Engineering chromosomal rearrangements in hematopoietic cells

by

Raymond Wu

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

September 2018

Andrea Ventura, MD, PhD	Date
Dissertation Mentor	

DEDICATION

I would like to dedicate this thesis to my parents Xiu Fang and Jian Hang Wu, and my sister Vanessa Wu. They helped me through this process even when things got tough. I hope that I make them proud with this work. I would also like to dedicate this thesis to Gabriel Blitz, who left the world too soon, as I made a promise many years ago to buck the trend and try to follow my dreams.

ABSTRACT

The widespread application of high throughput sequencing methods to human tumor samples is leading to the identification of a growing number of novel chromosomal rearrangements that result in the formation of gene fusions. A subset of these fusions is likely to be oncogenic drivers, yet few of them have been characterized in detail. For this reason, it is crucial to develop novel strategies to efficiently model and study this class of cancer-associated mutations. *Ex vivo* and *in vivo* transgenic overexpression models, while informative and relatively simple to generate, fail to fully recapitulate the genetic complexity of gene fusions. Early attempts to model the underlying chromosomal rearrangement using Cre-lox-based approaches had varying levels of success, and they have not been widely used due to their technical complexity and time-consuming nature. The situation has dramatically changed with the development of CRISPR-Cas9 based gene editing methods which provide a straightforward mean to introduce site specific double strand DNA breaks anywhere in the genome. Previous studies, including this one, have shown that by simultaneously expressing two single guide RNAs and Cas9, it is now possible to quickly engineer a wide range of intra- and inter-chromosomal rearrangements.

Using this novel approach, several rearrangements, have been successfully modeled in mice. However, in the context of hematological malignancies, where chromosomal rearrangements often occur, progress has been slower. To address this, I delivered tandem single guide RNAs and Cas9 by nucleofection into the mouse pro-B Ba/F3 cell line. This generated the chromosomal rearrangement that results in the *Npm1-Alk* gene fusion and resulted in Ba/F3 transformation. Using allograft mouse tumor models, I show that Ba/F3 cells engineered to harbor the murine *Npm1-Alk* rearrangement support tumor growth in nude mice, as well as forming frank leukemias with a transplantation model in BALB/c mice.

To better model gene fusions *in vivo*, I adapted and optimized a murine stem cell virus viral vector to concomitantly express CRISPR-Cas9 protein and tandem single guide RNAs to generate chromosomal rearrangements. This construct expresses Cas9 and tandem single guide RNAs in the convergent orientation to improve Pol III single guide transcription. Using this construct, I generated the *Npm1-Alk* gene fusion in mouse primary mouse embryonic fibroblasts. The tools, reagents, and general strategies I have developed in this study can be used to rapidly model newly identified gene fusions in the hematopoietic system.

BIOGRAPHICAL SKETCH

Raymond Wu was born in Brooklyn, New York on April 30, 1989. Raymond developed an interest in science at an early age during elementary and middle school science fairs. Raymond graduated from Stuyvesant High School in Manhattan where he had his first research experience at Columbia University studying the effect of radiation on DNA. Raymond then moved to Ithaca, New York for his Bachelor of Arts (B.A) degree in Biology at Cornell University, where he completed an honors thesis on RNA aptamers (*summa cum laude*) under the supervision of Dr. John T. Lis. In July 2011, Raymond then started his PhD studies at the Louis V. Gerstner Jr. Graduate School of Biomedical Sciences at Memorial Sloan Kettering. There he pursued his interests in non-coding RNA and mouse models of disease.

ACKNOWLEDGMENTS

I would like to thank my mentor, Dr. Andrea Ventura, for his mentorship throughout the years. Though not every project was successful, he was willing to help me through the tough times, and I am a better scientist because of it. I also want to thank my fellow lab members, past and present, for helping through this process. Their scientific and career advice helped shape who I am today. Particularly, I want to thank Dr. Danilo Maddalo, who generated the initial data for this project, Dr. Joana Vidigal, who developed the dual sgRNA guide cloning method, Dr. Turgut Dogruluk, who worked with me on the viability assays and showing me how to work with Ba/F3 cells, and Dr. Ram Kannan, who worked with me on the Western blots on the signaling pathways and who offered his career advice.

I would like to thank my Thesis Examination Committee members, Dr. Robert Benezra as chair of the committee and Dr. Lukas Dow as external examiner. I thank my Thesis Committee members, Dr. Richard White and Dr. Michael Kharas for their incredible patience and support. They helped me troubleshoot my experiments when things had not gone the way we had hoped. I also thank Dr. White for allowing me to use their Neon Transfection machine and Dr. Kharas for lending me the Ba/F3 cell line and his experimental suggestions.

I would also like to thank the Pelletier and Lodish labs for providing reagents that helped accelerate my research. Additionally, I thank the Molecular Cytogenetic (Dr. Gouri Nanjangud) and Comparative Pathology (Dr. Alessandra Piersigilli) core facilities for their help in preparing and analyzing my samples. I would also like to thank members of the 12th floor of Zuckerman, which include the Benezra, Mayr and Offit lab for reagents, suggestions, and general advice. I thank Dr. Robert Benezra for offering his advice in our joint lab meetings. They helped shape who I am as a scientist and to put things in perspective.

I thank our graduate school, especially our Dean, Dr. Kenneth Marians for his

leadership and advice. I also want to thank the Associate Dean Linda D. Burnley for support throughout the years. I would like to thank David L. McDonagh for his help through the thesis process along with Julie Masen, Stacy De La Cruz, and Dr. Tom Magaldi. I would also like to mention past Gerstner Sloan Kettering Graduate School staff including Maria Torres, Ivan Gerena, and Iwona Abramek. They made my time at the graduate school a better experience and s little less stressful. I also want to thank my classmates, Erman Karasu and Demetrius Dimucci who I had the pleasure of getting to know over the course of my research.

Lastly, I would like to thank our funding sources, which make our research possible and advance our understanding. My work was supported by National Institutes of Health (R01-CA149707) and the Cycle for Survival Grant has provided support on my work on modeling oncogenic fusions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ALCL: Anaplastic large cell lymphoma

ALK: Anaplastic Lymphoma Kinase

BCM: B cell medium

BCR: B cell receptor

βMe: beta-mercaptoethanol

Cas: CRISPR associated gene

CBh: Chicken beta-actin hybrid

cDNA: complementary DNA

CRISPR: Clustered Regularly Interspaced Palindromic Repeats

crRNA: CRISPR related RNA

DMEM: Dulbecco's Modified Eagle's Medium

DMSO: Dimethyl Sulfoxide

DNA: Deoxyribonucleic acid

DS: Double strand

DSB: Double Strand Break

EML4: Echinoderm microtubule associated protein like 4

FACS: Fluorescence activated cell sorting

FBS: Fetal Bovine Serum

FFPE: Formalin-Fixed/Paraffin embedded

FISH: Fluorescence *in situ* hybridization

FITC: Fluorescein isothiocyanate

FLHSC: Fetal Liver hematopoietic stem cell

GDNA: genomic DNA

GFP: Green Fluorescent Protein

h(): human

HDR: Homology Directed Repair

HPRT: Hypoxanthine Phosphoribosyltransferase

HRP: Horseradish peroxidase

HSPC: Hematopoietic stem and progenitor cell

IL3: Interleukin 3

Indels: Insertion Deletion Mutations

IRES: Internal Ribosomal Entry Site

LTRs: Long Terminal Repeats

m(): murine

mRNA: Messenger RNA

MSCV: Murine Stem Cell Virus

NHEJ: Non-Homologous End Joining

n-: nucleofected

N-A: NPM1-ALK (protein)

N-A: *Npm1-Alk* (murine)

NPM1: Nucleophosmin 1

nt: nucleotide

P2A: Peptide 2A

PAM: Protospacer Adjacent Motif

PBS: Phosphate Buffered Saline

PCR: Polymerase Chain Reaction

PGK: Phosphoglycerate Kinase

PNK: Polynucleotide Kinase

POL: Polymerase

PS: Penicillin-Streptomycin

RCF: Relative Centrifugal Force

RIPA: Radioimmunoprecipitation Assay

RNA: Ribonucleic Acid

RNAi: RNA interference

RNP: Ribonucleoprotein

RPMI: Roswell Memorial Institute Medium

RT: Room Temperature

RT: Reverse Transcriptase

SDS: Sodium Dodecyl Sulfate

sgRNA: single guide RNA

shRNA: short hairpin RNA

siRNA: short interfering RNA

ss: single stranded

SSA: Single strand annealing

SSC: Saline-sodium citrate

TCR: T cell receptor

tracrRNA: Trans-activating CRISPR-related RNA

VSVG: Vesicular Stomatitis Virus Glycoprotein

WT: wild-type

INTRODUCTION: Chromosomal rearrangements in cancer

A brief history of chromosomal rearrangements in human cancers

The idea that chromosomal abnormalities could promote cancer was first proposed over 100 years ago by Theodor Boveri (Bignold et al., 2006; Boveri, 2008). Building on his earlier work on the Chromosome Theory of Inheritance alongside with William Sutton (Crow and Crow, 2002), Boveri observed that sea urchins that inherited abnormal chromosomes had impaired development. Upon seeing similar chromosomal abnormalities in human cancers, he suggested a genetic cause for tumor formation. It was not until half a century later that a chromosomal rearrangement, a reciprocal translocation between the long arms of chromosome 9 and 22—t(9;22)(q34;q11), also known as the "Philadelphia Chromosome"—was directly linked to a human cancer (Forster et al., 2005; Nowell and Hungerford, 1960; Rowley, 1973). The realization that the Philadelphia Chromosome is found in virtually every chronic myeloid leukemia case spurred intense research efforts to identify the underlying molecular mechanisms. Those efforts ultimately led to the discovery that the chromosomal translocation results in an in-frame fusion between the Breakpoint Cluster Region (BCR) and the Abelson murine leukemia viral oncogene homolog 1 (ABL1) genes which in turn leads to constitutive activation and dimerization of the ABL1 kinase (Forster et al., 2005; Groffen et al., 1984; Heisterkamp et al., 1983; Shtivelman et al., 1985; Stam et al., 1985).

The next rearrangement to be identified was the t(8;14) (q24;q32) translocation, which brings the MYC proto-oncogene under the control of the Immunoglobulin heavy chain (IgH) enhancer (Eμ), and is observed in approximately 85% of Burkitt's lymphomas, a type of malignant B cell lymphomas (Erikson et al., 1982). Modelling this genetic event in the mouse led to the generation of the Eμ-Myc transgenic mouse model, a widely used model of B cell lymphomas and the very first 'oncomouse' (Adams et al., 1985; Hanahan et al., 2007).

Other highly recurrent translocations that were identified and characterized in early studies include the t(15;17) (q21;q22) and the t(2;5) (p23;q35) rearrangements. The t(15;17) (q21;q22) rearrangement generates the PML-RARα fusion and is observed in 95% of patients with acute promyelocytic leukemia (APL) (Alcalay et al., 1991; Borrow et al., 1990; de The et al., 1990). The t(2;5) (p23;q35) rearrangement, which underlies a fusion between Nucleophosmin1 and the Anaplastic Large Cell Lymphoma Kinase (NPM1-ALK), is commonly detected in Anaplastic Large Cell Lymphomas (ALCL) (Morris et al., 1994). Identification of the breakpoints and functional characterization of the fusions helped elucidate the mechanisms by which gene fusions cause malignancy and paved the way for targeted therapies using kinase inhibitors such as Imatinib (Druker et al., 2001) and Crizotinib (Crescenzo and Inghirami, 2015).

High-throughput sequencing identifies new fusions

Until recently, the study of cancer-associated chromosomal rearrangements was largely limited to recurrent events that could be detected cytogenetically and identifying the breakpoints and the genes involved was a painstaking process (Mertens et al., 2015; Mitelman, 2015). This has radically changed in the last decade due to the development of high throughput sequencing methods. RNA-sequencing (RNAseq) and whole genome sequencing (WGS), in particular, have led to the identification of thousands novel gene fusions, the majority occurring only in a small fraction of patients, and many involving chromosomal rearrangements not detectable by conventional cytogenetic methods (Mitelman, 2018; Mitelman et al., 2007) **(Figure 1)**. WGS allows the direct identification of the breakpoints of chromosomal rearrangements (Kumar-Sinha et al., 2015), while RNA-sequencing can identify chimeric messenger RNAs (mRNAs) encoded by gene fusions. Computational tools, such as FusionMap and MapSplice, use direct alignment of sequence reads to a reference genome to identify junction boundaries, while other programs, like TopHat-Fusion and FusionCatcher, attempt to detect fusions from paired-

end sequencing reads (Kumar et al., 2016). Much of this data is now readily available on online portals such as The Cancer Genome Atlas Project (TCGA) (Cancer Genome Atlas Research et al., 2013) or the Mitelman Online database (Schaefer et al., 2001).

Types of chromosomal rearrangements

Chromosomal rearrangements can be classified into two large groups: intra-chromosomal (i.e. deletions, inversions, tandem duplications), and inter-chromosomal (i.e. reciprocal, non-reciprocal translocations).

Interstitial deletions

Interstitial deletions, in which a segment of a chromosome is lost, are the simplest class of rearrangements and can create gene fusions through the juxtaposition of two previously separate genes **(Figure 2)**. The requires that the two fused genes are transcribed from the same strand and that the reading frame is maintained upon splicing of the chimeric transcript. One example is the interstitial deletion on chromosome 4q12 that generates an in-frame gene fusion between the Fip1 like protein 1 and Platelet derived growth factor receptor alpha genes and is responsible for a subset of hypereosinophilic syndromes and chronic eosinophilic leukemias (Cools et al., 2003).

Chromosomal Inversions

In chromosomal inversions, a segment of a chromosome is excised and re-inserted with inverted orientation. As for all chromosomal rearrangements, generation of a productive gene fusion requires that the reading frame is maintained in the chimeric transcript. In contrast to deletions, however, with inversions the two genes involved must be transcribed from opposite strands in the wild-type chromosome **(Figure 2)**. The echinoderm microtubule associated protein like 4–ALK (EML4-ALK) fusion oncogene–found in approximately 5% of lung adenocarcinomas–is a classic example of a gene fusion resulting from a chromosomal inversion (Soda et al., 2007).

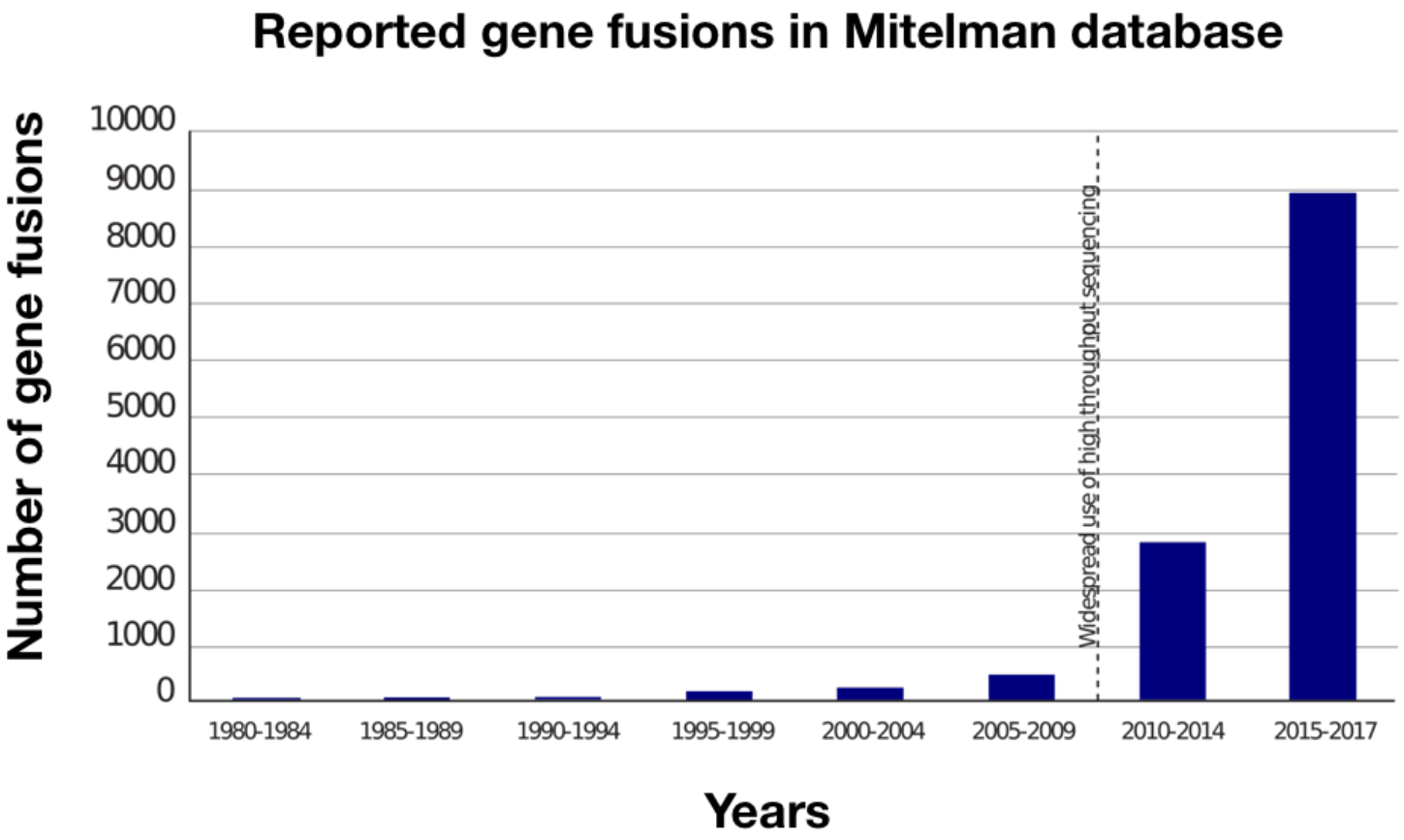


Figure 1. Number of fusions detected over time

In the last century, the number of reported gene fusions remained less then 1000 cases. In recent years, the number of gene fusions reported has vastly increased over time. This has coincided with the growth of high-throughput sequencing of patient tumor samples and detection of novel gene fusions.
Source: Mitelman database 2018 (https://cgap.nci.nih.gov/Chromosomes/Mitelman)

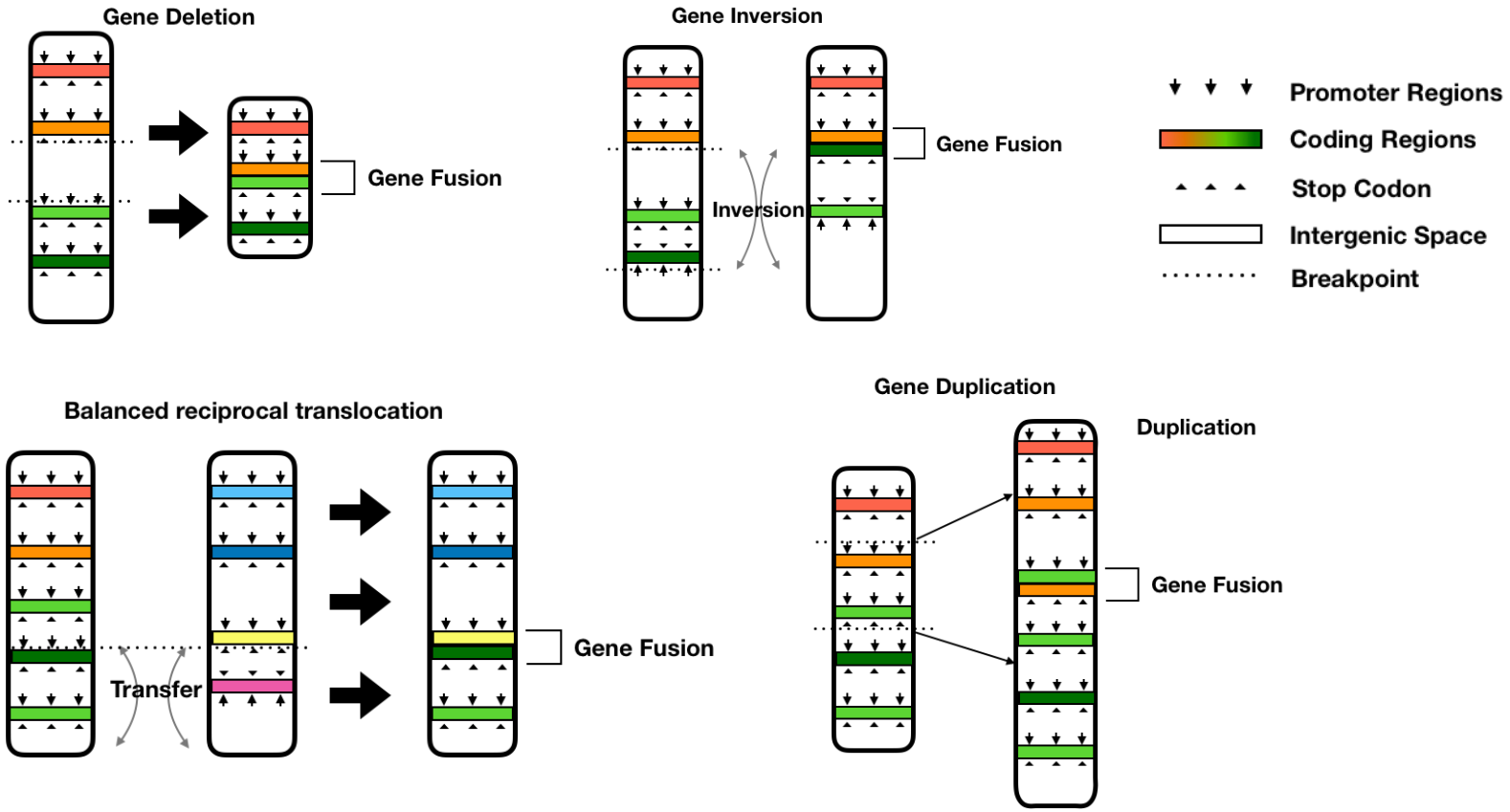


Figure 2. Types of chromosomal rearrangements

Chromosomal breakpoints can result in deletions, inversions, and gene duplications. With inter-chromosomal rearrangements such as balanced reciprocal translocations. When coding regions are joined together, and their transcriptional orientation is maintained, this results in expression of the gene fusion.

Tandem duplications

Tandem duplications are the third major type of intra-chromosomal rearrangements resulting in gene fusions and are characterized by the local duplication of a chromosomal segment **(Figure 2).** Mechanistically, they often result from mitotic recombination between two sister or homolog chromosomes, especially when the breakpoints are in areas of extensive repeats, but they can also be generated by NHEJ initiated by DSBs (Reams and Roth, 2015; Weckselblatt and Rudd, 2015)

Typical examples of tandem duplications resulting in oncogenic gene fusions are the fibroblast growth factor receptor 3 and transforming acidic coiled-coil containing protein 3 (FGFR3-TACC3) gene fusion, initially reported in glioblastoma multiforme (GBM) (Singh et al., 2012) and subsequently identified in a wide range of other solid tumors (Costa et al., 2016) and the KIAA1549-BRAF gene fusion found in the majority of pilocytic astrocytomas and diffuse leptomeningeal neuroglial tumors (Faulkner et al., 2015; Jones et al., 2008).

Chromosomal Translocations

Chromosomal translocations are the last group of common chromosomal rearrangements and are characterized by the fusion of fragments derived from two non-homologous chromosomes. Examples include reciprocal, non-reciprocal and Robertsonian translocations.

Reciprocal translocations involve the breakage and exchange of non-homologous chromosomes. They are considered balanced because the number of chromosomes and the amount of genetic material are preserved **(Figure 2)**. As the most prevalent form of translocation found in cancers, this class is the most extensively characterized in terms of mouse models, with examples including the translocations resulting in the formation of the BCR-ABL1 and NPM1-ALK gene fusions in CML and anaplastic large cell lymphomas, respectively.

Non-reciprocal translocations involve a one-way transfer of genetic material from one non-homologous chromosome to another. This is a destabilizing even, as it can affect chromosome segregation and trigger mitotic checkpoints due to mismatching chromosome pairs. In the presence of telomere dysfunction, fusion and breakage of chromosomes has been suggested to generate complex non-reciprocal translocations and promote tumorigenesis (Artandi et al., 2000).

Robertsonian translocations are characterized by the fusion of the long arms of telocentric or acrocentric chromosomes, generating a metacentric chromosome. As a consequence, the total number of chromosomes is reduced, but not the total number of chromosome arms. Robertsonian translocations are relatively rare in cancers but are occasionally associated with lymphomas and leukemias (Welborn, 2004), and no oncogenic gene fusion caused by a specific Robertsonian translocation has been reported to date.

Biological sources of DSBs and rearrangements

The molecular mechanisms responsible for the generation of chromosomal rearrangements have been extensively studied (Ghosh et al., 2018; Gothe et al., 2018; McCord and Balajee, 2018; Willis et al., 2015). There are several triggers for genomic rearrangements, and it has been observed that certain fusions are more prevalent depending on cancer type (Mitelman et al., 2007). Translocations are especially prevalent in hematological disorders due to the presence of DSBs generated in lymphocytes during V(D)J recombination and class switch mediated by the RAG and AID enzymes, respectively (Aplan, 2006; Iarovaia et al., 2014). Accordingly, many hematological fusion breakpoints involve the immunoglobulin locus **(Figure 3)**.

For most chromosomal rearrangements, the initiating event is thought to be the generation of a pair of double-strand breaks (DSBs), although for gene duplication and deletions, mitotic recombination between sister chromatids or homolog chromosomes also

plays a role (Reams and Roth, 2015). DSBs activates the DNA repair process, which can be categorized broadly as non-homologous end joining (NHEJ), or homology directed repair (HDR). In NHEJ, DSBs are repaired by directly religating the free DNA ends in an error prone manner which often results in insertions and deletions (indels) (Li et al., 2016b), but can also result in chromosomal rearrangements when the 'wrong' ends are ligated. In the S and G2 phase of the cell cycle, when sister chromatids are available as repair templates, the cell can use HDR for error free repair through a process of end resection, strand invasion, and DNA repair synthesis. A third mechanism through which DSBs can be repaired is known as Single Strand Annealing (SSA) and occurs when there is extensive sequence homology between the two resected ends. In contrast to HDR, SSA does not involve strand invasion.

Factors involved in chromosomal rearrangements

As mentioned earlier, a large body of evidence indicates that DSBs plays an important role in promoting chromosomal rearrangements. For example, ionizing radiations, a powerful source of DNA breaks, have been extensively used in *Drosophila melanogaster* to generate random chromosomal inversions, deletions, and translocations for balancing lethal mutations, mapping recessive mutations, and analyze chromosome segregation, respectively (Lindsley et al., 1992). Furthermore, pioneering work from the Jasin group using the I-SceI endonuclease has demonstrated that the introduction of a pair of DSBs in two non-homologous chromosomes is sufficient to induce reciprocal chromosomal translocations with high efficiency (Elliott et al., 2005; Richardson and Jasin, 2000; Rouet et al., 1994a, b; Weinstock et al., 2006). Although these studies demonstrate the essential role played by DSBs in causing chromosomal rearrangements—translocations in particular—dissecting the molecular mechanisms through which the two breakpoints are erroneously joined by the repair machinery proved more challenging and required the development of more flexible programmable nucleases (reviewed in (Brunet

and Jasin, 2018; Willis et al., 2015). The picture that has emerged from these studies is that chromosomal rearrangements occurring in somatic cells are usually mediated by the NHEJ machinery, with the exception of translocations involving conserved repeats, which are mediated by SSA (Elliott et al., 2005). Interestingly, these studies revealed a striking species-specific difference between mouse and humans; while in human cells translocations are generally mediated by the "canonical" NHEJ (cNHEJ) pathway, characterized by little or no end processing and near perfect joining of the two ends (Ghezraoui et al., 2014), in murine cells the "alternative" NHEJ (alt-NHEJ) pathway seems to be the predominant mechanism. This pathway is dependent on Ligase 3 (while the cNHEJ uses LIG4), is characterized by more extensive end resection, and frequently involves regions of microhomology (Simsek et al., 2011).

Topologic constraints and their role in chromosomal rearrangements

To be successfully joined by the NHEJ pathway, the two breakpoint need to physically interact and therefore it is not surprising that topologic constrains greatly affect the frequency and nature of chromosomal rearrangements. The ability of chromatin to move about the nucleus influences DSBs clustering, partner search, and synapsis during translocations. Two alternative models have been proposed: in the "breakage first" model, DSBs ends are thought to undergo extensive partner search (Aten et al., 2004), while in the "contact first" model DSBs remain positionally stable and can only join to other ends that are already in close proximity at the time they are generated (Soutoglou et al., 2007). This model is supported by elegant *in vivo* imaging studies by Misteli and colleagues who successfully tracked DSBs induced by the I-SceI endonuclease (Roukos et al., 2013). Consistent with this model are also early imaging studies showing that the fusion partners of BCR-ABL1 (Lukasova et al., 1997) and PML-RARα (Neves et al., 1999) are normally found near each other in the nucleus healthy bone marrow cells prior to formation of the translocation. Strikingly, one report suggests that where the DSBs are positioned in the

nuclear compartment can result in a change in mobility and repair pathway choice (Lemaitre et al., 2014). One interesting related idea is that transcriptional factories can bring DSBs together (clustering) where partner search can occur in a more permissive environment (McCord and Balajee, 2018).

The advent of high throughput sequencing and chromosome conformation capture technique (Hi-C) has allowed the systematic identification of long-range DNA interactions and the study of their effect on chromosomal rearrangements. In an elegant study in mouse pro-B cells, Dekker and colleagues showed that intra-chromosomal and inter-chromosomal long-range interactions detected by Hi-C correlate with increased frequency of intra- and interchromosomal rearrangements, respectively (Zhang et al., 2012).

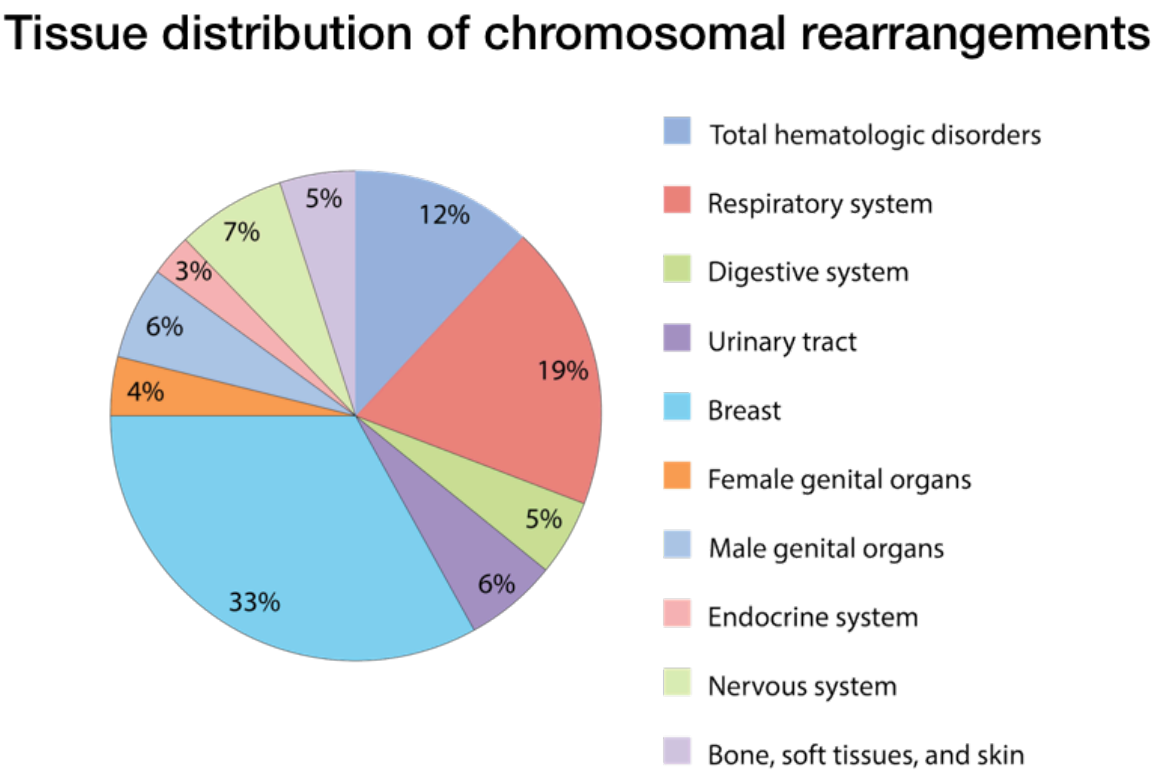


Figure 3. Tissue distribution of gene fusions identified as of 2018

The distribution of chromosomal rearrangements is spread throughout various tissues. Before high throughput sequencing, most were described in hematological disorders due to the limitations of detection. In recent years, many more fusions have been found in non-hematological tissues such as the breast due to the increasing amount of patient tumor sequencing (Source: (Mitelman, 2018)).

Consequences of chromosomal rearrangements

Chromosomal rearrangements can have profound biological consequences and contribute to tumor development through several mechanisms. Among them, the generation of in frame gene fusions with oncogenic properties is arguably the best characterized and a major focus of this thesis.

A variety of molecular mechanisms through which fusion proteins promote tumorigenesis have been identified **(Figure 4)**. For example, in oncogenic gene fusions involving kinases, kinase activity is commonly increased due to constitutive dimerization/oligomerization mediated by the fusion partner (EML4-ALK and BCR-ABL1 are classic examples). Increased expression levels of the of the fusion protein mediated by regulatory elements provided by one of the fusion partner can also play an important role as exemplified by gene fusions involving the ALK kinase, that is normally expressed at undetectable levels in most adult tissues (Lin et al., 2017). Fusion events can also result in constitutive activation if they lead to loss or disruption of an inhibitory domain. This alteration is typified by many fusions involving the BRAF oncogene, in which the fusion event typically results in deletion of its N-terminal autoregulatory domain (Ross et al., 2016). Fusion proteins often disrupt normal signaling pathways promoting proliferation, blocking differentiation, or suppressing apoptosis (Turner and Alexander, 2006). An example of this is NPM1-ALK in ALCL which drives STAT3 signaling as well other pro-survival and proliferation pathways (Zhang et al., 2002). Another major class of oncogenic gene fusions involves chromatin modifiers or transcription factors. In this case, altered gene regulation directly at the transcriptional level is the main mechanism through which these oncoprotein drive tumorigenesis (Faber et al., 2009; Krivtsov et al., 2006).

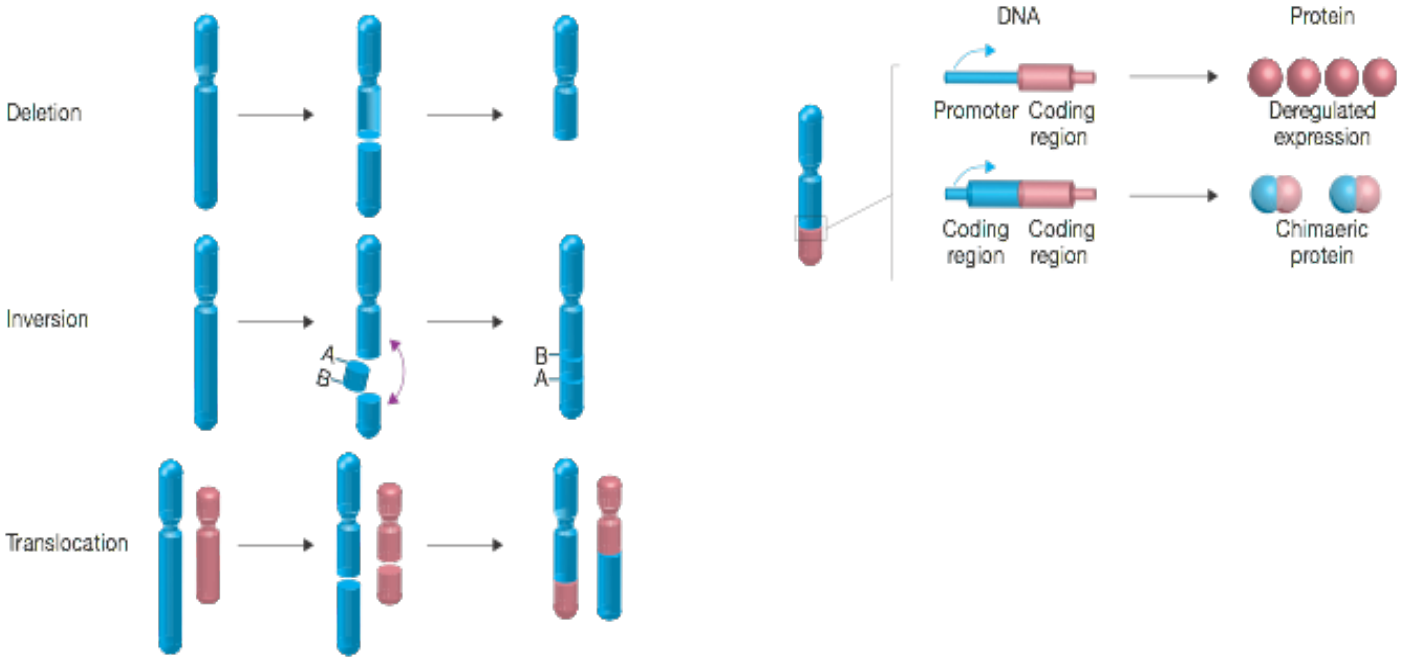


Figure 4. Types and consequences of chromosomal rearrangements

Deletions, inversions, and translocations can result in gene fusions. The two main molecular consequences are deregulated expression of coding genes due to control of different regulatory elements or generation of novel gene fusion protein with different properties then the wild type fusion partners (Reproduced with permission from: (Roukos and Misteli, 2014)).

An often-overlooked side effect of chromosomal rearrangements resulting the generation of a gene fusion is that they also result in the loss of one wild type allele of each of the genes involved in the rearrangement. This can be important if one the two genes is an haploinsufficient tumor suppressor. For example, reduced levels of wild type NPM1 in NPM1-ALK-positive lymphomas has been proposed to contribute to tumorigenesis, synergizing with constitutive signaling by the ALK (Mduff et al., 2011; Sportoletti et al., 2008). Similarly, in acute promyelocytic leukemias driven by the PML-RARα gene fusions, reduced dosage of PML (a tumor suppressor gene) seems to contribute to transformation independently from the transcriptional effects mediated by RARα (de The et al., 2017; Melnick and Licht, 1999).

In some cases, fusion oncogenes may promote tumorigenesis by acting as dominant negative alleles. For example, the AML1/ETO fusion results in downregulation of Acute myeloid leukemia 1 (AML1) target genes due to ETO recruiting corepressors to AML1 binding sites (Melnick et al., 2000). It has also been reported that the NPM1-ALK fusion protein heterodimerizes with wild type NPM1, leading to an additional reduction of the total NPM1 levels (Ceccon et al., 2016).

In the previous examples, fusion proteins mostly exert their function through a gain or loss of function mechanism, in some cases the fusion protein may have acquired neomorphic properties. For example, in acute megakaryoblastic leukemia (AMKL), the ETO2-GLIS2 fusion results in an altered DNA binding pattern which results in a unique transcriptional pattern that helps drive tumorigenesis (Wheat and Steidl, 2017).

Finally, some rearrangements drive tumorigenesis by causing aberrant expression of an otherwise wild-type proto-oncogene. A classic example is the Eμ-Myc translocation found in Burkitt's lymphoma where the IgH immunoglobulin enhancer drives oncogenic levels of Myc expression. Another example involves BCL2 translocations in follicular

lymphomas where overexpression of BCL2 suppresses apoptosis (Lackraj et al., 2018; Rowley, 2008). In addition, recent work has shown that rearrangements disrupting topologically associating domains (TADs) can have profound effects on gene expression (Lupianez et al., 2015). Whether such events can act as drivers in human cancers remains to be demonstrated.

Experimental approaches to modeling gene fusions

The identification of the first oncogenic gene fusion spurred intense research aimed at understanding how these genetic abnormalities promote tumorigenesis. Earlier studies took advantage of available patient-derived cancer cell lines harboring the chromosomal rearrangement of interest or were based on the ectopic expression of complementary DNAs (cDNAs) encoding for the gene fusion.

While these approaches provided fundamental insights into gene fusion biology (Cox and Der, 1994; Lu et al., 2017), the need for more sophisticated and genetically clean *in vivo* approaches quickly became apparent. Here I will briefly discuss the various methods that have been used to model and study cancer associated gene fusions. These approaches include *ex-vivo* manipulation, generation of transgenic mice, knock in strategies, and the use of recombinases to engineer the underlying chromosomal rearrangement.

Modeling gene fusions with transgenes

The generation of transgenic mice by injecting fertilized oocytes with exogenous DNA containing the gene of interest has been an established protocol since the early 1980s (Hofker and Breuer, 1998). Using this method, several copies of the transgene integrate randomly into the host genome and multiple ‘founders’ need to be screened to select the ones with the desired transgene expression pattern. To impart expression specificity, tissue specific promoters can be used to limit expression of the transgene to the cell or tissue of interest.

The earliest studies of transgenic mouse models provided fundamental insights into tumor biology. The Eμ-Myc strain, in which a MYC transgene was placed under the control of the Immunoglobulin heavy chain enhancer (Eμ) to drive its expression in B cell progenitors, was the first transgenic model of a chromosomal rearrangement (Adams et al., 1985). In another study, Huettner and colleagues showed that transgenic overexpression of BCR-ABL1 fusion was critical for tumor initiation. Using a tetracycline inducible promoter of BCR-ABL1, sustained expression of BCR-ABL1 was required for tumor maintenance as loss of expression results in tumor regression (Huettner et al., 2000; Ruggero and Rabbitts, 2015). Some patients with T cell acute lymphoblastic leukemia (ALL) have overexpression of LIM-domain-only proteins 2 (LMO2), a master regulator in hematopoiesis (Chambers and Rabbitts, 2015), due to aberrant transcriptional expression caused by LMO2 – T cell receptor (TCR) rearrangements. To identify the cell of origin, McCormack and colleagues showed in a mouse model that thymus specific transgene overexpression of LMO2 resulted in a block in differentiation and increased self-renewal in immature thymocytes, suggesting a potential cell of origin (McCormack et al., 2010).

Though these transgenic mouse models have provided better understanding of fusion biology, they are not without limitations. For example, mice that were generated to mimic the t(8;14) rearrangement found in Burkitt's lymphomas, develop pro-pre-B cell lymphomas that are histologically and biologically quite different from the human counterpart (Adams et al., 1985). On a practical note, the process of generating a transgenic mouse to model individual gene fusions is a costly and laborious process.

Hematopoietic Stem Cell Transplantation Models

Bone marrow ex vivo transplantation models

To overcome the limitations of transgenic mouse, *ex vivo* approaches were developed where hematopoietic stem cells can be experimentally manipulated with viral

vectors to deliver transgenes. These modified cells can then be transplanted back into lethally irradiated mice to allow for engraftment and long-term reconstitution of the bone marrow (Hemann, 2015). This non-germline approach has greatly facilitated dissecting the roles of gene fusions in hematologic malignancies.

Bone marrow transplantation strategies were first used to ectopically express the BCR-ABL1 gene fusion and accurately model the progression of chronic myelogenous leukemia (Daley et al., 1990; Kelliher et al., 1990). However, such a strategy is not always successful, as exemplified by attempts to model B cell precursor acute lymphoblastic leukemias driven by ETV6-RUNX1, in which only some features of the disease were observed (Bernardin et al., 2002).

Gene targeting strategies

The two strategies discussed so far are based on the ectopic expression of a cDNA encoding for the gene fusion product. As such, they share specific limitations that reduce their usefulness. First, since the cDNA is rarely expressed at levels comparable to what is observed in human tumors, unwanted consequences may be observed. For example, most of the data used to implicate oncogenic fusions are the result of overexpression of fusion cDNA. One possible interpretation is that the phenotypes observed with protein overexpression are potential artifacts. An equally important limitation is that neither approach models the underlying chromosomal rearrangement, thus failing to account for the contribution of reciprocal gene fusions, loss of one copy of the wild type alleles, and modifications of the local chromatin landscape.

The development of gene targeting by homologous recombination (Thomas and Capecchi, 1987) provided a series of novel strategies to more faithfully model gene fusion events and overcome some of these limitations. These can be grouped into two broad strategies: knock-in approaches, and Cre-lox based chromosomal engineering.

Knock-in strategies

The knock in approach can be used to model cDNA gene fusion expression in a ubiquitously expressed locus. For example, in a mouse model of synovial sarcoma, a SS18-SSX2 fusion transgene was integrated into the ROSA26 locus that contained a lox-STOP-lox cassette and allowed spatial induction of the gene fusion. When coupled with a tissue specific Cre, transgene induction caused mice to develop synovial-type sarcoma (Haldar et al., 2007). This conditional approach is especially useful for gene fusions whose expression may be embryonically lethal.

To recapitulate gene fusion expression in the native locus, one can knock-in the fusion cDNA of the 3' partner into the 5' partner gene locus. This keeps the gene fusion expression under the control of the endogenous promoter and models the loss of the wild type allele of the 5' partner. The first publication outlining this approach inserted the fusion specific portion of the mixed-lineage leukemia (MLL) gene into the AF9 intron locus in mouse embryonic stem cells (ESCs) (Corral et al., 1996) and formed acute leukemias. Building upon the fusion partner knock-in approach and Cre-lox system, one can create a conditional mouse translocation mimic. The cDNA, flanked by *loxP* sites, is inverted upon knock-in to the 5' partner gene to prevent transcription, but can be activated by Cre mediated inversion. This approach was used to model the EWS-ERG gene fusion in hematopoietic cells (Forster et al., 2005).

Compared to ectopic expression, these gene targeting approaches consider the contribution of wild type allele loss and gene fusion expression regulation by the 5' fusion partner in tumorigenesis. Expression of these gene fusions could be spatially controlled with lox-STOP-lox cassettes. However, these knock-in models do not address the contribution of the reciprocal fusion product or address changes in the chromatin structure. For example, recent work showed that the BCR-ABL1 cDNA fusion, when knocked into the native locus of BCR, fails to induce neoplasia (Foley et al., 2013).

Cre-lox-based methods

The Cre/loxP system has been extensively used to generate conditional knockout and knock in alleles, but it has proven useful also as a tool to engineer chromosomal rearrangements (Zheng et al., 2000). The basic idea, pioneered by Allan Bradley and colleagues, is to place individual loxP sites at the desired breakpoints and then let the Cre recombinase drive the formation of the rearrangement **(Figure 5)**. For example, by placing two loxP sites on the same chromosome, it is possible to induce the deletion of the intervening sequence—if the loxP sites are in the same orientation—or its inversion, if the two sites are inserted with opposite orientation. Analogously, loxP sites located on two different chromosomes can be used to generate reciprocal translocations.

This strategy was initially applied *in vitro* in murine ESCs (Ramirez-Solis et al., 1995; Smith et al., 1995; Van Deursen et al., 1995), and later adapted to an *in vivo* setting. The first "translocator" mouse generated using this approach was designed to model the t(9;11) chromosomal translocation which produces the MLL-AF9 gene fusion and is frequently observed in human mixed lineage leukemias (MLL) (Collins et al., 2000). Interestingly, although the authors were able to demonstrate generation of the rearrangement and expression of the fusion transcript in the brain, the mice did not develop leukemias as Cre expression could not be induced in the hematopoietic compartment. The same group later used a similar approach to generate the Mll-Enl gene fusion, and this time the mice developed myeloid leukemias at high penetrance and with short latency (Forster et al., 2003). By driving Cre-expression in different hematopoietic lineages, this model was later used to show that leukemias only arise if the translocation occurs in the stem cells or early progenitors, but not in more differentiated cells (Cano et al., 2008).

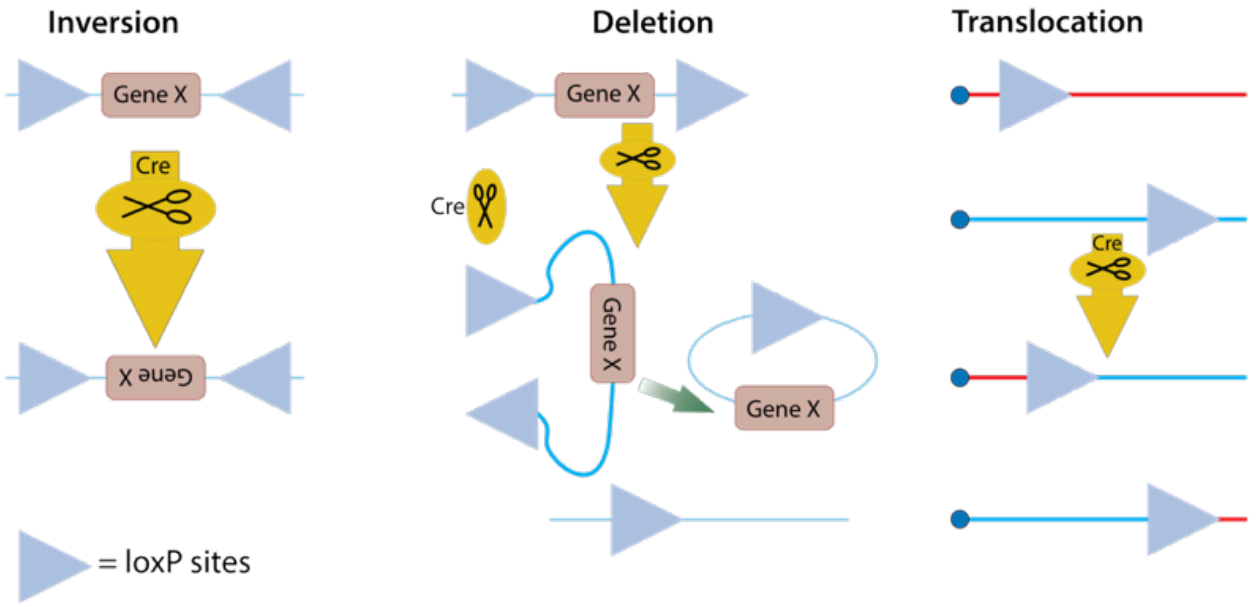


Figure 5. The use of Cre Recombinase to model chromosomal rearrangements

LoxP sites can be placed on specific genomic loci to model different types of rearrangements. Placing LoxP sites in the opposite orientation will result in an inversion of the gene locus while LoxP sites in the same orientation will result in a deletion upon Cre recombination. For modeling translocations, LoxP sites placed on nonhomologous chromosomes that face the same relative orientation will cause reciprocal translocations. This can also be used to cause loss of chromosomes by generating acentric/dicentric if LoxP sites are placed in the opposite orientation.

In principle, this approach is the one that most faithfully recapitulates the genetics of cancers driven by chromosomal rearrangements, as the gene fusion is expressed at appropriate levels and the reciprocal product is also modeled. However, several limitations have prevented its widespread use. First, the generation of these mice is technically challenging and time consuming, requiring multiple gene targeting events and complex breeding schemes to introduce the desired Cre transgene. Second, because the efficiency with which the rearrangement is produced upon Cre expression decreases as the distance between the loxP sites increases (Zheng et al., 2000), mice might fail to develop the desired tumors. Finally, not every gene fusion found in humans can be engineered using this strategy because the two genes involved in the rearrangement—or their relative orientation—may not be conserved in mice. A typical example is represented by the t(15;17) translocation that in humans generates the PML-RARα fusion and causes acute promyelocytic leukemia (Alcalay et al., 1991; Borrow et al., 1990; de The et al., 1990). The two genes are conserved in mice, but because of their relative orientation modeling the translocation would result in the generation of a dicentric chromosome **(Figure 6)**. However, one group was able to overcome this in a model of the *Pax3-Foxo1* translocation in mouse myoblasts by inverting the *Foxo1* locus with Cre recombination followed by CRISPR mediated translocation (Lagutina et al., 2015).

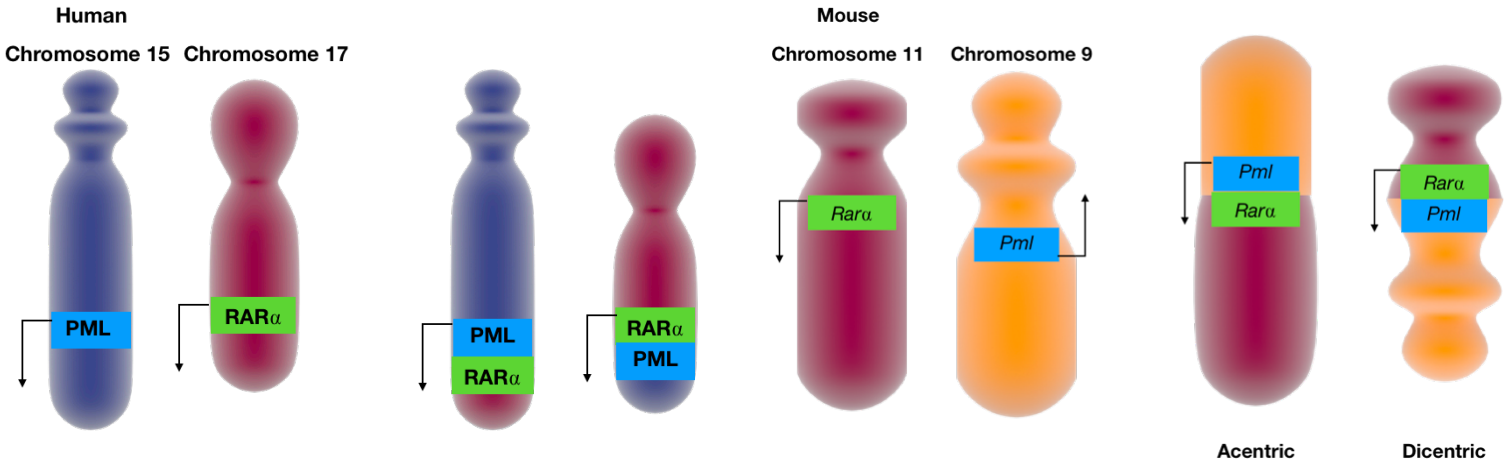


Figure 6. Transcriptional orientation and formation of chromosomal rearrangements

(Left) Shown here are the human chromosomes 15 and 17 containing PML and RARα genes respectively, and the reciprocal translocation products. Note the transcriptional orientation is preserved in the gene fusion. (Right) In the mouse, the homologous genes have opposite transcriptional orientations. To generate the correct gene fusion would result in an unstable structures such as acentric and dicentric chromosomes.
(Source: sommesault182)

The genome editing revolution

Natural or directed processes that produce DSBs are known to be able to generate a spectrum of mutations. DSBs are considered 'editogenic' as the DNA repair pathway can result in the change of genetic information (Jasin and Haber, 2016). In addition, the presence of two DSBs can results in inversions, deletions, and translocations though the ability to model this was hampered by lack of the appropriate tools. Ideally, a quick somatic *de novo* generation of gene fusions approach would address some of the shortcomings of previous translocation models.

Restriction enzymes

The first site-specific enzymes discovered that caused DSBs were restriction enzymes. Though used for molecular *in vitro* applications, the short recognition motif of most restriction enzymes precluded single site editing of the genome. The first tool for editing a specific locus were a subset of restriction enzymes called meganucleases. The I-SceI meganuclease, with its longer DNA recognition motif, cut site could be placed into the genome for site specific editing without the risk of off targets. This pioneering gene editing study showed that I-SceI could induce a site specific DSB in a cell and be used to repair a mutated green fluorescent protein (GFP) template (Rouet et al., 1994a, b). This was the first evidence that creating an exogenous DSB was able to modify a native genomic locus though the inability to modify the recognition site limited its application.

Zinc finger nucleases

Zinc fingers (ZFs) are naturally occurring motifs for DNA binding often found in transcription factors (Beerli et al., 1998), which could be rationally designed to bind a specific DNA sequence. To achieve this, Urnov and colleagues used an iterative selection process, to screen for sequence motifs that bind a sequence (Urnov et al., 2010). To adapt this for genome editing the non-specific Fok1 endonuclease was attached to ZFs (ZFNs) (Kim et al., 2009; Kim et al., 1996). However, as sequence binding had to be validated,

this process still requires laborious generation and screening of DNA binding motifs. Though labs were able to take advantage of ZFNs to edit several human genes such as CCR5, this was not widely adopted in research labs (Holt et al., 2010). On the clinical side though, this has several advantages as it has been shown to have no or minimal off targets and a long track record of safety in humans (Ando and Meyer, 2017).

Transcription activator-like effectors

Transcription activator-like effectors (TALEs) were discovered as secreted proteins from pathogenic plant bacteria (Cuculis and Schroeder, 2017), which were able to bind dsDNA and modulate host gene expression. The sequence binding is mediated by tandem repeats with each repeat's single nucleotide (nt) specificity determined by two amino acids motifs. Unlike ZFs, TALES provided set rules for design (Mussolino et al., 2011) and could be assembled with specific binding patterns (Bedell et al., 2012). By adding the Fok1 nuclease to TALEs (TALENs), this allowed site specific recruitment of nucleases. Besides cutting, one could attach proteins that alter DNA features, such as epigenetic marks (Maeder et al., 2013). Given the ability to design TALENs based on sequence alone, they were used more frequently to edit the genome (Bedell et al., 2012; Ding et al., 2012), despite the limitations of this method. For example, designing TALENs requires cloning multiple subunits, so generating a large library of TALENs is a labor-intensive task (Li et al., 2016b), even with the Golden Gate Cloning process (Cermak et al., 2015). To avoid the process of assemblying a modular protein for each target, would require the development of a gene editing tool that directly recruited effector proteins with sequence information alone.

Clustered Regularly Interspaced Short Palindromic Repeats

Brief History of CRISPR

The discovery of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems in bacteria and adaptation of CRISPR to use in mammals that made gene editing technology readily available to every lab equipped to do basic molecular biology (Barrangou and Doudna, 2016; Hsu et al., 2014; Wright et al., 2016).

The first encounter with CRISPR actually dates to the late '80s, when a region containing 5 direct nearly palindromic repeats of 29 nucleotides separated by 32 nucleotides non-repetitive spacers was identified downstream of the *iap* gene in *Escherichia coli* (Ishino et al., 1987; Nakata et al., 1989). The functions of this repeat array were to remain unknown until advances in sequencing and computational methods in the 90's led to the identification of similar short direct near palindromic repeats separated by constant-length non-repetitive spacers in many different species of bacteria and archaea (Bult et al., 1996; Groenen et al., 1993; Hoe et al., 1999; Masepohl et al., 1996; Mojica et al., 1995; Mojica et al., 1993). At the dawn of the 21st century, this large class of repeat was given the name 'CRISPR' and recognized as a single-family present in bacteria and archaea, but not in eukaryotes (Jansen et al., 2002a; Mojica et al., 2000).

These pioneering studies also revealed some peculiar features of CRISPR loci. First, most CRISPR loci were found to be flanked on one side by a several hundred base pairs long 'leader' sequence rich in single nucleotide repeats and with high AT content (Jansen et al., 2002b). Interesting, within a single CRISPR locus, the more distal a repeat was from the leader sequence, the more likely it was to harbor mutations (Jansen et al., 2002b). Second, a comparison of genes flanking CRISPR loci from different organisms led to the identification of several "CRISPR-associated genes" (*cas*) and to the hypothesis that they could be functionally linked to the formation and function of CRISPR loci (Jansen et al., 2002b; Mojica and Rodriguez-Valera, 2016).

The first indication of what such a function could be came when it was realized that some spacer sequences are derived from foreign genetic elements such as viruses, plasmids, or chromosomes from closely related species (Bolotin et al., 2005; Mojica et al., 2005). Crucially, Mojica also pointed out that: "*Although our current knowledge remains limited, multiple observations suggest that CRISPR could be involved in conferring specific immunity against foreign DNA…*"(Mojica et al., 2005), noticing in a given species presence of a spacer derived from a phage correlated with resistance to infection from the same phage, even when there was evidence that the phage could penetrate the cell (Mojica et al., 2005).

This correlative evidence was soon proven to underlie causality by Barrangou and colleagues, who showed that the development of resistance of *Sulfobolus solfataricus* to infection by the SIRV phage associated with the integration of spacers from the virus into a pre-existing CRISPR locus and that deletion of such spacers would restore sensitivity (Barrangou et al., 2007). Shortly thereafter, Marraffini and Sontheimer showed that CRISPR systems play an analogous role in protecting against the horizontal transfer of other mobile genetic elements, such as plasmids (Marraffini and Sontheimer, 2008).

Mechanistically, it was quickly discovered that the CRISPR array is transcribed and processed into short CRISPR related RNAs (called crRNAs) that are then used as 'guides' for cas proteins to cleave the DNA of the invading mobile DNA element (Brouns et al., 2008; Garneau et al., 2010). Evidence of how CRISPR was able to recognize targets for interference while avoiding self-targeting was first provided by the discovery of short conserved sequences named protospacer adjacent motifs (PAM) in CRISPR targets of *Streptococcus thermophilus*, and then found to be common to all CRISPR based systems (Mojica et al., 2009).

Stages of CRISPR-mediated adaptive immunity

The process of CRISPR-based immunity can be divided into three main stages: adaptation, maturation, and interference. In general terms, adaptation begins with identifying and processing foreign nucleic acids into spacer sequences followed by integration into a CRISPR array (Hille et al., 2018). These arrays provide the memory of prior infections which can then be recalled when specific CRISPR arrays are transcribed to produce long precursor crRNA (pre-crRNA). In the maturation phase, this is then processed to form mature crRNA which is then assembled into the interference machinery. During the interference phase, if there is a subsequent infection, then mature crRNA can be guided by an effector protein to cleave protospacer sequences found in foreign nucleic acids via complementary sequence binding **(Figure 7)**.

Survey of CRISPR systems

Briefly, CRISPR-Cas systems can be currently classified into two classes and divided into six types with additional subtypes. While all active CRISPR systems contain Cas1-2, which are involved in spacer integration into the CRISPR array, Class, Type, and Subtype are determined by the effector and accessory proteins under a unified system of classification. Class 1 CRISPR-Cas systems (types I, III, and IV) employ a multi-protein complex for interference, while Class 2 systems (type II, V, and VI) have a single unit as the effector, which is exemplified by Cas9 (Makarova et al., 2017a, b).

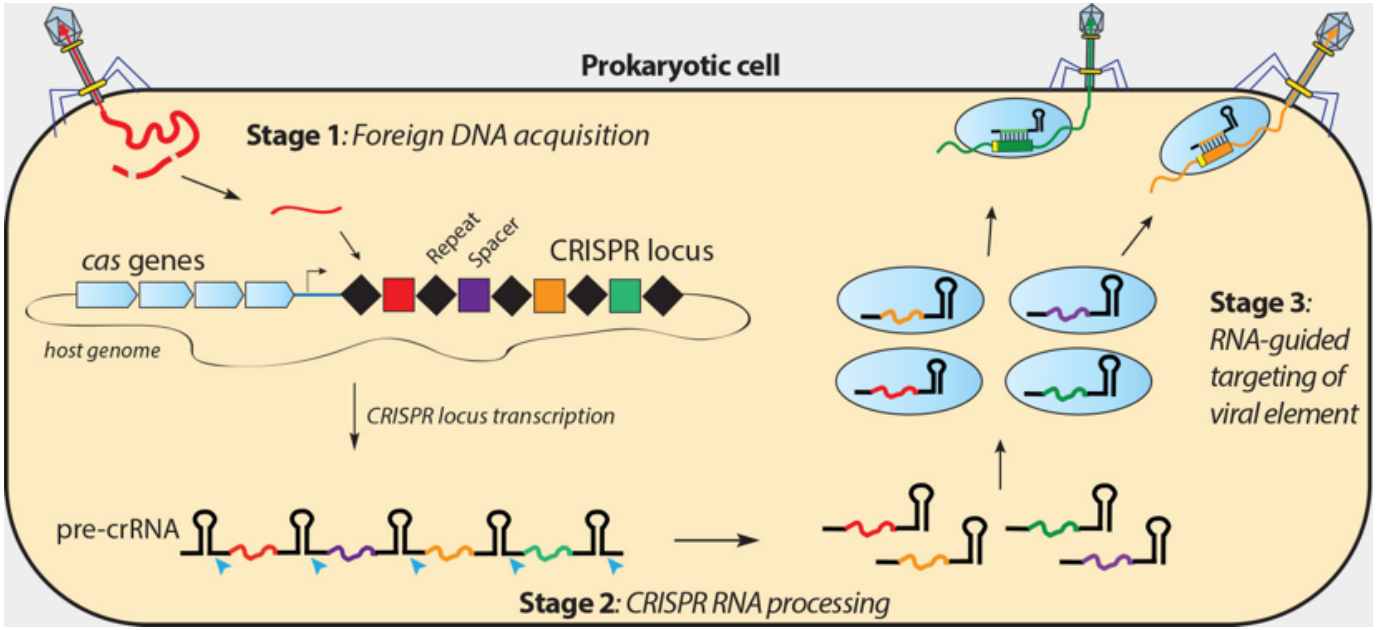


Figure 7. Diagram of CRISPR-Cas system

The Cas protein and the guide RNA act as an immunity system to defend against invading phage DNA. Foreign DNA is acquired and integrated as spacer regions into the CRISPR loci (Step 1). Pre-crRNA is transcribed from the CRISPR locus and is processed in mature crRNA (Step 2). Expression of cas effector genes results in the formation of the functional interference effector and crRNA complex for target interference. This can then be used to target invading DNA (Step 3).
(Reproduced with permission from Doudna Lab website: http://doudnalab.org/research_areas/crispr-systems/)

Spacer Acquisition

Cas1 and Cas2 are critical for spacer acquisition as it can act as an integrase of protospacer sequence into the CRISPR array (Koonin et al., 2017). Upon fragmentation of the DNA, the choice of spacer sequence for integration is influenced by the presence of a compatible PAM sequence. Spacer integration in type I-E systems, the best characterized occurs at the leader end of the CRISPR array where integrated host factor (IHF) bends the DNA. This allows the Cas1-2 bound protospacer sequence to nick and ligate to opposite ends of the repeat strand which results in the regeneration of the repeat for subsequent spacer acquisition. While in type I-A, spacer integration instead of using IHF, requires accessory proteins and leader-anchoring site motif on the CRISPR array (Heler et al., 2015).

crRNA biogenesis and maturation

Upon integration of spacer regions, transcription of pre-crRNA begins in the leader region and is subsequently processed into mature crRNA, which contain the spacer portion along with repeat elements, depending on the subtype, that are recognized by Cas proteins for target interference. In Class 1, the majority of type I and III systems typically use Cas6/Cas5 or homologous proteins, depending on the subtype, for cleavage of pre-crRNA into mature crRNA that typically remain bound to form part of the effector complex. With type IV, it has been postulated to have a similar mechanism though it still awaits experimental confirmation (Hille et al., 2018).

Class 2 systems differs significantly from Class 1 as the effector complex along with non-Cas proteins are directly involved in crRNA maturation. Type II and type V-B systems utilize *trans*-activating crRNA (tracrRNA) to form a duplex with pre-crRNA. This duplex is then recognized by an effector protein (Cas9 for type II-A) which is then processed by RNAse III into an intermediate step crRNA that is further processed to remove 5’ repeat sequences. This functional interference complex can then be used

CRISPR immunity. In type II-C, crRNA processing can proceed without RNAse III activity.

In type V and type VI systems, the effector proteins contain enzymatic activity for both crRNA processing and target interference without the need for a tracrRNA. In type V-A, Cas12a recognizes the pre-crRNA hairpin structure followed by cleavage of repeat to generate crRNA that is processed further by a yet identified RNAse to generate a functional complex. For type VI-A, Cas13a recognizes repeat sequence of the pre-crRNA for processing though this is not an absolute requirement for CRIPSR interference (East-Seletsky et al., 2017; Hille et al., 2018).

CRISPR Interference pathways

CRISPR interference involves guiding crRNAs to cleave foreign nucleic acids, though to avoid self-cleavage most CRISPR types have evolved a way to discriminate self and non-self through the presence of PAM sequences on their targets (Mojica et al., 2009; Shah et al., 2013).

With Class I systems, type I uses the CRISPR-associated complex for antiviral defense (Cascade) for target DNA recognition and by having the Cas3 gene for DNase activity. This multi-subunit Cascade complex can contain the Cas 5,7,8 and 11 gene families, which determines their subtypes (A-F, U) (Makarova et al., 2015). In type I-E, the best characterized, the Cascade binds to Cas6 bound crRNA where recognition of the PAM sequence results in DNA unwinding and crRNA binding. This allows Cas3 to be recruited for DNA target cleavage. For type III systems, there is a multi-subunit complex called Csm for III-A and Cmr for III-B which resembles the Cascade complex though in contrast to type I, they can target both DNA and RNA. In brief and with each subtype composed of different subunits, the complex assembles with Cas7, Cas5, Cas10, and Cas11 proteins along with mature crRNA where it binds to nascent target RNA transcript in a crRNA dependent manner. RNA and DNA cleavage then occur with Cas7 and Cas10 respectively. Cas10 can also trigger Csm6 RNAse to promote non-specific RNA

degradation. Type IV CRISPR mediated interference has not been fully characterized to date.

With Class 2 CRISPR-Cas systems, in contrast to Class 1 systems, a single effector protein accomplishes interference. The best characterized ones are type II, type V-A, and type VI with Cas9, Cas12a, and Cas13. In type II systems, the prototypical member is Cas9 which uses a crRNA-tracrRNA duplex for target interference. In brief, Cas9 with a mature duplex searches for compatible PAM sequences followed by crRNA binding to target DNA through sequence complementarity. Cas9 then generates a blunt double-strand break 3 bp upstream of the PAM (Garneau et al., 2010). For type V systems, Cas12a of type V-A requires only crRNA which upon PAM recognition and crRNA-DNA base pairing results in staggered double-strand breaks with 5 to 7 nucleotide overhangs upstream from the PAM. With type VI, Cas13 (C2c2) cleaves specific ssRNA upon crRNA complementary target binding without the need for tracrRNA (Abudayyeh et al., 2016), followed by non-specific RNA cleavage as well.

Use of CRISPR in gene editing

Even though CRISPR immunity evolved in Eubacteria and Archaea, the principles could be applied widely. This ability of a simple Watson-Crick pairing to mediate site-specific nuclease activity provided a simpler and programmable method of gene editing (Ventura and Dow, 2018). Needing only a ribonucleic protein complex that can be readily and economically synthesized, this was a game changer in genome editing.

The utility of type II systems

Given the diversity of CRISPR-Cas systems, why the research community coalesced onto the type II CRISPR-Cas system could be attributed to the fact that a single effector protein was responsible for targeted DNA cleavage (Lander, 2016). This made it ideal as a site-specific nuclease which could be expressed as a single transcriptional unit. Despite this knowledge, it wasn't until the discovery of tracrRNA transcripts in high throughput sequencing of Streptococcus pyogenes and its role in crRNA processing did the mechanism of CRISPR immunity become clear (Deltcheva et al., 2011). With this final piece, all the pieces were in place to adapt Cas9 for genome engineering.

Type II-A Cas9 is the most recognizable and studied Cas protein, derived from *Streptococcus pyogenes* (SpCas9), which is a single bilobal protein that contains two nuclease domains: RuvC near the amino terminus, and HNH near the middle of the protein (Gasiunas et al., 2012). The guide RNA consists of a tracrRNA, which is a scaffold of defined sequence, along with the mature crRNA containing the unique spacer. There are also other Cas9 alternatives, such as type II-C Cas9 found in *Staphylococcus aureus* (SaCas9), which is a smaller bilobal protein with DNA cleavage activity with different PAM recognition conditions (Jiang and Doudna, 2017; Ran et al., 2015) and no need for RNAse III processing step. Meanwhile, type II-B uses Cas9 for DNA targeting that leaves staggered DNA overhangs (Chen et al., 2017a).

Cas9 as a molecular tool

That system was adapted to work in other cell types and simplified the system from the original tracrRNA and crRNA to a single continuous guide RNA (Jinek et al., 2013). The single Cas9 protein is complexed with a single guide RNA (sgRNA) based on an 18-26 base pair (bp) recognition sequence (Fu et al., 2014), with nuclease activity only occurring when a sequence is adjacent to a broad protospacer motif (PAM) of NGG, **(Figure 8)** which is mediated by the HNH and RuvC domains. The seed sequence, defined as the first 10-20 bps adjacent to the PAM, is critical for efficient gene editing as alterations to the sequence results in loss of activity. The sgRNA structure contains several stem loops for Cas9 binding and allow the formation of a functional effector complex. that will survey DNA for the target sequence. Based on biophysical and crystallography studies with *in vitro* DNA templates, the key steps appear to be: (1) sgRNA binding switches Cas9 to an active form, (2) PAM identification and partial DNA unwinding, (3) RNA strand invasion and unwinding of the DNA, and (4) formation of the RNA-DNA hetero duplex followed by Cas9 conformational changes to allow DNA cleavage (Jiang and Doudna, 2017; Jinek et al., 2014; Raper et al., 2018; Singh et al., 2016). This ease of designing targets based on gene sequence alone made it readily adaptable for gene editing (Hsu et al., 2013).

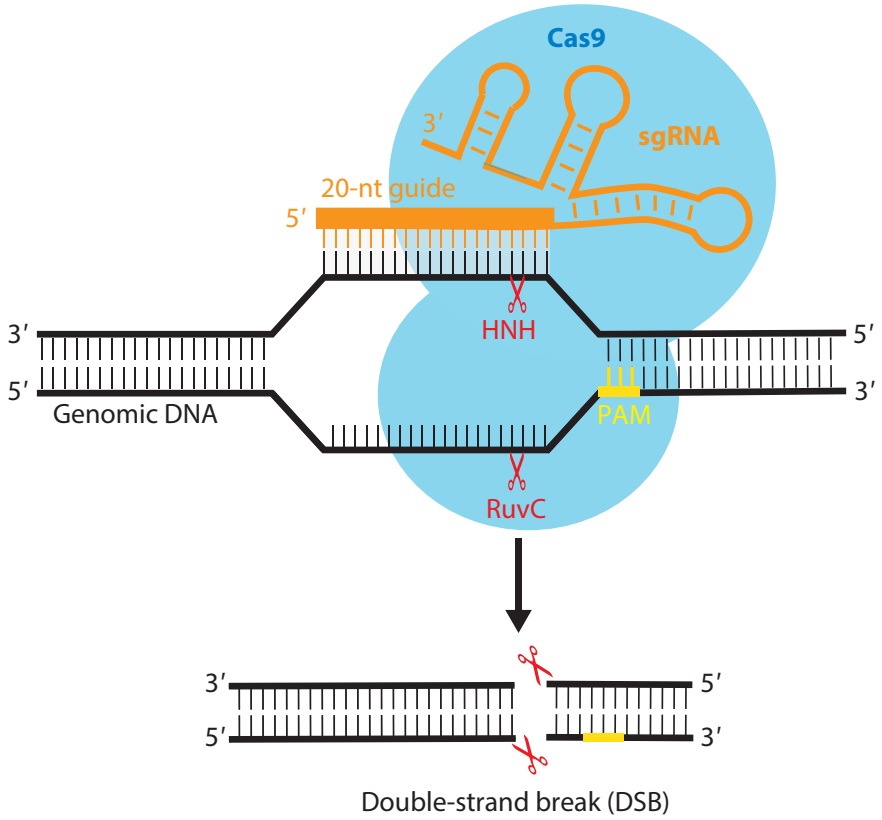


Figure 8. Cas9 mediated double strand break formation

An sgRNA bound to Cas9 will recognize a target sequence adjacent to a PAM motif through its 20 nt recognition site. Upon sgRNA binding, the DNA is unraveled and the HNH and RuvC nuclease domains of Cas9 mediate double strand cut of the DNA.
(Reproduced with permission from (Jiang and Doudna, 2017))

Delivery of CRISPR components for gene editing

Successful gene editing has two main components: (1) targeting of nucleases and (2) robust on-target delivery. CRISPR technology represents the latest innovation in targeted nucleases, yet delivery of CRISPR remains problematic.

CRISPR was first used for modeling indels, similar to previous work with site-specific nucleases. Instead of indels caused by NHEJ, specific mutations could be modeled via HDR by providing a repair template with the desired mutation along with a silent mutation of the target site to prevent re-cleaving (Yang et al., 2013). However, useful this tool has proven to be, delivery of CRISPR-Cas9 to the right cell remains a challenging issue.

CRISPR components (Cas9 & sgRNA) can be delivered through non-viral methods such as nanoparticles or cell penetrating peptides. Additionally, physical approaches such as electroporation have been effective for delivery for certain cell types. For viral delivery, several classes of viruses have been used, including adenoviruses, adeno-associated virus, and retroviruses (Wang et al., 2017).

In vitro Delivery Methods

Transfection

CRISPR-Cas9 was first delivered as plasmid DNA through lipid-based transfection reagents into tissue culture cell lines (Mali et al., 2013). This is often the simplest and most cost-effective method of delivery. Though sufficient for *in vitro* cell lines, delivery of plasmid DNA is thought to be less efficient due to potential toxicity and off-target effects (Liang et al., 2015; Peng et al., 2016). To address this, one can work directly with Cas9 and sgRNA, which involves a protein and RNA component (Kim et al., 2014; Kouranova et al., 2016). Cas9 can be delivered as an *in vitro* RNA transcript or a recombinant protein bound to the RNA portion of a ribonucleoprotein (RNP) complex. This has the benefit of transient expression when only the initial editing event is needed while minimizing the chances of

off target effects. However, approach is not scalable or cost effective when compared to DNA-based methods.

Nucleofection

Delivery methods based on cell deformation and electroporation utilizing the Amaxa Nucleofector or Neon Transfection machine are an attractive alternative (Kurosawa et al., 2012). Many cell types, including hematopoietic cells and human pre-B cells, are resistant to transfection-based reagents making electroporation-based transfection methods ideal (Han et al., 2015). Using this approach, CRISPR-Cas9 was readily delivered *in vitro* as DNA into human embryonic cell lines (Hsu et al., 2013). More recently, Cas9 RNPs were delivered to primary hematopoietic human CD4+ T cells (Hultquist et al., 2016). Though many cell types are amenable to this form of delivery, some cells such as primary stem cells have remained stubbornly resistant.

Viral Vector Delivery

Viral vectors have traditionally been used to deliver gene cargo to wide variety of cell types in a highly efficient manner. Depending on the application, viral vectors can provide transient or stable expression depending on whether the vector integrates into the genome. Though non-integrating viruses such as Adenoviruses (AV) and adeno associated viruses (AAV) can be used to delivery CRISPR-Cas9, they are primarily used in tissue where cells rarely divide to avoid diluting out CRISPR-Cas9. Meanwhile, viral vectors can be utilized to mark individual cells (i.e. cellular barcodes) in a genome-scale screening by integrating lentiviral CRISPR vectors (Sanjana et al., 2014; Shalem et al., 2014). Sequencing of transduced cells can be used to track sgRNAs in the starting population and to look for enrichment or depletion during the screen process. For *in vitro* applications, such as working with cell lines, integrating viral vectors are often used due to the ease of preparation of virus and transduction of target cells.

In vivo delivery Methods

Direct delivery of CRISPR payload into primary tissues

In cell lines, delivery of cargo is relatively straightforward, but this is not feasible or effective in the context of *in vivo* manipulation (Glass et al., 2018). One approach involves directly delivering CRISPR-Cas9 into cells of interest. For example, researchers were able to deliver Cas9 RNPs encapsulated in lipid particles directly into the ear of mice correcting genetic model of hearing loss (Gao et al., 2018a). Another approach is the hydrodynamic delivery of CRISPR plasmids into the liver (Yin et al., 2014). Recent work has shown that local delivery of Cas9 RNPs can also be achieved by injection into the muscle or brain of mice (Lee et al., 2017; Staahl et al., 2017) though this is not possible for all cell types.

Ex vivo manipulation of target cells

Though certain tissues preclude direct delivery of CRISPR due to their location, some cell types can be extracted, modified, and transplanted back into recipient organisms. For example, *ex vivo* delivery of CRISPR was achieved through electroporation using plasmid DNA in hematopoietic system (Song, 2017). One recent report showed successful ablation of the CCR5 gene, important in Human Immunodeficiency Virus (HIV) infection in human stem and progenitor cells (HSPCs) using the electroporation method. The report also showed that tandem guide delivery resulted in more consistent ablation, suggesting that simultaneous delivery of all the components is more efficient (Mandal et al., 2014).

Over the past few years, there have been improvements to delivery of CRISPR components, including the use of Cas9 RNP complexes (Gundry et al., 2016; Schumann et al., 2015), which have achieved similar or superior results to previous plasmid-based *ex vivo* manipulation (Song, 2017). However, each RNP must be manually prepared, which leads to restrictions in relation to large-scale screening where targets may number in the thousands. Despite these recent advances, delivery of the CRISPR-Cas9 complex

remains one of the greatest challenges to widespread adoption of CRISPR in mouse models.

Direct viral delivery into primary cells

Viral packages have been used to deliver genes of interest into cells for decades (Miller et al., 1984). With successful transduction, these modified cells have robust long-term expression of the delivered construct, either for therapeutic or research purposes. A key advantage of using viral delivery is in the ability of these constructs to cross the blood brain barrier, which precludes other lipid-based delivery methods. For gene editing in the brain, viral vectors remain the most reliable method of delivering CRISPR-Cas9 (Lentz et al., 2012).

Though viral integration may be desirable in some applications, long term expression of Cas9 may result in off target effects (Pattanayak et al., 2013). To avoid this, one group was able to construct an integrase deficient Cas9 lentivirus (Ortinski et al., 2017). Alternatively, non-integrating AAVs are one class of viruses that have been extensively used to deliver CRISPR to target cells while maintaining a high safety profile. A recent publication showed AAV could bring Cas9 to target the Duchene Muscular Dystrophy (DMD) gene mutation through excision of DMD exon 23 in muscle stem cells (Tabebordbar et al., 2016). Though AAVs can carry the least payload compared to other viruses, smaller Cas9 orthologues have been identified and adapted for AAV delivery for genome editing (Lau and Suh, 2017).

Applications of CRISPR

With CRISPR's ability to induce DSBs, this can be used to rapidly model known gene alterations (Cong et al., 2013; Ding et al., 2013; Li et al., 2013a). Cas9's ability to induce mutations stems from the fact that Cas9 can continue to cut or recut the region until the recognition sequence is no longer present. Hence, Cas9 promotes gene editing as this can only occur if the repair by the endogenous repair system disrupts the

recognition site (van Overbeek et al., 2016) **(Figure 9)**. This results in indels generated by the error-prone NHEJ. The locations of these indels can result in protein loss of function if it generates a truncated protein or disrupts key functional domains. If provided a repair template with nonsynonymous mutation in the recognition site to prevent re-cutting, then HDR (Yang et al., 2013) can be used for modeling specific mutations.

The initial CRISPR experiments were performed *in vitro* (Cho et al., 2013) and could be used to engineer specific mutations. To model mutations in an *in vivo* setting, it could be done in two ways: — germline, which is inherited, or somatic, where alterations cannot be passed down to future progeny. Owing to ethical concerns regarding human experimentation with CRISPR in the germline, most of the work has been done in model organisms (Ma and Liu, 2015).

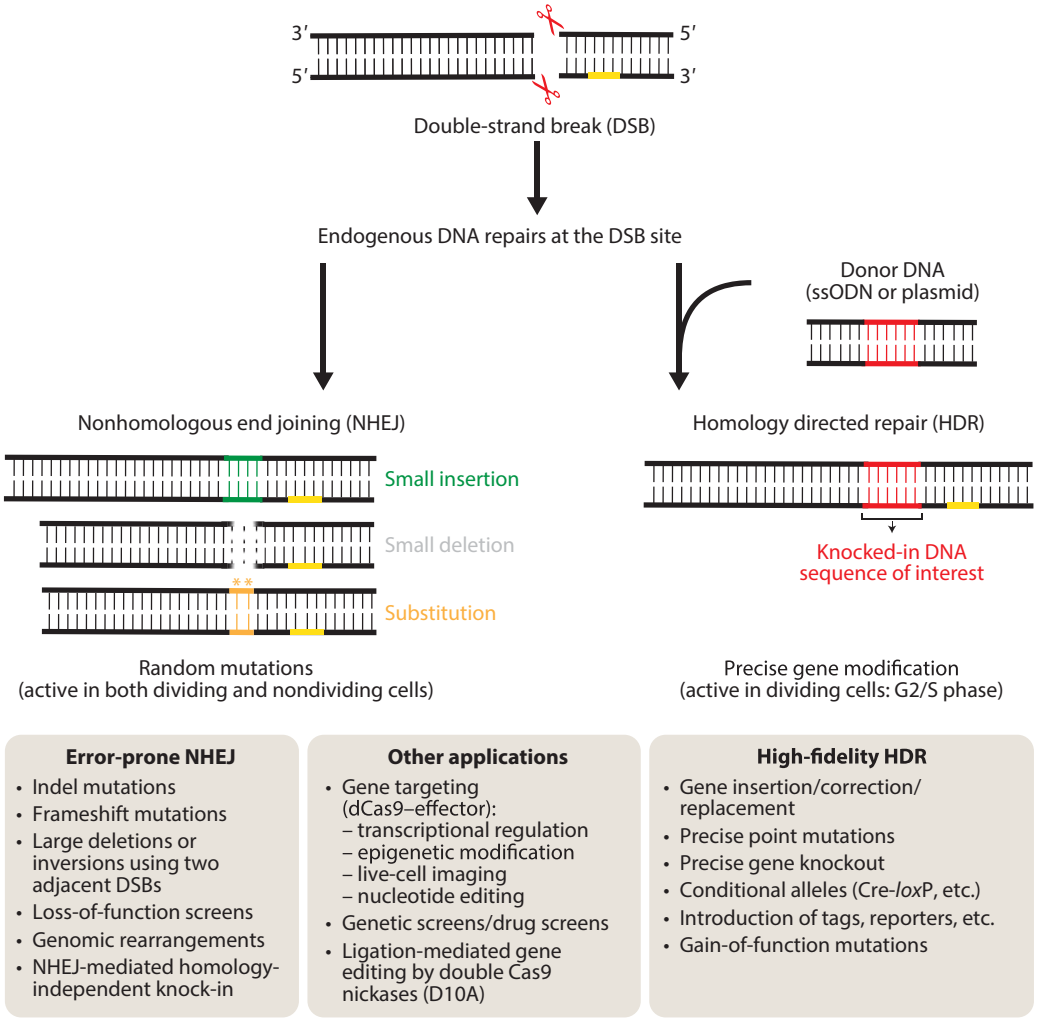


Figure 9. Use of Cas9 genome editing

Cas9 can induce site specific DSBs, but the mechanisms for repair remain largely the same. Double strand breaks can be repaired by mutagenic NHEJ repair that can result in loss of function and genomic rearrangements through illegitimate DNA repair. With HDR, one can get more precise alterations, which is ideal for introduction of genetic material such as markers. With dead Cas9, one can use it for specific gene targeting of protein effectors such as transcriptional regulators of gene expression. (Reproduced with permission from (Jiang and Doudna, 2017))

Germline modifications with CRISPR

Previously, the preferred method to generate mouse models was through template mediated homologous recombination (Capecchi, 1989). This was laborious with a slow throughput, making experiments quite costly. Recently, groups have shown that direct injection of mouse zygotes with CRISPR plasmids or Cas9 ribonucleoprotein (RNP) complexes caused inactivating mutations in target genes that were propagated through the progeny (Yang et al., 2013). In fact, groups have shown in mice and rats that multiple genes can be edited at once in ESCs (Li et al., 2013b).

The use of CRISPR has enabled efficient germline knock-in of specific mutations in the mouse. As these mutations are generated *in vivo* in the native locus, this could lead to a physiological model of gene alterations that was not feasible before. For example, groups were able to model single point mutants in Tet1 and Tet2 through HDR with a repair template in the mouse germline (Mou et al., 2015; Singh et al., 2015; Wang et al., 2013). Others were able to repair a point mutation in a mouse model of Duchene Muscular Dystrophy (DMD), demonstrating its potential in therapeutic gene editing (Long et al., 2014). CRISPR has also been used to tag genes with markers, which can be used to track proteins in live cells or used in downstream applications (Wang et al., 2013). Using HDR, one can insert a tag such as GFP or FLAG into the gene of interest (Lackner et al., 2015).

Somatic gene editing in cancer modeling

Though germline editing of cells (i.e. sperm or eggs) is useful for studying processes such as early development, there are ethical boundaries associated with germline modification. Given that most cell types do not propagate, genome editing could be used without the risk of germline transmission. One attractive target is the somatic stem cell, if modified, would be a self-renewing source of differentiated cell progeny with the same alteration. Currently, researchers have been able to edit *in vivo* a variety of cell types such as muscle, liver, lung, and many more (Song, 2017). For blood, *ex vivo* approaches

of modifying hematopoietic stem cells and transplanting them back into recipients have also been developed.

Using CRISPR-Cas to disrupt gene function in the mouse models of cancer takes advantage of NHEJ mediated repair and positive selection for mutated cells. For example, one group modeled the loss of p53 tumor suppressor in a transplantation model (Malina et al., 2013). Another group was able to model additional cooperating mutations with CRISPR-Cas9 editing and induction of mutant K-ras in lung adenocarcinoma (Sanchez-Rivera et al., 2014). Heckl and colleagues modeled several mutations found in myeloid leukemia through viral delivery of CRISPR-Cas9 (Heckl et al., 2014).

For modeling specific point mutations found in human cancer patient samples, CRISPR is invaluable in modeling these alterations. In the liver, point mutations in Beta-catenin were successfully modeled using a hydrodynamic delivery of Cas9 plasmid and repair template (Xue et al., 2014). In another case, oncogenic K-ras point mutations were successfully modeled in the lung by delivering adeno associated virus with Cas9 (Platt et al., 2014). Besides modeling oncogenic mutations, for many disease indications, CRISPR could provide a potential long-term treatment option by correcting mutations as well.

CRISPR-Cas9 variants

One of the early adaptations of CRISPR-Cas9 was conditional expression of CRISPR components using Cre recombinase for conditional, irreversible, expression of Cas9 in a transgenic mouse (Chu et al., 2016; Platt et al., 2014). For temporal flexibility, researchers have also created doxycycline inducible Cas9 in transgenic mouse models (Dow et al., 2015; Katigbak et al., 2018) or inducible sgRNA in lentiviral vectors (Aubrey et al., 2015), as constitutive expression of Cas9 may be detrimental to the organism and could increase the chance for off targets (Pattanayak et al., 2013). Acute editing would be beneficial for analyzing phenotypic output when chronic deletion is not desired. Combined with Cre-Lox recombination, CRISPR-Cas9 offers a chance for *in vivo,* temporal multistep

modeling of diseases.

Due to its large size which can impede efficient delivery, Cas9 has been split into multiple expression cassettes, which can be then reassembled in the cell through peptide, sgRNA recruitment, or rapamycin treatment (Truong et al., 2015; Wright et al., 2015; Zetsche et al., 2015b). Segmenting Cas9 has the added benefit of inducible expression and while each component can fit into AAV vectors which has one of the smallest carrying sizes (~4.5 kb). To bypass Cas9 delivery into cells altogether, several groups have made Cas9-expressing mice, which require only sgRNA expression (Chu et al., 2016; Dow et al., 2015; Katigbak et al., 2018; Platt et al., 2014). SgRNAs can be delivered, due to their small size, more efficiently.

Cas9 add-ons

As Cas9 target recognition and DNA cleavage are decoupled, it would be useful to attach additional proteins with specialized functions besides gene editing. This can be done by mutating the nuclease regions HNH and RuvC to create a catalytically non-functional or “dead” Cas9 (i.e. dCas9) that retains target recognition activity. Adding a fluorescent protein, such as GFP or mCherry, to Cas9 allowed targeting and imaging of specific regions of DNA such as the repetitive sequences of telomeres (Chen et al., 2013). Recently, researchers were able to modify sgRNA with MS2 binding sites to recruit fluorescent proteins, resulting in higher fluorescent signal and image resolution (Fu et al., 2016). Additionally, by adding a HALO tag to Cas9, various fluorescent ligands can be covalently linked for *in situ* labeling (Deng et al., 2015).

CRISPR-Cas9 can be repurposed as a DNA-binding protein to regulate gene expression. By adding a transcriptional activator, VP64, or repressor, KRAB, domain to dCas9, allows for tunable activation or repression of target genes (Gilbert et al., 2014; Gilbert et al., 2013; Konermann et al., 2015). The ability to turn on/off genes is useful in scenarios where genome editing is not desirable, such as expressing repressed genes.

Another area of intense interest is the role of epigenetic marks on DNA in transcriptional regulation. For example, adding the lysine-specific demethylase (LSD) to dCas9 allows the targeting of enhancer regions of pluripotency genes and repress their expression in response to histone H3 lysine 4 demethylation (Kearns et al., 2015). Recent research revealed that adding the catalytic core, acetyltransferase p300, resulted in acetylation of histone H3 lysine 27 and transcriptional activation (Hilton et al., 2015). Further, fusion of Tet1 demethylase or Dnmt3a methylase to dCas9 allows for activation or repression of target genes (Liu et al., 2016).

A recent appealing alternative to HDR and DSBs is the use of base editing with CRISPR-Cas9. For example, cytidine deaminases such as AID and APOBEC induce transitions by converting a cytidine to uridine (C -> T). A modified Cas9 capable of single stranded nicks (nickase), coupled with such cytidine deaminases, could be used as a programmable base editor. This can be used to model point mutations *in vivo* that would otherwise prove challenging with traditional nuclease-based methods (Komor et al., 2016). Meanwhile other base editors have been engineered to deaminate adenine to convert (A-T) to (G-C) base pairing (Gaudelli et al., 2017). Issues concerning specificity and activity have been recently addressed with improved base editors (Zafra et al., 2018).

Utility of using multiple sgRNAs at once

Single guide RNAs are sufficient for NHEJ mediated repair that creates small insertions and deletions. To model loss-of-function of coding genes, a single gRNA can be designed to target exons near the start site or critical protein domains (Shi et al., 2015). Due to indels and frame shift mutations, targeting exons often results in a non-productive protein product (Shi et al., 2015). For point mutations, a specific cut in the DNA by a single sgRNA is still sufficient for HDR with an appropriate repair template.

There are situations where the use of more than one sgRNA would be ideal. One such situation is when modeling two or more mutations as this requires simultaneous

expression of multiple sgRNAs as in the case of cancers with combinatorial genetic lesions (Heckl et al., 2014). Multiple sgRNAs would also be useful in dissecting the role of chromosomal topology in the function and pathogenicity associated with disruption of enhancer elements, CCCTC-binding factor (CTCF) sites, or topological associating domains (TADs) (Guo et al., 2015; Korkmaz et al., 2016; Lupianez et al., 2015). For some genes, such as non-coding RNAs, that are less sensitive to targeted disruption, assessment of loss-of-function requires whole gene deletion or poly A destabilization (Liu et al., 2017; Zhu et al., 2016). Finally, modeling of chromosomal rearrangements requires two DSBs to occur and with CRISPR, it is possible to induce intra or inter-chromosomal rearrangements by expressing tandem sgRNAs to fusion partner genes (Maddalo and Ventura, 2016).

Delivery of multiple sgRNAs into a single cell

Delivery of sgRNAs can be achieved through DNA, preassembled RNP complexes or viral transduction. However, each approach has varying levels of efficiency, dependent upon on the delivery vehicle and approach (Liang et al., 2015). For example, delivery with Cas9 mRNA results in quick and transient gene editing compared to plasmid-based delivery, as the need to transcribe the mRNA is circumvented. Even faster would be directly delivering functional Cas9 RNPs in the cell itself. However, these approaches require manual preparation of each complex for each unique target and delivery into the cell. In contrast, viral delivery is able to express multiple sgRNAs from the vector cassette, though need for provirus integration and transcription makes this the slowest process for assembling the Cas9 complex. In addition, co-delivery of single gRNAs in separate viral vectors unevenly distributes sgRNA among different cells, essentially diluting the pool making it less likely that each cells will contain each sgRNA.

CRISPR models of rearrangements

Several groups have been successfully modeled human cancers with *in vivo* mouse models using CRISPR (Heckl et al., 2014; Platt et al., 2014; Sanchez-Rivera et al., 2014). However, these studies focused on single locus mutations which could be modeled with a single sgRNA. Given the relative simplicity of the CRISPR approach, could CRISPR be used for more complex genetic alterations?

In vitro CRISPR modeling of rearrangements

While ZFNs and TALENs were used to model complex alterations including translocations (Brunet et al., 2009; Ghezraoui et al., 2014; Piganeau et al., 2013) and duplications (Lee et al., 2012), the use of gene editing to induce rearrangements remained limited to labs with the technical expertise. With the rise of CRISPR (Gaj et al., 2013), several groups were able to model deletion or inversion fusions in cancer cell lines (Blasco et al., 2014; Kannan and Ventura, 2015; Xiao et al., 2013). One group recently induced deletions, inversions, and duplications in mouse embryonic stem cells (ESCs), but not inter-chromosomal translocations (Kraft et al., 2015).

In vivo CRISPR modeling of rearrangements

The first *in vivo* evidence that CRISPR could model complex genetic alterations came from separate labs modeling the EML4-ALK inversion in lung adenocarcinoma (Blasco et al., 2014; Maddalo et al., 2014). Another study modeling interstitial deletion fusion of Brevican and Neurotrophic Receptor Tyrosine Kinase 1 (BCAN-NTRK1) in a mouse model of primary gliomas. They showed that *ex vivo* manipulation and delivery of adult neural stem cells or direct adenoviral transduction in the brain could recapitulate glioblastoma (Cook et al., 2017). Elegant work using an inducible transgenic Cas9 mouse was able to model R-Spondin rearrangements and demonstrate their role in colorectal cancer (Han et al., 2017). However, there has not been a *de novo* endogenous mouse model (autochthonous) for inter-chromosomal translocations involved in hematological

malignancies such as BCR-ABL1.

Why murine primary cells such as hematopoietic stem cells seem resistant to CRISPR mediated induction of certain rearrangements remains unclear. Experimental evidence suggests that primary murine leukocytes, such as non-activated murine T cells, are resistant to short interfering RNA (siRNA) mediated knockdown and viral delivery of CRISPR-Cas9 components (Seki and Rutz, 2018). Human cells appear to be more amenable to CRISPR induction/reversion of complex rearrangements based on several studies using xenograft mouse models (Breese et al., 2015; Choi and Meyerson, 2014; Lekomtsev et al., 2016; Reimer et al., 2017).

The need for models of Inter-chromosomal translocations

The only inter-chromosomal translocation linked to hematological malignancies and modeled with CRISPR *in vivo* is the MLL-ENL translocation, by modifying human CD34+ primary cells and transplanting back into immunodeficient mice (Reimer et al., 2017). Even so, this xenograft model does not recapitulate the endogenous fusion in the mouse and given the lack of an oncogenic phenotype, may not be the ideal experimental setup. Given the paucity of published data and need for better models, I aimed to model inter-chromosomal rearrangements in hematological malignancies in a murine setting. The lack of published tools motivated me to develop the toolbox needed to model chromosomal rearrangements in mice and would allow rapid screening of gene fusions found in human patients.

Limitations of CRISPR-Cas9

Although CRISPR-Cas9 technology has revolutionized cancer research by facilitating *in vitro* and *in vivo* models, its limitations should be considered during the design of CRISPR-based experiments. A crucial concern is the specificity of Cas9 and avoiding the possible introduction of undesired mutations.

Detecting/Reducing off-target effects

Off-target effects may be mitigated but not eliminated through multiple strategies. Using dual Cas9 'nickases' can reduce off-target effects as both single strand cuts must be close enough to generate a DSB (Ran et al., 2013a). Other techniques involve optimizing sgRNA selection/length to increase on-target activity (Dang et al., 2015; Doench et al., 2016) and de novo protein optimization of the Cas9 protein (eSpCas9) (Slaymaker et al., 2016). Ultimately, the only sure way to reduce off-targets in an experimental setting remains empirical validation after extensive screening to eliminate undesirable sgRNAs.

Identifying and mitigating potential off target DNA cuts is critical when this technology is used in a clinical setting (Tsai et al., 2015). The Joung lab has developed experimental methods to identify genome wide off-target effects in an unbiased manner. This group developed GUIDE-seq to identify the location of Cas9-mediated DSBs by inserting a DNA linker into cut sites that can be sequenced. They were able to identify highly promiscuous sgRNAs designed against the human Vascular endothelial growth factor A (VEGFA) locus, where off targets arose from mismatches near the distal part of the recognition sequence. This illustrates the need for better understanding of how CRISPR-Cas9 off targets occur and how best to mitigate them.

Improving Cas9 gene editing efficiency

The rules for optimizing efficiency of CRISPR-mediated gene editing remains largely uncharacterized. This is problematic as poor sgRNA design can or limit/prevent gene editing (Yuen et al., 2017). Several tools have identified sequence determinants of sgRNA efficiency though these models were on limited data sets (Moreno-Mateos et al., 2015; Xu et al., 2015). This is particularly significant when endeavoring to introduce specific mutations or repair mutated genes as the efficiency of HDR in most cells is low (often below 5%), making a more efficient process desirable (Liang et al., 2017). Some

progress has been made using temporally restricted Cas9 enzymes engineered to be expressed only during the S and G2 phases of the cell cycle, when HDR is a major repair pathway (Gutschner et al., 2016). Other strategies to improve HDR include blocking NHEJ DNA repair with small molecules, synchronizing the cell cycle to S and G2 phase (Lin et al., 2014; Maruyama et al., 2015), and localizing the donor template at the cleavage sites (Aird et al., 2017). However, these strategies have not been tested in primary tissues.

Biological challenges to using CRISPR to model rearrangements

The generation of chromosomal rearrangements using CRISPR-Cas9 poses unique challenges. Based on the published literature and mechanism of chromosomal rearrangements, deletion and inversion events are simpler to model. This is in part because the two breakpoints are within the same chromosome and therefore are more likely to interact. Consistent with this hypothesis, the shorter the distance between the two breakpoints, the more likely the rearrangement is to occur (Lieber et al., 2010). Strategies that enrich for modified cells use reporters to detect high nuclease activity (Ren et al., 2018). However, this approach is better suited for single site alterations instead of chromosomal rearrangements where nuclease activity does not correlate with translocation formation.

Chromosomal translocations are more challenging to model as the efficiency is lower than other types of rearrangements. In one study, the calculated inversion rate for EML4-ALK was a mere 4% in a cell line (Maddalo et al., 2014). Although precise rules have not emerged, it is likely that genome topology, micro-homology, and chromatin status play major roles in determining the probability that a translocation occurs in a given cell (Burman et al., 2015; Hogenbirk et al., 2016; Mathas et al., 2009). It has been suggested that delivering all the CRISPR components in single delivery vehicle may increase efficiency (Kannan and Ventura, 2015). However, the stochastic nature of the translocation process still results in unwanted indels, inversions or non-viable alterations

such as di/acentric chromosomes (Frock et al., 2015a; Frock et al., 2015b), which reduce overall efficiency. In fact, using two or more guides can induce off-target translocations and confound the true driver mutations (Perez et al., 2017).

Improving the efficiency of chromosomal rearrangements

Chromosomal rearrangements are rare and selected against, due to the detrimental effect of large genomic alteration on organisms. This is also due to the kinetics of fusion partner joining and DSB repair pathways account for this low efficiency (Lekomtsev et al., 2016), because it is far more likely to be resolved without a rearrangement (Maddalo and Ventura, 2016). The low efficiency of rearrangement generation may not be a major limitation if it provides a strong selective advantage and clonal outgrowth in humans over a long period, but this is not practical if the goal is to generate the desired rearrangement in a large fraction of cells in an experimental system. The inability to predictably direct DNA repair pathways to favor rearrangements is another roadblock to modeling rearrangements.

To address this, some groups incorporated selection markers which enrich for these rare genetic events (Spraggon et al., 2017; Vanoli and Jasin, 2017; Vanoli et al., 2017). However, these require the insertion of exogenous markers instead of exploiting the phenotypic output of certain gene fusions. For example, the well-established murine pro-B Ba/F3 line, has been used to screen oncogenic fusion kinases through withdrawal of the cytokine interleukin 3 (IL3). The Ba/F3 depends on IL3 cytokine for survival, but this can be replaced by kinase expression upon IL3 withdrawal. Only cells with successful expression of the fusion kinase will promote survival and IL3 independent growth. One downside is that this form of expression does not reflect the physiological rearrangement (Lu et al., 2017).

If selection methods are not possible or ideal, alternative approaches include single cell sorting and clonal analysis to identifying clones with rearrangements

(Lekomtsev et al., 2016). One promising technique uses digital droplet polymerase chain reaction (PCR) (BioRad) to screen large populations for edited cells (Findlay et al., 2016). Unfortunately, the presence of a marker or edited locus is only indicative of delivery of CRISPR components, while determining if a cell contains the desired rearrangement requires direct sequencing and detection of the breakpoint.

Functional significance of fusions modeled by CRISPR

Even if a gene fusion can be generated through CRISPR, it may be possible that some fusions are not the drivers of tumorigenesis (Reimer et al., 2017). Gene fusions may simply be passenger events or modulators of tumorigenesis. For example, one group elegantly modeled the *Pax3-Foxo1* reciprocal translocation found in human alveolar Rhabdomyosarcoma despite the opposite orientation of the fusion partners in mice. This was achieved through irreversible inversion of *Foxo1* followed by CRISPR meditated DSBs and translocation. Even with gene fusion expression and recapitulation of the chromosomal context, they were not able to see any oncogenic phenotype in the mice suggesting the need for secondary alterations (Lagutina et al., 2015).

It is also possible that some gene fusions caused by rearrangements contribute to a spectrum of chromosomal instability found in cancer. For example, in glioblastoma multiforme (GBM), EGFR rearrangements are often found in the context of EGFR gene amplification (Ozawa et al., 2010). In PDGFRα amplified gliomas, it was found that PDGFRα fusion is often present in samples. Given the vast number of identified fusions, developing a system for quickly identifying clinically promising candidates for further characterization is of utmost importance.

Goal of this thesis

The goal of this work is to develop approaches to model chromosomal rearrangements in the hematopoietic system using CRISPR-based genome editing. This was motivated, in part, by the current models of rearrangements that often present incomplete or conflicting observations, perhaps due to generation of gene fusion that does not recapitulate the endogenous genomic locus. Given that there are now thousands of uncharacterized gene fusions, the need for faster approaches for modeling is becoming urgent. To accomplish this, I developed a CRISPR-Cas9 γ-retroviral vector to facilitate delivery of CRISPR-Cas components. This work shows that CRISPR modeling can be used to recapitulate known gene fusions *in vitro*. In addition, I developed a simple cell-based platform with the murine pro-B Ba/F3 cell line that can be used to rapidly model and test fusions of relevance for hematologic malignancies. Although major technical challenges remain to be addressed, this work facilitates efforts to investigate the biology of gene fusions driving human cancers.

MATERIALS AND METHODS

Cloning of viral vectors

All viral vectors were maintained in Stbl3 or Stbl4 (Thermo Fisher Scientific) bacterial cell lines to minimize recombination. I used New England Biolabs (NEB) reagents for restriction digestion, Gibson Assembly, and Quick Ligase for cloning unless otherwise noted. For chemical transformation, I used standard heat shock and recovery with antibiotic selection. For bacterial electroporation, cells were electroporated by Gene Pulser (BioRad) according to manufacturer's protocol. Positive clones were screened by restriction digestion and sent for Sanger sequencing confirmation. Primers, sequences, and plasmids are available on request. Oligos were designed with Primer3 or Integrated DNA technologies (IDT). Sequences for single guide RNA (sgRNA) were determined with MIT CRISPR design site (Ran et al., 2013b).

Sequencing of cloning products

To confirm sequence identity of cloning products, I used primers upstream and downstream of target insert. Plasmids, PCR, or TOPO-TA cloning products were analyzed by Sanger DNA Sequencing with the MSKCC Integrated Genomic Operation Core Facility.

MSCV tandem sgRNA Cas9

Murine stem cell virus (MSCV) backbone was digested with XhoI and BamHI with intact long terminal repeat (LTR) promoter. The tandem hU6 Pol III promoters driving sgRNA expression and with a chicken beta hybrid promoter (CBh) driving expression of Cas9, internal ribosomal entry sequence (IRES) and enhanced green fluorescent protein (EGFP) was PCR cloned into the MSCV vector in the 5' to 3' orientation from the 5' LTR.

MSCV (cDNA overexpression constructs)

A generous gift of MSCV IRES EGFP with cDNA of human *NPM1-ALK* was provided by Dr. Cornelius Miething (Miething et al., 2003). This insert was driven by the viral LTR promoter and transcripts in the sense orientation.

pSUPER.Retro vector (Divergent/Convergent)

The pSUPER gammaretroviral MSCV vector backbone (Brummelkamp et al., 2002b) was used to add single or dual sgRNAs driven by hU6 or sU6 Pol III promoters by Gibson Assembly reaction that is divergent and upstream from the PGK promoter. Then Cas9 with a porcine teschovirus-1 2A self-cleaving peptide (P2A) EGFP or IRES EGFP DNA fragment was assembled by Gibson reaction (digestion method) downstream and in the sense orientation of the PGK promoter in the viral vector (pSRD). For the pSUPER.Retro vector with convergent cassettes (pSRC), sgRNAs to *Npm1-Alk*, placed downstream of Cas9 and in the convergent orientation. A similar strategy was used to generate a pSRC with sgRNA and mCherry only. Cells were cloned and transformed into Stbl4 via electroporation due to the difficult nature of the cloning.

pX viral vector cloning (Convergent)

PXZ201 was a generously gifted by the lab of Dr. Harvey Lodish with assistance from Dr. Hojun Li. Single guides were oligo cloned with Bbs1 into the convergent orientation in relation to the Cas9 gene but with different overhangs compared to PX330 and LentiCRISPR v2 plasmids. The dual cloning method described in by work in our lab and (Vidigal and Ventura, 2015) was adapted for this vector (Li et al., 2016a). A similar strategy was used to generate pX vectors with sgRNA and the mCherry marker. Plasmid DNA was extracted in Stbl3 transformed cells.

Viral Transduction and concentration

Supernatant Based Transduction

For non-concentrated virus, prior to transfection, packaging cells were changed to media of transduced cells. Platinum-Eco packaging cells (Cell BioLabs) were transfected with 2ug per well of a 6-well plate of plasmid DNA prepared by QIAGEN EndoFree Plasmid Maxiprep kits. Virus supernatant was collected 48 and/or 72 hours post transfection without prior media change and filtered through 0.45 μM filter (EMD Millipore). Virus

supernatant was applied to cells then subject to 1-2 rounds of spinoculation was done (689*g for 90 minutes at 25°C) with 4-8 mg/mL polybrene (Sigma-Aldrich). For hematopoietic cells, additional rounds of transduction with viral supernatant were performed as necessary.

For Retro-X concentration of viral supernatant

For our two-plasmid virus generation, GP2-293 (Clontech) cells were seeded on 4x15 cm tissue culture plates. Prior to transfection, packaging cells were changed to media of transduced cells. Packaging cells were transfected with MD2 (VSVG) and gammaretroviral vector with JetPrime transfection reagent (Polyplus+). Viral supernatant was then collected 48 and/or 72 hours post transfection without prior media change and filtered through 0.45 μM filter. Then supernatant was concentrated according to manufacturer's protocol of Retro-X solution (Clontech). The viral prep was then tested on NIH-3T3 by cell sorting. Concentrated virus was aliquoted and stored in phosphate buffered saline (PBS) at -80°C.

Mammalian cell culture

NIH-3T3 were grown in Dulbecco's Modified Eagle's Medium (DMEM) with 10% Fetal Bovine Serum (FBS) (Sigma), 10 mM L-glutamine, 10 nM Penicillin-Streptomycin (PS). GP2-293 cells were maintained under non-confluent densities and low passage number in DMEM media supplemented as mentioned above. Ba/F3 cells were maintained in B cell medium (BCM) (45% DMEM, 45% Roswell Memorial Institute Medium (RPMI), with 10% FBS (Hyclone), 20nM L-glutamine, 10nM P-S, and 50uM beta-mercaptoethanol (βMe) and split regularly to avoid the emergence of IL3 independent clones. WEHI cells were maintained in RPMI media supplemented with 10% FBS, 10mM L-glutamine, and 10 nM PS. WEHI supernatant containing secreted interleukin 3 (IL3) cytokine was collected and filtered through 0.45 μM filter and stored at -20°C. Viral packaging cells were maintained in DMEM media until virus production.

IL3 cytokine withdrawal

Ba/F3 cells were maintained in BCM supplemented with 5% WEHI filtered supernatant containing IL3 cytokine. To remove IL3 from media, cells were spun and washed three times with PBS at 0.2 Relative Centrifugal Force (RCF) for 5 minutes each. Ba/F3 cells were then resuspended in BCM without WEHI supplement. Cell number and viability were monitored by the Countess II automated cell counter (Thermo Scientific).

Delivery of CRISPR-Cas9 components and assessing CRISPR gene editing

Amaxa Nucleofection Protocol (Lonza)

For Ba/F3, I used Buffer SG with the suggested manufacturer's protocol in a 24 well format. Cells were allowed to recover for 24 hours post nucleofection. Nucleofection efficiency was tracked by cell sorting and Countess II Trypan Blue staining.

Polymerase chain reaction genotyping

To detect the *Npm1-Alk* translocation at the genomic level, I performed Phusion (Thermo Scientific) polymerase chain reaction (PCR) (Maddalo et al., 2014). PCR was TOPO (Invitrogen) cloned and sequenced by Sanger sequencing.

Assessing on target efficiency

Upon PCR amplification of the target locus, PCR band was either gel or PCR purified. The DNA was either directly sequenced or cloned into TOPO vector. Chromatograms of sequencing data were analyzed by Seqman Pro (DNASTAR).

Reverse transcription of complementary DNA

One microgram of total RNA which was quantified by Nanodrop (Thermo Scientific) was treated with DNase I (Thermo Scientific) according to manufacturer's protocol. Then the reaction was directly used in Superscript III reverse transcriptase (RT) kit (Thermo Scientific) using oligo DT primers using the standard protocol. For No RT controls, only buffer components were added. The reaction was then diluted with nuclease free double distilled water (ddH20) and stored at - 20 °C. I used primers to detect the breakpoint and

full length *Npm1-Alk* cDNA using Phusion polymerase (Maddalo et al., 2014).

Protein extraction and Western Blot analysis

Cells were harvested by spin down and washed 3 times with PBS. Pellets were lysed with Radioimmunoprecipitation assay (RIPA) lysis buffer supplemented with protease inhibitor (Roche) cocktail and PhosSTOP tablets (Roche). Protein was quantified using the Bradford Assay (Biorad) using a Bovine Serum Albumin (BSA) standard and read on a 96 well plate using 595 nM absorbance. 4X Lamenilli buffer with βMe were added to samples and heat denatured followed by loading onto 4-12% Bis-Tris gels (Invitrogen). Samples were transferred overnight (4°C) at 15 volts and then blocked with either 5% TBS-T with non-fat milk or BSA. Antibodies were incubated according to manufacturer's protocols. Secondary Horseradish peroxidase (HRP) antibodies were used to detect primary antibodies and ECL prime (General Electric - GE) was used for imaging on photographic films (GE).

Antibodies used

I used manufacturer's suggested protocol for Western Blot. (AbCam) Anti-ALK Cell Signaling Technologies (CST) Anti–human ALK, (Sigma) Anti-FLAG, Anti-β-actin

Flow cytometry analysis and fluorescence assisted cell sorting (FACS)

Cells were spun down and resuspended in FACS buffer (supplemented with FBS and EDTA). I used Fluorescein isothiocyanate (FITC) and Texas Red channels for analysis. Flow cytometry sorting of GFP+ cells was performed using FACSAria (BD Biosciences) flow cytometer using a DAPI stain. FACS data was analyzed using the FlowJo (v9) software.

RNA extraction and single guide RNA Northern Blot

Total RNA was extracted using Trizol (Life Technologies) according to manufacturer's protocol and kept in diethylpyrocarbonate (DEPC) treated double distilled water. Ten µg of total RNA was mixed with 2X RNA loading dye (Ambion). Samples were

loaded onto a 15% Urea polyacrylamide gel and separated at 20mA by gel electrophoresis. Gels were run until the lower marker (~10 bp) runs toward the end of the gel, to avoid missing premature termination transcription products. Gel was stained with Ethidium bromide and imaged to check loading consistency via ribosomal RNA signal. Gel was transferred to an Amersham Hybond membrane (GE) using a Bio-Rad transfer cassette. After transfer, membranes dried and are then cross-linked with 1100 volts for two 1-minute cycles. Membranes could also be stored at 4°C until ready for subsequent analysis. Then the membrane was pre-hybridized with Clontech hybridization buffer at 37°C shaking for one hour. DNA probes were designed by IDT as the reverse complement of the N20 specific sequence of each sgRNA. To end label DNA probes, I used gamma ATP and T4 DNA PNK (NEB) and labeled for 30 minutes at 37°C. Probes were cleaned with T25 spin columns (GE). Probe was incubated with the membrane in a rotating hybridization oven ON at 37°C. Membranes are washed first with 2X SSC (Saline-sodium citrate) buffer and then with 0.2X SSC buffer. Membranes were exposed with film from 1-72 hours (GE) in a cassette and kept at 4°C. To reprobe a membrane, it was stripped with boiling 0.1% Sodium Dodecyl Sulfate (SDS) and washed with 0.2X SSC until residual signal is removed. The membrane gets reblocked with prehybridization buffer and reprobed ON at 37°C. For RNA loading control, I used the murine small nuclear U6 (snU6) probe (Murchison et al., 2005).

In vitro cell line characterization

Cell Titer Glo Assay (Promega)

Cells were added to each well of a 96 well plate with flat bottom white colored plates at a density of 1x10^4/mL or 10,000 cells per well. For samples undergoing IL3 withdrawal, cells were maintained in non-IL3 media prior to adding reagent to the well. Crizotinib were prepared at 5X final concentration and then added to cells in the well at the appropriate concentration. Samples were read 3 days with Cell Titer Glo Assay

(Promega) and read on a Promega Glomax 96 well plate machine after drug treatment per manufacturer's instructions. For all assays, Crizotinib (CST) was prepared as a 10 mM stock solution in DMSO.

Mouse Husbandry/Experiments

Animal studies and procedures were reviewed by MSKCC Institutional Animal Care and Use Committee under protocol number 14-08-009. Athymic nude mice were ordered from Jackson Laboratories (JAX) in the (nu/nu) background. BALB/c mice for allograft studies were ordered from JAX. All mice were maintained in a Research Animal Resource Center facility under veterinary supervision.

Allograft transplantation of Ba/F3 cells via flank injection in nude mice

Protocol was partially adapted from a published report (Cuesta-Dominguez et al., 2012). One million live Ba/F3 cells were counted as determined by Countess II using Trypan Blue staining and resuspended into 1X Hank's Balanced Salt Solution (HBSS). Cells were put in a 1:1 mix with Matrigel (Corning) and kept on ice until ready for injection. Aythmic mice (JAX) were then injected on the on the right flank and monitored for tumor growth and overall survival. Tumors were measured with a digital caliper (Fowler) and volume calculated based on $V = (W(2) \times L)/2$. Mice were sacrificed once tumors were above 600 mm^3. After mice were sacrificed, tumors were extracted and saved for genomic DNA, RNA, and protein extraction and downstream analysis.

Allograft transplantation of Ba/F3 cells via tail vein injection in BALB/c mice

Allograft protocol was partially adapted from a previous report (Ma et al., 2014a). Six to eight-week-old BALB/c male mice were sub-lethally irradiated with 4.5 Grays (Gy) of radiation from a Gammacell 40 irradiator with a Cesium source and allowed to recover for 24 hrs. Then one million live Ba/F3 cells, as measured by Trypan Blue and Countess II quantification, were resuspended in PBS and injected in the tail vein of mice with a 28-gauge insulin syringe (Corning). Mice were then tracked for overall survival and were

sacrificed and processed when determined by a veterinarian to be morbid.

Statistical Analysis

For Ba/F3 cells treated with Crizotinib in Cell Glo Titer assays, unpaired T-test was calculated on GraphPad Prism 7 compared to parental Ba/F3 cell line. IC50 was calculated on GraphPad Prism 7 using a three parameter Inhibitor/Response non-linear regression analysis. For calculating overall survival, the Mantel Cox log rank test was used to compared between wild type mice and those injected with Ba/F3 mNPM1-ALK.

Tissue collection

Primary mouse embryonic fibroblasts (MEFs)

Head, arms, mid torso (with fetal liver and heart removed), legs, and tails were dissected from E14.5 post inception. embryos from timed mated females. The tissues were mechanically dissociated into smaller pieces and then incubated in Trypsin overnight with shaking at 4°C. Then suspension was placed in DMEM media supplemented as mentioned above. Cells were passaged at a low number (< 2) before being stored in DMEM media with 10% DMSO for long term storage.

Complete necroscopy and blood counts

Mice were dissected, and organs were viewed by trained pathologists for signs of overt disease. Blood was collected from mice before sacrifice and subjected to automatic and manual complete Blood quantification. The sample collection was performed by the Lab of Comparative Pathology Core.

Histology and Cytogenetics

Immunohistochemistry

Immunohistochemical staining was performed on 5 μM sections of formalin-fixed/paraffin embedded tissues with an automated staining processor (Discovery XT, Ventana Medical Systems) on positively charged glass slides. Staining was done by the Laboratory of Comparative Pathology (LCP) core facility (Maddalo et al., 2014). Antibodies

used for staining: GFP, anti ALK (AbCam), anti AKT, anti STAT3

Hematoxylin and Eosin Staining (H&E)

Samples were prepared as described previously and done by the LCP core (Maddalo et al., 2014). Images from each tumor were acquired with a Zeiss AxioCam Color camera at 10 - 40X magnification.

Fluorescence *in situ* hybridization (FISH) to detect translocations

FISH protocol was performed by the Molecular Cytogenetic Core Facility. Samples were grown and maintained to ensure robust cell division. Cell lines were harvested according to standard procedures and FISH analysis performed on the fixed cells using an in-house 2-color probe mix designed to detect *Npm1-Alk* fusion. The probe mix consisted of mouse BAC clones flanking the breakpoint on *Npm1* (11qA4) (RP23-96P14, RP23-413O19; labeled with Red dUTP) and *Alk* (17qE1.3) (RP23-397M18, RP23-12H17, labeled with Green dUTP). Probe labeling, hybridization, post-hybridization washing, and fluorescence detection were performed according to standard procedures. Slides were scanned using a Zeiss Axioplan 2i epifluorescence microscope equipped with CoolCube 1 CCD camera controlled by Isis 5.5.10 imaging software (MetaSystems Group Inc, Waltham, MA). The entire hybridized area was scanned under 63x objective and representative cells imaged. A minimum of 100 interphase nuclei and 25 metaphases were analyzed to assess the *Npm1-Alk* fusion status. A normal diploid cell is expected to show two Red (*Npm1*) and two Green (*Alk*) signals. A cell positive for *Npm1-Alk* fusion is expected to show at least one Fusion/Yellow signal (overlapping of *Alk* and *Npm1*), along with Red (normal *Npm1*) and Green (normal *Alk*) signal(s). Tissue FISH was performed by the Molecular Cytogenetic Core Facility. Samples (Splenic sections) were prepared on positively charged slides of 5 µM sections of formalin-fixed/paraffin embedded (FFPE) tissues. FISH analysis performed on the FFPE using an in-house 2-color probe mix designed to detect *Npm1-Alk* fusion as described above.

CHAPTER 1: Generation of Npm1-Alk fusion by CRISPR

Nucleophosmin 1

Nucleophosmin1 (NPM1) is a highly expressed protein that is typically located in the nucleolus, but it can also shuttle between the cytoplasm. NPM1 is involved in processes such as ribosome biogenesis, centrosome duplication, and protein chaperoning (Grisendi et al., 2005). Mutations within NPM1 are associated with leukemia and (Grisendi et al., 2006; La Starza et al., 2010) was identified as a haploinsufficient tumor suppressor (Sportoletti et al., 2008) that acts by interacting with ARF tumor suppressor and regulating p53 stability (Colombo et al., 2002; Itahana et al., 2003; Korgaonkar et al., 2005).

Anaplastic lymphoma kinase

Anaplastic lymphoma kinase (ALK) is a receptor tyrosine kinase that is expressed and functions predominantly in the brain during development, with knockout mice exhibiting mild behavioral defects (Bilsland et al., 2008; Iwahara et al., 1997). However, ALK is often found to be the fusion partner in a variety of cancers. For example, in lung adenocarcinoma the kinase portion of the ALK gene is fused with the echinoderm microtubule associated protein like 4 (EML4), forming the EML4-ALK fusion. Due to the loss of ALK regulatory regions, ALK ectopic expression occurs in the lung where it drives tumorigenesis (Soda et al., 2007; Soda et al., 2008). This mechanism involves dimerization followed by constitutive phosphorylation and activation which is similar for the other ALK fusion proteins (Hallberg and Palmer, 2013, 2016).

Anaplastic large cell lymphoma

Anaplastic large cell lymphoma (ALCL) is a rare form (1%) of non-Hodgkin lymphoma (NHL) that can be found on the skin or lymph nodes. The disease usually strikes younger patients as aggressive lymphomas and include ALK+ and ALK- groups. NPM1-ALK is a fusion protein found in a majority (> 70%) of patients diagnosed with ALK+ ALCL. Morphological features of ALK+ tumors include hallmark cells with enlarged nucleus and express CD30 and ALK (Drexler et al., 2000; Hapgood and Savage, 2015; Inghirami et al., 1994; Le Beau et al., 1989). The cell of origin is not currently well characterized, because they express either B, T, or null cell markers, though data suggests a majority are T cell lymphomas (Malcolm et al., 2016). It has been suggested that abnormal expression pattern of T cell markers in T cell lymphomas may be due to NPM1-ALK fusion itself (Ambrogio et al., 2009). Mutational profiling of ALK+ ALCL shows that few possess p53 mutations, suggesting that NPM1-ALK is the oncogenic driver (Rassidakis et al., 2005) and ALK would be an attractive therapeutic target.

Crizotinib

Current treatments for ALK+ ALCL involve using ALK inhibitors such as Crizotinib (Gambacorti-Passerini et al., 2011; Sahu et al., 2013). Crizotinib is a c-Met/ALK inhibitor that was first approved for the treatment of ALK+ non-small cell lung cancer (NSCLC). However, it has also been shown to be effective against ROS1 kinase fusions found in NSCLC (Davies and Doebele, 2013). Crizotinib prevents the activation of the ALK kinase by binding to the adenosine triphosphate (ATP) binding pocket. Common resistance mechanisms involve gatekeeper mutations that block Crizotinib binding at the ATP pocket via steric hinderance (Crescenzo and Inghirami, 2015).

NPM1-ALK translocation

NPM1-ALK is a reciprocal translocation that occurs between chromosomes 2 (ALK) and 5 (NPM1) in humans (Morris et al., 1994) while in mice, the syntenic regions are in chromosomes 11 (*Npm1*) and 17 (*Alk*) (Mathew et al., 1997). This fusion results in the joining of the dimerization domain of NPM1 and the kinase domain of ALK. As the NPM1 promoter now drives the ALK gene, this results in constitutive activation of the kinase and drives proliferation and anti-apoptosis (Kutok and Aster, 2002; Pearson et al., 2012) (**Figure 10**).

Previous models of NPM1-ALK fusion do not recapitulate disease

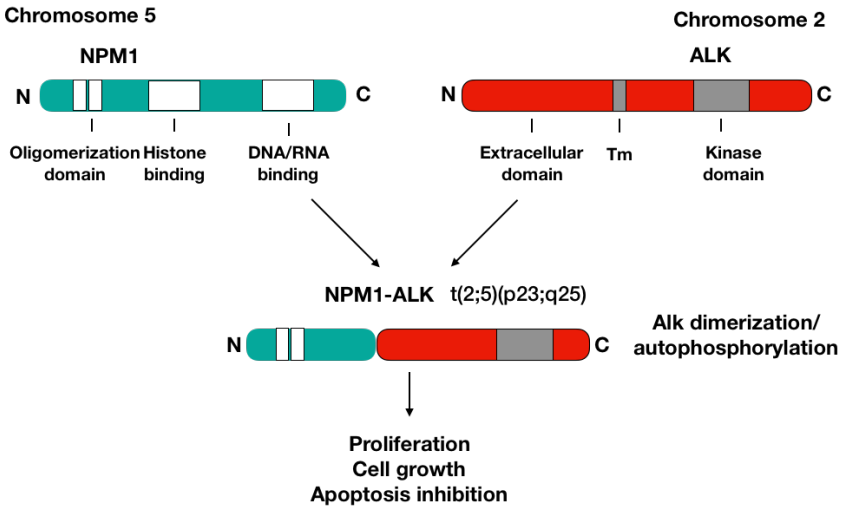
Groups previously modeled NPM1-ALK in mice to understand how it drives tumorigenesis. One study used retroviral vectors expressing cDNA of human NPM1-ALK to transduce bone marrow and transplant back into irradiated mice. This resulted in B cell lymphomas; something not seen in patients (Kuefer et al., 1997). Later work using modified fetal liver HSCs with retroviral expression of human NPM1-ALK cDNA resulted in myeloid or B cell neoplasia, but not the T cell lymphomas seen in patients (Miething et al., 2003). A mouse model combining IL9 transgenic mouse with NPM1-ALK cDNA overexpression also failed to recapitulate the disease (Lange et al., 2003).

Other groups worked to more accurately model using transgenic mouse models. One group generated a T cell specific expression of NPM1-ALK transgenic mouse, resulting in a mixed population of B and immature non-recombined T cell lymphomas, in contrast with patients that have T cell lymphomas with TCR recombination (Chiarle et al., 2003). A transgenic mouse model using the *Vav* hematopoietic promoter to express NPM1-ALK in hematopoietic stem cells resulted in lymphomas with B cell characteristics (Turner et al., 2003) while using T cell specific promoters resulted in lymphoblastic lymphomas (Jager et al., 2005). Though informative, these results do not faithfully recapitulate the human disease which was already noted (Turner and Alexander, 2005).

The need for faithful recapitulation of the endogenous gene fusion was not addressed until the advent of gene editing technology that could model those complex rearrangements.

Expression of Npm1-Alk sgRNAs and Cas9 can induce Npm1-Alk translocation

Previous work shows that tandem expression of Cas9 and sgRNAs could model the *Eml4-Alk* inversion event in mice. In that same paper, the *Npm1-Alk* fusion could be modeled as well, indicating CRISPR could be used to model inter-chromosomal translocation (Maddalo et al., 2014). Using the MIT CRISPR site (Hsu et al., 2013), several pairs of sgRNAs were initially designed against the intronic regions of murine Nucleophosmin (*Npm1*) and Anaplastic Lymphoma Kinase (*Alk*) and cloned into the pX330 vector to recapitulate a known tyrosine kinase fusion found in anaplastic large cell lymphoma (Morris et al., 1994), by inducing the *Npm1-Alk* inter-chromosomal rearrangement in murine cells. PX330 plasmids expressed a FLAG tagged Cas9 with a 3X nuclear localization sequence (3X NLS) along with a chicken beta hybrid promoter (CBh) (Cong et al., 2013). PX330 plasmids expressing these Npm1-Alk guides can induce *Npm1-Alk* translocations in transfected NIH-3T3 as it was detected at level of genomic DNA (gDNA) and cDNA (Maddalo et al., 2014) (**Figure 11**).



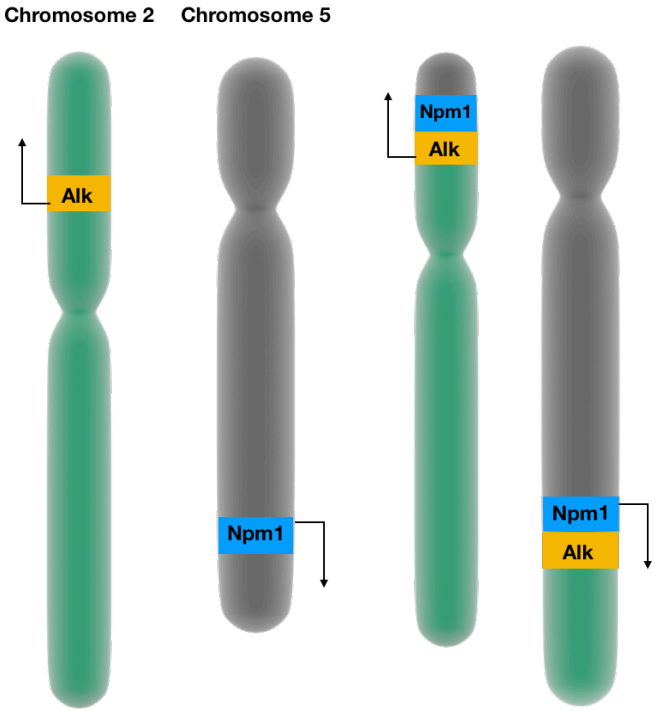


Figure 10. Schematic of the NPM1-ALK genomic translocation

(Left) Schematic of NPM1 and ALK human genes located on chromosomes 5 and 2 respectively. The 5' partner NPM1 contains the oligomerization, histone binding, and nucleic acid binding domains. The 3' fusion partner ALK contains the extracellular domain which regulates activation followed by the transmembrane (Tm) and kinase domain. Upon t(2;5)(p23;q25) reciprocal translocation where the NPM1 promoter drives expression of the fusion product that contains the oligomerization domain and the tyrosine kinase domain of ALK. This allows for constitutive dimerization and activation of the tyrosine kinase and signaling. (Right) Shown here is the human chromosomes 2 and 5 containing the Npm1 and Alk gene respectively. Upon reciprocal translocation, the NPM1-ALK and ALK-NPM1 gene fusions are created. Note that the transcriptional orientation of the fusion is preserved.
(Source: sommersault1824.com)

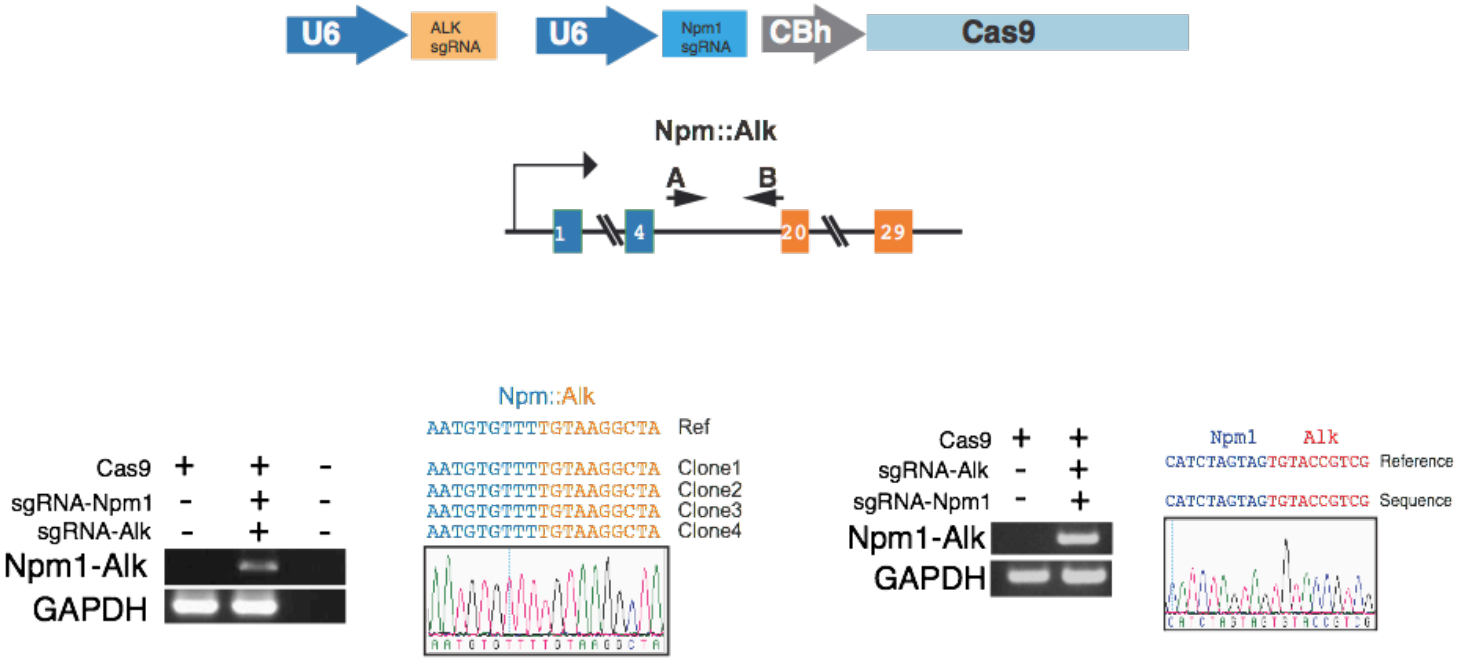


Figure 11. Tandem expression of Npm1-Alk sgRNAs and Cas9

(Top) Here the human U6 Pol III promoter drives sgRNA expression targeting Npm1 and Alk while chicken beta hybrid Pol II promoter drives Cas9 in pX330. (Top and Left) Using primer pairs that span the Npm1-Alk genomic fusion breakpoint, only co-transfection of Npm1 and Alk sgRNA plasmids results in induction of Npm1-Alk genomic fusion that was confirmed by Sanger sequencing. (Right) Npm1-Alk fusion transcript was confirmed by PCR of cDNA and was confirmed by Sanger sequencing. (Reproduced with permission from (Maddalo et al., 2014))

Modeling chromosomal rearrangements with CRISPR-Cas9

Npm1-Alk induction in Ba/F3 cells

Using our Ba/F3 cell line, a versatile cell that depends on the cytokine interleukin 3 (IL3) for survival, I tried to take advantage of this property to enrich for *Npm1-Alk* expressing cells. This cell line has been widely used to screen for kinase activity of gene fusions by withdrawing IL3 (Warmuth et al., 2007). Ba/F3 cells will die upon IL3 withdrawal, but if another kinase can provide pro-survival signaling, the cell can survive without IL3. As oncogenic kinase fusions provide strong pro-survival and anti-apoptotic signaling, this is an ideal system to screen for novel gene fusions for kinase activity. To increase my chance of generating *Npm1-Alk* in Ba/F3, I used the Amaxa nucleofector to deliver *Npm1-Alk* with Cas9 plasmid DNA. A longer recovery period was required before selection due to increased cell death associated with plasmid delivery but IL3 independent clones were visible within 2-3 weeks.

Detection of *Npm1-Alk* rearrangement

As shown by PCR, I observed that both Ba/F3 cell IL3 independent outgrowth contained the *Npm1-Alk* translocation at the genomic level. However, the PCR amplicon size of the nucleofected Ba/F3 (n-N-A) does not match the predicted *in silico* size and shows more variability in the edited locus with a range of gaps near the breakpoint. Considering that our guides target the intron, there is no selective pressure to maintain the sequence, because the cDNA transcript will splice out this region. I then confirmed the predicted fusion transcript breakpoint of *Npm1-Alk* in both cell lines by PCR. I also detected the whole length cDNA that matched the predicted sequence, which shows that CRISPR editing did not alter the exons **(Figure 12)**.

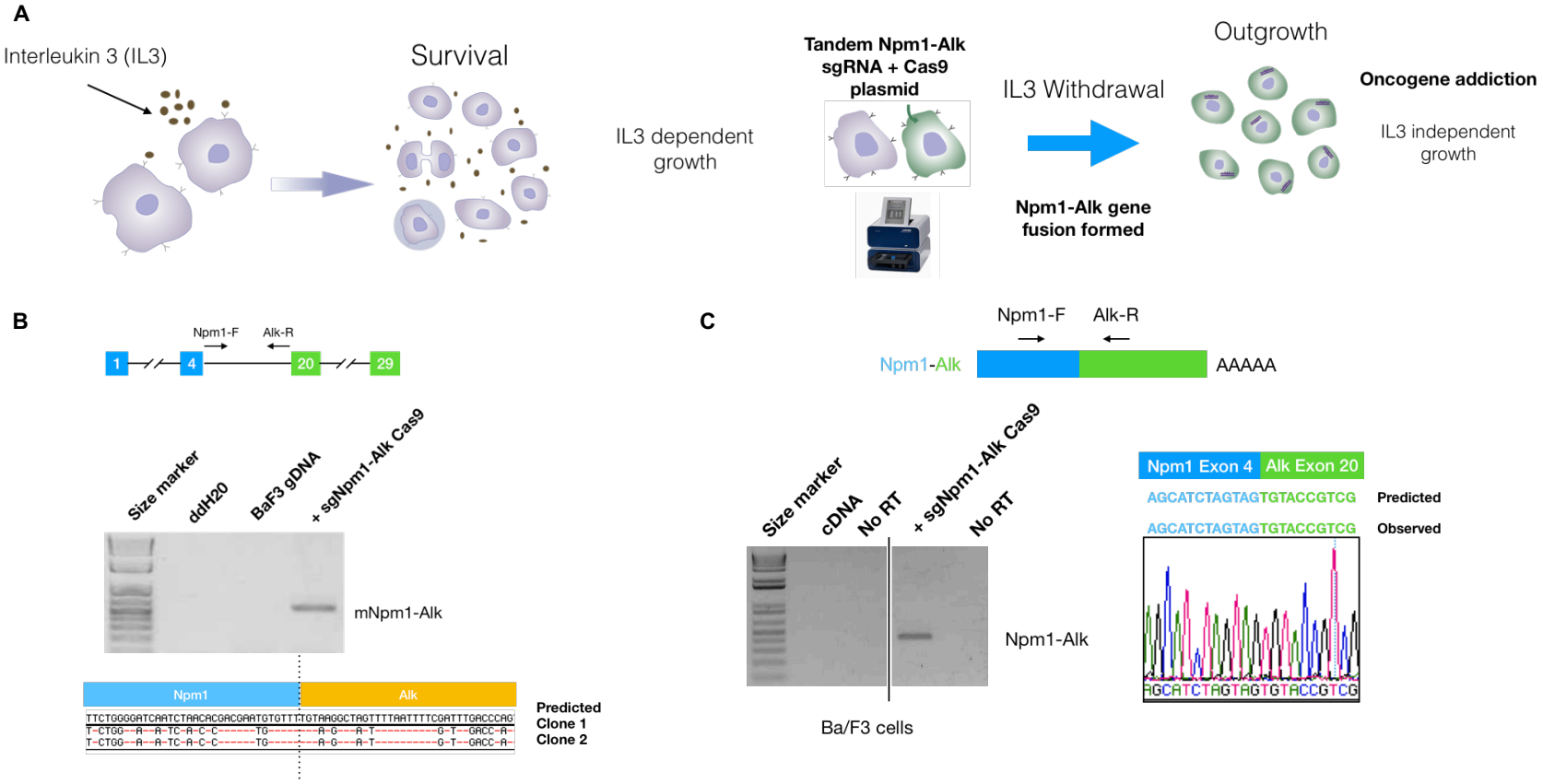


Figure 12. Delivery of Npm1-Alk sgRNAs and Cas9 into Ba/F3 cells

A) Shown here is the murine pro-B hematopoietic cell line Ba/F3 which depends on interleukin 3 for continued survival and proliferation. To select for cells with the Npm1-Alk rearrangement, I removed IL3 so only cells with Npm1-Alk gene fusion and become IL3 independent and oncogene addicted for growth. B) In IL3 independent Ba/F3 population, I could detect the Npm1-Alk fusion partners in genomic DNA with PCR though there are large gaps near the breakpoint. C) With PCR primers specific for the fusion transcript breakpoint, I could detect the predicted cDNA transcript which is not affected by the gaps in the intronic region of the gDNA.

Cytogenetic characterization of Npm1-Alk Ba/F3 cell lines

I observed the cytogenetic changes in our *Npm1-Alk* Ba/F3 cell line (N-A) to confirm CRISPR mediated inter-chromosomal translocation. I performed Fluorescence in situ hybridization (FISH) to detect murine *Npm1* and *Alk* genes. A two-probe break apart FISH was used to detect the *Npm1-Alk* translocation (Mathew et al., 1997) and karyotyping was performed in parallel to check the ploidy of our cell line. Unfortunately, the Ba/F3 parental line used for modeling our rearrangement is tetraploid, possessing 4 copies of *Alk* and *Npm1*. In addition, two of the chromosomes containing *Alk* fused as a Robertsonian translocation (**Figure 13**). It should be noted that gross chromosomal changes are known to happen in cell lines grown in culture (Mamaeva, 1998). For the N-A line, I could detect the *Npm1-Alk* and *Alk-Npm1* reciprocal translocation partner (**Figure 14**). However, the Alk fusion partner is the Robertsonian translocation with two *Alk* signals.

Though it would be ideal for the Ba/F3 line to be diploid, I conducted a karyotype and FISH analysis of individual cells to determine if there was any gross chromosome difference and heterogeneity between the parental and N-A lines. Based on the analysis of 20-25 cells per line, there is no significant change in karyotype between our *Npm1-Alk* and parental line **(APPENDIX 1.1: Karyotypes of Ba/F3 Cells)**. For our n-*Npm1-Alk* line, nearly every cell analyzed had the same cytogenetic alteration suggesting clonal outgrowth **(APPENDIX 1: FISH quantification of Npm1-Alk Ba/F3 line)**. This data supports that the fusion is selected for, rather than a non-specific effect, accounting for the phenotypic effect observed in our assays.

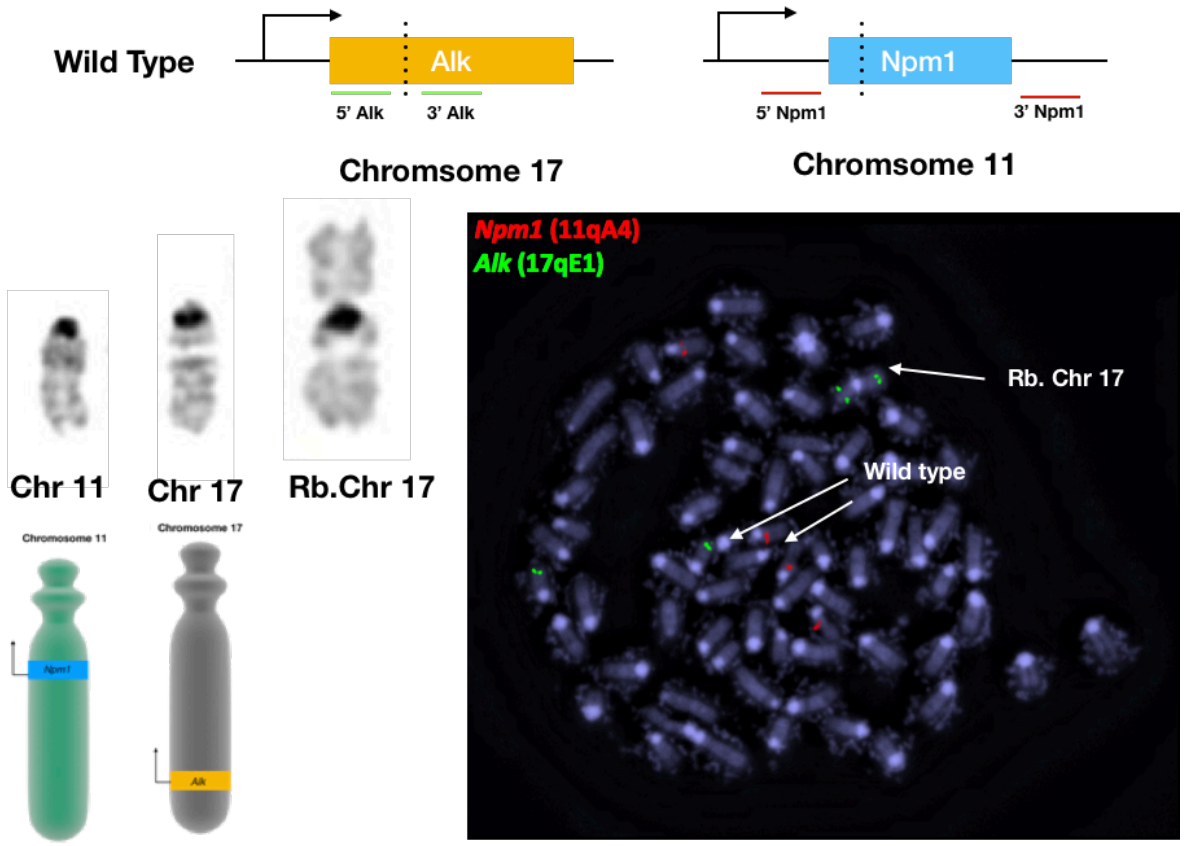


Figure 13. Cytogenetic analysis of the Ba/F3 parental line

(Top) Shown here is the FISH strategy for detecting Alk and Npm1 in Ba/F3. Green fluorescent probes were designed flanking the Alk target site (dotted line). Red fluorescent probes were designed flanking the Npm1 gene and the target site. In the cells, you can see Alk and Npm1 signals on chromosome 17 and 11 respectively, with FISH showing that our parental line is mostly tetraploid with four copies of Npm1 (red) and Alk (green) (White arrows). Of note, two chromosome 17 have formed a Robertsonian translocation which has two Alk signals. (Left) Here the karyotype of chromosome 11 and 17 along with the Robertsonian translocation 17 in Ba/F3 is shown and a schematic showing the location of the Npm1 and Alk gene.

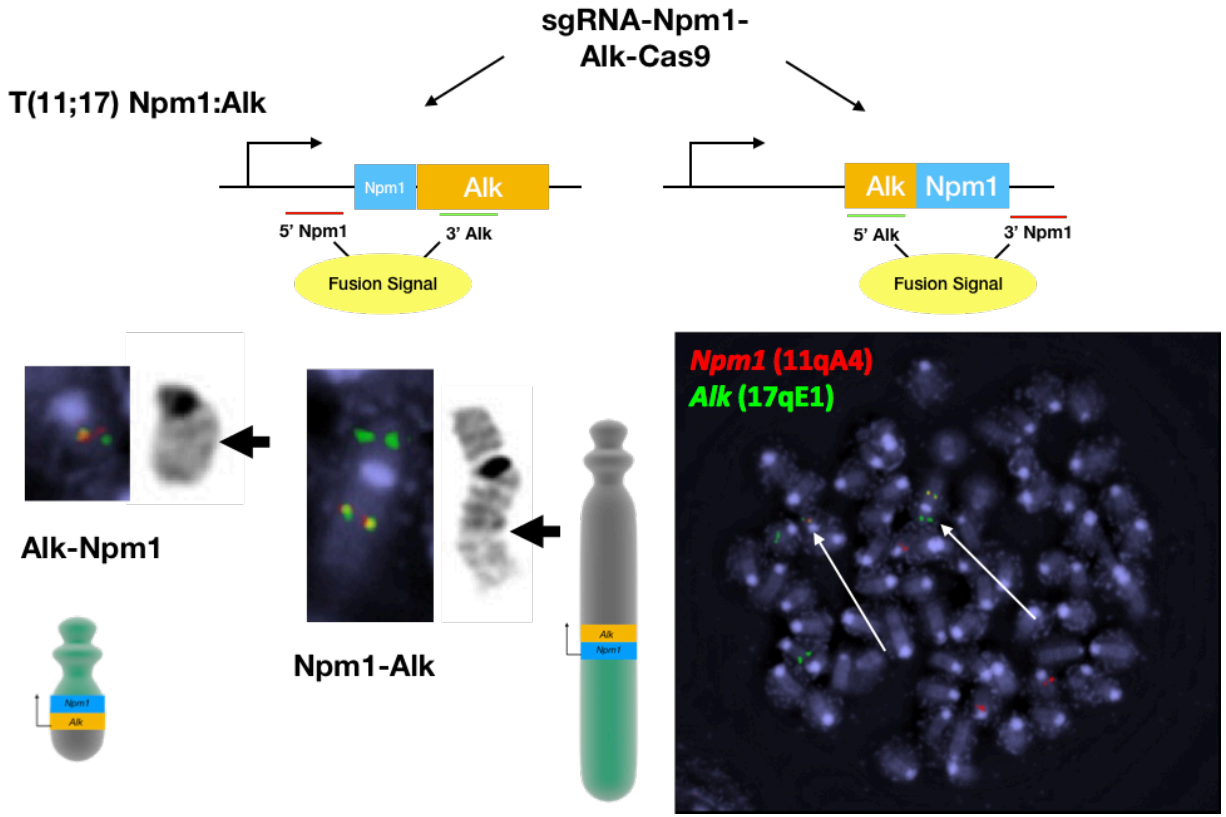


Figure 14. Cytogenetic Analysis of Ba/F3 Npm1-Alk cell line

(Top) Here the delivery of sgRNA targeting Npm1 and Alk along with Cas9 can induce the Npm1-Alk and Alk-Npm1 t(11;17) reciprocal translocation. This results in the juxtaposition of the Npm1 and Alk probes which can be then detected as a yellow fusion signal. (Right) On the FISH, there is a fusion signal seen for both the Alk-Npm1 and Npm1-Alk translocation (white arrows). (Left) The Alk-Npm1 translocation has the expected karyotype as illustrated by the diagram. For Npm1-Alk, the Robertsonian translocation involving two chromosomes 17 is present and has fused with the edited chromosome 11 containing Npm1, hence the wild type Alk and fusion signals in the FISH. This is illustrated by the karyotype with the black arrow showing the breakpoint. The region below the centromere (black dot) corresponds to the expected gene fusion as shown by the diagram.

NPM1-ALK fusion protein

Once the presence of *Npm1-Alk* was confirmed at the genomic level, I wanted confirm gene fusion protein expression. I screened several ALK and NPM1 antibodies because most are specific for the human ALK (hALK) but not murine ALK (mALK) (Bonzheim et al., 2015). I identified a polyclonal anti-ALK antibody (Abcam) that recognizes a common motif on the C terminus of the ALK. I detected the predicted fusion protein size of mNPM1-ALK (mN-A) as well as our control hNPM1-ALK cDNA (hN-A). I also treated the cells to see if *Alk* inhibitor Crizotinib treatment leads to degradation of the fusion protein which has been documented before (Pearson et al., 2012) (**Figure 15**).

ALK inhibition results in reduced cell viability of NPM1-ALK fusions

To quantify the sensitivity of our Ba/F3 fusion cell lines to drug treatment, I performed a titration of Crizotinib treatment to our mNPM1-ALK Ba/F3 lines. This was conducted using a commercial Cell Titer Glo Assay that measures cell survival (George et al., 2008). To rule out nonspecific effects of Crizotinib treatment, IL3 independent NPM1-ALK cell lines were also grown in IL3 during treatment. Because IL3 promotes survival, Crizotinib treated cell lines should be more viable with the presence of IL3 compared to drug alone. This would show that mNPM1-ALK generated by CRISPR-Cas9 is the driving oncogene for survival, which would be partially compensated by IL3 add back.

For the CRISPR generated mN-A cell line, treatment with Crizotinib caused specific reduction in growth compared to WT. This effect is most pronounced at lower concentrations while at higher concentrations of Crizotinib there is a non-specific toxic effect. This effect was partially ameliorated with add back of IL3 suggesting that NPM1-ALK fusion is the driver oncogene in these cell lines. These results indicate that NPM1-ALK can drive transformation of Ba/F3 cells in agreement with previous results with cDNA overexpression (**Figure 15**).

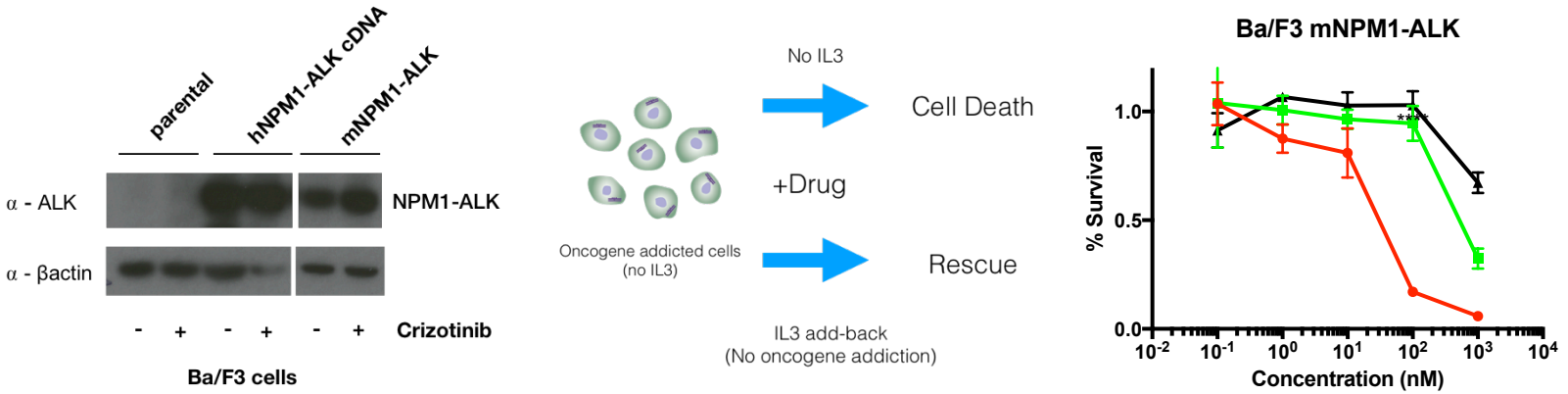


Figure 15. Detection of the NPM1-ALK fusion protein and oncogene dependence

(Left) Parental Ba/F3 cells along with those with human cDNA overexpression of Npm1-Alk and CRISPR mNPM1-ALK were collected for protein lysates and probed for NPM1-ALK fusion protein. ALK is not normally expressed in Ba/F3 cells while our cDNA and CRISPR Ba/F3 had detectable and predicted size of the NPM1-ALK fusion protein. Treatment with ALK inhibitor Crizotinib did not affect fusion protein stability. (Center/Right) To determine if our CRISPR generated mNPM1-ALK caused oncogene dependence on IL3 independent Ba/F3 cells, I performed an IL3 addback experiment upon specific inhibition of NPM1-ALK with the ALK inhibitor Crizotinib. NPM1-ALK cell lines were kept in either +IL3 (green) or -IL3 (red) and cell viability measured with Promega Cell Titer Glo Assay after three days post drug treatment. Parental Ba/F3 with IL3 grow in the presence of drug though there are concentration dependent nonspecific toxic effects. NPM1-ALK cell lines grown without IL3 have a dose dependent sensitivity to ALK inhibition. Addback of IL3 partially ameliorates the effect of Crizotinib treatment in both the cDNA and CRISPR forms of NPM1-ALK.

Ba/F3 mNPM1-ALK (-IL3) IC50 = 30.15nM
Ba/F3 hNPM1-ALK cDNA (-IL3) IC50 = 66.85 nM

Allograft flank injection of Ba/F3 mNPM1-ALK into nude mice

Because our *in vitro* data showed that CRISPR induced N-A fusion transformed Ba/F3 cells, I characterized our cell line in mice to test its ability to support *in vivo* tumor growth. Nude athymic mice that have a depleted immune system were used, so that tumor growth could be measured and tracked in an allograft model. One million mN-A or parental Ba/F3 cells were each prepared in Matrigel and injected into the flanks of nude mice (Cuesta-Dominguez et al., 2012). Rapid tumor growth was observed in mice injected with N-A compared to parental Ba/F3 where there was no measurable tumor growth. N-A injected mice had similar levels of growth and overall survival. None of the mice injected with WT Ba/F3 showed any tumor growth or died during the duration of the experiment. Upon collection of tumor masses, the *Npm1-Alk* fusion transcript was detected, indicating that our Ba/F3 mN-A line could support tumor growth (**Figure 16**).

Allograft tail vein injection of Ba/F3 mNPM1-ALK into BALB/c mice

To test the effect of our Ba/F3 cell lines *in vivo*, I employed the BALB/c mouse strain which was previously used to assess the tumorigenicity of Ba/F3 lines in syngeneic engraftment tumor models (Ilaria and Van Etten, 1995). Compared to the nude mouse, this mouse model still has an intact immune system. To accomplish this, I intravenously injected Ba/F3 resuspended in PBS into the tail vein of young adult BALB/c mice (~6 weeks). This would allow us to track tumor engraftment in the blood, marrow, and spleen and assess the tumorigenicity of our cell lines (**Figure 17**).

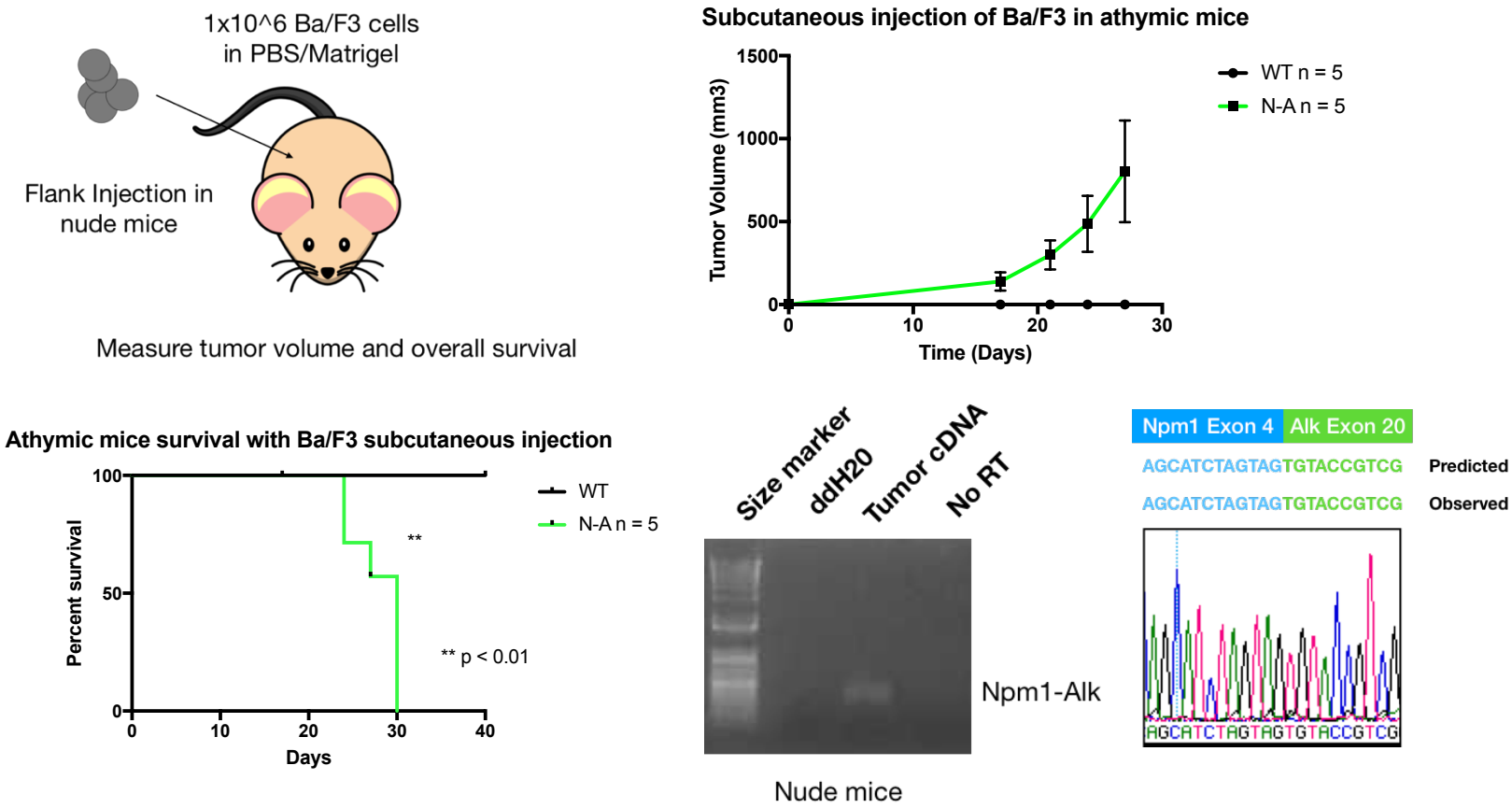
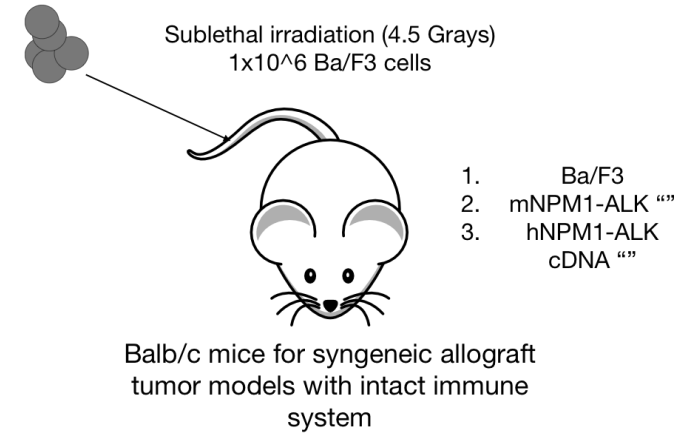


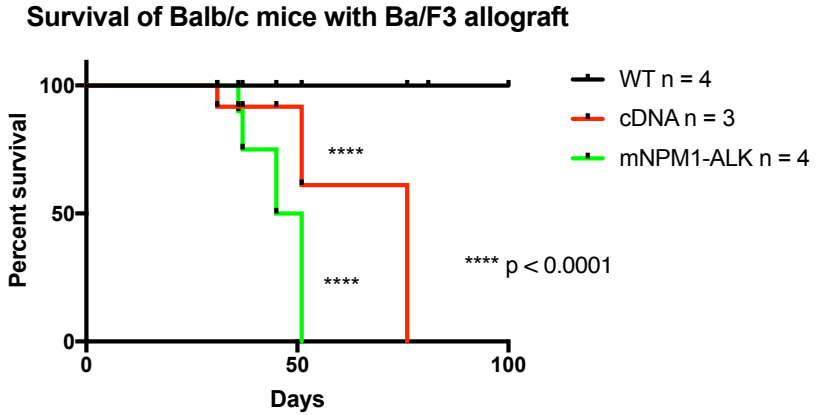
Figure 16. NPM1-ALK Ba/F3 cells support tumor growth in nude mice

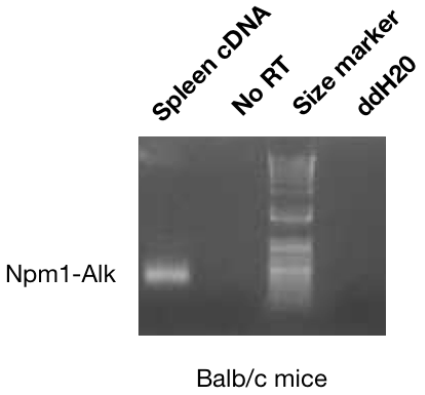
(Upper Left) To assess the transformation of our Ba/F3 cells containing the mN-A fusion, I used the immunocompromised nude mice as a mouse model of tumor growth. I prepared a PBS/Matrigel mix of 1 million parental or mN-A Ba/F3 cells and subcutaneously injected into the flank of nude mice. (Upper Right) I saw a rapid tumor growth which was not evident in those injected with parental cell line. (Lower Left) I also saw a rapid onset of disease and which required euthanasia compared to parental Ba/F3 cells. I was able to collect RNA and detect the presence of the Npm1-Alk fusion transcript.

Upon sub lethal irradiation of mice (4.5 grays), one million parental, hNPM1-ALK cDNA or mN-A Ba/F3 cells were injected through the tail vein of BALB/c mice and tracked for overall survival. Compared to earlier studies using cDNA overexpression of hNPM1-ALK, the mNPM1-ALK injected cell line resulted in a more rapid onset of disease (Slupianek et al., 2001) (**Figure 17**). Our mN-A cell line progressed faster than the hN-A cDNA overexpression positive control used (Miething et al., 2003), which may suggest a more aggressive phenotype with the murine translocation. However, there are still three copies of *Npm1* and *Alk* in our N-A lines compared to four in the parental Ba/F3 line. Mice injected with our parental Ba/F3 line showed no symptoms during the experiment. This data suggests that expression of the NPM1-ALK gene fusion has transformed Ba/F3 cells.

After mice were sacrificed, a complete necroscopy was conducted by a trained pathologist. Blood counts indicated a large increase in levels of nucleated lymphocytes and presentation of frank leukemia (**APPENDIX 1.2: Pathologist Reports**). After harvesting the spleen of a mouse injected with Ba/F3 expressing mN-A, I detected the *Npm1-Alk* fusion transcript (**Figure 17**). To analyze the effect of Ba/F3 allograft on the tissue level, splenic sections were placed on slides for immunohistochemistry (IHC) and cytogenetic analysis. I saw mALK staining on splenic sections which is not normally seen in wild type mice. There was also indication of widespread infiltration in the spleen by nucleated lymphocytes as determined by Hematoxylin & Eosin (H&E) staining (**Figure 18**).







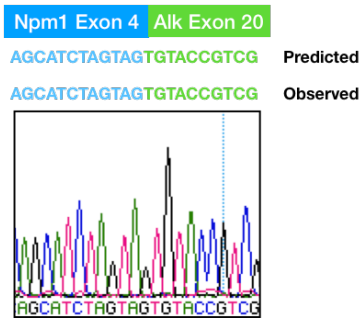


Figure 17. NPM1-ALK Ba/F3 cells cause accelerated disease onset

(Upper Left and Right) To assess the transformation of our Ba/F3 expressing NPM1-ALK in a syngeneic allograft tumor model with an intact immune system. WT BALB/c mice were sub lethally irradiated and injected with parental Ba/F3, cDNA overexpression of human N-A or CRISPR murine N-A. I then tracked overall survival and organ infiltration. Strikingly, CRISPR generated NPM1-ALK caused a faster acceleration of disease compared to the cDNA Ba/F3 cell line. It should be noted that this line has one less wild type copy of NPM1 and ALK due to CRISPR mediated translocation along with the reciprocal product. (Lower Left) I was able to extract RNA from the spleen and see the presence of the Npm1-Alk fusion transcript breakpoint.

Detection of Npm1-Alk fusion in organ infiltration

To confirm that these infiltrating cells were positive for the *Npm1-Alk* translocation, I conducted a tissue FISH for murine *Npm1* and *Alk* on splenic sections. I delineated the diploid tissue and the infiltrating lymphocytes, which are tetraploid and contain the *Npm1-Alk* translocation. These results suggest that Ba/F3 cells containing the mN-A fusion can infiltrate in the tissue and cause frank leukemia (**Figure 18**).

Conclusions

Through a series of experiments conducted *in vitro* and *in vivo,* I demonstrated that CRISPR can induce chromosomal rearrangements in a hematopoietic cell line. *In vitro* characterization showed that murine rearrangements generated by CRISPR recapitulates previous overexpression studies modeling the *Npm1-Alk* gene fusion. I demonstrated tumorigenicity with our transformed NPM1-ALK Ba/F3 cell line in an allograft mouse model. I should note that the polyclonal and hyperploid nature of my Ba/F3 mNPM1-ALK cell line do not reflect the underlying genetic context of the chromosomal rearrangement and should be optimized. Though informative, Ba/F3 cells do not reflect the cell of origin for ALCL and new models will be needed to better understand the biology of *Npm1-Alk* mediated transformation. This spurred me to develop a CRISPR viral vector that could be used for *ex vivo* modification of hematopoietic cells, where I could generate autochthonous mouse tumor models with the Npm1-Alk rearrangement.

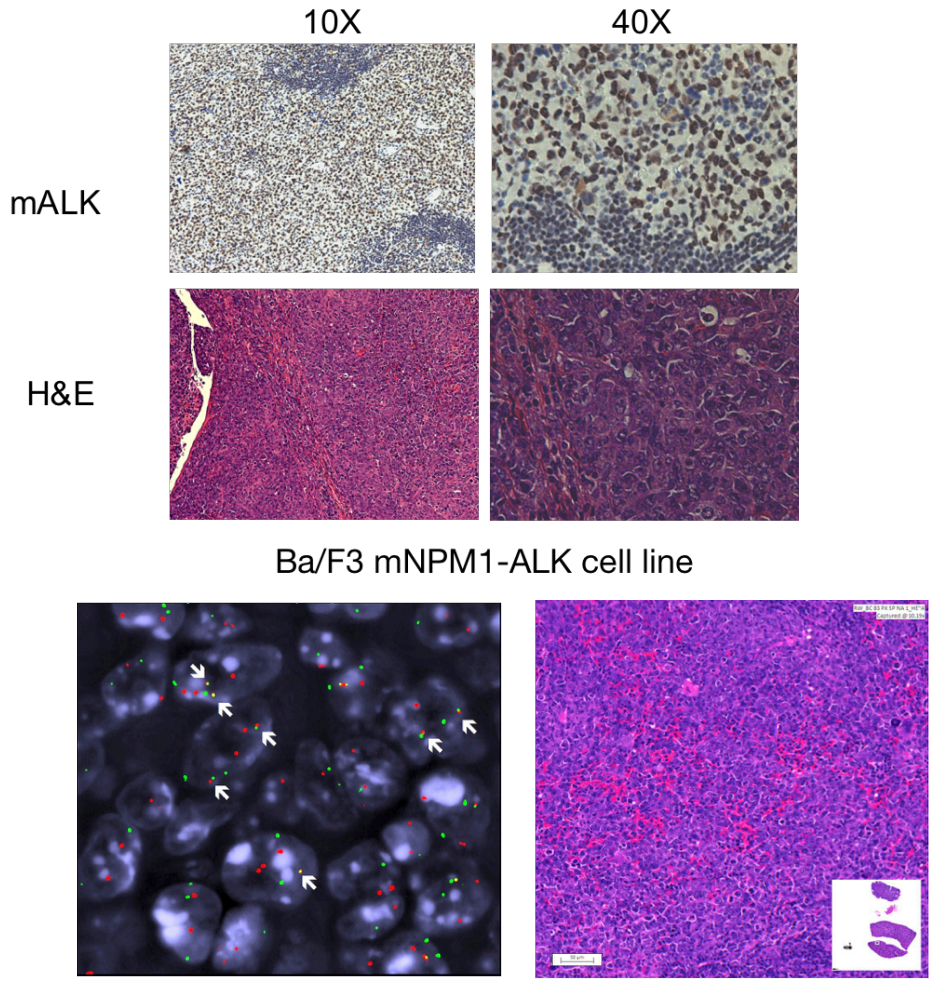


Figure 18. BALB/c splenic sections of mice injected with Ba/F3 NPM1-ALK

(Upper Section) Splenic sections show that there is ALK+ staining of nucleated cells suggesting infiltration. In addition, the H&E stain, shows that the normal splenic structure is no longer present in mice injected with Ba/F3 cells containing the fusion, but shows signs of massive infiltration by lymphocytic cells.
(Lower Section) Tissue FISH of splenic section for Npm1 and Alk confirm the presence of cells with the Npm1-Alk gene fusion (white arrows). The corresponding H&E is shown for reference.

CHAPTER 2: Optimizing CRISPR-Cas9 expression in viral vectors

Through evolutionary selection, viruses can efficiently deliver their genome by infecting cells and using the cellular machinery of the infected cells to produce the components needed for replication (Hu and Pathak, 2000). The discovery that viruses can carry host DNA and uptake of cellular oncogenes resulted in transformation upon transduction was efficient means of gene transfer between mammalian cells (Oppermann et al., 1979). This led to the idea of using viruses as vectors to deliver exogenous genes to study gene function (Bouard et al., 2009). The first viral vectors were derived from the simple retroviruses, dsRNA viruses from the genus *Gammaretroviruses* (γ-retroviruses). Later on, retroviruses derived from the Human Immunodeficiency virus called lentiviruses were developed as viral vectors as well.

Reverse transcription and integration

As the RNA retroviral genome enters the cell, reverse transcriptase generates the dsDNA template. This process was elucidated in the 1970's by Temin and colleagues (Temin and Mizutani, 1970). They discovered that tRNA from the host cell binds to the primer binding site on the RNA genome and acts as a primer for reverse transcription. Reverse transcriptase then produces a ssDNA template resulting in RNA/DNA hybrid intermediate step. Then the newly formed 3' LTR DNA acts as the template for dsDNA template formation through a series of jumps that results in the degradation of the initial RNA template and formation of the double stranded DNA viral genome. This dsDNA template is then integrated in the genome through viral integrase as a provirus (Ryu, 2017)

Transcriptional units of the γ-retroviral genome

Once integrated, the provirus becomes a transcriptional unit for viral mRNA production. The basic construction of a simple retrovirus consists of flanking 5' and 3' identical long terminal repeats (LTRs) that each contain the U3, R and U5 regions. The

U3 region contains enhancer and regulatory signals that mediate Pol II driven transcription of the viral genes gag, pol, and env. Though identical, the virus strictly uses the 5' LTR and 3' LTR for transcriptional initiation and termination respectively. This is achieved through the use of the TATA box and polyadenylation site, which is used by the 5' LTR and 3' LTR respectively. The U5 region contains elements involved in reverse transcription of the RNA viral genome. During the retroviral viral life cycle, the U3 and U5 regions of 5' LTR in the viral RNA genome is regenerated from the 3' LTR during reverse transcription (Delviks-Frankenberry et al., 2011).

Retroviruses are a versatile tool

The γ-retroviruses present a robust delivery system in mouse models owing to several factors. As they were also the first to be adapted as experimental tools with established protocols for design and production at low cost. As the viruses also integrate into transduced cells, there is strong long-term expression that can be propagated and tracked in dividing cells. Viral vectors can also hold multiple transgenes that can range from proteins, small RNAs, and selectable markers. Given these traits, the γ-retrovirus was quickly adapted for gene transfer by researchers. Following this, the first clinical trials used γ-retroviruses, though that method has waned more recently due to the propensity of the provirus to cause insertional mutagenesis near oncogenes such LMO2 (Baum et al., 2004).

Viral vectors offer a facile method of interrogating gene function that can be used in most laboratory settings. The versatility of viruses as a tool can be attributed to the ease of engineering the viral vector genome (Heinz et al., 2011). Viral vectors can be propagated as a plasmid in bacterial stocks, which allows for molecular cloning and long-term storage. This allows for scalable design and production of viral vector libraries with cloning protocols such as Gibson Assembly (Gibson et al., 2009). There are now

established protocols for preparing genetic screens using short hairpin (shRNA) and CRISPR retroviral libraries (Sanjana et al., 2014; Sims et al., 2011).

Viral vectors are a robust and cost-effective method of transducing cells *in vitro* and *in vivo*. Preparing virus involves supplying packaging cells like human embryonic kidney 293 SV40 Large T-antigen (HEK293T) with the necessary replication factors *in trans*, followed by viral supernatant collection for direct transduction or further processing. Protocols were established where viral vectors can be used to modify challenging primary tissue such as hematopoietic stem cells *ex vivo* followed by adoptive transfer into irradiated hosts (Haviernik et al., 2008). For example, researchers can alter the tropism and functional properties of the virus through pseudotyping, in which a different viral envelope is supplied during viral production and allows for expanded target cell transduction and virus concentration (Walther and Stein, 2000). Retroviral transduced cells retain long term expression of the viral transgene and can be used to track cells over time (Persons et al., 1997). Viral transduction remains one of the most cost effective to delivery genes into cells.

Though there are reports that simple and lentiviruses have equal transduction efficiencies, the γ-retrovirus system was predominantly used to study the murine hematopoietic system (Hawley et al., 1992). Anecdotal and personal experiences indicate that the lentivirus is less efficient at transducing murine hematopoietic stem cells (Simmons and Alberola-Ila, 2016). On the contrary, it has been shown that lentivirus can transduce human CD34+ stem cells efficiently (Ramezani et al., 2000) though no mechanism has been proposed to explain this difference. Many of the steps in producing lentivirus, such as envelope pseudotyping, can also be adapted for simple retroviruses.

Adapting gammaretroviruses for CRISPR-Cas9 delivery and gene editing

The most suitable delivery system of nucleases for modeling chromosomal rearrangements would, ideally, be easy to prepare on a large scale to model the gene fusion using tandem sgRNAs. The vector needs to be highly efficient in targeting cells and once inside would express the CRISPR components to induce chromosomal rearrangements. I wanted to target the murine hematopoietic system, so I focused on γ-retroviruses because most of the viral transduction protocols of murine HSCs used γ-retroviruses. As γ-retroviral vectors are adapted to infect a variety of cell types and robustly express their transgenes, they are preferred for applications where the cell type is not amenable to other means of gene delivery. Though Cas9 DNA transfection and Cas9 ribonucleic proteins (RNPs) were successfully used in a variety of cell types, primary hematopoietic stem cells remain resistant to these methods (Aubrey et al., 2015; Modarai et al., 2018). To adapt γ-retroviruses as tools for CRISPR genetic dissection, I searched for existing retroviral vectors.

In the Retroviridae family, one gammaretrovirus, the Murine stem cell virus (MSCV) was extensively used in transducing murine hematopoietic stem cells (Cherry et al., 2000; Hawley et al., 1992) which was derived from the Murine embryonic stem cell virus (MESV)(Grez et al., 1990). It was found to transduce and express in the stem cell population more efficiently than other γ-retroviral vectors such as MLV. The MSCV expression cassette maintains expression in hematopoietic stem cells and in differentiated progeny while other retroviral vectors are more prone to transcriptional silencing (Ellis, 2005). This is critical when modeling hematological malignancies as it allows for continued transgene expression and long-term tracking of modified cells.

All in one Cas9 γ-retroviral vectors

With the advent of CRISPR libraries (Chen et al., 2015) and sgRNA design software (Perez et al., 2017), I sought to develop an all-in-one viral vector suitable for the hematopoietic system and amenable to large scale vector library production. Instead of targeting just one gene, I wanted to target two sites simultaneously to generate genomic rearrangements. In addition, though my previous worked showed that nucleofection of CRIPSR plasmids into Ba/F3 could induce the *Npm1-Alk* translocation, this may not be applicable to primary cell types. Viral vectors have been shown to transduce many cell types and successfully deliver gene cargo. Work in our lab showed that adenovirus expressing tandem sgRNAs and Cas9 could induce chromosomal rearrangements, though it has not been widely reported that adenoviruses can transduce murine hematopoietic stem cells efficiently. Taking all these factors into consideration, I endeavored to make a MSCV vector with tandem sgRNA and Cas9 expression.

In generating a MSCV Cas9 vector, I noticed that the total size of Cas9, sgRNAs and GFP marker made it larger than the typical genome size of MSCV (~8kb) which is known to affect titer (Kumar et al., 2001). In addition, optimization of CRISPR expression would be crucial if I were to use my vector to model rearrangements. Conceptually, the ability to induce gene fusions will depend heavily on the kinetics of CRISPR gene editing and DNA repair, which is largely beyond my control. Despite these limitations, I felt that an all-in-one viral vector, once optimized would be a much simpler method for modeling gene fusions (Maddalo and Ventura, 2016).

Cloning of gammaretroviral CRISPR Cas9 vector

To express multiples sgRNAs in an all in one vector, members from our lab modified the pX330 by adding a second hU6 promoter and sgRNA structure with a different target sequence cloning site, BsaI, instead of Bbs1, which was called pX333 (Maddalo et al., 2014). SgRNAs to *Npm1* and *Alk* were then sequentially cloned into our pX333 vector in the tandem orientation. In theory, expressing two sgRNAs in a single vector would be easier and provide a more consistent payload when delivered into cells. Because the goal was to induce rearrangements, having the desired guides in one vector with equimolar expression would be more efficient for one step generation of fusions.

Cloning CRISPR into the MSCV backbone (MSCV Cas9)

Having the dual sgRNA cassette in pX333 allowed us to directly subclone into our vector of choice. Our first generation of the MSCV Cas9 vector, derived from the MSCV puromycin IRES EGFP (PIG) vector backbone (Schmitt et al., 2002), contained tandem sgRNAs driven by human U6 Polymerase III (Pol III) promoters (Goomer and Kunkel, 1992). This was followed by CBh promoter with N-terminal FLAG and NLS tagged Cas9 and with internal ribosomal entry sequence (IRES) and green fluorescent protein (GFP) **(Figure 19)**. The integrated provirus maintains the viral LTR Pol II promoter while Cas9 and sgRNAs is driven by the internal promoters. This use of heterologous promoters was seen previously when small interfering RNAs (siRNA), were first driven by Pol III promoters in the MSCV backbone (Devroe and Silver, 2002; Fellmann et al., 2013). However, as shRNA can be expressed as a Pol II transcript, and then subsequently processed into siRNA, Pol II promoters are used more often (Fellmann et al., 2013). That supported the idea that γ-retroviral vectors could support heterologous promoters, though this has not been systematically tested in CRISPR viral vectors.

MSCV Cas9 vector failed to induce Npm1-Alk translocation

Transduction efficiency was quite low with MSCV Cas9 vector nor could Npm1-Alk translocation be detected, so I enriched for a GFP-positive population with fluorescence activated cell sorting (FACS) from transduced NIH-3T3. I could then attempt to determine the level of *Npm1-Alk* sgRNA expression as I hypothesized that expression of sgRNA may have been affected. To test this, a Northern blot for the *Npm1* and *Alk* sgRNAs showed no expression in the sorted transduced cell population. In the packaging cells there was robust sgRNA signal, indicating that the Pol III promoter is functional when transfected into packaging cells. That suggested that the expression of sgRNA was being disrupted at the level of the provirus. I began to consider that the viral LTR transcribes all along the expression cassette and may interfere with intervening sequence if it is not also a Pol II transcript. In addition, using tandem homologous U6 promoters in viral vectors is not advisable due to sequence homology and subsequent viral recombination of expression cassettes (Vidigal and Ventura, 2015) (**Figure 19**).

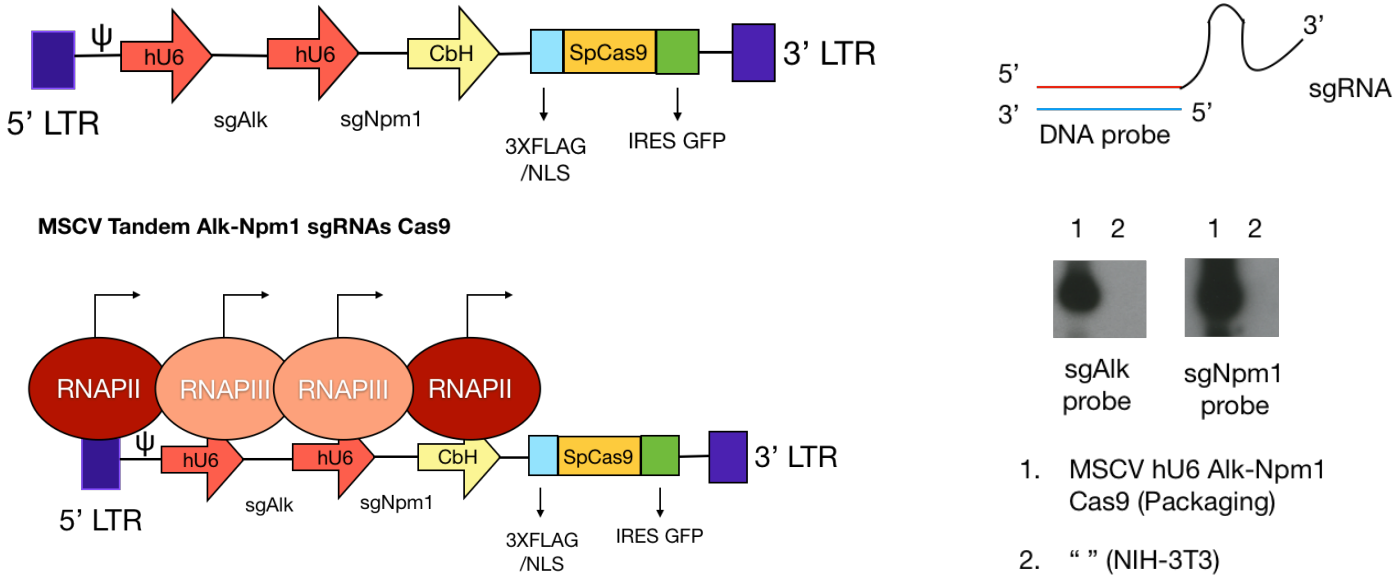


Figure 19. Gammaretroviral CRISPR vector fails to express sgRNA expression

(Upper Left) Shown here is tandem sgRNA expression with Cas9 MSCV retroviral vector. Here Pol III human U6 promoters that drives expression of the Alk and Npm1 sgRNA. This is followed by the CBh Pol II promoter that expresses a 3X FLAG tagged with Nuclear localization sequence (NLS) Cas9 bicistronic transcript with an IRES GFP. (Right) I assessed sgRNA expression through a small RNA Northern Blot with a DNA probe with reverse complementarity toward the unique targeting sequencing of the sgRNA. Surprisingly when NIH-3T3 cells are transduced with our MSCV vector with CBh driven Cas9 IRES EGFP and human U6 Pol III promoter driving Alk and Npm1 sgRNA (hU6 A-N), I do not see expression in a GFP sorted transduced population containing the provirus (Lane 2). However, in Lane 1, expression of sgRNA from the vector is confirmed as this is directly transcribed from the plasmid transfected into HEK293T packaging cells. (Lower Left) Shown here is the transcriptional output of the integrated provirus. Note that our Pol III transcripts is in the same orientation as the viral mRNA driven by the active 5' LTR Pol II promoter (in tandem). followed by the Pol II CBh promoter. This configuration may lead to promoter or transcriptional interference and reduce the efficiency of CRISPR mediated gene editing.

Retroviral vector design

To rule out the possibility that viral promoter or downstream promoters were affecting sgRNA expression, I looked into vectors that inactivate the viral promoter of the integrated provirus. These self-inactivating gammaretroviruses (Schambach et al., 2009) are generated by deletion of the U3 promoter region in the 3' LTR of the viral vector. In the process of replication and integration, the 3' LTR becomes the template of the 5' LTR sequence in the provirus. This results in an inactive viral promoter and requires an internal promoter for transgene expression. These self-inactivating retroviruses have been used previously, but one downside is that they have lower titers (Maetzig et al., 2011). Their main advantage is their improved safety profile as the viral promoter is less likely to drive ectopic expression of genes near the insertion site of the retrovirus (Uren et al., 2005).

Promoter interference in viral expression cassettes

It has been long observed in early viral vectors that the presence of the viral promoter disrupts the expression of downstream genes (Cullen et al., 1984; Emerman and Temin, 1984; Nakajima et al., 1993; Schambach et al., 2006). There remains a possibility that Pol II and Pol III undergo steric hindrance which results in inefficient Pol III transcription (Nie et al., 2010). Although there are viral vectors that express shRNA or other small RNAs without any reported issues, these used Pol II promoters. Furthermore, the effect may not have been noticeable if the reduction in expression did not affect phenotypic output. Many of the early vectors initially had issues with expression though this has improved over time (Curtin et al., 2008; Ginn et al., 2003). However, CRISPR expression in viral vectors has not been thoroughly studied and may require further optimization. A recent report mentioned that Pol III promoters such as U6 can drive Pol II expression of markers (Gao et al., 2018c).

Design of a divergent self-inactivating MSCV viral vector (pSUPER.Retro)

To alleviate this promoter interference, I once again referred to the literature for designing expression cassettes that reduces promoter interference. Several designs were considered, including placing the expression cassette in a divergent manner. Divergent transcription is a common feature found in metazoans (Seila et al., 2009). I hypothesized that having the promoters on different strands could possibly alleviate promoter occlusion by RNA polymerase (Greger et al., 1998).

Orientation of promoters

The orientation of transcriptional units depends on how the gene and promoter is placed on the DNA strand. For example, promoters in the convergent orientation face each other with the direction of transcription converging towards each other. However, their transcripts do not overlap as they are transcribed on different strands of DNA. In tandem promoters, one gene is downstream of the other on the same DNA strand. This results in a transcriptional output that can lead to readthrough of downstream genes like our earlier vectors. In divergent promoters, there may be partial or no overlap of sequence, but the transcripts transcribe away from each other (Shearwin et al., 2005). These orientations can alter the efficiency of transcription as RNA polymerase loading and processivity can be affected due to transcriptional interference. I sought to find a viral vector that minimized these transcriptional effects to generate the highest titer and expression of CRISPR-Cas9.

I surmised that expressing CRISPR-Cas9 transcripts from the same viral vector could look toward the early days of RNA interference (RNAi) for inspiration, where Pol III promoters were used to express small RNAs. One of the earliest retroviral vectors developed for shRNA expression is called pSUPER.Retro (pSR), which is based off the MSCV backbone, though modified to be a self-inactivating version (Brummelkamp et al., 2002a, b). This vector has shRNA driven by the Pol III H1 promoter in the non-overlapping

divergent orientation from the Pol II human phosphoglycerate kinase (PGK) promoter that drives puromycin expression (**Figure 20**). The PGK promoter is often used for *in vivo* applications because it can be stably expressed *in vivo* compared to other promoters that can express transgenes to toxic levels (Norrman et al., 2010; Qin et al., 2010).

I hypothesized that any transcriptional interference of our Pol III transcripts from the Pol II promoter would be alleviated as promoter occlusion is minimized while the LTR in the provirus is no longer functional to drive expression. I adapted the self-inactivating MSCV vector (pSRD) to include sgRNA driven by U6 promoters in the divergent orientation (Kabadi et al., 2014). However, I decided against using H1 promoters in the future as it has been reported to have variable expression of small RNAs (Ma et al., 2014b). One study reported that murine U6 promoter is less efficient in human and murine progenitor cells than the human homologue for shRNA expression in a lentiviral vector (Roelz et al., 2010). To address viral recombination and expression concerns, I replaced one of the sgRNA Pol III promoters with a synthetic U6 (sU6) promoter which has robust expression and sequence divergence from the other U6 promoters (Vidigal and Ventura, 2015) (**Figure 20**).

Optimizing Cas9 EGFP expression

Despite the improvements in sgRNA expression, I needed to assess expression of Cas9 and fluorescent marker because there are several fusion protein designs. I had initially used the Cas9 IRES GFP, while other labs have used the Cas9 EGFP with 2A self-cleaving peptide (P2A) variant. Previous work has suggested that P2A fusion proteins result in more stable expression then IRES sequences (Mizuguchi et al., 2000; Szymczak and Vignali, 2005). Maintaining Cas9 protein expression would be critical in ensuring that genome editing occurs by binding to all available sgRNAs and forming functional complexes. To address this, I replaced the puromycin gene with either FLAG-tagged 3X NLS Cas9 IRES GFP or FLAG tagged 3X NLS Cas9 P2A EGFP (all in one) or a mCherry

only that were each driven by the human PGK Pol II promoter **(Figure 20)**.

MSCV divergent Cas9 vectors improve sgRNA expression

When I transduced NIH-3T3 cells with pSRD vectors containing Cas9 P2A EGFP or Cas9 IRES EGFP, I saw comparable transduction efficiency as my original MSCV tandem sgRNA Cas9 vector. This suggests that this orientation did not significantly improve viral titer though this is something that will need to be optimized. However, when I measured expression of *Npm1* and *Alk* sgRNA by Northern Blot, I observed that sgRNA expression was present in only our pSRD vectors in transduced NIH-3T3 cells **(Figure 21)**. This suggests that our self-inactivating MSCV vectors with divergent expression cassettes may have partially relived transcriptional interference of our sgRNA expression. For simplicity, I decided to keep future iterations of our vectors to the Cas9 P2A EGFP version.

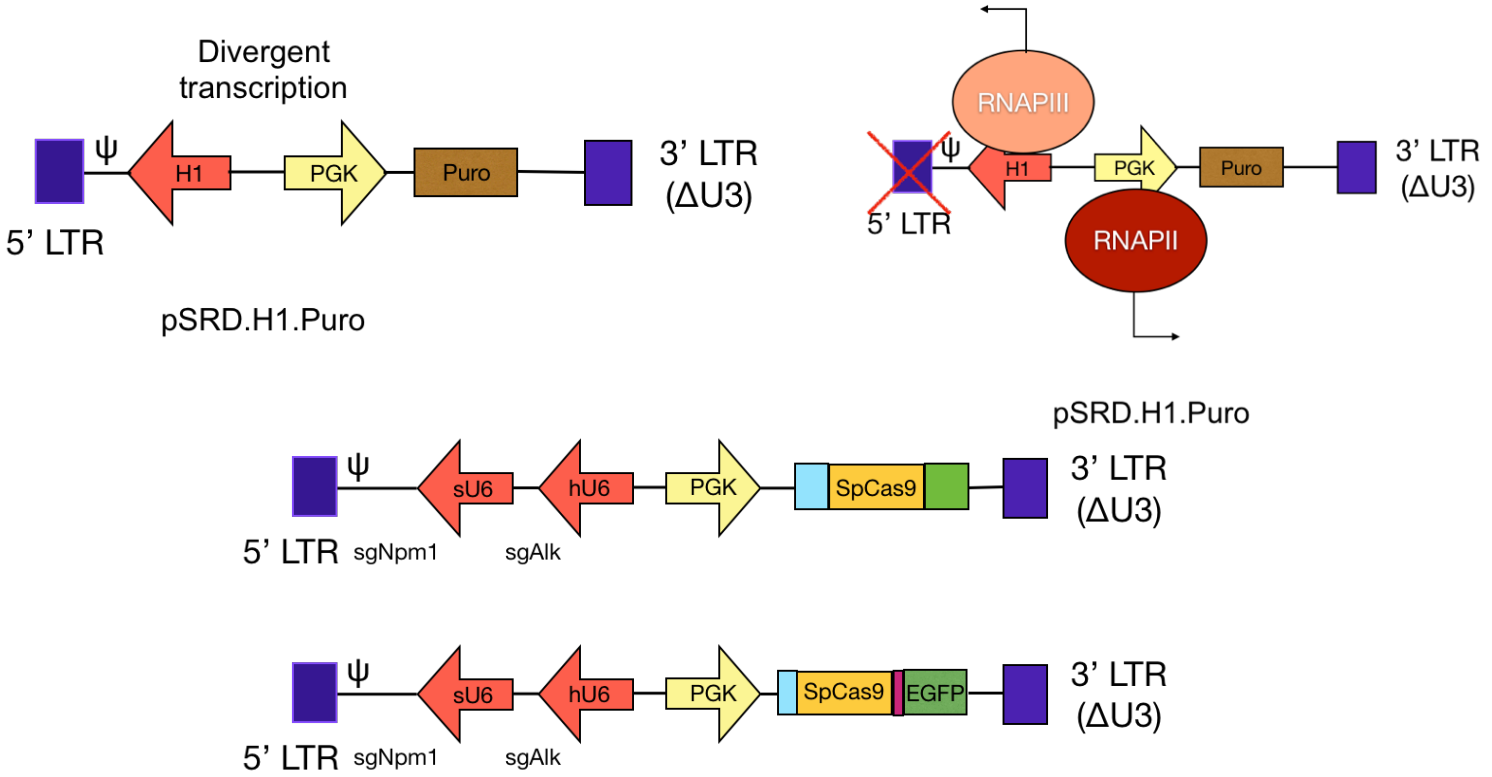


Figure 20. pSuper.Retro Cas9 divergent all in one vector (pSRD)

(Upper Left) Shown here is the pSUPER.Retro vector that addresses some of the shortcomings of our previous vector. Here the MSCV backbone has a 3' LTR with the self-inactivating U3 deletion. When this virus integrates as a provirus the 3' LTR regenerates the 5' LTR and is no longer capable of Pol II transcription. Here the H1 Pol III promoter transcribes toward the 5' LTR and is divergent from the PGK Pol II promoter. (Upper Right) Illustrated here is the transcriptional orientation of the various expression cassettes. (Bottom) The key change from the earlier version is that the hU6 promoter is in the divergent transcriptional orientation in relation to the PGK driven Cas9 to minimize potential promoter interference. This was achieved by inverting the dual sgRNA expression cassette. Note that I showed a synthetic U6 (sU6) Pol III promoter as it has less homology to the hU6 promoter yet maintains robust Pol III expression. I also made versions containing Cas9 IRES GFP or Cas9 P2A EGFP to compare transduction efficiency.

.

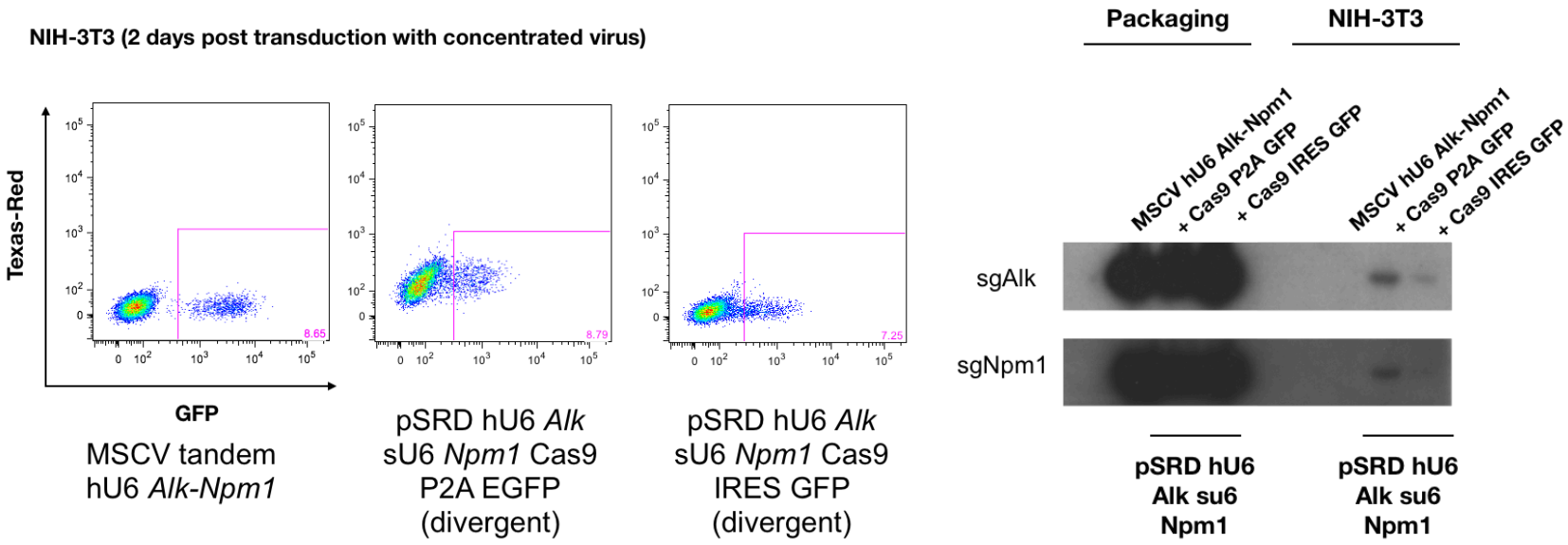


Figure 21. pSRD vectors relieve transcriptional repression in transduced cells

(Left) Our original MSCV tandem CRISPR vector transduce at comparable rates to our pSRD vectors CRISPR vectors in murine NIH-3T3 cells and suggests that the vector size plays a large role in determining transduction efficiency. (Right) Expression of the sgRNAs between the vectors is comparable among transfected packaging cells via Northern Blot. Strikingly, there is a clear signal for sgRNA in our pSRD vector with our P2A EGFP and IRES GFP version not present in the MSCV version.

Generation of MSCV convergent CRISPR vector

Given that our pSRD vectors had low viral titer, I thought that it could still be improved. As viral mRNA is transcribed from the 5' LTR of the provirus via Pol II, having any expression cassette with a Pol III promoter directly downstream of the 5' LTR may still be disruptive to transcription. I was inspired by a publication that used a "double copy" MSCV vector to express a shRNA from the U3 region of the 3' LTR (Luo et al., 2004) to boost expression while remaining away from the 5' LTR during viral mRNA transcription. This same group the developed a MSCV vector with an LTR driven Cas9 P2A EGFP and hU6 driven sgRNA in the convergent orientation (Li et al., 2016a). This vector is called pXZ201 (pX), and it also contained Bbs1 cloning sites which is compatible with our lab's dual cloning method (Vidigal and Ventura, 2015). This also saves several hundred base pairs in the size of the vector, which should help with increasing viral titer **(Figure 22)**.

To this end, I generated a pX vector with sgRNAs targeting *Npm1* and *Alk* with this orientation to maximize viral mRNA transcription and minimize potential transcriptional interference. I also adapted our self-inactivating pSRD vector to have *Npm1* and *Alk* sgRNAs expressed in the convergent orientation (pSRC) with an internal PGK promoter to drive expression of Cas9 P2A EGFP **(Figure 22).** I was able to show that our pX and pSRC vectors can transduce NIH-3T3 with unconcentrated virus (**Figure 23**).

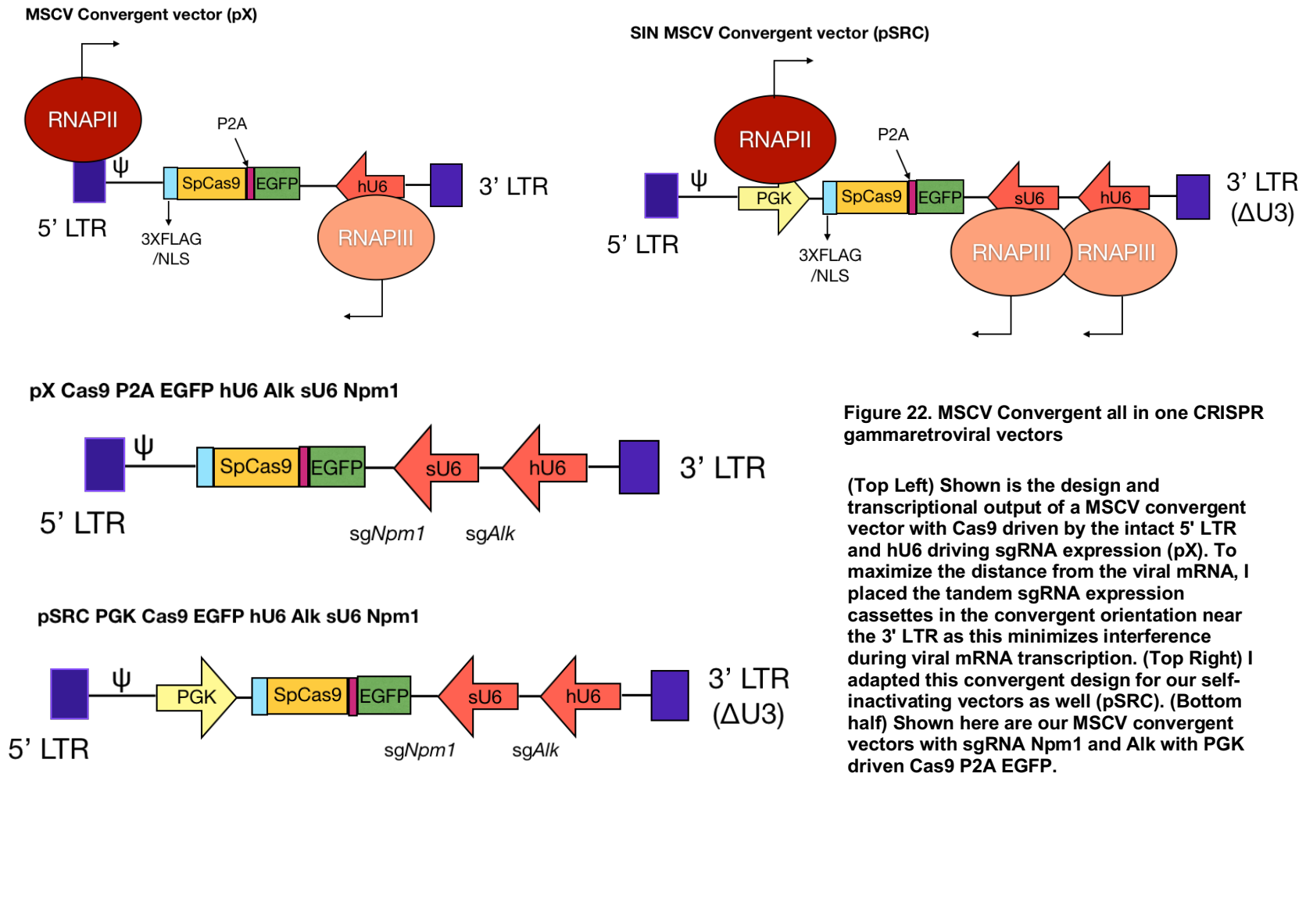


Figure 22. MSCV Convergent all in one CRISPR gammaretroviral vectors

(Top Left) Shown is the design and transcriptional output of a MSCV convergent vector with Cas9 driven by the intact 5' LTR and hU6 driving sgRNA expression (pX). To maximize the distance from the viral mRNA, I placed the tandem sgRNA expression cassettes in the convergent orientation near the 3' LTR as this minimizes interference during viral mRNA transcription. (Top Right) I adapted this convergent design for our self-inactivating vectors as well (pSRC). (Bottom half) Shown here are our MSCV convergent vectors with sgRNA Npm1 and Alk with PGK driven Cas9 P2A EGFP.

Concentration of pseudotyped retrovirus

Due to the low viral titer I observed with my previous viral vector using only viral supernatant with spinfection, I determined that viral concentration was necessary as the size of the CRISPR components reduces viral titer significantly. Though many applications that require viral production can use packaging cell lines for fresh viral supernatant, they cannot be stored long term (Burns et al., 1993) or concentrated due to the fragile nature of the viral envelope. That is problematic for low viral titer preparations where larger volumes of virus are required for concentration.

To increase the concentration of my γ-retrovirus preparations, I used a method called viral pseudotyping. This procedure involves replacing the viral envelope involved in viral recognition of the cell receptor and transduction (Ichim and Wells, 2011). I used the Vesicular Stomatitis Virus Glycoprotein (VSVG) envelope due to its ability to withstand the polyethylene glycol (PEG) and ultracentrifugation steps needed to concentrate virus (Ichim and Wells, 2011). Though commonly used to concentrate lentiviruses, pseudotyping of simple retroviruses has been performed (Ichim and Wells, 2011; Reya et al., 2003). For producing pseudotyped virus, I used GP2-293 cells (Clontech), for viral production as it already has integrated gag/pol, so only the VSVG envelope and transfer plasmids are required (Kines et al., 2006). I then concentrated virus using the commercially available Retro-X concentrator (Clontech) reagent, which also allows for long term storage of viral titer preparations. I transduced NIH-3T3 cells with the Retro-X concentrated viruses from our pX and pSRC vectors, and I found most of the cells to be GFP positive **(Figure 23).** I confirmed that sgRNA expression was stable in transduced cells and detected *Npm1-Alk* cDNA fusion transcript **(Figure 24)**.

MSCV convergent vectors induce Npm1-Alk translocation in primary tissues

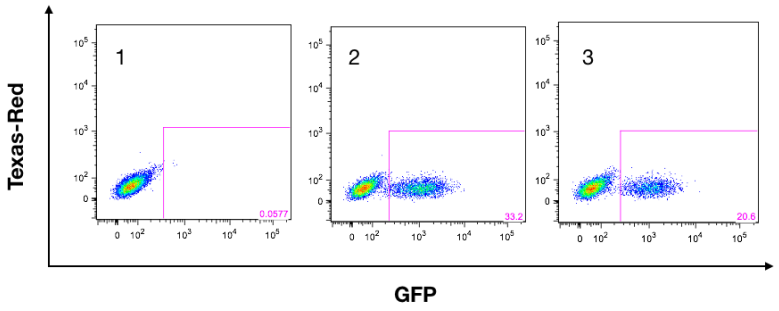
Though our vectors could induce the *Npm1-Alk* translocation in the immortalized NIH-3T3 cell line *in vitro,* this may not translate to success in diploid primary cells. Previous

results show that NIH-3T3 is tetraploid, which may markedly ease induction of the *Npm1-Alk* fusion due to the presence of additional target sites (Maddalo et al., 2014). To test this, I used concentrated pX and pSRC virus on low passage p53 null primary mouse embryonic fibroblasts from C57BL/6 mice. I was able to detect the *Npm1-Alk* translocation in genomic DNA for the first time in primary tissue **(Figure 24)**, which I may be due to the fact that there is a threshold level of sgRNA expression for optimal editing to induce the correct rearrangement (Yuen et al., 2017).

Conclusions

I showed that our gammaretroviral vector can be used to induce inter-chromosomal translocations *in vitro* in NIH-3T3 cells and primary p53 null MEFs. This suggests that our vector can be used to model chromosomal rearrangements though more cell types will have to be tested. I also note that having the convergent orientation for Cas9 and sgRNA expression cassettes in MSCV based vectors minimizes interference and maximizes viral titer. Currently, there are no reports of transcriptional interference in retroviral vectors with promoters expressing sgRNAs. This suggests that current designs based off shRNA expressions systems may not be optimized for CRISPR expression.

1. NIH-3T3
2. “” pX empty Cas9 P2A EGFP
3. “” pX hU6 Alk sU6 Npm1 Cas9 P2A EGFP



1. NIH-3T3
2. “” pSRC Cas9 P2A EGFP only
3. “” pSRC Cas9 P2A EGFP hU6 Alk sU6 Npm1

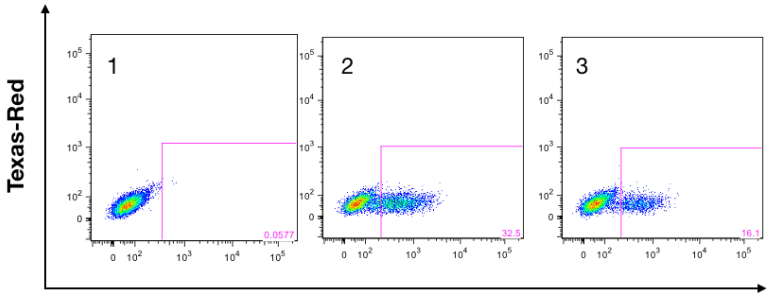
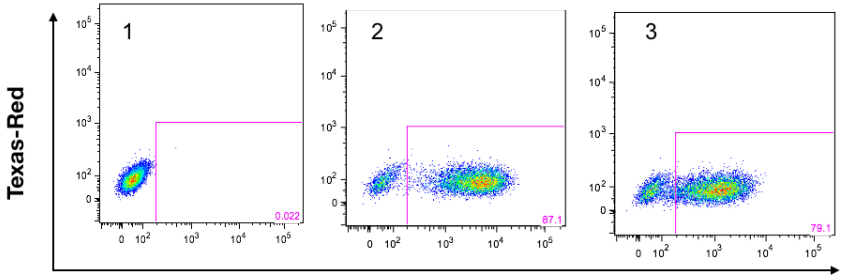


Figure 23. MSCV Convergent all in one vectors can transduce NIH-3T3

(Top Left) Our MSCV based retroviral vector for CRISPR-Cas9 expression (pX) can transduce NIH-3T3. (Bottom Left) Our pSRC vector can transduce NIH-3T3 as unconcentrated virus shows improved transduction over previous divergent versions. Adding sgRNA reduced titer compared to empty Cas9 MSCV vectors. (Bottom Right) To increase titer, I concentrated our virus using Retro-X commercial reagent and transduced NIH-3T3. I saw an increase in transduction efficiency which I then harvested for downstream analysis.



1. NIH-3T3
2. “ “ pX hU6 Alk sU6 Npm1
3. “ “ pSRC hU6 Alk sU6 Npm1

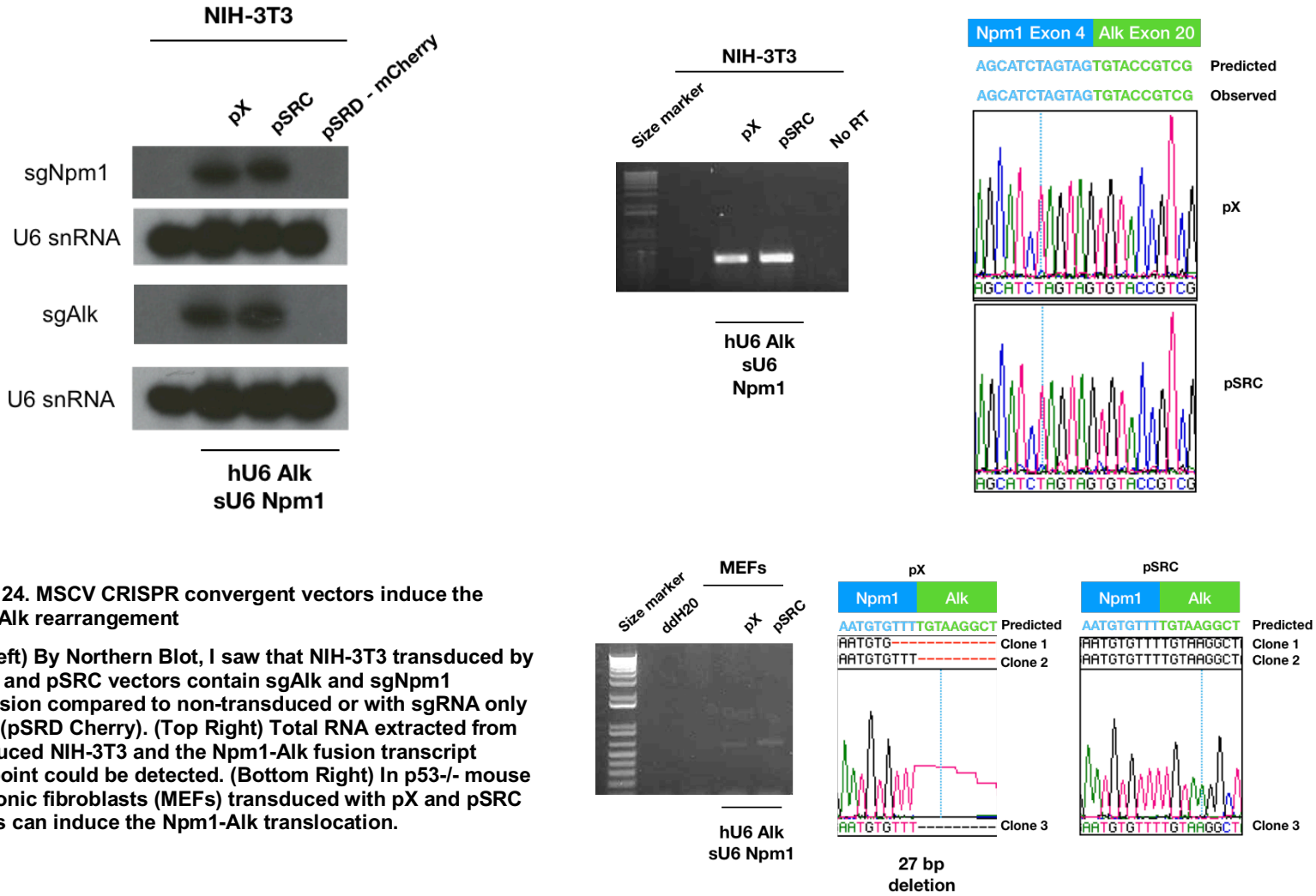


Figure 24. MSCV CRISPR convergent vectors induce the Npm1-Alk rearrangement

(Top Left) By Northern Blot, I saw that NIH-3T3 transduced by our pX and pSRC vectors contain sgAlk and sgNpm1 expression compared to non-transduced or with sgRNA only vector (pSRD Cherry). (Top Right) Total RNA extracted from transduced NIH-3T3 and the Npm1-Alk fusion transcript breakpoint could be detected. (Bottom Right) In p53-/- mouse embryonic fibroblasts (MEFs) transduced with pX and pSRC vectors can induce the Npm1-Alk translocation.

CHAPTER 3: DISCUSSION AND FUTURE DIRECTIONS

Generation of Npm1-Alk gene fusion in Ba/F3 cells

Using the Amaxa nucleofection based delivery of plasmid DNA expressing tandem sgRNAs targeting *Npm1* and *Alk* with Cas9, I was able to induce and select for a polyclonal population with the *Npm1-Alk* gene fusion in the Ba/F3 cell line. I was able to show that transplanting Ba/F3 expressing mNPM1-ALK were able to infiltrate the spleen of BALB/c syngeneic hosts. This suggests that *Npm1-Alk* expression is sufficient to transform Ba/F3 cells. However, I should note that the hyperploidy nature of the Ba/F3 cell line does not recapitulate the diploid genome of where chromosomal rearrangements normally occur.

Optimizing an CRISPR-Cas9 retroviral vector

The growing number of novel oncogenic fusions identified from high throughput sequencing data underlies the need for a toolbox to elucidate their function. The emergence of CRISPR gene editing technology made it possible to model chromosomal rearrangements in a quick and cost-effective manner. To accomplish this, I worked on developing a viral vector system to deliver CRISPR-Cas9 into hematopoietic cells. Though non-viral delivery methods such as Cas9 RNPs may suffice for small-scale studies, the low throughput of these methods would be a hindrance in doing large scale screening assays from a DNA library.

After several iterations, I generated a gammaretroviral MSCV vector that optimized expression and titer through a convergent orientation of the CRISPR components. I also adapted oligo dual sgRNA cloning into our pX viral vector. As a test case, I characterized the *Npm1-Alk* gene fusion, using our viral vector in a primary p53 null mouse embryonic fibroblasts. Unfortunately, the viral vector was unable to generate chromosomal rearrangements in Ba/F3 or primary fetal liver hematopoietic stem cells. The difference in the level of expression between nucleofection or viral approaches can be quite substantial and could play a role in how efficient the translocation is formed.

Generation of inter-chromosomal translocations is a rare event

Based on previous research, the formation of a translocation via CRISPR-Cas9 in a cell is theoretically a one-time event. For example, once the indel has formed in a cell, it can no longer form the translocation, as the CRISPR recognition site would be altered. Recent work has suggested that Cas9-induced DSBs are repaired with different kinetics than normal DSBs, which is another variable to consider (Brinkman et al., 2018). This may explain the insertions, large deletions and rearrangements reported with Cas9 (Kosicki et al., 2018; Shou et al., 2018). Taking this into account, chromosomal translocations would likely be the least probable event involving sgRNAs editing multiple sites **(Figure 25)**.

Because there is currently no way to direct DNA repair without perturbation of the cell's normal DNA damage response (DDR), delivery of CRISPR components in a cell does not guarantee formation of the translocation and limits the ability to model inter-chromosomal translocations. However, one report has suggested that DSBs with 5' overhangs skew toward intra and inter-chromosomal translocations through alternative end joining (Ling et al., 2018). As a start to better understand how translocations form, one can use reporter tools to quantify the various repair outcomes for easy readout during experimental manipulation (Jayavaradhan et al., 2018; Ren et al., 2018).

Improving chromosomal engineering

Some of the difficulties encountered thus far are the result of adapting CRISPR-Cas9 to viral vectors that worked well for smaller RNAi constructs. I think the most pressing issue is further optimization of CRISPR technology, similar to the path RNAi technology (Barrangou et al., 2015) went through. The challenges with target specificity (Kleinstiver et al., 2015), efficiency (Doench et al., 2016), and delivery (Wang et al., 2016) will have to be improved before CRISPR technology can be used reliably.

One way to improve CRISPR delivery is to use different CRISPR variants. There is a plethora of new CRISPR proteins, such as Cpf1 and SaCas9, which are smaller than

the Cas9 protein (Ran et al., 2015; Zetsche et al., 2015a) or eSpCas9 with enhanced specificity being discovered or engineered (Slaymaker et al., 2016). Smaller Cas effectors would help viral packaging in my vectors while CRISPR variants with increased on target editing could aid in making DSBs more accessible for rearrangements.

Besides the effector protein, the guide itself can be improved. For example, sgRNA structure and sequence could be optimized, which would increase sgRNA stability, expression and editing efficiency (Dang et al., 2015; Gao et al., 2018b; Hendel et al., 2015; Malina et al., 2014). Though this may not be a major issue for indel formations as Cas9 will continue cutting until the recognition sequence is mutated, for more complex modeling, each factor that can be optimized will have a cumulative effect on the efficiency of chromosomal engineering. I envision that CRISPR will be readily adapted to model translocations and my attempts may have been premature due to the current technical limitations for gene delivery and editing.

Even if CRISPR and guides were to be improved, the selection of optimal targets play a role as well. Recent work suggests that DSBs induced by Cas9 can take hours to resolve (Brinkman et al., 2018) and repair outcomes can be influenced by the spacer sequence (van Overbeek et al., 2016). It has been shown that base composition of the sgRNA affects editing potency, with high G+C content disfavored in highly potent sgRNAs (Wang et al., 2014). Mutations to the PAM or seed sequence (10-12 base pairs in the PAM proximal region) have a large (Wu et al., 2014) effect on target binding and editing. Several groups have optimized sgRNA selection criteria, which has improved targeting efficiency (Dang et al., 2015; Doench et al., 2016; Moreno-Mateos et al., 2015). However, the efficiency of the guides can be affected by multiple factors such as expression level, chromatin accessibility, PAM site density and target sequence, which are not addressed by current sgRNA selection algorithms.

CRISPR mediated induction of translocations

For modeling chromosomal translocations, I think the initial burst of CRISPR activity in the cell determines how genome-editing event will resolve itself. For example, it was suggested that the non-homologous end joining repair pathway is critical for translocation formation during DSB formation (Ghezraoui et al., 2014; Zhang and Jasin, 2011). In addition, several repair pathways and factors such as the CtIP DNA resection factor, are believed to be crucial for inducing chromosomal translocations using CRISPR-Cas9 (Brunet and Jasin, 2018; Cheong et al., 2018). Failure to induce the translocation after the resolution of the DDR will make the cell refractory to further gene editing attempts. In my Ba/F3 system, the tetraploid nature of my cell line may have increased the chance for formation of the *Npm1-Alk* translocation due to the presence of more target sites.

Though the incidence of chromosomal translocations is quite low, are there approaches that can enrich for cells that have the desired rearrangement. Several groups have successfully altered repair outcomes through inhibiting DNA repair proteins, altering cell cycle or timing CRISPR delivery (Chu et al., 2015; Gutschner et al., 2016; Lin et al., 2014). Though these approaches work, it introduces additional variables to your experimental system and no paper so far has shown that they can be used to enrich for rearrangements.

As a way to enrich for modified cells, several groups have developed reporter assays to detect cells that receive a nuclease and undergo genome editing (Ren et al., 2018). However, these approaches are limited to site alterations such as indels. The presence of indels is not indicative of chromosomal rearrangement but in fact that the cell is refractory toward rearrangement due to the loss of the cutting site. To overcome this, the Jasin group published a method to improve rates of modeling chromosomal rearrangements *in vitro* using a selection marker that is expressed only upon successful rearrangement (Vanoli and Jasin, 2017; Vanoli et al., 2017). However, those methods may

be difficult to implement *in vivo,* unless non-toxic delivery methods to primary cells are improved (Zuris et al., 2015).

In vivo rearrangements present challenges

Even with improvements to the CRISPR system, the biology of editing mammalian genomes presents additional obstacles. For example, it has been recently shown that mice have an immune response to Cas9 protein which would reduce the efficacy of genome editing (Chew et al., 2016). The efficiency of Cas9 to edit the genomic locus of a target can vary from cell type making it difficult to discern the rules for high efficiency gene editing (Ren et al., 2018).This may explain why gene fusions have been modeled more frequently in some tissues compared to other types.

The cell and tumor surveillance response present one of the barriers to modeling chromosomal rearrangements. Recent reports show that CRISPR-Cas9 induces the p53 response upon DNA damage which can induce cell death and reduce editing efficiency (Haapaniemi et al., 2018; Ihry et al., 2018). This would partly explain the difficulty in modeling with CRISPR-Cas9 in primary stem cells as DNA damage would be selected against, especially when inducing oncogenic chromosomal rearrangements. The selection for p53 mutant cells during gene editing to overcome this phenomenon could lead to oncogenic transformation, which can confound downstream analysis. However, transient knockdown of p53 may be an approach to increase the efficiency of chromosomal rearrangements.

One large impediment with primary cells has been targeted delivery of cargo. Though efficient delivery protocols for many cell lines are available, some cell types such as those from the immune system have been resistant to standardization. During my attempts to model my chromosomal rearrangements using viral and non-viral approaches, I was continually frustrated by the lack of consistent protocols for delivering CRISPR into certain cell types such as hematopoietic stem cells. It is known that plasmid transfection

causes acute toxicity in target cells (Muerdter et al., 2018; Ren et al., 2018). To overcome this, Cas9 RNPs have been used (Brunetti et al., 2018; Gundry et al., 2016). A recent report reported that viral vectors are not efficient in primary murine hematopoietic cells, but require nucleofection-based methods (Seki and Rutz, 2018). Surprisingly, another group reported that all current Cas9 RNP-based delivery systems (Neon and Amaxa) fail to induce gene editing in human hematopoietic stem cells, due to RNPs accumulating the cell or nuclear membrane (Modarai et al., 2018). This suggests that the field has not coalesced on a standard protocol for delivering CRISPR-Cas9 to primary immune cells, as there are still conflicting reports.

However, reports from other labs showed robust transduction in murine hematopoietic stem cells and their progenitors with minimal virus manipulation (Chen et al., 2014; Katigbak et al., 2016; Li et al., 2016a; Malina et al., 2013). While other groups using lentiviral approaches report low transduction efficiency and the need for virus concentration (Aubrey et al., 2015; Heckl et al., 2014)(Dr. Aubrey - Personal Communication). Interestingly, the Weissman group found that lentiviral vectors can transduce human CD34+ HSPCs with CRISPR-Cas with high efficiency (Zhu et al., 2018). Altogether, I might be reaching the technical limits of inducing chromosomal translocations with CRISPR in primary hematopoietic stem cells. This is not to say that Cas9 delivery is not possible, as there several reports of Cas9 or sgRNA being delivered in hematopoietic cells through nanoparticles and lipid vesicles (Platt et al., 2014; Zuris et al., 2015).

Despite the rapid advances in CRISPR, there has been only one report of an inter-chromosomal translocation in the hematological malignancy. This was a xenograft mouse model of the MLL-ENL where the gene fusion did not drive disease but acts as a cooperating mutation (Reimer et al., 2017) in contrast to an earlier study using a Cre-lox generated MLL-ENL mouse model (Forster et al., 2003). Though human xenografts may model patient alterations, studies show that modeling the murine form of the chromosomal

rearrangement can recapitulate the disease (Maddalo et al., 2015). The lack of studies demonstrating this *in vivo* so far may reflect the difficulty in efficiently generating translocations in murine HSPCs. One potential explanation is that transformation of murine HSPCs with gene fusion expression could be selected against. One study showed that murine HSCs unlike differentiated progentiors are resistant to MLL-ENL mediated transformation (Ugale et al., 2014). This suggests that targeting the differentiated progeny, instead of murine HSPCs, with CRISPR could model transformation by gene fusions.

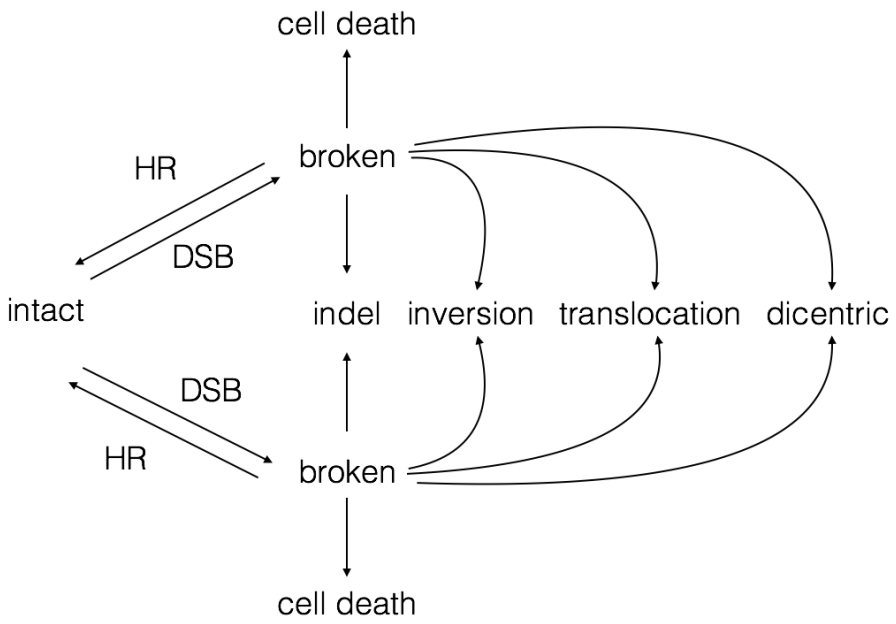


Figure 25. Potential outcomes of double-strand breaks on two genomic loci

As the intact DNA is broken by double stand breaks (DSBs), it can be resolved in several outcomes. The DNA can be faithfully repaired by homologous recombination (HR) though if the DSB was induced by a site-specific nuclease, it will be recut until it no longer contains the recognition site. An unresolved DSB on either locus can trigger cell death if it cannot be repaired. The most likely outcome will be insertions/deletions (indels) where the recognition site is no longer present. However, this means that it can no longer be recut so that locus will be unable to generation any other chromosomal rearrangements. For sites that are on the same chromosome, targeting two sites can result in a genomic inversion and then relegation. This also abolishes the original target sites and hence is irreversible. For sites that are on different chromosomes, there is a partner search where open ends can find each other and then re-ligate resulting in reciprocal exchange or translocation. This is also an irreversible process but much less likely to occur due to the need to move large genomic pieces in the nucleus. Another possible outcome is dicentric chromosomes where there is aberrant relegation of chromosomes resulting in multiple centromeres to be present in one chromosome that causes mis segregation during cell division and possible cell death (Brinkman et al., 2018; Zeisig and So, 2009).

Future Directions

Databases to generate large libraries of sgRNAs for rearrangements

On the technological front, the first CRISPR screens began by using single guides, which is enough for mutations to disrupt gene function. However, modeling fusions presents additional complications, which has slowed adaption of CRISPR to dissect these targets. To successfully model fusions, one needs two sgRNAs to cut simultaneously and induce successful chromosomal rearrangements. My viral vector attempts to provide a solution to this problem.

Even with that innovation, the ability to design sgRNAs at a large scale was still limited. The MIT CRISPR site was initially used to design our guides but there were issues with designing optimal guides due to large gaps in coverage of the mouse genome. In addition, the score presented by the site was not correlated to the presence of off target effects. In fact, some guides cut multiple regions and when combined with multiple guides were able to create off target chromosomal translocations. This could potentially invalidate any study using more than one guide.

To address this, GUIDESCAN was developed to generate a master database of guides that are unique up to 3 mismatches on the target site. The database was also designed to generate multiple pairs of guides with an output that can be readily ordered and cloned (Perez et al., 2017). My viral vectors can be readily adapted to this workflow for CRISPR library screening for potential targets. These guides would be targeted to the expected breakpoints and ideally be trained on sgRNAs used in several cell types.

Testing new oncogenic gene fusions

There is a plethora of uncharacterized candidates to look into as there have been numerous recent studies detailing novel fusions in sequencing data, but no attempt to decipher any functional significance (Andersson et al., 2015; Chen et al., 2018; Nelson et al., 2017). Our platform can be adapted for large scale screening experiments in

hematopoietic cell lines. That means being amenable to large scale cloning of synthesized oligos generated by a bioinformatic platform for designing libraries of sgRNAs (GUIDESCAN) (Perez et al., 2017). Based on work with previous gene fusion screens using cDNA overexpression in Ba/F3 cells, CRISPR is an ideal screening tool for gene fusions (Lu et al., 2017). The screening of these fusions would provide strong evidence for follow up characterization. One of the benefits of this approach is the rapid turnaround because IL3 withdrawal assays can quickly generate positive hits that can be validated and further characterized.

Biological questions of interest

With the plethora of fusions that have not been fully characterized, there is a strong incentive for continuing modeling fusions and discover new biological properties of cancer (Nelson et al., 2017). For example, how drug resistance of NPM1-ALK tumors to Crizotinib arises *de novo* is still not well characterized (Ceccon et al., 2013). One exciting possibility with CRISPR is targeting chromosomal rearrangements with a suicide gene as a potential treatment for cancer (Chen et al., 2017b).

Fusion variants of BCR-ABL1 were identified and reported to have different properties (Melo, 1997). With CRISPR, it is now feasible to study these variants because they can be rapidly generated. Though translocations mostly generate fusion proteins, one report suggests a role for circRNAs in oncogenic transformation (Guarnerio et al., 2016). In fact, a recent report showed that CRISPR generated *Npm1-Alk* translocations in Ba/F3 cells can induce de novo formation of circRNAs (Babin et al., 2018). Unfortunately, the most recent report of a Npm1-Alk mouse model still used a transgenic cDNA overexpression so it could not be used to address this question (Klingeberg et al., 2013; Kreutmair et al., 2017).

As a proof of concept, I showed that our vector could model oncogenic gene fusions. However, our viral vectors can be adapted to screen additional targets that benefit

from tandem sgRNA expression. Examples include non-coding RNAs (Canver et al., 2017; Zhu et al., 2016), enhancers, and other elements that are not amenable to single site disruption.

Final Thoughts

Given the advances in gene editing, I generated a tool that I think will be useful for the broader research community. Though I faced setbacks in achieving an accurate *in vivo* mouse model of an oncogenic inter-chromosomal translocation, I was able to make some progress toward this goal. I generated a set of MSCV based gammaretroviral vectors, active and self-inactivating, with an optimal configuration. This work shows that sgRNA expression can be hindered by promoter interference in a viral vector, an observation which has yet to be discussed in the literature in detail. I hope others will improve upon the vector design, so it has improved titer and gene expression.

I showed that testing a transformed Ba/F3 line, expressing a CRISPR generated Npm1-Alk gene fusion, into an allograft model can cause engraftment and disease presentation. To date, there are no published syngeneic allograft tumor murine models of cells containing the murine inter-chromosomal translocation in the hematopoietic system. Although the ultimate goal in the future is autochthonous induction of the rearrangement and tumor using *ex vivo* modification of murine HSCs.

I envision that my viral vector can be used to screen potential oncogenic fusions based on patient sequencing data in an optimized workflow protocol with our Ba/F3 cell line. Once viral vectors can induce chromosomal rearrangements in primary murine HSPCs with high enough efficiency, I envision rapid generation of *ex vivo* transplantation models of oncogenic gene fusions. This would provide a robust platform for understanding the biology of gene fusions and as a discovery platform to screen for novel drug targets.

APPENDIXES

APPENDIX 1: FISH quantification of Npm1-Alk Ba/F3 line

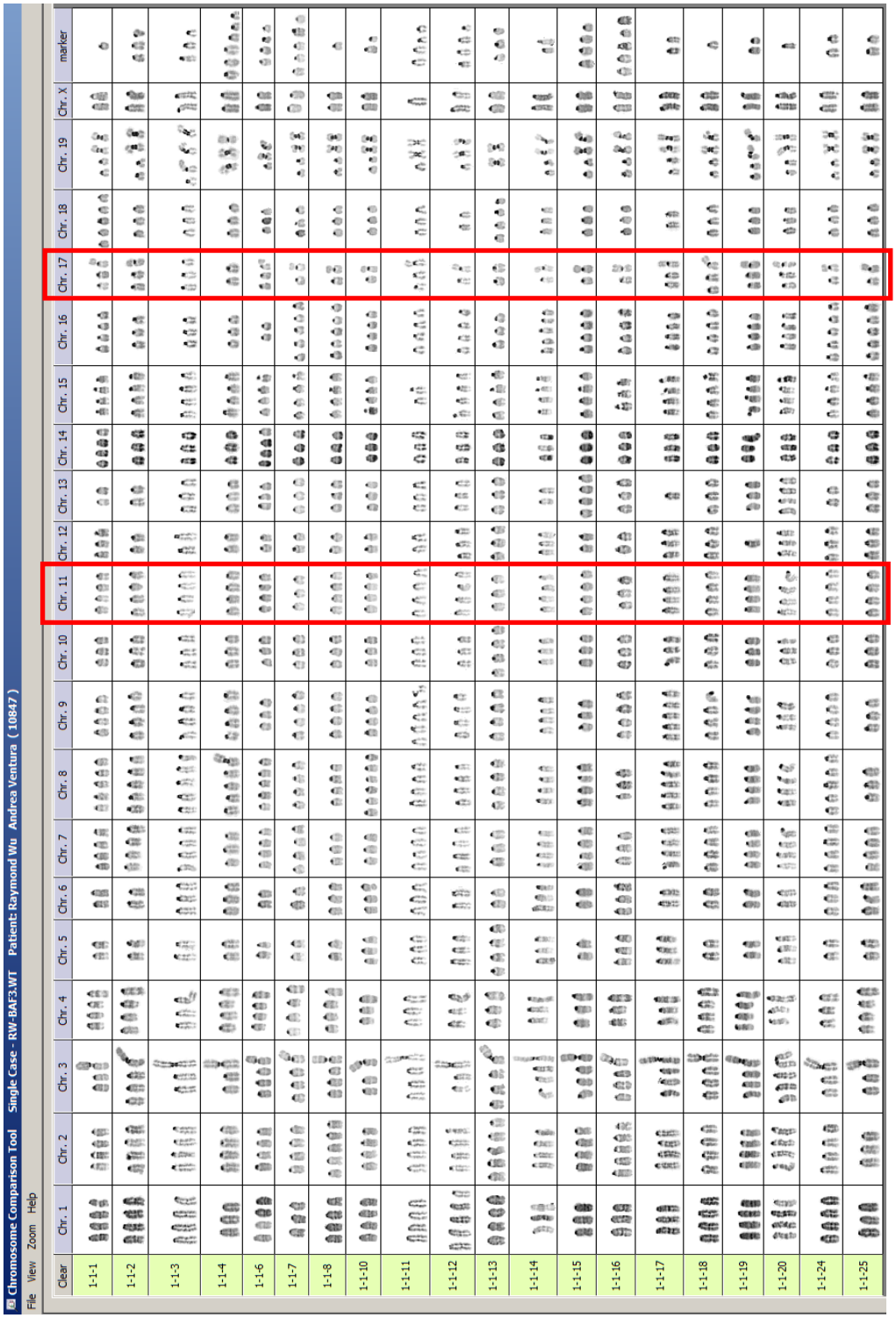
	Heterogenous with variable copy number for Alk/Npm1/Fusion but majority show two copies of fusion (F) signal					
Sample ID	**Ba/F3.nuc.p.IL3-**			**Rosa26**		
Lab ID	**10120**			**Lab-Control**		
Image #	Through scope			Through scope		
S.No	NPM	ALK	Fusion	NPM	ALK	Fusion
	Red	Green	R/G	Red	Green	R/G
1	3	3	**2**	2	2	0
2	3	3	2	2	2	0
3	2	3	2	2	2	0
4	3	3	2	2	2	0
5	3	3	2	2	2	0
6	2	3	1	2	2	0
7	2	3	2	2	2	0
8	2	3	2	2	2	0
9	3	3	2	2	2	0
10	2	3	2	2	2	0
11	2	2	1	2	2	0
12	3	3	2	2	2	0
13	3	3	2	2	2	0
14	3	3	2	2	2	0
15	3	3	1	2	2	0
16	3	2	2	2	2	0
17	2	3	2	2	2	0
18	2	3	2	2	2	0
19	3	3	2	2	2	0
20	1	3	2	2	2	0
21	3	3	2	2	2	0
22	2	1	3	2	2	0
23	2	2	1	2	2	0
24	2	2	2	2	2	0
25	3	2	0	2	2	0
26	3	3	2	2	2	0
27	3	3	2	2	2	0
28	3	3	2	2	2	0
29	3	3	2	2	2	0
30	3	3	2	2	2	0
31	2	3	2	2	2	0
32	2	2	2	2	2	0
33	3	2	2	2	2	0

34	3	3	1	2	2	0
35	2	3	2	2	2	0
36	3	3	2	2	2	0
37	2	2	2	2	2	0
38	3	2	2	2	2	0
39	2	3	1	2	2	0
40	3	3	2	2	2	0
41	3	3	2	2	2	0
42	3	3	1	2	2	0
43	3	3	2	2	2	0
44	3	3	2	2	2	0
45	2	2	2	2	2	0
46	2	3	1	2	2	0
47	2	3	2	2	2	0
48	3	3	2	2	2	0
49	2	4	1	2	2	0
50	2	2	1	2	2	0
51	3	3	2	2	2	0
52	3	2	2	2	2	0
53	2	3	2	2	2	0
54	3	2	2	2	2	0
55	2	2	2	2	2	0
56	3	3	1	2	2	0
57	3	3	2	2	2	0
58	2	3	2	2	2	0
59	2	2	2	2	2	0
60	2	3	1	2	2	0
61	1	3	2	2	2	0
62	3	3	2	2	2	0
63	3	3	2	2	2	0
64	2	2	3	2	2	0
65	3	3	2	2	2	0
66	3	3	2	2	2	0
67	2	3	2	4	4	0
68	3	4	1	2	2	0
69	3	3	0	2	2	0
70	2	3	2	2	2	0
71	2	3	2	2	2	0
72	3	3	2	2	2	0
73	3	3	1	2	2	0
74	3	3	1	2	2	0
75	2	2	2	2	2	0
76	3	3	2	2	2	0
77	4	4	0	2	2	0
78	2	4	0	2	2	0
79	4	2	2	2	2	0
80	1	2	3	2	2	0
81	3	3	2	2	2	0

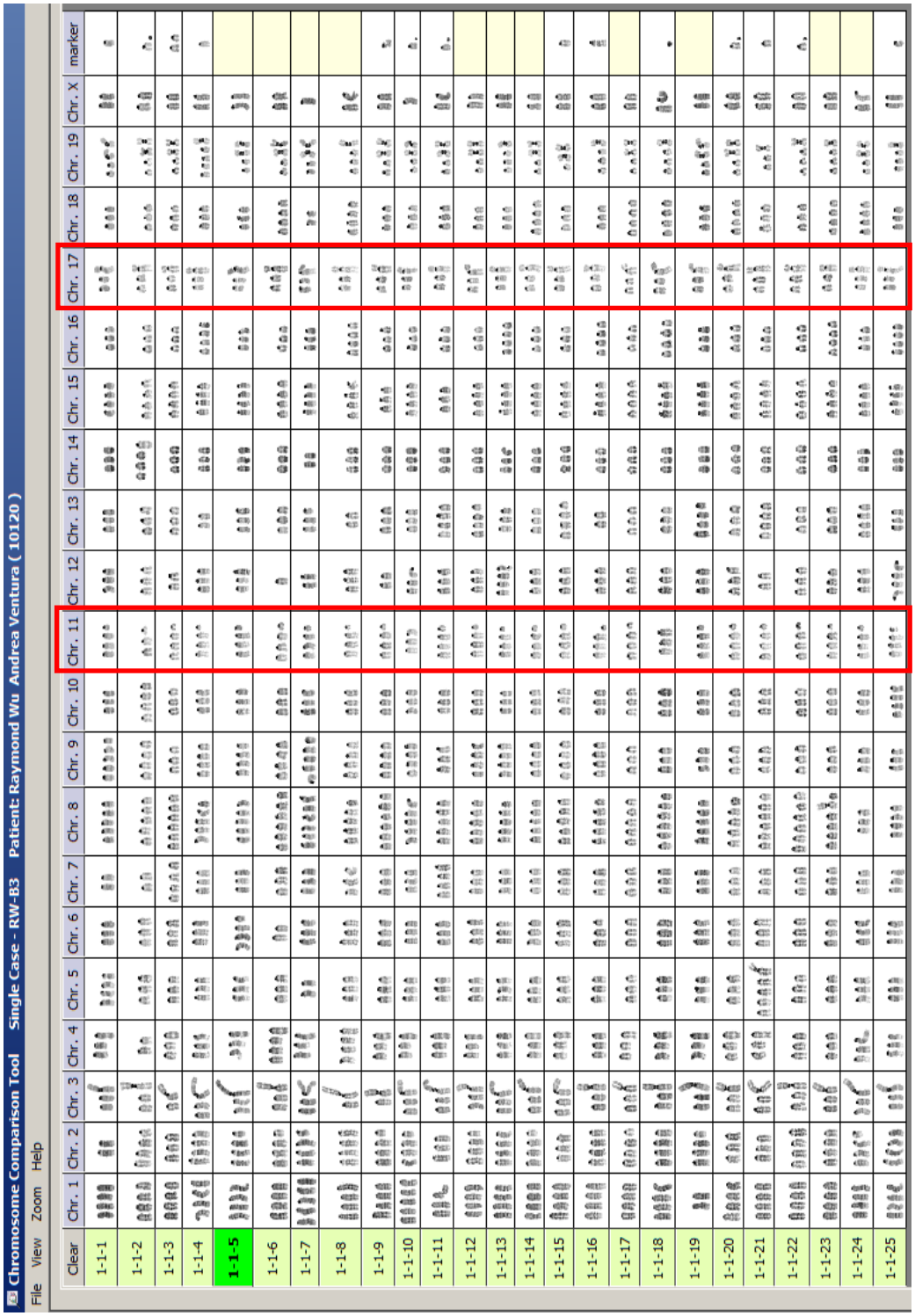
82	3	3	1	2	2	0
83	2	3	2	2	2	0
84	2	3	2	2	2	0
85	2	2	2	2	2	0
86	2	3	2	2	2	0
87	2	3	2	2	2	0
88	3	3	2	2	2	0
89	3	2	2	2	2	0
90	3	3	1	2	2	0
91	2	3	2	2	2	0
92	3	3	1	2	2	0
93	2	3	1	2	2	0
94	3	4	1	2	2	0
95	3	4	1	2	2	0
96	3	2	1	2	2	0
97	3	3	2	2	2	0
98	3	3	2	2	2	0
99	2	2	2	2	2	0
100	4	4	0	2	2	0
Total	257	282	171	202	202	0
Mean	3	3	2	2	2	0
Min	1	1	0	2	2	0
Max	4	4	3	4	4	0
% Positive for fusion			95			0
	% positive for 3R-3G-2F		33			
	% positive with 2F		70			

APPENDIX 1.1: Karyotypes of Ba/F3 Cells

Wild Type Ba/F3



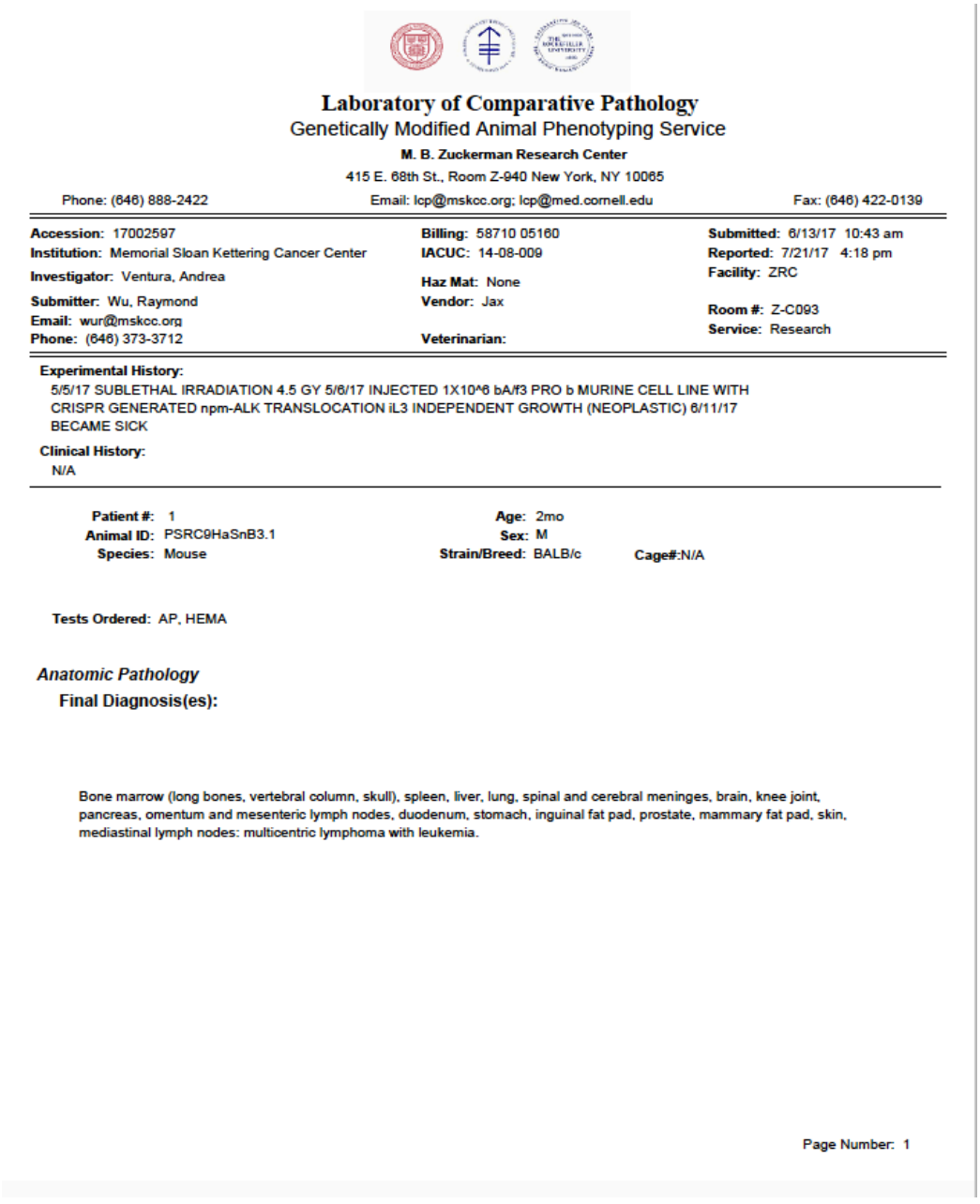
Nucleofected Ba/F3 Npm1-Alk Cas9 Plasmid



APPENDIX 1.2: Pathologist Reports

NPM1-ALK Ba/F3 (Nucleofected)

(pSR.C9hAsNB3 =Nucleofected – mislabeled in report)



Laboratory of Comparative Pathology
Genetically Modified Animal Phenotyping Service
M. B. Zuckerman Research Center
415 E. 68th St., Room Z-940 New York, NY 10065
Phone: (646) 888-2422 Email: lcp@mskcc.org; lcp@med.cornell.edu Fax: (646) 422-0139

Accession: 17002597
Institution: Memorial Sloan Kettering Cancer Center
Investigator: Ventura, Andrea
Submitter: Wu, Raymond
Email: wur@mskcc.org
Phone: (646) 373-3712

Billing: 58710 05160
IACUC: 14-08-009
Haz Mat: None
Vendor: Jax
Veterinarian:

Submitted: 6/13/17 10:43 am
Reported: 7/21/17 4:18 pm
Facility: ZRC
Room #: Z-C093
Service: Research

Experimental History:
5/5/17 SUBLETHAL IRRADIATION 4.5 GY 5/6/17 INJECTED 1X10^6 bA/f3 PRO b MURINE CELL LINE WITH CRISPR GENERATED npm-ALK TRANSLOCATION iL3 INDEPENDENT GROWTH (NEOPLASTIC) 6/11/17 BECAME SICK

Clinical History:
N/A

Patient #: 1
Animal ID: PSRC9HaSnB3.1
Species: Mouse
Age: 2mo
Sex: M
Strain/Breed: BALB/c
Cage#:N/A

Tests Ordered: AP, HEMA

Anatomic Pathology
Final Diagnosis(es):

Bone marrow (long bones, vertebral column, skull), spleen, liver, lung, spinal and cerebral meninges, brain, knee joint, pancreas, omentum and mesenteric lymph nodes, duodenum, stomach, inguinal fat pad, prostate, mammary fat pad, skin, mediastinal lymph nodes: multicentric lymphoma with leukemia.

Page Number: 1

Gross Finding(s):

Weights(g):

Body:	20.911 grams	**Heart:**	0.159 grams	**Liver:**	1.365 grams	**Spleen:**	0.461 gram

A 2 month-old BALB/c male mouse that was sublethally irradiated on 5/5/17 and then transplanted the following day with murine pro-B cell line (Ba/F3) that have the murine NPM1-ALK inter chromosomal translocation, was submitted alive for necropsy to assess the cause of the generalized sickness and the spread of the injected cells.
The body condition score was 2 (underconditioned).
At necropsy the carcass was diffusely pale.
Along the ventral midline of the abdomen there are 2 erythematous and raised round nodular lesions on the skin, measuring 0.3 cm in diameter, that extend into the subcutis. The same lesion is present also in the xiphoideal region.
All mammary lymph nodes are very prominent along the mammary lines on both sides of the cadaver, with the largest one measuring about 0.4 cm in diameter.
The spleen is moderately enlarged and has a marbled appearance.
Liver and kidneys are diffusely pale.
The lumbar tract of the spinal cord and the adjacent layers of soft tissue are diffusely thickened and appear like a plaque lesion, measuring 1 x 0.7 cm, with firm texture and off white discoloration.
In the popliteal region there is a tan and firm mass that measures up to 0.5 cm in diameter.
Slides examined include the following tissues. All tissues are normal unless otherwise described.
1. Heart, lungs, thymus, mediastinal lymph node.
2. Kidneys, liver, gallbladder.
3. Stomach, duodenum, jejunum, ileum, cecum, colon, salivary glands, mandibular lymph node.
4. Uterus, urinary bladder.
5. Adrenal, ovaries, oviducts, pancreas, spleen
6. Trachea, esophagus, skin
7. Hind limb (femur, tibia, bone marrow, stifle joint, skeletal muscle, peripheral nerves), vertebral column with spinal cord, sternum
8,9. Head, coronal sections (including brain, eye, ears, nasal and oral cavities, teeth).
10. Abdominal cutaneous masses
11. Mammary gland, and skin
12. Right popliteal mass and lymph node.
13. Lumbar mass

Microscopic Finding(s):

Mediastinal lymph nodes: the lymph node is expanded and the cortical and paracortical architecture obliterated by compact sheets of tightly packed round to oval cells measuring 14 to 20 µm, with a scant to moderate amount of basophilic cytoplasm, and a round to oval nucleus. Nucleoli can be single and central or multiple (usually 3) and submembranous. Mitoses are between 5 to 7 per high power field in 400X magnification. Anisocytosis and anisokaryosis are mild to moderate. The neoplastic lymphocytes exhibit occasionally plasmacytoid features.
Lungs: Multifocally the lumen of the alveolar capillaries is often dilated by single or short rows of neoplastic lymphocytes.
Liver: diffusely neoplastic lymphocytes infiltrate the heaptic parenchyma around the central vein sections and the portal triads. Teliolymphocytes are present in the biliary epithelium. Bile ducts exhibit hyperplastic features ands a low grade reactive dysplasia. Neoplastic cells are seen also in the lumen of the hepatic sinusoids.
Stomach: multifocal infiltration of neoplastic lymphocytes in the lamina propria of the mucosa and in the submucosa, with minimal distortion of the organ architecture.
Small intestine (duodenum): focal extensive neoplastic infiltration of the lamina propria and of the tunica submucosa.
Small intestine (jejunum and ileum): the Peyer's patches contain one germinal center with tingible bodies.
Pancreas and omentum: focal extensive infiltration of sheets of neoplastic cells. A large part of the pancreas is replaced by

neoplastic lymphocytes with partial sparing of the pancreatic duct.
Testes: increased number of Leydig's cells (interstitial cell hyperplasia).
Prostate: partial infiltration of neoplastic cells with focal compression of the gland.
Inguinal fat pad: an oval mass that measures about 0.3 x 0.2 cm is composed of sheets of neoplastic cells with few scattered entrapped adipocytes.
Spleen: most of the white and red pulp are replaced by the infiltrating neoplastic lymphocyte population with diffuse subgross enlargement of the organ.
Cutaneous mass in xiphoideal region: a well demarcated but infiltratively growing fusiform mass composed of sheets of neoplatic lymphocytes extends from the mid-dermis down into the abdominal muscle planes.
Cutaneous mass: the subcutis is multifocally expanded by 2 fusiform, well demarcated masses which are delimited above by the dermis and below by the fascia. The masses are composed of neoplastic lymphocytes.
Bone marrow (femur and tibia): diffusely infiltrated by neoplastic lymphocytes with significant replacement of the normal myeloid and erythroid populations.
Femur: the periosteum along the diaphisis is diffusely infiltrated and thickneed by neoplastic lymphocytes with multifocal penetration into the adjacent skeletal muscle bundles.
Tibia: diffuse periosteal infiltration by neoplastic cells.
Knee joint: diffuse infiltration by sheets of neoplastic lymphocytes.
Skull: the marrow cavity of all flat bones is diffusely infiltrated by neoplastic lymphocytes with multifocal invasion also of the adjacent and contiguous soft tissues.
Mid- and inner ear: diffuse obliteration through neoplastic lymphocytes.
Brain: diffuse and severe infiltration of the meninges by neoplastic lymphocytes with compression and atrophy of the cerebral emispheres. The malignant lymphocytes also invade the sagittal sinus and almost completely efface the pituitary gland.
Spinal cord and lumbar mass: the mass is composed of neoplastic lymphocytes that heavily infiltrate the skeletal muscles, the bone marrow in the vertebral bodies with multifocal necrosis, the spinal meninges and the cauda equina.
Mammary gland: diffuse effacement of the fat pad through neoplastic lymphocytes.
Popliteal mass: focal well demarcated and partially encapsulated mass consisting of tightly packed round cells arranged in sheets (neoplastic lymphocytes). In the local lymph nodes, within the paracortical and medullary sinuses there are few single large lymphocytes consistent with neoplastic cells.

Non Tissue Collection

Non Tissue Collection	EC

Hematology

CBC with automatic differential

Comments and Interpretation:

The cause of the observed generalized sickness is the multiorgan infiltration of neoplastic lymphocytes with extensive destruction and possible impaired (reduced) or lost organ function. The presence of neoplastic cells in the bone marrow of different bone compartments, as well as in the hepatic sinusoids and in the lumen of the alveolar capillary confirms a concurrent and ongoing leukemic phase. Morphologically the majority of the neoplastic lymphocytes are centroblasts, although immunoblast are also well represented. Plasmacytoid features of the malignant lymphocytes were evident especially in the bone marrow smears. The neoplastic lymphocytes were negative for immunostaining for B220 antigen and positive for ALK-1. In the examined organs and tissues there is no morphologic evidence of any infectious process.

Tissue: Stored by LCP **Blocks:** Stored by LCP **Slides:** Returned to investigator

Pathologist: Alessandra Piersigilli, DVM, PhD, Dipl. ECVP

MSCV human NPM1-ALK cDNA overexpression

(MSCV NPM1-ALk IRES EGFP cDNA - not translocation - mislabeled in report)

Laboratory of Comparative Pathology
Genetically Modified Animal Phenotyping Service
M. B. Zuckerman Research Center
415 E. 68th St., Room Z-940 New York, NY 10065
Phone: (646) 888-2422 Email: lcp@mskcc.org; lcp@med.cornell.edu Fax: (646) 422-0139

Accession: 17002720
Institution: Memorial Sloan Kettering Cancer Center
Investigator: Ventura, Andrea
Submitter: Wu, Raymond
Email: wur@mskcc.org
Phone: (646) 373-3712

Billing: 58710 05160
IACUC: 1408009
Haz Mat: None
Vendor: Jax
Veterinarian:

Submitted: 6/21/17 2:09 pm
Reported: 7/13/17 5:33 pm
Facility: ZRC
Room #: C0-93
Service: Research

Experimental History:
irradiated mice with 4.5 Gyinjected BaF3 with NPM ALK translocation

Clinical History:
hunched brusied tail

Patient #: 1
Animal ID: NAIE.B3.050617
Species: Mouse

Age: 1mo
Sex: M
Strain/Breed: BALB/c
Cage#:2784735

Tests Ordered: AP, HEMA

Anatomic Pathology

Final Diagnosis(es):

Spleen, liver, small intestine, bone marrow, femur, tibia, bulla tympanica, spine: multicentric lymphoma with leukemia.
Testes: oligozoospermia
Eyes: bilateral mild cataract.
Bone marrow: left shift.
Tail: multifocal coagulative necrosis and ulceration with intralesional bacterial colonies and subcutaneous or full thickness dry gangrene, amputation, severe, subacute.

Gross Finding(s):

Weights(g):

Body:	19.840 grams	**Liver:**	1.39 grams	**Spleen:**	0.39 grams

A 1 month old male BALB/c mouse previously irradiated (4.5 Gy) and injected with BaF with NPM ALK translocation wassubmitted alive for euthanasia and necropsy as it presented a hunched and bruised tail. The body condition score was 1 (emaciated).
Spleen: discoidal polar enlargement with pale red-grey discoloration and focal splenomegaly.
Liver: mild diffuse enlargement and randomly scattered pin point white foci about 1 mm in diameter.
Vertebral columns and paravertebral tissues: within the ventral aspect of lumbo-sacral tract there is a focal extensive enlargement measuring 2 x 1 x 0.5 cm, rather firm and pale grey.
Tail: circumferential and segmental area of necrosis at the base. Necrosis and amputation of the tip of the tail. Multifocal ulcerations.
Slides examined include the following tissues. All tissues are normal unless otherwise described.
1. Heart, lungs, thymus, mediastinal lymph node.
2. Kidneys, liver, gallbladder.
3. Stomach, duodenum, jejunum, ileum, cecum, colon, salivary glands, mandibular lymph node.
4. Testes, urinary bladder, prostate, seminal vesicles, epididymidis.
5. Adrenal, pancreas, spleen
6. Trachea, esophagus, skin
7. Hind limb (femur, tibia, bone marrow, stifle joint, skeletal muscle, peripheral nerves), vertebral column with spinal cord, sternum
8,9. Head, coronal sections (including brain, eye, ears, nasal and oral cavities, teeth).
10. Tail
11. Vertebral column (lumbar tract)

Microscopic Finding(s):

Spleen: subgrossly diffusely and moderately enlarged. Most of the white and red pulps are replaced by sheets of large round cells. the neoplastic cells are basophilic, with scant to moderate amount of cytoplasm, an oval nucleus with one or 2 prominent nucleoli, which is located in the center or at the periphery of the cell. Their size is quite variable, however it ranges from 10 to 18 µm. Anisocytosis and anisokaryosis are moderate to high. Mitoses are 1-3 per high power field in 400X magnification. Small islets of erythroid precursors are entrapped in between adjacent sheets of neoplastic cells.
Liver: multifocal infiltration of the parenchyma with neoplastic solid emboli within the lumen of blood vessels and single cells in the sinusoids (peripheral leukemia). Areas of hepatic parenchyma are necrotic and sharply demarcated when infiltration of neoplastic cells is present.
Small intestine: focal lymphomatous infiltration of the Peyer's patches.
Testes: rare mature sperms in the lumen of the seminiferous tubules.
Adrenal glands: bilateral subcapsular cells hyperplasia.
Right rear limb: the bone marrow in the diaphysis of the femur and tibia is diffusely replaced by the continuous infiltration of neoplastic cells. In the thigh the tumor cells invade the cortical bone and the adjacent soft tissue with subgross expansion of the cranial surface of the Rectus femoris muscle.
Tympanic bulla: focal effacement of the bone amrrow and cortical bone by neoplastic cells.
Eyes: few Morgagni globules along the posterior aspect of the lens.
Lumbar mass: the mass consists of neoplastic round cells arranged in sheets and lobules which diffusely efface the vertebral bodies, infiltrate the meninges, nerve roots and ganglia and expand the local muscle bundles.
Lumbal bone marrow: in the vertebral segments spared by the neoplastia the bone marrow has an increased percentage of leukocytes (myeloid and granulocytic precursors).
Tail: the distal part of the tail is diffusely necrotic, with sharp demarcation at the border with the intact tissue (gangrene). Within the remaining part there are multifocal areas of epidermal, dermal and subcutaneous coagulative necrosis with severe infiltration of neutrophils and multifocal bacterial colonies (cocci)

Non Tissue Collection

Non Tissue Collection JC

Hematology

CBC with automatic differential

Finalized by: BACANIC Date: 6/29/2017 4:34:26PM

Comments and Interpretation:

The most remarkable finding in this animal is the lymphomatous infiltration of multiple organs with limited leukemia. The infiltration of the bone marrow in different bone segments explains the leukocytopenia. Anemia however is not so pronounced likely due to the different lifespan of erythrocytes compared to leukocytes. The bilateral mild cataract is an effect of the irradiation as otherwise it is a typical age-related finding. The oligozoospermia is likely due to a delay of the sperm maturation, again as possible effect of the total body irradiation, as usually mice reach the sexual maturity around 32 days, with slight variation depending on the strain. The presence of immature forms of granulocytes is consistent with the extensive neutrophilic infiltration and degeneration in the tail lesions. Concerning the pathogenesis of the tail ulcerations, necrosis with amputation, several mechanisms and causes can be considered. Radiation, especially when repeated, leads to tail wounds, depending also on time and intensity of the exposure. In the sections several bacterial colonies are seen within the crusts and in the area of ulceration. Staphyloccocus spp. especially xylosus is known to cause severe necrotizing dermatitis even dry gangrene in mice. Infection can occur due to escoriation of the tail or iatrogenic manipulation (eg. puncture, injection).

Tissue: Stored by LCP **Blocks:** Stored by LCP **Slides:** Stored by LCP

Pathologist: Alessandra Piersigilli, DVM, PhD, Dipl. ECVP

BIBLIOGRAPHY

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L.*, et al.* (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D., and Brinster, R.L. (1985). The C-Myc Oncogene Driven by Immunoglobulin Enhancers Induces Lymphoid Malignancy in Transgenic Mice. Nature *318*, 533-538.

Aird, E.J., Lovendahl, K.N., St. Martin, A., Harris, R.S., and Gordon, W.R. (2017). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. bioRxiv.

Alcalay, M., Zangrilli, D., Pandolfi, P.P., Longo, L., Mencarelli, A., Giacomucci, A., Rocchi, M., Biondi, A., Rambaldi, A., Lococo, F.*, et al.* (1991). Translocation Breakpoint of Acute Promyelocytic Leukemia Lies within the Retinoic Acid Receptor-Alpha Locus. Proceedings of the National Academy of Sciences of the United States of America *88*, 1977-1981.

Ambrogio, C., Martinengo, C., Voena, C., Tondat, F., Riera, L., di Celle, P.F., Inghirami, G., and Chiarle, R. (2009). NPM-ALK oncogenic tyrosine kinase controls T-cell identity by transcriptional regulation and epigenetic silencing in lymphoma cells. Cancer Res *69*, 8611-8619.

Ando, D., and Meyer, K. (2017). Gene Editing: Regulatory and Translation to Clinic. Hematol Oncol Clin North Am *31*, 797-808.

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

Artandi, S.E., Chang, S., Lee, S.L., Alson, S., Gottlieb, G.J., Chin, L., and DePinho, R.A. (2000). Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. Nature *406*, 641-645.

Aten, J.A., Stap, J., Krawczyk, P.M., van Oven, C.H., Hoebe, R.A., Essers, J., and Kanaar, R. (2004). Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. Science *303*, 92-95.

Aubrey, B.J., Kelly, G.L., Kueh, A.J., Brennan, M.S., O'Connor, L., Milla, L., Wilcox, S., Tai, L., Strasser, A., and Herold, M.J. (2015). An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo. Cell Rep *10*, 1422-1432.

Babin, L., Piganeau, M., Renouf, B., Lamribet, K., Thirant, C., Deriano, L., Mercher, T., Giovannangeli, C., and Brunet, E.C. (2018). Chromosomal Translocation Formation Is Sufficient to Produce Fusion Circular RNAs Specific to Patient Tumor Cells. iScience *5*, 19-29.

Barrangou, R., Birmingham, A., Wiemann, S., Beijersbergen, R.L., Hornung, V., and Smith, A. (2015). Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference. Nucleic Acids Res *43*, 3407-3419.

Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. Nat Biotechnol *34*, 933-941.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Baum, C., von Kalle, C., Staal, F.J., Li, Z., Fehse, B., Schmidt, M., Weerkamp, F., Karlsson, S., Wagemaker, G., and Williams, D.A. (2004). Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. Mol Ther *9*, 5-13.

Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., Tan, W., Penheiter, S.G., Ma, A.C., Leung, A.Y.*, et al.* (2012). In vivo genome editing using a

high-efficiency TALEN system. Nature.
Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F., 3rd (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci U S A *95*, 14628-14633.
Bernardin, F., Yang, Y., Cleaves, R., Zahurak, M., Cheng, L., Civin, C.I., and Friedman, A.D. (2002). TEL-AML1, expressed from t(12;21) in human acute lymphocytic leukemia, induces acute leukemia in mice. Cancer Res *62*, 3904-3908.
Bignold, L.P., Coghlan, B.L., and Jersmann, H.P. (2006). Hansemann, Boveri, chromosomes and the gametogenesis-related theories of tumours. Cell Biol Int *30*, 640-644.
Bilsland, J.G., Wheeldon, A., Mead, A., Znamenskiy, P., Almond, S., Waters, K.A., Thakur, M., Beaumont, V., Bonnert, T.P., Heavens, R.*, et al.* (2008). Behavioral and neurochemical alterations in mice deficient in anaplastic lymphoma kinase suggest therapeutic potential for psychiatric indications. Neuropsychopharmacology *33*, 685-700.
Blasco, R.B., Karaca, E., Ambrogio, C., Cheong, T.C., Karayol, E., Minero, V.G., Voena, C., and Chiarle, R. (2014). Simple and rapid in vivo generation of chromosomal rearrangements using CRISPR/Cas9 technology. Cell Rep *9*, 1219-1227.
Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551-2561.
Bonzheim, I., Steinhilber, J., Fend, F., Lamant, L., and Quintanilla-Martinez, L. (2015). ALK-positive anaplastic large cell lymphoma: an evolving story. Front Biosci (Schol Ed) *7*, 248-259.
Borrow, J., Goddard, A.D., Sheer, D., and Solomon, E. (1990). Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. Science *249*, 1577-1580.
Bouard, D., Alazard-Dany, D., and Cosset, F.L. (2009). Viral vectors: from virology to transgene expression. Br J Pharmacol *157*, 153-165.
Boveri, T. (2008). Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. J Cell Sci *121 Suppl 1*, 1-84.
Breese, E.H., Buechele, C., Dawson, C., Cleary, M.L., and Porteus, M.H. (2015). Use of Genome Engineering to Create Patient Specific MLL Translocations in Primary Human Hematopoietic Stem and Progenitor Cells. PLoS One *10*, e0136644.
Brinkman, E.K., Chen, T., de Haas, M., Holland, H.A., Akhtar, W., and van Steensel, B. (2018). Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. Mol Cell.
Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.
Brummelkamp, T.R., Bernards, R., and Agami, R. (2002a). Stable suppression of tumorigenicity by virus-mediated RNA interference. Cancer Cell *2*, 243-247.
Brummelkamp, T.R., Bernards, R., and Agami, R. (2002b). A system for stable expression of short interfering RNAs in mammalian cells. Science *296*, 550-553.
Brunet, E., and Jasin, M. (2018). Induction of Chromosomal Translocations with CRISPR-Cas9 and Other Nucleases: Understanding the Repair Mechanisms That Give Rise to Translocations. Adv Exp Med Biol *1044*, 15-25.
Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V.M., Gregory, P., Urnov, F., Weinstock, D.M., and Jasin, M. (2009). Chromosomal translocations induced at specified loci in human stem cells. Proc Natl Acad Sci U S A *106*, 10620-10625.
Brunetti, L., Gundry, M.C., Kitano, A., Nakada, D., and Goodell, M.A. (2018). Highly

Efficient Gene Disruption of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. J Vis Exp.
Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D.*, et al.* (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science *273*, 1058-1073.
Burman, B., Zhang, Z.Z., Pegoraro, G., Lieb, J.D., and Misteli, T. (2015). Histone modifications predispose genome regions to breakage and translocation. Genes Dev *29*, 1393-1402.
Burns, J.C., Friedmann, T., Driever, W., Burrascano, M., and Yee, J.K. (1993). Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. Proc Natl Acad Sci U S A *90*, 8033-8037.
Cancer Genome Atlas Research, N., Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet *45*, 1113-1120.
Cano, F., Drynan, L.F., Pannell, R., and Rabbitts, T.H. (2008). Leukaemia lineage specification caused by cell-specific Mll-Enl translocations. Oncogene *27*, 1945-1950.
Capecchi, M.R. (1989). Altering the genome by homologous recombination. Science *244*, 1288-1292.
Ceccon, M., Merlo, M.E.B., Mologni, L., Poggio, T., Varesio, L.M., Menotti, M., Bombelli, S., Rigolio, R., Manazza, A.D., Di Giacomo, F.*, et al.* (2016). Excess of NPM-ALK oncogenic signaling promotes cellular apoptosis and drug dependency. Oncogene *35*, 3854-3865.
Ceccon, M., Mologni, L., Bisson, W., Scapozza, L., and Gambacorti-Passerini, C. (2013). Crizotinib-resistant NPM-ALK mutants confer differential sensitivity to unrelated Alk inhibitors. Mol Cancer Res *11*, 122-132.
Cermak, T., Starker, C.G., and Voytas, D.F. (2015). Efficient design and assembly of custom TALENs using the Golden Gate platform. Methods Mol Biol *1239*, 133-159.
Chambers, J., and Rabbitts, T.H. (2015). LMO2 at 25 years: a paradigm of chromosomal translocation proteins. Open Biol *5*, 150062.
Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S.*, et al.* (2013). Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. Cell *155*, 1479-1491.
Chen, C., Liu, Y., Rappaport, A.R., Kitzing, T., Schultz, N., Zhao, Z., Shroff, A.S., Dickins, R.A., Vakoc, C.R., Bradner, J.E.*, et al.* (2014). MLL3 Is a Haploinsufficient 7q Tumor Suppressor in Acute Myeloid Leukemia. Cancer Cell *25*, 652-665.
Chen, F., Ding, X., Feng, Y., Seebeck, T., Jiang, Y., and Davis, G.D. (2017a). Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. Nat Commun *8*, 14958.
Chen, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R.*, et al.* (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell *160*, 1246-1260.
Chen, Z.H., Yu, Y.P., Zuo, Z.H., Nelson, J.B., Michalopoulos, G.K., Monga, S., Liu, S., Tseng, G., and Luo, J.H. (2017b). Targeting genomic rearrangements in tumor cells through Cas9-mediated insertion of a suicide gene. Nat Biotechnol *35*, 543-550.
Cheong, T.C., Blasco, R.B., and Chiarle, R. (2018). The CRISPR/Cas9 System as a Tool to Engineer Chromosomal Translocation In Vivo. Adv Exp Med Biol *1044*, 39-48.
Cherry, S.R., Biniszkiewicz, D., van Parijs, L., Baltimore, D., and Jaenisch, R. (2000). Retroviral expression in embryonic stem cells and hematopoietic stem cells. Mol Cell Biol *20*, 7419-7426.
Chew, W.L., Tabebordbar, M., Cheng, J.K., Mali, P., Wu, E.Y., Ng, A.H., Zhu, K., Wagers,

A.J., and Church, G.M. (2016). A multifunctional AAV-CRISPR-Cas9 and its host response. Nat Methods *13*, 868-874.
Chiarle, R., Gong, J.Z., Guasparri, I., Pesci, A., Cai, J., Liu, J., Simmons, W.J., Dhall, G., Howes, J., Piva, R.*, et al.* (2003). NPM-ALK transgenic mice spontaneously develop T-cell lymphomas and plasma cell tumors. Blood *101*, 1919-1927.
Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol.
Choi, P.S., and Meyerson, M. (2014). Targeted genomic rearrangements using CRISPR/Cas technology. Nat Commun *5*, 3728.
Chu, V.T., Graf, R., Wirtz, T., Weber, T., Favret, J., Li, X., Petsch, K., Tran, N.T., Sieweke, M.H., Berek, C.*, et al.* (2016). Efficient CRISPR-mediated mutagenesis in primary immune cells using CrispRGold and a C57BL/6 Cas9 transgenic mouse line. Proc Natl Acad Sci U S A *113*, 12514-12519.
Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kuhn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543-548.
Collins, E.C., Pannell, R., Simpson, E.M., Forster, A., and Rabbitts, T.H. (2000). Inter-chromosomal recombination of Mll and Af9 genes mediated by cre-loxP in mouse development. EMBO Rep *1*, 127-132.
Colombo, E., Marine, J.C., Danovi, D., Falini, B., and Pelicci, P.G. (2002). Nucleophosmin regulates the stability and transcriptional activity of p53. Nat Cell Biol *4*, 529-533.
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.
Cook, P.J., Thomas, R., Kannan, R., de Leon, E.S., Drilon, A., Rosenblum, M.K., Scaltriti, M., Benezra, R., and Ventura, A. (2017). Somatic chromosomal engineering identifies BCAN-NTRK1 as a potent glioma driver and therapeutic target. Nat Commun *8*, 15987.
Cools, J., DeAngelo, D.J., Gotlib, J., Stover, E.H., Legare, R.D., Cortes, J., Kutok, J., Clark, J., Galinsky, I., Griffin, J.D.*, et al.* (2003). A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. N Engl J Med *348*, 1201-1214.
Corral, J., Lavenir, I., Impey, H., Warren, A.J., Forster, A., Larson, T.A., Bell, S., McKenzie, A.N., King, G., and Rabbitts, T.H. (1996). An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. Cell *85*, 853-861.
Costa, R., Carneiro, B.A., Taxter, T., Tavora, F.A., Kalyan, A., Pai, S.A., Chae, Y.K., and Giles, F.J. (2016). FGFR3-TACC3 fusion in solid tumors: mini review. Oncotarget *7*, 55924-55938.
Cox, A.D., and Der, C.J. (1994). Biological assays for cellular transformation. Methods Enzymol *238*, 277-294.
Crescenzo, R., and Inghirami, G. (2015). Anaplastic lymphoma kinase inhibitors. Curr Opin Pharmacol *23*, 39-44.
Crow, E.W., and Crow, J.F. (2002). 100 years ago: Walter Sutton and the chromosome theory of heredity. Genetics *160*, 1-4.
Cuculis, L., and Schroeder, C.M. (2017). A Single-Molecule View of Genome Editing Proteins: Biophysical Mechanisms for TALEs and CRISPR/Cas9. Annu Rev Chem Biomol Eng *8*, 577-597.
Cuesta-Dominguez, A., Ortega, M., Ormazabal, C., Santos-Roncero, M., Galan-Diez, M., Steegmann, J.L., Figuera, A., Arranz, E., Vizmanos, J.L., Bueren, J.A.*, et al.* (2012). Transforming and tumorigenic activity of JAK2 by fusion to BCR: molecular mechanisms of action of a novel BCR-JAK2 tyrosine-kinase. PLoS One *7*, e32451.

Cullen, B.R., Lomedico, P.T., and Ju, G. (1984). Transcriptional interference in avian retroviruses--implications for the promoter insertion model of leukaemogenesis. Nature *307*, 241-245.
Curtin, J.A., Dane, A.P., Swanson, A., Alexander, I.E., and Ginn, S.L. (2008). Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct. Gene Ther *15*, 384-390.
Daley, G.Q., Van Etten, R.A., and Baltimore, D. (1990). Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science *247*, 824-830.
Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P., and Wu, H. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. Genome Biol *16*, 280.
Davies, K.D., and Doebele, R.C. (2013). Molecular pathways: ROS1 fusion proteins in cancer. Clin Cancer Res *19*, 4040-4045.
de The, H., Chomienne, C., Lanotte, M., Degos, L., and Dejean, A. (1990). The t(15;17) translocation of acute promyelocytic leukaemia fuses the retinoic acid receptor alpha gene to a novel transcribed locus. Nature *347*, 558-561.
de The, H., Pandolfi, P.P., and Chen, Z. (2017). Acute Promyelocytic Leukemia: A Paradigm for Oncoprotein-Targeted Cure. Cancer Cell *32*, 552-560.
Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.
Delviks-Frankenberry, K., Galli, A., Nikolaitchik, O., Mens, H., Pathak, V.K., and Hu, W.S. (2011). Mechanisms and factors that influence high frequency retroviral recombination. Viruses *3*, 1650-1680.
Deng, W., Shi, X., Tjian, R., Lionnet, T., and Singer, R.H. (2015). CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. Proc Natl Acad Sci U S A *112*, 11870-11875.
Devroe, E., and Silver, P.A. (2002). Retrovirus-delivered siRNA. BMC Biotechnol *2*, 15.
Ding, Q., Lee, Y., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T.*, et al.* (2012). A TALEN Genome-Editing System for Generating Human Stem Cell-Based Disease Models. Cell stem cell.
Ding, Q., Regan, S.N., Xia, Y., Oostrom, L.A., Cowan, C.A., and Musunuru, K. (2013). Enhanced Efficiency of Human Pluripotent Stem Cell Genome Editing through Replacing TALENs with CRISPRs. Cell stem cell *12*, 393-394.
Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R.*, et al.* (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol *34*, 184-191.
Dow, L.E., Fisher, J., O'Rourke, K.P., Muley, A., Kastenhuber, E.R., Livshits, G., Tschaharganeh, D.F., Socci, N.D., and Lowe, S.W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nat Biotechnol *33*, 390-394.
Drexler, H.G., Gignac, S.M., von Wasielewski, R., Werner, M., and Dirks, W.G. (2000). Pathobiology of NPM-ALK and variant fusion genes in anaplastic large cell lymphoma and other lymphomas. Leukemia *14*, 1533-1559.
Druker, B.J., Talpaz, M., Resta, D.J., Peng, B., Buchdunger, E., Ford, J.M., Lydon, N.B., Kantarjian, H., Capdeville, R., Ohno-Jones, S.*, et al.* (2001). Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med *344*, 1031-1037.
East-Seletsky, A., O'Connell, M.R., Burstein, D., Knott, G.J., and Doudna, J.A. (2017). RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. Mol Cell *66*, 373-383 e373.

Elliott, B., Richardson, C., and Jasin, M. (2005). Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. Mol Cell *17*, 885-894.
Ellis, J. (2005). Silencing and variegation of gammaretrovirus and lentivirus vectors. Hum Gene Ther *16*, 1241-1246.
Emerman, M., and Temin, H.M. (1984). Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism. Cell *39*, 449-467.
Erikson, J., Finan, J., Nowell, P.C., and Croce, C.M. (1982). Translocation of immunoglobulin VH genes in Burkitt lymphoma. Proceedings of the National Academy of Sciences of the United States of America *79*, 5611-5615.
Faber, J., Krivtsov, A.V., Stubbs, M.C., Wright, R., Davis, T.N., Heuvel-Eibrink, M., Zwaan, C.M., Kung, A.L., and Armstrong, S.A. (2009). HOXA9 is required for survival in human MLL-rearranged acute leukemias. Blood *113*, 2375-2385.
Faulkner, C., Ellis, H.P., Shaw, A., Penman, C., Palmer, A., Wragg, C., Greenslade, M., Haynes, H.R., Williams, H., Lowis, S.*, et al.* (2015). BRAF Fusion Analysis in Pilocytic Astrocytomas: KIAA1549-BRAF 15-9 Fusions Are More Frequent in the Midline Than Within the Cerebellum. J Neuropathol Exp Neurol *74*, 867-872.
Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D.Y., Barbosa, I.A., Kwon, J.S., Guan, Y.*, et al.* (2013). An optimized microRNA backbone for effective single-copy RNAi. Cell Rep *5*, 1704-1713.
Findlay, S.D., Vincent, K.M., Berman, J.R., and Postovit, L.M. (2016). A Digital PCR-Based Method for Efficient and Highly Specific Screening of Genome Edited Cells. PLoS One *11*, e0153901.
Foley, S.B., Hildenbrand, Z.L., Soyombo, A.A., Magee, J.A., Wu, Y., Oravecz-Wilson, K.I., and Ross, T.S. (2013). Expression of BCR/ABL p210 from a knockin allele enhances bone marrow engraftment without inducing neoplasia. Cell Rep *5*, 51-60.
Forster, A., Pannell, R., Drynan, L., Cano, F., Chan, N., Codrington, R., Daser, A., Lobato, N., Metzler, M., Nam, C.H.*, et al.* (2005). Chromosomal translocation engineering to recapitulate primary events of human cancer. Cold Spring Harb Symp Quant Biol *70*, 275-282.
Forster, A., Pannell, R., Drynan, L.F., McCormack, M., Collins, E.C., Daser, A., and Rabbitts, T.H. (2003). Engineering de novo reciprocal chromosomal translocations associated with Mll to replicate primary events of human cancer. Cancer Cell *3*, 449-458.
Frock, R.L., Hu, J., and Alt, F.W. (2015a). Mechanisms of Recurrent Chromosomal Translocations. 27-51.
Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015b). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol *33*, 179-186.
Fu, Y., Rocha, P.P., Luo, V.M., Raviram, R., Deng, Y., Mazzoni, E.O., and Skok, J.A. (2016). CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. Nat Commun *7*, 11707.
Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol *32*, 279-284.
Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol *31*, 397-405.
Gambacorti-Passerini, C., Messa, C., and Pogliani, E.M. (2011). Crizotinib in anaplastic large-cell lymphoma. N Engl J Med *364*, 775-776.
Gao, X., Tao, Y., Lamas, V., Huang, M., Yeh, W.H., Pan, B., Hu, Y.J., Hu, J.H., Thompson, D.B., Shu, Y.*, et al.* (2018a). Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. Nature *553*, 217-221.
Gao, Z., Herrera-Carrillo, E., and Berkhout, B. (2018b). Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. Mol Ther Nucleic Acids *10*,

36-44.
Gao, Z., Herrera-Carrillo, E., and Berkhout, B. (2018c). RNA Polymerase II Activity of Type 3 Pol III Promoters. Mol Ther Nucleic Acids *12*, 135-145.
Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.
Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A *109*, E2579-2586.
Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A.T to G.C in genomic DNA without DNA cleavage. Nature *551*, 464-+.
George, R.E., Sanda, T., Hanna, M., Frohling, S., Luther, W., 2nd, Zhang, J., Ahn, Y., Zhou, W., London, W.B., McGrady, P.*, et al.* (2008). Activating mutations in ALK provide a therapeutic target in neuroblastoma. Nature *455*, 975-978.
Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C.*, et al.* (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. Mol Cell *55*, 829-842.
Ghosh, R., Das, D., and Franco, S. (2018). The Role for the DSB Response Pathway in Regulating Chromosome Translocations. Adv Exp Med Biol *1044*, 65-87.
Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods *6*, 343-U341.
Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C.*, et al.* (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647-661.
Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A.*, et al.* (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442-451.
Ginn, S.L., Fleming, J., Rowe, P.B., and Alexander, I.E. (2003). Promoter interference mediated by the U3 region in early-generation HIV-1-derived lentivirus vectors can influence detection of transgene expression in a cell-type and species-specific manner. Hum Gene Ther *14*, 1127-1137.
Glass, Z., Lee, M., Li, Y., and Xu, Q. (2018). Engineering the Delivery System for CRISPR-Based Genome Editing. Trends Biotechnol *36*, 173-185.
Goomer, R.S., and Kunkel, G.R. (1992). The Transcriptional Start Site for a Human U6 Small Nuclear-Rna Gene Is Dictated by a Compound Promoter Element Consisting of the Pse and the Tata Box. Nucleic Acids Res *20*, 4903-4912.
Gothe, H.J., Minneker, V., and Roukos, V. (2018). Dynamics of Double-Strand Breaks: Implications for the Formation of Chromosome Translocations. Adv Exp Med Biol *1044*, 27-38.
Greger, I.H., Demarchi, F., Giacca, M., and Proudfoot, N.J. (1998). Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. Nucleic Acids Res *26*, 1294-1301.
Grez, M., Akgun, E., Hilberg, F., and Ostertag, W. (1990). Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. Proc Natl Acad Sci U S A *87*, 9202-9206.
Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K., and Pandolfi, P.P. (2005). Role of nucleophosmin in embryonic development and tumorigenesis. Nature *437*, 147-153.

Grisendi, S., Mecucci, C., Falini, B., and Pandolfi, P.P. (2006). Nucleophosmin and cancer. Nat Rev Cancer *6*, 493-505.
Groenen, P.M., Bunschoten, A.E., van Soolingen, D., and van Embden, J.D. (1993). Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. Mol Microbiol *10*, 1057-1065.
Groffen, J., Stephenson, J.R., Heisterkamp, N., de Klein, A., Bartram, C.R., and Grosveld, G. (1984). Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. Cell *36*, 93-99.
Guarnerio, J., Bezzi, M., Jeong, J.C., Paffenholz, S.V., Berry, K., Naldini, M.M., Lo-Coco, F., Tay, Y., Beck, A.H., and Pandolfi, P.P. (2016). Oncogenic Role of Fusion-circRNAs Derived from Cancer-Associated Chromosomal Translocations. Cell *165*, 289-302.
Gundry, M.C., Brunetti, L., Lin, A., Mayle, A.E., Kitano, A., Wagner, D., Hsu, J.I., Hoegenauer, K.A., Rooney, C.M., Goodell, M.A.*, et al.* (2016). Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. Cell Rep *17*, 1453-1461.
Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y.*, et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900-910.
Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G.F., and Chin, L. (2016). Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. Cell Rep *14*, 1555-1566.
Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. Nat Med.
Haldar, M., Hancock, J.D., Coffin, C.M., Lessnick, S.L., and Capecchi, M.R. (2007). A conditional mouse model of synovial sarcoma: insights into a myogenic origin. Cancer Cell *11*, 375-388.
Hallberg, B., and Palmer, R.H. (2013). Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. Nat Rev Cancer *13*, 685-700.
Hallberg, B., and Palmer, R.H. (2016). The role of the ALK receptor in cancer biology. Ann Oncol *27 Suppl 3*, iii4-iii15.
Han, T., Schatoff, E.M., Murphy, C., Zafra, M.P., Wilkinson, J.E., Elemento, O., and Dow, L.E. (2017). R-Spondin chromosome rearrangements drive Wnt-dependent tumour initiation and maintenance in the intestine. Nat Commun *8*, 15945.
Han, X., Liu, Z., Jo, M.C., Zhang, K., Li, Y., Zeng, Z., Li, N., Zu, Y., and Qin, L. (2015). CRISPR-Cas9 delivery to hard-to-transfect cells via membrane deformation. Sci Adv *1*, e1500454.
Hanahan, D., Wagner, E.F., and Palmiter, R.D. (2007). The origins of oncomice: a history of the first transgenic mice genetically engineered to develop cancer. Gene Dev *21*, 2258-2270.
Hapgood, G., and Savage, K.J. (2015). The biology and management of systemic anaplastic large cell lymphoma. Blood *126*, 17-25.
Haviernik, P., Zhang, Y., and Bunting, K.D. (2008). Retroviral transduction of murine hematopoietic stem cells. Methods Mol Biol *430*, 229-241.
Hawley, R.G., Fong, A.Z., Burns, B.F., and Hawley, T.S. (1992). Transplantable myeloproliferative disease induced in mice by an interleukin 6 retrovirus. J Exp Med *176*, 1149-1163.
Heckl, D., Kowalczyk, M.S., Yudovich, D., Belizaire, R., Puram, R.V., McConkey, M.E., Thielke, A., Aster, J.C., Regev, A., and Ebert, B.L. (2014). Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. Nat Biotechnol *32*, 941-946.
Heinz, N., Schambach, A., Galla, M., Maetzig, T., Baum, C., Loew, R., and Schiedlmeier,

B. (2011). Retroviral and transposon-based tet-regulated all-in-one vectors with reduced background expression and improved dynamic range. Hum Gene Ther *22*, 166-176.
Heisterkamp, N., Stephenson, J.R., Groffen, J., Hansen, P.F., de Klein, A., Bartram, C.R., and Grosveld, G. (1983). Localization of the c-ab1 oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. Nature *306*, 239-242.
Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature *519*, 199-202.
Hemann, M. (2015). Reconstitution of Mice with Modified Hematopoietic Stem Cells. Cold Spring Harb Protoc *2015*, 679-684.
Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., Steinfeld, I., Lunstad, B.D., Kaiser, R.J., Wilkens, A.B.*, et al.* (2015). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nat Biotechnol *33*, 985-989.
Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. Cell *172*, 1239-1259.
Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol *33*, 510-517.
Hoe, N., Nakashima, K., Grigsby, D., Pan, X., Dou, S.J., Naidich, S., Garcia, M., Kahn, E., Bergmire-Sweat, D., and Musser, J.M. (1999). Rapid molecular genetic subtyping of serotype M1 group A Streptococcus strains. Emerg Infect Dis *5*, 254-263.
Hofker, M.H., and Breuer, M. (1998). Generation of transgenic mice. Methods Mol Biol *110*, 63-78.
Hogenbirk, M.A., Heideman, M.R., de Rink, I., Velds, A., Kerkhoven, R.M., Wessels, L.F., and Jacobs, H. (2016). Defining chromosomal translocation risks in cancer. Proc Natl Acad Sci U S A *113*, E3649-3656.
Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G.M., Kohn, D.B., Gregory, P.D., Holmes, M.C.*, et al.* (2010). Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. Nat Biotechnol *28*, 839-847.
Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.
Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol.
Hu, W.S., and Pathak, V.K. (2000). Design of retroviral vectors and helper cells for gene therapy. Pharmacol Rev *52*, 493-511.
Huettner, C.S., Zhang, P., Van Etten, R.A., and Tenen, D.G. (2000). Reversibility of acute B-cell leukaemia induced by BCR-ABL1. Nat Genet *24*, 57-60.
Hultquist, J.F., Schumann, K., Woo, J.M., Manganaro, L., McGregor, M.J., Doudna, J., Simon, V., Krogan, N.J., and Marson, A. (2016). A Cas9 Ribonucleoprotein Platform for Functional Genetic Studies of HIV-Host Interactions in Primary Human T Cells. Cell Rep *17*, 1438-1452.
Iarovaia, O.V., Rubtsov, M., Ioudinkova, E., Tsfasman, T., Razin, S.V., and Vassetzky, Y.S. (2014). Dynamics of double strand breaks and chromosomal translocations. Mol Cancer *13*, 249.
Ichim, C.V., and Wells, R.A. (2011). Generation of high-titer viral preparations by concentration using successive rounds of ultracentrifugation. J Transl Med *9*, 137.
Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S.,

Chen, J., Sondey, M., Ye, C.*, et al.* (2018). p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nat Med.
Ilaria, R.L., Jr., and Van Etten, R.A. (1995). The SH2 domain of P210BCR/ABL is not required for the transformation of hematopoietic factor-dependent cells. Blood *86*, 3897-3904.
Inghirami, G., Macri, L., Cesarman, E., Chadburn, A., Zhong, J., and Knowles, D.M. (1994). Molecular characterization of CD30+ anaplastic large-cell lymphoma: high frequency of c-myc proto-oncogene activation. Blood *83*, 3581-3590.
Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol *169*, 5429-5433.
Itahana, K., Bhat, K.P., Jin, A., Itahana, Y., Hawke, D., Kobayashi, R., and Zhang, Y. (2003). Tumor suppressor ARF degrades B23, a nucleolar protein involved in ribosome biogenesis and cell proliferation. Mol Cell *12*, 1151-1164.
Iwahara, T., Fujimoto, J., Wen, D., Cupples, R., Bucay, N., Arakawa, T., Mori, S., Ratzkin, B., and Yamamoto, T. (1997). Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. Oncogene *14*, 439-449.
Jager, R., Hahne, J., Jacob, A., Egert, A., Schenkel, J., Wernert, N., Schorle, H., and Wellmann, A. (2005). Mice transgenic for NPM-ALK develop non-Hodgkin lymphomas. Anticancer Res *25*, 3191-3196.
Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002a). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol *43*, 1565-1575.
Jansen, R., van Embden, J.D., Gaastra, W., and Schouls, L.M. (2002b). Identification of a novel family of sequence repeats among prokaryotes. OMICS *6*, 23-33.
Jasin, M., and Haber, J.E. (2016). The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. DNA Repair (Amst) *44*, 6-16.
Jayavaradhan, R., Pillis, D., and Malik, P. (2018). A Versatile Tool for the Quantification of CRISPR/Cas9-Induced Genome Editing Events in Human Hematopoietic Cell Lines and Hematopoietic Stem/Progenitor Cells. J Mol Biol.
Jiang, F., and Doudna, J.A. (2017). CRISPR-Cas9 Structures and Mechanisms. Annu Rev Biophys *46*, 505-529.
Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2*, e00471.
Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S.*, et al.* (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. Science.
Jones, D.T., Kocialkowski, S., Liu, L., Pearson, D.M., Backlund, L.M., Ichimura, K., and Collins, V.P. (2008). Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas. Cancer Res *68*, 8673-8677.
Kabadi, A.M., Ousterout, D.G., Hilton, I.B., and Gersbach, C.A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. Nucleic Acids Res *42*, e147.
Kannan, R., and Ventura, A. (2015). The CRISPR revolution and its impact on cancer research. Swiss medical weekly *145*, w14230.
Katigbak, A., Cencic, R., Robert, F., Senecha, P., Scuoppo, C., and Pelletier, J. (2016). A CRISPR/Cas9 Functional Screen Identifies Rare Tumor Suppressors. Sci Rep *6*, 38968.
Katigbak, A., Robert, F., Paquet, M., and Pelletier, J. (2018). Inducible Genome Editing with Conditional CRISPR/Cas9 Mice. G3 (Bethesda).
Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat Methods *12*, 401-403.

Kelliher, M.A., McLaughlin, J., Witte, O.N., and Rosenberg, N. (1990). Induction of a chronic myelogenous leukemia-like syndrome in mice with v-abl and BCR/ABL. Proc Natl Acad Sci U S A *87*, 6649-6653.
Kim, H.J., Lee, H.J., Kim, H., Cho, S.W., and Kim, J. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res *19*, 1279-1288.
Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res *24*, 1012-1019.
Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A *93*, 1156-1160.
Kines, K.J., Mann, V.H., Morales, M.E., Shelby, B.D., Kalinna, B.H., Gobert, G.N., Chirgwin, S.R., and Brindley, P.J. (2006). Transduction of Schistosoma mansoni by vesicular stomatitis virus glycoprotein-pseudotyped Moloney murine leukemia retrovirus. Exp Parasitol *112*, 209-220.
Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R.*, et al.* (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481-485.
Klingeberg, C., Illert, A.L., Schneider, N., Peschel, C., Miething, C., and Duyster, J. (2013). Generation and characterization of a CD30 positive T-cell lymphoma mouse model resembling human Anaplastic Large Cell Lymphoma (ALCL). Onkologie *36*, 195-195.
Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420-424.
Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H.*, et al.* (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583-588.
Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol *37*, 67-78.
Korgaonkar, C., Hagen, J., Tompkins, V., Frazier, A.A., Allamargot, C., Quelle, F.W., and Quelle, D.E. (2005). Nucleophosmin (B23) targets ARF to nucleoli and inhibits its function. Mol Cell Biol *25*, 1258-1271.
Korkmaz, G., Lopes, R., Ugalde, A.P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat Biotechnol *34*, 192-198.
Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol.
Kouranova, E., Forbes, K., Zhao, G., Warren, J., Bartels, A., Wu, Y., and Cui, X. (2016). CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos. Hum Gene Ther *27*, 464-475.
Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M.*, et al.* (2015). Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. Cell Reports *10*, 833-839.
Kreutmair, S., Klingeberg, C., Andrieux, G., Keller, A., Miething, C., Pfeifer, D., Follo, M., Busch, H., Boerries, M., Duyster, J.*, et al.* (2017). Identification and characterisation of the lymphoma initiating cell (LIC) population in an anaplastic large cell lymphoma (ALCL) mouse model. Oncol Res Treat *40*, 109-109.
Krivtsov, A.V., Twomey, D., Feng, Z., Stubbs, M.C., Wang, Y., Faber, J., Levine, J.E., Wang, J., Hahn, W.C., Gilliland, D.G.*, et al.* (2006). Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature *442*, 818-822.
Kuefer, M.U., Look, A.T., Pulford, K., Behm, F.G., Pattengale, P.K., Mason, D.Y., and

Morris, S.W. (1997). Retrovirus-mediated gene transfer of NPM-ALK causes lymphoid malignancy in mice. Blood *90*, 2901-2910.
Kumar, M., Keller, B., Makalou, N., and Sutton, R.E. (2001). Systematic determination of the packaging limit of lentiviral vectors. Hum Gene Ther *12*, 1893-1905.
Kumar, S., Vo, A.D., Qin, F., and Li, H. (2016). Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. Sci Rep *6*, 21597.
Kumar-Sinha, C., Kalyana-Sundaram, S., and Chinnaiyan, A.M. (2015). Landscape of gene fusions in epithelial cancers: seq and ye shall find. Genome Med *7*, 129.
Kurosawa, A., Saito, S., Mori, M., and Adachi, N. (2012). Nucleofection-based gene targeting in human pre-B cells. Gene *492*, 305-308.
Kutok, J.L., and Aster, J.C. (2002). Molecular biology of anaplastic lymphoma kinase-positive anaplastic large-cell lymphoma. J Clin Oncol *20*, 3691-3702.
La Starza, R., Matteucci, C., Gorello, P., Brandimarte, L., Pierini, V., Crescenzi, B., Nofrini, V., Rosati, R., Gottardi, E., Saglio, G.*, et al.* (2010). NPM1 deletion is associated with gross chromosomal rearrangements in leukemia. PLoS One *5*, e12855.
Lackner, D.H., Carre, A., Guzzardo, P.M., Banning, C., Mangena, R., Henley, T., Oberndorfer, S., Gapp, B.V., Nijman, S.M., Brummelkamp, T.R.*, et al.* (2015). A generic strategy for CRISPR-Cas9-mediated gene tagging. Nat Commun *6*, 10237.
Lackraj, T., Goswami, R., and Kridel, R. (2018). Pathogenesis of follicular lymphoma. Best Pract Res Clin Haematol *31*, 2-14.
Lagutina, I.V., Valentine, V., Picchione, F., Harwood, F., Valentine, M.B., Villarejo-Balcells, B., Carvajal, J.J., and Grosveld, G.C. (2015). Modeling of the human alveolar rhabdomyosarcoma Pax3-Foxo1 chromosome translocation in mouse myoblasts using CRISPR-Cas9 nuclease. PLoS Genet *11*, e1004951.
Lander, E.S. (2016). The Heroes of CRISPR. Cell *164*, 18-28.
Lange, K., Uckert, W., Blankenstein, T., Nadrowitz, R., Bittner, C., Renauld, J.C., van Snick, J., Feller, A.C., and Merz, H. (2003). Overexpression of NPM-ALK induces different types of malignant lymphomas in IL-9 transgenic mice. Oncogene *22*, 517-527.
Lau, C.H., and Suh, Y. (2017). In vivo genome editing in animals using AAV-CRISPR system: applications to translational research of human disease. F1000Res *6*, 2153.
Le Beau, M.M., Bitter, M.A., Larson, R.A., Doane, L.A., Ellis, E.D., Franklin, W.A., Rubin, C.M., Kadin, M.E., and Vardiman, J.W. (1989). The t(2;5)(p23;q35): a recurring chromosomal abnormality in Ki-1-positive anaplastic large cell lymphoma. Leukemia *3*, 866-870.
Lee, H.J., Kweon, J., Kim, E., Kim, S., and Kim, J.S. (2012). Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. Genome Res *22*, 539-548.
Lee, K., Conboy, M., Park, H.M., Jiang, F.G., Kim, H.J., Dewitt, M.A., Mackley, V.A., Chang, K., Rao, A., Skinner, C.*, et al.* (2017). Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. Nat Biomed Eng *1*, 889-901.
Lekomtsev, S., Aligianni, S., Lapao, A., and Burckstummer, T. (2016). Efficient generation and reversion of chromosomal translocations using CRISPR/Cas technology. BMC Genomics *17*, 739.
Lemaitre, C., Grabarz, A., Tsouroula, K., Andronov, L., Furst, A., Pankotai, T., Heyer, V., Rogier, M., Attwood, K.M., Kessler, P.*, et al.* (2014). Nuclear position dictates DNA repair pathway choice. Genes Dev *28*, 2450-2463.
Lentz, T.B., Gray, S.J., and Samulski, R.J. (2012). Viral vectors for gene delivery to the central nervous system. Neurobiol Dis *48*, 179-188.
Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., Li, Y., Gao, N., Wang, L., Lu, X.*, et al.* (2013a). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system.

Nat Biotechnol *31*, 681-683.
Li, H., Shi, J., Huang, N.J., Pishesha, N., Natarajan, A., Eng, J.C., and Lodish, H.F. (2016a). Efficient CRISPR-Cas9 mediated gene disruption in primary erythroid progenitor cells. Haematologica.
Li, W., Teng, F., Li, T., and Zhou, Q. (2013b). Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. Nat Biotechnol *31*, 684-686.
Li, X., Wu, R., and Ventura, A. (2016b). The present and future of genome editing in cancer research. Hum Genet *135*, 1083-1092.
Liang, X., Potter, J., Kumar, S., Ravinder, N., and Chesnut, J.D. (2017). Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. J Biotechnol *241*, 136-146.
Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S*., et al.* (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J Biotechnol *208*, 44-53.
Lieber, M.R., Gu, J., Lu, H., Shimazaki, N., and Tsai, A.G. (2010). Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans. Subcell Biochem *50*, 279-296.
Lin, J.J., Riely, G.J., and Shaw, A.T. (2017). Targeting ALK: Precision Medicine Takes on Drug Resistance. Cancer Discov *7*, 137-155.
Lin, S., Staahl, B.T., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife *3*, e04766.
Lindsley, D.L., Zimm, G.G., and Lindsley, D.L. (1992). The genome of Drosophila melanogaster (San Diego: Academic Press).
Ling, A.K., So, C.C., Le, M.X., Chen, A.Y., Hung, L., and Martin, A. (2018). Double-stranded DNA break polarity skews repair pathway choice during intrachromosomal and interchromosomal recombination. Proc Natl Acad Sci U S A.
Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233-247 e217.
Liu, Y., Han, X., Yuan, J., Geng, T., Chen, S., Hu, X., Cui, I.H., and Cui, H. (2017). Biallelic insertion of a transcriptional terminator via the CRISPR/Cas9 system efficiently silences expression of protein-coding and non-coding RNA genes. J Biol Chem *292*, 5624-5633.
Long, C., McAnally, J.R., Shelton, J.M., Mireault, A.A., Bassel-Duby, R., and Olson, E.N. (2014). Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. Science *345*, 1184-1188.
Lu, H., Villafane, N., Dogruluk, T., Grzeskowiak, C.L., Kong, K., Tsang, Y.H., Zagorodna, O., Pantazi, A., Yang, L., Neill, N.J*., et al.* (2017). Engineering and Functional Characterization of Fusion Genes Identifies Novel Oncogenic Drivers of Cancer. Cancer Res *77*, 3502-3512.
Lukasova, E., Kozubek, S., Kozubek, M., Kjeronska, J., Ryznar, L., Horakova, J., Krahulcova, E., and Horneck, G. (1997). Localisation and distance between ABL and BCR genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia. Hum Genet *100*, 525-535.
Luo, B., Heard, A.D., and Lodish, H.F. (2004). Small interfering RNA production by enzymatic engineering of DNA (SPEED). Proc Natl Acad Sci U S A *101*, 5494-5499.
Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R*., et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.
Ma, D., and Liu, F. (2015). Genome Editing and Its Applications in Model Organisms.

Genomics Proteomics Bioinformatics *13*, 336-344.
Ma, H., Nguyen, B., Li, L., Greenblatt, S., Williams, A., Zhao, M., Levis, M., Rudek, M., Duffield, A., and Small, D. (2014a). TTT-3002 is a novel FLT3 tyrosine kinase inhibitor with activity against FLT3-associated leukemias in vitro and in vivo. Blood *123*, 1525-1534.
Ma, H., Wu, Y., Dang, Y., Choi, J.G., Zhang, J., and Wu, H. (2014b). Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. Mol Ther Nucleic Acids *3*, e161.
Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E.*, et al.* (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature *516*, 423-427.
Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E.*, et al.* (2015). Corrigendum: In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature *524*, 502.
Maddalo, D., and Ventura, A. (2016). Somatic Engineering of Oncogenic Chromosomal Rearrangements: A Perspective. Cancer Res *76*, 4918-4923.
Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E.*, et al.* (2013). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat Biotechnol.
Maetzig, T., Galla, M., Baum, C., and Schambach, A. (2011). Gammaretroviral vectors: biology, technology and application. Viruses *3*, 677-713.
Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H.*, et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol *13*, 722-736.
Makarova, K.S., Zhang, F., and Koonin, E.V. (2017a). SnapShot: Class 1 CRISPR-Cas Systems. Cell *168*, 946-946 e941.
Makarova, K.S., Zhang, F., and Koonin, E.V. (2017b). SnapShot: Class 2 CRISPR-Cas Systems. Cell *168*, 328-328 e321.
Malcolm, T.I., Villarese, P., Fairbairn, C.J., Lamant, L., Trinquand, A., Hook, C.E., Burke, G.A., Brugieres, L., Hughes, K., Payet, D.*, et al.* (2016). Anaplastic large cell lymphoma arises in thymocytes and requires transient TCR expression for thymic egress. Nat Commun *7*, 10087.
Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., Dicarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-Guided Human Genome Engineering via Cas9. Science (New York, NY).
Malina, A., Katigbak, A., Cencic, R., Maiga, R.I., Robert, F., Miura, H., and Pelletier, J. (2014). Adapting CRISPR/Cas9 for functional genomics screens. Methods Enzymol *546*, 193-213.
Malina, A., Mills, J.R., Cencic, R., Yan, Y., Fraser, J., Schippers, L.M., Paquet, M., Dostie, J., and Pelletier, J. (2013). Repurposing CRISPR/Cas9 for in situ functional assays. Genes Dev *27*, 2602-2614.
Mamaeva, S.E. (1998). Karyotypic evolution of cells in culture: a new concept. Int Rev Cytol *178*, 1-40.
Mandal, P.K., Ferreira, L.M., Collins, R., Meissner, T.B., Boutwell, C.L., Friesen, M., Vrbanac, V., Garrison, B.S., Stortchevoi, A., Bryder, D.*, et al.* (2014). Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. Cell Stem Cell *15*, 643-652.
Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol *33*, 538-542.
Masepohl, B., Gorlitz, K., and Bohme, H. (1996). Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium Anabaena sp. PCC 7120. Biochim Biophys Acta *1307*, 26-30.
Mathas, S., Kreher, S., Meaburn, K.J., Johrens, K., Lamprecht, B., Assaf, C., Sterry, W., Kadin, M.E., Daibata, M., Joos, S.*, et al.* (2009). Gene deregulation and spatial genome reorganization near breakpoints prior to formation of translocations in anaplastic large cell lymphoma. Proc Natl Acad Sci U S A *106*, 5831-5836.
Mathew, P., Sanger, W.G., Weisenburger, D.D., Valentine, M., Valentine, V., Pickering, D., Higgins, C., Hess, M., Cui, X., Srivastava, D.K.*, et al.* (1997). Detection of the t(2;5)(p23;q35) and NPM-ALK fusion in non-Hodgkin's lymphoma by two-color fluorescence in situ hybridization. Blood *89*, 1678-1685.
McCord, R.P., and Balajee, A. (2018). 3D Genome Organization Influences the Chromosome Translocation Pattern. Adv Exp Med Biol *1044*, 113-133.
McCormack, M.P., Young, L.F., Vasudevan, S., de Graaf, C.A., Codrington, R., Rabbitts, T.H., Jane, S.M., and Curtis, D.J. (2010). The Lmo2 oncogene initiates leukemia in mice by inducing thymocyte self-renewal. Science *327*, 879-883.
Mduff, F.K., Hook, C.E., Tooze, R.M., Huntly, B.J., Pandolfi, P.P., and Turner, S.D. (2011). Determining the contribution of NPM1 heterozygosity to NPM-ALK-induced lymphomagenesis. Lab Invest *91*, 1298-1303.
Melnick, A., Carlile, G.W., McConnell, M.J., Polinger, A., Hiebert, S.W., and Licht, J.D. (2000). AML-1/ETO fusion protein is a dominant negative inhibitor of transcriptional repression by the promyelocytic leukemia zinc finger protein. Blood *96*, 3939-3947.
Melnick, A., and Licht, J.D. (1999). Deconstructing a disease: RARalpha, its fusion partners, and their roles in the pathogenesis of acute promyelocytic leukemia. Blood *93*, 3167-3215.
Melo, J.V. (1997). BCR-ABL gene variants. Baillieres Clin Haematol *10*, 203-222.
Mertens, F., Johansson, B., Fioretos, T., and Mitelman, F. (2015). The emerging complexity of gene fusions in cancer. Nat Rev Cancer *15*, 371-381.
Miething, C., Grundler, R., Fend, F., Hoepfl, J., Mugler, C., Schilling, C., Morris, S.W., Peschel, C., and Duyster, J. (2003). The oncogenic fusion protein nucleophosmin-anaplastic lymphoma kinase (NPM-ALK) induces two distinct malignant phenotypes in a murine retroviral transplantation model. Oncogene *22*, 4642-4647.
Miller, A.D., Eckner, R.J., Jolly, D.J., Friedmann, T., and Verma, I.M. (1984). Expression of a retrovirus encoding human HPRT in mice. Science *225*, 630-632.
Mitelman, F. (2015). A Short History of Chromosome Rearrangements and Gene Fusions in Cancer. In Chromosomal Translocations and Genome Rearrangements in Cancer, J.D. Rowley, M.M. Le Beau, and T.H. Rabbitts, eds. (Cham: Springer International Publishing), pp. 3-11.
Mitelman, F. (2018). Cancer Gene Fusions Detected by Massive Parallel Sequencing (European Cytogeneticists Association), pp. 11 - 14.
Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer *7*, 233-245.
Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E., and Hayakawa, T. (2000). IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. Mol Ther *1*, 376-382.
Modarai, S.R., Man, D., Bialk, P., Rivera-Torres, N., Bloh, K., and Kmiec, E.B. (2018). Efficient Delivery and Nuclear Uptake Is Not Sufficient to Detect Gene Editing in CD34+ Cells Directed by a Ribonucleoprotein Complex. Mol Ther Nucleic Acids *11*, 116-129.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.
Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol *60*, 174-182.
Mojica, F.J., Diez-Villasenor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol *36*, 244-246.
Mojica, F.J., Ferrer, C., Juez, G., and Rodriguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. Mol Microbiol *17*, 85-93.
Mojica, F.J., Juez, G., and Rodriguez-Valera, F. (1993). Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Mol Microbiol *9*, 613-621.
Mojica, F.J., and Rodriguez-Valera, F. (2016). The discovery of CRISPR in archaea and bacteria. FEBS J *283*, 3162-3169.
Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods *12*, 982-988.
Morris, S.W., Kirstein, M.N., Valentine, M.B., Dittmer, K.G., Shapiro, D.N., Saltman, D.L., and Look, A.T. (1994). Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science *263*, 1281-1284.
Mou, H., Kennedy, Z., Anderson, D.G., Yin, H., and Xue, W. (2015). Precision cancer mouse models through genome editing with CRISPR-Cas9. Genome Med *7*, 53.
Muerdter, F., Boryn, L.M., Woodfin, A.R., Neumayr, C., Rath, M., Zabidi, M.A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R.R.*, et al.* (2018). Resolving systematic errors in widely used enhancer activity assays in human cells. Nature Methods *15*, 141-+.
Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S., and Hannon, G.J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc Natl Acad Sci U S A *102*, 12135-12140.
Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., and Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res *39*, 9283-9293.
Nakajima, K., Ikenaka, K., Nakahira, K., Morita, N., and Mikoshiba, K. (1993). An improved retroviral vector for assaying promoter activity. Analysis of promoter interference in pIP211 vector. FEBS Lett *315*, 129-133.
Nakata, A., Amemura, M., and Makino, K. (1989). Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. J Bacteriol *171*, 3553-3556.
Nelson, K.N., Peiris, M.N., Meyer, A.N., Siari, A., and Donoghue, D.J. (2017). Receptor Tyrosine Kinases: Translocation Partners in Hematopoietic Disorders. Trends Mol Med *23*, 59-79.
Neves, H., Ramos, C., da Silva, M.G., Parreira, A., and Parreira, L. (1999). The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. Blood *93*, 1197-1207.
Nie, L., Das Thakur, M., Wang, Y., Su, Q., Zhao, Y., and Feng, Y. (2010). Regulation of U6 promoter activity by transcriptional interference in viral vector-based RNAi. Genomics Proteomics Bioinformatics *8*, 170-179.

Norrman, K., Fischer, Y., Bonnamy, B., Wolfhagen Sand, F., Ravassard, P., and Semb, H. (2010). Quantitative comparison of constitutive promoters in human ES cells. PLoS One *5*, e12413.
Nowell, P.C., and Hungerford, D.A. (1960). Minute Chromosome in Human Chronic Granulocytic Leukemia. Science *132*, 1497-1497.
Oppermann, H., Levinson, A.D., Varmus, H.E., Levintow, L., and Bishop, J.M. (1979). Uninfected vertebrate cells contain a protein that is closely related to the product of the avian sarcoma virus transforming gene (src). Proc Natl Acad Sci U S A *76*, 1804-1808.
Ortinski, P.I., O'Donovan, B., Dong, X.Y., and Kantor, B. (2017). Integrase-Deficient Lentiviral Vector as an All-in-One Platform for Highly Efficient CRISPR/Cas9-Mediated Gene Editing. Mol Ther-Meth Clin D *5*, 153-164.
Ozawa, T., Brennan, C.W., Wang, L., Squatrito, M., Sasayama, T., Nakada, M., Huse, J.T., Pedraza, A., Utsuki, S., Yasui, Y.*, et al.* (2010). PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas. Genes Dev *24*, 2205-2218.
Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol.
Pearson, J.D., Lee, J.K., Bacani, J.T., Lai, R., and Ingham, R.J. (2012). NPM-ALK: The Prototypic Member of a Family of Oncogenic Fusion Tyrosine Kinases. J Signal Transduct *2012*, 123253.
Peng, R., Lin, G., and Li, J. (2016). Potential pitfalls of CRISPR/Cas9-mediated genome editing. FEBS J *283*, 1218-1231.
Perez, A.R., Pritykin, Y., Vidigal, J.A., Chhangawala, S., Zamparo, L., Leslie, C.S., and Ventura, A. (2017). GuideScan software for improved single and paired CRISPR guide RNA design. Nat Biotechnol *35*, 347-349.
Persons, D.A., Allay, J.A., Allay, E.R., Smeyne, R.J., Ashmun, R.A., Sorrentino, B.P., and Nienhuis, A.W. (1997). Retroviral-mediated transfer of the green fluorescent protein gene into murine hematopoietic cells facilitates scoring and selection of transduced progenitors in vitro and identification of genetically modified cells in vivo. Blood *90*, 1777-1786.
Piganeau, M., Ghezraoui, H., De Cian, A., Guittat, L., Tomishima, M., Perrouault, L., Rene, O., Katibah, G., Zhang, L., Holmes, M.*, et al.* (2013). Cancer translocations in human cells induced by zinc finger and TALE nucleases. Genome Res.
Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M.*, et al.* (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440-455.
Qin, J.Y., Zhang, L., Clift, K.L., Hulur, I., Xiang, A.P., Ren, B.Z., and Lahn, B.T. (2010). Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. PLoS One *5*, e10611.
Ramezani, A., Hawley, T.S., and Hawley, R.G. (2000). Lentiviral vectors for enhanced gene expression in human hematopoietic cells. Mol Ther *2*, 458-469.
Ramirez-Solis, R., Liu, P., and Bradley, A. (1995). Chromosome engineering in mice. Nature *378*, 720-724.
Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S.*, et al.* (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186-191.
Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y.*, et al.* (2013a). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell.
Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.
Raper, A.T., Stephenson, A.A., and Suo, Z. (2018). Functional Insights Revealed by the

Kinetic Mechanism of CRISPR/Cas9. J Am Chem Soc *140*, 2971-2984.
Rassidakis, G.Z., Thomaides, A., Wang, S., Jiang, Y., Fourtouna, A., Lai, R., and Medeiros, L.J. (2005). p53 gene mutations are uncommon but p53 is commonly expressed in anaplastic large-cell lymphoma. Leukemia *19*, 1663-1669.
Reams, A.B., and Roth, J.R. (2015). Mechanisms of gene duplication and amplification. Cold Spring Harb Perspect Biol *7*, a016592.
Reimer, J., Knoss, S., Labuhn, M., Charpentier, E.M., Gohring, G., Schlegelberger, B., Klusmann, J.H., and Heckl, D. (2017). CRISPR-Cas9-induced t(11;19)/MLL-ENL translocations initiate leukemia in human hematopoietic progenitor cells in vivo. Haematologica *102*, 1558-1566.
Ren, C., Xu, K., Segal, D.J., and Zhang, Z. (2018). Strategies for the Enrichment and Selection of Genetically Modified Cells. Trends Biotechnol.
Reya, T., Duncan, A.W., Ailles, L., Domen, J., Scherer, D.C., Willert, K., Hintz, L., Nusse, R., and Weissman, I.L. (2003). A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature *423*, 409-414.
Richardson, C., and Jasin, M. (2000). Frequent chromosomal translocations induced by DNA double-strand breaks. Nature *405*, 697-700.
Roelz, R., Pilz, I.H., Mutschler, M., and Pahl, H.L. (2010). Of mice and men: human RNA polymerase III promoter U6 is more efficient than its murine homologue for shRNA expression from a lentiviral vector in both human and murine progenitor cells. Exp Hematol *38*, 792-797.
Ross, J.S., Wang, K., Chmielecki, J., Gay, L., Johnson, A., Chudnovsky, J., Yelensky, R., Lipson, D., Ali, S.M., Elvin, J.A.*, et al.* (2016). The distribution of BRAF gene fusions in solid tumors and response to targeted therapy. Int J Cancer *138*, 881-890.
Rouet, P., Smih, F., and Jasin, M. (1994a). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci U S A *91*, 6064-6068.
Rouet, P., Smih, F., and Jasin, M. (1994b). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol *14*, 8096-8106.
Roukos, V., and Misteli, T. (2014). The biogenesis of chromosome translocations. Nat Cell Biol *16*, 293-300.
Roukos, V., Voss, T.C., Schmidt, C.K., Lee, S., Wangsa, D., and Misteli, T. (2013). Spatial dynamics of chromosome translocations in living cells. Science *341*, 660-664.
Rowley, J.D. (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature *243*, 290-293.
Rowley, J.D. (2008). Chromosomal translocations: revisited yet again. Blood *112*, 2183-2189.
Ruggero, K., and Rabbitts, T.H. (2015). Pre-clinical Modelling of Chromosomal Translocations and Inversions. In Chromosomal Translocations and Genome Rearrangements in Cancer, J.D. Rowley, M.M. Le Beau, and T.H. Rabbitts, eds. (Cham: Springer International Publishing), pp. 429-445.
Ryu, W.S. (2017). Molecular Virology of Human Pathogenic Viruses. Molecular Virology of Human Pathogenic Viruses, 1-423.
Sahu, A., Prabhash, K., Noronha, V., Joshi, A., and Desai, S. (2013). Crizotinib: A comprehensive review. South Asian J Cancer *2*, 91-97.
Sanchez-Rivera, F.J., Papagiannakopoulos, T., Romero, R., Tammela, T., Bauer, M.R., Bhutkar, A., Joshi, N.S., Subbaraj, L., Bronson, R.T., Xue, W.*, et al.* (2014). Rapid modelling of cooperating genetic events in cancer through somatic genome editing. Nature *516*, 428-431.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods *11*, 783-784.
Schaefer, C., Grouse, L., Buetow, K., and Strausberg, R.L. (2001). A new cancer genome anatomy project web resource for the community. Cancer J *7*, 52-60.
Schambach, A., Bohne, J., Chandra, S., Will, E., Margison, G.P., Williams, D.A., and Baum, C. (2006). Equal potency of gammaretroviral and lentiviral SIN vectors for expression of O6-methylguanine-DNA methyltransferase in hematopoietic cells. Mol Ther *13*, 391-400.
Schambach, A., Swaney, W.P., and van der Loo, J.C. (2009). Design and production of retro- and lentiviral vectors for gene expression in hematopoietic cells. Methods Mol Biol *506*, 191-205.
Schmitt, C.A., Fridman, J.S., Yang, M., Baranov, E., Hoffman, R.M., and Lowe, S.W. (2002). Dissecting p53 tumor suppressor functions in vivo. Cancer cell *1*, 289-298.
Schumann, K., Lin, S., Boyer, E., Simeonov, D.R., Subramaniam, M., Gate, R.E., Haliburton, G.E., Ye, C.J., Bluestone, J.A., Doudna, J.A.*, et al.* (2015). Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. Proc Natl Acad Sci U S A *112*, 10437-10442.
Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell cycle *8*, 2557-2564.
Seki, A., and Rutz, S. (2018). Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. J Exp Med.
Shah, S.A., Erdmann, S., Mojica, F.J., and Garrett, R.A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. RNA biology *10*, 891-899.
Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G.*, et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84-87.
Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference--a crash course. Trends Genet *21*, 339-345.
Shi, J., Wang, E., Milazzo, J.P., Wang, Z., Kinney, J.B., and Vakoc, C.R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat Biotechnol *33*, 661-667.
Shou, J., Li, J., Liu, Y., and Wu, Q. (2018). Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. Mol Cell.
Shtivelman, E., Lifshitz, B., Gale, R.P., and Canaani, E. (1985). Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature *315*, 550-554.
Simmons, A., and Alberola-Ila, J. (2016). Retroviral Transduction of T Cells and T Cell Precursors. Methods Mol Biol *1323*, 99-108.
Sims, D., Mendes-Pereira, A.M., Frankum, J., Burgess, D., Cerone, M., Lombardelli, C., Mitsopoulos, C., Hakas, J., Murugaesu, N., Isacke, C.M.*, et al.* (2011). High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing. Genome biology *12*, R104.
Simsek, D., Brunet, E., Wong, S.Y., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J.*, et al.* (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet *7*, e1002080.
Singh, D., Chan, J.M., Zoppoli, P., Niola, F., Sullivan, R., Castano, A., Liu, E.M., Reichel, J., Porrati, P., Pellegatta, S.*, et al.* (2012). Transforming fusions of FGFR and TACC genes in human glioblastoma. Science *337*, 1231-1235.
Singh, D., Sternberg, S.H., Fei, J., Doudna, J.A., and Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. Nat Commun *7*, 12778.
Singh, P., Schimenti, J.C., and Bolcun-Filas, E. (2015). A mouse geneticist's practical

guide to CRISPR applications. Genetics *199*, 1-15.
Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84-88.
Slupianek, A., Nieborowska-Skorska, M., Hoser, G., Morrione, A., Majewski, M., Xue, L., Morris, S.W., Wasik, M.A., and Skorski, T. (2001). Role of phosphatidylinositol 3-kinase-Akt pathway in nucleophosmin/anaplastic lymphoma kinase-mediated lymphomagenesis. Cancer Res *61*, 2194-2199.
Smith, A.J., De Sousa, M.A., Kwabi-Addo, B., Heppell-Parton, A., Impey, H., and Rabbitts, P. (1995). A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. Nat Genet *9*, 376-385.
Soda, M., Choi, Y.L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H.*, et al.* (2007). Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature *448*, 561-566.
Soda, M., Takada, S., Takeuchi, K., Choi, Y.L., Enomoto, M., Ueno, T., Haruta, H., Hamada, T., Yamashita, Y., Ishikawa, Y.*, et al.* (2008). A mouse model for EML4-ALK-positive lung cancer. Proc Natl Acad Sci U S A *105*, 19893-19897.
Song, M. (2017). The CRISPR/Cas9 system: Their delivery, in vivo and ex vivo applications and clinical development by startups. Biotechnol Prog *33*, 1035-1045.
Soutoglou, E., Dorn, J.F., Sengupta, K., Jasin, M., Nussenzweig, A., Ried, T., Danuser, G., and Misteli, T. (2007). Positional stability of single double-strand breaks in mammalian cells. Nat Cell Biol *9*, 675-682.
Sportoletti, P., Grisendi, S., Majid, S.M., Cheng, K., Clohessy, J.G., Viale, A., Teruya-Feldstein, J., and Pandolfi, P.P. (2008). Npm1 is a haploinsufficient suppressor of myeloid and lymphoid malignancies in the mouse. Blood *111*, 3859-3862.
Spraggon, L., Martelotto, L.G., Hmeljak, J., Hitchman, T.D., Wang, J., Wang, L., Slotkin, E.K., Fan, P.D., Reis-Filho, J.S., and Ladanyi, M. (2017). Generation of conditional oncogenic chromosomal translocations using CRISPR-Cas9 genomic editing and homology-directed repair. J Pathol *242*, 102-112.
Staahl, B.T., Benekareddy, M., Coulon-Bainier, C., Banfal, A.A., Floor, S.N., Sabo, J.K., Urnes, C., Munares, G.A., Ghosh, A., and Doudna, J.A. (2017). Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. Nat Biotechnol *35*, 431-434.
Stam, K., Heisterkamp, N., Grosveld, G., de Klein, A., Verma, R.S., Coleman, M., Dosik, H., and Groffen, J. (1985). Evidence of a new chimeric bcr/c-abl mRNA in patients with chronic myelocytic leukemia and the Philadelphia chromosome. N Engl J Med *313*, 1429-1433.
Szymczak, A.L., and Vignali, D.A. (2005). Development of 2A peptide-based strategies in the design of multicistronic vectors. Expert Opin Biol Ther *5*, 627-638.
Tabebordbar, M., Zhu, K., Cheng, J.K.W., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C., Wu, E.Y., Xiao, R., Ran, F.A.*, et al.* (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science *351*, 407-411.
Temin, H.M., and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature *226*, 1211-1213.
Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.
Truong, D.J., Kuhner, K., Kuhn, R., Werfel, S., Engelhardt, S., Wurst, W., and Ortiz, O. (2015). Development of an intein-mediated split-Cas9 system for gene therapy. Nucleic Acids Res *43*, 6450-6458.
Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P.*, et al.* (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol *33*, 187-197.

Turner, S.D., and Alexander, D.R. (2005). What have we learnt from mouse models of NPM-ALK-induced lymphomagenesis? Leukemia *19*, 1128-1134.
Turner, S.D., and Alexander, D.R. (2006). Fusion tyrosine kinase mediated signalling pathways in the transformation of haematopoietic cells. Leukemia *20*, 572-582.
Turner, S.D., Tooze, R., Maclennan, K., and Alexander, D.R. (2003). Vav-promoter regulated oncogenic fusion protein NPM-ALK in transgenic mice causes B-cell lymphomas with hyperactive Jun kinase. Oncogene *22*, 7750-7761.
Ugale, A., Norddahl, G.L., Wahlestedt, M., Sawen, P., Jaako, P., Pronk, C.J., Soneji, S., Cammenga, J., and Bryder, D. (2014). Hematopoietic stem cells are intrinsically protected against MLL-ENL-mediated transformation. Cell Rep *9*, 1246-1255.
Uren, A.G., Kool, J., Berns, A., and Lohuizen, M. (2005). Retroviral insertional mutagenesis: past, present and future. Oncogene *24*, 7656-7672.
Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nature reviews Genetics *11*, 636-646.
Van Deursen, J., Fornerod, M., Van Rees, B., and Grosveld, G. (1995). Cre-mediated site-specific translocation between nonhomologous mouse chromosomes. Proc Natl Acad Sci U S A *92*, 7376-7380.
van Overbeek, M., Capurso, D., Carter, M.M., Thompson, M.S., Frias, E., Russ, C., Reece-Hoyes, J.S., Nye, C., Gradia, S., Vidal, B.*, et al.* (2016). DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. Mol Cell *63*, 633-646.
Vanoli, F., and Jasin, M. (2017). Generation of chromosomal translocations that lead to conditional fusion protein expression using CRISPR-Cas9 and homology-directed repair. Methods *121-122*, 138-145.
Vanoli, F., Tomishima, M., Feng, W., Lamribet, K., Babin, L., Brunet, E., and Jasin, M. (2017). CRISPR-Cas9-guided oncogenic chromosomal translocations with conditional fusion protein expression in human mesenchymal cells. Proc Natl Acad Sci U S A *114*, 3696-3701.
Ventura, A., and Dow, L.E. (2018). Modeling Cancer in the CRISPR Era. Annual Review of Cancer Biology *2*, 111-131.
Vidigal, J.A., and Ventura, A. (2015). Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat Commun *6*, 8083.
Walther, W., and Stein, U. (2000). Viral vectors for gene transfer - A review of their use in the treatment of human diseases. Drugs *60*, 249-271.
Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910-918.
Wang, H.X., Li, M., Lee, C.M., Chakraborty, S., Kim, H.W., Bao, G., and Leong, K.W. (2017). CRISPR/Cas9-Based Genome Editing for Disease Modeling and Therapy: Challenges and Opportunities for Nonviral Delivery. Chem Rev *117*, 9874-9906.
Wang, L., Li, F., Dang, L., Liang, C., Wang, C., He, B., Liu, J., Li, D., Wu, X., Xu, X.*, et al.* (2016). In Vivo Delivery Systems for Therapeutic Genome Editing. Int J Mol Sci *17*.
Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80-84.
Warmuth, M., Kim, S., Gu, X.J., Xia, G., and Adrian, F. (2007). Ba/F3 cells and their use in kinase drug discovery. Curr Opin Oncol *19*, 55-60.
Weckselblatt, B., and Rudd, M.K. (2015). Human Structural Variation: Mechanisms of Chromosome Rearrangements. Trends Genet *31*, 587-599.
Weinstock, D.M., Elliott, B., and Jasin, M. (2006). A model of oncogenic rearrangements: differences between chromosomal translocation mechanisms and simple double-strand break repair. Blood *107*, 777-780.
Welborn, J. (2004). Acquired Robertsonian translocations are not rare events in acute

leukemia and lymphoma. Cancer Genet Cytogenet *151*, 14-35.
Wheat, J.C., and Steidl, U. (2017). ETO2-GLIS2: A Chimeric Transcription Factor Drives Leukemogenesis through a Neomorphic Transcription Network. Cancer Cell *31*, 307-308.
Willis, N.A., Rass, E., and Scully, R. (2015). Deciphering the Code of the Cancer Genome: Mechanisms of Chromosome Rearrangement. Trends Cancer *1*, 217-230.
Wright, A.V., Nunez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell *164*, 29-44.
Wright, A.V., Sternberg, S.H., Taylor, D.W., Staahl, B.T., Bardales, J.A., Kornfeld, J.E., and Doudna, J.A. (2015). Rational design of a split-Cas9 enzyme complex. Proc Natl Acad Sci U S A *112*, 2984-2989.
Wu, X., Kriz, A.J., and Sharp, P.A. (2014). Target specificity of the CRISPR-Cas9 system. Quant Biol *2*, 59-70.
Xiao, A., Wang, Z., Hu, Y., Wu, Y., Luo, Z., Yang, Z., Zu, Y., Li, W., Huang, P., Tong, X.*, et al.* (2013). Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. Nucleic Acids Res *41*, e141.
Xu, H., Xiao, T., Chen, C.H., Li, W., Meyer, C.A., Wu, Q., Wu, D., Cong, L., Zhang, F., Liu, J.S.*, et al.* (2015). Sequence determinants of improved CRISPR sgRNA design. Genome Res *25*, 1147-1157.
Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N.S., Cai, W., Yang, G., Bronson, R., Crowley, D.G.*, et al.* (2014). CRISPR-mediated direct mutation of cancer genes in the mouse liver. Nature *514*, 380-384.
Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. Cell.
Yin, H., Xue, W., Chen, S., Bogorad, R.L., Benedetti, E., Grompe, M., Koteliansky, V., Sharp, P.A., Jacks, T., and Anderson, D.G. (2014). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. Nat Biotechnol *32*, 551-553.
Yuen, G., Khan, F.J., Gao, S., Stommel, J.M., Batchelor, E., Wu, X., and Luo, J. (2017). CRISPR/Cas9-mediated gene knockout is insensitive to target copy number but is dependent on guide RNA potency and Cas9/sgRNA threshold expression level. Nucleic Acids Res *45*, 12039-12053.
Zafra, M.P., Schatoff, E.M., Katti, A., Foronda, M., Breinig, M., Schweitzer, A.Y., Simon, A., Han, T., Goswami, S., Montgomery, E.*, et al.* (2018). Optimized base editors enable efficient editing in cells, organoids and mice. Nat Biotechnol.
Zeisig, B.B., and So, C.W. (2009). Retroviral/Lentiviral transduction and transformation assay. Methods Mol Biol *538*, 207-229.
Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A.*, et al.* (2015a). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.
Zetsche, B., Volz, S.E., and Zhang, F. (2015b). A split-Cas9 architecture for inducible genome editing and transcription modulation. Nat Biotechnol *33*, 139-142.
Zhang, Q., Raghunath, P.N., Xue, L., Majewski, M., Carpentieri, D.F., Odum, N., Morris, S., Skorski, T., and Wasik, M.A. (2002). Multilevel dysregulation of STAT3 activation in anaplastic lymphoma kinase-positive T/null-cell lymphoma. J Immunol *168*, 466-474.
Zhang, Y., and Jasin, M. (2011). An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat Struct Mol Biol *18*, 80-84.
Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell *148*, 908-921.
Zheng, B., Sage, M., Sheppeard, E.A., Jurecic, V., and Bradley, A. (2000). Engineering

mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. Mol Cell Biol *20*, 648-655.
Zhu, F., Feng, M., Sinha, R., Seita, J., Mori, Y., and Weissman, I.L. (2018). Screening for genes that regulate the differentiation of human megakaryocytic lineage cells. Proc Natl Acad Sci U S A.
Zhu, S., Li, W., Liu, J., Chen, C.H., Liao, Q., Xu, P., Xu, H., Xiao, T., Cao, Z., Peng, J.*, et al.* (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. Nat Biotechnol *34*, 1279-1286.
Zuris, J.A., Thompson, D.B., Shu, Y., Guilinger, J.P., Bessen, J.L., Hu, J.H., Maeder, M.L., Joung, J.K., Chen, Z.Y., and Liu, D.R. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat Biotechnol *33*, 73-80.