Tiling assay to identify genetic element(s) functionally linked to *IL7R* expression 175

List of Abbreviations

A-EJ: Alternative end joining

aNSC: adult neural stem cell

ATM: Ataxia telangiectasia mutated

BFB: Breakage-fusion-bridge

BRCA1/2: Breast cancer gene 1/2

Cas9: CRISPR associated protein 9

CBS: CTCF-binding site

cGAMP: 2',3'-cyclic GMP-AMP

cGAS: Cyclic GMP-AMP synthase

CIN: Chromosomal instability

CLL: Chronic lymphocytic leukemia

CNN-LOH: Copy number neutral loss of heterozygosity

CRISPR: Clustered regularly interspaced short palindromic repeats

crRNA: CRISPR RNA

CSI: Chromosomal structural instability

CTCF: CCCTC-binding factor

DLEU2: Deleted in leukemia 2

DSB: Double strand break

ecDNA: Extrachromosomal DNA

EGFR: Epidermal growth factor receptor

FACS: Fluorescence-activated cell sorting

FISH: Fluorescence *in situ* hybridization

GFP: Green fluorescent protein

HR: Homologous recombination

HSPC: hematopoietic stem and progenitor cell

IL7R: Interleukin 7 receptor

KO: Knock-out

LNGFR: Low-affinity nerve growth factor receptor

LOF: Loss of function

LOH: Loss of heterozygosity

MDR: Minimally deleted region

mESC: Mouse embryonic stem cell

MRN: MRE11/RAD50/NBS1

MYC: Myelocytomatosis

NHEJ: Non-homologous end joining

sgRNA: single guide RNA

ssODN: single-stranded oligodeoxynucleotide

STING: Stimulator of interferon genes

TAD: Topologically associating domain

TBK1: TANK binding kinase 1

tracrRNA: transactivating crRNA

TSG: Tumor suppressor gene

1 Chapter 1: Introduction

1.1 Chromosomal structural changes are important drivers of cancer

One of the defining hallmarks of cancer is genome instability (Hanahan and Weinberg 2011). Of the different manifestations of genome instability, an important form is chromosomal instability (CIN). Chromosomal instability, as opposed to nucleotide-level mutations such as point mutations or microsatellite instability, affects the structure and number of chromosomes. They are thus more likely to affect the expression of multiple genes at once as opposed to the targeted deregulation of a single gene.

The link between chromosomal abnormalities and cancer development was first proposed by Theodore Boveri more than a century ago (Holland and Cleveland 2009). While prescient, the first example of a recurrent chromosomal abnormality in cancer was not identified until nearly 50 years later by Nowell and Hungerford using cytogenetic techniques (Nowell 1962). In studying karyotypes of patients with chronic myeloid leukemia (CML), they noticed an abnormally small chromosome in most samples. This minute chromosome became known as the “Philadelphia Chromosome”, and though it was initially presumed to be the result of a large interstitial deletion, later work from Rowley revealed that the Philadelphia Chromosome was actually the result of a reciprocal translocation between chromosomes 9 and 22 (Rowley 1973). This translocation leads to the formation

of an in-frame gene fusion between the breakpoint cluster region (*BCR*) gene on chromosome 22 and the Abelson murine leukemia viral oncogene homolog 1 (*ABL1*) gene on chromosome 9, resulting in the expression of a novel oncogene, *BCR-ABL1*. BCR-ABL1 was soon established to be the etiological cause of leukemogenesis due to the constitutive tyrosine kinase signaling activity of ABL1 (Kurzrock, Gutterman, and Talpaz 1988).

Since this initial discovery, hundreds of recurrent balanced rearrangements have now been identified across a broad spectrum of cancers (Mitelman, Johansson, and Mertens 2007). New discoveries have been greatly expedited with advances in high throughput sequencing methods, which have allowed entire cancer genomes to be sequenced. This has provided an unprecedented view of the genomic landscape of cancer and shed light on the fact that chromosomal rearrangements are more numerous, varied, and complex than was previously appreciated (Stephens et al. 2009) (**Fig 1**). Ultimately, cancer is a genetic disease driven by aberrations in two broad classes of genes: the overexpression of genes that allow cells to proliferate without restraint (oncogenes) and/or the downregulation of genes that exert regulatory brakes on cell growth (tumor suppressors). Chromosomal rearrangements are a common method by which cancer cells can achieve this molecular imbalance. Though a variety of different types of rearrangements can be found in cancer genomes, they ultimately converge on common mechanisms that lead to tumorigenesis by inactivating tumor

suppressor genes or upregulating oncogenes. In some cases, rearrangement can generate hypermorphic gene fusions, as illustrated by the case of BCR-ABL1.

Cancer cells also accumulate multiple types of chromosomal alterations during tumor progression. As a result of this, cancer genomes generally become progressively more unstable, directly contributing to accelerated genomic evolution and thus enhancing cancer cells' abilities to adapt to changing environments throughout tumor progression, dissemination, and metastasis (Abbas, Keaton, and Dutta 2013).

1.1.1 Complexity and diversity of cancer-associated chromosomal structural changes

Structural chromosomal abnormalities can be broadly categorized into two general types: balanced (reciprocal) and unbalanced (nonreciprocal). In balanced alterations, the chromosome structure is altered without the net gain or loss of genetic material. The translocation that underpins the Philadelphia Chromosome is one such example of a balanced rearrangement. Conversely, unbalanced or nonreciprocal alterations lead to the net loss or gain of genetic material (e.g. deletions, duplications). Overlaying this classification, structural changes can be additionally characterized as interchromosomal (involving different chromosomes, as in the case of translocations) or intrachromosomal (affecting the same chromosome, as in deletions, inversions, and duplications).

1.1.1.1 Translocations

A chromosomal translocation occurs when a double strand break (DSB) event occurs on a chromosome and is resolved by reattachment to a heterologous chromosome wherein a concurrent DSB also occurred (Roukos and Misteli 2014). Translocations can result in disease by causing transcriptional deregulation (for instance, by repositioning a proto-oncogene next to an active promoter or enhancer), or if the fusion of two chromosomal segments produces a new hybrid gene whose gene product confers a fitness advantage to the cell. An example of the first case is the t(8;14)(q24;q32) rearrangement commonly found in Burkitt lymphomas. As a result of this translocation, *MYC* is placed adjacent to the immunoglobulin heavy chain (*IgH*) enhancer (*Eμ*) (Dalla-Favera et al. 1982; Taub et al. 1982). Under the control of a potent enhancer that normally regulates immunoglobulin expression in B cells, *MYC* becomes overexpressed in the entire B-cell lineage. Indeed, overexpression of oncogenes due to their translocation to the vicinity of immunoglobulin enhancers is particularly frequent in hematological malignancies (Mitelman, Johansson, and Mertens 2007). The Philadelphia Chromosome, introduced above, is an example of how translocation can generate an in-frame fusion gene that encodes for an oncoprotein with a novel function or that signals constitutively.

There are different types of translocations. Balanced translocations involve the reciprocal exchange of genetic segments between two chromosomes, e.g. the telomeric fragment fuses to the centromeric fragment of the other chromosome,

and vice versa. In contrast, nonreciprocal translocations describe the unequal, one-way exchange of DNA, resulting in net genomic loss (**Fig 1.1A**).

Recurrent translocations are particularly common in leukemias and lymphomas, where often only a single, balanced translocation is sufficient to drive tumorigenesis (Aplan 2006). Their prevalence in hematological malignancies is hypothesized to be due to the fact that normal B and T cell development involves the controlled induction and resolution of DSBs during the process of V(D)J recombination and class switch recombination (CSR), mediated by the RAG endonuclease and activation-induced cytidine deaminase (AID) respectively (Nussenzweig and Nussenzweig 2010). It is thought that errors in these somatic events are the driving force of chromosomal translocations in the B and T cell lineages (Kuppers 2005). In contrast, recurrent translocations have been more difficult to identify in solid tumors as they often contain many structural rearrangements involving multiple chromosomes and breakpoints, and feature more complex rearrangements suggestive of multiple rounds of chromosome breakage and fusion (Thompson and Compton 2011).

1.1.1.2 Deletions

In a recent pan-cancer study investigating the patterns of somatic structural variation across >2,500 cancer genomes, deletions were found to be the most common form of structural variation (Li et al. 2020). Deletions typically result in a net loss of genetic material, and most recurrent deletions likely contribute to

malignancy by disrupting the function of tumor suppressor genes (TSGs). This can occur through three mechanisms: Firstly, if the deletion occurs on one chromosome but is repaired by using its homologue as a template, this can result in a copy number neutral loss of heterozygosity (CNN-LOH), which can unmask a recessive mutation if it is present on the retained allele, à la Knudson's "two-hit hypothesis" (Knudson 1971). Alternatively, the deletion could result in hemizygosity if not repaired, resulting in net genomic loss (**Fig 1.1B**). In addition to revealing a recessive mutation, a hemizygous deletion can also lead to haploinsufficiency, which can result in decreased TSG function, as in the case of *CDKN1B, TP53, PTEN, ATM, RB1,* among others (Santarosa and Ashworth 2004; Philipp-Staheli, Payne, and Kemp 2001). Lastly, deletions that affect the same chromosomal segment in both homologues can lead to complete ablation of a TSG. It is worthwhile to note that monoallelic deletions can also lead to the complete inactivation of tumor suppressors if they are located on the X chromosome, because only one copy of the X chromosome is functional. This is illustrated by deletions affecting the *WTX* tumor suppressor, located on Xq11.1, which is found in a subset of patients with Wilms' tumors (Rivera et al. 2007).

Loss of chromosomal segments containing TSGs are routinely found across a broad spectrum of cancers. The most frequently deleted tumor suppressor genes and their associated cancers are summarized in **Table 1.**

Tumor suppressor gene (TSG)	Genetic change	Cancer type
APC	del(5)(q21-22)	Colorectal cancer
PAX5	del(9)(p13)	ALL
CDKN2A/CDKN2B	del(9)(p21)	Various
PTEN	del(10)(p23.3)	Various
ATM	del(11)(q22-q23)	Various
ETV6	del(12)(p13)	AML, ALL
RB1	del(13)(q14.2)	Retinoblastoma
NF1	del(17)(q11.2)	NF1
TP53	del(17)(p13.1)	Various
WTX	del(X)(q11.1)	Wilms' tumor

Table 1. Examples of commonly deleted tumor suppressor genes.
Adapted from Frohling and Dohner, 2008.

While chromosomal deletions are most often associated with loss of gene expression, they may have opposing effects (i.e. increase gene expression) when the affected loci contain regulatory elements. Advances in our understanding of how the genome is organized and folded within the nucleus and how this relates to gene expression have led to the insight that enhancers, promoters, and their cognate gene targets are brought into close proximity within chromatin loops, also known as topologically associating domains (TADs). The boundaries of these TADs are demarcated by DNA-bound CCCTC-binding factor (CTCF)-cohesin complexes, which prevent interaction between elements housed in distinct TADs (Szabo, Bantignies, and Cavalli 2019). Disruption of these boundaries can lead to dysregulation of gene expression by allowing enhancers to interact with non-cognate genes (Ghirlando and Felsenfeld 2016). In cancer, CTCF binding has

been found to be perturbed through various mechanisms that inhibit effective binding of the CTCF-cohesin complex, including hypermethylation of the CTCF-binding site or deletion of such sites altogether (Hnisz et al. 2016; Weischenfeldt et al. 2017).

Finally, deletions can also result in the formation of new fusion genes by removing an intervening chromosomal segment, resulting in the fusion of two genes located on the same chromosome. Indeed, such interstitial deletion is the proposed mechanism behind the *TMPRSS2-ERG* fusion commonly found in prostate cancer (Tomlins et al. 2005; Linn et al. 2016) as well as the *BCAN-NTRK1* fusion oncogene that drives glioblastoma and that was recently engineered in mice by our lab (Cook et al. 2017).

1.1.1.3 Duplications

Duplication of genetic material, particularly of oncogenes, is another common alteration in cancer, and it has been repeatedly demonstrated to be correlated with oncogene overexpression as well as acquired therapy resistance. Duplicated genes may be organized as repeated units occurring at a single locus (tandem duplications or homogeneously staining regions) or distributed throughout the genome (Albertson 2006). They may also be found on acentric, atelomeric structures known as extrachromosomal circular DNA (ecDNA; also commonly known as double minutes) (**Fig 1.1C**). It is worthwhile to note, however, that genomic amplification is not always correlated with increased gene expression

(Santarius et al. 2010), as disruptive breakpoints can occur within genes, thus disturbing their normal expression. Chromosomal tandem duplications can also result in formation of a gene fusion by juxtaposition of the duplicated segment, as in the case of *BRAF-KIAA1549* in pilocytic astrocytomas (Jones et al. 2008).

The members of four families of oncogenes are often amplified: *MYC*, *CCND1, EGFR,* and *RAS* (Croce 2008). *MYC*, the master regulator of various cellular pathways, is the most frequently amplified oncogene across a broad spectrum of cancers (Beroukhim et al. 2010). *ERBB2*, encoding HER2, a member of the epidermal growth factor receptor (*EGFR*) family, is amplified in 30% of breast cancers and constitutes a distinct subtype with different clinical implications, including treatment with HER2-targeting antibody, trastuzumab (Slamon et al. 1987; Carter et al. 1992).

Interestingly, given the inherent redundancy of many protein-coding genes in the human genome, many oncogenes have paralogs with similar functions located elsewhere in the genome. For example, while there are three D-type cyclins that bind to CDK4 and promote the transition from G1 to S phase in the cell cycle, only *CCDN1*, encoding cyclin D1, is frequently amplified across tumors. This may be due to non-overlapping functions amongst the D-type cyclins. An alternative theory is that particular genes may be more susceptible to amplification. Specific chromosomal loci have been determined to be fragile sites that are inherently more susceptible to replication stress, an event that can trigger DSB formation, and the underlying molecular basis for their increased fragility has been

attributed to interference between the replication and transcription machinery particularly at large genes (Helmrich, Ballarino, and Tora 2011). In diffuse large B cell lymphomas, more than 50% of recurrent amplifications and deletions map to these fragile sites (Barlow et al. 2013). Additionally, there is some evidence that suggests that the proximity of the *MET* oncogene to fragile sites can promote its amplification by promoting DSBs that subsequently drive repeated breakage-fusion-bridge (BFB) cycles (described in more detail later), leading to *MET* amplification (Hellman et al. 2002).

1.1.1.4 Inversions

Chromosomal inversions are another type of structural abnormality wherein two DSBs occur on the same chromosome, generating a segment that reverses orientation before being stitched back into the parent chromosome. Inversions can be balanced or imbalanced (involving a net change in the amount of genetic material), pericentric (containing the centromere) or paracentric (not containing the centromere) (**Fig 1.1D**).

Compared to other types of structural abnormalities found in cancer, inversions are relatively infrequent, though there are a few well known and characterized examples. For example, *EML4-ALK* is an oncogenic fusion gene found in a subset of non-small cell lung carcinomas, generated by a recurrent paracentric inversion on the short arm of chromosome 2: inv(2)(p21p23) (Soda et al. 2007). The inversion stitches the intracellular kinase domain of *ALK*, a receptor

tyrosine kinase, to various breakpoints within *EML4*, a microtubule-associated protein, leading to the constitutive activation of *ALK.* The transforming role of *EML4-ALK* has furthermore been verified by somatically inducing the inversion *in vivo* (Maddalo et al. 2014). Another example of an inversion shown to promote tumorigenesis is inv(6)(q22;q22), found in a subset of colorectal cancers without mutations in APC or the downstream WNT pathway, in which an R-Spondin family member, *RSPO3*, becomes fused to *PTPRK,* thus driving overexpression of *RSPO3* (Seshagiri et al. 2012). Again, recreation of this rearrangement *in vivo* in the mouse intestine has verified its identity as a *bona fide* fusion oncogene (Han et al. 2017).

1.1.1.5 Chromoanagenesis

The advent of high-throughput genome sequencing applied to cancer genomes has led to the discovery that some tumor cells can harbor tens to hundreds of rearrangements within localized regions, which seemingly arise through a single catastrophic event. When initially discovered, this phenomenon challenged the then-prevailing theory that cancer initiation and progression is a step-wise evolutionary process guided by progressive cycles of mutation and selection. Although initially thought to occur in only 2-3% of cancers, whole genome sequencing efforts have revealed that chromothripsis is much more prevalent, especially in certain types of cancers, namely sarcomas, melanomas, glioblastomas, and lung adenocarcinomas, where frequency of occurrence can be

more than 50% (Cortes-Ciriano et al. 2020). Significant efforts to understand the cause and consequences of this new type of chromosomal damage led to the discovery that there are at least three distinct but related mechanisms capable of producing these complex, localized rearrangements: chromothripsis, chromoanasynthesis, and chromoplexy. These mechanisms are grouped under the term ‘chromoanagenesis’ (‘chromo’ for chromosomes and ‘anagenesis’ for rebirth) to refer to this type of complex structural reorganization regardless of the etiological mechanism (Holland and Cleveland 2012; Pellestor 2019).

Chromothripsis is characterized by multiple DSBs occurring in a single catastrophic event on one or a few chromosomes, followed by reassembly of the resulting DNA segments in a random order to form derivative chromosomes (**Fig 1.1E**). It was first described by Stephens et al., who performed paired-end sequencing on tumor cells of a patient with chronic lymphocytic leukemia (CLL) and a small-cell lung cancer (SCLC) cell line, SCLC-21H, to study their patterns of chromosomal rearrangements. They found that the breakpoints of the rearrangements occurred within a localized region on the affected chromosomes, and the resulting derivative chromosomes were composed entirely of retained or lost alleles, without evidence of duplications (Stephens et al. 2011). Intriguingly, Stephens et al. also observed the existence of extrachromosomal circular DNAs composed of DNA fragments that were not incorporated into derivative chromosomes. Importantly, while the derivative chromosomes and ecDNAs comprised genetic material from one of the homologous chromosomes, the other

copy remained intact (Stephens et al. 2011). Chromothripsis is thought to occur as a result of the sequestration of lagging chromosomes or acentric fragments into micronuclei, which are self-contained nuclei-like structures separate from the major nucleus. These micronuclei, due to their defective nuclear envelope structure, experience delayed DNA replication and are also more prone to nuclear envelope rupture, exposing the DNA contents to the cytoplasm. Both these events promote massive DSBs in the micronuclear DNA, thus “shattering” it. Upon repair of the DSBs by non-homologous end joining (NHEJ), the resulting rearranged chromosome is then reincorporated into the major nucleus (Holland and Cleveland 2012; Krupina, Goginashvili, and Cleveland 2021).

In contrast to chromothripsis, another form of complex chromosomal reorganization, chromoanasynthesis, was discovered in patients with congenital and developmental abnormalities, whose patterns of rearrangements deviated from what was found in chromothriptic cancer genomes (Liu et al. 2011). In contrast to the pattern of two copy number states evidenced in chromothripsis (with the lower copy number state representing heterozygous deletion of a DNA segment and the higher copy number state representing retention of an allele), chromoanasynthetic chromosomes contain multiple duplications and triplications, and breakpoint sequencing identified templated insertions and stretches of microhomology. Instead of being the result of cataclysmic shattering followed by

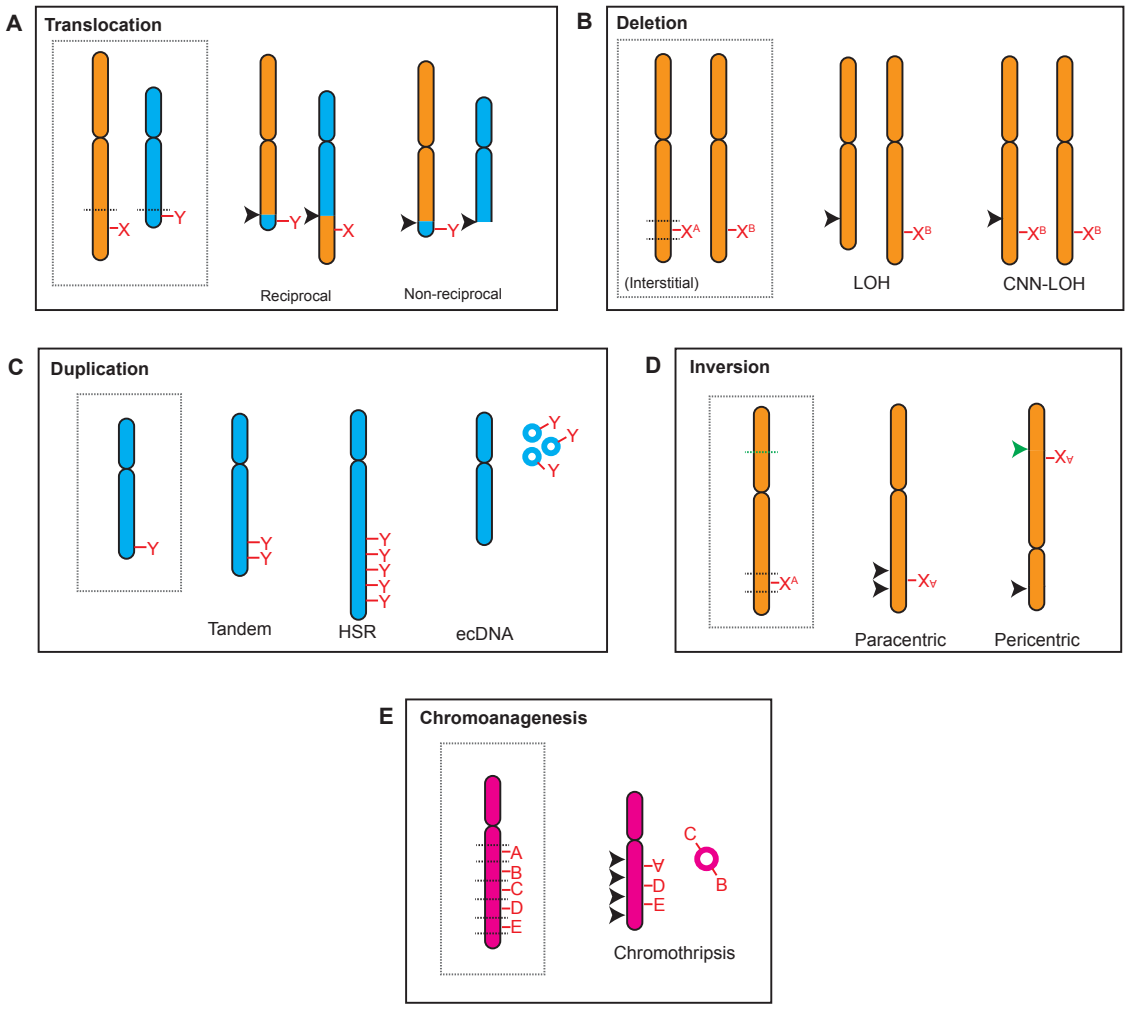


Fig 1.1. Cancer genomes harbor a diverse and complex array of chromosomal rearrangements.

Examples of chromosomal rearrangements found in patient samples. For each panel, the originating double strand break(s) are marked by black dotted lines superimposed on the affected chromosomes, boxed in gray. 'X' and 'Y' are example genes to help illustrate the effect of different rearrangements on affected genes. The different outcomes of repair are shown to the right of each panel. Breakpoint junctions on repaired chromosomes are indicated by black arrows. LOH: loss of homozygosity; CNN-LOH: copy number neutral loss of homozygosity; HSR: homogeneously staining regions.

restitching, chromoanasynthesis is more likely explained by DNA replication-based mechanisms of repair, such as fork-stalling and template switching (FoSTes) and microhomology-mediated break-induced replication (MMBIR) (Lee, Carvalho, and Lupski 2007; Hastings, Ira, and Lupski 2009).

Chromoplexy represents the third class of complex chromosomal remodeling, and has primarily been found in prostate cancer, lung cancer, and melanoma (Baca et al. 2013). It was first discovered in whole genome sequencing studies of primary prostate tumors, where it was revealed that multiple inter- and intra-chromosomal translocations and deletions occur interdependently, forming "chains" of rearrangements that can involve up to 8 chromosomes in a single "chain" (Baca et al. 2013). Thus, unlike chromothripsis, where one or a few chromosomes can incur hundreds of DSBs in the "shattering" event, the DSBs found in chromoplexy are not clustered, are fewer in number, and involve multiple chromosomes (Shen 2013).

1.2 Chromosomal structural instability (CSI) can be caused by improper repair of DNA double strand breaks (DSBs)

While it is clear that different kinds of chromosomal rearrangements drive tumorigenesis through shared mechanisms, i.e. generating driver mutations that lead to oncogene activation and/or tumor suppressor loss, understanding the mechanistic events underlying these rearrangements has also been an active and sustained field of research.

Regardless of type, all chromosomal rearrangements begin with the formation of one or more DNA ends. Here I will mainly focus on how the erroneous repair of two DNA ends resulting from a double strand break can lead to chromosomal rearrangements, but it is also important to note that DNA ends arising from other lesions, such as stalled or collapsed replication forks during DNA synthesis, which result in a one-ended DSB, can also be substrates for chromosomal rearrangement. Indeed, dysfunctional telomeres resulting from extensive telomeric attrition leads to the fusion of deprotected telomeric ends, generating dicentric chromosomes that can engender more chromosomal instability by feeding into breakage-fusion-bridge (BFB) cycles (Maser and DePinho 2002). Replication fork collapse is of particular relevance in cancer, as oncogene activation can result in a phenomenon termed oncogene-induced replication stress, characterized by increased stalled/collapsed replication forks. (Halazonetis, Gorgoulis, and Bartek 2008; Bartkova et al. 2006).

Perhaps the most direct evidence linking the formation of DSBs to chromosomal rearrangement is the fact that a pair of defined chromosomal DSBs, generated by the site-specific endonuclease I-SceI, was sufficient to induce chromosomal translocations in mouse embryonic stem cells (Richardson and Jasin 2000). Reflecting the potentially catastrophic consequences of these events on genome integrity, multiple evolutionarily conserved pathways exist to repair DSBs. The three main pathways of DSB repair that have been implicated in the generation of chromosomal rearrangements, and which are the focus of this

section of the Introduction, are non-homologous end joining (NHEJ), homologous recombination (HR), and alternative end joining (A-EJ).

Highlighting their importance in tumorigenesis, mutations in DSB repair genes underlie many hereditary cancers. For example, germline mutations in ataxia telangiectasia mutated (*ATM)*, a central regulator in the cell's response to DSB damage, lead to ataxia-telangiectasia, an autosomal recessive condition characterized by hypersensitivity to radiation and a predisposition to cancer (Savitsky et al. 1995). Carriers of germline mutations in *BRCA1* or *BRCA2*, two critical proteins required for HR-mediated DSB repair, are predisposed to developing breast and ovarian cancers (Miki et al. 1994; Wooster et al. 1995).

1.2.1 Major sources of pathological DSBs

The formation of a DSB requires the simultaneous breakage of the phosphate backbones of two complementary DNA strands. It is estimated that, based on metaphase analysis of early passage dividing primary mammalian fibroblasts, ten DSBs occur per day per cell (Lieber et al. 2003; Martin et al. 1985). In cycling cells, a major cause of DSBs is the stalling/collapse of the replication fork due to collision with protein complexes (e.g. the transcriptional machinery) or at existing DNA lesions. Other major causes of DSBs include reactive oxygen species (ROS) produced by mitochondria during normal oxidative respiration, ionizing radiation from the environment, and the accidental action of endogenous nucleases. For example, while the lymphoid-specific RAG complex and activation-induced

deaminase (AID) create physiological, programmed DSBs during antigen receptor rearrangement, they can also act on off-target sites (Mahowald, Baron, and Sleckman 2008). Indeed, these erroneous DSBs account for half of all the chromosomal translocations found in lymphomas; as expected, most of the rearrangements involve the immunoglobulin or T-cell receptor loci (Lieber 2010).

1.2.2 Paradoxical roles of DNA damage repair pathways in cancer formation and development

While undoubtedly important, the contribution of DSB repair in the generation of chromosomal rearrangements can also be considered somewhat paradoxical. Evidence from the study of cancer genomes has helped to demonstrate that the inappropriate repair of DSBs by these pathways can lead to chromosomal rearrangements.

The main pathways of repair of DSBs thought to be involved in generating chromosomal structural instability (CSI) are nonhomologous end joining (NHEJ), homologous recombination (HR), and alternative end-joining (A-EJ). Whole genome sequencing studies of rearrangement junctions in cancer cells have uncovered signatures indicating pathological DNA repair by these pathways.

1.2.3 Non-homologous end joining (NHEJ)

In mammalian cells, NHEJ is the predominant DSB repair pathway. To distinguish it from other end joining pathways that have recently been discovered, NHEJ is also sometimes designated with the prefix ‘canonical’ (C-NHEJ). Unlike with

homologous recombination repair, NHEJ does not require a homologous template to effect repair, thus remaining active throughout all phases of the cell cycle, particularly in G0/G1 (Hustedt and Durocher 2016). C-NHEJ is the dominant pathway involved in the repair of ionizing-radiation (IR)-induced DSBs. Outside of repairing pathological breaks, it also plays an important role in rejoining the physiological DSBs generated during V(D)J and class switch recombination in lymphocyte development (Malu et al. 2012). Indeed, germline deficiency in C-NHEJ leads to severe combined immunodeficiency (SCID) owing to improper lymphocyte development (Moshous et al. 2001; Schwarz et al. 2003).

Repair of DSB by C-NHEJ is a multistep process that begins with the binding of the Ku70 and Ku80 heterodimer to the broken DNA ends. Once bound, the Ku complex functions as a docking station for other NHEJ components to load. If DNA end processing is required, the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) is recruited along with the nuclease Artemis. DNA gaps are filled in by the Pol X family of polymerases, Pol μ and Pol λ. As the terminal step, LIG4 and its co-factors XRCC4 and XLF, complete ligation of the DNA ends (**Fig 1.2A**).

DSB repair by C-NHEJ is traditionally characterized as the rejoining of broken DNA ends independent of the genetic sequence at the break. When the two ends of the DSB are composed of perfectly compatible overhangs, or are blunt ends, NHEJ repair is generally error-free. However, most DSBs generate incompatible ends that prevent direct ligation. Therefore, C-NHEJ often involves

minor end resection (< 20 nucleotides) which can expose small regions of microhomology (≤ 4 nucleotides) that aid in end joining (Pannunzio, Watanabe, and Lieber 2018). The end resection observed in C-NHEJ is distinct from the more extensive end resection (≥ 20 nucleotides) that initiates homology-directed repair pathways, such as HR and A-EJ (Grabarz et al. 2012). Due to the iterative processes of end resection and template-independent nucleotide addition, the outcome of C-NHEJ-mediated repair is small deletions (typically 1-4 base pairs) or insertions (due to the activity of fill-in polymerases) (Chang et al. 2017). These small genomic scars left by NHEJ are often referred to as “indels” (insertions or deletions) and are why C-NHEJ is often considered an “error-prone” repair pathway.

1.2.3.1 The role of C-NHEJ in chromosomal rearrangement

The oncogenic capacity of C-NHEJ is due to its ability to join non-contiguous sequences when there is more than one DSB. This can lead to the range of chromosomal rearrangements seen in cancer, including deletions, insertions, inversions, or translocations. The formation of dicentric chromosomes, particularly as a result of deprotected telomeric ends, has also been attributed to the fusion of two telomeric ends by C-NHEJ (Smogorzewska et al. 2002), although more recent studies suggest that dysfunctional telomeres arising from normal telomere attrition are fused by A-EJ (Capper et al. 2007; Rai et al. 2010).

This oncogenic role of C-NHEJ in generating chromosomal rearrangements is supported by studies where the breakpoint junctions of rearrangements in cancer genomes have been sequenced. Based on the fact that little to no homology is found at these junctions, it is thought that C-NHEJ is responsible for generating the rearrangements (Campbell et al. 2008; Stephens et al. 2009), although there is some debate about the potential involvement of alternative end joining (A-EJ) in the rearrangement process (Pannunzio et al. 2014). In addition, the massive chromosomal rearrangement generated by chromothripsis is also thought to be mediated by C-NHEJ (Kloosterman et al. 2011; Stephens et al. 2011; Malhotra et al. 2013; Ly et al. 2017) and/or A-EJ.

Compelling and formal proof for the role of C-NHEJ in chromosomal rearrangement is provided by studies that that employed nucleases to induce site-specific DSBs at endogenous loci in human cell lines (Brunet et al. 2009; Piganeau et al. 2013; Ghezraoui et al. 2014). One such study found that chromosomal translocation frequency was greatly reduced in C-NHEJ-deficient (LIG4 or XRCC4 knock out) cells, indicating that the formation of chromosomal translocations in human cells is dependent on C-NHEJ (Ghezraoui et al. 2014). Surprisingly, this finding directly contrasted the results of an earlier study in mouse ES cells, where translocation was suppressed by C-NHEJ and instead arose through A-EJ (Simsek and Jasin 2010). Thus, there appears to be a species-specific difference in the oncogenic capacity of A-EJ versus C-NHEJ. While in the murine system, chromosomal translocation formation primarily arises by A-EJ and is repressed by

a functional C-NHEJ pathway, the opposite appears to be true in human cells. Nevertheless, A-EJ is still operative in human cells and is capable of forming chromosomal translocations in the absence of LIG4 or XRCC4 (Ghezraoui et al. 2014). A role for A-EJ in cancer may be especially evident in contexts where other DNA damage repair pathways are impaired, as is often the case in cancer (see below).

While a role for C-NHEJ in chromosomal translocation has been extensively studied, its role in the formation of other types of rearrangements, such as gene amplification, are less clear. One possible mechanism by which C-NHEJ may contribute to gene amplification is through the formation of dicentric chromosomes as a result of telomere-telomere fusion, which can subsequently act as a substrate for the breakage-fusion-bridge (BFB) process (Smogorzewska et al. 2002). A classic model for gene amplification, BFB was initially described by Barbara McClintock when studying the fate of dicentric chromosomes during meiotic mitosis in maize endosperm development (McClintock 1941). BFB cycles are characterized by the fusion of sister chromatids due to the loss of telomeric repeats at one of the chromatids prior to replication. As a result of the fusion, a dicentric chromosome is formed. Subsequently, during anaphase, the two centromeres of the dicentric chromosome are pulled in opposite directions, causing a bridge-like structure, composed of the joined segment of the chromosome, to form. As a result of the physical mitotic spindle stress exerted on the bridge structure as the poles pull apart, a DSB will result. If the DSB does not occur at the exact location where

the two chromosomes fused, then there will be an unequal segregation of genetic material between the daughter cells, with one receiving a chromosome containing an inverted duplication and the other the corresponding deletion. Importantly, this BFB cycle produces chromosomes that also lack telomeres, setting the stage for the cycle to perpetuate, resulting in further structural rearrangements (Bohlander, Kakadiya, and Coysh 2019; Lo et al. 2002).

1.2.4 Homologous recombination (HR)

In contrast to NHEJ, which can operate in any phase of the cell cycle, homologous recombination (HR) repair is restricted to the S and G2 phases because it requires the use of a homologous template to mediate faithful repair. During and shortly after replication, these homologous sequences are readily available in the form of sister chromatids. Unlike other DSB repair pathways which are error-prone, HR-mediated repair is thought to be faithful and error-free, thus representing an important mechanism for protecting genome integrity. Indeed, though other mechanisms for DSB repair can in principle also occur during S and G2 phase, HR appears the preferred pathway for repair during these phases of the cell cycle (Moynahan and Jasin 2010).

Mechanistically, HR begins with the binding of the MRE11/RAD50/NBS1 (MRN) complex along with its interacting partner C-terminal binding protein interacting protein (CtIP) to the broken ends of the DSB. The endonuclease activity of MRE11 performs an initial resection, leaving short 3’ single stranded DNA

overhangs at the break points. Once resection is initiated by MRN/CtIP, more extensive end resection is carried out by the nuclease EXO1 or BLM (Bloom syndrome RecQ-like helicase) in complex with DNA2 to generate longer 3' ssDNA tails. These single stranded tails are then coated with RPA to form an RPA-ssDNA nucleoprotein filament, which prevents formation of DNA hairpins from the nascent ssDNA (Chen, Lisby, and Symington 2013). Mediator proteins RAD52, BRCA2, and PALB2 subsequently promote displacement of RPA by the RAD51 recombinase, which is indispensable for HR and essential for strand invasion. The RAD51-coated nucleoprotein filament then performs strand invasion at the homologous site in the sister chromatid, forming the characteristic displacement loop ('D'-loop), thus priming the ssDNA for DNA synthesis. BRCA1, a major HR factor, is thought to be involved at multiple steps of the pathway, including end resection, RPA displacement, and strand invasion (Scully et al. 2019). Once formed, the D-loop intermediate can then be resolved via different pathways. In synthesis-dependent strand annealing (SDSA), which is the predominant pathway in mitotic HR, the newly synthesized strand is displaced and anneals to the other DNA end to form a non-crossover, and thus the template DNA remains unchanged (Paques and Haber 1999). In the classical DSB repair pathway, which is prevalent in meiotic recombination, the second DNA end may be 'captured' to form a double Holliday junction (dHJ), which, depending on the mechanism of resolution can result in non-crossover or crossover (Paques and Haber 1999) (**Fig 1.2C**). A crossover event can be particularly deleterious in a mitotic cell, as it could result in

a loss of heterozygosity. However, the BLM helicase in complex with topoisomerase III has been shown to suppress mitotic crossovers, thereby decreasing the risk of genomic instability when dHJs forms during HR repair (Chu and Hickson 2009; Bizard and Hickson 2014).

1.2.4.1 The role of HR in chromosomal rearrangement

If a DSB is repaired by HR using the sister chromatid as the homologous template, the repair is error-free due to the perfect identity of the sister chromatid, regardless of the outcome of the double Holliday junction. However, if the homologous chromosome is instead used as the repair template in HR (inter-homologue HR), this could in principle lead to loss of heterozygosity, and the extent of the loss would be dependent on whether the inter-homologue HR is resolved by crossover or non-crossover, as well as whether the recombinant sister chromatids segregate into different daughter cells (Moynahan and Jasin 2010).

In both mouse and human, HR between homologous chromosomes occurs much less frequently than HR between sister chromatids (Stark and Jasin 2003). This is likely due to the close vicinity of sister chromatids to each other within the nuclear space, compared to the distance between homologues. In addition, most inter-homologue HR events are resolved by non-crossovers (Stark and Jasin 2003), limiting the extent of the LOH. However, there is evidence that inter-homologue HR can play a role in tumorigenesis, as a fraction of hereditary

retinoblastoma patients have been reported to harbor loss of the wild type allele through recombination-mediated LOH (Hagstrom and Dryja 1999).

Another pathway through which HR could in principle lead to chromosomal structural change is by utilizing homologous sequences located at non-allelic chromosomal positions in a process called non-allelic homologous recombination (NAHR). NAHR commonly occurs between low-copy repeats (LCRs), which are blocks of DNA, ranging from 10 – 300 kilobases, that share sequence similarity within the human genome. However, because crossing over is a rare outcome of mitotic HR, genomic rearrangements rarely occur even as a result of NAHR (Elliott, Richardson, and Jasin 2005; Richardson and Jasin 2000).

The importance of HR in guarding against genomic instability is evidenced by the fact that mutations in genes in this pathway predispose individuals to cancer. For example, lesions in the genes *BRCA1* or *BRCA2* predispose individuals to breast, ovarian, and other cancers (Miki et al. 1994; Wooster et al. 1995). Indeed, germline mutations in these two genes are associated with 3-4% of hereditary breast and ovarian cancers, and individuals with a germline mutation have a 50-80% lifetime risk of developing breast cancer and 30-50% for ovarian cancer (Roy, Chun, and Powell 2011; Tung et al. 2020). Importantly, tumor cells from *BRCA1* mutation carriers display genomic instability characterized by aneuploidy and chromosomal rearrangements (Moynahan and Jasin 2010). This is presumably because, in the absence of a functional HR pathway, most DSBs are instead

repaired by other error-prone pathways, such as C-NHEJ and A-EJ (Mateos-Gomez et al. 2015; Ceccaldi et al. 2015).

1.2.5 Alternative DSB repair pathways

When other DSB repair pathways are compromised, an alternative end joining (A-EJ) repair mechanism is thought to take effect, thus serving as a "backup" mechanism. A-EJ was first discovered in studies of Ku70-mutant yeast cells, which are rendered defective in C-NHEJ. A-EJ displays slower kinetics than C-NHEJ, with some studies suggesting that A-EJ is about 10-fold less efficient than C-NHEJ (Han and Yu 2008; Chang et al. 2017). Unlike C-NHEJ, which does not rely on any microhomology, short stretches of microhomology are required for A-EJ, typically in the range of 2 – 20 bp; at the lower end of this range, there is overlap between C-NHEJ and A-EJ, making it sometimes difficult to truly discern if an observed junctional microhomology is the result of C-NHEJ or A-EJ repair. Due to the requirement for microhomology-based annealing, end resection is the initiating step of A-EJ. The heterologous 3' tails generated during the pairing step are removed, resulting in deletions flanking the initial break point. For this reason, A-EJ is an inherently mutagenic pathway of repair.

Mechanistically, A-EJ initiates with the recognition of DSBs by PARP1. This is followed by binding of the MRN/CtIP complex to begin end resection, which as it progresses, reveals microhomology sequences at the repair site. The DNA ends are then bridged and aligned according to the microhomologies, and the non-

homologous 3' tails are digested by ERCC1/XPF nucleases (Sfeir and Symington 2015). Pol q fills in the resulting gaps within the duplex, and finally the DSB is joined by LIG3/XRCC1 (Simsek et al. 2011). The less efficient LIG1 is also able to replace LIG3 at this step (Simsek et al. 2011) (**Fig 1.2B**).

1.2.5.1 A-EJ and cancer

A role for A-EJ in generating chromosomal rearrangements came from analysis of C-NHEJ- and p53-deficient mice, which lack mature lymphocytes because of impaired V(D)J recombination, but which invariably developed pro-B cell lymphomas harboring oncogenic chromosomal translocations between the *Igh* and *Myc* loci, of which the majority harbored microhomologies at the breakpoints (Zhu et al. 2002; Difilippantonio et al. 2000). These studies implicated the involvement of another end joining pathway capable of joining non-contiguous DNA ends. Subsequent studies then showed that at least in mice, chromosomal translocations are stimulated by C-NHEJ deficiency and arise through A-EJ. However, as discussed above, the converse is observed in human cells.

A-EJ appears to be particularly important in specific contexts where other DNA damage repair pathways are compromised, which is often the case in cancer cells. The evidence for A-EJ involvement is strongest in HR-deficient cancers. *BRCA2*-mutated ovarian cancer cells express elevated levels of *POLQ* (the gene encoding Pol q), suggesting that A-EJ may be active in these cells as a backup pathway for DSB repair (Ceccaldi et al. 2015). Notably, treatment of these cells

with a Pol q inhibitor sensitized these cells to cytotoxic agents (Ceccaldi et al. 2015). In addition, mouse embryos deficient in both Pol q and HR are not viable, further demonstrating a synthetic lethal relationship between A-EJ and HR (Mateos-Gomez et al. 2015).

Expression analyses have also found that chronic myeloid leukemia (CML) cell lines show increased levels of A-EJ factors with a concomitant decrease in C-NHEJ factor expression (Sallmyr, Tomkinson, and Rassool 2008). This effect appears to be even greater in therapy-resistant lines, in which treatment with A-EJ inhibitors (i.e. DNA ligase and PARPi) leads to decreased survival (Tobin et al. 2013).

One possibility is that upon reaching a more advanced stage, characterized by the loss of cell cycle control and acquired deficiencies in DNA damage response and repair pathways, tumor cells may shunt DNA repair into the more mutagenic A-EJ, leading to generation of more complex karyotypes. Indeed, loss of LIG4 or XRCC4 has recently reported to be associated with more frequent complex genome rearrangements in glioblastoma (Ratnaparkhe et al. 2018).

1.3 Historical perspective on modeling chromosomal rearrangements in mouse

While massive sequencing efforts of cancer genomes in recent years have uncovered a wealth of information about the genetic landscape of cancer, shining

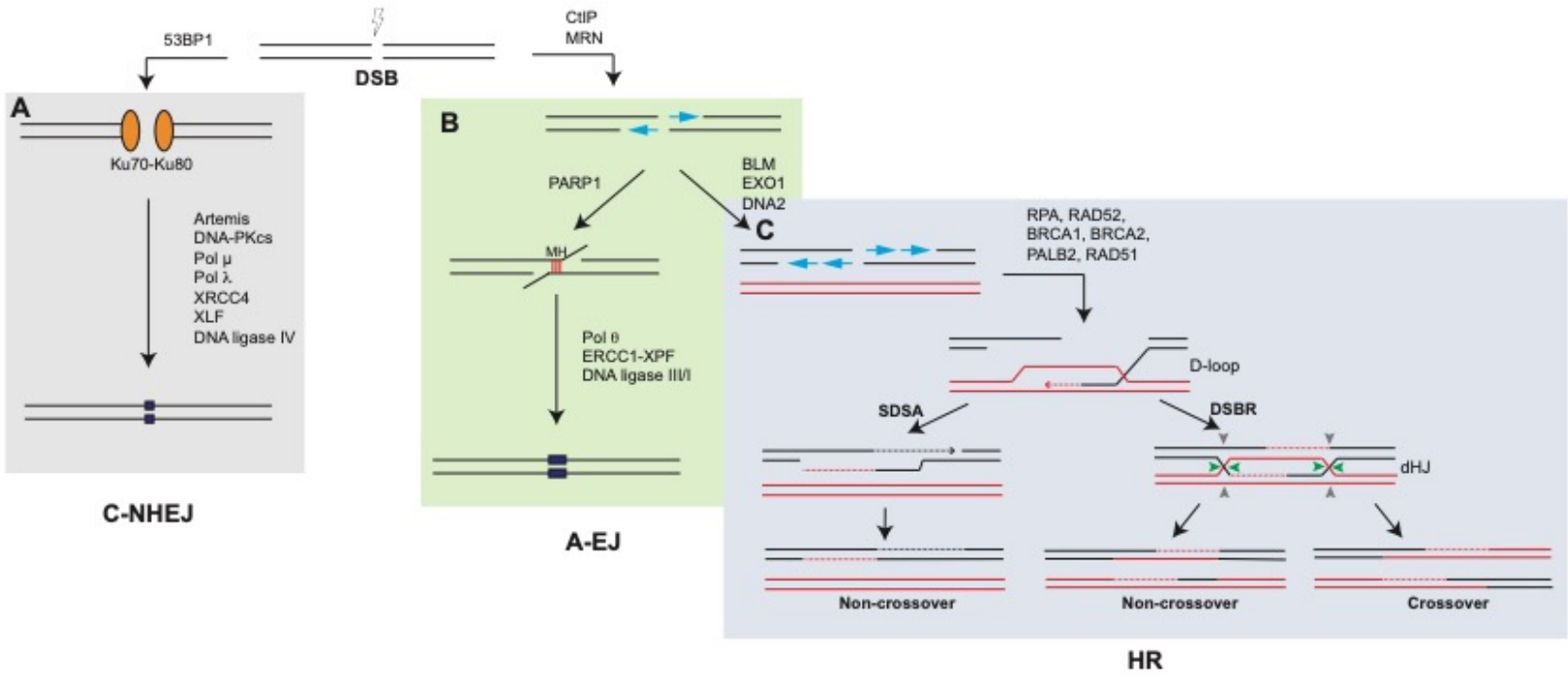


Fig 1.2. Different DSB repair pathways in mammalian cells.
(A) C-NHEJ is the major pathway for DSB repair in mammalian cells. Unlike the other pathways, C-NHEJ does not require major end resection. C-NHEJ initiates with the binding of the Ku70-Ku80 heterodimer, which protects the broken DNA ends and functions as a docking station for the other components of the pathway. The end result of C-NHEJ is often the presence of indels (dark blue boxes) at the breakpoint. **(B)** End resection by CtIP/MRN complex generates ssDNA for A-EJ, which functions as a backup repair pathway. The shallow end resection mediated by CtIP/MRN reveals short stretches of microhomology (MH), which is a requirement for A-EJ. Heterologous 3' flaps are removed following MH-based annealing. A-EJ repair breakpoints often contain more extensive insertions/deletions than those found in C-NHEJ. **(C)** HR is marked by more extensive resection. ssDNA is bound by RPA until displacement by RAD51 and the initiation of strand invasion of a homologous template. HR can proceed via either synthesis-dependent strand annealing (SDSA) or DSB repair (DSBR). Differential resolution of the double Holliday junction (dHJ) in DSBR can result in non-crossover (green arrowheads at both HJs) or crossover (green arrowheads at one HJ and gray arrowheads at the other).

a light on the underlying genomic complexity across a broad spectrum of cancers, experimental validation is required to distinguish true driver mutations from passenger mutations acquired because of the inherent genomic instability of cancer. In this regard, *in vivo* modeling of cancers is crucial for understanding and verifying the genetic lesions that underlie the molecular causes of cancer formation and progression. Once generated, these animal models can also serve as useful preclinical platforms for drug development and testing. Mouse models in particular have been the workhorse of cancer research for many decades. The anatomical, physiological, and genomic similarities between mice and humans make them an essential tool in the study of the biology of cancer.

Here, I offer a brief historical perspective on how mouse models have been traditionally generated and used to study the consequences of oncogene expression or tumor suppressor deletion, which, as explained above, can result from different kinds of chromosomal rearrangements. Historically, the most common strategies employed to imitate the consequences of these genetic modifications can be grouped into three categories: transgenesis, knock-in/knock-out, and Cre-lox-based methods.

1.3.1 The use of transgenic animals to study chromosomal rearrangements

In the earliest studies on the consequences of chromosomal rearrangements in mice, the rearrangements were induced by exposure to chemical or physical mutagens, such as ionizing radiation. However, the breakpoints and types of

rearrangements generated using this method are random and cannot be predetermined, thus limiting the usefulness of this technique.

One of the earliest directed methods of testing the oncogenic potential of a gene was through the generation of transgenic mice. Of the methods available today, it is the simplest and perhaps most widely used approach to quickly determine if a gene has oncogenic potential. One of the first examples of modeling a fusion gene generated by a chromosomal translocation was a model of B-cell lymphoma resulting from the translocation between *MYC* and the IgH enhancer *Eμ* (Adams et al. 1985). However, the major drawback of transgenic animals is that the transgene is randomly integrated into the mouse genome. Often, the gene is placed under control of an exogenous promoter, which does not recapitulate physiological expression levels. In some instances, the non-physiological, ubiquitous expression of a gene may result in embryonic lethality (Heisterkamp et al. 1991). In addition, the untargeted, random insertion of the transgene may also inadvertently disrupt the expression of other genes. With regards to the study of fusion genes, one must also take into consideration that there is no effect on the expression of the endogenous genes involved in the rearrangement. Therefore, transgenic animals do not model the potential haploinsufficiency of the wild type genes that occurs as a result of the rearrangement. Importantly, this reduced dosage might contribute to the oncogenic potential of the rearrangement, classic examples being the *NPM1-ALK* and *PML-RARa* fusions; both NPM1 and PML have been shown to be tumor suppressors (Grisendi et al. 2006; Salomoni and

Bellodi 2007). Moreover, translocations and inversions may result in the generation of reciprocal fusions, which may also have oncogenic potential. To test the cooperative effect of multiple fusions using traditional transgenic approaches would be time consuming, requiring the interbreeding of separate strains.

1.3.2 Homologous recombination-mediated targeting strategies

The realization that cultured mouse fibroblasts could mediate homologous recombination between endogenous DNA and exogenously introduced DNA (Folger et al. 1982) ushered in a new era of mouse modeling, where transgenes were able to be introduced into site-specific locations in the mouse genome to effect knock-in or knock-out. Furthermore, transgenes could be placed under the control of the endogenous promoter. However, knock-in strategies still did not address the limitations of transgenic animals, i.e. reciprocal gene fusion products from translocations and inversions are not modeled, and expression of one of the partner genes is not affected.

1.3.3 Cre-loxP-mediated chromosomal engineering

The development of the Cre-loxP site-specific recombination system, used in combination with HR-based targeting, greatly expanded the types of chromosomal rearrangements that could be modeled in mice. The Cre-loxP system has now been a mainstay of mouse genetics for almost three decades. Using this system, two loxP sites are targeted to desired loci in a sequential fashion into mouse

embryonic stem cells (mESCs). The expression of Cre recombinase in the system then induces recombination between the two loxP sites. The orientation of the loxP sites relative to one another determines the outcome of the rearrangement. In this way, deletions, duplications, inversions, and translocations can be generated.

The usefulness of the Cre-loxP system lies in the fact that chromosomal rearrangements can be spatially and/or temporally controlled. By interbreeding mice harboring loxP sites with strains where Cre recombinase is driven from a tissue-specific promoter, tissue-specific rearrangement can be achieved, limiting the molecular consequences of the rearrangement to a specific cell type. A further layer of temporal control can be included by using Cre recombinase variants that are only activated by the addition of another factor, e.g. CreERT2 and its dependency on tamoxifen for activation. Spatiotemporal control is particularly useful in the cases where the chromosomal rearrangement would otherwise cause developmental defects or embryonic lethality, which is often the case with large-scale deletions. The ability to somatically induce the rearrangement also more faithfully captures the stochastic nature of mutations in cancer.

A major drawback of HR-mediated gene targeting is that successful recombination occurs at a very low frequency (1 in 10^3-10^9 cells), as the efficiency of targeting relies on sporadic formation of a DSB at the target site (Capecchi 1989). Secondly, its dependence on HR restricts its application to dividing cells, limiting its application to certain cell types.

1.4 CRISPR-based methods of modeling chromosomal rearrangements

The low efficiency and lengthy generation times associated with these aforementioned methods of generating genetically engineered mice has hindered the generation of mouse models to study chromosomal rearrangement. The discovery that the introduction of a DSB at the target site increased the frequency of HR-mediated gene targeting by several orders of magnitude fueled the consequent journey to achieve site-specific DSBs (Rouet, Smih, and Jasin 1994). Early studies used meganucleases, such as the 18-bp cutter I-SceI, to introduce DSBs into the mouse genome.

The discovery and adaptation of zinc finger nucleases (ZFNs) and later, TAL effector nucleases (TALENs) to induce concurrent DSBs at specific endogenous loci that, upon repair, gave rise to intra- and interchromosomal rearrangements offered a glimpse into the possibility of a new method of gene editing to model chromosomal rearrangements (Brunet et al. 2009). However, because both technologies relied on protein-based recognition of target DNA sequence, new fusion proteins had to be re-designed and re-engineered for each new DNA target. These rate-limiting steps have largely prevented their widespread adoption.

1.4.1 The discovery and adaptation of bacterial CRISPR systems has revolutionized the field of genetic research

In the relatively short amount of time since it was discovered and successfully adapted for genome editing, CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR associated proteins) has revolutionized biomedical research. The simplicity and elegance of the system is due to its versatile and modular nature. At its core, only two components are needed to induce a DSB at a desired locus: the Cas9 endonuclease and a chimeric single guide RNA (sgRNA) created from the fusion of crRNA (CRISPR RNA) and tracrRNA (transactivating crRNA). The former mediates the specificity to the DNA sequence being targeted and the latter is a backbone sequence required for proper assembly with Cas9 to form a functional editing complex (Jinek et al. 2012). Because the specificity of the Cas9 endonuclease is determined by the sequence complementarity between the crRNA sequence and target DNA, by simply altering this portion of the sgRNA, Cas9 can, in principle, be directed to any locus of interest in the genome. This RNA-mediated recognition bestows the system with much greater flexibility and ease-of-use compared to protein-based recognition systems like ZFNs and TALENs .

CRISPR systems were first discovered as a form of adaptive immunity in bacteria and archaea to protect against invading foreign nucleic acids (Barrangou et al. 2007). There are two major CRISPR/Cas systems (which are further divided into types and subtypes) characterized by the organization of their effector molecules. Class I systems comprise multiprotein effector complexes, whereas class II systems use a single effector. The most widely used Cas9 endonuclease

belongs to this latter class. Cleavage by Cas9 occurs when the crRNA/sgRNA hybridizes to a target sequence flanked at the 3' end by a protospacer adjacent motif (PAM); for Cas9, this PAM sequence requirement is NGG, though different Cas proteins have different PAM requirements.

Following cleavage by Cas9, the resulting DSB is repaired by endogenous repair pathways, C-NHEJ or HR, depending on the cell cycle. Because repair by C-NHEJ often results in the introduction of small insertions or deletions at the cut site, it is useful for generating loss-of-function mutations in protein-coding genes. Repair by HR can be harnessed to insert exogenous sequences, introduced as a donor DNA template harboring homology to the sequences flanking the cut site.

1.4.2 The promise of CRISPR for *in vivo* modeling of chromosomal rearrangements

The promise of CRISPR for the *in vivo* modeling of chromosomal rearrangements is already beginning to be realized. CRISPR has significantly accelerated the generation time of genetically engineered mice by making it facile to target specific genes in mESCs or directly in zygotes. In order to study collaborative effects of different genes in tumorigenesis, mice with mutations in multiple genes have traditionally been generated by sequential targeting in mESCs and/or laborious intercrossing of different strains, each carrying a single mutation. Wang et al. demonstrated that multiplexed gene editing was possible with CRISPR by simultaneously disrupting multiple genes in mESCs. Additionally, they

demonstrated one-step generation of double knockout mice by co-injecting two different targeting sgRNAs directly into zygotes (Wang et al. 2013). Subsequently, it was demonstrated that conditional mice could also be generated through a one-step process: the co-delivery to zygotes of Cas9, two sgRNAs, and two donor DNA templates containing loxP sites led to the generation of animals with floxed alleles (Yang et al. 2013).

Other studies have demonstrated how CRISPR gene editing can be used in tandem with more traditional Cre-loxP strategies: Sanchez-Rivera et al. simultaneously induced Cre-dependent expression of mutant *Kras* and inactivated TSGs by delivery of a single vector containing Cas9, sgRNA, and Cre to *Rosa26*$^{LSL\text{-}KrasG12D}$ mice, thus allowing rapid investigation of the cooperative effects of concurrent tumor suppressor inactivation and oncogene expression in tumor formation and progression (Sanchez-Rivera et al. 2014). Indeed, the combination of Cre-loxP with CRISPR-based techniques presents new possibilities for investigating the temporality of genetic events in tumor progression.

In addition to aiding the generation of transgenic animals by greatly simplifying germline manipulation, arguably the most important technical leap that CRISPR has enabled is the ability to modify somatic cells in the mouse *in situ*. This allows the genome of adult animals to be directly engineered, obviating the need for time-consuming germline manipulation and/or animal breeding. Delivering CRISPR components using hydrodynamic gene transfer, Xue et al. engineered loss-of-function (LOF) mutations in *Pten* and *Trp53* in mouse hepatocytes, leading

to the formation of liver tumors that reproduced the disease characteristics of previous models generated using classical approaches (Xue et al. 2014). *In vivo* somatic engineering has also been successfully applied to engineer chromosomal rearrangements by tandem delivery of Cas9 and two sgRNAs to generate concurrent DSBs at desired breakpoints. Using this technique, it has been possible to model the intrachromosomal inversion that generates the EML4-ALK fusion oncogene in non-small cell lung cancer (NSCLC) as well as the chromosomal deletion that leads to the formation of the BCAN-NTRK1 fusion oncogene in glioblastoma (Maddalo et al. 2014; Cook et al. 2017; Blasco et al. 2014) (**Fig 1.3**). Importantly, both mouse models recapitulated the phenotypes seen in human patients. More significantly, both models also replicated the exact genetic mechanism underlying the transformative molecular changes that drove tumorigenesis.

CRISPR targeting has also enabled other kinds of rearrangements to be modeled. Multiplexed, electroporation-mediated CRISPR targeting of the pancreas in an *in vivo* forward genetic screen led to the recovery of interstitial deletions inactivating $p16^{Ink4a}$ and $p19^{Arf}$ as well as translocations in the resulting pancreatic tumors (Maresch et al. 2016) (**Fig 1.3**). Where *in vivo* transduction approaches have been limited, for example with hematopoietic stem and progenitor cells (HSPCs), success has been achieved with *ex vivo* manipulation, where cells are retrieved and the rearrangement generated before subsequent transplantation back into lethally irradiated recipients. Recently, the t(11;19) *MLL-ENL*

translocation found in acute myeloid leukemia was modeled in human CD34+ HSPCs in this way; upon engraftment in immunodeficient mice, the animals developed a monocytic leukemia-like disease, which upon secondary transplant, resulted in the development of B-cell acute lymphoblastic leukemia (B-ALL), which was postulated to be due to the acquisition of additional genetic lesions during *in vivo* passage (Reimer et al. 2017) (**Fig 1.3**).

Given that the majority of chromosomal rearrangements are initiated by DSBs (with the concurrence of multiple DSBs giving rise to different kinds of rearrangements), it is not difficult to envision that other kinds of rearrangements that heretofore seemed out of reach with conventional transgenic methods, may now begin to be investigated *in vivo*.

1.4.3 Limitations of CRISPR

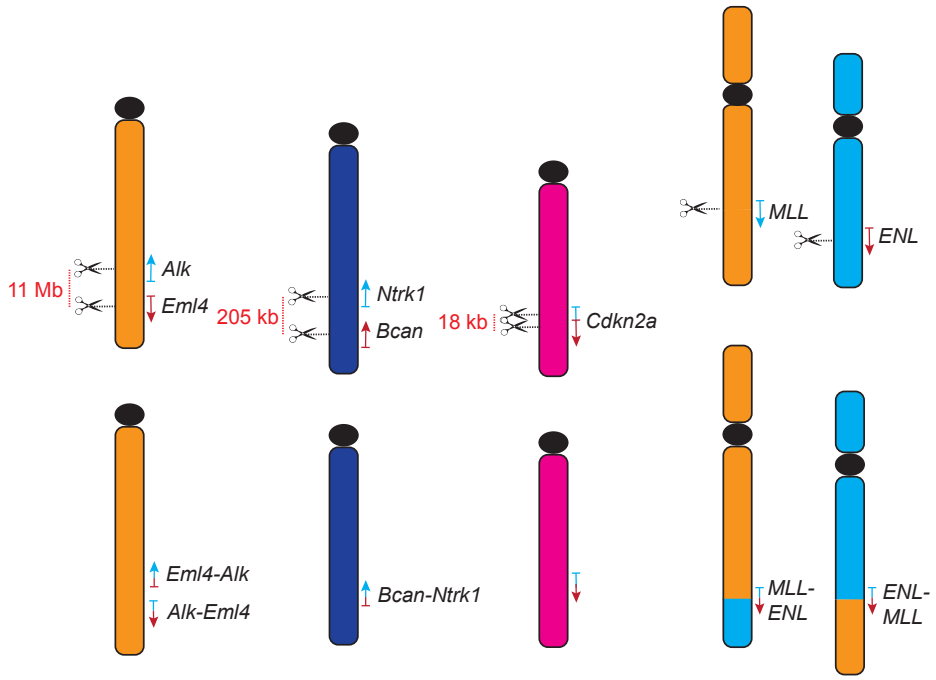
Though CRISPR has been employed to great success to model cancer *in vivo*, it still faces some significant technical limitations. One of the most salient challenges facing researchers who wish to use CRISPR to induce somatic engineering is the methodology of delivering CRISPR components to target tissue in the mouse. Indeed, the ability to deliver the guide RNAs and Cas9 protein to the desired cell or tissue of interest is one of the significant deciding factors in the success of generating any model.

Hydrodynamic gene transfer was the first delivery method used to target naked DNA plasmid harboring CRISPR components to the hepatocytes of mice,

resulting in *Pten* and *Trp53* LOF with subsequent liver tumor formation. The same researchers also showed success in engineering a point mutation in *Ctnnb1* by co-delivery of a single-stranded oligodeoxynucleotide (ssODN) donor (Xue et al. 2014). Though hydrodynamic tail vein injection has demonstrated efficacy for targeting the liver, targeting other organs/tissue types has largely relied on other methods. For example, malignancies of the blood and brain have been modeled using *ex vivo* manipulation of stem cells followed by transplantation into recipient mice. Nucleofection (Cook et al. 2017), electroporation (Chen et al. 2014) and lentiviral transduction (Heckl et al. 2014; Reimer et al. 2017) have all been successfully used to generate LOF mutations as well as chromosomal rearrangements in adult neural stem cells (aNSCs) and in HSPCs that have led to new models of glioma and AML.

Viral delivery of CRISPR components remains the most popular method of targeting mouse tissues. Commonly used delivery systems include lentiviral vectors, but their limited packaging capacity makes it challenging to incorporate Cas9, sgRNAs, and additional elements (e.g. promoters, reporters) within the same viral payload. Nonetheless, lentiviruses have been used to deliver CRISPR to the lung (Sanchez-Rivera et al. 2014; Blasco et al. 2014), pancreas (Chiou et al. 2015), mammary gland (Annunziato et al. 2016), and the colonic epithelium (Roper et al. 2017).

In contrast to lentiviral vectors, recombinant adenoviral vectors have a higher packaging capacity and the additional benefit of not integrating into the host



Rearrangement	*Inversion*	*Interstitial deletion*	*Interstitial deletion*	*Translocation*
Loci involved	*inv(17)*	*del(3)*	*del(2)*	*t(11;19)*
Tumorigenic event	Eml4-Alk fusion oncogene	Bcan-Ntrk1 fusion oncogene	Inactivation of $p19^{Arf}/p16^{Ink4a}$	MLL-ENL fusion oncogene
Reference	Maddalo et al, 2014 Blasco et al., 2014	Cook et al., 2017	Maresch et al., 2016	Reimer et al., 2017

Fig 1.3. Examples of oncogenic chromosomal rearrangements that have been modeled using CRISPR somatic gene editing *in vivo*.
Schematic representation of some of the different types of chromosomal rearrangements that have been generated *in vivo* by inducing CRISPR-mediated double strand breaks directly in the relevant tissue. Each example includes details of type of rearrangement, loci involved, tumorigenic mechanism, and reference paper.

genome, thus decreasing the potential of generating off-target lesions from long-term Cas9 and sgRNA expression. Adenovirus-mediated somatic genome editing has been successful in targeting the lung (Maddalo et al. 2014), liver (Cheng et al. 2014), and brain (Cook et al. 2017). However, adenoviral vectors can also elicit stronger immunogenic responses in mice (Wang et al. 2015; Schirmbeck et al. 2008).

Nonviral delivery methods may need to be developed to address some of the shortcomings of viral delivery methods, especially with respect to reaching inaccessible organ/tissue sites. In this regard, electroporation-based CRISPR delivery has been applied to the pancreas for plasmid-mediated combinatorial gene editing (Maresch et al. 2016), but the requirement for surgery to the animal in order to achieve localization may present a technical limitation.

Recently, development in nanoparticle technologies has been driven by a desire to deliver CRISPR components as gene therapy, with an eye to correcting the genetic lesions found in human diseases in the future. This motivation has led to efforts to engineer nanoparticles with tissue-specific tropisms. Mice constitutively expressing Cas9 were administered endothelial-targeting nanoparticles via tail vein injection; these nanoparticles carried sgRNA against the endothelial-specific gene *Icam2* and resulted in *Icam2*-specific lesions and corresponding decrease in protein expression in pulmonary and cardiovascular endothelial cells (Platt et al. 2014). Most recently, advances have been made in designing lipid nanoparticles capable of selectively targeting specific tissues; these

nanoparticles were shown to be compatible with CRISPR components of various modalities (e.g. Cas9 mRNA and sgRNA or Cas9 ribonucleoprotein complexes) and able to reach the lung, spleen, or livers of mice and lead to gene editing in the lung and liver (Cheng et al. 2020).

1.5 Objectives of this thesis

The advent of CRISPR genome editing has made it feasible to faithfully recapitulate various kinds of chromosomal alterations found in cancer cells. As whole genome sequencing efforts continue to shed light on the genetic landscape of cancer across many different types, potentiating new questions and avenues of research into how different kinds of genetic lesions can lead to tumorigenesis, it will become increasingly important to distinguish driver from passenger mutations. CRISPR offers an unprecedented opportunity to manipulate the genome and experimentally determine the transforming capability of many different types of chromosomal alterations, both *in vitro* as well as *in vivo.*

My thesis explores three inter-related approaches to using CRISPR/Cas9 to manipulate the mouse genome in order to model genetic alterations found in different types of cancer. The bulk of my thesis (Chapters 2, 3, 4 and 5) centers on a novel approach that pairs conventional Cre-lox targeting strategy with CRISPR-mediated editing to engineer extrachromosomal circular DNA amplicons (ecDNAs) harboring oncogenes commonly amplified in cancer. In Chapter 3, I describe our targeting strategy in detail, including the steps of targeting human and mouse cell

lines to carry out proof-of-concept *in vitro* experiments, and our observations following ecDNA formation in these cell lines. Having demonstrated the feasibility of our strategy *in vitro*, Chapter 4 describes the generation of new mouse models to translate our strategy to an *in vivo* system in order to directly address the question of whether oncogene amplification via ecDNAs is sufficient to drive tumorigenesis. In Chapter 5, I discuss the questions raised by our experiments in the previous two chapters, and offer some possible avenues for future investigation.

I round out my thesis with two appendices containing work I carried out on two other projects during my studies. In Appendix I, I report on the results of using engineered bacteria as an alternative delivery vector for CRISPR to mediate mammalian somatic gene editing. In Appendix II, I investigate the role of a large deletion commonly found in chronic lymphocytic leukemia, specifically exploring the possibility that the deletion leads to a restructuring of the local chromatin conformation with consequent de-regulation of gene expression *in cis*.

2 Chapter 2: Materials and Methods

2.1 Cell culture and drug treatment.

The human colorectal carcinoma cell line HCT116 was maintained in RPMI-1640 supplemented with 10% FBS and 1% penicillin-streptomycin (PS). The human embryonic kidney 293T cell line, used for validating the cassettes, was maintained in DMEM supplemented with 10% FBS and 1% PS. The mouse V6.5 embryonic stem cell line (a B6/129 F1 hybrid line) was maintained on a layer of irradiated DR4 mouse embryonic fibroblasts (MEFs) seeded on gelatinized cell culture plates and fed daily with DMEM supplemented with 15% FBS, nonessential amino acids, 0.1 mM beta-mercaptoethanol, 2 mM L-glutamine, and 1,000 U/ml LIF. Mouse adult neural stem cells (aNSCs) were maintained on cell culture plates coated with laminin (10 µg/ml; Sigma) in Neurocult Stem Cell Basal Media with Proliferation Supplements, 20 ng/ml EGF, 10 ng/ml FGF and 2 µg/ml heparin (Stem Cell Technologies), as described previously (Cook et al. 2017).

For selection of targeted cells, the following antibiotics were used at differing concentrations for HCT116, V6.5, or aNSCs (in that order): puromycin (0.25 µg/ml or 2 µg/ml or 0.5 µg/ml), hygromycin (200 µg/ml or 75 µg/ml or 100 µg/ml). For testing the effect of selective pressure on recombined cells, hygromycin was used at the indicated concentrations.

2.2 Cloning and plasmid construction.

NEBuilder HiFi DNA Assembly Cloning Kit and KLD Enzyme Mix (both New England BioLabs) were used for all steps of plasmid construction. For testing the split-GFP minigene, a loxP-containing SV40 small T intron was inserted into the GFP-encoding gene in an existing vector in the lab (Bilenti-GFP, a gift from Dr. Brian Brown, Mount Sinai) as well as within the GFP gene of a promoter-less NeoR-P2A-GFP vector. For constructing the 5' cassette, the hPGK promoter was amplified from the Bilenti-GFP vector and inserted into the NeoR-P2A-GFP-loxP-intron vector digested with KpnI. An SV40 polyA sequence was inserted downstream of GFP as an ultramer. To remove the N-terminus of GFP, divergent primers were used to linearize the vector, followed by ligation using KLD Enzyme Mix (New England BioLabs). To construct the final version of the 5' cassette, NeoR was replaced by PuroR from a pre-existing vector in the lab.

For constructing the 3' cassette, the PuroR was first replaced in the 5' cassette with HygroR amplified from pBABE-hygro (Addgene #1765). The promoter was replaced with the EF1a promoter from a pre-existing vector in the lab. The C-terminus was removed as described above. To construct the final version of the 3' cassette, H2B was amplified from a pre-existing vector in the lab and inserted upstream of the N-terminus of GFP.

2.3 Cassette targeting.

To achieve targeting, cassettes were delivered to cells in the form of a PCR-amplified linear DNA containing homology arms of 40 – 80 bp, in tandem with Cas9

protein, crRNA, and tracrRNA (all from Integrated DNA Technologies) pre-assembled into Cas9-RNP. The components were introduced into HCT116 and V6.5 cells via transfection using Lipofectamine CRISPRMAX Cas9 Transfection Reagent (Thermo Fisher Scientific). Two days after transfection, cells were selected using puromycin (for 5' cassette targeting) or hygromycin (for 3' cassette targeting) at concentrations indicated above. For aNSCs, the components were introduced into cells by nucleofection using the Amaxa Nucleofector X Unit (Lonza). Cells were resuspended in P3 buffer and nucleofected using program DN-100.

crRNA (referred to in Fig 3.3 and Fig 4.1)	Sequence (5' – 3')
hg1	CAGTCTGCGTGATGTTACAG
hg4	ACAATGTCGCCAATGTACAG
mg1	TAACTTCTTCCACCTATGGG
mg5	GCCCGTATAGTATCGCCTGT
mg6	AGATGCGCACAGAAAAGTGG
mg10	ATCATGAGTTGAGTTCACTC

Table 2: crRNAs used in Chapters 3 and 4 of this study

To verify successful targeting, genomic DNA was extracted from cells using standard protocols. Primers spanning the cassettes were used for diagnostic PCR genotyping to identify positive clones as well as for Sanger sequencing to verify the insertional junctions and to check the integrity of the inserted cassette sequence.

Species	Locus	Sequence (5' – 3')

Human	*EGFR* proximal (forward)	ATATGCTAAAAGGAAGTGAGCT
Human	*EGFR* proximal (reverse)	GGAAGCAGTGAGATGACTCTAG
Human	*EGFR* distal (forward)	ACTAAGCAGGTCCTAGGGCC
Human	*EGFR* distal (reverse)	TCCTGCCCGGCTAATTTTTAGT
Mouse	*Egfr* proximal (forward)	AGAGGTGTGTAGCATTTGAAAT
Mouse	*Egfr* proximal (reverse)	AAGGCCCAGATTATCGTGCA
Mouse	*Egfr* distal (forward)	CAGTTATAGCAACTAGCAATCG
Mouse	*Egfr* distal (reverse)	GCAGATGTTGTGACACTTTACA
Mouse	*Myc* proximal (forward)	CCAACAGATTGAGAGAGAGAGA
Mouse	*Myc* proximal (reverse)	GAGTTGGATTCAGAATAACCAC
Mouse	*Myc* distal (forward)	GGGGACTGTAACTTAGTATGGA
Mouse	*Myc* distal (reverse)	AGGGCTTATGCTTAAAATCTAC

Table 3: Primers used for genotyping PCR and Sanger sequencing to confirm cassette insertion in human and mouse cell lines

Adenoviral transduction of Cre recombinase.

Recombinant adenoviruses expressing Cre recombinase (1 x 10^{12} particles/mL) were purchased from ViraQuest. Induction of Cre-mediated recombination was achieved in HCT116 by transducing cells with approximately 1 x 10^{8} particles/mL of Ad-Cre. For V6.5 cells and aNSCs, approximately 1 x 10^{9} particles/mL was used. To verify recombination, genotyping PCR was performed using multiplexed primer

combinations that would allow simultaneous detection of the reconstituted reporter and the unrecombined 3' cassette (refer to **Fig 3.4** for illustrative diagram).

Species	Locus	Sequence (5' – 3')
Human/Mouse	Reconstituted GFP forward (anneals to GFP-N) (primer H in **Fig 3.4B, C**)	TCGAGCTGAAGGGCATCGAC
Human	Reconstituted GFP reverse (primer B in **Fig 3.4B, C**)	GGAAGCAGTGAGATGACTCTAG
Human	Unrecombined 3' cassette reverse (primer F in **Fig 3.4B**)	TCCTGCCCGGCTAATTTTTAGT
Mouse	Reconstituted GFP reverse	AAGGCCCAGATTATCGTGCA
Mouse	Unrecombined 3' cassette reverse	GCAGATGTTGTGACACTTTACA

Table 4: Primers used for detecting Cre-mediated recombination resulting from split-GFP strategy in human and mouse cell lines

2.4 GFP expression analysis and isolation of GFP-positive cells.

GFP expression was analyzed by flow cytometry (BD LSRFortessa) using 4'6-diamidino-2-phenylindole (DAPI) as a live/dead stain. To isolate GFP-positive cells, fluorescence-activated cell sorting (FACS) was performed with assistance from the MSKCC Flow Cytometry Core Facility on BD FACSAria cell sorter machines. All raw data was analyzed and quantified using FlowJo software.

2.5 Fluorescence in situ hybridization (FISH) analysis.

FISH experiments were performed by the MSKCC Molecular Cytogenetics Core Facility. For metaphase harvest, HCT116 and murine ES and NS cells were treated with colcemid at a final concentration of 0.1 μg/mL (human) or 0.05 μg/ml (murine). Following 30-60 min incubation, the cells were trypsinized according to standard procedures, incubated in 0.075M KCl for 10 minutes at 37°C and fixed in methanol-acetic acid (3:1). FISH analysis was performed on fixed cells using a 2-color probe. BAC/PAC clones containing human *EGFR* locus (RP5-1091E12, RP11-339F13) were labeled with Red dUTP and a centromeric repeat plasmid for Chromosome 7 (P7t1), labeled with Green dUTP served as the control. BAC clones containing murine *Egfr* locus (RP23-295E4, RP23-263C13) were labeled with Red dUTP and RP23-173M6 (11qA1), labeled with Green dUTP served as the control. All RP11 and RP23 clones were purchased from the Roswell Park Cancer Institute Genomics Shared Resource (Buffalo, NY). Probe labelling, hybridization, post-hybridization washing, and fluorescence detection were performed according to procedures established at the Molecular Cytogenetics Core Facility. Slides were scanned using a Zeiss Axioplan 2i epifluorescence microscope (Carl Zeiss Microscopy, Thornwood, NY) equipped with Isis imaging software (MetaSystems Group Inc, Waltham, MA). The entire hybridized area was scanned through 63X objective lens to assess the quality of hybridization and signal pattern. Following initial scan, for each cell line, a minimum of 50 intact nuclei and 20 metaphase spreads were analyzed and signal pattern recorded. In the HCT116 cell line, a minimum of consecutive 1000 cells were scored and the

entire hybridized slide also analyzed to assess the frequency of micronuclei and record the signal pattern in all FISH-positive micronuclei. To the extent possible, apoptotic cells/bodies were excluded. Criteria for scoring micronuclei were as follows: 1) morphologically identical to but smaller than main-nuclei (< 1/3 the mean diameter of main-nuclei); 2) non-refractile and DAPI-stain intensity comparable to main-nuclei; 3) present outside or minimally touching the main-nucleus.

2.6 Southern blotting.

Southern blotting was performed using neutral transfer protocol adapted for DIG-labeled probes. For each sample, 10 µg of digested genomic DNA was resolved in a 0.7% agarose gel and transferred onto a Hybond-N^+ nylon membrane (GE Healthcare). Probes were designed to be 400 – 500 bp in length and amplified from wild type V6.5 genomic DNA (for external probes) or vectors containing the cassettes (for internal probes). PCR DIG Probe Synthesis Kit (Sigma-Aldrich) was used to generate DIG-dUTP labeled probes.

2.7 Quantitative PCR.

qPCR was performed using SYBR Green Real-Time PCR reagents (Thermo Fisher Scientific) on an Applied Biosystems 7900HT Fast Real-Time PCR instrument (Thermo Fisher Scientific). Genomic DNA was prepared from samples using standard methods. Primers were designed to amplify 150 – 250 bp

amplicons. Each sample was represented by three technical replicates for each target gene. Normalized fold change was calculated using the delta-delta Ct method (Livak and Schmittgen 2001).

Target	Oligos (5' – 3')
Reconstituted GFP	**For**: GGGCACAAGCTGGAGTACAA **Rev**: TGAAGTTCACCTTGATGCCG
Internal control	**For:** GCTCCGTTAAAGCTTGCTCCT **Rev:** AATGGCTGTCACACCTCATCAA

Table 5: qPCR primers used in Chapter 3

2.8 Zygote electroporation.

Zygote electroporation was carried out by the MSKCC Mouse Genetics Core Facility based on published protocols (Hashimoto, Yamashita, and Takemoto 2016) using crRNAs experimentally validated in mESCs. Briefly, multiple zygotes placed in an electrode chamber were subjected to electroporation at one time. Each electroporation mixture contained the 5' and 3' breakpoint targeting crRNAs (25 ng/µl), Cas9 protein (100 ng/µl; IDT), and two ssODNs of 159 bp with asymmetric homology arms (200 ng/µl) in a solution of 0.1% polyvinyl alcohol. Electoporated zygotes were cultured in KSOM medium at 37°C and 5% CO_2 until the two-cell stage, whereupon they were transferred to the oviducts of pseudopregnant females on the day of the vaginal plug. N0 animals generated from the zygotes were genotyped and Sanger sequenced to confirm insertion of both loxP sites.

Locus	Sequence (5' – 3')

Egfr proximal (forward)	TGCTGGTTGTGCTCAGTCTTC
Egfr proximal (reverse)	TCTGCATTAGCTCCTGCCTCC
Egfr distal (forward)	TCCTGGTCAAATTAGTGTGGC
Egfr distal (reverse)	GCTTACAAGGCCTCAACCCTAG
Myc proximal (forward)	CATGTTGAACCAGAGTACAC
Myc proximal (reverse)	AGCTTAGCTGAGAAATGAAGAGC
Myc distal (forward)	TGGAGTTGTCTCTGGTCTGTC
Myc distal (reverse)	GGATAACCGTGAGCTCCCAGC

Table 6: Primers used in Chapter 4 to identify double-targeted N0 and F1 mice

2.9 Generation of novel floxed mouse strains.

Double-targeted N0 mice were mated to C57BL/6J wild type mice to generate $Egfr^{fl/+}$ and $Myc^{fl/+}$ F1 progeny. Genotyping PCRs using primers listed in Table 6 were used to identify mice with loxP sites inserted *in cis* or a single loxP site. Tail fibroblasts were isolated from $Myc^{fl/+}$ F1 mice of several weeks old for *in vitro* culture; these fibroblasts were transduced with Ad-Cre to test for recombination using the following primers (from Table 6): (For circularization) *Myc* proximal (reverse) + *Myc* distal (forward); (for deletion) *Myc* proximal (forward) + *Myc* distal (reverse).

2.10 Statistical analysis and quantification.

All statistical analyses were performed using Prism software. Unless otherwise noted, data were analyzed using unpaired two-tailed t-test. *, $p<0.05$.

3 Chapter 3: *In vitro* modeling of ecDNA

3.1 Introduction

As discussed in Chapter 1, one of the most frequent alterations driving cancer formation is oncogene amplification, where the copy number increase of key oncogenes provides cells with selective growth advantages. Homogeneously staining regions (HSRs) are a common mechanism of copy number amplification. HSRs are comprised of repeated segments within a chromosome; its name derives from its characteristic extended and uniform staining pattern in karyotypes. Extrachromosomal circular DNA (ecDNA) has long been known to be another common mechanism through which tumor cells can achieve gene amplification. However, the role of these oncogene-harboring, non-chromosomal elements in tumor formation and progression has only begun to be carefully studied in recent years.

Though non-chromosomal material has been recognized to exist in both normal and malignant cells (Moller et al. 2018; Shibata et al. 2012; Paulsen et al. 2018), there are a number of important differences between the extrachromosomal circular amplicons found in normal somatic cells and those associated with cancers. Firstly, ecDNAs found in normal tissues do not contain full genes; a class of these, called “microDNAs”, are 200-400 bp long, derived from unique sites in the genome and enriched in 5’ UTRs and CpG islands (Shibata et al. 2012). These structures were directly visualized with electron microscopy following their isolation from

exonuclease-treated mouse brain tissues (Shibata et al. 2012). Secondly, these ecDNA amplicons are not recurrently found in cells, suggesting they are not under active selective pressure.

In contrast, the ecDNAs found in cancer are usually a few megabases in size and contain one or multiple full genes as well as regulatory regions (Wu et al. 2019). Indeed, their large size allowed Spriggs and colleagues to visualize them by simple light microscopy when studying metaphase spreads of neuroblastoma cells from patients (Cox, Yuncken, and Spriggs 1965). They referred to these chromatin bodies as “double minutes”, in reference to the fact that they were often found in a paired conformation. Later, Alt and colleagues showed that oncogenes (specifically *MYCN*) could be carried on these double minutes (Kohl et al. 1983). With recent advances in sequencing, computational, and cytogenetic methods shedding more light on ecDNA biology, the term “double minutes” may now be considered a bit of a misnomer, as these circular DNA structures have now been shown to exist in either a single body form or the conventional double body form that gave them their initial name. In fact, we now know that only 30% of ecDNAs exist in a paired conformation (Turner et al. 2017). In addition to their large size, ecDNAs are found to contain recurrent oncogenes and to exist at high copy numbers in cancer cells, suggesting they are actively maintained through positive selection (Kim et al. 2020).

A role for oncogene-encoding ecDNA in tumor progression only began to be appreciated when Nathanson and colleagues were studying drug resistance of

targeted therapies in glioblastoma. They found that *EGFR*, which is frequently mutated in glioblastoma as the deletion variant *EGFRvIII*, is mostly amplified on ecDNA in the patient-derived cell line, GBM39. Importantly, GBM39 cells harboring *EGFRvIII*-ecDNA were able to develop resistance to anti-EGFR therapies through the reversible loss of ecDNAs. Nathanson et al. found that during drug treatment, the number of *EGFRvIII*-containing ecDNA rapidly decreased in cells, but once treatment was withdrawn, the number of ecDNAs rebounded (Nathanson et al. 2014). Since then, the systematic analysis of different cancer cell lines has shown that ecDNAs are widespread in many tumor types and is a common mechanism through which oncogenes are amplified (Kim et al. 2020; Turner et al. 2017). Furthermore, ecDNAs have been shown to promote intratumoral genetic heterogeneity, presumably due to their lack of centromeres (deCarvalho et al. 2018). Recent structural and epigenetic studies of ecDNAs have also demonstrated that while ecDNAs are also packed into nucleosomes, the chromatin found in ecDNAs are much less tightly compacted, resulting in high chromatin accessibility, which contributes to the high level of transcriptional activity of oncogenes found on ecDNA (Wu et al. 2019). Thus, ecDNAs represent a mechanism by which tumor cells can quickly respond to changes in the environment (e.g. drug treatment), promote genomic evolution, and increase oncogene transcription.

Currently, the biogenesis of ecDNAs is not well understood, although chromothripsis is hypothesized to be one possible mechanism through which they

form (Holland and Cleveland 2012). It is thought that the linear DNA fragments produced as a result of chromothripsis may self-ligate to form circular structures through non-homologous end joining (NHEJ). Sequencing performed on the breakpoint junctions of circular amplicons in multiple tumor types display no or minimal (<5 bp) sequence homology, implicating NHEJ in their breakpoint repair (Kim et al. 2020). In a recent study investigating the mechanism underlying early acquisition of resistance to methotrexate in HeLa cells, Shoshani et al. found that spontaneously arising resistant clones contained *DHFR* (resistant gene) within ecDNAs, and that inhibition of NHEJ significantly decreased the frequency of resistant colony formation and double minute production (Shoshani et al. 2021).

Despite the growing interest in ecDNAs, there are many outstanding questions about their role in tumor formation and progression that cannot yet be addressed due to a lack of methods to engineer their formation *in vitro* or *in vivo*. For example, though the prevalence of ecDNAs in many tumor types is now well documented, the ability of ecDNAs to initiate tumorigenesis has not yet been investigated. It is not known if ecDNA-mediated oncogene amplification is an early event, or a manifestation of the inherent genomic instability of cancer cells as they advance. It is also unknown whether other genetic requirements, such as concurrent loss of tumor suppressors or other factors, need to be fulfilled in order for ecDNAs to be maintained and propagated. To definitively answer these questions, methods to model these structures both *in vitro* and *in vivo* are urgently needed.

We have attempted to develop a strategy to induce circularization of any genomic locus of choice by taking advantage of the Cre-loxP system, a mainstay of mouse genetics, in conjunction with our lab's expertise in CRISPR/Cas9 mediated genome editing. Our strategy entails inserting two cassettes, each containing a complementary but nonfunctional half of a GFP reporter gene, as well as a loxP site, into pre-determined breakpoints in the genome. Recombination between the loxP sites circularizes the intervening region and reunites the two halves of the GFP reporter, thus reconstituting a functional reporter gene. In this chapter, we demonstrate the feasibility of our system *in vitro* in human and mouse cell lines.

3.2 Results

3.2.1 A split-GFP strategy for modeling circularization

We envisaged a straightforward method to induce circularization, taking advantage of the Cre-loxP system in conjunction with our lab's expertise in CRISPR/Cas9 mediated genome editing. Our strategy involves flanking a region of interest (containing oncogenes and/or functional enhancers) with a pair of cassettes containing loxP sites (**Fig 3.1A**). When the loxP sites are inserted in *cis* and in the same orientation, adding Cre recombinase will induce circularization of the intervening region concurrent with its excision from the chromosome (**Fig 3.1A**).

We sought to enhance this simple circularization strategy with additional features that would facilitate downstream *in vitro* and *in vivo* studies. Ideally, our system would allow us to identify and interrogate the behavior of the recombined circle(s) within cells. To accomplish this, we decided to split a GFP reporter cassette between the proximal and distal loxP-containing cassettes, such that upon recombination, the full-length GFP reporter would be reconstituted and lead to GFP expression, and thus allow us to visualize recombinant cells (**Fig 3.1C**). The loxP sites within these cassettes are incorporated into an SV40 intron (**Fig 3.1B**). The dynamic range of this split-GFP strategy, which would manifest as variation in GFP signal intensity, would furthermore allow us to monitor the abundance of circles and to distinguish between cells with few versus many copies

of circular DNA. This would serve as a useful readout to study changes in ecDNA copy number over time or in response to selective pressure.

Because we wished to implement our strategy for *in vivo* studies by generating germline mouse models harboring the cassettes, it was of utmost importance to us to be able to exert spatiotemporal control over the formation of the ecDNAs. The conditional nature of the Cre-loxP system is advantageous in this regard, as we can control the tissue-specificity of ecDNA formation by crossing our mice with strains where the Cre recombinase is driven from a tissue-specific promoter. For an added layer of temporal control, our mice can also be crossed to strains expressing modified versions of Cre, such as the hormone-dependent CreERT2, so that ecDNA formation is only induced upon addition of tamoxifen. Additionally, the reconstituted GFP reporter should not elicit an immunogenic response *in vivo* once recombined, because each half of the reporter is constitutively expressed from its own promoter (**Fig 3.1B**).

Because there should only be one or two copies of the circle in each cell immediately after recombination, we reasoned that adding an N-terminus H2B tag to the reporter half within the 3' cassette would help to concentrate the reconstituted signal within the nucleus, which would aid us in the identification of recombined cells (**Fig 3.1B**). Moreover, studies have shown that ecDNAs found in cancer cells are also packaged into chromatin (Wu et al. 2019). Therefore, the H2B-fused reporter may potentially serve as a way to directly visualize the recombined circles within GFP-positive cells.

Finally, each cassette contains a different resistance gene to allow for the selection and isolation of targeted clones (**Fig 3.1B**). The selection genes are placed within the cassettes in such a way that the hygromycin resistance gene will be retained within the circle upon recombination. We reasoned that this would allow us to test the maintenance, propagation, and accumulation of the circles when selective pressure (i.e. hygromycin) is applied to the system.

3.2.2 Generation and testing of cassettes

We first set out to test the split GFP reporter by inserting a loxP-containing SV40 intron into the middle of a GFP reporter gene located within a previously experimentally validated bidirectional lentiviral vector (see Appendix II). We ascertained by fluorescence imaging that this did not negatively impact the intensity of GFP expression within transfected cells (**Fig 3.2A**).

Having determined that the expression of the split-GFP minigene is not affected, we next cloned it downstream of NeoR-P2A from another plasmid within the lab's repository, making a promoter-less NeoR-P2A-GFP-loxP-intron construct, within which we inserted the hPGK promoter amplified from our bidirectional vector. We tested the two constructs (promoter-less and with hPGK) to confirm that GFP expression is promoter-dependent (data not shown). An SV40 polyA sequence was cloned downstream of the GFP to further enhance its expression (data not shown). To remove the N-terminus of GFP to create the 5' cassette, divergent primers were designed to amplify the entire vector excluding the N terminus; the

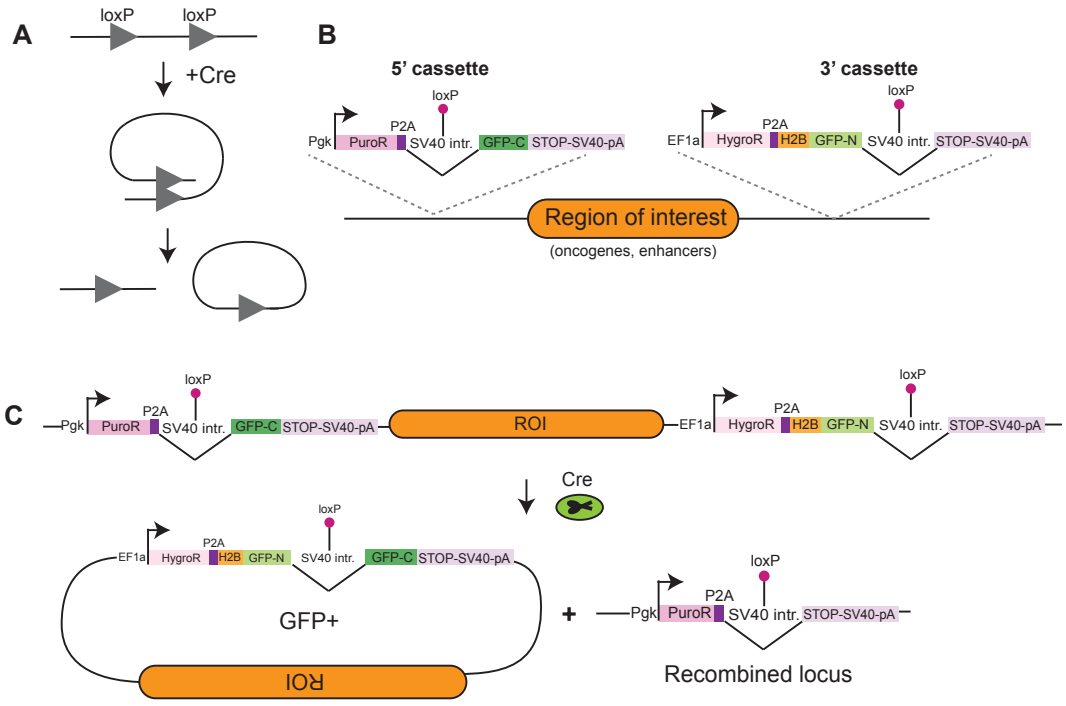


Figure 3.1. Overview of the design of the cassettes and expected outcomes upon recombination.
(A) Cre recombinase mediates a site-specific recombination event between two loxP sites inserted in the same orientation *in cis* (on the same chromosome). This leads to the deletion of the intervening DNA with concomitant circularization of the excised region. (**B)** Schematic of the loxP-containing cassettes for our split H2B-GFP strategy. The cassettes are inserted at proximal and distal desired breakpoints marking the boundaries of a region of interest. (**C)** Upon addition of Cre recombinase, the intervening region will be excised as a piece of circular DNA containing the region of interest. In addition, recombined cells will display nuclear expression of GFP due to the reconstitution of the H2B-GFP reporter.
ROI = region of interest.

resulting linearized vector was ligated using a one-step KLD (kinase, ligase, and DpnI) mix. We confirmed that the C-terminus of GFP does not produce a fluorescent signal (**Fig 3.2B**). These steps constituted an early version of the 5' cassette. In a later version, we replaced *NeoR* with *PuroR*, as our transfected cells did not survive G418 treatment when tested for antibiotic resistance with the early version of the cassette. We confirmed robust survival of cells transfected with the *PuroR* upon puromycin treatment (data not shown).

Several modifications were made to the 5' cassette to convert it into the 3' cassette. First, *NeoR* was replaced with *HygroR*. The EF1a promoter was amplified from an existing lentiviral vector and used to replace hPGK. The new construct, containing EF1a, *HygroR*, and full-length, split-GFP (with loxP intron) was used to test for GFP expression and hygromycin resistance, which were both confirmed (**Fig 3.2C** and not shown). The C-terminus was then removed using the same method employed to remove the N-terminus. H2B was inserted into both constructs (full-length and N-terminus GFP) to generate the fused reporter. Upon transfection into cells, we confirmed nuclear localization of the full-length split reporter signal, and complete abrogation of signal when only the N-terminus was expressed (**Fig 3.2C**). Finally, we also confirmed that the two halves of the reporter expressed in the same cell generates no fluorescent signal (**Fig 3.2D**).

3.2.3 Generating double-targeted clones flanking the *EGFR* locus

For our *in vitro* studies, we chose to target the *EGFR* locus. *EGFR* amplifications

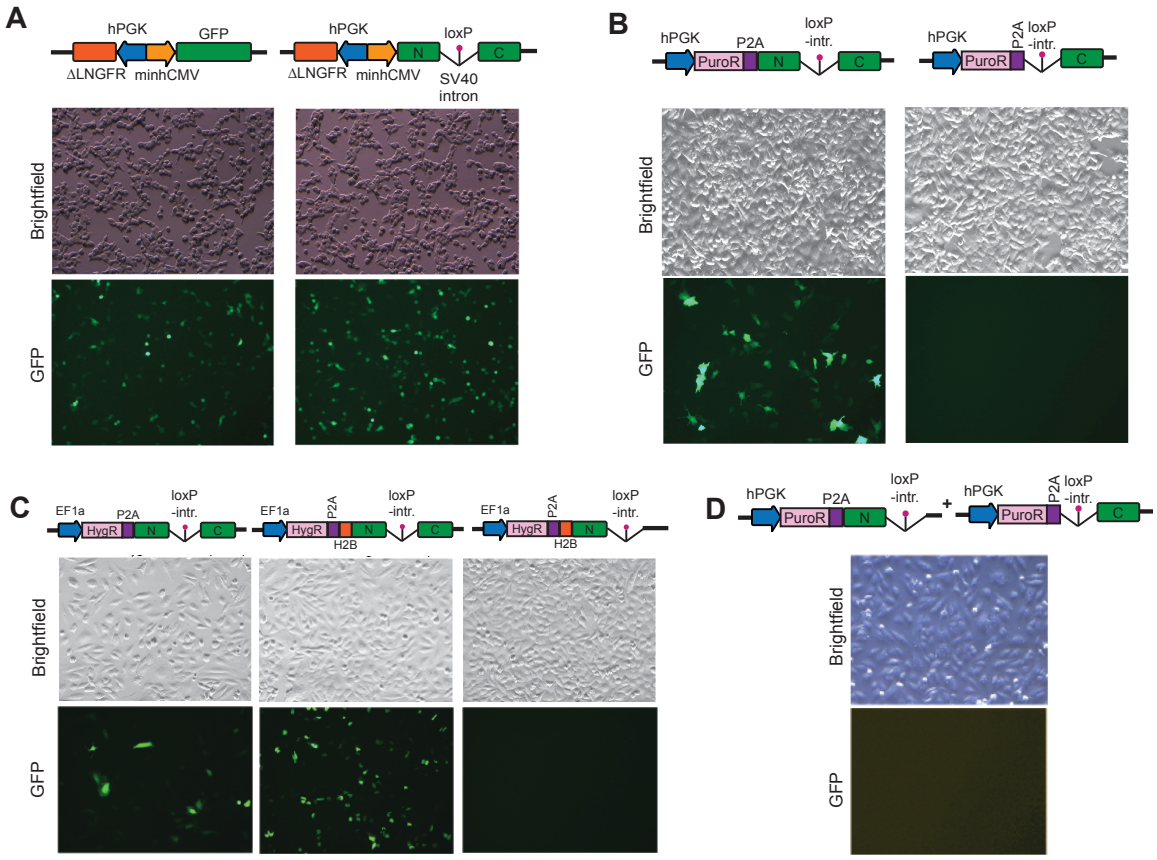


Fig 3.2. Generation and testing of cassettes for split-GFP strategy.
(A) A loxP-containing SV40 intron was cloned into the middle of a GFP reporter gene encoded within a previously validated lentiviral construct. Fluorescence imaging of cells transfected with constructs with uninterrupted GFP or intron-containing GFP confirmed no negative impact on expression of the reporter. **(B)** The C-terminus of GFP does not produce a fluorescent signal. **(C)** Addition of H2B to the 3' cassette localizes the reporter signal to the nucleus. Similar to (B), the N-terminus of GFP has no fluorescent signal. **(D)** GFP expression is specific to the unified gene, as co-transfection of the two halves has no signal.

are commonly carried on ecDNAs in glioblastoma and its sequence structure has been recently characterized from a patient-derived glioblastoma cell line, GBM39 (Nathanson et al. 2014; Wu et al. 2019). We therefore sought to recreate the *EGFR*-containing ecDNA found in GBM39 by designing CRISPR guides that target the locus at positions close to the sequenced breakpoints from GBM39. As our choice of breakpoints is limited by guide-cutting efficiency, we did not induce double strand breaks at exactly the same breakpoints as that in GBM39; nevertheless, the guides we chose capture most of the sequence found in *EGFR*-ecDNAs and result in a circular structure of 1.26 Mb (compared to 1.29 Mb in GBM39). In GBM39, the distal breakpoint is located within the first intron of *PSPH*, whereas our distal guide targets upstream within the third intron of the same gene; the difference between the two breakpoints is ~26 kb. There is also a small (<1 kb) difference in the proximal breakpoint between our modeled ecDNA and the ecDNA in GBM39, which occurs within the intergenic region between *SEC61g* and *EGFR* (**Fig 3.3A**).

For our *in vitro* experiments, we used HCT116, a human colorectal cancer cell line with a stable karyotype that is readily amenable to transfection. While testing the cutting efficiency of our guides, we demonstrated that by simply co-transfecting both guides, we could detect by genotyping PCR a product suggestive of circularization of the intervening region (**Fig 3.3B**), implying that the concurrent introduction of two double strand breaks followed by ligation repair is in principle sufficient to induce circularization of a genomic locus. However, we must note that

the same product, with the same junction breakpoints, would also be generated by tandem duplication. Therefore, genotyping PCR analysis by itself is insufficient to differentiate between these two outcomes.

Having confirmed the cutting efficiencies of our guides, we next targeted our cassettes into the proximal and distal breakpoints by sequential targeting. We co-transfected HCT116 with Cas9-RNPs targeting the desired insertion site and the 3' donor cassette in the form of a PCR-amplified product with 40 - 80 bp homology arms (**Fig 3.3C**). Positive clones were selected with hygromycin. After isolating a clone targeted at the first breakpoint, the 5' cassette was targeted to a second breakpoint, followed by selection with puromycin. Using this strategy, we generated 8 independent double-targeted clones and validated the insertion of the cassettes by genotyping and sequencing across the entire cassette (**Fig 3.3D**). By genotyping, all clones appeared to harbor a heterozygous insertion of each cassette.

3.2.4 Cre-induced recombination leads to reconstitution of the GFP reporter

Having obtained double-targeted isogenic clones, we next tested the ability of our system to reconstitute the GFP reporter upon expression of the Cre recombinase (**Fig 3.4A**). It is known that Cre-loxP recombination efficiency decreases over increasing genetic distances and that the frequency of recombination between *trans* loxP sites is also several degrees of magnitude smaller than for sites placed in *cis* (Zheng et al. 2000; Nagy 2000; Yu and Bradley 2001).

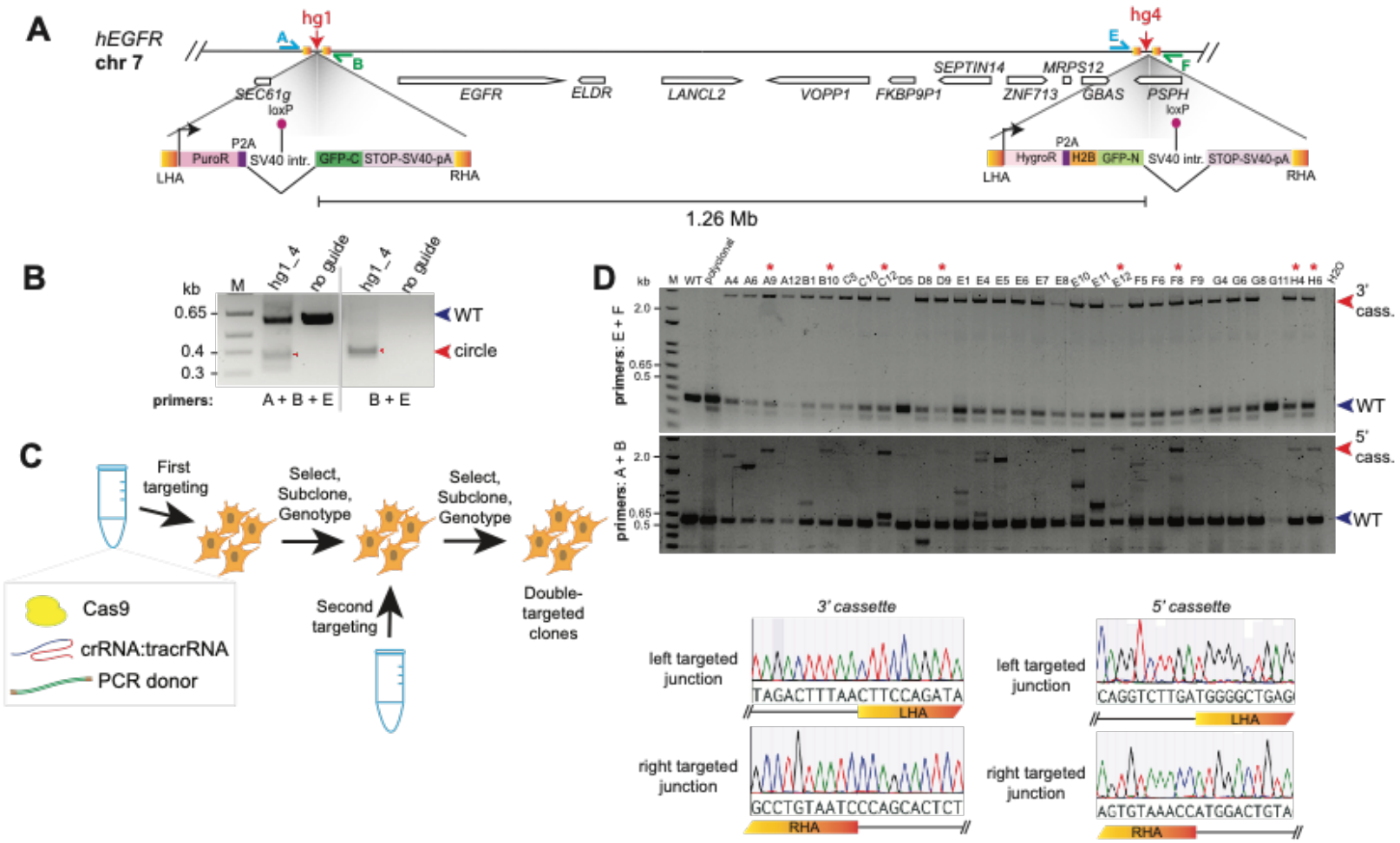


Fig 3.3. Generation and verification of HCT116 double-targeted clones.

Fig 3.3. (cont'd). (A) Overview of the *EGFR* genomic locus targeted in our *in vitro* studies. Red arrowheads mark the locations of the breakpoints created by our gRNAs and indicate where the cassettes are inserted. In double-targeted clones where the cassettes are inserted *in cis*, Cre-mediated recombination will induce circularization of the 1.26 Mb region. **(B)** Co-transfection of the guide pair is sufficient to induce formation of PCR products suggestive of circularization. **(C)** Schematic illustrating our sequential targeting strategy to generate double-targeted clones. **(D)** 8 independent double-targeted clones (indicated by red asterisks) were generated with our strategy. PCR genotyping using primers that span across the entire cassette (refer to (A)) confirm insertion of the 5' and 3' cassettes. The insertion junctions in all 8 clones were validated by sequencing.

Due to their lack of centromeres, ecDNAs are thought to segregate unevenly to daughter cells following mitosis (Turner et al. 2017; Verhaak, Bafna, and Mischel 2019; deCarvalho et al. 2018). We reasoned that our system would allow us to capture this behavior, as differences in copy number of the reconstituted reporter would translate to differences in fluorescence intensities between cells. Importantly, because of the unequal division of ecDNAs during each mitosis, a daughter cell may not receive any copies of the reporter from the parent cell, and therefore become GFP-negative. With regards to our system, we therefore expected that a single GFP-positive cell, harboring one or two copies of the reconstituted, circularized reporter, would eventually give rise to a mixed population of GFP-positive and GFP-negative cells (**Fig 3.4A**). Furthermore, within this colony population, GFP-positive cells should display a mosaic intensity of GFP expression. Importantly, this behavior would be distinct from the reconstitution of the reporter via tandem duplication, as in this case GFP expression would instead be passed on via Mendelian chromosomal inheritance, resulting in each daughter cell having a copy of the reporter, and thus a population of uniformly GFP-positive cells.

The addition of Cre recombinase to our cells resulted in GFP expression in some, but not all, of the 8 tested clones. Clones A9, B10, C12, E12, F8, H4, and H6 generated comparable GFP expression levels (i.e. no or low expression). We thus chose to focus on the first four clones of the 8 total (A9, B10, C12, D9). By PCR genotyping analysis using divergent primer pairs and multiplexed primer

combinations, we confirmed the formation of the recombined junction in clones A9 and D9, although as previously described, genotyping PCR analysis cannot distinguish between true circularization or tandem duplication, the latter of which can occur due to recombination between homologous chromosomes or sister chromatids (**Fig 3.4C**). By flow cytometry analysis, the expression of the reconstituted GFP reporter was confirmed in the clones (**Fig 3.4D**). The percentage of GFP-positive cells can be taken as an indicator of recombination efficiency, and the relatively robust induction of GFP expression in clone D9, compared to clone A9, corroborates the results from genotyping PCR, where the band corresponding to the recombined product in D9 is several times stronger than in A9 (**Fig 3.4B**). We noted some variation in the percentage of GFP-positive cells in clone D9 following Cre treatment in independent experiments, with efficiencies ranging from 1.2 – 6.3% (not shown), though a GFP-positive population was always consistently produced upon addition of Cre.

3.2.5 GFP-positive cells lose GFP expression over time and increase GFP expression in response to selective pressure

In order to interrogate the behavior of recombined cells, we isolated the GFP-positive population from clones A9 and D9 by FACS. Five days post-sorting, we observed that the population that arose from sorted D9 cells showed a mosaic pattern of GFP intensity (**Fig 3.5A**). Importantly, a significant number of cells were GFP-negative, which is in concordance with the notion that the reporter becomes

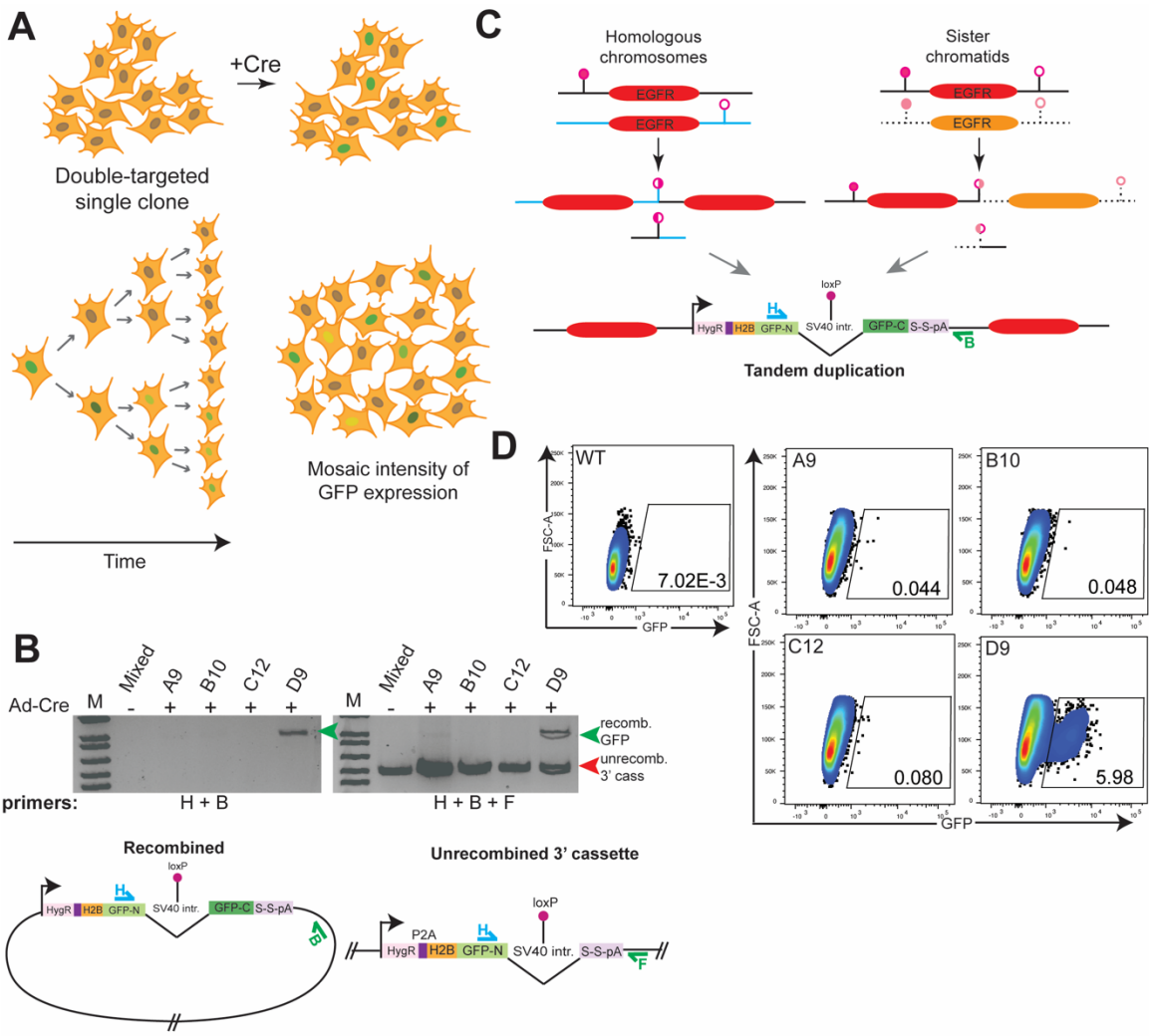


Fig 3.4. Cre recombinase induces unification and expression of the reporter in double-targeted clones.

(A) Schematic illustrating the expected outcome of Cre treatment on double-targeted cells is GFP expression in a subset of cells (top). If the reporter is autonomously replicated and unequally partitioned between daughter cells after each cell division, starting with a single GFP-positive cell, the resulting population after many rounds of division should display variation in GFP expression; importantly, a fraction of cells should lose expression of the reporter (bottom). **(B)** Infection of cells by recombinant adenoviruses expressing Cre recombinase (Ad-Cre) leads to formation of the recombined product in two of the 8 clones (A9 and D9; only four clones are shown for simplicity). The schematic at bottom illustrates the position of the primers used in the genotyping PCR to detect the recombined product. **(C)** Genotyping PCR analysis cannot distinguish between true circularization or tandem duplication, which can arise through recombination between homologous chromosomes or sister chromatids. **(D)** Cre treatment induces robust GFP expression in D9 compared to other clones. Plots show results from a representative flow cytometry analysis.

lost due to unequal segregation during cell division. Indeed, flow cytometry analysis of the D9 population a week after initial sorting revealed that the majority of cells within the population had indeed “reverted” back to a GFP-negative state (**Fig 3.5B**). In contrast, while only 0.12% of cells from the isogenic line A9 were initially GFP-positive, the majority of cells (~90%) remained GFP-positive despite further passaging (**Fig 3.5B**).

We also subjected sorted D9 cells to increasing concentrations of hygromycin to observe their response to selective pressure. Because the hygromycin resistance gene is retained within the circle and expressed from the same promoter as reconstituted GFP, we reasoned that increasing the concentration of hygromycin would lead to an increase in GFP intensity as cells that accumulate multiple copies of the circle (i.e. multiple copies of *hygroR*) would better survive increasing intensity of selection. A few days after exposure to different concentrations of hygromycin, we observed an increase in the percentage of GFP-positive cells in proportion to hygromycin concentration (**Fig 3.5C**). Moreover, the mean fluorescence intensity of the GFP-positive subset also increased in step with drug concentration (**Fig 3.5C**). Notably, the persistent presence of a GFP-negative subset at all concentrations indicated that there is probably another copy of the 3’ cassette located on the other allele, which is providing resistance to hygromycin. We note that wild type, non-targeted HCT116 did not survive any of the tested hygromycin concentrations (not shown).

In order to determine if the increase in GFP expression was due to copy number increase of the reporter, we performed a quantitative PCR (qPCR) analysis on genomic DNA extracted from the GFP-positive subset of cells from each concentration to investigate changes in gene copy number. We did not observe a fold change increase in GFP copy number in hygromycin-treated cells relative to the no hygromycin condition (**Fig 3.5D**), suggesting that the increased reporter expression is instead due to an upregulation in transcriptional output, and that this increased output is also likely modulating the cells' ability to survive higher concentrations of hygromycin.

3.2.6 Cre treatment leads to different outcomes in double-targeted HCT116 clones

To directly visualize ecDNA in our cell lines, we performed metaphase and interphase fluorescence in situ hybridization (FISH) on Cre-treated, sorted cells. Cre-mediated recombination can lead to heterogeneous karyotypic outcomes depending on the cell cycle stage at which recombination occurs. At the G1 phase of the cell cycle, recombination between two loxP sites *in cis* will result in chromosomal deletion and circularization of the excised locus (the desired outcome). If the loxP sites are targeted in *trans*, the karyotype of the cell will reflect a tandem duplication on one allele with a concurrent deletion of the other allele copy. Cre-mediated recombination can also take place after DNA replication has occurred (S/G2 phase). After replication, four loxP sites are present in a double-

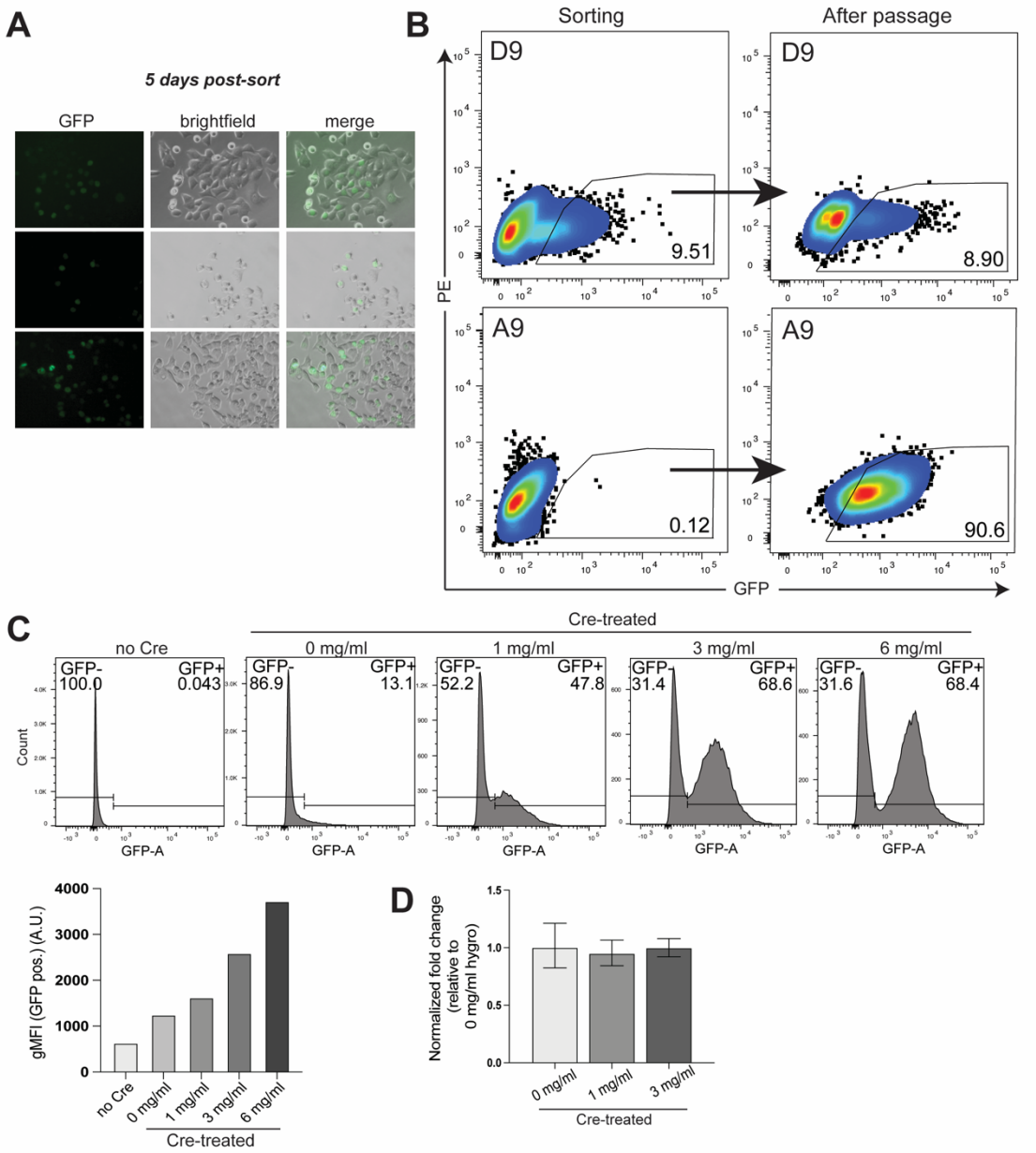
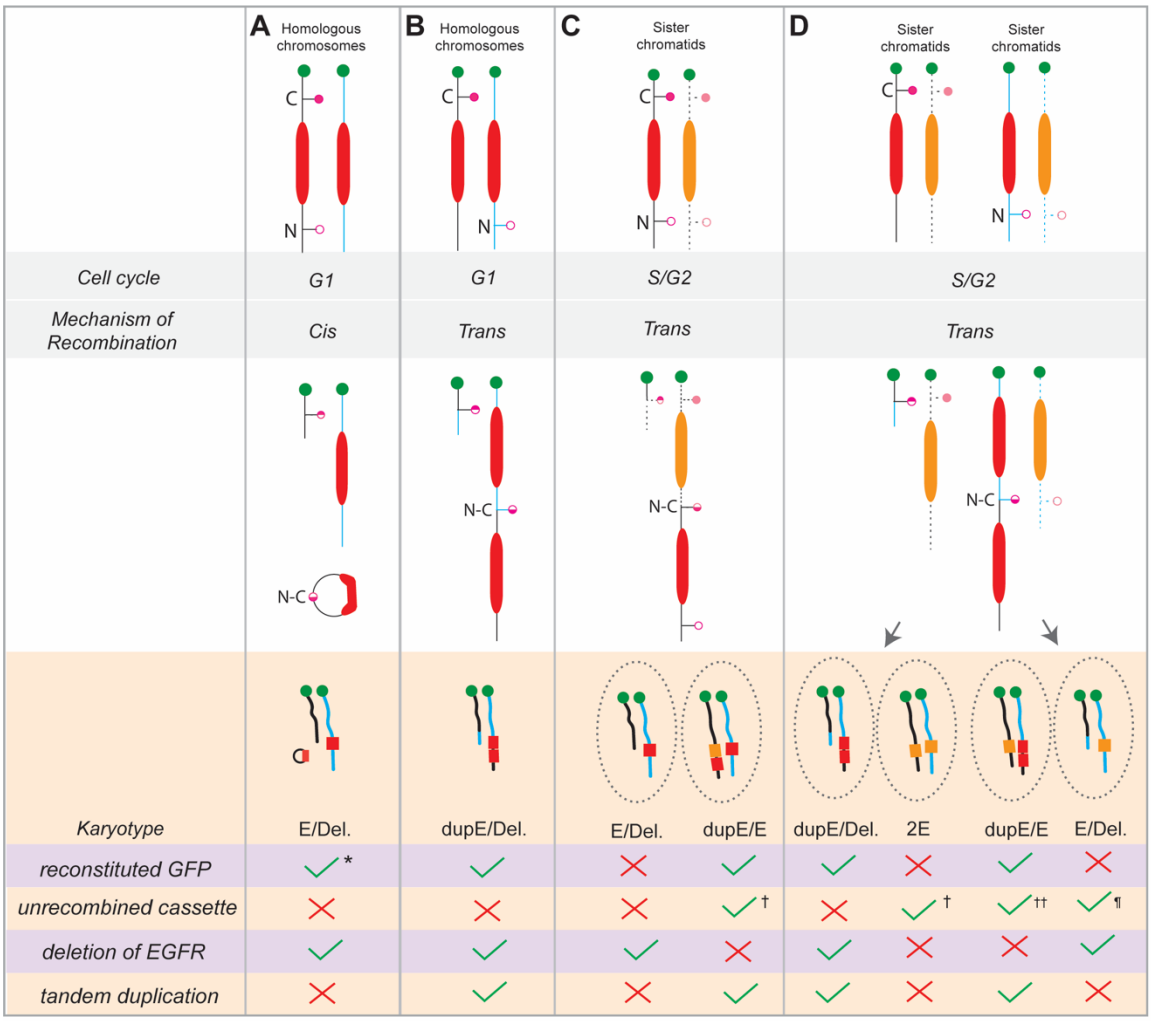


Fig 3.5. Recombined cells lose reporter expression when passaged under neutral selection and increase reporter expression in response to selective pressure.
(A) GFP-positive D9 cells display mosaic expression of GFP upon further culture, with notable loss of GFP expression in a subset of cells. **(B)** The majority of cells that grew out from an initially GFP-positive sorted population of D9 cells no longer express GFP, whereas cells that grew out from a very small GFP-positive fraction of Cre-treated A9 cells retained GFP expression. **(C)** Exposure of sorted D9 cells to increasing concentrations of hygromycin resulted in a corresponding increase in the percentage of GFP-positive cells (top), as well as an increase in the geometric mean fluorescence intensity (gMFI) of the GFP-positive subset (bottom). **(D)** No change in GFP copy number was detected in genomic DNA extracted from GFP-positive cells exposed to increasing concentrations of hygromycin. Error bars reflect s.d. of each measurement.

targeted cell, and recombination can occur between sister or non-sister chromatids depending on whether the loxP sites are *in cis* or *in trans*. These may lead to several distinct recombination outcomes, as illustrated in **Fig 3.6**.

Unexpectedly, our FISH analysis revealed that our parental HCT116 cell line is tetraploid, with four copies of chromosome 7, thus deviating from the published diploid karyotype (**Fig 3.7A**) (Thompson and Compton 2008; Lengauer, Kinzler, and Vogelstein 1997). Nevertheless, FISH analysis of metaphase spreads and interphase cells from our D9 line confirmed a deletion of one of the four copies of *EGFR* after Cre treatment (3E/Del.) (**Fig 3.7B**). This suggests Cre successfully mediated recombination between two loxP sites located in *cis*. In contrast, the majority of cells from clone A9 possessed a tandem duplication of the *EGFR* locus on one copy of chr7 with a concomitant deletion of the locus on another copy (2E;1dupE/Del.), indicating that *trans* recombination had occurred between two loxP sites located on homologous chromosomes (**Fig 3.7B**). Despite evidence of Cre-mediated excision of the locus in clone D9, we were not able to recover *EGFR*-containing ecDNA structures in any metaphase spreads.

In spite of the absence of ecDNA structures, these findings are in line with our earlier results. In particular, the persistent reporter expression in passaged A9 cells after initial isolation of rare GFP-positive cells is explained by the fact that the GFP reporter was reconstituted by tandem duplication, which is thus propagated to all daughter cells. In addition, the low frequency of GFP expression following Cre treatment – a proxy for recombination efficiency – is an indication that the



* conditional on presence of the circle
† both 5' and 3' cassettes
†† 5' cassette only
¶ 3' cassette only

Fig 3.6. Cre-mediated recombination can result in different recombination outcomes in our double-targeted cells.
(A) Recombination between two loxP sites located *in cis* during G1 phase results in chromosomal deletion and circularization of the excised locus (the desired outcome of our system). **(B)** Recombination between loxP sites targeted *in trans* will result in tandem duplication on one allele, with concurrent deletion of the other copy of *EGFR*. **(C-D)** Cre-mediated recombination during the S/G2 phase can result in additional outcomes. **(C)** Recombination between sister chromatids can occur if loxP sites are targeted *in cis*, leading to different karyotypes in daughter cells (signified by dashed gray ellipses). **(D)** Recombination between non-sister chromatids can occur if loxP sites are targeted *in trans*. The karyotypic makeup of the daughter cells will depend on which chromatids segregate into the same cell (illustrated by the gray arrows).

cassettes are targeted in *trans* in this line. In contrast, the loxP sites are likely located *in cis* in clone D9, as the recombination efficiency is higher than in A9, and Cre clearly mediates excision of the intervening DNA. It is also likely that the D9 clone harbors at least one other copy of the 3' cassette on another copy of chr7, as the GFP-negative fraction that arose after sorting for GFP-positive cells is able to survive hygromycin selection (**Fig 3.5C**).

Intriguingly, we noted the presence of *EGFR*-positive micronuclei in our samples, though they were also present in D9 cells not treated with Cre (**Fig 3.7C**). Micronuclei are markers of DNA damage and form around acentric chromosomal fragments or entire lagging chromosomes generated during defective mitoses (Krupina, Goginashvili, and Cleveland 2021). Some reports suggest that ecDNA can also be sequestered into micronuclei (Bailey et al. 2020), and that these ecDNA-containing micronuclei are distinct from micronuclei containing chromosomal fragments (Shimizu 2011). Thus, we characterized the micronuclei in our samples by either their inclusion or exclusion of Cen7, as the inclusion of Cen7 is more suggestive of the presence of the whole chromosome, while the exclusion of Cen7 may be indicative of acentric chromosomal fragments or ecDNAs. We found that Cre-treated D9 samples had a higher occurrence of micronuclei without Cen7 compared to untreated D9 (**Fig 3.7C**). Within these Cen7-absent micronuclei, there were instances of micronuclei with more than 4 *EGFR*-positive signals only in the Cre-treated condition (**Fig 3.7C**).

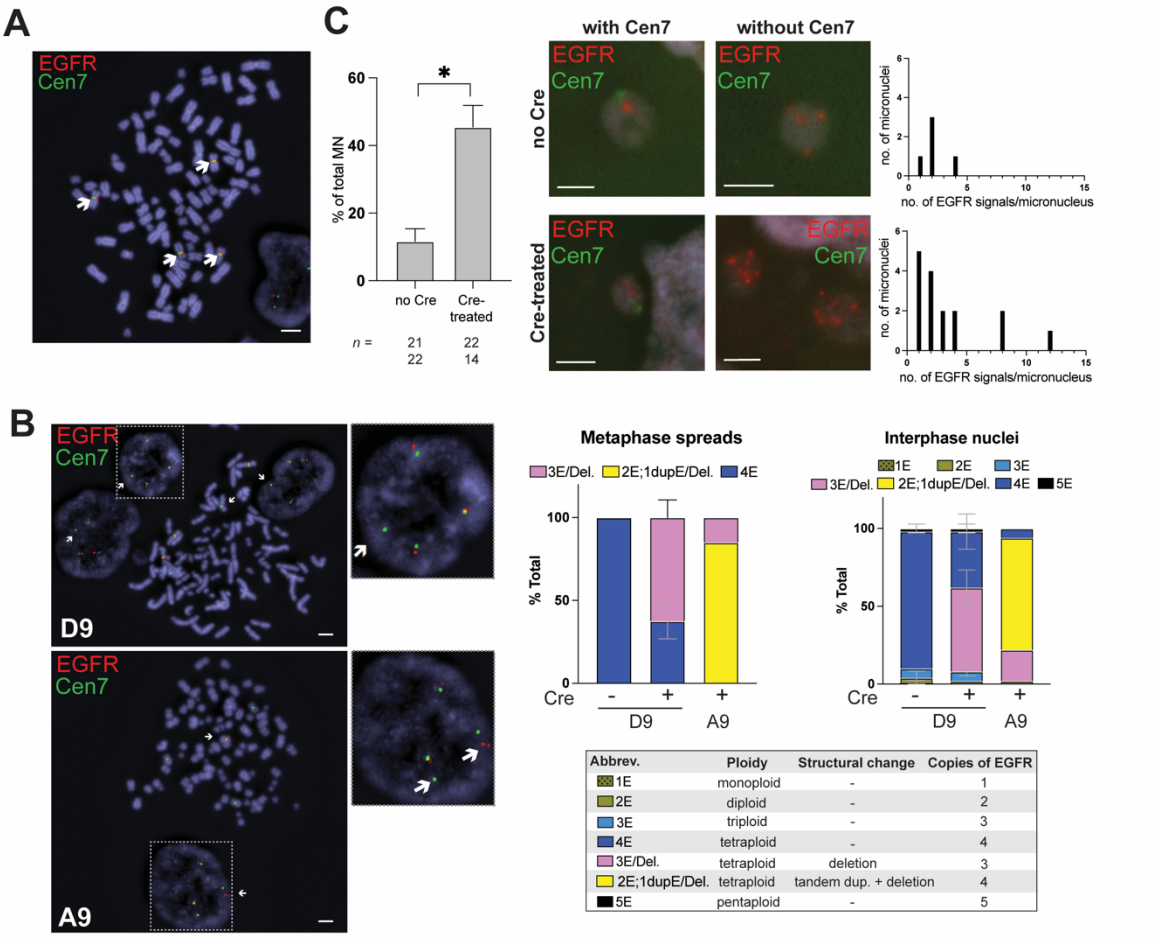


Fig 3.7. Cre-treatment results in excision or tandem duplication of the *EGFR* locus in double-targeted clones.

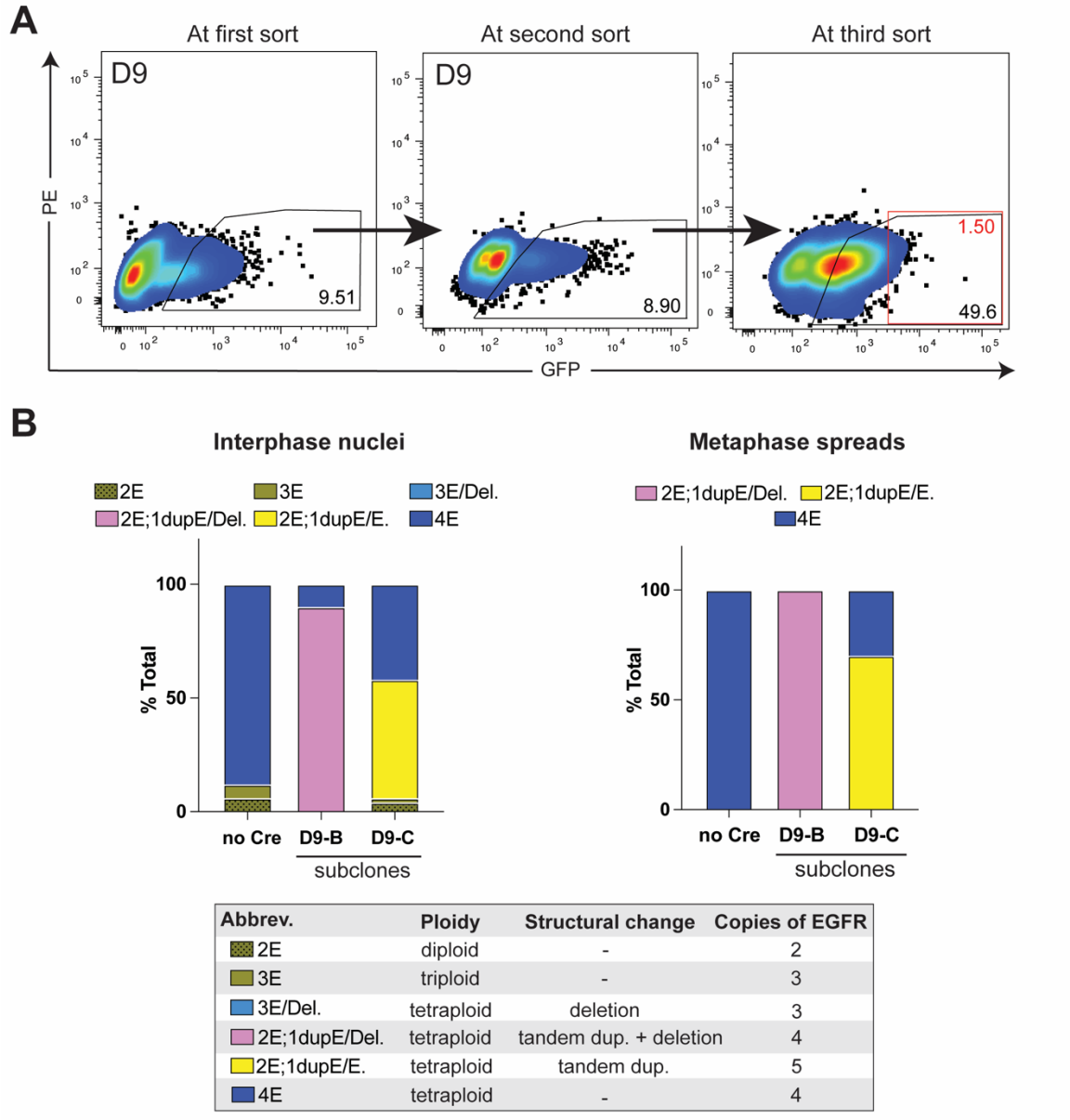
(A) FISH analysis revealed that our parental cell line is cytogenetically abnormal, displaying near-tetraploidy with four copies of chromosome 7 (white arrowheads). **(B)** (Left) Metaphase and interphase FISH analysis showed that Cre recombinase induced deletion of one copy of *EGFR* in D9 cells (inset, white arrowhead) and tandem duplication with concurrent deletion in A9 cells (inset, white arrowhead). (Right) Quantification of pooled counts from separate metaphase and interphase FISH analyses. n = 1 or 2 different analyses. Error bars represent one s.d. from the mean. **(C)** (Left) Cre-treated D9 cells displayed a higher occurrence of micronuclei without Cen7 compared to untreated D9. Quantification reflects counts from 2 independent samples per condition. Error bars represent one s.d. from the mean. *, $p<0.05$, unpaired two-tailed t-test. (Middle) Representative FISH images of micronuclei with or without Cen7 in Cre-treated and untreated D9 samples. (Right) Frequency distribution of the number of EGFR-positive signals within Cen7-absent micronuclei in Cre-treated and untreated D9 samples (pooled from independent samples). All scale bars in this figure = 5 microns.

3.2.7 Sequential sorting of GFP-positive D9 cells selects for tandem duplication

We attempted to enrich for circles by sequentially sorting GFP-positive D9 cells. We reasoneds that if the circles were segregating in a random fashion at each division, then the cells that accumulated the highest copy number of circles should also have the highest GFP intensity, and could be identified and isolated by fluorescence-assisted cell sorting (FACS). We thus subjected Cre-treated D9 cells to three rounds of sorting. In the first two rounds, we isolated all GFP-expressing cells. As stated above, the majority of cells that arose from an initial GFP-positive population lost GFP expression upon additional passage, such that the majority of cells at the time of the second sort were GFP-negative. However, at the time of the third sort, almost half of the cells had retained GFP expression (**Fig 3.8A**). The brightest GFP-expressing cells (top 1.5%) were isolated and expanded for FISH analysis. Unlike at earlier timepoints, the cells that grew out from this thrice-sorted population remained uniformly GFP-positive. FISH analysis of two subclones revealed tandem duplication in both, explaining the uniformity of GFP expression (**Fig 3.8B**).

3.2.8 Cre-mediated recombination leads to formation and maintenance of ecDNAs in murine *p53-/-* adult neural stem cells

Our results in HCT116 cells so far indicated that our system was working as expected and inducing a looping-out event of the targeted *EGFR* locus upon



Abbrev.	Ploidy	Structural change	Copies of EGFR
2E	diploid	-	2
3E	triploid	-	3
3E/Del.	tetraploid	deletion	3
2E;1dupE/Del.	tetraploid	tandem dup. + deletion	4
2E;1dupE/E.	tetraploid	tandem dup.	5
4E	tetraploid	-	4

Fig 3.8. Sequential sorting of GFP-positive D9 cells selects for tandem duplication.
(A) Cre-treated D9 cells were subjected to three rounds of sorting to isolate the GFP-positive fraction. After each sort, cells were passaged to allow expansion of the population. The brightest GFP-positive cells (top 1.5%) were isolated from the third sort and subcloned for FISH analysis. **(B)** FISH analysis of two subclones from this GFP^{hi} population revealed tandem duplication in both.

exposure to Cre recombinase. However, cells were seemingly unable to maintain these induced ecDNAs, as we failed to select or enrich for them and were not able to observe them in metaphase FISH analyses. Our findings suggest the existence, in HCT116 cells, of mechanisms preventing the accumulation and propagation of ecDNAs, either via their active elimination, or through impaired survival of ecDNA-containing cells. Early studies on gene amplification that used drug selection to promote spontaneous amplification of endogenous target genes demonstrated that homozygous loss of p53 was required for gene amplification to occur (Livingstone et al. 1992; Yin et al. 1992). HCT116 cells have an intact DNA damage checkpoint, and numerous studies that involve knocking out p53 have been carried out in this cell line (Bunz et al. 1998; Kaeser, Pebernard, and Iggo 2004; Abu El Maaty et al. 2017). Indeed, previous work has shown that tetraploidy can be spontaneously acquired in p53 wild type HCT116 and that tetraploidy tolerance is mediated by overexpression of cyclin D1 (Dewhurst et al. 2014; Crockford et al. 2017).

To explore a potential role for p53 loss in ecDNA tolerance, we targeted our cassettes into adult neural stem cells (aNSCs) derived from a *p53-/-* mouse. This system has been previously employed in our lab to model the chromosomal rearrangement that generates the *Bcan-Ntrk1* fusion oncogene (Cook et al. 2017). Because we wanted to rapidly assess ecDNA formation in the aNSCs, Cre infection was performed on the polyclonal targeted population. Cre-mediated reconstitution of the reporter was confirmed by flow cytometry and fluorescence

imaging. As with our HCT116 cells, GFP-positive aNSCs were isolated by FACS. We then subjected these GFP-positive recombined cells and a Cre-untreated control to FISH analysis.

Analysis of metaphase spreads revealed the presence of DAPI-positive extrachromosomal DNA even in Cre-untreated cells; importantly, however, these ecDNAs did not contain *Egfr*. The fact that these cells already harbor ecDNAs and are cytogenetically abnormal is perhaps not surprising given their *p53-/-* status. Strikingly, we did observe the presence of *Egfr*-positive extrachromosomal signals only within the metaphase spreads of aNSCs treated with Cre (**Fig 3.9**). Furthermore, the appearance of these *Egfr*-ecDNAs was associated with the loss of chromosomal *Egfr* within the same cell, indicating that the ecDNAs were induced by Cre-mediated recombination of the targeted locus. Further experiments are needed to confirm these preliminary results. If confirmed, they would constitute the first direct evidence of genetically-engineered site-specific ecDNAs *in vitro.*

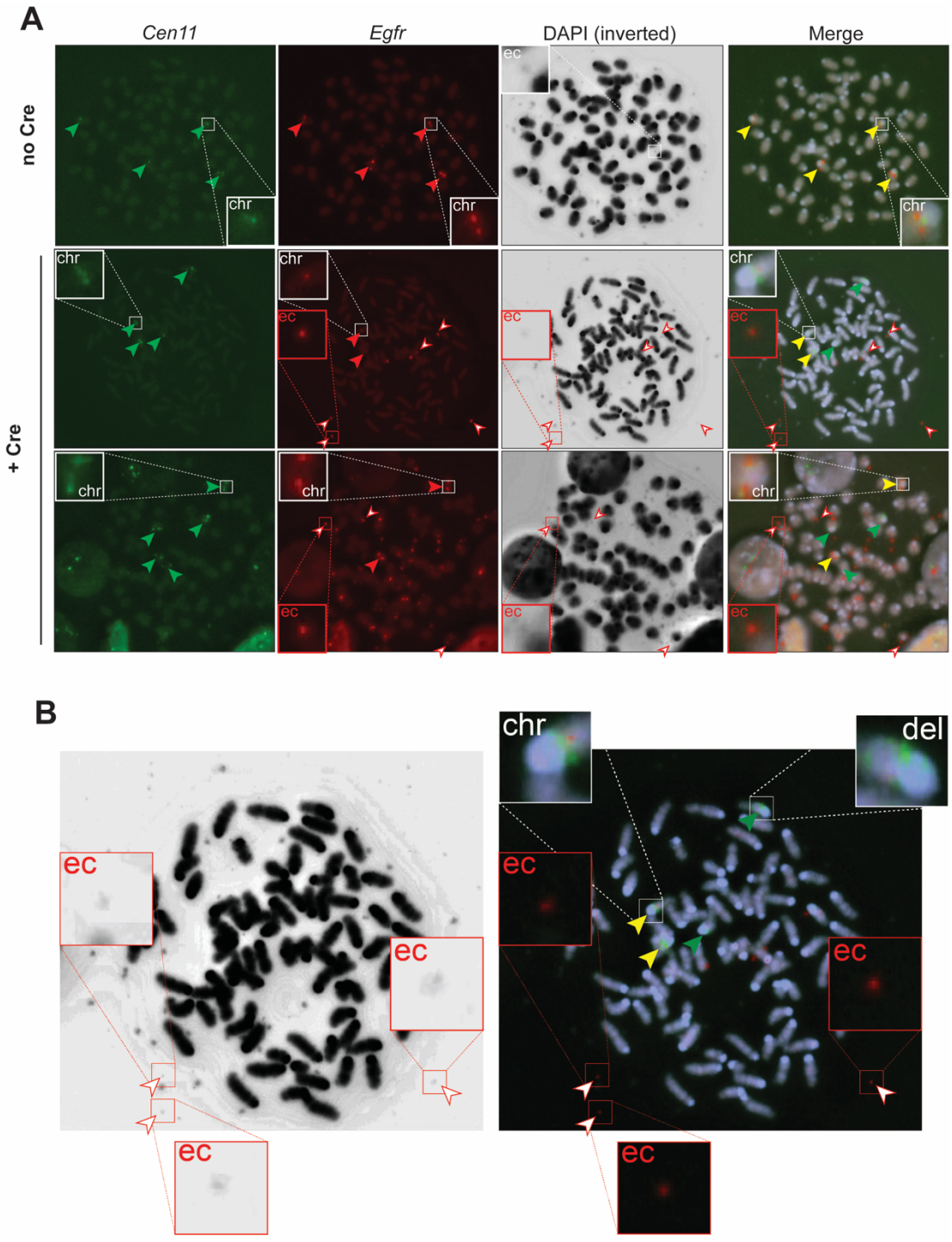


Fig 3.9. Murine *p53-/-* aNSCs are able to form and maintain ecDNAs upon induction by Cre recombinase.

Fig 3.9. (cont'd) (A) Metaphase FISH analysis of Cre-treated and untreated targeted aNSCs demonstrate that upon infection with Cre recombinase, double-targeted aNSCs are able to form and maintain *Egfr*-ecDNAs. The *p53-/-* aNSCs are cytogenetically heterogeneous and abnormal, displaying near-tetraploidy and presence of pre-existing ecDNAs of unknown identity (though *Egfr*-negative). **(Top)** A representative metaphase spread of untreated aNSCs. Green arrowheads indicate four copies of chr11. Red arrowheads indicate chromosomal *Egfr* copies. White arrowheads indicate pre-existing ecDNAs. Yellow arrowheads in the merged image indicate that the *Egfr* signal is chromosomal in nature (that is, red and green signals localize on the same chromosome). Insets are zoomed in on the areas circumscribed by white or red squares; chromosomal and extrachromosomal signals are indicated by 'chr' and 'ec', respectively. **(Middle and bottom)** Representative metaphase spreads of Cre-treated aNSCs. As above, green arrowheads indicate chr11, and red arrowheads indicate chromosomal *Egfr*. White-filled, red-bordered arrowheads indicate extrachromosomal *Egfr* signals that colocalize with DAPI-positive foci. Many displaced (i.e. non-chr11) *Egfr* signals colocalize with, or are located adjacent to, chromosomes. **(B)** Magnified view of the Cre-treated cell shown in (A), middle row. **(The work represented in this figure was carried out by Dr. Davide Pradella.)**

3.3 Discussion

In this chapter, I have described in detail our strategy for modeling ecDNAs harboring oncogenes *in vitro*, which can be easily adapted for the manipulation of mouse embryonic stem cells (mESCs) for the generation of germline mouse models to study the impact of ecDNA formation on tumor initiation (see Chapter 4). We note that while we have chosen to target the *EGFR* locus in our proof-of-concept experiments, our strategy would in principle allow any genomic locus of choice to be targeted. Our strategy therefore opens up new research opportunities for investigating the role of additional ecDNA-amplified elements, such as other oncogenes commonly found on ecDNAs, as well as enhancers, which have also been found to be localized on ecDNAs and may function as mobile regulatory elements to drive genome-wide transcriptional amplification from chromosomal oncogenes (Zhu et al. 2021).

We further demonstrate the generation of double-targeted human and mouse cell lines harboring our loxP-containing cassettes via CRISPR/Cas9 induced double strand breaks and tandem delivery of donor sequences. Importantly, we have verified that our inserted cassettes are able to undergo Cre-mediated recombination to result in the reconstitution of a H2B-GFP reporter. Using H2B-GFP as a positive marker for recombined cells, we have been able to follow the outcomes of the reporter in recombined cells as they progress through additional passages and are exposed to selective pressure. In particular, we

observed that the behavior of two of our isogenic double-targeted HCT116 lines were markedly different following Cre treatment. In our D9 clone, we observed a loss of the reporter signal over time, in concordance with the predicted behavior of unequally-segregating ecDNAs due to their acentric nature. In contrast, we observed persistent reporter signal in our A9 clone throughout continued passaging.

When subjected to selective pressure in the form of hygromycin treatment, a higher percentage of Cre-treated D9 cells retained reporter signal, and the average reporter intensity also increased in this subset of GFP-positive cells. This suggested that the recombined cells are better able to survive high selective pressure, perhaps due to accumulation of multiple copies of the circle, which would also explain the increase in average reporter signal. However, the increase instead appears to be due to upregulated transcriptional output from the recombined allele, and not an increase in gene copy number. In theory, this increased promoter activity could originate from already-formed circles within a cell, but equally likely is the possibility that the transcriptional output originates from a chromosomal tandem duplication that reconstitutes the reporter. Our experiment is also somewhat confounded by the probable presence of an additional *hygroR*-containing cassette located on another copy of chromosome 7, which would explain the persistence of a GFP-negative population at all drug concentrations. Future improvements can be made to our strategy by replacing our split-GFP design with one that instead splits a resistance gene, such as blasticidin, between

the two cassettes. In this way, only cells that have undergone recombination, whether by tandem duplication or circularization, should persist under selection, thus eliminating other karyotypes from analysis.

Metaphase and interphase FISH analysis shed further light on our results by showing that Cre treatment in the D9 clone results in excision of the intervening *EGFR* locus, implicating recombination *in cis*. In contrast, Cre treatment resulted in a tandem duplication in the A9 clone due to recombination *in trans*, corroborating the persistence of the GFP reporter signal that we observed by flow cytometry. Surprisingly, we did not observe *EGFR*-ecDNAs in the metaphase spreads of D9 cells despite evidence of excision of the recombined junction. Our attempts to enrich for ecDNAs by sequentially sorting cells based on GFP expression resulted in selection of cells harboring the tandem duplication. Collectively, our data indicated that our system leads to the formation of ecDNAs, but that they are not maintained in HCT116 cells.

This raises several interesting possibilities: First, ecDNAs may be actively eliminated by the cells, possibly through sequestration into micronuclei. Indeed, in FISH analysis of interphase nuclei from Cre-treated D9 cells, we observed a small frequency of micronuclei enriched for clustered, ecDNA-like *EGFR* signals. Evidence from the literature supports micronuclei formation as a mechanism by which oncogene-amplified ecDNAs are excluded from the nucleus in tumor cells (Valent et al. 2001; Ambros et al. 1997; Shimizu et al. 1994). These micronuclei have the propensity to rupture due to defects in nuclear envelope structure,

causing the release of ecDNAs into the cytoplasm. This cytoplasmic DNA may then be targeted for degradation by the cellular innate immune pathway, cGAS-STING, which senses cytoplasmic double-stranded DNA and consequently activates pro-apoptotic and pro-senescence programs (Nassour et al. 2019; Rello-Varona et al. 2012; Sun et al. 2013; Mackenzie et al. 2017). Secondly, a permissive genetic background may be required for the long-term maintenance and/or propagation of ecDNAs. Loss of tumor suppressors, such as p53, may be required for cells to tolerate ecDNAs and continue proliferating in their presence.

To quickly explore this possibility, we targeted murine *p53-/-* aNSCs with our cassettes and subjected the polyclonal targeted population to Cre recombinase infection. Our results from this preliminary experiment suggest that p53 loss may help to promote ecDNA tolerance, as we indeed observed Cre-dependent formation of *Egfr*-ecDNAs in this cell line. Future experiments are needed to validate the role of p53 in the regulation of ecDNA tolerance. There may also be other regulators at play whose identities are as yet unknown, but we can adapt our *in vitro* system to screen for gene knockouts that might allow cells to maintain the induced circle. Potential additional candidates for knock-out studies include members of the cGAS-STING pathway. In addition, future experiments should also aim at replicating the targeting strategy in additional permissive cell lines, such as COLO320 DM, a human colorectal cancer cell line in which a fraction of cells stably harbor *MYC*-amplified ecDNAs (Quinn et al. 1979). A comparison of the global gene expression profiles between permissive cell lines may help to converge on

common transcriptional programs underlying their tolerance to ecDNA and thus achieve a greater understanding of the mechanisms underpinning ecDNA maintenance and propagation.

4 Chapter 4: *In vivo* modeling of ecDNAs

4.1 Introduction

Having demonstrated the feasibility of our approach in human and mouse cell lines, we next sought to target the mouse germline genome with a similar strategy in order to ask the question of whether amplification of an oncogene via ecDNAs is sufficient to initiate tumor formation *in vivo*. While compelling evidence demonstrates that oncogene-harboring ecDNAs are involved in tumor progression and drug resistance, it is unknown whether the formation of these ecDNAs occurs as an early and initiating event, or are instead formed as a result of the increased genomic instability of tumor genomes after disease has already been initiated. In order to answer this question, it is crucial to be able to induce the formation of these ecDNAs *in vivo*.

We undertook two complementary and parallel approaches to generate novel mouse models of inducible ecDNAs. In the first approach, undertaken in collaboration with the Mouse Genetics Core Facility at MSKCC, we targeted the insertion of loxP sites into mouse early pronuclear zygotes via *in vitro* electroporation with pre-assembled Cas9 protein and guide RNA complexes (Cas RNPs) in tandem with short single-stranded oligonucleotides (ssODNs) containing loxP sequences as DNA repair templates for precise CRISPR-mediated transgenesis (Hashimoto, Yamashita, and Takemoto 2016). The resulting pups

(founders) are then mated with wild type mice to generate progeny that either have loxP sites inserted *in cis* or a single loxP site; these latter animals can be used as controls for downstream experiments (**Fig 4.2A**). In the second approach, carried out in parallel with the first, the 5' and 3' cassettes comprising the split-GFP strategy (discussed in Chapter 3) were sequentially targeted into mESCs, followed by injection into 8-cell stage blastocysts to generate chimeric mice (in collaboration with the Mouse Genetics Core Facility), which will then be mated with wild type mice to establish germline transmission of double-targeted loci (**Fig 4.3A**).

Additionally, to complement our *in vitro* experiments in human cell lines, we carried out a panel of related experiments to characterize our double-targeted mESCs. We find that our results in mESCs largely corroborates our human data that suggests that ecDNA are unstable structures and that there are likely cell-intrinsic mechanisms that prevent their maintenance and propagation.

4.2 Results

4.2.1 Modeling *Egfr*- and *Myc*-ecDNA *in vivo*

In selecting our oncogenes of interest for our germline mouse models, we chose to target *Egfr* and *Myc*. As described in the previous chapter, *Egfr* is commonly amplified in ecDNAs in glioblastoma, and is encompassed within a 1.29 Mb locus that overlaps other downstream genes including *ELDR, LANCL2, VOPP1, SEPT14,* and others (Wu et al. 2019). In humans, the affected region spans 54.7M – 56M (hg38) on chr7. An interrogation into the degree of conservation between the mouse and human genomes at this locus revealed only partial synteny between the two species. Specifically, the shared syntenic block within this locus extends to chr7 55.3M in the human genome, which falls short of the entire region encompassed in *EGFR*-ecDNA (**Fig 4.1B**). Therefore, it is only possible to achieve partial recapitulation of the sequence structure found in human *EGFR*-ecDNAs in a mouse model. We thus selected guides that would allow us to capture the maximum amount of homology between the two species within the affected region (that is, inclusion of *Egfr* and *Eldr*). Our 5' guide targets a breakpoint downstream of *Sec61g*, as in the human ecDNA, while our 3' guide targets an intergenic region between *Fbxo48* and *Plek* (**Fig 4.1A**). While *Fbxo48* is not found within the human ecDNA, it is located immediately downstream of *Eldr* in the mouse genome and the two genes are divergently oriented, suggesting that a common promoter drives

the simultaneous expression of both genes. Therefore, to avoid any potential confounding effects that may result from disrupting *Fbxo48* expression, we chose to avoid targeting this region, leading to the inclusion of *Fbxo48* in our mouse *Egfr*-ecDNA. Altogether, Cre recombination will result in the circularization of a 0.44 Mb region.

In addition to modeling *EGFR*-ecDNA, we also chose to target the *Myc* locus. Almost 30% of *MYC* amplifications in human cancers are carried on ecDNAs (Kim et al. 2020), and the sequences of *MYC*-containing ecDNAs from COLO320 DM, a human colorectal cancer cell line, have recently been reconstructed from long-read sequencing (Hung et al. 2020). The mouse homologue of *MYC* is also a well-established tumor driver in multiple models of cancer, including liver and ovarian cancers (Kawate et al. 1999; Reyes-Gonzalez and Vivas-Mejia 2021). In contrast to *EGFR*-ecDNAs, which display little structural variation between individual amplicons, *MYC*-containing ecDNAs show a high degree of sequence diversity resulting from differential rearrangement of the *MYC* locus (Hung et al. 2020). One of the most prevalent variants found in COLO320 DM, and the basis of our mouse *Myc*-ecDNA model, is a 1.58 Mb ecDNA amplicon containing the full-length *MYC* coding sequencies that retains several distal regulatory elements, including H3 lysine 27 acetylation (H3K27ac)-enriched sites indicative of active enhancers (Hung et al. 2020) (**Fig 4.1C**). This locus shares high epigenetic conservation between the mouse and human genomes, with H3K27ac marks being preserved (**Fig 4.1C**). We thus chose to target breakpoints that preserve the

full breadth of putative enhancer elements within this locus, including multiple intragenic enhancers located within the long noncoding RNA, *Pvt1*. The resultant ecDNA thus contains the full-length sequence of *Fam84b, Myc, and Pvt1* and is 1.7 Mb in size (**Fig 4.1A**).

4.2.2 Generation of novel mouse models via direct zygote electroporation of CRISPR components

In our first approach to generate novel mouse models of inducible ecDNAs, we collaborated with the Mouse Genetics Core Facility at MSKCC to target loxP sites into murine C57BL/6 in vitro fertilized (IVF) zygotes via electroporation-mediated delivery of CRISPR components. Each electroporation contained a pair of validated crRNAs, generic tracrRNA, and Cas9 protein — pre-assembled into a Cas9-RNP complex — and two single-stranded oligonucleotides (ssODNs) each containing a loxP site, flanked by asymmetric homology arms, the use of which has been shown to increase the frequency of HDR-mediated repair (Richardson et al. 2016). Our choice to generate transgenic animals using this method functions as a complementary approach to our gene targeting experiments in mESCs, which were carried out in parallel. Additionally, we reasoned that insertion of only the loxP sites would avoid any potential confounding effects that may result from using the full cassette, although as discussed earlier in Chapter 3, we do not expect the reconstitution of the GFP reporter to elicit an immunogenic reaction *in vivo*. Importantly, direct zygote electroporation circumvents the need for ES cell

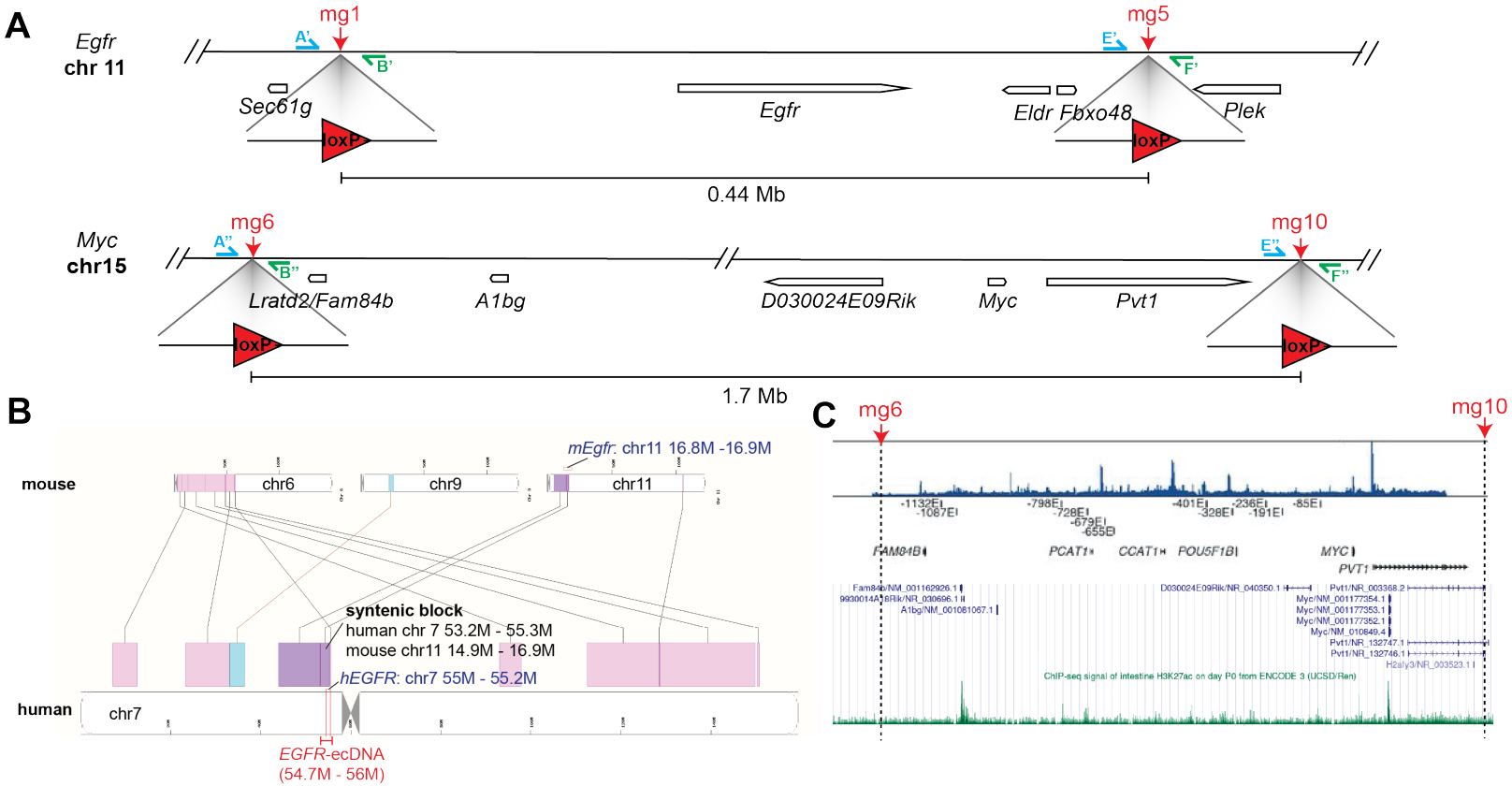


Fig 4.1 Overview of the targeted murine *Egfr* and *Myc* loci in our germ line mouse models.
(A) Schematic showing the mouse Egfr and Myc loci targeted in our germ line mouse models. Red arrowheads mark the locations of the breakpoints created by our guide RNAs and indicate where the loxP sites/cassettes are inserted. Cre-mediated recombination will induce circularization of a 0.44 Mb region in the *Egfr* locus and a 1.7 Mb region in the *Myc* locus. **(B)** Ensembl view of syntenic regions between mouse and human genomes centered on human chr7 (bottom). Colored boxes represent syntenic blocks between the two species. The human *EGFR*-ecDNA (red delimited line) spans multiple syntenic blocks that align to different chromosomes in the mouse genome. **(C)** (Top) H3K27ac ChIP-seq peaks identified by Hung et al. in *MYC*-ecDNA in COLO320 DM cells (Hung et al., 2020). (Bottom) UCSC Genome Browser view of the orthologous mouse locus, showing H3K27ac ChIP-seq signals from day P0 mouse intestinal tissue (green). The mouse-human conserved genes in this region are *Fam84b*, *Myc* (both protein-coding) and *Pvt1* (long non-coding RNA). In both the mouse and human genomes, the region circumscribed by the protein-coding genes constitutes a gene desert.

manipulation and characterization, thus significantly cutting down the time to generate transgenic founders.

Successfully electroporated zygotes bred to viability were genotyped using primers spanning the loxP insertion to identify double-positive founders. The PCR products were sequenced to ascertain correct orientation of the inserted loxP sequences. Because the configuration of the loxP sites (either in *cis* or in *trans*) is unknown in the founders, they require mating to wild type mice to obtain progeny that either will have a single loxP site, or both sites, due to Mendelian segregation of alleles (**Fig 4.2A**). Progeny mice with both loxP sites can therefore be interpreted to have come from a founder where the sites were inserted in *cis*. In contrast, progeny resulting from mice with *trans* insertion of loxP sites harbor only one copy of loxP, which function as useful controls for downstream experiments.

Altogether, seven double-positive, sequence-verified founders for the *Myc* locus (chr15) and one double-positive founder for the *Egfr* locus (chr11) were mated to wild type mice. We obtained experimental and control F1 mice with our desired loxP configurations (**Fig 4.2B**). One pup from each mating was sequenced across the entire incorporated ssODN sequence to confirm correct germline transmission of the loxP site. In addition, we confirmed successful Cre-mediated recombination in tail fibroblasts derived from an F1 mouse harboring a floxed *Myc* allele, and showed that we could detect both recombination products (**Fig 4.2C**).

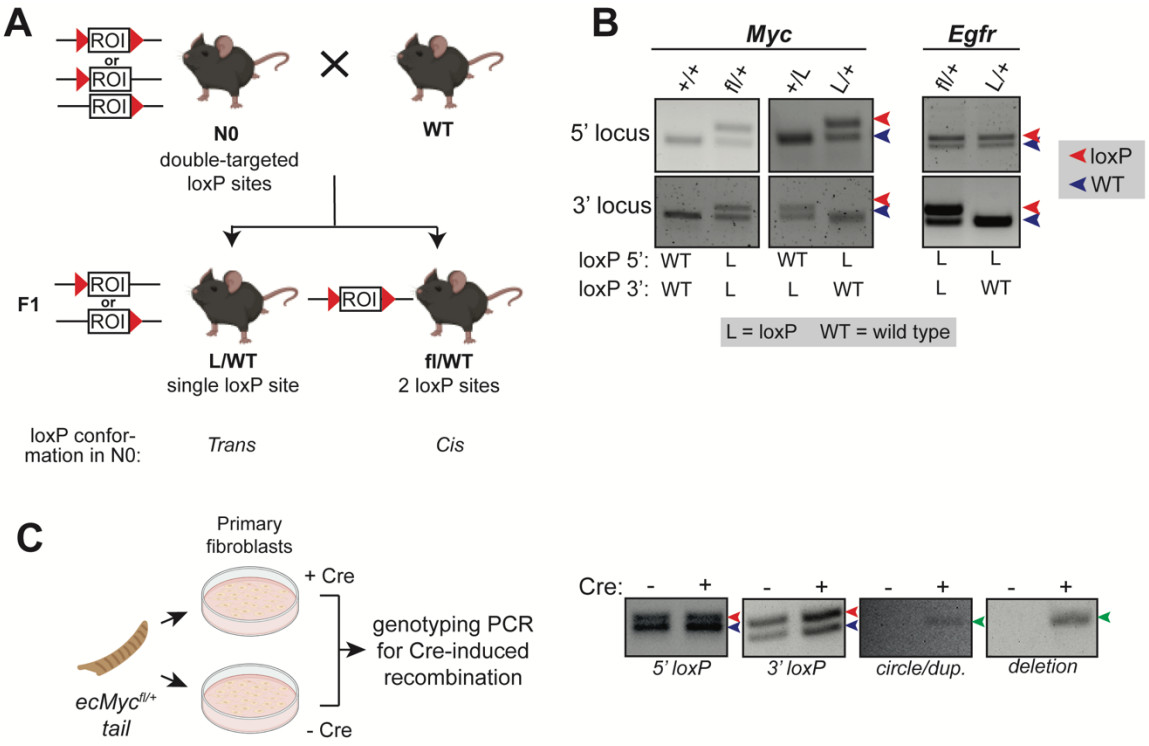


Fig 4.2 Generation of loxP-targeted germ line mouse models using zygote electroporation. **(A)** LoxP sites are targeted into the mouse *Egfr* and *Myc* loci via electroporation of mouse zygotes with Cas9-RNPs and loxP-containing ssODN repair templates. N0 founders were genotyped to confirm double-targeted status, then mated with wild type mice to generate progeny (F1) with either loxP sites *in cis* or a single loxP site. N0 mice with loxP sites targeted *in cis* can also generate wild type progeny (no loxP sites) when mated to wild type animals, which is not shown for simplicity. **(B)** Experimental and control F1 mice with desired loxP configurations were generated from crossing founders to wild type mice. F1 progeny were verified by genotyping PCR. (C) Tail fibroblasts were isolated from a *Myc*$^{fl/+}$ F1 mouse, cultured, and infected with recombinant adenoviruses expressing Cre recombinase. Genotyping PCR confirmed the formation of expected recombination products upon Cre treatment (green arrowheads).

4.2.3 Generation of double cassette-targeted mouse embryonic stem cells by sequential targeting

In parallel, we undertook a second approach to engineer additional, complementary mouse models by utilizing CRISPR/Cas9 to insert our characterized cassettes into mouse embryonic stem cells (mESCs) (**Fig 4.3A**). To accomplish this, mouse V6.5 ES cells were transfected with pre-assembled Cas9-RNPs in tandem with the 5' donor cassette in the form of a PCR-amplified product with 80 bp homology arms. Positive clones were selected with puromycin and sequenced to verify insertion of the cassette. Subsequently, these clones were targeted at the second site to insert the 3' cassette, followed by hygromycin selection. We obtained two independent double-targeted clones each for the *Myc* (hereafter referred to as *ecMyc-1 and* -2) and *Egfr* (hereafter referred to as *ecEgfr-1 and* -2) loci which were sequenced through the entire 5' and 3' cassettes to verify correct loxP orientation and preservation of the two reporter halves (**Fig 4.3B**).

4.2.4 Cre-induced recombination leads to reconstitution of the GFP reporter in mESCs

As we did for the HCT116 and aNSC cell lines, we tested the ability of our system targeted in mESCs to reconstitute the GFP reporter upon addition of Cre recombinase. We verified that Cre induced GFP expression in both *ecEgfr-1* and *ecEgfr-2* as assessed by flow cytometry (**Fig 4.3C**). In comparison, the recombination efficiency appeared to be lower for the *ecMyc* clones, which is

perhaps attributable to the increased distance between the two targeted loxP sites. By genotyping PCR analysis using multiplexed primers, we confirmed the formation of the recombined junction in both *ecEgfr* clones following Cre treatment, but could only detect the recombined product in one of the *ecMyc* clones (*ecMyc-1*) (**Fig 4.3D**). Due to this, we chose not to carry *ecMyc*-2 forward to downstream experiments. To confirm correct loxP-mediated recombination, we sequenced across both of the recombination products (i.e. the reconstituted reporter as well as the deletion) (**Fig 4.3D**).

4.2.5 Confirmation of ES cell targeting by Southern blot analysis

To confirm site-specific and single integration targeting in our ES cell clones, we performed Southern blot analysis using external probes specific for the proximal and distal breakpoints of each locus as well as internal probes against the 5' and 3' cassettes. Although we were able to confirm site-specific targeting in the clones by PCR amplification across the cassette into the genomic locus, Southern blot analysis provides invaluable information by allowing interrogation of the genomic structure of the non-targeted allele. It is also an essential method for verifying integration copy number, as this information cannot be provided by PCR analysis.

We confirmed heterozygous insertion of the 5' and 3' cassettes in both *ecEgfr* clones (**Fig 4.4A, B**). Our analysis also revealed a deletion (<1 kb) spanning the 5' breakpoint in the non-targeted allele. Because both clones were derived from the same parental, 5' cassette targeted clone (F11), both harbor the same deletion.

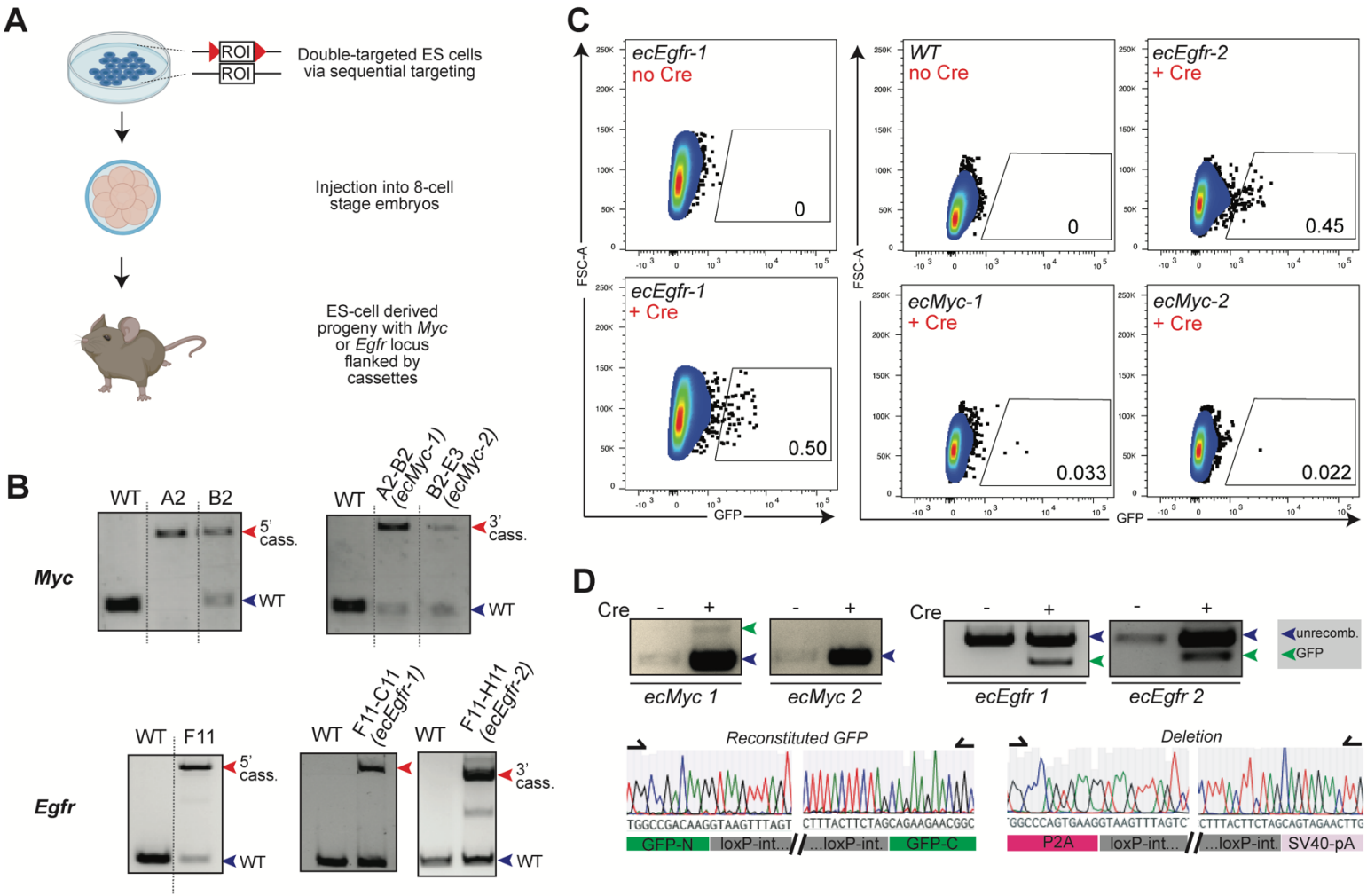


Fig 4.3 Generation and verification of cassette-targeted mESCs.

Fig 4.3 (cont'd). (A) In a parallel approach to generate novel germ line mouse strains, mouse embryonic stem cells are targeted with our split-GFP cassettes through sequential targeting. Following characterization of these ES cell clones, they will be injected into 8-cell stage embryos and transferred to pseudopregnant foster mothers. The resulting progeny will be ES-cell derived mice with *Myc* or *Egfr* loci flanked by our cassettes. **(B)** Genotyping PCR of mESC clones generated by sequential targeting. Genotyping primers anneal within genome sequences that flank the cassette insertion sites, and thus span the entire cassette in targeted cells. Because the mESCs are cultured on mouse embryonic fibroblasts (MEFs), the wild type band can come from genomic DNA of the mESCs or MEFs. **(C)** Cre recombinase induces GFP expression in our mESC clones as assessed by flow cytometry. **(D)** (Top) Genotyping PCR analysis using multiplexed primers confirms the formation of the reconstituted reporter in both *Egfr* clones and one of the *Myc* clones. (Bottom) DNA chromatogram showing sequencing results spanning the recombined junctions in both recombination products (reconstituted GFP and deleted allele).

As we will be selecting for germline transmission of our targeted alleles, this deletion allele can be selected out within the resulting progeny. Within the *ecMyc* clone, the 3' cassette was confirmed to be heterozygously inserted, whereas the 5' cassette appears to be homozygously targeted (**Fig 4.4A, C**). Furthermore, our analysis confirmed single-site integration of the cassettes in both *ecEgfr* and *ecMyc* clones (**Fig 4.4B, C**).

4.2.6 Recombined mESCs show mosaic expression of the reconstituted reporter and display heterogeneous karyotypes

Following Cre-induced expression of the reporter in *ecEgfr* clone 1, we isolated the GFP-positive population by FACS in order to follow the behavior of these recombined cells through further passages. By fluorescence microscopy, we observed that the resulting colonies that grew out from sorted, GFP-positive single cells were mosaic mixtures of reporter-positive and -negative cells (**Fig 4.5A**), similar to what was observed in our HCT116 cell lines. Furthermore, within a single colony, the intensity of GFP expression also differed between cells.

We next subjected the sorted cells to different concentrations of hygromycin to observe their response to selective pressure. By both fluorescence microscopy and flow cytometry, we observed an increase in the geometric mean fluorescence intensity (gMFI) of cells subjected to the highest amount of selective pressure (1 mg/ml hygromycin) compared to cells treated with 0.15 mg/ml hygromycin (**Fig**

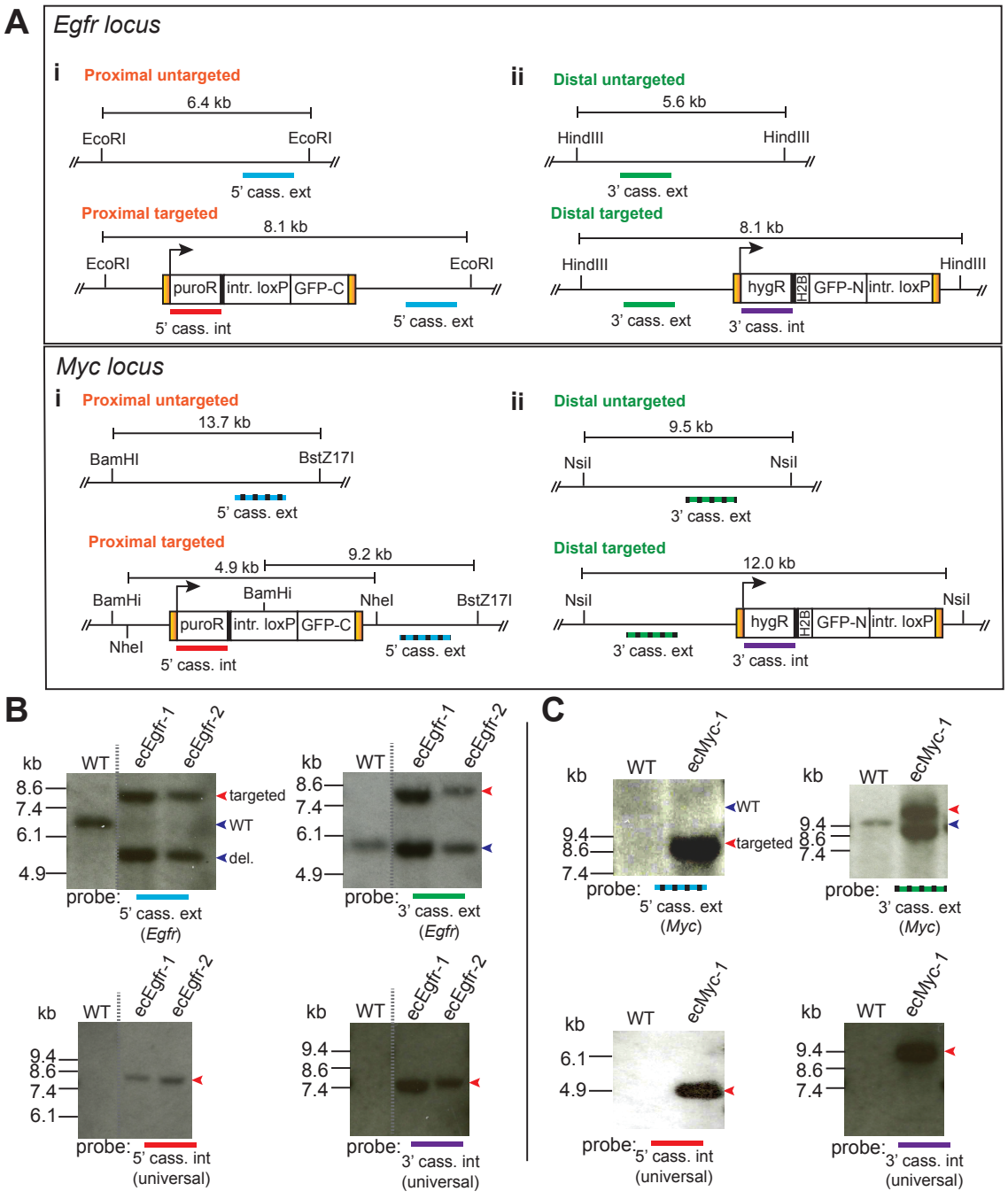


Fig 4.4 Confirmation of mESC targeting by Southern blot analysis.
(A) Schematic of proximal and distal targeted loci in the *Egfr* and *Myc* loci showing the expected fragment sizes, restriction enzymes, and probes used in Southern blot analysis of mESC clones. **(B)** Southern blot analysis of *Egfr* clones showing site-specific and single integration targeting. **(C)** Southern blot analysis of the *Myc* clone showing site-specific and single integration targeting.

4.5B, C). As expected, cells that did not receive Cre did not show any GFP expression and did not survive past 0.15 mg/ml hygromycin (not shown).

Interphase and metaphase FISH using *Egfr*-specific probes was performed on sorted cells that were either treated with the highest concentration of hygromycin, or were not exposed to selective pressure. Heterogeneous karyotypes were observed in cells not exposed to hygromycin. Nearly 50% of analyzed cells harbored a deletion of one copy of *Egfr* (E/Del.), while the majority of the remaining cells harbored a tandem duplication of *Egfr* on one allele without a concomitant loss of *Egfr* (dupE/E; **Fig 4.5D**). Notably, treatment with hygromycin strongly selected for cells with the tandem duplication (**Fig 4.5D**), with the majority of analyzed cells possessing this karyotype.

The results of the FISH analysis suggest that the loxP sites are targeted *in cis* in this clone. The karyotypes can be explained by Cre-mediated recombination occurring in G1 phase in some cells within the population, and during S/G2 in others. Post-replication recombination would give rise to daughter cells with differing karyotypes (**Fig 3.6C**), about half of which would contain the tandem duplication (dupE/E) that reunites the two reporter fragments, and the other half would contain deletion of one copy of *Egfr* (E/Del.). Importantly, these cells would never express the reporter. Because our cells were sorted for GFP expression after Cre-induced recombination, it is therefore likely that the sorted population initially contained a mixture of cells harboring the tandem duplication (due to post-replication recombination) or the circularized reporter (recombination during G1).

Both types of cells would be GFP-positive in the beginning; however, with the addition of selective pressure, the cells carrying the tandem duplication are selected over those harboring the circle. Similar to our findings in HCT116, there appears to be an intrinsic suppressive mechanism against the maintenance and/or propagation of these circular structures in cells. The genetic basis for this mechanism is as yet unclear, but is of great interest for future investigation.

We further note that the karyotypes isolated from our cells could also technically be explained by post-replication recombination between non-sister chromatids when the loxP sites are inserted *in trans* (**Fig 3.6D**). Again, because our cells were sorted for GFP expression, we find this explanation unlikely. In addition, the efficiency of recombination been non-sister chromatids (a *trans* configuration) is several orders of magnitude lower than recombination *in cis* (Yu and Bradley 2001).

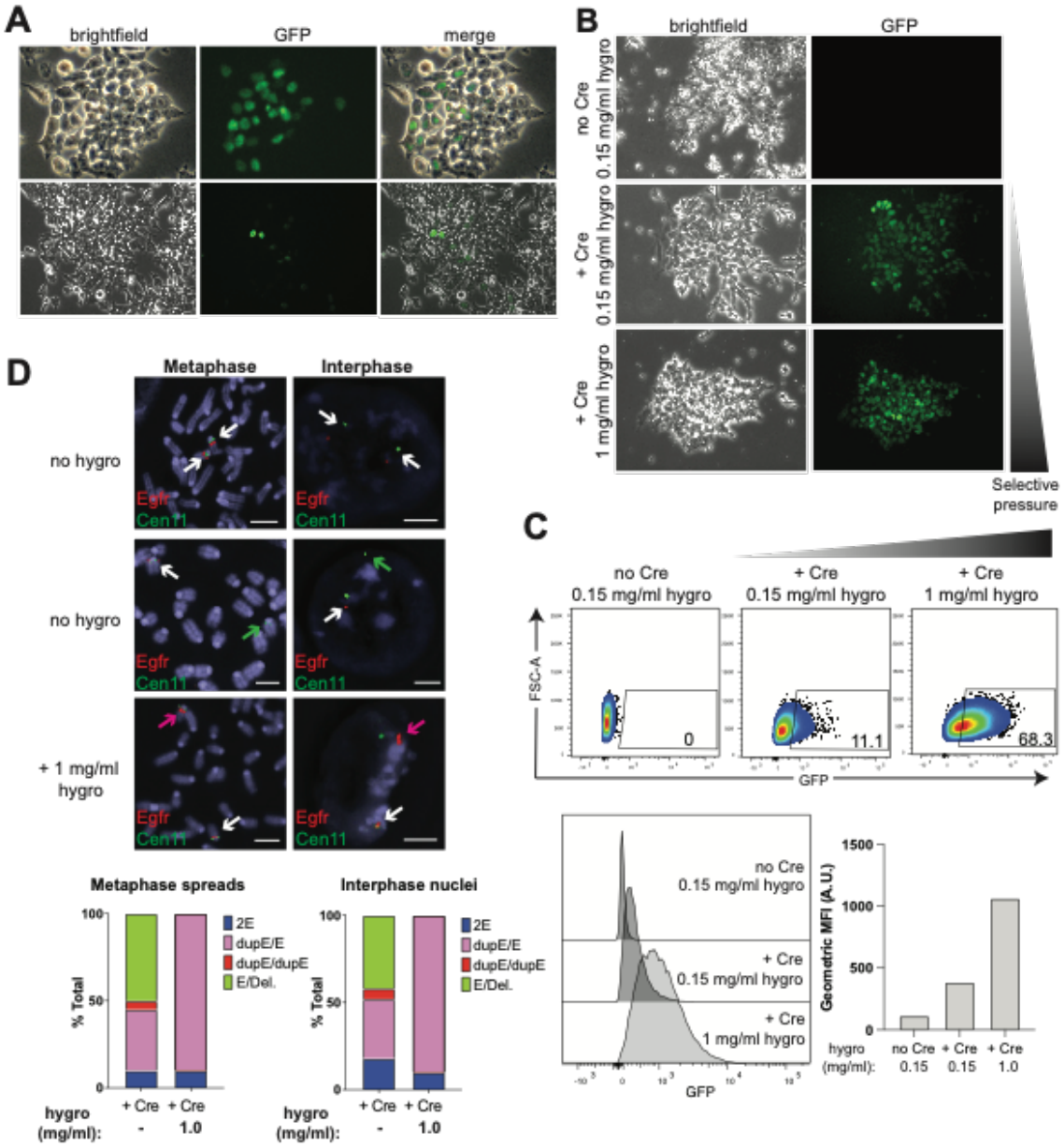


Fig 4.5 mESC clones show mosaic GFP expression after Cre-mediated recombination and are biased against retention of the circle.

(A) Colonies growing out of sorted, GFP-positive single cells show mosaic expression of GFP, with some cells losing GFP expression altogether. **(B)** Fluorescence imaging shows an increase in GFP expression at higher hygromycin concentrations. All surviving cells show GFP expression. Cells not treated with Cre did not survive past 0.15 mg/ml hygro. **(C)** Flow cytometry analysis confirms an increase in the percentage of GFP-expressing cells as well as an increase in GFP intensity when cells were subjected to higher selective pressure. **(D)** Heterogeneous karyotypes were observed in FISH analysis of hygro-treated and untreated cells. (Top) V6.5 cells were confirmed to be diploid (white arrowheads). Untreated (no hygro) cells harbored a deletion of one copy of *Egfr* (green arrowheads), while hygro treatment strongly selected for cells with the tandem duplication (pink arrowheads). (Bottom) Quantification of karyotypes observed by FISH.

4.3 Discussion

In this chapter, I have described the generation of novel mouse strains using complementary approaches that can be used to formally test our hypothesis that ecDNA-mediated oncogene amplification is sufficient to initiate tumorigenesis *in vivo*. The first approach takes advantage of recent advancements in mouse transgenic technology for one-step generation of conditional alleles by electroporation of CRISPR/Cas9 and donor templates directly into zygotes. In the second approach, we relied on more conventional methods of germline targeting, which involves homologous recombination to sequentially target our cassette strategy into mESCs, followed by injection into 8-cell-stage embryos to generate ES-cell derived animals.

Using these parallel approaches, we have targeted the *Myc* and *Egfr* loci, two of the most recurrently amplified oncogenes frequently found on ecDNAs (Kim et al. 2020). By mating the founders generated by the first approach (zygote electroporation) to wild type mice, we have already obtained cohorts of experimental and control mice harboring floxed alleles of both loci, which will be immediately subjected to Cre recombinase using *in situ* delivery methods. Using fibroblasts derived from these mice, we have confirmed by PCR genotyping that Cre treatment induces recombination between the loxP sites, leading to the formation of the circle and the concomitant deletion.

In collaboration with the laboratory of Scott Lowe, we intend to use the *Myc*$^{fl/+}$ mice to test the transformative potential of *Myc*-ecDNA within the context of hepatocellular carcinoma (HCC), where *Myc* is often overexpressed or amplified (Kawate et al. 1999; Zheng, Cubero, and Nevzorova 2017). Somatic mutations in *TP53* are also common in HCC (Hussain et al. 2007). In these experiments, a CRISPR construct (pX330) where Cas9 is co-expressed with a sgRNA targeting *Trp53* will first be targeted to the mouse liver by hydrodynamic tail vein injection (HTVI). After allowing for recovery for a few days, Cre recombinase is then delivered via retro-orbital injection of recombinant adeno-associated viruses where Cre is driven from the hepatocyte-specific promoter of thyroxin binding globulin (*Tbg*) (**Fig 4.6A**).

To test ecDNA-mediated *Myc* amplification in an additional disease setting, we are further collaborating with the Lowe lab to employ their novel murine model of high-grade serous ovarian carcinoma (HGSOC), where genetic elements are directly introduced into the ovary by tissue electroporation (Paffenholz et al., unpublished). Using their model, they have tested various combinations of oncogene overexpression and tumor suppressor loss in the background of *Trp53* loss, which is found in more than 95% of HGSOC tumors (Cancer Genome Atlas Research 2011). In particular, they have shown that transposon-driven *Myc* overexpression cooperates strongly with ablation of p53, and that this effect is further enhanced by concomitant inactivation of *Pten*, another tumor suppressor commonly mutated in HGSOC. Thus, this represents an ideal setting to test the

ability of our *in vivo* engineered ecDNAs (*ecMyc*) to phenocopy this result. Briefly, abdominal surgery will be performed to expose the ovaries of female $Myc^{fl/+}$ mice, followed by co-injection of a targeting plasmid (expressing Cas9 and sgRNAs against *Trp53* and *Pten*) and a Cre-expressing vector (**Fig 4.6B**).

Our $Egfr^{fl/+}$ mice can be used to model glioblastoma, where ecDNA-mediated amplification of *Egfr* occurs at high frequency (An et al. 2018; Kim et al. 2020). Inactivation of p53 is also a common event in glioblastoma (Brennan et al. 2013). Thus, we can model ecDNA-mediated amplification of *Egfr* within the context of p53 ablation by crossing our novel strain to mice with floxed *p53* ($p53^{fl/fl}$), followed by another round of backcrossing to generate $Egfr^{fl/+};p53^{fl/fl}$ mice. Injection of these mice with recombinant adenoviruses expressing Cre will lead to circularization of the *Egfr* locus with concomitant loss of *p53* (**Fig 4.6C**). Importantly, inducing ecDNA formation *in situ* in this manner would recapitulate the presumed steps of ecDNA formation and accumulation in tumors, i.e. the acquisition of a single ecDNA amplicon (resulting from chromothripsis or other genetic events) in a cell, followed by replication and propagation to daughter cells. In parallel to these studies, we can also model *Egfr*-ecDNAs via *ex vivo* chromosomal engineering of primary adult neural stem cells (aNSCs) isolated from *p53-/-* mice, followed by orthotopic implantation into nude mice. This will allow us to rapidly assess the oncogenic potential of our circular amplicon. Indeed, our lab has previously generated mouse glioma models by recreating the *Bcan-Ntrk1* fusion oncogene

via *ex vivo* manipulation of aNSCs (Cook et al. 2017); thus, this pipeline is well-established in our lab.

Lastly, we can take a non-directed approach and induce whole-body formation of *Myc* or *Egfr*-ecDNAs by crossing our floxed mice with strains where CreERT2 is driven from the ubiquitous *Rosa26* promoter (**Fig 4.6D**). CreERT2 is a variant of the recombinase that is fused to a mutated ligand-binding domain of the estrogen receptor and thus requires tamoxifen for activity. CreERT2 activity is shown to be tightly regulated by tamoxifen treatment with very low background. Thus, by delivering tamoxifen via intraperitoneal injection to our *R26-CreERT2*$^{Tg/0}$*;Myc*$^{fl/+}$ or *R26-CreERT2*$^{Tg/0}$*;Egfr*$^{fl/+}$ strains, we will be able to induce circularization in a ubiquitous manner and follow the mice for signs of tumor development.

We also carried out *in vitro* experiments to characterize our double-targeted mESCs and showed that they display similar behavior as our HCT116 cell lines. Specifically, we see a mosaic GFP expression pattern in cells that grew out from an initially GFP-positive population after recombination, including a reversion back to a GFP-negative state, as well as varying intensity of GFP expression. When these cells were subjected to increasing concentrations of hygromycin, they displayed a dose-dependent increase in reporter expression. However, FISH analysis revealed that Cre-mediated recombination resulted in a population of cells with heterogeneous karyotypes, including tandem duplication, which were preferentially selected with hygromycin treatment. Within the population of cells

not exposed to hygromycin, we found some cells harboring a deletion of one copy of *Egfr*, indicating that a looping-out event occurred between loxP sites located *in cis*. However, the fact that we were not able to recover any ecDNA structures in our metaphase spreads, coupled with the preferential selection for the tandem duplication, indicates that there are cell-intrinsic mechanisms that suppress the maintenance of ecDNA amplicons.

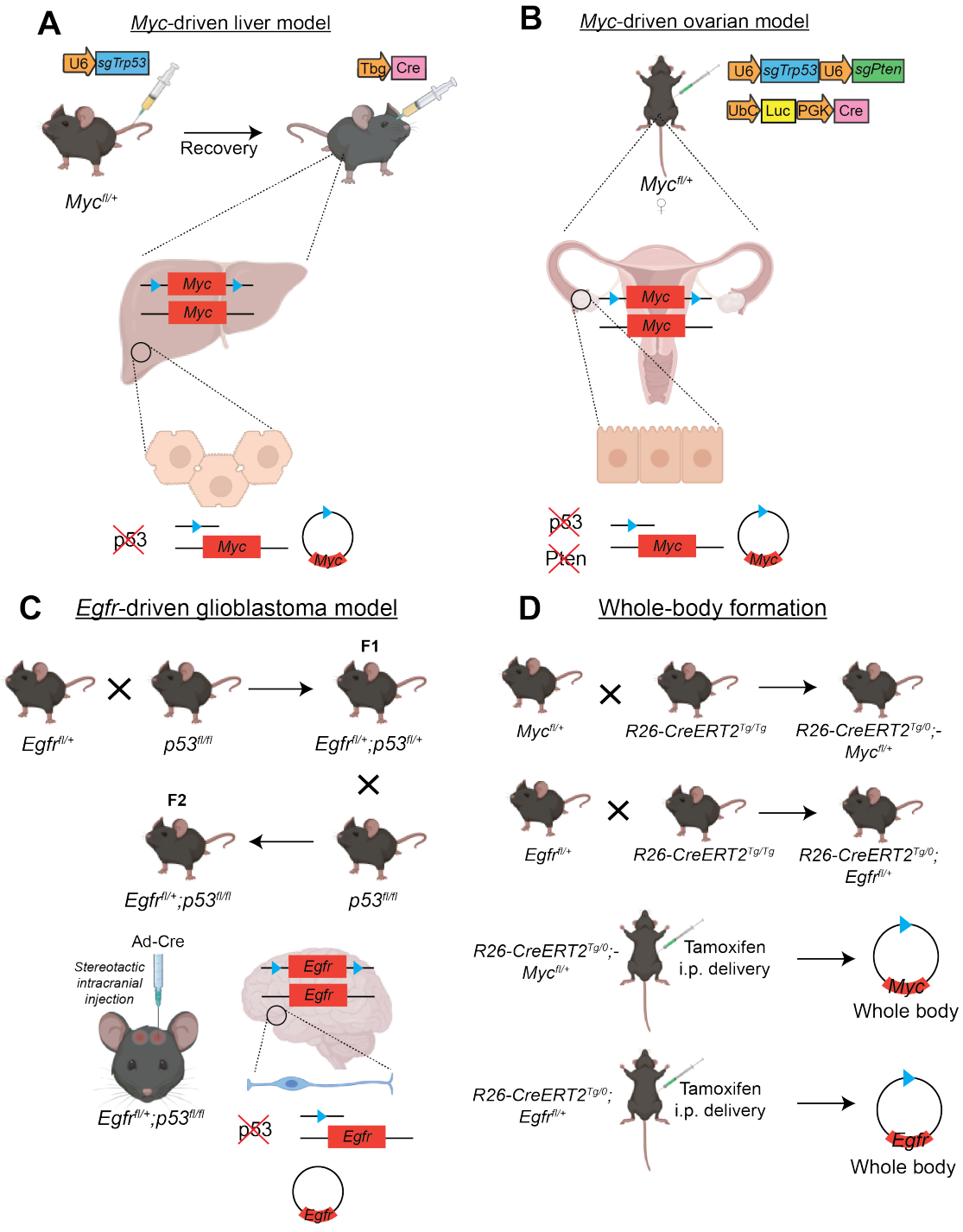


Fig 4.6 Schematic illustrating proposed mouse experiments to test the oncogenic potential of ecDNAs in vivo.

Fig 4.6 (cont'd). (A) *Myc*$^{fl/+}$ mice will first be targeted with *sgTrp53* using HTVI for liver-specific delivery followed by administration of AAVs expressing hepatocyte-specific Cre. This results in *ecMyc* formation with concomitant p53 loss in a subset of liver cells. **(B)** *sgTrp53*, *sgPten*, and a Cre encoding plasmid will be simultaneously delivered to the ovarian tissue of *Myc*$^{fl/+}$ female mice by electroporation, resulting in *ecMyc* formation concurrent with p53 and Pten loss, two TSGs that cooperate in ovarian carcinogenesis. **(C)** Autochthonous model of *Egfr*-driven glioblastoma. Two rounds of crossing introduces the *Egfr*^{fl} allele into a *p53*$^{fl/fl}$ background. Intracranial delivery of Ad-Cre will result in *ecEgfr* formation with concomitant p53 loss. **(D)** Whole-body induction of *ecMyc* or *ecEgfr* can be achieved by crossbreeding to *R26-CreERT2* strains. Tamoxifen delivery is expected to induce Cre activity in all major organ systems.

5 Chapter 5: Discussion

5.1 Conclusions

In this thesis, I have presented work on three projects using CRISPR/Cas9 as a linchpin to interrogate a diverse array of biological questions. In Chapters 3 and 4, I delineated the generation of a novel system that pairs conventional Cre-loxP targeting with CRISPR/Cas9 gene editing to engineer extrachromosomal circular DNAs in cells and in animals. We demonstrated the feasibility of this system in both human and mouse cell lines and have employed it to model ecDNAs containing two oncogenes recurrently amplified in human cancers: *EGFR* and *MYC*. In a series of *in vitro* experiments, I have demonstrated the successful targeting of our system in the HCT116 cancer cell line and in mouse adult neural stem cells and embryonic stem cells. As part of our system, a reporter gene is expressed and localized within the engineered ecDNA upon successful circularization, allowing us to follow the dynamics of ecDNAs within recombined cells. The behavior of our reporter after multiple cell divisions and in response to selective pressure is compatible with both the predicted and empirically observed behavior of ecDNAs in cancer cell lines. Specifically, we have found that our GFP reporter can be lost from daughter cells as a population progresses through multiple rounds of mitoses, in line with the notion that ecDNAs are segregated randomly at each cell division. In addition, the intensity of reporter expression

within GFP-positive cells is also variable, again in agreement with heterogeneity of ecDNA copy number as an outcome of random segregation. Genotyping PCR and FISH analysis confirm that Cre-mediated recombination induces the deletion of one copy of *EGFR* in our targeted cells, implicating that the excised region is looped out. Intriguingly, we did not observe *EGFR*-staining ecDNA structures within metaphase spreads obtained from our cells, although we did observe the presence of micronuclei containing *EGFR*-positive ecDNA-like foci in our HCT116 cells. Altogether, our data suggests that HCT116 and mESCs may have cell-intrinsic mechanisms that prevent the maintenance and propagation of ecDNAs. Interestingly, our preliminary results in mouse *p53-/-* aNSCs suggest that p53 loss may be a genetic requirement for cells to tolerate ecDNA formation and maintenance.

Additionally, using a two-pronged approach, we have generated novel mouse models containing floxed *Myc* and *Egfr* loci that will allow us to study the *in vivo* significance of ecDNA formation in tumorigenesis. Using the first approach, we have already obtained cohorts of control and experimental mice that can be immediately subjected to Cre recombinase to investigate the oncogenic potential of engineered ecDNAs in various cancer settings, which we have outlined. Using cells isolated from these animals, we have confirmed the generation of the expected recombination products after exposure to Cre recombinase. With the second complementary approach, we generated double-targeted mESC clones that will be used to generate mouse strains containing our split-reporter cassettes,

confirmed them by PCR genotyping and Southern blot analyses, and further characterized one of the clones with a panel of *in vitro* experiments, including FISH analysis, which confirmed targeting *in cis*.

In Appendix I, I described our efforts to adapt genetically engineered bacteria as a novel delivery vector for CRISPR in applications of *in vivo* somatic engineering. I demonstrated that intracellular, noninvasive *E. coli* can be used to deliver the CRISPR/Cas9 platform into mammalian cells and effect gene editing *in vitro*. With further experimentation and optimization, our system may be able to be adapted for successful applications *in vivo*. Finally, in Appendix II, I interrogated the transcriptional impact of deleting a conserved CTCF-binding site in the context of chronic lymphocytic leukemia by generating a series of isogenic cell lines harboring various deletions within a minimally deleted locus in patients. Comprehensive analysis of the global gene expression profiles of independent lines revealed that while no specific expression changes could be linked to deletion of the CTCF-binding site, deletion of the long noncoding RNA, *DLEU2*, generates a specific transcriptional signature marked by significant overexpression of the interleukin-7 receptor *IL7R*. Interestingly, complementation studies using lentiviral vectors to re-express *DLEU2* transcripts did not affect *IL7R* expression, suggesting that there may be an as yet unknown genetic element within the minimally deleted region that may be functionally linked to IL7R.

5.2 Future directions

5.2.1 *In vitro* and *in vivo* modeling of ecDNA

While the role of ecDNAs in cancer development has attracted much attention and research in recent years, these studies are limited by their usage of cancer cell lines already harboring ecDNA, some of which have been maintained for many years. While these studies are informative in illuminating certain aspects of ecDNA biology, they are unable to address key biological questions. For example, it is not known if ecDNAs harboring oncogenes are sufficient to initiate tumor formation or if their formation is a late event during tumorigenesis. While the possession of multiple copies of ecDNAs carrying oncogenes likely confers a fitness advantage to a cell growing within the context of a tumor by allowing it to easily modulate its genomic composition in response to environmental changes, such as drug treatment, the initial stages of cancer-associated ecDNA formation and propagation remains largely a black box. It is unknown whether a single copy of an ecDNA residing in a single cell, can continue to propagate through subsequent divisions to result in a population of progeny cells with heterogeneous ecDNA copy numbers. Because the bulk of studies carried out on ecDNA biology has been done on cancer cells already harboring these structures, it is also unknown how normal cells will respond to the presence of a large acentric amplicon. In order for a cell

to maintain an ecDNA, a suitable genetic background, e.g. tumor suppressor loss, may be a requirement to bypass cell cycle checkpoints.

Our experiments in mouse and human cell lines suggest that cells may be inherently biased against maintaining ecDNA structures. There are several possibilities to explain our results. Firstly, we may be missing genetic element(s) required for the stabilization and/or replication of ecDNA within our circle. Within the context of our human cell line experiments, this is unlikely, because the ecDNA we are modeling in HCT116 captures the majority of the ecDNA sequence found in the patient-derived cell line GBM39. A second, and in our view more likely, possibility is that a permissive genetic background is required for ecDNAs to be maintained and propagated. HCT116 cells may not represent a genetically permissive cell type, and additional genetic perturbations may be required to allow ecDNAs to be retained within this cell line. Indeed, our preliminary experiment in *p53-/-* murine aNSCs suggests that this may be the case, but further studies are needed to definitively establish a relationship between p53 loss and ecDNA tolerance. Below, I discuss some factors that may contribute to ecDNA intolerance and suggest directions for future experiments based on our findings.

5.2.1.1 Elimination of ecDNAs via sequestration into micronuclei

One possibility is that the ecDNAs, once formed, are eliminated from cells via sequestration into micronuclei. Our results from FISH analysis of interphase HCT116 cells suggest that this is the case. Evidence from the literature also

supports this as a known mechanism by which ecDNA can be excluded from tumor cells. Micronuclei are small, membrane-bound structures containing DNA content (often lagging chromosomes as a result of merotelic spindle attachments or damaged, acentric chromosome fragments that fail to be incorporated into the primary nucleus after mitosis) and are spatially distinct from the primary nucleus. Because micronuclei formation are frequently generated as a result of defective mitosis, which is often accompanied by genome damage, micronuclei are often used as a marker of genotoxic stress. Additionally, the nuclear envelope that encapsulates micronuclei are often defective, leading to micronuclear rupture that releases DNA content into the cytoplasm, which can then induce a cyclic GMP-AMP synthase (cGAS)-mediated innate immune response through the activation of the cGAS-STING pathway (Krupina, Goginashvili, and Cleveland 2021). Micronuclei are often found in cancer cells, and recent evidence has implicated them as a source of complex chromosomal rearrangements, including chromothripsis, due to their propensity to undergo extensive DNA damage (Zhang et al. 2015; Ly et al. 2017). Thus, micronuclei are considered a hallmark of chromosomal instability.

Interestingly, DNA replication can continue to occur within micronuclei. However, DNA replication within micronuclei exhibits delayed kinetics compared with that in the major nucleus, with some micronuclei still undergoing replication when the major nucleus is in G2 phase (Crasta et al. 2012). This entry of still-replicating micronuclear content into the subsequent mitosis produces extensive

DSBs in the micronuclear DNA, which, upon repair by NHEJ, can lead to the complex genomic rearrangements seen in chromothripsis. Additionally, ecDNA-harboring micronuclei were also found to be transcriptionally active, as transcripts could be detected within micronuclei as assessed by RNA FISH and bromouridine incorporation (Utani, Kawamoto, and Shimizu 2007).

Intriguingly, ecDNAs have been found to be excluded into micronuclei in cancer cells, eventually leading to loss of amplified oncogenes and a consequent decrease in oncogenicity. For example, the human promyelocytic leukemia cell line, HL-60, which carries *MYC* amplification on ecDNAs, has been observed to undergo spontaneous differentiation which was postulated to be through an active elimination process via micronuclei (Shimizu et al. 1994). Von Hoff et al. demonstrated a similar finding in COLO320 DM cells, showing that treatment with low concentrations of hydroxyurea (HU) resulted in increased loss of *MYC*-containing ecDNAs, and that these ecDNAs were being preferentially entrapped within micronuclei (Von Hoff et al. 1992). A similar effect was found upon exposure of tumor cells to radiation therapy, which resulted in the loss of extrachromosomally amplified drug resistance genes (Schoenlein et al. 2003). Later, other groups reported similar findings with other types of tumor cells, indicating this phenomenon is not oncogene-dependent, including neuroblastoma cell lines harboring extrachromosomally amplified *MYCN*, in which micronuclei packed with *MYCN*-positive ecDNAs were observed (Ambros et al. 1997; Valent et al. 2001). This was accompanied by downregulation of *MYCN* expression and

decrease of proliferative ability (Ambros et al. 1997). A similar finding was also later reported in neuroblastoma tumors isolated from patients (Valent et al. 2001). Surprisingly, siRNA-mediated knockdown of extrachromosomally amplified genes were shown to be correlated with an increase in ecDNA-containing micronuclei formation in colorectal and neuroectodermal cell lines with a concomitant decrease in proliferation rates and invasiveness in *in vitro* studies (Ji et al. 2014).

A possible mechanism for how ecDNAs may be preferentially sequestered into micronuclei is proposed to be due to the unique intracellular behavior of ecDNAs during mitosis. While the acentric nature of ecDNAs mean that they are segregated unevenly to daughter cells, several groups have reported that ecDNA can “hitchhike” onto chromosomal DNA through a close association with chromosomes during cell division via an unknown mechanism (Shimizu, Misaka, and Utani 2007; Kanda, Otter, and Wahl 2001; Lange et al. 2021). At the same time, aggregated ecDNAs have also been observed to lag behind the main chromatin mass during anaphase. By inserting the lac operator sequence into artificial ecDNAs and directly visualizing them through co-expression of a lac repressor-GFP fusion protein in the same cell, Shimizu et al. showed that these ecDNAs can form clusters that lagged behind in anaphase and were shuttled into micronuclei at the next interphase (Shimizu, Misaka, and Utani 2007). This observation was increased upon induction of DNA damage by treatment with a low concentration of HU, which resulted in the persistence of gH2AX foci specifically on ecDNAs even after gH2AX disappeared from chromosomal DNA, suggesting

that DNA repair mechanisms act differently on ecDNAs compared to chromosomes. This persistence of gH2AX on ecDNAs appeared to promote their aggregation, lagging during mitosis, and consequent exclusion into micronuclei (Shimizu, Misaka, and Utani 2007).

Notably, supercoiled plasmid DNA microinjected into cell nuclei rapidly form aggregates that are then excluded into micronuclei after mitosis (Shimizu, Kamezaki, and Shigematsu 2005), although the mechanism for this is unknown. Altogether, the exclusion of ecDNAs into micronuclei may be an intrinsic cellular mechanism resulting from the unique properties of acentric ecDNAs, including their propensity to form nuclear aggregates during cell division. Further studies are needed to elucidate the mechanism of this aggregative behavior.

5.2.1.2 Activation of the cGAS-STING pathway by micronuclei

What is the fate of ecDNA-containing micronuclei, or cells that harbor these micro-organelle structures? Once within micronuclei, ecDNA content may be exposed to the cytoplasm through micronuclear envelope breakdown. Due to defective nuclear lamina organization, micronuclei are more prone to rupture (Hatch et al. 2013). The release of DNA content into the cytoplasm can then trigger an innate immune response via the cGAS-STING pathway.

cGAS (cyclic GMP-AMP synthase) is a pattern recognition receptor that recognizes and binds to dsDNA in the cytoplasm and within micronuclei. Originally thought to primarily serve as an immune mechanism against microbial invasion

such as viral infection, during which viral genomic content is released into the cell, this canonical function of cGAS has now been expanded to include roles in tumorigenesis, senescence, and autophagy.

Briefly, cGAS is activated through interaction with double-stranded DNA in a sequence-independent manner. Activated cGAS then catalyzes ATP and GTP into 2',3'-cyclic GMP-AMP (cGAMP), the secondary messenger in the pathway. Subsequently, cGAMP activates the downstream effector, stimulator of interferon genes (STING), at the endoplasmic reticulum. Activated STING leads to the downstream activation of TANK binding kinase 1 (TBK1), which transphosphorylates STING. The functional output of the pathway is the eventual activation of interferon regulatory 3 (IRF3), a transcriptional factor that leads to the expression of immune-stimulated genes (ISGs) and type 1 interferons (IFNs). IKK is also activated in a parallel branch of the pathway, driving the induction of NFkB inflammatory genes (Kwon and Bakhoum 2020).

Recent work has shown that self-DNA contained within micronuclei triggers activation of the cGAS-STING pathway. Indeed, Mackenzie et al. showed that cGAS is recruited directly to micronuclei upon nuclear envelope rupture, independent of the type of cargo carried by the micronuclei. Single-cell RNA-seq confirmed the upregulation of downstream ISGs, indicating activation of the pathway, specifically in micronucleated cells (Mackenzie et al. 2017). In response to cytosolic DNA, the activation of the cGAS-STING pathway can lead to senescence, characterized by permanent cell cycle arrest, and trigger the

production of inflammatory factors including cytokines, chemokines, proteases and growth factors, collectively known as the senescence-associated secretory phenotype (SASP) (Gluck et al. 2017). Deletion of cGAS abrogated this SASP response and accelerated the spontaneous immortalization of mouse embryonic fibroblasts (Yang et al. 2017).

Recognition of self-DNA by the cGAS-STING pathway also activates an autophagy pathway that can lead to cell death. Nassour and colleagues demonstrated that cGAS-STING plays an essential role in inducing autophagy-mediated cell death in response to replicative crisis brought on by telomere deprotection. When cGAS or STING were attenuated, this licensed RB- and p53-deficient cells to bypass crisis, despite the increased accumulation of micronuclei and cytoplasmic chromatin fragments, and continue proliferating. Interestingly, the genomes of surviving cells that bypassed crisis were found to harbor chromosomal aberrations (Nassour et al. 2019).

The fact that cGAS-STING can recognize self-DNA raises the question of how this pathway responds to chromosomal DNA during mitosis, when the nuclear envelope is disassembled. Work by Zierhut et al. showed that cGAS binds preferentially to nucleosomal DNA than to naked DNA by direct interaction with nucleosomes, but that this nucleosome binding suppresses cGAS activation, possibly by engaging it in a configuration that precludes optimal catalytic activitiy. Indeed, a direct comparison of the catalytic activity of nucleosome- versus DNA-bound cGAS showed that the former had a reduced catalytic rate (Zierhut et al.

2019). Thus, this protective mechanism sufficiently dampens cGAS signaling so that normal mitosis does not trigger an aberrant immune response, but the extended presence of self-DNA in the micronucleus or the cytoplasm in interphase, when cells are already in a transcriptionally-primed state, is presumed to lead to the accumulation of enough phosphorylated IRF that can immediately feed into downstream gene activation. These results also explain why cGAS activation by self-DNA is less potent than by transfected DNA (Harding et al. 2017; Spektor, Umbreit, and Pellman 2017).

A report suggests that micronuclei themselves may be degraded by autophagy, as evidenced by colocalization of autophagosomal and autolysosomal markers with micronuclei, though the mechanism for how the autophagic machinery can selectively recognize micronuclei and whether this occurs through a cGAS-STING dependent pathway was not addressed in this paper (Rello-Varona et al. 2012). More recently, separate research showed that genetic ablation of cGAS increased the frequency of micronuclei, corroborating earlier results from Nassour et al., while overexpression of cGAS had the converse effect (Zhao et al. 2021).

Taken together, the activation of the cGAS-STING pathway, mediated by the release of self-DNA from micronuclei, may represent a mechanism by which cells can detect and consequently suppress the early accumulation of ecDNAs. An experimentally testable prediction of this model is that the inactivation of this

pathway by genetic inactivation of cGAS or STING in non-permissive cells should promote the maintenance and propagation of ecDNAs.

5.2.1.3 Permissive genetic backgrounds for the maintenance and/or propagation of ecDNAs: inactivation of p53

Early studies showed that normal cells of both human and rodent origin fail to develop resistance to drugs like methotrexate, hydroxyurea, or N-phosphonacetyl-L-aspartate (PALA) by amplifying the gene coding for the target enzyme, while many immortalized and tumor cell lines were readily able to (Tlsty 1990). Instead, loss of p53 was required for gene amplification to occur and for cells to develop resistance (Livingstone et al. 1992; Yin et al. 1992). As a critical tumor suppressor that is mutated in over 50% of all cancers, p53 plays a major role in safeguarding against tumorigenesis by inducing apoptosis and cell cycle arrest, such as senescence, in response to cellular insults and DNA damage, (Michalak et al. 2005). The successful amplification of oncogenes on ecDNAs may thus also require a nonfunctional p53 pathway to provide a permissive genetic background to allow cells to bypass this cell cycle control checkpoint.

Our preliminary experiments in aNSCs are suggestive of this, though further studies will be required to establish a direct relationship between p53 loss and ecDNA maintenance. One obvious experiment would be to target our system into aNSCs expressing wild type p53 (rederived from a wild type mouse) and investigate if ecDNAs are able to be maintained in the presence of a functional p53

pathway. Because our *p53-/-* aNSCs have been maintained *in vitro* for some length of time, it is likely that the line has accumulated additional genetic events that may be linked to their ability to tolerate ecDNAs. Therefore, it may also be necessary to ablate p53 within the context of a system with wild type p53 already targeted with our strategy, such as in our HCT116 cell lines; the generation of such isogenic lines will help to discern if p53 loss is the sole regulator of ecDNA tolerance. Recent reports have also suggested a putative link between mutant, gain-of-function p53 (mtp53) and cGAS-STING; mtp53 was shown to bind to TBK1 and prevent downstream activation of IRF3, thus promoting tolerance for cytoplasmic DNA (Ghosh et al. 2021). It will thus be interesting to examine this relationship in the context of our systems and to investigate if concurrent p53 and cGAS-STING inactivation can cooperate to promote ecDNA tolerance.

5.2.1.4 Cellular fitness cost for maintaining ecDNAs

Interestingly, ecDNA elements are lost over time when patient-derived tumor cells are brought into culture, suggesting there is a fitness cost for cells to maintain ecDNA in an *in vitro* context (Nikolaev et al. 2014; Pandita et al. 2004; Yost et al. 2013; deCarvalho et al. 2018). In contrast, analysis of glioma samples at diagnosis and relapse show that oncogene-containing ecDNAs can be maintained longitudinally, suggesting that their maintenance adequately benefits a growing tumor *in vivo* to overcome any potential fitness cost (deCarvalho et al 2018). Corroborating evidence from COLO320 DM cells propagated in culture versus as

a tumor *in vivo* demonstrate similar evidence; the number of ecDNAs in cells propagated *in vivo* was found to be 3- to 4-fold higher compared to their counterparts in cell culture (Von Hoff et al. 1992). Notably, double minutes were initially present in almost all metaphases of the COLO320 DM cell line (Quinn et al. 1979), but is now only present in ~20% of cultured cells, according to the American Type Culture Collection (ATCC) and in concordance with our own observations. In a similar vein, the prostate cancer PC-3 cell line, which harbors a low copy number of ecDNAs, showed increased amplification of *Myc*-ecDNA after *in vivo* propagation and selection (Fukumoto, Shevrin, and Roninson 1988).

5.2.1.5 Mathematical modeling predictions of ecDNA behavior

Mathematical simulations of predicted ecDNA behavior can help to deduce some rules of ecDNA inheritance and shed light on their functional consequences. While chromosomal segregation during mitosis ensures that each daughter cell receives the same genomic content, ecDNAs are predicted to segregate randomly. This random partitioning should lead to a binomial distribution in the number of ecDNA per cell within the total population (Shimizu et al. 1994; Lange et al. 2021). Indeed, quantification of ecDNAs by FISH analysis in cancer cell lines of different histological types reveals a Gaussian distribution of ecDNA copy number per cell, confirming this prediction (Lange et al. 2021). Using CRISPR, some groups have inserted TetO arrays into ecDNAs to allow visualization of their dynamics during cell division upon the expression of TetR-fluorophore fusion proteins in the same

cell. These live cell videos directly confirm the random inheritance pattern of ecDNA (Yi et al. 2020; Lange et al. 2021).

Mathematical simulations can also help us make *a priori* predictions about the distribution of ecDNA copy number within a population of cells (i.e. variance), which has important implications for intratumor heterogeneity and therapy resistance. These simulations begin with a single cell possessing a single ecDNA amplicon and proceed through its expansion into a population of cells, terminating at a specified population size. The models stipulate that under neutral selection (that is, where a cell is neither advantaged nor disadvantaged for having ecDNA), cells with low ecDNA copy number frequently give rise to a daughter cell without ecDNA. Because a cell that has lost ecDNA cannot regain them, ecDNA prevalence within this population will quickly decay so that only a small minority of cells will possess ecDNAs. This prediction may help to explain the behavior of our engineered ecDNAs in the mouse V6.5 and human HCT116 systems. In particular, the frequent loss of our reporter signal may be attributable to the fact that *EGFR* amplification may not confer a fitness advantage to HCT116 cells or mESCs. In contrast, when cells harboring ecDNAs are positively selected (e.g. when the ecDNA harbors an oncogene that does confer a proliferative or growth advantage to the cell), it is predicted that ecDNA prevalence will remain frequent in the population, with the majority of cells possessing them. Indeed, comparison of simulated outcomes to empirical measurements of ecDNA prevalence in cell lines and patient samples indicates that the ecDNAs found in these biological samples

are under strong positive selection (Lange et al. 2021). This may explain our results in our targeted aNSCs, where the majority of recombined cells retained the reporter signal. Because these cells are cultured in EGF-supplemented medium, cells harboring amplified *Egfr* via ecDNAs may be conferred a selective advantage.

Finally, mathematical simulations predict that oncogenes can reach much higher copy numbers by being amplified on ecDNA versus intrachromosomally. This is corroborated by empirical data showing that ecDNA-amplified copies of *MYC* and *EGFR* achieve higher copy numbers than chromosomal amplification of the same oncogenes in cancer cell lines and patient tumor samples. Moreover, intratumor heterogeneity is predicted to be enhanced and maintained for longer in tumors with ecDNAs compared to those with chromosomal amplifications, as the greater amplicon copy number that can be achieved due to random segregation is predicted to directly contribute to greater genetic diversity between cells (Turner et al. 2017).

5.2.2 *In vivo* somatic engineering by genetically engineered *E. coli*

In Appendix I, we attempted to induce somatic gene editing in an *in vivo* context by using genetically engineered *E. coli* as a delivery vector for CRISPR. We demonstrated that *E. coli* strain BM4570, transformed with CRISPR plasmids, could infect colon epithelial cells and induce gene editing *in vitro*. Because the mammalian large intestine represents the major site of colonization for most commensal strains of *E. coli* (Conway, Krogfelt, and Cohen 2004*)*, we reasoned

that our *E. coli*, once administered to mice, could localize to the large intestine and achieve targeting of the CRC-associated gene, *Apc*, in intestinal epithelial cells. However, our results indicated that our *E. coli* was not able to invade the intestinal epithelial lining. Studies on humans and mice have shown that the endogenous intestinal microbiota resists colonization by exogenous strains (Anderson, Gillespie, and Richmond 1973; Freter et al. 1983). Therefore, most mouse studies on *E. coli* colonization, including of the common lab strain K12, have been performed using streptomycin-resistant bacterial strains colonizing streptomycin-treated mice. Streptomycin treatment enables effective colonization of *E. coli* by eliminating a significant portion of the mouse intestinal microbiome (Hentges et al. 1984). One possible reason for why we were not able to induce gene editing in the large intestine of our mice could be due to ineffective colonization of our genetically engineered strain of *E. coli*. Thus, in order to achieve promote colonization, pre-treatment of our mice via the administration of streptomycin to their drinking water may be required, as well as further genetic modification to our *E. coli* to confer streptomycin resistance.

Another consideration is the presence of the mucus layer that lines the intestinal tract and is turned over every hour in mice (Johansson and Hansson 2016). In order to bypass the digestive tract, enema administration can be used to directly introduce our *E. coli* to the colon. This method has been used in a recent study investigating the effect of fecal microbiota transplantation on dextran sulfate sodium (DSS)-induced colitis (Zhou et al. 2019).

5.2.3 Tiling array to elucidate genetic element with functional link to *IL7R* in CLL

Although our results from Appendix II suggest that *DLEU2* expression is not functionally linked to *IL7R*, there may be an unidentified element within the region whose deletion drives the upregulation of *IL7R*. Future experiments could investigate this by creating a screening library of guide RNAs targeting roughly 1-kb increments along the locus targeted in our experiments. Transduction of this library of gRNAs in cassette-targeted, GFP-expressing MEC1 cells would generate a polyclonal population of cells harboring deletions of varying sizes, which would cause loss of GFP expression. The readout of this screen would be increased IL7R expression within GFP-negative cells. Next-generation sequencing of $IL7R^{hi}$ and $IL7R^{lo}$ populations would identify the gRNAs that are enriched in $IL7R^{hi}$ cells, and thus help to elucidate the minimally deleted region underlying the $IL7R^{hi}$ phenotype. This assay could therefore help to narrow the search for an unknown element within this region that may be functionally linked to *IL7R*.

APPENDIX I: Exploring the use of engineered bacteria as an alternative delivery vector for mammalian somatic gene editing

Introduction

Though next generation sequencing efforts continue to nominate an increasing number of disease-relevant genes, experimental validation is required to distinguish driver mutations from non-relevant passenger mutations. Genetically engineered mouse models (GEMMs) constitute the gold standard for validating a genetic lesion's oncogenicity by allowing its transforming potential to be studied *in vivo*. These efforts have been aided by the development of CRISPR/Cas9 technology, which allows genetic alterations to be engineered somatically in mice, thus circumventing time-consuming germline manipulation. For example, Maddalo et al. modeled the EML4-ALK fusion protein, found in a subset of human non-small cell lung cancers (NSCLC) and caused by an inversion on the short arm of chromosome 2, by intratracheal delivery of recombinant adenoviral vectors simultaneously expressing Cas9 and two sgRNAs targeting the inversion breakpoints (Maddalo et al. 2014). The resulting mice developed tumors harboring the *Eml4-Alk* inversion, expressed the fusion protein, and displayed histopathological and molecular features similar to human ALK-driven NSCLCs. Cook et al. used a similar strategy to model another gene fusion, *BCAN-NTRK1*, which is found in high-grade gliomas and results from the deletion of the

intervening region between *Brevican* (*Bcan*) and *Ntrk1 (Cook et al. 2017)*. They engineered the *Bcan-Ntrk1* rearrangement directly in the brain of adult $p53^{fl/fl}$ mice by intracranial injection of an all-in-one adenoviral vector expressing Cas9 and the targeting sgRNAs, and concomitantly induced the loss of *p53* by co-injection with recombinant adenoviruses expressing the Cre recombinase. Most of the injected mice developed high-grade gliomas that were invariably *Bcan-Ntrk1*-positive. Both examples demonstrate that genetic lesions can be successfully engineered somatically in mice, allowing for rapid interrogation of their oncogenic potential.

However, an important limitation of somatic gene editing is that because it relies on viral delivery of CRISPR/Cas9 components, its applicability is restricted to tissues that can be easily accessed by current methods of viral delivery, thus limiting the types of cancers that can be modeled with this strategy. To target certain tissues in the mouse with CRISPR/Cas9 that are not accessible by viral delivery methods, complex surgical procedures may be required to localize the delivery of reagents to the desired tissue type. For example, recent efforts by Roper et al. to generate mouse models of colorectal cancer (CRC) using *in situ* gene editing required colonoscopy-guided mucosal injection to deliver sgRNAs targeting *Apc* and/or *Trp53* to the colonic epithelial cells of $Rosa26^{LSL\text{-}Cas9\text{-}eGFP/+}$;$Villin^{CreER}$ mice with concomitant tamoxifen administration (Roper et al. 2017). Though they were able to induce tumorigenesis in the relevant cell of origin, the tumors did not progress beyond high-grade dysplasia.

We were inspired by earlier studies investigating bacteria as a gene therapy delivery vector to ask if bacteria could be harnessed as an alternative vector to viruses to deliver the CRISPR/Cas9 system to the intestinal tract. This idea was also informed by the fact that CRISPR was, in fact, initially discovered as a bacterial adaptive immune strategy, but that furthermore, humans have recently been found to have an anti-Cas immune response due to exposure to CRISPR/Cas systems via the gut microbiome (Charlesworth et al. 2019). We therefore envisioned that bacteria harboring CRISPR components could be delivered to mice through oral gavage, circumventing the need for surgical procedures. Compared to viral vectors, bacteria are significantly cheaper to produce and can be easily scaled up. *E. coli* carrying a therapeutic *TGF-β1* genetic payload was used to treat mice with DNBS-induced colitis, resulting in reduction of disease severity in treated mice (Castagliuolo et al. 2005). Strikingly, Xiang et al. showed that *E. coli* expressing shRNA against *Ctnnb1* mediated downregulation of β-catenin specifically in the intestinal epithelium, a phenomenon termed “trans-kingdom RNAi” (Xiang, Fruehauf, and Li 2006).

Motivated by these collective findings, we attempted to use a genetically engineered strain of *E. coli* (*BM4570*) to deliver CRISPR/Cas9 to cells *in vitro* and *in vivo* to induce gene editing. We showed that *BM4570* carrying sgRNA-encoding constructs was able to mediate gene editing in mammalian cells *in vitro* at different loci and that this was dependent on the invasive property of the strain. In addition, we investigated different approaches to deliver the sgRNA, and found that guides

expressed by the host cell machinery led to more efficient gene editing than guides expressed from a bacterial promoter. Finally, we tested the feasibility of the system *in vivo* by administrating *BM4570* carrying targeting constructs against *Apc* to wild type C57BL/6J mice.

Results

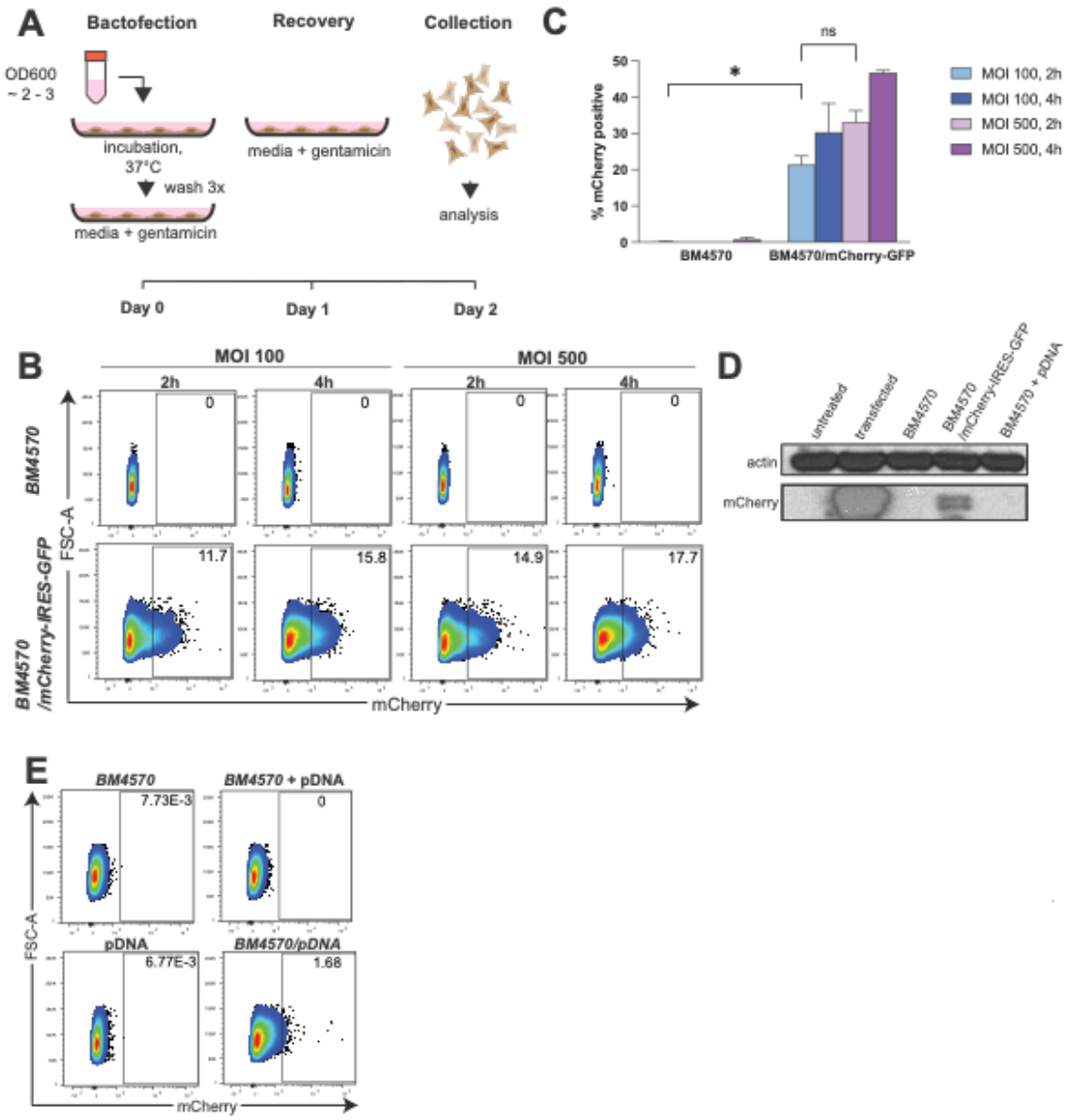
Bactofection of *BM4570* carrying reporter constructs leads to reporter expression in colonic epithelial cells *in vitro*.

BM4570 is a genetically engineered *E. coli* strain created by Grillot-Courvalin et al. (Grillot-Courvalin et al. 1998). Derived from the common lab strain K12, it was conferred unique properties normally found in pathogenic bacteria through the insertion of the *inv* and *hly* genes (from *Yersinia pseudotuberculosis* and *Listeria monocytogenes* respectively) into its genome. *Inv* encodes invasin, which allows the bacteria to invade nonphagocytic cells expressing β1-integrin. Listeriolysin O (encoded by *hly*) permits bacterial escape from the entry vesicle (Grillot-Courvalin et al. 1998). The insertion of the *inv* and *hly* loci disrupts the expression of *dapA/dapB*, two essential bacterial genes. *dapA* and *dapB* catalyze the first two steps of diaminopimelate (or dap, as it is henceforth referred) synthesis, an amino acid required for bacterial cell wall synthesis. Altogether, *BM4570* is able to invade and lyse within cells, and also cannot survive in any environment where dap (which is not synthesized in the mammalian milieu) is absent. This combination of factors ensures that although *BM4570* is invasive, it is nonpathogenic.

We first set out to test the optimal conditions for bactofection *in vitro*. We chose CMT93, a mouse colorectal carcinoma cell line, as it is closely related to our target tissue of interest. *BM4570* was transformed with a bacterial mCherry-IRES-GFP reporter plasmid, then co-incubated with CMT93 monolayers at varying

multiplicities of infection (MOIs) for 2 or 4 hours. During the co-incubation period, the media was supplemented with dap. The bactofected CMT93 monolayers were then washed multiple times and further cultured in media containing gentamicin but dap-removed for an additional day before being collected for analysis (**App lFig 1A**). The addition of gentamicin serves to remove extracellular bacteria. We observed that CMT93 cells co-incubated with *BM4570* harboring the mCherry-IRES-GFP reporter (*BM4570/mCherry-IRES-GFP*) were mCherry-positive in a MOI- and time-dependent manner (**App I Fig 1B**, **C**). Furthermore, mCherry protein was found only in cells that were infected with *BM4570/mCherry-IRES-GFP* or that were transfected with the construct (**App I Fig 1D**). Importantly, we ascertained that the mCherry signal was directly transmitted to cells from *BM4570*, because cells that were co-incubated with a mixture of *BM4570* and naked mCherry-IRES-GFP plasmid DNA showed no positive signal (**App I Fig 1E**).

While it was clear that *BM4570* were able to transmit the reporter signal to CMT93, we wanted to demonstrate that this was done through the transfer of genetic material from bacteria that had invaded and lysed within the cell. We transformed *BM4570* with a mammalian expression construct encoding H2B-mCherry (*BM4570/H2B-mCherry*). We confirmed that 24 hours following co-incubation with bacteria harboring the H2B-mCherry construct, CMT93 were positive for the nuclear reporter (**App I Fig 2A, B**). The signal was not due to leaky expression from intracellular bacteria, as CMT93 showed no fluorescent expression immediately after the co-incubation period (**App I Fig 2A**).



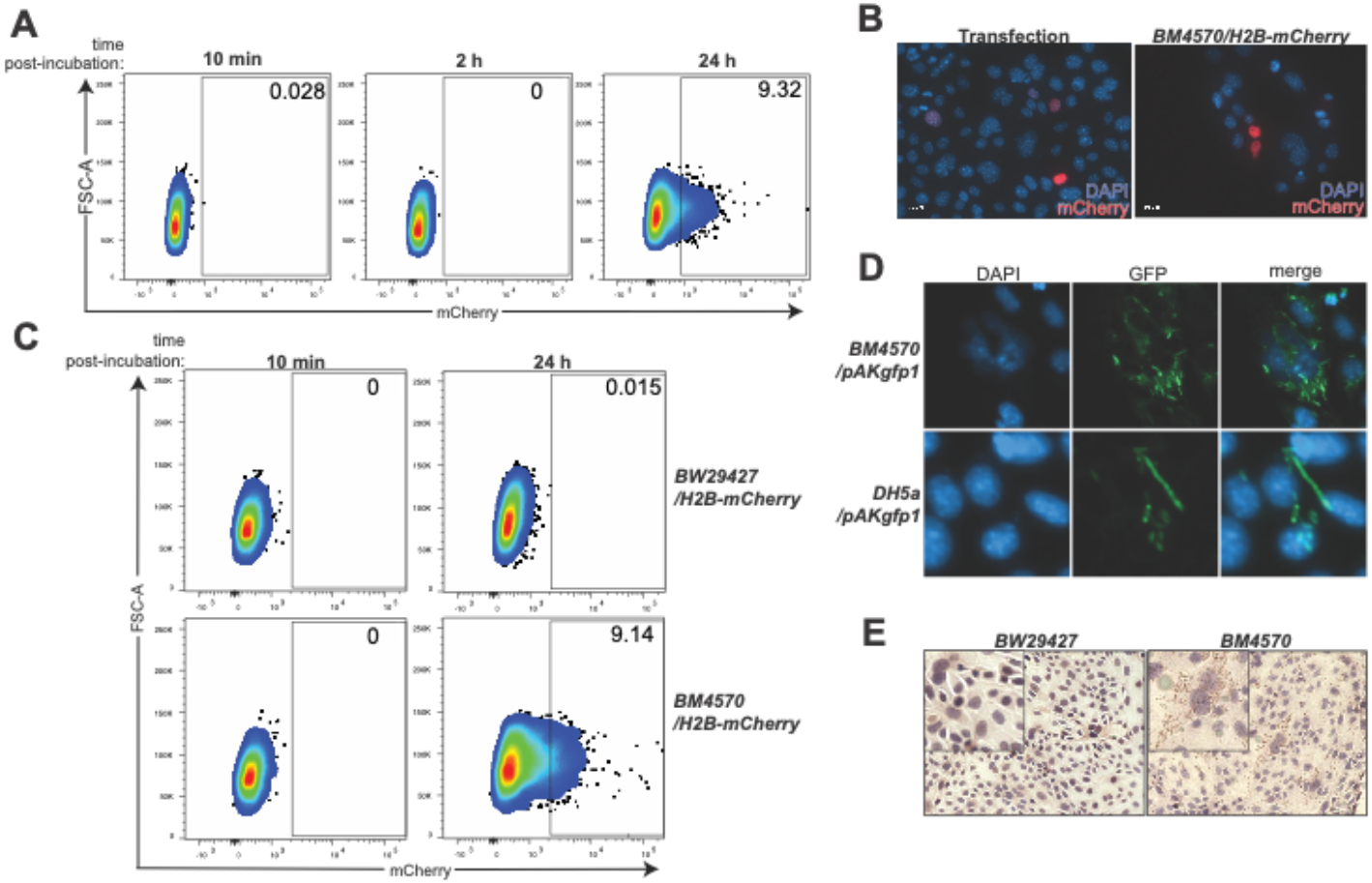
App I Fig 1. Infection of CMT93 with transformed *BM4570* leads to reporter expression in cells.

(A) Schematic of the bactofection protocol. **(B)** Bactofection of CMT93 cells with BM4570/mCherry-IRES-GFP for different periods of incubation and at different MOIs leads to mCherry expression in a time- and dose-dependent manner. Flow cytometry plots taken from one representative experiment. **(C)** Quantitation of experiment in (B). Error bars represent one standard deviation (s.d.) from the mean. n=2 independent experiments. ns, not significant; *, $p<0.05$ (unpaired two-tailed t-test). **(D)** mCherry protein was found in the lysate of only cells transfected with the reporter plasmid or bactofected with BM4570 transformed with the same. **(E)** mCherry reporter plasmid is transmitted to cells directly through BM4570 and not through other means, such as naked plasmid DNA in the media.

To show that the transfer of exogenous genes to cells is specific to *BM4570*, we obtained another *E. coli* strain (*BW29427*) that is also auxotrophic for dap but does not express *inv* or *hly*. Using the same protocol, we bactofected CMT93 with *BW29427*. No fluorescent signal was detected in cells 24 h following co-incubation with *BW29427* carrying the H2B-mCherry reporter (**App I Fig 2C**). Furthermore, we observed by fluorescence microscopy and immunohistochemistry that *BM4570* displayed intracellular cytoplasmic localization within cells following incubation, which was not observed with other bacterial strains (**App I Fig 2D, E**). Thus, due to its expression of the *inv* and *hly* genes, *BM4570* is uniquely able to invade mammalian cells and deliver genetic cargo.

Design of a bacteria delivery system for gene editing.

Having established that *BM4570* is able to deliver exogenous genes to colonic epithelial cells, we next sought to test if *BM4570* could deliver targeting guides to cells to induce gene editing. To reduce the size of the genetic cargo carried by the bacteria, we generated CMT93 cells that stably expressed Cas9 (**App I Fig 3A**). We then designed two parallel approaches to deliver single guide RNAs (sgRNAs) to Cas9-expressing CMT93 (**App I Fig 3B**). The approaches differ in the source of the targeting guides. In one approach, the guide sequence is placed downstream of a bacterial promoter (BBa_J23119) that leads to its constitutive expression in transformed bacteria. Following co-incubation with Cas9-expressing cells, bacteria carrying transcribed guides are expected to release the guides after invasion and



App I Fig 2. The ability of *BM4570* to transfer reporter plasmid to cells is specific to the strain.
(A) Cells bactofected with *BM4570/H2B-mCherry* showed nuclear mCherry expression only 24 h after incubation and wash-out. (B) Nuclear expression of the mCherry reporter is confirmed in bactofected cells by fluorescence imaging. (C) Another dap-auxotrophic strain, *BW29427*, with no invasive properties, is not able to deliver reporter plasmid to cells. (D) *BM4570* transformed with a bacterially-expressed GFP reporter, but not the common *DH5a* strain, display intracellular localization by fluorescence imaging. (E) Immunohistochemistry stains show intracellular localization of *BM4570* but not of *BW29427*.

intracellular lysis. In the other approach, the guide sequence is placed downstream of the U6 mammalian promoter and will therefore only be expressed upon successful delivery of the construct to the nucleus. In the first approach, *BM4570* acts as a delivery system for transcribed sgRNA, while in the second approach, it

is a delivery system for the construct. In addition, both constructs also contain the H2B-mCherry reporter driven from the CMV promoter as a marker for bactofected cells. By Northern blot analysis, we confirmed the presence of *sgAlk* in cells co-incubated with *BM4570* carrying both types of constructs at 24, 48, and 72 hours post-incubation (**App I Fig 3C**).

To test the efficiency of the delivery systems, we chose guides targeting the *Apc*, *Eml4*, and *Alk* loci (*sgApc, sgEml4, sgAlk*), which have been experimentally validated by our lab and others (Maddalo et al. 2014; Roper et al. 2017). *sgApc* targets exon 16 of the *Apc* gene, mimicking the clinically relevant mutation found in colorectal cancer. Following bactofection using our experimentally optimized protocol, we sorted for mCherry-positive cells using flow cytometry, extracted genomic DNA from these cells, and assessed gene editing efficiency by T7 endonuclease I-mediated digestion and high throughput sequencing.

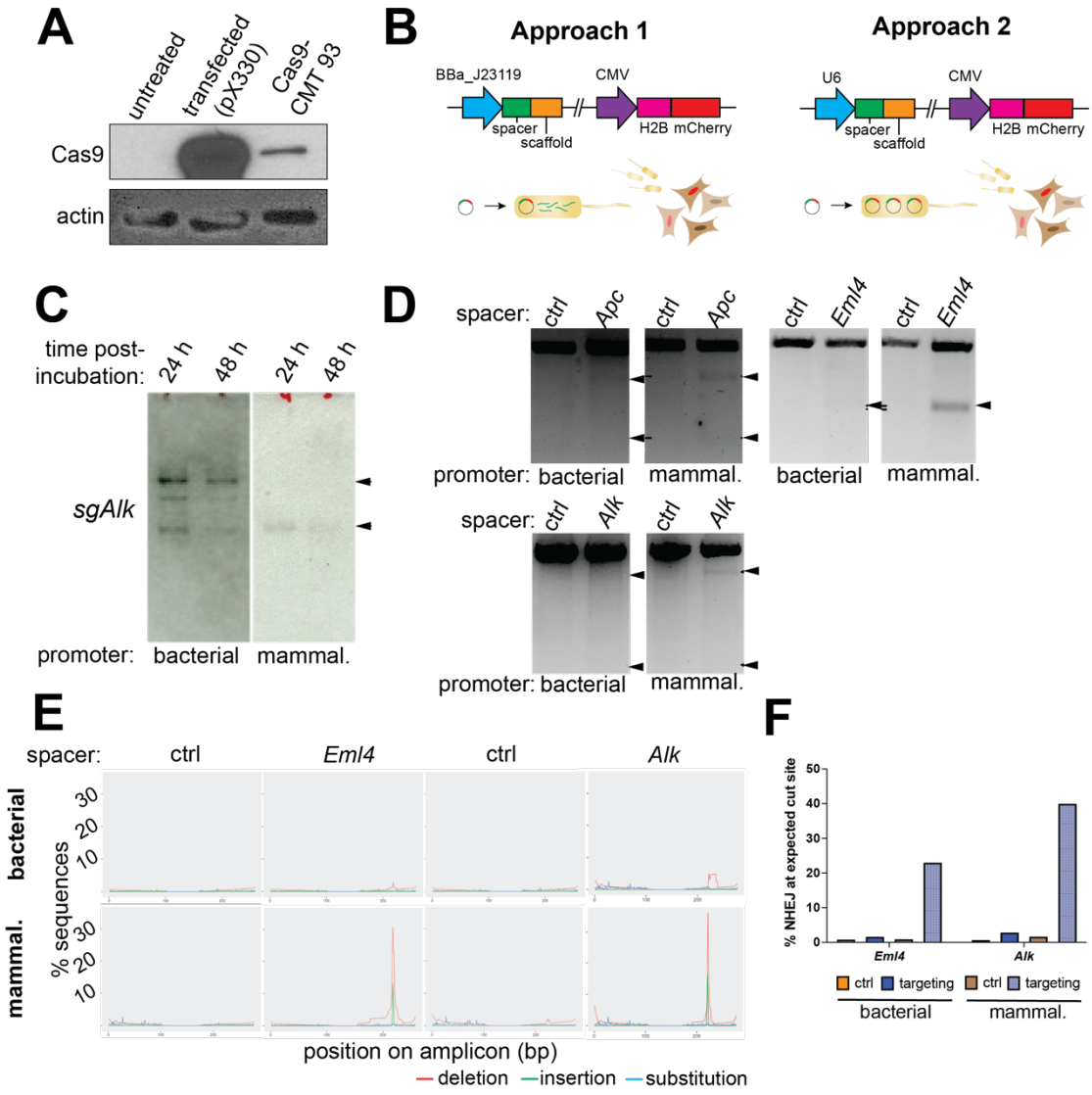
We were able to detect evidence of gene editing using both approaches by T7 endonuclease I-mediated digestion at all three loci (**App I Fig 3D**). Noticeably, gene editing efficiency was higher using the second approach, where sgRNA expression is driven from the U6 promoter. This was further validated by CRISPR

sequencing, which showed that while using *BM4570* to deliver transcribed sgRNAs achieved 3-4% gene editing efficiency, the second approach achieved >30% gene editing at the *Eml4* and *Alk* loci (**App I Fig 3E, F**). The majority of gene edits were deletions mediated by non-homologous end joining (NHEJ). These data demonstrate that *BM4570* is able to deliver CRISPR components to cells *in vitro* to induce gene editing.

Testing the efficiency of bacterial-mediated gene editing *in vivo*.

Having demonstrated successful gene editing with our bacterial delivery system *in vitro*, we next sought to test the feasibility of this system for *in vivo* use. As a first step, we attempted to investigate the bacteria's *in vivo* tropism. Xiang et al. previously reported that mice orally administered with *E. coli* expressing shRNA against *Ctnnb1* showed gene silencing in the intestinal epithelium. *BM4570* and *BW29427* strains carrying the H2B-mCherry marker construct were orally administered to wildtype C57BL/6J mice on a daily basis for up to 5 days. The small intestine and colon were resected from mice after the last oral dosage and immunohistochemistry (IHC) was performed on sections of tissue comprising the duodenum, jejunum, ileum, and colon. We were not able to discern any differences in IHC staining results in tissues resected from mice administered with *BM4570* or *BW29427* (**App I Fig 4A**).

In parallel, we inoculated a small cohort of 12 mice with bacteria carrying targeting constructs against *Apc* to investigate if gene editing could be detected in

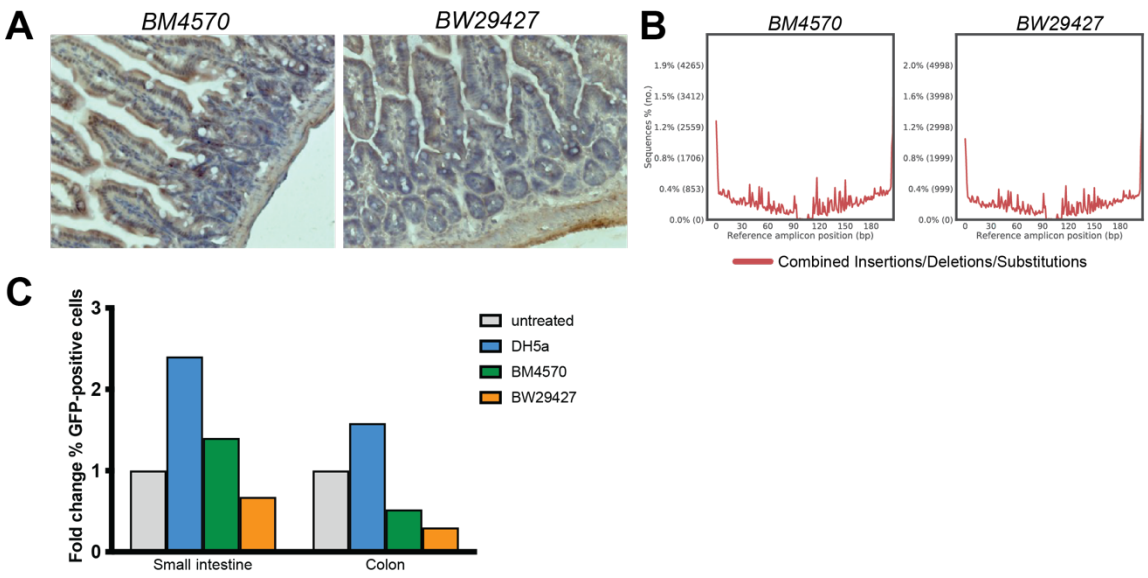


App 1 Fig 3. Delivery of CRISPR components by *BM4570* leads to gene editing *in vitro*.
(A) Cas9 expression was confirmed by Western blot in a CMT93 cell line generated to stably express Cas9. **(B)** Schematic illustrating two parallel approaches to deliver sgRNA to Cas9-CMT93 cells. In Approach 1, the guide spacer sequence is placed downstream of a bacterial promoter, leading to its constitutive expression in transformed bacteria. In Approach 2, the spacer is placed downstream of the mammalian U6 promoter; the bacteria acts as a delivery system for the construct. Both constructs contain H2B-mCherry driven from the CMV promoter as a positive marker for bactofection. **(C)** Northern blot analysis confirmed the presence of *sgAlk* (black arrowheads) in cells co-incubated with *BM4570* carrying both types of constructs. Arrowheads indicate processed and unprocessed sgRNA species. **(D)** Both approaches led to gene editing as assessed by T7 endonuclease I-mediated digestion (black arrowheads). **(E)** Gene editing was validated by CRISPR sequencing. Furthermore, the second approach leads to more efficient editing as indicated by higher percentage of sequences harboring NHEJ marks at the expected cut site. **(F)** Quantification of data in (E).

intestinal tissue. We used mice that were homozygous for a Cas9 knock-in allele (*Rosa26*$^{Cas9\text{-}P2A\text{-}GFP}$). *BM4570* and *BW29427* strains carrying *Apc*-targeting constructs (described earlier) were orally administered to mice daily for 6 days. We did not observe evidence of gene editing by T7 endonuclease I digestion in any of the tissues examined (stomach, proximal/distal intestines, colon). CRISPR sequencing of the tissues also did not reveal evidence of gene editing (**App I Fig 4B**).

Because the intestine constitutes a complex environment characterized by constant interaction between the host immune system and the gut microbiota, we decided to assess if we could detect successful infection of the intestinal epithelium by *BM4570* after *ex vivo* incubation of resected tissue with bacteria before proceeding with further experiments. The small intestine and colon were resected from a homozygous *Rosa26*$^{Cas9\text{-}P2A\text{-}GFP}$ mouse and incubated with different bacterial strains expressing a GFP marker. The epithelial layer was then removed from the tissues, dissociated through enzymatic treatment, and analyzed by flow cytometry for GFP expression. We did not observe an increase in GFP-positive cells from either the small intestine or the colon after incubation with *BM4570* compared to the noninvasive strains *BW29427* and *DH5α* (**App I Fig 4C**). Collectively, these data suggest that to achieve *in vivo* gene editing at an appreciable efficiency in the mouse intestinal epithelium using bacteria as a CRISPR delivery vector requires further optimization and experimentation, which may include perturbing the native gut microbiome and/or removing the protective

mucous layer that lines the intestinal tract (Johansson and Hansson 2016).



App I Figure 4. *In vivo* administration of *BM4570* does not lead to effective infection of intestinal epithelial cells.

(A) Immunohistochemistry staining of colon tissue sections from mice administered *BM4570* or *BW29427* strains transformed with H2B-mCherry construct. **(B)** CRISPR sequencing did not reveal evidence of editing at the *Apc* locus in intestinal tissues isolated from mice administered with bacteria carrying *Apc*-targeting constructs. **(C)** Ex vivo incubation of resected small intestine and colon tissue with our invasive (BM4570) or noninvasive strains (DH5a, BW29427) showed no fold change in GFP expression.

Discussion

In these sets of experiments, we have demonstrated that *E. coli* expressing invasion and listeriolysin can deliver components of the CRISPR/Cas9 platform into mammalian cells and induce gene editing *in vitro*. While the use of bacteria as delivery vectors for exogenous genes is not new, to our knowledge, no one has previously utilized bacteria as an alternative to viral vectors to mediate CRISPR-based gene editing. As invasive bacteria can efficiently transfer genes to nondividing differentiated cells, such as the epithelial cells lining the intestinal mucosa, we envisioned that this could be an attractive alternative to using viruses to deliver CRISPR/Cas9 to the intestinal tissue to generate mouse models of colorectal cancer. While recent studies have shown success in using lentiviral vectors to induce somatic *Apc* mutations in mouse intestinal epithelial cells, leading to tumor formation, the procedure involved a colonoscopy-guided mucosal injection system (Roper et al. 2017). We reasoned that if we could demonstrate efficient gene editing resulting from *in vitro* bactofection of a representative cell type, then we would be able to adapt our approach to mediate somatic gene editing *in vivo*. We used a genetically engineered strain of *E. coli* which was conferred invasive properties yet remained nonpathogenic due to its auxotrophicity for dap. We ascertained that this strain, *BM4570*, was indeed capable of delivering exogenous reporter constructs to our chosen cell type, leading to reporter gene expression in infected cells. This ability was unique to *BM4570*, as bactofection with another dap auxotroph strain, *BW29427,* having no *inv* or *hly* expression,

showed no such effect. However, *in vitro* experiments involving monolayer cultured cells do not recapitulate the complex environment of the intestine and/or colon, which includes a protective mucus lining. Thus, our *in vivo* and *ex vivo* experiments were not able to demonstrate successful infection of the intestinal epithelial cell layer. In order to adapt this system for efficient somatic gene editing *in vivo*, further experimentation and optimization steps will be required.

Materials and Methods

Bacterial cell culture and bactofection.

Non-invasive *E. coli* strain *BW29427* was obtained from the Yale E. Coli Genetic Stock Center. Invasive *E. coli* strain *BM4570* was cultured as described by Grillot-Courvalin et al. (Grillot-Courvalin et al. 1998). Both *E. coli* strains are cultured while shaking overnight at 30°C in brain heart infusion broth (BHI) supplemented with 0.5 mM diaminopimelic acid (dap). The next day, bacteria were harvested by centrifugation in the mid-logarithmic phase of growth (OD600 of ~0.7) and resuspended in DMEM media containing 0.5 mM dap and 10% FBS at 5 x 10^7 c.f.u../mL for bactofection experiments. Pre-seeded cells were incubated with different volumes of resuspended bacteria depending on indicated MOI. Plates were incubated for 2 or 4 h at 37°C, washed with three rounds of DMEM and incubated in complete medium containing 20 µg/ml gentamicin. Other bacterial strains used in this study were cultured in Luria Broth (LB) medium at 37°C while shaking overnight. Expression of reporter constructs in bactofected cells was assessed by flow cytometry (BD LSRFortessa) with DAPI as live/dead stain as well as fluorescence imaging.

Bacterial transformation.

BM4570 was made electrocompetent via published protocols (Warren 2011), and transformed using 1 mm-electroporation cuvettes in an electroporator set at 1.8

kV, 200 ohms and 25 μF. *BW29427* is chemically competent and transformed using standard heat shock protocols.

Mammalian cell culture.

The mouse colon carcinoma CMT93 cell line was purchased from the American Type Culture Collection (ATCC) and cultured in DMEM supplemented with 10% FBS. Cas9-expressing cell lines were obtained through transduction with lentiCRISPR v2 (Addgene #52961) using standard procedures for lentivirus production and infection; transduced cells were selected with puromycin at 2 μg/ml to establish stable expression cell lines.

Plasmid construction.

The mammalian expression vector H2B-mCherry (Addgene #20972) was a kind gift from Dr. Robert Benezra. The psgRNAcos plasmid, where gRNA expression is driven from the constitutive bacterial promoter BBa_J23119 (Addgene #114005) was modified by cloning the H2B-mCherry expression cassette (including promoter and polyA sequence) into SacI-digested psgRNAcos (psgRNAcos-H2B-mCherry). To facilitate cloning of sgRNAs into psgRNAcos-H2B-mCherry, an extra BsaI site within the H2B-mCherry expression cassette was removed by site-directed mutagenesis. To adapt this vector for mammalian expression of gRNA, psgRNAcos-H2B-mCherry was cut with EcoRI and MluI to release the BBa_J23119-gRNA fragment and replaced with U6-gRNA fragment cloned from

pX330 (Addgene #42230) using NEBuilder HiFi DNA Assembly Kit (New England BioLabs). For ease of cloning gRNAs, the spacer region was made the same in both vectors, such that both vectors can be cut with BsaI to insert annealed gRNA oligos with the appropriate BsaI-overhangs against *Apc*, *Eml4*, or *Alk*.

Northern blot analysis.

To detect gRNA expression, bactofected cells were collected in TRIZOL (Invitrogen) and total RNA was isolated according to manufacturers' protocols. For each sample, 5 µg of RNA were resolved in a 15% Urea-PAGE gel and blotted onto a Hybond-N^+ nylon membrane (GE Healthcare). Membranes were UV-cross-linked and hybridized overnight with g^{32}P-labelled probes against *sgAlk*.

Genome editing analysis.

Gene editing was detected from PCR products using the T7 endonuclease I assay (New England BioLabs) according to the manufacturer's protocols. CRISPR sequencing was performed by the MSKCC Integrated Genomics Operations Core Facility. Paired-end reads were analyzed and graphical reports generated using CRISPresso (Pinello et al. 2016).

Immunocytochemistry (ICC) and Immunohistochemistry (IHC).

Cells were grown in monolayer in culture chamber slides (Millipore). Following bactofection, cells were fixed with 4% paraformaldehyde (PFA) and permeabilized

with 0.5% Triton X-100. Cells were blocked with 1% BSA in PBST followed by primary antibody incubation (rabbit anti-GFP, Invitrogen) overnight at 4°C followed by incubation with a biotinylated secondary antibody (goat anti-rabbit IgG, Vectastain Elite ABC Kit) for 1 h at room temperature.

For IHC on mouse tissue, the stomach, small intestine, and colon were resected from mice. Before fixing, the small intestine and colon were cleared of fecal matter. Tissues were fixed in 4% PFA and cryopreserved using the sucrose gradient method. Tissues were embedded in OCT sectioning medium and stored at -80°C until sectioned. Tissues were sectioned at 10 μm thickness onto Superfrost microscope slides and permeabilized with 0.5% Triton X-100 and blocked with goat serum in PBS followed by primary antibody incubation (rabbit anti-mCherry, Invitrogen) overnight at 4°C. Endogenous peroxidase was quenched with 3% H2O2 followed by incubation with biotinylated secondary antibody (goat anti-rabbit IgG, Vectastain Elite ABC Kit).

All slides were incubated with Vectastain Elite ABC Reagent (Vector Labs) containing horseradish peroxidase followed by staining with the ImmPACT DAB Kit containing diaminobenzidine (DAB) chromagen (Vector Labs). Slides were counterstained with hematoxylin, dehydrated, then cleared in xylene, and mounted with an aqueous-based mounting medium (Aqua-Mount, Lerner Laboratories).

Western blotting.

Whole cell extracts were isolated using 1X RIPA buffer supplemented with protease and phosphatase inhibitors. Protein concentration was determined using BCA Protein Assay Reagent (Pierce). 30 µg of protein were separated on a 4-12% Tris-acetate gel and electrophoretically transferred to PVDF membranes. Membranes were incubated with primary antibody against Cas9-FLAG (mouse anti-FLAG, Sigma) or against mCherry (rabbit anti-mCherry, Abcam) followed by incubation with horseradish peroxidase-linked secondary antibody (anti-mouse IgG or anti-rabbit IgG, Amersham). Membranes were developed with ECL prime (Bio Rad).

Administration of live bacteria to mice via oral gavage.

For investigating the *in vivo* localization of bacteria, four *Rosa26*$^{Cas9\text{-}P2A\text{-}GFP}$ mice were randomly assigned to receive either *BM4570* or *BW29427*, with an additional mouse that received only PBS. Each group was administered 1 x 10^9 c.f.u. *E. coli* transformed with H2B-mCherry reporter plasmid using 22G oral gavage needles. The treatment was carried out daily for five days. Mice were killed on the last day, 3 hours after the last inoculation, and tissues were resected. For investigating gene editing *in vivo*, a total of twelve *Rosa26*$^{Cas9\text{-}P2A\text{-}GFP}$ mice were assigned to receive *BM4570* harboring *Apc*-targeting plasmids (sgRNA driven from either bacterial or mammalian promoter). For each group, two mice received the control spacer plasmid, while the other four mice received the targeting plasmid. Each group was further subdivided to receive either 1 x 10^9 c.f.u. of bacteria or 5 x 10^9 c.f.u. As

above, the treatment was carried out daily for five days, and tissue was collected on the last day of treatment.

***Ex vivo* intestine tissue incubation with bacteria.**

Briefly, the small intestine and colon were resected from *Rosa26*$^{Cas9\text{-}P2A\text{-}GFP}$ mice. Fecal matter was flushed from the intestinal lumen and 5 x 10^9 c.f.u. of bacteria resuspended in PBS supplemented with 0.5 mM dap was used to fill the luminal space. After 2 h of co-incubation, the bacteria were flushed from the lumen with three rounds of washing. The small intestines and colons were then cut longitudinally to open the tissue and minced into a smooth paste-like texture with a scalpel. Pieces were incubated with a collagenase/dispase digestion mix (75 U/ml collagenase and 20 µg/ml dispase) for 3 h at 37°C under shaking conditions. After disaggregation, FBS was added to terminate the dispase reaction (to 5% final concentration). Cells were then sequentially filtered through 100 µm, 70 µm and 40 µm cell strainers. Single cells were collected by centrifugation and resuspended in 0.1% BSA, 1 mM EDTA FACS buffer for flow cytometry analysis.

Statistical analysis.

All statistical analyses were performed using Prism software. Unless otherwise noted, data were analyzed using unpaired two-tailed t-test. *, $p<0.05$; n.s., not significant.

APPENDIX II: Dissecting the oncogenic role of del13q14 in chronic lymphocytic leukemia

Introduction

Over the past decade, The Cancer Genome Atlas (TCGA) and other large-scale cancer genome sequencing efforts have offered an unprecedented view of the genetic landscape of human cancers. This has expedited the discovery and characterization of cancer-causing abnormalities at the genomic, transcriptomic, proteomic, and epigenetic levels. Nevertheless, the molecular mechanisms underlying highly recurrent chromosomal abnormalities in some cancer types still remain incompletely characterized, hampering the development of effective therapies, and highlighting the need for careful, molecular dissection of affected loci/genes.

One such example is the chromosome 13 deletion at cytoband 13q14 in chronic lymphocytic leukemia (CLL). CLL is the most common adult leukemia in the Western world. It is characterized by the clonal expansion of CD5+ CD23+ B cells in blood, bone marrow, and secondary lymphoid tissues. Its genome usually harbors 0-2 copy number alterations, with recurrent chromosomal abnormalities within chromosome 11q22-q23, chromosome 17p13, chromosome 13q14, and in rare cases, trisomy of chromosome 12 (Zhang and Kipps 2014). Interstitial deletions involving chr13q14 (commonly known as del13q14) are by far the most

common alterations found in CLL, with over 50% of patients harboring this abnormality. In addition, del13q14 is often monoallelic, with no somatic mutations in genes mapping to the non-deleted chromosome (Mertens et al. 2002; Migliazza et al. 2001). Clonal analysis suggests that when present, del13q14 is the initiating event (Landau et al. 2015), yet despite more than two decades of work to pinpoint the responsible element(s) within del13q14, the molecular mechanisms through which this highly recurrent deletion promotes CLL remain unclear and no protein coding tumor suppressor genes have been identified in the region.

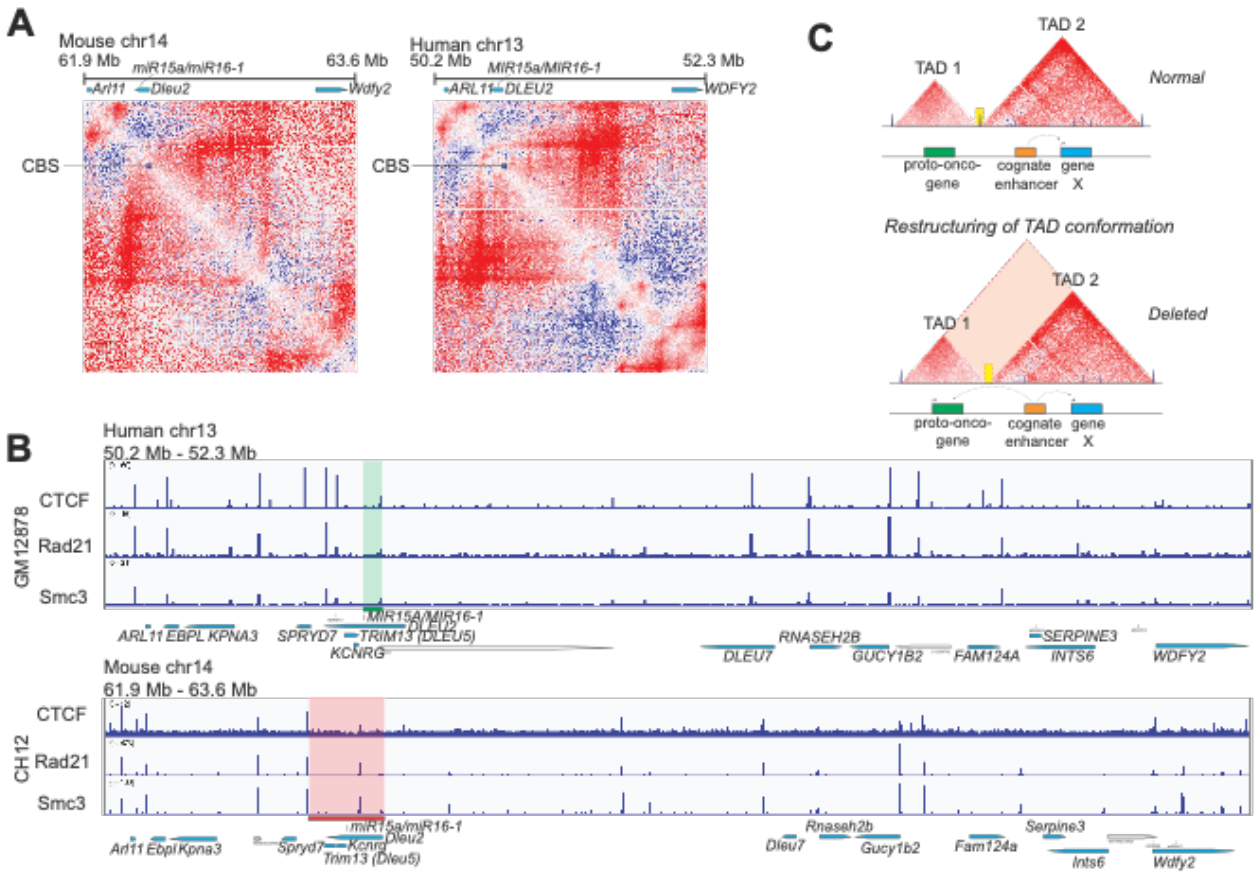
Among the genes mapping to this region, the miRNA cluster *miR15a~16-1* has been proposed to be the responsible tumor suppressor element because of its ability to target BCL2 in *in vitro* studies (Calin et al. 2002; Cimmino et al. 2005). However, BCL2 expression levels in patient samples are highly variable and not correlated with del13q14 status (Ouillette et al. 2008). Furthermore, the recent generation of knock-out mice to model del13q14 strongly suggests that deletion of *miR15a~16-1* only partially explains CLL pathogenesis. In particular, mice with heterozygous deletion of the miRNA cluster do not develop lymphoproliferative disease; however, the heterozygous deletion of a 120 Kb locus, which encompasses the miRNA cluster and mimics the “minimally-deleted region” (MDR) that is observed in patients, is sufficient to promote the development of a lymphoproliferative disease (Klein et al. 2010). Additionally, although homozygous animals in both conditions developed CLL at significant frequencies, *MDR-/-* animals had twice the rate of disease compared to mice harboring homozygous

loss of the miRNA cluster only, and succumb to the disease at an earlier age (Klein et al. 2010). Collectively, these recent studies in mice indicate that an additional element exists within the MDR that is essential for CLL pathogenesis.

In addition to somatic mutations that drive cancer initiation through altering protein-coding genes, it is becoming increasingly clear that mutations affecting regulatory elements of the genome can also play an important role. Among these elements, CTCF/cohesin-binding sites (CBS) have been proposed to define the boundaries of so called topologically associating domains (TADs). These high-order chromatin structures have been proposed to play an important role in transcriptional regulation by bringing distal promoter and enhancer elements into close proximity as well as preventing interaction between elements housed in distinct TADs (Szabo, Bantignies, and Cavalli 2019). TADs can furthermore be broken down into kilobase-scale clusters, called sub-TADs, within which gene expression can be further fine-tuned at a local level (Matthews and Waxman 2018). The fact that perturbation of CBSs has been shown to lead to aberrant gene expression in development (Lupianez et al. 2015) and disease (Flavahan et al. 2019; Flavahan et al. 2016) demonstrates the importance of maintaining the integrity of these boundary sites. Despite the wealth of correlative evidence and cell-based studies, however, no conclusive evidence that perturbation of individual CBS can promote tumor initiation and progression in vivo has been reported so far. Through an analysis of TCGA, MSK IMPACT, and publicly available Hi-C datasets, we found that although the minimally-deleted region in del13q14 lacks conserved

protein-coding genes that might explain its role in CLL, it contains a CBS that maps near the boundary between two putative TADs (**App II Fig 1A)**. Both the CBS and the two TADs are conserved in humans and mice, where they are part of a large syntenic block (**App II Fig 1B**). Crucially, this CBS is included within the 120 Kb region that was deleted in the aforementioned mouse model of CLL. Based on these findings, we hypothesized that deletion of this CTCF/cohesin-binding site and the consequent rewiring of local promoter-enhancer interactions is a key mechanism through which del13q14 promotes CLL (**App II Fig 1C**).

We generated an allelic series of isogenic human cell lines using the MEC1 background (a patient-derived CLL line without deletions in chr13q14) to directly test our hypothesis that that the loss of the conserved CBS in del13q14 is the crucial genetic event that leads to rewiring of local promoter-enhancer interactions. Our design also allowed us to detect any effects arising from the deletion of other elements within this region.



App II Figure 1. Genomic and 3D organization of the del13q14 locus in humans and mice.
(A) Snapshots of Hi-C maps showing the syntenic block in activated mouse B-cells (left panel) and a human lymphoblastoid cell line (right panel). Note the conservation of TAD structure within this region in the two species. The position of the CBS that is targeted for deletion is indicated on both maps. **(B)** Genomic overview of the del13q14 region depicting orthologous genes between mouse and human in light blue, and CTCF, Rad21, and Smc3 ChIP-seq peaks in dark blue. The human ChIP-seq peaks are from GM12878, and the mouse ChIP-seq peaks are from CH12 (a B-cell lymphoma cell line). The syntenic block is bounded by the genes *ARL11* and *WDFY2*. The green box highlights the 36 Kb region that was deleted in MEC1 cell lines. For comparison, the red box depicts the MDR that was deleted in the knock-out mouse models generated by Klein et al. (Klein et al., 2010). **(C)** Proposed mechanism of how deletion of a CBS in del14q14+ CLL disrupts local TAD configurations and rewires promoter-enhancer interactions, leading to aberrant expression of a proto-oncogene.

Results

Generation of a series of del13q14 alleles in human cells.

To dissect the functional consequences of del13q14, we utilized CRISPR/Cas9-based methods to generate an allelic series of isogenic human cell lines harboring various deletions of the del13q14 region. These deletions were designed to directly test our working hypothesis that loss of the conserved CBS (described in the introduction) is a crucial genetic event, but would also allow us to detect a possible role for other elements within the minimally deleted region (**App II Fig 2A**). We chose the MEC1 cell line to carry out our experiments, as it is a near-diploid, patient-derived CLL line that does not harbor del13q14.

While we were able to easily generate MEC1 clones harboring small homozygous deletions encompassing either the *miR-15a~16-1* cluster ($miR^{-/-}$) or the CBS ($CBS^{-/-}$) by transfecting Cas9 protein and two gRNAs targeting the desired breakpoints, this approach proved too inefficient to generate the larger deletions (>36 Kb). To overcome this problem, we took advantage of a novel strategy developed by the laboratory of Scott Lowe (Francisco Barriga, personal communication) termed 'MACHETE' (Molecular Alteration of Chromosomes with Engineered Tandem Elements) to generate arbitrarily large deletions with high efficiency (**App II Fig 2B)**. This method is performed in two steps. First, a cassette encoding thymidine kinase (TK), EGFP, and the puromycin resistance gene is inserted within the region that is to be targeted for deletion. The cassette, in the

form of a PCR-amplified donor sequence with microhomology arms to the target region, was delivered in tandem with Cas9 and a single gRNA designed to cleave at the integration site. Cells harboring the correctly integrated cassette were then selected with puromycin.

The presence of the TK gene in the cassette confers sensitivity to ganciclovir. Upon delivery of Cas9 and gRNA pairs designed to generate large deletions surrounding the integrated cassette, rare cells harboring the desired deletion could therefore be readily selected by growing them in the presence of ganciclovir (**App II Fig 2B)**.

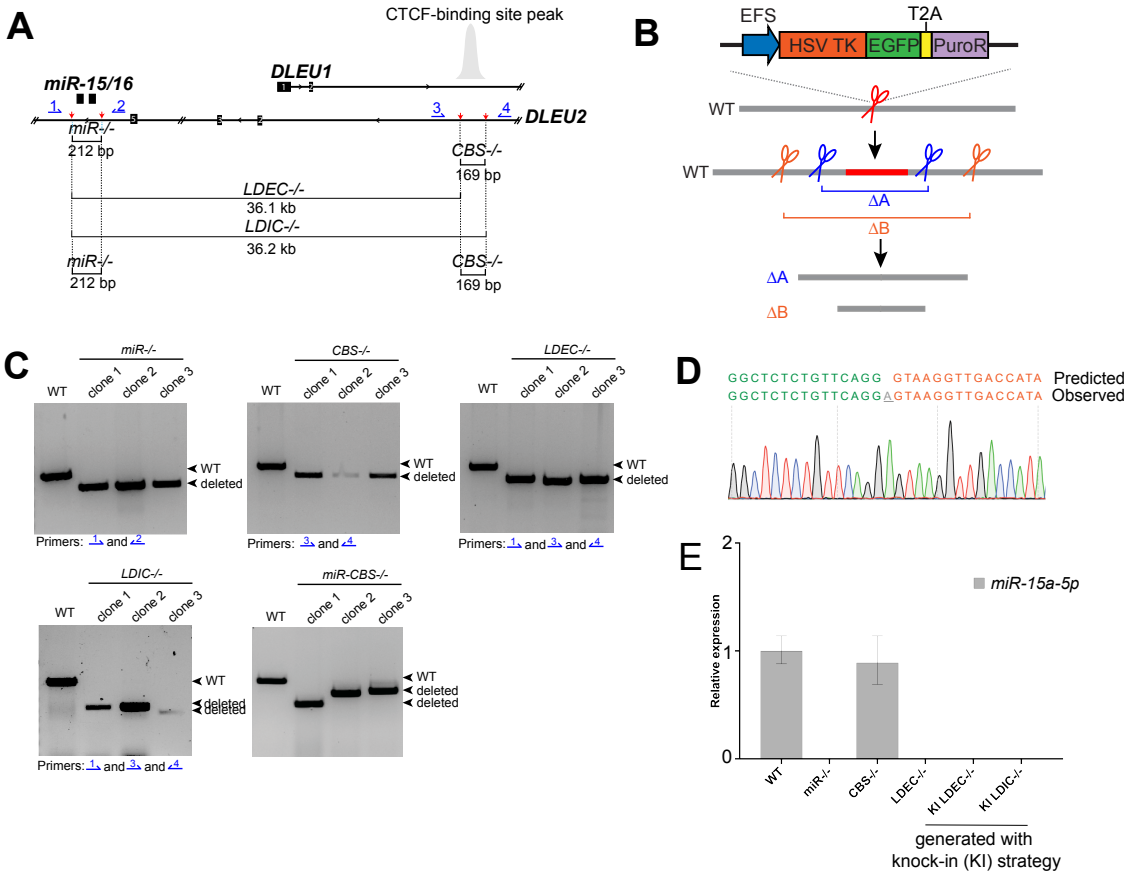
Ultimately, we isolated three independent MEC1 clones featuring each genotype: (a) deletion of *miR-15a~16-1*, (b) deletion of the CBS, (c) a large deletion that encompassed both the *miR-15a/16-1* cluster and the CBS, (d) a slightly smaller deletion that included the *miR-15a~16-1* cluster but terminated just before the CBS, and (e) deletion of only the *miR* cluster and the CBS but not the intervening region (**App II Fig 2A, C)**. For each method, we also isolated three independent control clones: for the two small deletions, these were cells that were targeted but not edited, and for the knock-in strategy, these were cells from which only the integrated selection cassette was excised, restoring the wild-type configuration of the locus.

In addition to verifying the clones by sequencing the deletion boundaries (**App II Figure 2D)**, we also confirmed the selective and complete abrogation of *miR-15a* expression in clones harboring deletions encompassing the *miR-15a~16-*

1 locus (**App II Figure 2E**). Interestingly, despite deletion of the entire cluster, the levels of mature *miR-16* were not significantly affected in these clones (not shown), likely due to the presence of a *miR-16* paralogue (*miR-16-2*) on chromosome 3 whose mature sequence is identical to *miR-16-1*.

Consequences on gene expression.

Having successfully generated three independent clones for each deletion, we first examined their consequences on the local and global transcriptome by generating RNA-seq datasets from each clone. We postulated *a priori* three alternative scenarios for expected results from the analysis: (a) The only important genetic element in the MDR is the *miR-15a~16-1* cluster; (b) Loss of the CBS rewires local promoter-enhancer interactions, affecting gene expression *in cis*; (c) An additional important genetic element is included in the region between *miR-15a~16-1* and the CBS. In scenario A, we would expect to see changes in gene expression imputable to de-repression of *miR-15* and possibly *miR-16* targets in *miR-/-* clones, but no additional changes in clones harboring larger deletions extending beyond the *miR* cluster. This would be at odds with the experimental data from the mouse knock-out models discussed in the introduction, and thus would be unlikely. In scenario B, clones harboring deletion of the CBS, alone or in combination with deletion of the *miR* cluster, would display de-regulated expression of one or more genes residing within 2-3 Mbp surrounding the deleted region. Importantly, expression of these genes should not be affected in clones harboring deletion of



App II Figure 2. Generation of an allelic series of human cell lines harboring various deletions in chr13q14.

(A) Schematic showing the position of the deletions relative to *DLEU2*, *miR-15a~16-1* and the CBS. gRNAs targeting the desired breakpoints are indicated with red arrows; primers used to detect the deletions are indicated by blue arrows. **(B)** Overview of the knock-in strategy used to generate the larger deletions. The gRNA targeting the integration site of the cassette is shown as red scissors; scissors in blue and orange indicate gRNA pairs designed to target desired breakpoints in the second step of the method. **(C)** Genotyping PCR using primers indicated in panel A showing successful isolation of three independent clones for each indicated deletion. **(D)** Representative chromatogram of one of the clones showing predicted and observed sequences at the deletion boundaries. **(E)** RT-qPCR showing complete abrogation of miR-15a expression in clones harboring deletions encompassing the *miR-15a~16-1* locus. The results are shown as the averaged value of clones with each deletion.

the *miR* cluster alone, including the larger deletion that terminates immediately before the CBS. This scenario would assign a clear role to the CBS and immediately suggest a potential mechanism for the pathogenesis of del13q14+ CLL. Lastly, in scenario C, we would observe gene expression changes in clones harboring larger deletions extending beyond the *miR* cluster, even when the CBS is not included in the deletion. If genes within 2-3 Mbp from the deletion were affected, then it would be reasonable to hypothesize that one or more regulatory elements acting *in cis* were affected by the deletion. If, on the other hand, the changes in gene expression were restricted to genes located far away from the deletion or on other chromosomes, then the most plausible interpretation would implicate a genetic element present in the region between *miR-15a~16-1* and the CBS that is acting *in trans* (for example, a long non-coding RNA or a small peptide). The obvious candidate in this case would be *DLEU2*, the host gene for *miR-15a~16-1,* and its role could be directly examined with 'rescue' experiments performed by ectopically expressing its cDNA in the deleted clones.

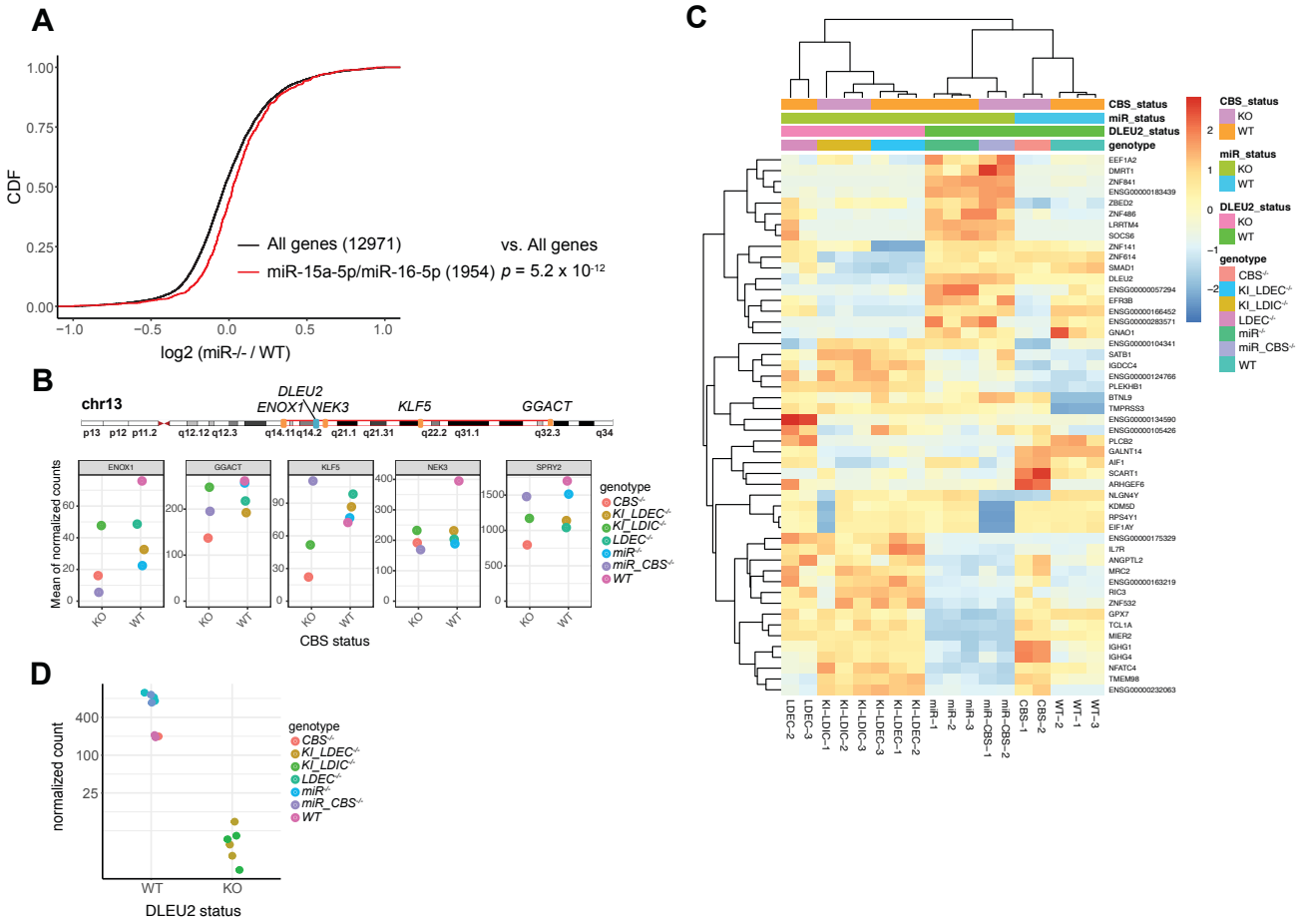
The results of our bioinformatic analysis was suggestive of scenario C. Notably, in clones harboring deletion of the *miR15a~16-1* cluster, we observed a mild but statistically significant change in gene expression of its targets (**App II Fig 3A**), thus ruling out scenario A. Of the genes on chromosome 13 whose expression changed significantly, none were unique to the CBS-deleted clones (**App II Fig 3B**); this ruled out scenario B. However, clustering analysis on the data showed that clones harboring the larger, 36 Kb deletions clustered together regardless of

whether the CBS was included in the deletion (**App II Fig 3C**). This finding signified to us that the deletion of a genetic element, or elements, in the region between the *miR15a~16-1* cluster and the CBS was the driving force of gene expression changes in clones harboring the larger deletions.

Identifying genes regulated by *DLEU2*.

One of the candidates immediately implicated was *DLEU2*, the long noncoding RNA that houses *miR15a~16-1*. Exons 2 to 6 of *DLEU2* are deleted as part of the 36 Kb deletion, resulting in complete abrogation of *DLEU2* expression in these clones. *DLEU2* expression was not affected in other clones, with the exception of those harboring the *miR* deletion, where *DLEU2* expression was upregulated (**App II Fig 3D**). To identify what other genes were negatively correlated with *DLEU2* expression, we performed a correlation analysis and identified the interleukin-7 receptor *IL7R* as significantly upregulated in *DLEU2*-KO cells but not in the other conditions (**App II Fig 4A**). IL7R and its ligand, IL7, play important roles in B cell lineage commitment and early development by feeding into downstream signaling pathways including PI3K, MAPK, and JAK-STAT pathways (Reth and Nielsen 2014). Expression of *IL7* and *IL7R* is also known to promote T cell acute lymphoblastic leukemia (T-ALL) development (Zenatti et al. 2011; Oliveira et al. 2019).

To find other genes that were also upregulated in *DLEU2*-KO cells but not the other genotypes, we performed gene clustering to identify groups of genes with



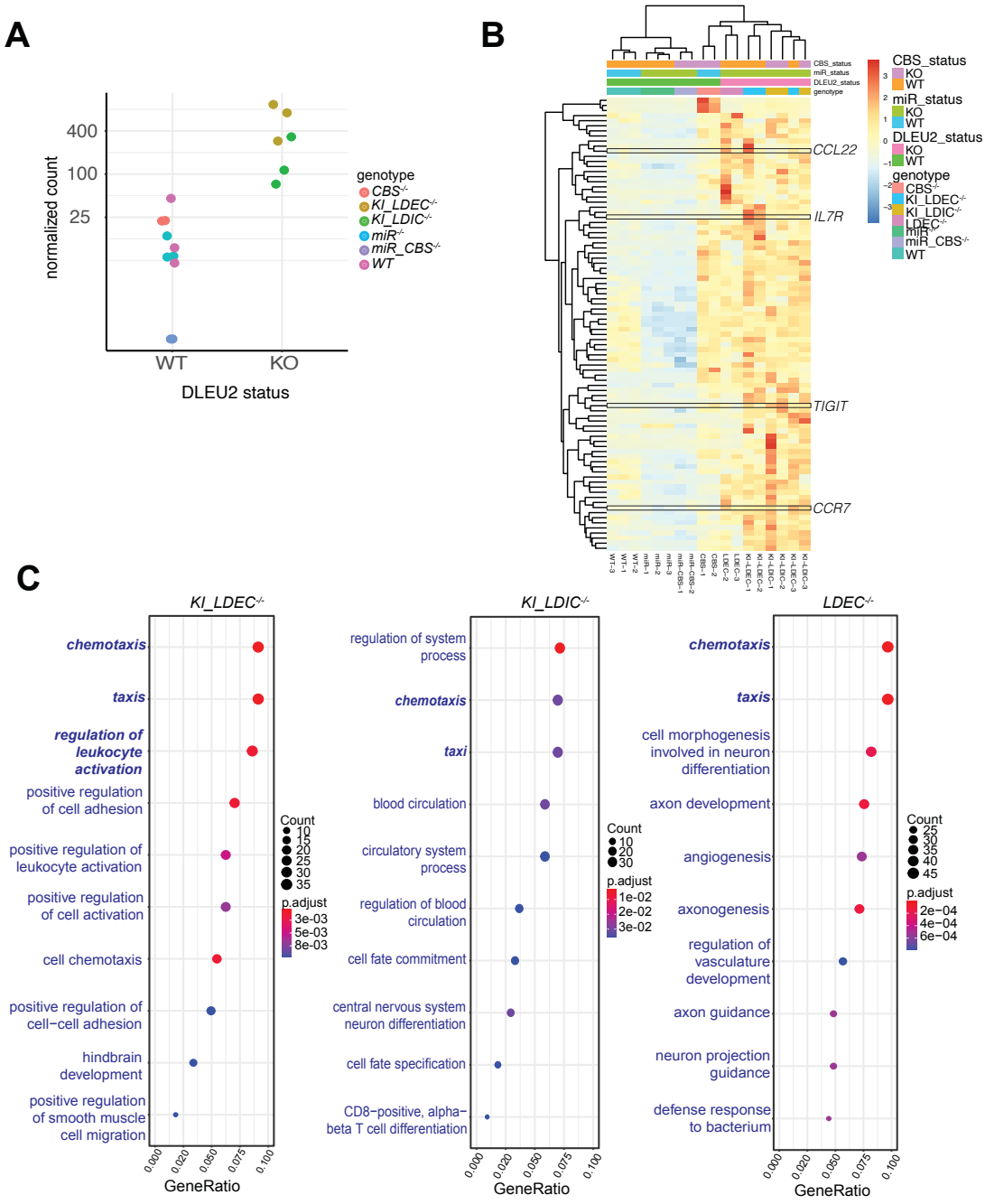
App II Figure 3. Bioinformatic analysis on RNA-seq data from clones harboring chr13q14 deletions is suggestive of an important genetic element between *miR-15a~16-1* and the CBS.
(A) Bioinformatic analysis of RNA-seq data from *miR-/-* clones show modest de-repression of *miR-15a~16-1* targets. p value, two-sided Kolmogorov-Smirnov test. **(B)** Schematic displaying the location of genes mapping to chr13 that were significantly differentially expressed in *CBS-/-* clones (top) and the average of their normalized expression in CBS-WT and CBS-KO clones. **(C)** Heatmap displaying the clustering pattern of the top 50 most variably expressed genes using an unsupervised clustering method. **(D)** Dot plot showing *DLEU2* expression in the different genotypic backgrounds. Expression is plotted as normalized counts on a log10 scale.

similar expression patterns across conditions. As a result of this analysis, we were able to identify a number of other genes which were also upregulated in *DLEU2*-KO clones, some of which have known roles in the progression of other types of leukemia, such as *CCL22*, *CCR7* and *TIGIT* (**App II Fig 4B**). Furthermore, gene ontology analysis demonstrated that *DLEU2*-KO clones were enriched in GO processes such as regulation of leukocyte activation, migration, and chemotaxis (**App II Fig 4C**). This is in line with our finding that genes like *CCR7* and *CCL22* are also upregulated in these clones. Notably, these genes have been implicated in the ability of leukemic cells to infiltrate the CNS and in mediating T reg immigration into tumor tissue to inhibit antitumor immunity (Buonamici et al. 2009; Rapp et al. 2019).

Investigating a potential relationship between *DLEU2* and *IL7R*.

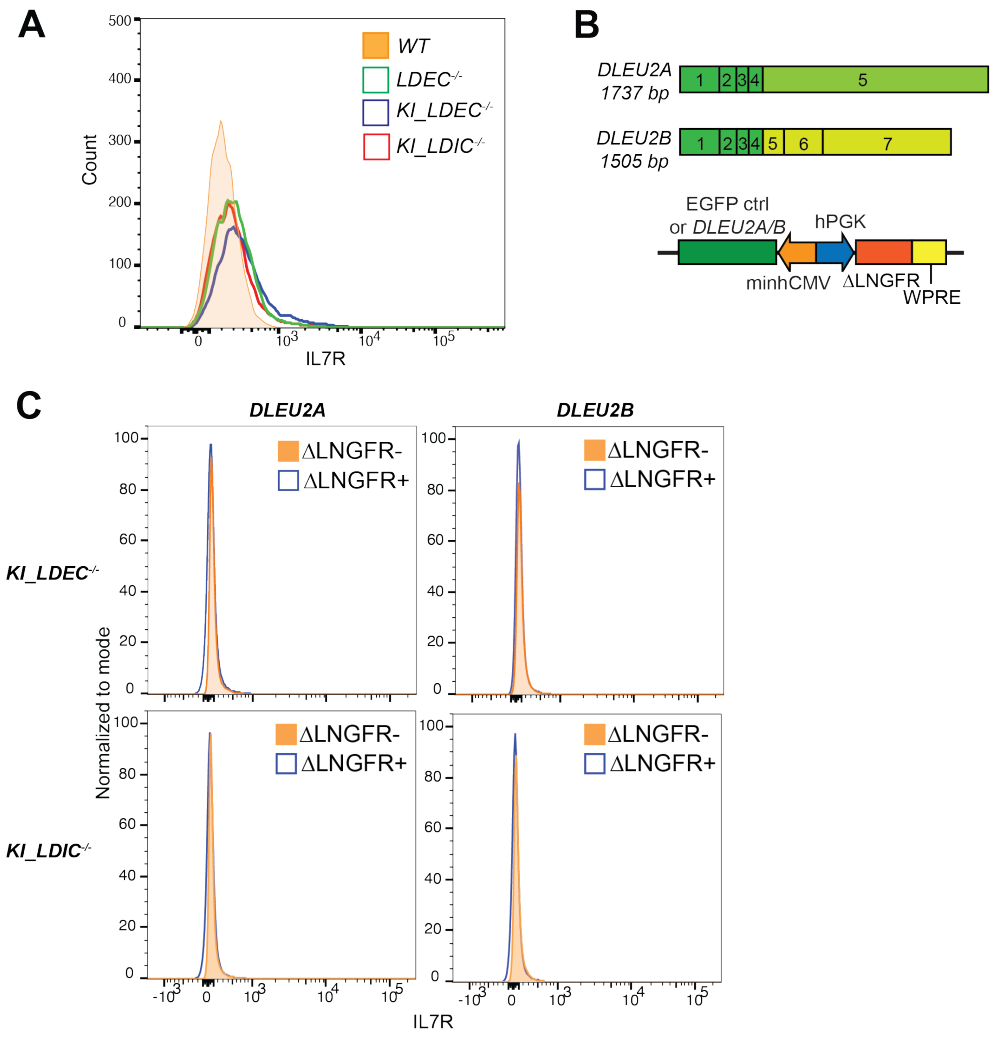
Because our RNA-seq data showed that *IL7R* expression was orders of magnitude higher in *DLEU2*-KO clones, with low to no expression in *DLEU2* wild type clones, we decided to focus on investigating a potential relationship between *DLEU2* and *IL7R*. We were able to validate the specific upregulation of *IL7R* in *DLEU2*-deleted clones at the proteomic level by flow cytometry analysis, which verified that the cell surface receptor was indeed expressed at a higher level in *DLEU2*-KO cells (**App II Fig 5A**).

Having verified that we could detect significant changes in *IL7R* expression at the protein level, we next set out to determine if re-expressing *DLEU2* in the



App II Figure 4. *IL7R* and other genes are negatively correlated with *DLEU2* expression. **(A)** Correlation analysis revealed *IL7R* is specifically upregulated in *DLEU2*-KO clones only. The dot plot shows *IL7R* expression plotted as normalized counts on a log10 scale. **(B)** Heatmap showing 89 significantly differentially expressed genes with similar gene expression profiles as *IL7R*. Genes implicated in the progression of leukemia, namely *CCL22*, *CCR7*, and *TIGIT*, are highlighted with boxes. **(C)** Over-representation analysis on biological processes associated with significant genes in the *DLEU2*-KO showed enrichment in genes involved in leukocyte activation and migration.

knock-out clones would result in decreased expression of *IL7R*. To do this, we cloned two isoforms of *DLEU2* that we identified through RACE-PCR to be expressed in the *DLEU2*-WT clones, which we refer to as *DLEU2A* and *DLEU2B*. These transcripts are ~1.7- and ~1.5-kb respectively. We cloned the cDNAs into a lentiviral vector with a bi-directional promoter to allow simultaneous expression of the transgene as well as a truncated form of LNGFR (low-affinity nerve growth factor receptor), which served as a transduction marker (**App II Fig 5B**). We titrated the amount of the lentiviral vector used in the transduction to ensure a subpopulation of cells would not be transduced and thus would remain ΔLNGFR-negative. This allowed us to compare the effects on IL7R in transduced and non-transduced cells within the same sample. If re-expression of *DLEU2* does indeed downregulate IL7R, we would expect that the subpopulation of cells positive for ΔLNGFR would have lower IL7R expression compared to ΔLNGFR-negative cells. However, when we transduced *DLEU2*-KO cells with vectors encoding either isoform of *DLEU2*, we did not see a difference in IL7R levels between transduced and non-transduced cells (**App II Fig 5C**).



App II Figure 5. Investigating a potential regulatory relationship between DLEU2 and IL7R. **(A)** Flow cytometry analysis verified that IL7R protein is overexpressed in *DLEU2*-KO clones compared to WT cells. **(B)** Two identified isoforms of *DLEU2* were subcloned into a bi-directional lentiviral vector. Transduced cells are identified by the expression of ΔLNGFR. **(C)** Ectopic expression of either isoform of *DLEU2* in *DLEU2*-KO clones showed no difference in IL7R levels between transduced (ΔLNGFR+) and non-transduced (ΔLNGFR-) cells.

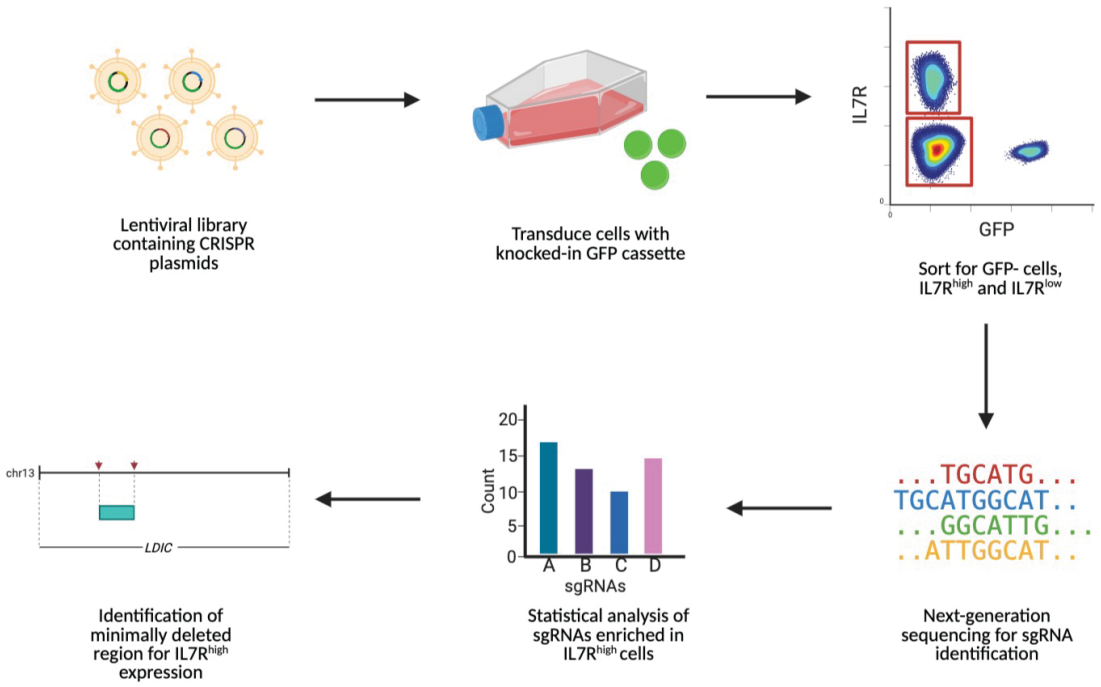
Discussion

In these sets of experiments, we have generated an allelic series of isogenic human cell lines carrying various deletions within the region affected by del13q14 in order to interrogate the impact of deleting a conserved CTCF-binding site (CBS) on the transcriptome. In parallel, we agnostically surveyed the transcriptomic effects of combinatorial and/or additional deletions within the region, including the intervening locus between *miR-15a~16-1* and the CBS.

The results of our analysis suggest that, in line with published mouse data, deletion of the *miR-15a~16-1* cluster appears to have a modest effect on its targets, leading to their de-repression. Deletion of the CBS did not appear to exert any specific effects on the expression of genes *in cis*. However, clones harboring the large deletion, all of which showed complete abrogation of *DLEU2* expression, clustered together regardless of inclusion or exclusion of the CBS. Furthermore, we found that *IL7R* was significantly and highly upregulated in these *DLEU2*-KO clones, along with other genes with known roles in the progression of other types of leukemia. We next sought to explore a potential regulatory relationship between *DLEU2* and *IL7R* by ectopically restoring the expression of *DLEU2* in KO clones. However, we did not observe a downregulation in *IL7R* expression upon re-expression of *DLEU2*.

While we cannot rule out the possibility that the increased IL7R expression we observed in *DLEU2*-KO clones stem from an experimental artifact, the likelihood is very low given we saw the same upregulation in all three independent

clones for both genotypes with *DLEU2* deletion. Our results collectively suggest that *DLEU2* expression is not linked to IL7R, but there is still a possibility that deletion of an unidentified element between the *miR-15a~16-1* cluster and the CBS is driving the upregulation of IL7R. Future experiments could investigate this by generating a screening library of guide RNAs targeting roughly 1-kb increments along the 36-kb region. After transduction of this library of gRNAs into cassette-targeted MEC1 cells, one would expect to generate a polyclonal population of cells carrying deletions of varying sizes. Regardless of size, all deletions would lead to loss of the cassette; thus, cells harboring deletions could be sorted based on loss of GFP expression. From this pool of cells, a subpopulation of cells with increased IL7R expression should arise, which can be detected by flow cytometry analysis using our experimentally-verified conjugated antibodies. By subjecting both populations of GFP-negative, $IL7R^{high}$ and $IL7R^{low}$ cells to next-generation sequencing followed by bioinformatic analysis, we could determine which gRNAs are enriched in the $IL7R^{high}$ cells, and thus identify a minimally deleted region underlying the phenotype (**App II Fig 6**). The results from these experiments may feasibly help to narrow the search for an unknown element within the region that may be functionally linked to *IL7R*.



App II Figure 6. Tiling assay to identify genetic element(s) functionally linked to *IL7R* expression.
Schematized outline of a proposed experiment using a tiled array approach to screen for a genetic element linked to IL7R expression. Briefly, a screening library comprised of guide RNAs tiling across the region is transduced into our cassette-targeted cells. Successful deletion resulting from the action of two guides will lead to loss of GFP expression; within thisGFP-negative subset of cells, the desired readout will be increased IL7R expression. Next-generation sequencing and bioinformatic analysis will aid in the identification of guides enriched in IL7R^{high} vs IL7R^{low} populations. A minimally deleted region underlying IL7R^{high} expression will be nominated by the location of the guides.

Materials and Methods

Generating isogenic cell lines with deletions.

The human chronic lymphocytic leukemia cell line MEC1 was cultured in RPMI with 10% FBS and 1% PS. To generate the *miR-/-* and *CBS-/-* deletions, pre-assembled Cas9-RNP complexes containing the targeting crRNA pair were co-nucleofected into cells suspended in Solution SF (Lonza Bioscience) with the 4D-Nucleofector with X Unit attachment (Lonza Bioscience). The *miR-CBS-/-* deletion was generated by targeting the CBS in *miR-/-* cells. To generate the larger deletions, pFB2-E (a kind gift from Dr. Francisco Barriga) was PCR amplified with primers containing 40 bp homology arms and co-transfected in tandem with Cas9-RNPs using the same nucleofection condition as above to generate the knock-in cells. Cells containing the knock-in cassette were selected with 2 μg/ml puromycin. Targeting Cas9-RNPs were then used to generate the *LDEC-/-* and *LDIC-/-* deletions in knock-in cells; cells harboring the deletions were counter-selected with 3 μg/ml ganciclovir. Single cell lines were established using the limiting dilution method in 96 well plates. PCR genotyping analysis confirmed the presence of the deletions and Sanger sequencing used to verify the junctional breakpoints.

RT-qPCR for miRNA expression analysis.

Total RNA was extracted from cells using TRIZOL according to the manufacturer's instructions. cDNA was synthesized from 2 μg of total RNA using the TaqMan

Advanced miRNA cDNA Synthesis Kit (Applied Biosystems). TaqMan Advanced miRNA assays were used to detect *miR-15a* and *miR-16-1* expression levels. *miR-17* was used as an internal control. Expression differences were calculated as normalized fold change to WT using the delta-delta Ct method.

Flow cytometry analyses.

To assess *IL7R* expression levels, cells were incubated with a fluorophore conjugated mouse antibody against human CD127 (Invitrogen, #67-1278-42). To assess transduction efficiency, cells were co-incubated with a fluorophore conjugated mouse antibody against human CD271 (Miltenyi Biotec, #130-113-421). Both antibodies were incubated with cells for 1 h at 4°C followed by two rounds of washes before flow cytometry analysis.

Rapid amplification of cDNA ends (RACE)-PCR.

The full-length sequences of *DLEU2* transcript isoforms was obtained using the SMARTer RACE 5'/3' Kit (Takara Bio). Total RNA was extracted from *miR-/-* clones using TRIZOL according to manufacturer's instructions. 1 µg of total RNA was used for first-strand cDNA synthesis to make 5'- and 3'-RACE-Ready cDNA. To generate 5' and 3' cDNA fragments, gene-specific primers were used in the RACE PCR reactions. For the 5' cDNA fragment, the GSP was 5'- ACAGGTCAAAACCGACTGCG-3'. For the 3' cDNA fragment, the GSP was 5'- GTAGCAGAGAACCAATTCTG-3'. The fragments were gel-purified and cloned

into linearized pRACE vector for sequencing. The full-length cDNA was generated from the 5'-RACE-Ready cDNA using 5' GGCGGGGTTGGCTCTAACGA-3' and 5'- GTGTACCTCTCTATATATAA-3' for *DLEU2A* and 5'-GGCGGGGTTGGCTCTAACGA-3' and 5'-CTTTGAAATATCTTAAATTTATTC-3' for *DLEU2B*.

Lentiviral transduction.

cDNA of *DLEU2* isoforms were cloned into the bi-directional lentiviral vector (a kind gift from Dr. Brian Brown, Mount Sinai) using NEBuilder HiFi DNA Assembly (New England BioLabs). The vector was first linearized by XhoI, followed by one-step assembly of minCMV, *DLEU2A/*B, and SV40 poly A into the linearized backbone. Lentiviruses were produced by standard transfection methods. Briefly, 293T cells were co-transfected with the lentiviral vector, pMDLg/pRRE (Addgene #12251), pRSV-Rev (Addgene #12253) and VSV-G envelope expressing plasmid (Addgene #12259) using Lipofectamine 3000 (Thermo Fisher Scientific). Virus-containing supernatants were collected 48 h and 72 h after transfection and filtered through a 0.45 μm membrane. Cells were infected in the presence of 8 μg/ml polybrene and "spinfected" by centrifugation for 1 h at RT.

RNA-sequencing and bioinformatic analyses.

Total RNA was extracted from each of the three independent lines representing each of the six genotypes using TRIZOL (Invitrogen) according to manufacturer's

protocols and subjected to DNase I (Qiagen) treatment. After quantification and quality control by Agilent BioAnalyzer, 500 ng of total RNA with RIN values of 7.0-10 underwent polyA selection and TruSeq library preparation according to instructions provided by Illumina (TruSeq Stranded mRNA LT Kit). Samples were barcoded and run on a HiSeq 4000 in a PE50/50 run. Reads were aligned to the human genome (hg38) using STAR (Dobin et al. 2013). Differential gene expression analysis was performed using the DESeq2 package (Love, Huber, and Anders 2014). Gene ontology over-representation analysis was performed using the clusterProfiler package (Yu et al. 2012). All bioinformatic analyses were carried out on RStudio unless otherwise indicated.

6 References

Abbas, T., M. A. Keaton, and A. Dutta. 2013. 'Genomic instability in cancer', *Cold Spring Harb Perspect Biol*, 5: a012914.

Abu El Maaty, M. A., W. Strassburger, T. Qaiser, Y. Dabiri, and S. Wolfl. 2017. 'Differences in p53 status significantly influence the cellular response and cell survival to 1,25-dihydroxyvitamin D3-metformin cotreatment in colorectal cancer cells', *Mol Carcinog*, 56: 2486-98.

Adams, J. M., A. W. Harris, C. A. Pinkert, L. M. Corcoran, W. S. Alexander, S. Cory, R. D. Palmiter, and R. L. Brinster. 1985. 'The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice', *Nature*, 318: 533-8.

Albertson, D. G. 2006. 'Gene amplification in cancer', *Trends Genet*, 22: 447-55.

Ambros, I. M., S. Rumpler, A. Luegmayr, C. M. Hattinger, S. Strehl, H. Kovar, H. Gadner, and P. F. Ambros. 1997. 'Neuroblastoma cells can actively eliminate supernumerary MYCN gene copies by micronucleus formation--sign of tumour cell revertance?', *Eur J Cancer*, 33: 2043-9.

An, Z., O. Aksoy, T. Zheng, Q. W. Fan, and W. A. Weiss. 2018. 'Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies', *Oncogene*, 37: 1561-75.

Anderson, J. D., W. A. Gillespie, and M. H. Richmond. 1973. 'Chemotherapy and antibiotic-resistance transfer between Enterobacteria in the human gastro-intestinal tract', *J Med Microbiol*, 6: 461-73.

Annunziato, S., S. M. Kas, M. Nethe, H. Yucel, J. Del Bravo, C. Pritchard, R. Bin Ali, B. van Gerwen, B. Siteur, A. P. Drenth, E. Schut, M. van de Ven, M. C. Boelens, S. Klarenbeek, I. J. Huijbers, M. H. van Miltenburg, and J. Jonkers. 2016. 'Modeling invasive lobular breast carcinoma by CRISPR/Cas9-mediated somatic genome editing of the mammary gland', *Genes Dev*, 30: 1470-80.

Aplan, P. D. 2006. 'Causes of oncogenic chromosomal translocation', *Trends Genet*, 22: 46-55.

Baca, S. C., D. Prandi, M. S. Lawrence, J. M. Mosquera, A. Romanel, Y. Drier, K. Park, N. Kitabayashi, T. Y. MacDonald, M. Ghandi, E. Van Allen, G. V. Kryukov, A. Sboner, J.

P. Theurillat, T. D. Soong, E. Nickerson, D. Auclair, A. Tewari, H. Beltran, R. C. Onofrio, G. Boysen, C. Guiducci, C. E. Barbieri, K. Cibulskis, A. Sivachenko, S. L. Carter, G. Saksena, D. Voet, A. H. Ramos, W. Winckler, M. Cipicchio, K. Ardlie, P. W. Kantoff, M. F. Berger, S. B. Gabriel, T. R. Golub, M. Meyerson, E. S. Lander, O. Elemento, G. Getz, F. Demichelis, M. A. Rubin, and L. A. Garraway. 2013. 'Punctuated evolution of prostate cancer genomes', *Cell*, 153: 666-77.

Bailey, C., M. J. Shoura, P. S. Mischel, and C. Swanton. 2020. 'Extrachromosomal DNA-relieving heredity constraints, accelerating tumour evolution', *Annals of Oncology*, 31: 884-93.

Barlow, J. H., R. B. Faryabi, E. Callen, N. Wong, A. Malhowski, H. T. Chen, G. Gutierrez-Cruz, H. W. Sun, P. McKinnon, G. Wright, R. Casellas, D. F. Robbiani, L. Staudt, O. Fernandez-Capetillo, and A. Nussenzweig. 2013. 'Identification of early replicating fragile sites that contribute to genome instability', *Cell*, 152: 620-32.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. 2007. 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science*, 315: 1709-12.

Bartkova, J., N. Rezaei, M. Liontos, P. Karakaidos, D. Kletsas, N. Issaeva, L. V. Vassiliou, E. Kolettas, K. Niforou, V. C. Zoumpourlis, M. Takaoka, H. Nakagawa, F. Tort, K. Fugger, F. Johansson, M. Sehested, C. L. Andersen, L. Dyrskjot, T. Orntoft, J. Lukas, C. Kittas, T. Helleday, T. D. Halazonetis, J. Bartek, and V. G. Gorgoulis. 2006. 'Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints', *Nature*, 444: 633-7.

Beroukhim, R., C. H. Mermel, D. Porter, G. Wei, S. Raychaudhuri, J. Donovan, J. Barretina, J. S. Boehm, J. Dobson, M. Urashima, K. T. Mc Henry, R. M. Pinchback, A. H. Ligon, Y. J. Cho, L. Haery, H. Greulich, M. Reich, W. Winckler, M. S. Lawrence, B. A. Weir, K. E. Tanaka, D. Y. Chiang, A. J. Bass, A. Loo, C. Hoffman, J. Prensner, T. Liefeld, Q. Gao, D. Yecies, S. Signoretti, E. Maher, F. J. Kaye, H. Sasaki, J. E. Tepper, J. A. Fletcher, J. Tabernero, J. Baselga, M. S. Tsao, F. Demichelis, M. A. Rubin, P. A. Janne, M. J. Daly, C. Nucera, R. L. Levine, B. L. Ebert, S. Gabriel, A. K. Rustgi, C. R. Antonescu, M. Ladanyi, A. Letai, L. A. Garraway, M. Loda, D. G. Beer, L. D. True, A. Okamoto, S. L. Pomeroy, S. Singer, T. R. Golub, E. S. Lander, G. Getz, W. R. Sellers, and M. Meyerson. 2010. 'The landscape of somatic copy-number alteration across human cancers', *Nature*, 463: 899-905.

Bizard, A. H., and I. D. Hickson. 2014. 'The dissolution of double Holliday junctions', *Cold Spring Harb Perspect Biol*, 6: a016477.

Blasco, R. B., E. Karaca, C. Ambrogio, T. C. Cheong, E. Karayol, V. G. Minero, C. Voena, and R. Chiarle. 2014. 'Simple and rapid in vivo generation of chromosomal rearrangements using CRISPR/Cas9 technology', *Cell Rep*, 9: 1219-27.

Bohlander, Stefan K., Purvi M. Kakadiya, and Alix Coysh. 2019. 'Chromosome Rearrangements and Translocations.' in Paolo Boffetta and Pierre Hainaut (eds.), *Encyclopedia of Cancer (Third Edition)* (Academic Press: Oxford).

Brennan, C. W., R. G. Verhaak, A. McKenna, B. Campos, H. Noushmehr, S. R. Salama, S. Zheng, D. Chakravarty, J. Z. Sanborn, S. H. Berman, R. Beroukhim, B. Bernard, C. J. Wu, G. Genovese, I. Shmulevich, J. Barnholtz-Sloan, L. Zou, R. Vegesna, S. A. Shukla, G. Ciriello, W. K. Yung, W. Zhang, C. Sougnez, T. Mikkelsen, K. Aldape, D. D. Bigner, E. G. Van Meir, M. Prados, A. Sloan, K. L. Black, J. Eschbacher, G. Finocchiaro, W. Friedman, D. W. Andrews, A. Guha, M. Iacocca, B. P. O'Neill, G. Foltz, J. Myers, D. J. Weisenberger, R. Penny, R. Kucherlapati, C. M. Perou, D. N. Hayes, R. Gibbs, M. Marra, G. B. Mills, E. Lander, P. Spellman, R. Wilson, C. Sander, J. Weinstein, M. Meyerson, S. Gabriel, P. W. Laird, D. Haussler, G. Getz, L. Chin, and Tcga Research Network. 2013. 'The somatic genomic landscape of glioblastoma', *Cell*, 155: 462-77.

Brunet, E., D. Simsek, M. Tomishima, R. DeKelver, V. M. Choi, P. Gregory, F. Urnov, D. M. Weinstock, and M. Jasin. 2009. 'Chromosomal translocations induced at specified loci in human stem cells', *Proc Natl Acad Sci U S A*, 106: 10620-5.

Bunz, F., A. Dutriaux, C. Lengauer, T. Waldman, S. Zhou, J. P. Brown, J. M. Sedivy, K. W. Kinzler, and B. Vogelstein. 1998. 'Requirement for p53 and p21 to sustain G2 arrest after DNA damage', *Science*, 282: 1497-501.

Buonamici, Silvia, Thomas Trimarchi, Maria Grazia Ruocco, Linsey Reavie, Severine Cathelin, Brenton G. Mar, Apostolos Klinakis, Yevgeniy Lukyanov, Jen-Chieh Tseng, Filiz Sen, Eric Gehrie, Mengling Li, Elizabeth Newcomb, Jiri Zavadil, Daniel Meruelo, Martin Lipp, Sherif Ibrahim, Argiris Efstratiadis, David Zagzag, Jonathan S. Bromberg, Michael L. Dustin, and Iannis Aifantis. 2009. 'CCR7 signalling as an essential regulator of CNS infiltration in T-cell leukaemia', 459: 1000-04.

Calin, G. A., C. D. Dumitru, M. Shimizu, R. Bichi, S. Zupo, E. Noch, H. Aldler, S. Rattan, M. Keating, K. Rai, L. Rassenti, T. Kipps, M. Negrini, F. Bullrich, and C. M. Croce. 2002. 'Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia', *Proc Natl Acad Sci U S A*, 99: 15524-9.

Campbell, P. J., P. J. Stephens, E. D. Pleasance, S. O'Meara, H. Li, T. Santarius, L. A. Stebbings, C. Leroy, S. Edkins, C. Hardy, J. W. Teague, A. Menzies, I. Goodhead, D.

J. Turner, C. M. Clee, M. A. Quail, A. Cox, C. Brown, R. Durbin, M. E. Hurles, P. A. Edwards, G. R. Bignell, M. R. Stratton, and P. A. Futreal. 2008. 'Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing', *Nat Genet*, 40: 722-9.

Cancer Genome Atlas Research, Network. 2011. 'Integrated genomic analyses of ovarian carcinoma', *Nature*, 474: 609-15.

Capecchi, M. R. 1989. 'Altering the genome by homologous recombination', *Science*, 244: 1288-92.

Capper, R., B. Britt-Compton, M. Tankimanova, J. Rowson, B. Letsolo, S. Man, M. Haughton, and D. M. Baird. 2007. 'The nature of telomere fusion and a definition of the critical telomere length in human cells', *Genes Dev*, 21: 2495-508.

Carter, P., L. Presta, C. M. Gorman, J. B. Ridgway, D. Henner, W. L. Wong, A. M. Rowland, C. Kotts, M. E. Carver, and H. M. Shepard. 1992. 'Humanization of an anti-p185HER2 antibody for human cancer therapy', *Proc Natl Acad Sci U S A*, 89: 4285-9.

Castagliuolo, I., E. Beggiao, P. Brun, L. Barzon, S. Goussard, R. Manganelli, C. Grillot-Courvalin, and G. Palu. 2005. 'Engineered E. coli delivers therapeutic genes to the colonic mucosa', *Gene Ther*, 12: 1070-8.

Ceccaldi, R., J. C. Liu, R. Amunugama, I. Hajdu, B. Primack, M. I. Petalcorin, K. W. O'Connor, P. A. Konstantinopoulos, S. J. Elledge, S. J. Boulton, T. Yusufzai, and A. D. D'Andrea. 2015. 'Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair', *Nature*, 518: 258-62.

Chang, H. H. Y., N. R. Pannunzio, N. Adachi, and M. R. Lieber. 2017. 'Non-homologous DNA end joining and alternative pathways to double-strand break repair', *Nat Rev Mol Cell Biol*, 18: 495-506.

Charlesworth, C. T., P. S. Deshpande, D. P. Dever, J. Camarena, V. T. Lemgart, M. K. Cromer, C. A. Vakulskas, M. A. Collingwood, L. Zhang, N. M. Bode, M. A. Behlke, B. Dejene, B. Cieniewicz, R. Romano, B. J. Lesch, N. Gomez-Ospina, S. Mantri, M. Pavel-Dinu, K. I. Weinberg, and M. H. Porteus. 2019. 'Identification of preexisting adaptive immunity to Cas9 proteins in humans', *Nat Med*, 25: 249-54.

Chen, C., Y. Liu, A. R. Rappaport, T. Kitzing, N. Schultz, Z. Zhao, A. S. Shroff, R. A. Dickins, C. R. Vakoc, J. E. Bradner, W. Stock, M. M. LeBeau, K. M. Shannon, S. Kogan, J.

Zuber, and S. W. Lowe. 2014. 'MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia', *Cancer Cell*, 25: 652-65.

Chen, H., M. Lisby, and L. S. Symington. 2013. 'RPA coordinates DNA end resection and prevents formation of DNA hairpins', *Mol Cell*, 50: 589-600.

Cheng, Q., T. Wei, L. Farbiak, L. T. Johnson, S. A. Dilliard, and D. J. Siegwart. 2020. 'Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR-Cas gene editing', *Nat Nanotechnol*, 15: 313-20.

Cheng, R., J. Peng, Y. Yan, P. Cao, J. Wang, C. Qiu, L. Tang, D. Liu, L. Tang, J. Jin, X. Huang, F. He, and P. Zhang. 2014. 'Efficient gene editing in adult mouse livers via adenoviral delivery of CRISPR/Cas9', *FEBS Lett*, 588: 3954-8.

Chiou, S. H., I. P. Winters, J. Wang, S. Naranjo, C. Dudgeon, F. B. Tamburini, J. J. Brady, D. Yang, B. M. Gruner, C. H. Chuang, D. R. Caswell, H. Zeng, P. Chu, G. E. Kim, D. R. Carpizo, S. K. Kim, and M. M. Winslow. 2015. 'Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing', *Genes Dev*, 29: 1576-85.

Chu, W. K., and I. D. Hickson. 2009. 'RecQ helicases: multifunctional genome caretakers', *Nat Rev Cancer*, 9: 644-54.

Cimmino, Amelia, George Adrian Calin, Muller Fabbri, Marilena V. Iorio, Manuela Ferracin, Masayoshi Shimizu, Sylwia E. Wojcik, Rami I. Aqeilan, Simona Zupo, Mariella Dono, Laura Rassenti, Hansjuerg Alder, Stefano Volinia, C. g Liu, Thomas J. Kipps, Massimo Negrini, and Carlo M. Croce. 2005. 'miR-15 and miR-16 induce apoptosis by targeting BCL2', *Proceedings of the National Academy of Sciences*, 102: 13944-49.

Conway, T., K. A. Krogfelt, and P. S. Cohen. 2004. 'The Life of Commensal Escherichia coli in the Mammalian Intestine', *EcoSal Plus*, 1.

Cook, P. J., R. Thomas, R. Kannan, E. S. de Leon, A. Drilon, M. K. Rosenblum, M. Scaltriti, R. Benezra, and A. Ventura. 2017. 'Somatic chromosomal engineering identifies BCAN-NTRK1 as a potent glioma driver and therapeutic target', *Nat Commun*, 8: 15987.

Cortes-Ciriano, I., J. J. Lee, R. Xi, D. Jain, Y. L. Jung, L. Yang, D. Gordenin, L. J. Klimczak, C. Z. Zhang, D. S. Pellman, Pcawg Structural Variation Working Group, P. J. Park, and Pcawg Consortium. 2020. 'Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing', *Nat Genet*, 52: 331-41.

Cox, D., C. Yuncken, and A. I. Spriggs. 1965. 'Minute Chromatin Bodies in Malignant Tumours of Childhood', *Lancet*, 1: 55-8.

Crasta, K., N. J. Ganem, R. Dagher, A. B. Lantermann, E. V. Ivanova, Y. Pan, L. Nezi, A. Protopopov, D. Chowdhury, and D. Pellman. 2012. 'DNA breaks and chromosome pulverization from errors in mitosis', *Nature*, 482: 53-8.

Croce, C. M. 2008. 'Oncogenes and cancer', *N Engl J Med*, 358: 502-11.

Crockford, A., L. P. Zalmas, E. Gronroos, S. M. Dewhurst, N. McGranahan, M. E. Cuomo, V. Encheva, A. P. Snijders, J. Begum, S. Purewal, J. Cerveira, H. Patel, M. J. Renshaw, and C. Swanton. 2017. 'Cyclin D mediates tolerance of genome-doubling in cancers with functional p53', *Annals of Oncology*, 28: 149-56.

Dalla-Favera, R., M. Bregni, J. Erikson, D. Patterson, R. C. Gallo, and C. M. Croce. 1982. 'Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells', *Proc Natl Acad Sci U S A*, 79: 7824-7.

deCarvalho, A. C., H. Kim, L. M. Poisson, M. E. Winn, C. Mueller, D. Cherba, J. Koeman, S. Seth, A. Protopopov, M. Felicella, S. Zheng, A. Multani, Y. Jiang, J. Zhang, D. H. Nam, E. F. Petricoin, L. Chin, T. Mikkelsen, and R. G. W. Verhaak. 2018. 'Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma', *Nat Genet*, 50: 708-17.

Dewhurst, S. M., N. McGranahan, R. A. Burrell, A. J. Rowan, E. Gronroos, D. Endesfelder, T. Joshi, D. Mouradov, P. Gibbs, R. L. Ward, N. J. Hawkins, Z. Szallasi, O. M. Sieber, and C. Swanton. 2014. 'Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution', *Cancer Discov*, 4: 175-85.

Difilippantonio, M. J., J. Zhu, H. T. Chen, E. Meffre, M. C. Nussenzweig, E. E. Max, T. Ried, and A. Nussenzweig. 2000. 'DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation', *Nature*, 404: 510-4.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, 29: 15-21.

Elliott, B., C. Richardson, and M. Jasin. 2005. 'Chromosomal translocation mechanisms at intronic alu elements in mammalian cells', *Mol Cell*, 17: 885-94.

Flavahan, W. A., Y. Drier, B. B. Liau, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suva, and B. E. Bernstein. 2016. 'Insulator dysfunction and oncogene activation in IDH mutant gliomas', *Nature*, 529: 110-4.

Flavahan, William A., Yotam Drier, Sarah E. Johnstone, Matthew L. Hemming, Daniel R. Tarjan, Esmat Hegazi, Sarah J. Shareef, Nauman M. Javed, Chandrajit P. Raut, Benjamin K. Eschle, Prafulla C. Gokhale, Jason L. Hornick, Ewa T. Sicinska, George D. Demetri, and Bradley E. Bernstein. 2019. 'Altered chromosomal topology drives oncogenic programs in SDH-deficient GIST', *Nature*.

Folger, K. R., E. A. Wong, G. Wahl, and M. R. Capecchi. 1982. 'Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules', *Mol Cell Biol*, 2: 1372-87.

Freter, R., H. Brickner, M. Botney, D. Cleven, and A. Aranki. 1983. 'Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora', *Infect Immun*, 39: 676-85.

Fukumoto, M., D. H. Shevrin, and I. B. Roninson. 1988. 'Analysis of gene amplification in human tumor cell lines', *Proc Natl Acad Sci U S A*, 85: 6846-50.

Ghezraoui, H., M. Piganeau, B. Renouf, J. B. Renaud, A. Sallmyr, B. Ruis, S. Oh, A. E. Tomkinson, E. A. Hendrickson, C. Giovannangeli, M. Jasin, and E. Brunet. 2014. 'Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining', *Mol Cell*, 55: 829-42.

Ghirlando, R., and G. Felsenfeld. 2016. 'CTCF: making the right connections', *Genes Dev*, 30: 881-91.

Ghosh, M., S. Saha, J. Bettke, R. Nagar, A. Parrales, T. Iwakuma, A. W. M. van der Velden, and L. A. Martinez. 2021. 'Mutant p53 suppresses innate immune signaling to promote tumorigenesis', *Cancer Cell*, 39: 494-508 e5.

Gluck, S., B. Guey, M. F. Gulen, K. Wolter, T. W. Kang, N. A. Schmacke, A. Bridgeman, J. Rehwinkel, L. Zender, and A. Ablasser. 2017. 'Innate immune sensing of cytosolic chromatin fragments through cGAS promotes senescence', *Nat Cell Biol*, 19: 1061-70.

Grabarz, A., A. Barascu, J. Guirouilh-Barbat, and B. S. Lopez. 2012. 'Initiation of DNA double strand break repair: signaling and single-stranded resection dictate the

choice between homologous recombination, non-homologous end-joining and alternative end-joining', *Am J Cancer Res*, 2: 249-68.

Grillot-Courvalin, C., S. Goussard, F. Huetz, D. M. Ojcius, and P. Courvalin. 1998. 'Functional gene transfer from intracellular bacteria to mammalian cells', *Nat Biotechnol*, 16: 862-6.

Grisendi, S., C. Mecucci, B. Falini, and P. P. Pandolfi. 2006. 'Nucleophosmin and cancer', *Nat Rev Cancer*, 6: 493-505.

Hagstrom, S. A., and T. P. Dryja. 1999. 'Mitotic recombination map of 13cen-13q14 derived from an investigation of loss of heterozygosity in retinoblastomas', *Proc Natl Acad Sci U S A*, 96: 2952-7.

Halazonetis, T. D., V. G. Gorgoulis, and J. Bartek. 2008. 'An oncogene-induced DNA damage model for cancer development', *Science*, 319: 1352-5.

Han, L., and K. Yu. 2008. 'Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV-deficient B cells', *J Exp Med*, 205: 2745-53.

Han, T., E. M. Schatoff, C. Murphy, M. P. Zafra, J. E. Wilkinson, O. Elemento, and L. E. Dow. 2017. 'R-Spondin chromosome rearrangements drive Wnt-dependent tumour initiation and maintenance in the intestine', *Nat Commun*, 8: 15945.

Hanahan, D., and R. A. Weinberg. 2011. 'Hallmarks of cancer: the next generation', *Cell*, 144: 646-74.

Harding, Shane M., Joseph L. Benci, Jerome Irianto, Dennis E. Discher, Andy J. Minn, and Roger A. Greenberg. 2017. 'Mitotic progression following DNA damage enables pattern recognition within micronuclei', *Nature*, 548: 466-70.

Hashimoto, M., Y. Yamashita, and T. Takemoto. 2016. 'Electroporation of Cas9 protein/sgRNA into early pronuclear zygotes generates non-mosaic mutants in the mouse', *Dev Biol*, 418: 1-9.

Hastings, P. J., G. Ira, and J. R. Lupski. 2009. 'A microhomology-mediated break-induced replication model for the origin of human copy number variation', *PLoS Genet*, 5: e1000327.

Hatch, E. M., A. H. Fischer, T. J. Deerinck, and M. W. Hetzer. 2013. 'Catastrophic nuclear envelope collapse in cancer cell micronuclei', *Cell*, 154: 47-60.

Heckl, D., M. S. Kowalczyk, D. Yudovich, R. Belizaire, R. V. Puram, M. E. McConkey, A. Thielke, J. C. Aster, A. Regev, and B. L. Ebert. 2014. 'Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing', *Nat Biotechnol*, 32: 941-6.

Heisterkamp, N., G. Jenster, D. Kioussis, P. K. Pattengale, and J. Groffen. 1991. 'Human bcr-abl gene has a lethal effect on embryogenesis', *Transgenic Res*, 1: 45-53.

Hellman, A., E. Zlotorynski, S. W. Scherer, J. Cheung, J. B. Vincent, D. I. Smith, L. Trakhtenbrot, and B. Kerem. 2002. 'A role for common fragile site induction in amplification of human oncogenes', *Cancer Cell*, 1: 89-97.

Helmrich, A., M. Ballarino, and L. Tora. 2011. 'Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes', *Mol Cell*, 44: 966-77.

Hentges, DJ, JU Que, SW Casey, and AJ Stein. 1984. 'The influence of streptomycin on colonization resistance in mice', *Microecol. Ther*, 14: 53-62.

Hnisz, D., A. S. Weintraub, D. S. Day, A. L. Valton, R. O. Bak, C. H. Li, J. Goldmann, B. R. Lajoie, Z. P. Fan, A. A. Sigova, J. Reddy, D. Borges-Rivera, T. I. Lee, R. Jaenisch, M. H. Porteus, J. Dekker, and R. A. Young. 2016. 'Activation of proto-oncogenes by disruption of chromosome neighborhoods', *Science*, 351: 1454-58.

Holland, A. J., and D. W. Cleveland. 2009. 'Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis', *Nat Rev Mol Cell Biol*, 10: 478-87.

———. 2012. 'Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements', *Nat Med*, 18: 1630-8.

Hung, King L., Kathryn E. Yost, Liangqi Xie, Sihan Wu, Joshua T. Lange, Connor V. Duffy, Katerina Kraft, Jun Tang, Quanming Shi, John C. Rose, M. Ryan Corces, Jeffrey M. Granja, Rui Li, Utkrisht Rajkumar, Robert Tjian, Vineet Bafna, Paul S. Mischel, Zhe Liu, and Howard Y. Chang. 2020. 'EcDNA hubs drive cooperative intermolecular oncogene expression'.

Hussain, S. P., J. Schwank, F. Staib, X. W. Wang, and C. C. Harris. 2007. 'TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer', *Oncogene*, 26: 2166-76.

Hustedt, N., and D. Durocher. 2016. 'The control of DNA repair by the cell cycle', *Nat Cell Biol*, 19: 1-9.

Ji, W., Z. Bian, Y. Yu, C. Yuan, Y. Liu, L. Yu, C. Li, J. Zhu, X. Jia, R. Guan, C. Zhang, X. Meng, Y. Jin, J. Bai, J. Yu, K. Y. Lee, W. Sun, and S. Fu. 2014. 'Expulsion of micronuclei containing amplified genes contributes to a decrease in double minute chromosomes from malignant tumor cells', *Int J Cancer*, 134: 1279-88.

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

Johansson, M. E., and G. C. Hansson. 2016. 'Immunological aspects of intestinal mucus and mucins', *Nat Rev Immunol*, 16: 639-49.

Jones, D. T., S. Kocialkowski, L. Liu, D. M. Pearson, L. M. Backlund, K. Ichimura, and V. P. Collins. 2008. 'Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas', *Cancer Research*, 68: 8673-7.

Kaeser, M. D., S. Pebernard, and R. D. Iggo. 2004. 'Regulation of p53 stability and function in HCT116 colon cancer cells', *J Biol Chem*, 279: 7598-605.

Kanda, T., M. Otter, and G. M. Wahl. 2001. 'Mitotic segregation of viral and cellular acentric extrachromosomal molecules by chromosome tethering', *J Cell Sci*, 114: 49-58.

Kawate, S., T. Fukusato, S. Ohwada, A. Watanuki, and Y. Morishita. 1999. 'Amplification of c-myc in hepatocellular carcinoma: correlation with clinicopathologic features, proliferative activity and p53 overexpression', *Oncology*, 57: 157-63.

Kim, H., N. P. Nguyen, K. Turner, S. Wu, A. D. Gujar, J. Luebeck, J. Liu, V. Deshpande, U. Rajkumar, S. Namburi, S. B. Amin, E. Yi, F. Menghi, J. H. Schulte, A. G. Henssen, H. Y. Chang, C. R. Beck, P. S. Mischel, V. Bafna, and R. G. W. Verhaak. 2020. 'Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers', *Nat Genet*, 52: 891-97.

Klein, U., M. Lia, M. Crespo, R. Siegel, Q. Shen, T. Mo, A. Ambesi-Impiombato, A. Califano, A. Migliazza, G. Bhagat, and R. Dalla-Favera. 2010. 'The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia', *Cancer Cell*, 17: 28-40.

Kloosterman, W. P., M. Hoogstraat, O. Paling, M. Tavakoli-Yaraki, I. Renkens, J. S. Vermaat, M. J. van Roosmalen, S. van Lieshout, I. J. Nijman, W. Roessingh, R. van 't Slot, J. van de Belt, V. Guryev, M. Koudijs, E. Voest, and E. Cuppen. 2011. 'Chromothripsis

is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer', *Genome Biol*, 12: R103.

Knudson, A. G., Jr. 1971. 'Mutation and cancer: statistical study of retinoblastoma', *Proc Natl Acad Sci U S A*, 68: 820-3.

Kohl, N. E., N. Kanda, R. R. Schreck, G. Bruns, S. A. Latt, F. Gilbert, and F. W. Alt. 1983. 'Transposition and amplification of oncogene-related sequences in human neuroblastomas', *Cell*, 35: 359-67.

Krupina, K., A. Goginashvili, and D. W. Cleveland. 2021. 'Causes and consequences of micronuclei', *Curr Opin Cell Biol*, 70: 91-99.

Kuppers, R. 2005. 'Mechanisms of B-cell lymphoma pathogenesis', *Nat Rev Cancer*, 5: 251-62.

Kurzrock, R., J. U. Gutterman, and M. Talpaz. 1988. 'The molecular genetics of Philadelphia chromosome-positive leukemias', *N Engl J Med*, 319: 990-8.

Kwon, J., and S. F. Bakhoum. 2020. 'The Cytosolic DNA-Sensing cGAS-STING Pathway in Cancer', *Cancer Discov*, 10: 26-39.

Landau, Dan A., Eugen Tausch, Amaro N. Taylor-Weiner, Chip Stewart, Johannes G. Reiter, Jasmin Bahlo, Sandra Kluth, Ivana Bozic, Mike Lawrence, Sebastian Böttcher, Scott L. Carter, Kristian Cibulskis, Daniel Mertens, Carrie L. Sougnez, Mara Rosenberg, Julian M. Hess, Jennifer Edelmann, Sabrina Kless, Michael Kneba, Matthias Ritgen, Anna Fink, Kirsten Fischer, Stacey Gabriel, Eric S. Lander, Martin A. Nowak, Hartmut Döhner, Michael Hallek, Donna Neuberg, Gad Getz, Stephan Stilgenbauer, and Catherine J. Wu. 2015. 'Mutations driving CLL and their evolution in progression and relapse', *Nature*, 526: 525-30.

Lange, Joshua T., Celine Y. Chen, Yuriy Pichugin, Liangqi Xie, Jun Tang, King L. Hung, Kathryn E. Yost, Quanming Shi, Marcella L. Erb, Utkrisht Rajkumar, Sihan Wu, Charles Swanton, Zhe Liu, Weini Huang, Howard Y. Chang, Vineet Bafna, Anton G. Henssen, Benjamin Werner, and Paul S. Mischel. 2021. 'Principles of ecDNA random inheritance drive rapid genome change and therapy resistance in human cancers', *bioRxiv*: 2021.06.11.447968.

Lee, J. A., C. M. Carvalho, and J. R. Lupski. 2007. 'A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders', *Cell*, 131: 1235-47.

Lengauer, C., K. W. Kinzler, and B. Vogelstein. 1997. 'Genetic instability in colorectal cancers', *Nature*, 386: 623-7.

Li, Y., N. D. Roberts, J. A. Wala, O. Shapira, S. E. Schumacher, K. Kumar, E. Khurana, S. Waszak, J. O. Korbel, J. E. Haber, M. Imielinski, Pcawg Structural Variation Working Group, J. Weischenfeldt, R. Beroukhim, P. J. Campbell, and Pcawg Consortium. 2020. 'Patterns of somatic structural variation in human cancer genomes', *Nature*, 578: 112-21.

Lieber, M. R. 2010. 'The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway', *Annu Rev Biochem*, 79: 181-211.

Lieber, M. R., Y. Ma, U. Pannicke, and K. Schwarz. 2003. 'Mechanism and regulation of human non-homologous DNA end-joining', *Nat Rev Mol Cell Biol*, 4: 712-20.

Linn, D. E., K. L. Penney, R. T. Bronson, L. A. Mucci, and Z. Li. 2016. 'Deletion of Interstitial Genes between TMPRSS2 and ERG Promotes Prostate Cancer Progression', *Cancer Research*, 76: 1869-81.

Liu, P., A. Erez, S. C. Nagamani, S. U. Dhar, K. E. Kolodziejska, A. V. Dharmadhikari, M. L. Cooper, J. Wiszniewska, F. Zhang, M. A. Withers, C. A. Bacino, L. D. Campos-Acevedo, M. R. Delgado, D. Freedenberg, A. Garnica, T. A. Grebe, D. Hernandez-Almaguer, L. Immken, S. R. Lalani, S. D. McLean, H. Northrup, F. Scaglia, L. Strathearn, P. Trapane, S. H. Kang, A. Patel, S. W. Cheung, P. J. Hastings, P. Stankiewicz, J. R. Lupski, and W. Bi. 2011. 'Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements', *Cell*, 146: 889-903.

Livak, K. J., and T. D. Schmittgen. 2001. 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods*, 25: 402-8.

Livingstone, L. R., A. White, J. Sprouse, E. Livanos, T. Jacks, and T. D. Tlsty. 1992. 'Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53', *Cell*, 70: 923-35.

Lo, A. W., L. Sabatier, B. Fouladi, G. Pottier, M. Ricoul, and J. P. Murnane. 2002. 'DNA amplification by breakage/fusion/bridge cycles initiated by spontaneous telomere loss in a human cancer cell line', *Neoplasia*, 4: 531-8.

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

Lupianez, D. G., K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, and S. Mundlos. 2015. 'Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions', *Cell*, 161: 1012-25.

Ly, P., L. S. Teitz, D. H. Kim, O. Shoshani, H. Skaletsky, D. Fachinetti, D. C. Page, and D. W. Cleveland. 2017. 'Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining', *Nat Cell Biol*, 19: 68-75.

Mackenzie, K. J., P. Carroll, C. A. Martin, O. Murina, A. Fluteau, D. J. Simpson, N. Olova, H. Sutcliffe, J. K. Rainger, A. Leitch, R. T. Osborn, A. P. Wheeler, M. Nowotny, N. Gilbert, T. Chandra, M. A. M. Reijns, and A. P. Jackson. 2017. 'cGAS surveillance of micronuclei links genome instability to innate immunity', *Nature*, 548: 461-65.

Maddalo, D., E. Manchado, C. P. Concepcion, C. Bonetti, J. A. Vidigal, Y. C. Han, P. Ogrodowski, A. Crippa, N. Rekhtman, E. de Stanchina, S. W. Lowe, and A. Ventura. 2014. 'In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system', *Nature*, 516: 423-7.

Mahowald, G. K., J. M. Baron, and B. P. Sleckman. 2008. 'Collateral damage from antigen receptor gene diversification', *Cell*, 135: 1009-12.

Malhotra, A., M. Lindberg, G. G. Faust, M. L. Leibowitz, R. A. Clark, R. M. Layer, A. R. Quinlan, and I. M. Hall. 2013. 'Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms', *Genome Res*, 23: 762-76.

Malu, S., V. Malshetty, D. Francis, and P. Cortes. 2012. 'Role of non-homologous end joining in V(D)J recombination', *Immunol Res*, 54: 233-46.

Maresch, R., S. Mueller, C. Veltkamp, R. Ollinger, M. Friedrich, I. Heid, K. Steiger, J. Weber, T. Engleitner, M. Barenboim, S. Klein, S. Louzada, R. Banerjee, A. Strong, T. Stauber, N. Gross, U. Geumann, S. Lange, M. Ringelhan, I. Varela, K. Unger, F. Yang, R. M. Schmid, G. S. Vassiliou, R. Braren, G. Schneider, M. Heikenwalder, A. Bradley, D. Saur, and R. Rad. 2016. 'Multiplexed pancreatic genome engineering and cancer induction by transfection-based CRISPR/Cas9 delivery in mice', *Nat Commun*, 7: 10770.

Martin, G. M., A. C. Smith, D. J. Ketterer, C. E. Ogburn, and C. M. Disteche. 1985. 'Increased chromosomal aberrations in first metaphases of cells isolated from the kidneys of aged mice', *Isr J Med Sci*, 21: 296-301.

Maser, R. S., and R. A. DePinho. 2002. 'Connecting chromosomes, crisis, and cancer', *Science*, 297: 565-9.

Mateos-Gomez, P. A., F. Gong, N. Nair, K. M. Miller, E. Lazzerini-Denchi, and A. Sfeir. 2015. 'Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination', *Nature*, 518: 254-7.

Matthews, Bryan J., and David J. Waxman. 2018. 'Computational prediction of CTCF/cohesin-based intra-TAD loops that insulate chromatin contacts and gene expression in mouse liver', *eLife*, 7.

McClintock, B. 1941. 'The Stability of Broken Ends of Chromosomes in Zea Mays', *Genetics*, 26: 234-82.

Mertens, Daniel, Stephan Wolf, Petra Schroeter, Claudia Schaffner, Hartmut Döhner, Stephan Stilgenbauer, and Peter Lichter. 2002. 'Down-regulation of candidate tumor suppressor genes within chromosome band 13q14.3 is independent of the DNA methylation pattern in B-cell chronic lymphocytic leukemia', *Blood*, 99: 4116-21.

Michalak, E., A. Villunger, M. Erlacher, and A. Strasser. 2005. 'Death squads enlisted by the tumour suppressor p53', *Biochem Biophys Res Commun*, 331: 786-98.

Migliazza, Anna, F. Bosch, H. Komatsu, E. Cayanis, S. Martinotti, E. Toniato, E. Guccione, X. Qu, M. Chien, V. V. Murty, G. Gaidano, G. Inghirami, P. Zhang, S. Fischer, S. M. Kalachikov, J. Russo, I. Edelman, A. Efstratiadis, and R. Dalla-Favera. 2001. 'Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia', *Blood*, 97: 2098-104.

Miki, Y., J. Swensen, D. Shattuck-Eidens, P. A. Futreal, K. Harshman, S. Tavtigian, Q. Liu, C. Cochran, L. M. Bennett, W. Ding, and et al. 1994. 'A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1', *Science*, 266: 66-71.

Mitelman, F., B. Johansson, and F. Mertens. 2007. 'The impact of translocations and gene fusions on cancer causation', *Nat Rev Cancer*, 7: 233-45.

Moller, H. D., M. Mohiyuddin, I. Prada-Luengo, M. R. Sailani, J. F. Halling, P. Plomgaard, L. Maretty, A. J. Hansen, M. P. Snyder, H. Pilegaard, H. Y. K. Lam, and B. Regenberg. 2018. 'Circular DNA elements of chromosomal origin are common in healthy human somatic tissue', *Nat Commun*, 9: 1069.

Moshous, D., I. Callebaut, R. de Chasseval, B. Corneo, M. Cavazzana-Calvo, F. Le Deist, I. Tezcan, O. Sanal, Y. Bertrand, N. Philippe, A. Fischer, and J. P. de Villartay. 2001. 'Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency', *Cell*, 105: 177-86.

Moynahan, M. E., and M. Jasin. 2010. 'Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis', *Nat Rev Mol Cell Biol*, 11: 196-207.

Nagy, A. 2000. 'Cre recombinase: the universal reagent for genome tailoring', *Genesis*, 26: 99-109.

Nassour, J., R. Radford, A. Correia, J. M. Fuste, B. Schoell, A. Jauch, R. J. Shaw, and J. Karlseder. 2019. 'Autophagic cell death restricts chromosomal instability during replicative crisis', *Nature*, 565: 659-63.

Nathanson, D. A., B. Gini, J. Mottahedeh, K. Visnyei, T. Koga, G. Gomez, A. Eskin, K. Hwang, J. Wang, K. Masui, A. Paucar, H. Yang, M. Ohashi, S. Zhu, J. Wykosky, R. Reed, S. F. Nelson, T. F. Cloughesy, C. D. James, P. N. Rao, H. I. Kornblum, J. R. Heath, W. K. Cavenee, F. B. Furnari, and P. S. Mischel. 2014. 'Targeted Therapy Resistance Mediated by Dynamic Regulation of Extrachromosomal Mutant EGFR DNA'.

Nikolaev, S., F. Santoni, M. Garieri, P. Makrythanasis, E. Falconnet, M. Guipponi, A. Vannier, I. Radovanovic, F. Bena, F. Forestier, K. Schaller, V. Dutoit, V. Clement-Schatlo, P. Y. Dietrich, and S. E. Antonarakis. 2014. 'Extrachromosomal driver mutations in glioblastoma and low-grade glioma', *Nat Commun*, 5: 5690.

Nowell, P. C. 1962. 'The minute chromosome (Phl) in chronic granulocytic leukemia', *Blut*, 8: 65-6.

Nussenzweig, A., and M. C. Nussenzweig. 2010. 'Origin of chromosomal translocations in lymphoid cancer', *Cell*, 141: 27-38.

Oliveira, M. L., P. Akkapeddi, D. Ribeiro, A. Melao, and J. T. Barata. 2019. 'IL-7R-mediated signaling in T-cell acute lymphoblastic leukemia: An update', *Adv Biol Regul*, 71: 88-96.

Ouillette, Peter, Harry Erba, Lisa Kujawski, Mark Kaminski, Kerby Shedden, and Sami N. Malek. 2008. 'Integrated Genomic Profiling of Chronic Lymphocytic Leukemia Identifies Subtypes of Deletion 13q14', *Cancer Research*, 68: 1012-21.

Pandita, A., K. D. Aldape, G. Zadeh, A. Guha, and C. D. James. 2004. 'Contrasting in vivo and in vitro fates of glioblastoma cell subpopulations with amplified EGFR', *Genes Chromosomes Cancer*, 39: 29-36.

Pannunzio, N. R., S. Li, G. Watanabe, and M. R. Lieber. 2014. 'Non-homologous end joining often uses microhomology: implications for alternative end joining', *DNA Repair (Amst)*, 17: 74-80.

Pannunzio, N. R., G. Watanabe, and M. R. Lieber. 2018. 'Nonhomologous DNA end-joining for repair of DNA double-strand breaks', *J Biol Chem*, 293: 10512-23.

Paques, F., and J. E. Haber. 1999. 'Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae', *Microbiol Mol Biol Rev*, 63: 349-404.

Paulsen, Teressa, Pankaj Kumar, M. Murat Koseoglu, and Anindya Dutta. 2018. 'Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells', *Trends in Genetics*, 34: 270-78.

Pellestor, F. 2019. 'Chromoanagenesis: cataclysms behind complex chromosomal rearrangements', *Mol Cytogenet*, 12: 6.

Philipp-Staheli, J., S. R. Payne, and C. J. Kemp. 2001. 'p27(Kip1): regulation and function of a haploinsufficient tumor suppressor and its misregulation in cancer', *Exp Cell Res*, 264: 148-68.

Piganeau, M., H. Ghezraoui, A. De Cian, L. Guittat, M. Tomishima, L. Perrouault, O. Rene, G. E. Katibah, L. Zhang, M. C. Holmes, Y. Doyon, J. P. Concordet, C. Giovannangeli, M. Jasin, and E. Brunet. 2013. 'Cancer translocations in human cells induced by zinc finger and TALE nucleases', *Genome Res*, 23: 1182-93.

Pinello, L., M. C. Canver, M. D. Hoban, S. H. Orkin, D. B. Kohn, D. E. Bauer, and G. C. Yuan. 2016. 'Analyzing CRISPR genome-editing experiments with CRISPResso', *Nat Biotechnol*, 34: 695-7.

Platt, R. J., S. Chen, Y. Zhou, M. J. Yim, L. Swiech, H. R. Kempton, J. E. Dahlman, O. Parnas, T. M. Eisenhaure, M. Jovanovic, D. B. Graham, S. Jhunjhunwala, M. Heidenreich, R. J. Xavier, R. Langer, D. G. Anderson, N. Hacohen, A. Regev, G. Feng, P. A. Sharp,

and F. Zhang. 2014. 'CRISPR-Cas9 knockin mice for genome editing and cancer modeling', *Cell*, 159: 440-55.

Quinn, L. A., G. E. Moore, R. T. Morgan, and L. K. Woods. 1979. 'Cell lines from human colon carcinoma with unusual cell products, double minutes, and homogeneously staining regions', *Cancer Research*, 39: 4914-24.

Rai, R., H. Zheng, H. He, Y. Luo, A. Multani, P. B. Carpenter, and S. Chang. 2010. 'The function of classical and alternative non-homologous end-joining pathways in the fusion of dysfunctional telomeres', *EMBO J*, 29: 2598-610.

Rapp, Moritz, Maximilian W. M. Wintergerst, Wolfgang G. Kunz, Viola K. Vetter, Max M. L. Knott, Dominik Lisowski, Sascha Haubner, Stefan Moder, Raffael Thaler, Stephan Eiber, Bastian Meyer, Natascha Röhrle, Ignazio Piseddu, Simon Grassmann, Patrick Layritz, Benjamin Kühnemuth, Susanne Stutte, Carole Bourquin, Ulrich H. Von Andrian, Stefan Endres, and David Anz. 2019. 'CCL22 controls immunity by promoting regulatory T cell communication with dendritic cells in lymph nodes', *Journal of Experimental Medicine*, 216: 1170-81.

Ratnaparkhe, M., J. K. L. Wong, P. C. Wei, M. Hlevnjak, T. Kolb, M. Simovic, D. Haag, Y. Paul, F. Devens, P. Northcott, D. T. W. Jones, M. Kool, A. Jauch, A. Pastorczak, W. Mlynarski, A. Korshunov, R. Kumar, S. M. Downing, S. M. Pfister, M. Zapatka, P. J. McKinnon, F. W. Alt, P. Lichter, and A. Ernst. 2018. 'Defective DNA damage repair leads to frequent catastrophic genomic events in murine and human tumors', *Nat Commun*, 9: 4760.

Reimer, J., S. Knoss, M. Labuhn, E. M. Charpentier, G. Gohring, B. Schlegelberger, J. H. Klusmann, and D. Heckl. 2017. 'CRISPR-Cas9-induced t(11;19)/MLL-ENL translocations initiate leukemia in human hematopoietic progenitor cells in vivo', *Haematologica*, 102: 1558-66.

Rello-Varona, S., D. Lissa, S. Shen, M. Niso-Santano, L. Senovilla, G. Marino, I. Vitale, M. Jemaa, F. Harper, G. Pierron, M. Castedo, and G. Kroemer. 2012. 'Autophagic removal of micronuclei', *Cell Cycle*, 11: 170-6.

Reth, M., and P. Nielsen. 2014. 'Signaling circuits in early B-cell development', *Adv Immunol*, 122: 129-75.

Reyes-Gonzalez, J. M., and P. E. Vivas-Mejia. 2021. 'c-MYC and Epithelial Ovarian Cancer', *Front Oncol*, 11: 601512.

Richardson, C. D., G. J. Ray, M. A. DeWitt, G. L. Curie, and J. E. Corn. 2016. 'Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA', *Nat Biotechnol*, 34: 339-44.

Richardson, C., and M. Jasin. 2000. 'Frequent chromosomal translocations induced by DNA double-strand breaks', *Nature*, 405: 697-700.

Rivera, M. N., W. J. Kim, J. Wells, D. R. Driscoll, B. W. Brannigan, M. Han, J. C. Kim, A. P. Feinberg, W. L. Gerald, S. O. Vargas, L. Chin, A. J. Iafrate, D. W. Bell, and D. A. Haber. 2007. 'An X chromosome gene, WTX, is commonly inactivated in Wilms tumor', *Science*, 315: 642-5.

Roper, J., T. Tammela, N. M. Cetinbas, A. Akkad, A. Roghanian, S. Rickelt, M. Almeqdadi, K. Wu, M. A. Oberli, F. J. Sanchez-Rivera, Y. K. Park, X. Liang, G. Eng, M. S. Taylor, R. Azimi, D. Kedrin, R. Neupane, S. Beyaz, E. T. Sicinska, Y. Suarez, J. Yoo, L. Chen, L. Zukerberg, P. Katajisto, V. Deshpande, A. J. Bass, P. N. Tsichlis, J. Lees, R. Langer, R. O. Hynes, J. Chen, A. Bhutkar, T. Jacks, and O. H. Yilmaz. 2017. 'In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis', *Nat Biotechnol*, 35: 569-76.

Rouet, P., F. Smih, and M. Jasin. 1994. 'Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease', *Mol Cell Biol*, 14: 8096-106.

Roukos, V., and T. Misteli. 2014. 'The biogenesis of chromosome translocations', *Nat Cell Biol*, 16: 293-300.

Rowley, J. D. 1973. 'Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining', *Nature*, 243: 290-3.

Roy, R., J. Chun, and S. N. Powell. 2011. 'BRCA1 and BRCA2: different roles in a common pathway of genome protection', *Nat Rev Cancer*, 12: 68-78.

Sallmyr, A., A. E. Tomkinson, and F. V. Rassool. 2008. 'Up-regulation of WRN and DNA ligase IIIalpha in chronic myeloid leukemia: consequences for the repair of DNA double-strand breaks', *Blood*, 112: 1413-23.

Salomoni, P., and C. Bellodi. 2007. 'New insights into the cytoplasmic function of PML', *Histol Histopathol*, 22: 937-46.

Sanchez-Rivera, F. J., T. Papagiannakopoulos, R. Romero, T. Tammela, M. R. Bauer, A. Bhutkar, N. S. Joshi, L. Subbaraj, R. T. Bronson, W. Xue, and T. Jacks. 2014. 'Rapid modelling of cooperating genetic events in cancer through somatic genome editing', *Nature*, 516: 428-31.

Santarius, T., J. Shipley, D. Brewer, M. R. Stratton, and C. S. Cooper. 2010. 'A census of amplified and overexpressed human cancer genes', *Nat Rev Cancer*, 10: 59-64.

Santarosa, M., and A. Ashworth. 2004. 'Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way', *Biochim Biophys Acta*, 1654: 105-22.

Savitsky, K., A. Bar-Shira, S. Gilad, G. Rotman, Y. Ziv, L. Vanagaite, D. A. Tagle, S. Smith, T. Uziel, S. Sfez, M. Ashkenazi, I. Pecker, M. Frydman, R. Harnik, S. R. Patanjali, A. Simmons, G. A. Clines, A. Sartiel, R. A. Gatti, L. Chessa, O. Sanal, M. F. Lavin, N. G. Jaspers, A. M. Taylor, C. F. Arlett, T. Miki, S. M. Weissman, M. Lovett, F. S. Collins, and Y. Shiloh. 1995. 'A single ataxia telangiectasia gene with a product similar to PI-3 kinase', *Science*, 268: 1749-53.

Schirmbeck, R., J. Reimann, S. Kochanek, and F. Kreppel. 2008. 'The immunogenicity of adenovirus vectors limits the multispecificity of CD8 T-cell responses to vector-encoded transgenic antigens', *Mol Ther*, 16: 1609-16.

Schoenlein, P. V., J. T. Barrett, A. Kulharya, M. R. Dohn, A. Sanchez, D. Y. Hou, and J. McCoy. 2003. 'Radiation therapy depletes extrachromosomally amplified drug resistance genes and oncogenes from tumor cells via micronuclear capture of episomes and double minute chromosomes', *Int J Radiat Oncol Biol Phys*, 55: 1051-65.

Schwarz, K., Y. Ma, U. Pannicke, and M. R. Lieber. 2003. 'Human severe combined immune deficiency and DNA repair', *Bioessays*, 25: 1061-70.

Scully, R., A. Panday, R. Elango, and N. A. Willis. 2019. 'DNA double-strand break repair-pathway choice in somatic mammalian cells', *Nat Rev Mol Cell Biol*, 20: 698-714.

Seshagiri, S., E. W. Stawiski, S. Durinck, Z. Modrusan, E. E. Storm, C. B. Conboy, S. Chaudhuri, Y. Guan, V. Janakiraman, B. S. Jaiswal, J. Guillory, C. Ha, G. J. Dijkgraaf, J. Stinson, F. Gnad, M. A. Huntley, J. D. Degenhardt, P. M. Haverty, R. Bourgon, W. Wang, H. Koeppen, R. Gentleman, T. K. Starr, Z. Zhang, D. A. Largaespada, T. D. Wu, and F. J. de Sauvage. 2012. 'Recurrent R-spondin fusions in colon cancer', *Nature*, 488: 660-4.

Sfeir, A., and L. S. Symington. 2015. 'Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?', *Trends Biochem Sci*, 40: 701-14.

Shen, M. M. 2013. 'Chromoplexy: a new category of complex rearrangements in the cancer genome', *Cancer Cell*, 23: 567-9.

Shibata, Y., P. Kumar, R. Layer, S. Willcox, J. R. Gagan, J. D. Griffith, and A. Dutta. 2012. 'Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues', *Science*, 336: 82-6.

Shimizu, N. 2011. 'Molecular mechanisms of the origin of micronuclei from extrachromosomal elements', *Mutagenesis*, 26: 119-23.

Shimizu, N., F. Kamezaki, and S. Shigematsu. 2005. 'Tracking of microinjected DNA in live cells reveals the intracellular behavior and elimination of extrachromosomal genetic material', *Nucleic Acids Res*, 33: 6296-307.

Shimizu, N., N. Misaka, and K. Utani. 2007. 'Nonselective DNA damage induced by a replication inhibitor results in the selective elimination of extrachromosomal double minutes from human cancer cells', *Genes Chromosomes Cancer*, 46: 865-74.

Shimizu, N., H. Nakamura, T. Kadota, K. Kitajima, T. Oda, T. Hirano, and H. Utiyama. 1994. 'Loss of amplified c-myc genes in the spontaneously differentiated HL-60 cells', *Cancer Research*, 54: 3561-7.

Shoshani, O., S. F. Brunner, R. Yaeger, P. Ly, Y. Nechemia-Arbely, D. H. Kim, R. Fang, G. A. Castillon, M. Yu, J. S. Z. Li, Y. Sun, M. H. Ellisman, B. Ren, P. J. Campbell, and D. W. Cleveland. 2021. 'Chromothripsis drives the evolution of gene amplification in cancer', *Nature*, 591: 137-41.

Simsek, D., E. Brunet, S. Y. Wong, S. Katyal, Y. Gao, P. J. McKinnon, J. Lou, L. Zhang, J. Li, E. J. Rebar, P. D. Gregory, M. C. Holmes, and M. Jasin. 2011. 'DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation', *PLoS Genet*, 7: e1002080.

Simsek, D., and M. Jasin. 2010. 'Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation', *Nat Struct Mol Biol*, 17: 410-6.

Slamon, D. J., G. M. Clark, S. G. Wong, W. J. Levin, A. Ullrich, and W. L. McGuire. 1987. 'Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene', *Science*, 235: 177-82.

Smogorzewska, A., J. Karlseder, H. Holtgreve-Grez, A. Jauch, and T. de Lange. 2002. 'DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2', *Curr Biol*, 12: 1635-44.

Soda, M., Y. L. Choi, M. Enomoto, S. Takada, Y. Yamashita, S. Ishikawa, S. Fujiwara, H. Watanabe, K. Kurashina, H. Hatanaka, M. Bando, S. Ohno, Y. Ishikawa, H. Aburatani, T. Niki, Y. Sohara, Y. Sugiyama, and H. Mano. 2007. 'Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer', *Nature*, 448: 561-6.

Spektor, Alexander, Neil T. Umbreit, and David Pellman. 2017. 'Cell Biology: When Your Own Chromosomes Act like Foreign DNA', *Current Biology*, 27: R1228-R31.

Stark, J. M., and M. Jasin. 2003. 'Extensive loss of heterozygosity is suppressed during homologous repair of chromosomal breaks', *Mol Cell Biol*, 23: 733-43.

Stephens, P. J., C. D. Greenman, B. Fu, F. Yang, G. R. Bignell, L. J. Mudie, E. D. Pleasance, K. W. Lau, D. Beare, L. A. Stebbings, S. McLaren, M. L. Lin, D. J. McBride, I. Varela, S. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, M. A. Quail, J. Burton, H. Swerdlow, N. P. Carter, L. A. Morsberger, C. Iacobuzio-Donahue, G. A. Follows, A. R. Green, A. M. Flanagan, M. R. Stratton, P. A. Futreal, and P. J. Campbell. 2011. 'Massive genomic rearrangement acquired in a single catastrophic event during cancer development', *Cell*, 144: 27-40.

Stephens, P. J., D. J. McBride, M. L. Lin, I. Varela, E. D. Pleasance, J. T. Simpson, L. A. Stebbings, C. Leroy, S. Edkins, L. J. Mudie, C. D. Greenman, M. Jia, C. Latimer, J. W. Teague, K. W. Lau, J. Burton, M. A. Quail, H. Swerdlow, C. Churcher, R. Natrajan, A. M. Sieuwerts, J. W. Martens, D. P. Silver, A. Langerod, H. E. Russnes, J. A. Foekens, J. S. Reis-Filho, L. van 't Veer, A. L. Richardson, A. L. Borresen-Dale, P. J. Campbell, P. A. Futreal, and M. R. Stratton. 2009. 'Complex landscapes of somatic rearrangement in human breast cancer genomes', *Nature*, 462: 1005-10.

Sun, L., J. Wu, F. Du, X. Chen, and Z. J. Chen. 2013. 'Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway', *Science*, 339: 786-91.

Szabo, Quentin, Frédéric Bantignies, and Giacomo Cavalli. 2019. 'Principles of genome folding into topologically associating domains', *Science Advances*, 5: eaaw1668.

Taub, R., I. Kirsch, C. Morton, G. Lenoir, D. Swan, S. Tronick, S. Aaronson, and P. Leder. 1982. 'Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells', *Proc Natl Acad Sci U S A*, 79: 7837-41.

Thompson, S. L., and D. A. Compton. 2008. 'Examining the link between chromosomal instability and aneuploidy in human cells', *J Cell Biol*, 180: 665-72.

———. 2011. 'Chromosomes and cancer cells', *Chromosome Res*, 19: 433-44.

Tlsty, T. D. 1990. 'Normal diploid human and rodent cells lack a detectable frequency of gene amplification', *Proc Natl Acad Sci U S A*, 87: 3132-6.

Tobin, L. A., C. Robert, A. P. Rapoport, I. Gojo, M. R. Baer, A. E. Tomkinson, and F. V. Rassool. 2013. 'Targeting abnormal DNA double-strand break repair in tyrosine kinase inhibitor-resistant chronic myeloid leukemias', *Oncogene*, 32: 1784-93.

Tomlins, S. A., D. R. Rhodes, S. Perner, S. M. Dhanasekaran, R. Mehra, X. W. Sun, S. Varambally, X. Cao, J. Tchinda, R. Kuefer, C. Lee, J. E. Montie, R. B. Shah, K. J. Pienta, M. A. Rubin, and A. M. Chinnaiyan. 2005. 'Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer', *Science*, 310: 644-8.

Tung, N. M., J. C. Boughey, L. J. Pierce, M. E. Robson, I. Bedrosian, J. R. Dietz, A. Dragun, J. B. Gelpi, E. W. Hofstatter, C. J. Isaacs, I. Jatoi, E. Kennedy, J. K. Litton, N. A. Mayr, R. D. Qamar, M. G. Trombetta, B. E. Harvey, M. R. Somerfield, and D. Zakalik. 2020. 'Management of Hereditary Breast Cancer: American Society of Clinical Oncology, American Society for Radiation Oncology, and Society of Surgical Oncology Guideline', *J Clin Oncol*, 38: 2080-106.

Turner, K. M., V. Deshpande, D. Beyter, T. Koga, J. Rusert, C. Lee, B. Li, K. Arden, B. Ren, D. A. Nathanson, H. I. Kornblum, M. D. Taylor, S. Kaushal, W. K. Cavenee, R. Wechsler-Reya, F. B. Furnari, S. R. Vandenberg, P. N. Rao, G. M. Wahl, V. Bafna, and P. S. Mischel. 2017. 'Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity', *Nature*, 543: 122-25.

Utani, K., J. K. Kawamoto, and N. Shimizu. 2007. 'Micronuclei bearing acentric extrachromosomal chromatin are transcriptionally competent and may perturb the cancer cell phenotype', *Mol Cancer Res*, 5: 695-704.

Valent, A., J. Benard, B. Clausse, M. Barrois, D. Valteau-Couanet, M. J. Terrier-Lacombe, B. Spengler, and A. Bernheim. 2001. 'In vivo elimination of acentric double minutes containing amplified MYCN from neuroblastoma tumor cells through the formation of micronuclei', *Am J Pathol*, 158: 1579-84.

Verhaak, R. G. W., V. Bafna, and P. S. Mischel. 2019. 'Extrachromosomal oncogene amplification in tumour pathogenesis and evolution', *Nat Rev Cancer*, 19: 283-88.

Von Hoff, D. D., J. R. McGill, B. J. Forseth, K. K. Davidson, T. P. Bradley, D. R. Van Devanter, and G. M. Wahl. 1992. 'Elimination of extrachromosomally amplified MYC genes from human tumor cells reduces their tumorigenicity', *Proc Natl Acad Sci U S A*, 89: 8165-9.

Wang, D., H. Mou, S. Li, Y. Li, S. Hough, K. Tran, J. Li, H. Yin, D. G. Anderson, E. J. Sontheimer, Z. Weng, G. Gao, and W. Xue. 2015. 'Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses', *Hum Gene Ther*, 26: 432-42.

Wang, H., H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang, and R. Jaenisch. 2013. 'One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering', *Cell*, 153: 910-8.

Warren, D. J. 2011. 'Preparation of highly efficient electrocompetent Escherichia coli using glycerol/mannitol density step centrifugation', *Anal Biochem*, 413: 206-7.

Weischenfeldt, Joachim, Taronish Dubash, Alexandros P. Drainas, Balca R. Mardin, Yuanyuan Chen, Adrian M. Stütz, Sebastian M. Waszak, Graziella Bosco, Ann Rita Halvorsen, Benjamin Raeder, Theocharis Efthymiopoulos, Serap Erkek, Christine Siegl, Hermann Brenner, Odd Terje Brustugun, Sebastian M. Dieter, Paul A. Northcott, Iver Petersen, Stefan M. Pfister, Martin Schneider, Steinar K. Solberg, Erik Thunissen, Wilko Weichert, Thomas Zichner, Roman Thomas, Martin Peifer, Aslaug Helland, Claudia R. Ball, Martin Jechlinger, Rocio Sotillo, Hanno Glimm, and Jan O. Korbel. 2017. 'Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking', *Nature Genetics*, 49: 65-74.

Wooster, R., G. Bignell, J. Lancaster, S. Swift, S. Seal, J. Mangion, N. Collins, S. Gregory, C. Gumbs, and G. Micklem. 1995. 'Identification of the breast cancer susceptibility gene BRCA2', *Nature*, 378: 789-92.

Wu, Sihan, Kristen M. Turner, Nam Nguyen, Ramya Raviram, Marcella Erb, Jennifer Santini, Jens Luebeck, Utkrisht Rajkumar, Yarui Diao, Bin Li, Wenjing Zhang, Nathan Jameson, M. Ryan Corces, Jeffrey M. Granja, Xingqi Chen, Ceyda Coruh, Armen Abnousi, Jack Houston, Zhen Ye, Rong Hu, Miao Yu, Hoon Kim, Julie A. Law, Roel G. W. Verhaak, Ming Hu, Frank B. Furnari, Howard Y. Chang, Bing Ren, Vineet Bafna, and Paul S. Mischel. 2019. 'Circular ecDNA promotes accessible chromatin and high oncogene expression', *Nature*, 575: 699-703.

Xiang, S., J. Fruehauf, and C. J. Li. 2006. 'Short hairpin RNA-expressing bacteria elicit RNA interference in mammals', *Nat Biotechnol*, 24: 697-702.

Xue, W., S. Chen, H. Yin, T. Tammela, T. Papagiannakopoulos, N. S. Joshi, W. Cai, G. Yang, R. Bronson, D. G. Crowley, F. Zhang, D. G. Anderson, P. A. Sharp, and T. Jacks. 2014. 'CRISPR-mediated direct mutation of cancer genes in the mouse liver', *Nature*, 514: 380-4.

Yang, H., H. Wang, J. Ren, Q. Chen, and Z. J. Chen. 2017. 'cGAS is essential for cellular senescence', *Proc Natl Acad Sci U S A*, 114: E4612-E20.

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

Yi, Eunhee, Amit D. Gujar, Molly Guthrie, Hoon Kim, Kevin C. Johnson, Samirkumar B. Amin, Sunit Das, Patricia A. Clow, Albert W. Cheng, and Roel GW Verhaak. 2020. 'Live-cell imaging shows uneven segregation of extrachromosomal DNA elements and transcriptionally active extrachromosomal DNA clusters in cancer', *bioRxiv*: 2020.10.20.335216.

Yin, Y., M. A. Tainsky, F. Z. Bischoff, L. C. Strong, and G. M. Wahl. 1992. 'Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles', *Cell*, 70: 937-48.

Yost, S. E., S. Pastorino, S. Rozenzhak, E. N. Smith, Y. S. Chao, P. Jiang, S. Kesari, K. A. Frazer, and O. Harismendy. 2013. 'High-resolution mutational profiling suggests the genetic validity of glioblastoma patient-derived pre-clinical models', *PLOS ONE*, 8: e56185.

Yu, G., L. G. Wang, Y. Han, and Q. Y. He. 2012. 'clusterProfiler: an R package for comparing biological themes among gene clusters', *OMICS*, 16: 284-7.

Yu, Y., and A. Bradley. 2001. 'Engineering chromosomal rearrangements in mice', *Nat Rev Genet*, 2: 780-90.

Zenatti, P. P., D. Ribeiro, W. Li, L. Zuurbier, M. C. Silva, M. Paganin, J. Tritapoe, J. A. Hixon, A. B. Silveira, B. A. Cardoso, L. M. Sarmento, N. Correia, M. L. Toribio, J. Kobarg, M. Horstmann, R. Pieters, S. R. Brandalise, A. A. Ferrando, J. P. Meijerink, S. K. Durum, J. A. Yunes, and J. T. Barata. 2011. 'Oncogenic IL7R gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia', *Nat Genet*, 43: 932-9.

Zhang, C. Z., A. Spektor, H. Cornils, J. M. Francis, E. K. Jackson, S. Liu, M. Meyerson, and D. Pellman. 2015. 'Chromothripsis from DNA damage in micronuclei', *Nature*, 522: 179-84.

Zhang, Suping, and Thomas J. Kipps. 2014. 'The pathogenesis of chronic lymphocytic leukemia', *Annual review of pathology*, 9: 103-18.

Zhao, Mengmeng, Fei Wang, Juehui Wu, Yuanna Cheng, Yajuan Cao, Xiangyang Wu, Mingtong Ma, Fen Tang, Zhi Liu, Haipeng Liu, and Baoxue Ge. 2021. 'CGAS is a micronucleophagy receptor for the clearance of micronuclei', *Autophagy*: 1-17.

Zheng, Binhai, Marijke Sage, Elizabeth A. Sheppeard, Vesna Jurecic, and Allan Bradley. 2000. 'Engineering Mouse Chromosomes with Cre-loxP: Range, Efficiency, and Somatic Applications', *Molecular and Cellular Biology*, 20: 648-55.

Zheng, K., F. J. Cubero, and Y. A. Nevzorova. 2017. 'c-MYC-Making Liver Sick: Role of c-MYC in Hepatic Cell Function, Homeostasis and Disease', *Genes (Basel)*, 8.

Zhou, J., Z. Zhou, P. Ji, M. Ma, J. Guo, and S. Jiang. 2019. 'Effect of fecal microbiota transplantation on experimental colitis in mice', *Exp Ther Med*, 17: 2581-86.

Zhu, C., K. D. Mills, D. O. Ferguson, C. Lee, J. Manis, J. Fleming, Y. Gao, C. C. Morton, and F. W. Alt. 2002. 'Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations', *Cell*, 109: 811-21.

Zhu, Y., A. D. Gujar, C. H. Wong, H. Tjong, C. Y. Ngan, L. Gong, Y. A. Chen, H. Kim, J. Liu, M. Li, A. Mil-Homens, R. Maurya, C. Kuhlberg, F. Sun, E. Yi, A. C. deCarvalho, Y. Ruan, R. G. W. Verhaak, and C. L. Wei. 2021. 'Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription', *Cancer Cell*.

Zierhut, C., N. Yamaguchi, M. Paredes, J. D. Luo, T. Carroll, and H. Funabiki. 2019. 'The Cytoplasmic DNA Sensor cGAS Promotes Mitotic Cell Death', *Cell*, 178: 302-15 e23.