

ELUCIDATING MECHANISMS OF TOPOISOMERASE 2-INDUCED  
GENE FRAGILITY AND THE FORMATION OF ONCOGENIC,  
THERAPY-RELATED *MLL* TRANSLOCATIONS

**Dissertation**

Zur Erlangung des Grades  
Doktor der Naturwissenschaften

Am Fachbereich Biologie  
Der Johannes Gutenberg-Universität Mainz

**Henrike Johanna Gothe**

geboren am 10.01.1991

in Soltau, Deutschland

Mainz, 2020



Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 23. Juni 2020



# Contents

<b>Zusammenfassung</b>	<b>VII</b>
<b>Summary</b>	<b>IX</b>
<b>Preface</b>	<b>XI</b>
<b>List of publications</b>	<b>XIII</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
<b>1.1 DNA damage and cancer</b>	<b>1</b>
<b>1.2 Genomic rearrangements as cancer drivers</b>	<b>2</b>
<b>1.3 Therapy-related carcinogenesis</b>	<b>4</b>
<b>1.4 Topoisomerases are a source of genomic instability</b>	<b>6</b>
1.4.1 Topoisomerases relieve topological stress but are inherently risky	6
1.4.2 Isoform specific functions of TOP2A and TOP2B	8
1.4.3 Roles of TOP2 isoforms in ETO-induced toxicity and DNA fragility	10
1.4.4 Chemotherapies targeting TOP2s are associated with secondary leukemia and oncogenic <i>MLL</i> translocations	10
<b>1.5 Mechanisms of TOP2cc conversion to DNA breaks</b>	<b>12</b>
1.5.1 Etoposide induces genomic instability by conversion of TOP2ccs into DNA breaks	12
1.5.2 Transcription and replication contribute to TOP2-dependent DSB formation	12
1.5.3 Genome architecture as a new player in TOP2-induced genomic instability	14
<b>1.6 Repair of TOP2-DNA covalent complexes</b>	<b>15</b>
1.6.1 Repair of protein-bound DNA breaks	15
1.6.2 Reversal and removal of covalent TOP2-DNA complexes	16
1.6.3 Repair of TOP2-derived DSBs	18
<b>1.7 Methods to study genomic instability and the occurrence of oncogenic fusions</b>	<b>19</b>
1.7.1 Genome-wide profiling of DSBs	19
1.7.2 Methods to study rare chromosome translocations	21
1.7.3 Polymerase chain reaction to quantify specific chromosome translocations	21
1.7.4 Unbiased approaches by next-generation sequencing techniques	22
1.7.5 Microscopic detection of recurrent chromosome translocations by FISH	23
<b>1.8 Aims</b>	<b>24</b>

**Chapter 2 Results** \_\_\_\_\_ **27**

**2.1 Etoposide induces gene fragility and oncogenic translocations in hematopoietic cells** \_\_\_\_\_ **27**

- 2.1.1 Genome-wide profiling of etoposide-induced DSBs in human hematopoietic cells by sBLISS 27
- 2.1.2 sBLISS identifies DSB hotspots within *MLL* and recurrent fusion partners \_\_\_\_\_ 30
- 2.1.3 C-Fusion 3D: a novel tool to study nuclear positioning and rare chromosome rearrangements in 3D \_\_\_\_\_ 32
- 2.1.4 C-Fusion 3D detects oncogenic *MLL* translocations with high accuracy and sensitivity 36
- 2.1.5 TOP2-induced *MLL* fusion are rare and occur at frequencies of  $10^{-3}$  to  $10^{-4}$  \_\_\_\_\_ 38

**2.2 Transcription and spatial chromosome folding jointly induce genomic instability** \_\_\_\_\_ **42**

- 2.2.1 ETO-induced fragility is linked to transcriptional activity \_\_\_\_\_ 42
- 2.2.2 *MLL* and frequent partner genes are highly active and enriched at chromatin loop boundaries \_\_\_\_\_ 45
- 2.2.3 Transcriptional activity, localization at loop anchors and TOP2-induced breakage are interlinked \_\_\_\_\_ 48
- 2.2.4 DSBs at chromatin loop anchors are formed by transcriptional activity and directionality 50
- 2.2.5 The *MLL* DSB hotspot is located in proximity to a chromatin loop \_\_\_\_\_ 54
- 2.2.6 Formation of *MLL* translocations is mainly transcription-dependent \_\_\_\_\_ 58
- 2.2.7 *MLL* gene fragility is directly linked to its expression \_\_\_\_\_ 60

**2.3 The TOP2 isoforms alpha and beta both contribute to *MLL* fragility** \_\_\_\_\_ **64**

- 2.3.1 Cell models to study the contributions of TOP2 isoforms to ETO-induced DNA damage 64
- 2.3.2 Both TOP2A and TOP2B isoforms contribute to ETO-induced DNA damage and *MLL* translocations \_\_\_\_\_ 66
- 2.3.3 *MLL* fragility is cell cycle dependent \_\_\_\_\_ 68

**2.4 Repair of TOP2ccs is mediated by MRE11, TDP2 and NHEJ** \_\_\_\_\_ **71**

- 2.4.1 MRE11, TDP2 and NHEJ suppress ETO-induced DNA damage and the formation of *MLL* translocations \_\_\_\_\_ 71
- 2.4.2 Transcription and replication-dependent roles of MRE11 and TDP2 in the repair or TOP2ccs \_\_\_\_\_ 74

**2.5 Screening for new factors influencing TOP2-induced *MLL* fragility** \_\_\_\_\_ **77**

- 2.5.1 Loss-of-function screen identifies novel regulators of *MLL* fragility \_\_\_\_\_ 77
- 2.5.2 VCP and its cofactor PLAA might contribute to *MLL* fragility \_\_\_\_\_ 80
- 2.5.3 VCP mediates ETO-induced DNA damage and cytotoxicity \_\_\_\_\_ 83

<b>Chapter 3 Discussion</b>	<b>89</b>
<b>3.1 Etoposide induces DSB hotspots and leads to oncogenic <i>MLL</i> fusions</b>	<b>89</b>
<b>3.2 The dangerous interplay of transcription, DNA topology and genome architecture</b>	<b>90</b>
3.2.1 Transcription drives TOP2-linked gene fragility	90
3.2.2 Transcription and chromosome organization are interlinked in generating DNA damage	91
3.2.3 Breakage at the <i>MLL</i> DSB hotspot is enriched in proximity to a chromatin loop anchor and depends on gene activity	94
3.2.4 Transcription- dependent <i>MLL</i> fusions and spatial proximity	95
<b>3.3 Implications of TOP2 isoform specificity</b>	<b>97</b>
3.3.1 Both TOP2 isoforms contribute to ETO-induced DNA fragility	97
3.3.2 Therapeutic implications of TOP2 isoform specificity	98
<b>3.4 The multifaceted repair of TOP2-DNA complexes</b>	<b>98</b>
3.4.1 MRE11 and TDP2 are suppressors of <i>MLL</i> fragility and fusions	99
3.4.2 NHEJ counteracts the formation of oncogenic <i>MLL</i> fusions	100
3.4.3 VCP as a novel processing factor of TOP2ccs?	101
3.4.4 Targeting of TOP2cc repair in cancer therapy	103
<b>3.5 C-Fusion 3D advances the detection of rare chromosome rearrangements and identifies candidate regulators of <i>MLL</i> fragility</b>	<b>105</b>
3.5.1 Quantification of rare chromosome translocations by C-Fusion 3D	105
3.5.2 Additional candidate factors to modulate TOP2-dependent <i>MLL</i> gene fragility	106
<b>3.6 Summary</b>	<b>108</b>
<b>Chapter 4 Material and Methods</b>	<b>109</b>
<b>4.1 Maintenance of cell lines</b>	<b>109</b>
<b>4.2 Lentiviral transduction of human cell lines</b>	<b>110</b>
<b>4.3 Generation of knockout and engineered cell lines</b>	<b>111</b>
4.3.1 Generation of a stable TK6-Cas9 cell line	111
4.3.2 Generation of HCT116-TOP2A-mAID-TOP2B <sup>-/-</sup> cells	111
4.3.3 Generation of HTETOP-TOP2B <sup>-/-</sup> cells	112
4.3.4 Generation of TK6 ZFP64 <sup>-/-</sup> cells	113
4.3.5 Generation of the TK6 <i>MLL</i> -CTCF <sup>mut</sup> cell line	113
<b>4.4 Chemical inhibition of transcription and replication</b>	<b>114</b>
<b>4.5 Cell cycle synchronization and release</b>	<b>114</b>

<b>4.6</b>	<b>siRNA knockdown</b>	<b>115</b>
4.6.1	RNAi of TOP2A and TOP2B in Cal51 cells	115
4.6.2	RNAi screen for 150 factors in Cal51 cells (refers to Chapter 2.5.1)	115
4.6.3	RNAi screen for 33 VCP adaptors in Cal51 cells (refers to Chapter 2.5.2).	117
<b>4.7</b>	<b>Western blotting</b>	<b>119</b>
<b>4.8</b>	<b>ChIP-qPCR of CTCF in TK6 cells</b>	<b>120</b>
<b>4.9</b>	<b>Immunofluorescence</b>	<b>121</b>
<b>4.10</b>	<b>Fluorescence <i>in situ</i> hybridization (FISH)</b>	<b>122</b>
<b>4.11</b>	<b>High-Throughput Imaging</b>	<b>124</b>
<b>4.12</b>	<b>Image analysis</b>	<b>124</b>
4.12.1	Nuclei segmentation, intensity measurements for immunofluorescence analysis	124
4.12.2	C-Fusion 3D	125
<b>4.13</b>	<b>Cell harvesting and fixation for sBLISS</b>	<b>127</b>
<b>4.14</b>	<b>Supplementary methods: sBLISS</b>	<b>130</b>
4.14.1	sBLISS template preparation	130
4.14.2	sBLISS <i>in vitro</i> transcription (IVT) and NGS library preparation	131
4.14.3	sBLISS sequencing and data processing	132
4.14.4	sBLISS limitations on quantifying DSBs between different cell lines	132
<b>4.15</b>	<b>Supplementary methods: Bioinformatics analyses</b>	<b>133</b>
4.15.1	Detection of sBLISS hotspots	133
4.15.2	Definition of genomic regions	133
4.15.3	ENCODE data analysis	134
4.15.4	CTCF motifs	135
4.15.5	Chromatin Loop analyses	135
4.15.6	RNA-seq data analysis	136
4.15.7	GRO-seq data analysis	136
4.15.8	ChIP-seq data analysis	137
4.15.9	Nascent RNA-seq (transcription factories)	137
4.15.10	Overlap of genomic features	137
4.15.11	Statistical tests and corrections	138
4.15.12	Multiple correlation test	138
<b>4.16</b>	<b>Data availability and Software</b>	<b>138</b>
<b>References</b>		<b>141</b>
<b>List of Abbreviations</b>		<b>155</b>



<i>List of Figures</i>	157
<i>List of Tables</i>	161
<i>Acknowledgements</i>	163
<i>Curriculum Vitae</i>	164



## Zusammenfassung

Die Integrität des Genoms ist fortlaufend durch Schäden der DNA bedroht, welche zu Mutationen, Chromosomenveränderungen und Krankheiten führen können. Genomische Instabilität ist ein Kennzeichen von Krebszellen, welche oftmals von strukturellen Chromosomenveränderungen gekennzeichnet sind. Fehler bei der Reparatur von DNA-Schäden können zu Translokationen von Chromosomen führen, die oft schwere Mutationen hervorbringen und in Verbindung mit Krebserkrankungen stehen. Viele erfolgreiche Chemotherapien greifen zelluläre Topoisomerasen des Typ II (TOP2) an, wodurch kovalente Verbindungen zwischen TOP2 und der DNA entstehen. Topoisomerasen sind essentiell, um Torsionsspannungen der genomischen DNA aufzuheben, doch die Stabilisierung von TOP2 in kovalenten Komplexen mit der DNA führt zu toxischen DNA-Doppelstrangbrüchen und tötet Krebszellen dadurch ab. Diese Chemotherapeutika, sogenannte Topoisomerase-Gifte, werden jedoch auch in Verbindung mit der Entstehung von Leukemien gebracht, welche häufig durch wiederkehrenden Translokationen des Gens *mixed lineage leukemia (MLL)* gekennzeichnet werden. Die zellulären Vorgänge, welche auf die Umwandlung von TOP2-DNA-Komplexen zu DNA-Schäden hinwirken, sind allerdings kaum bekannt.

Wir haben die Methode C-Fusion 3D entwickelt, um in einzelnen Zellen Brüche und Umlagerungen von Chromosomen zu quantifizieren und somit Genomschäden, die von TOP2 verursacht werden, zu untersuchen. Durch die zusätzliche Anwendung eines Sequenzierungsverfahrens, das Doppelstrangbrüche im gesamten Genom erkennt, identifizieren wir wiederkehrende Hotspots für DNA-Brüche, die hauptsächlich in transkribierten Genomabschnitten zu finden sind. Wir erkennen eine Wechselbeziehung zwischen der Brüchigkeit von DNA, der Aktivität und Richtung der Transkription und dem Zusammentreffen mit Ankerpunkten der Genomarchitektur. Das *MLL*-Gen und Translokationspartner bilden Genfusionen mit Häufigkeiten im Bereich  $10^{-3}$  bis  $10^{-4}$ , die von der Aktivität beider TOP2-Isozyme abhängen. Die Interaktion von der Transkription mit Chromatinschlaufen trägt zu der Zerbrechlichkeit von Genen und der Entstehung von *MLL*-Translokationen bei. Ein mehrstufiges System von Reparaturprozessen wirkt der Entstehung von Genfusionen entgegen und schließt die Aktivität von TDP2, MRE11 und der Nicht-Homologen Endverknüpfung (NHEJ) ein. Durch ein systematisches Testen erkennen wir, dass die ATPase VCP zu der TOP2-induzierten Zerbrechlichkeit von *MLL* beiträgt.

Zusammenfassend stellen wir fest, dass sich durch die Transkription und die räumliche Strukturierung des Genoms ein Risiko von Instabilität ergibt, welche zu der Entstehung von wiederkehrenden, krebserregenden *MLL*-Translokationen beiträgt.



## Summary

The genome integrity is constantly challenged by sources of DNA damage that can lead to mutagenesis, rearrangements, cell death or disease. Genomic instability is a main characteristic of cancer and most cancers harbor structural rearrangements of the genome. Erroneous DNA repair can cause the formation of chromosome translocations, which often harbor severe mutagenic potential and are linked to cancer morbidity.

Therapies of cancers most often include the treatment with DNA damage inducing chemotherapeutics. Many successful chemotherapeutics target topoisomerases type II (TOP2) and lead to the accumulation of covalent TOP2-DNA complexes. Topoisomerases are essential enzymes that relieve torsional stress within genomic DNA, but upon trapping by chemotherapeutics, TOP2-DNA complexes lead to the formation of toxic DNA double-strand breaks (DSBs) to kill cancer cells. However, TOP2 poisons are frequently associated with the occurrence of therapy-related leukemia, often driven by recurrent translocations of the *mixed lineage leukemia (MLL)* gene. The molecular events that lead to the conversion of trapped topoisomerases to DSBs and to the formation of *MLL* translocations remain elusive.

To study TOP2-induced gene fragility, we developed C-Fusion 3D, a new methodology to quantify chromosome breaks and rearrangements in single cells. Upon combination with genome-wide DSB profiling, we find that abortive TOP2 complexes generate recurrent DSB hotspots at transcriptionally active regions, they often coincide with chromatin looping boundaries and positively associate with transcription output and directionality. *MLL* and translocation partner genes are highly transcribed, enriched at chromatin loop anchors and form fusions at frequencies ranging from  $10^{-3}$  to  $10^{-4}$  that depend on both TOP2 isoforms. It is the interplay of transcription with loop anchor sites that contributes to gene fragility and the formation of *MLL* fusions. A multilayer network of repair factors suppresses the formation of oncogenic fusions by mediating DSB repair that include activities of TDP2, MRE11 and non-homologous end joining. Screening for novel factors that contribute to TOP2-induced genomic instability reveals a role of VCP in DSB formation, cytotoxicity and *MLL* fragility.

We propose that transcription and spatial chromatin folding jointly pose a threat to genome stability by contributing to chromosome breakage and the formation of oncogenic, therapy-related *MLL* translocations.



## Preface

The research that is presented in this dissertation has been performed in the research group of Dr. Vassilis Roukos at the Institute of Molecular Biology (IMB) in Mainz. Additional contributions were made by the IMB Bioinformatics Core Facility, the group of Dr. Nicola Crosetto at the Karolinska Institute Stockholm, and the group of Prof. Dr. Argyris Papantonis, University of Göttingen. Parts of the work presented here were published in the research article:

Gothe HJ, Bouwman BAM, Gusmao EG, Piccinno R, Petrosino G, Sayols S, Drechsel O, Minneker V, Josipovic N, Mizi A, Nielsen CF, Wagner EM, Takeda S, Sasanuma H, Hudson DF, Kindler T, Baranello L, Papantonis A, Crosetto N and Roukos V (2019) Spatial chromosome folding and active transcription drive DNA fragility and formation of oncogenic *MLL* translocations. *Molecular Cell*, Jul 25;75(2):267-283.e12

The thesis was independently written by myself and all resources that have been used are listed in the Reference section. All experiments and analysis were performed by me with the following exceptions:

- Dr. Oliver Drechsel developed custom made scripts for the analysis of C-Fusion 3D, IMB Bioinformatics Core Facility, Mainz, Germany
- sBLISS experiments were done by Dr. Britta Annika Maria Bouwman under supervision of Dr. Nicola Crosetto, Karolinska Institute Stockholm, Sweden
- The detection of sBLISS hotspots and genome analysis were done by Dr. Giuseppe Petrosino and Dr. Sergi Sayols, IMB Bioinformatics Core Facility, Mainz, Germany
- Dr. Eduardo Gade Gusmao performed the analysis of CTCF motifs and chromatin looping under supervision of Prof. Dr. Argyris Papantonis, Department of Pathology, University Göttingen, Germany
- Natasa Josipovic and Dr. Athanasia Mizi performed and analyzed nascent RNA-seq experiments in TK6 cells under supervision of Prof. Dr. Argyris Papantonis.
- Generation of HCT116-TOP2A-mAID cells was contributed by Dr. Christian F. Nielsen and Dr. Damien F. Hudson from the Murdoch Childrens Research Institute, Melbourne, Australia
- TK6 repair deficient cell lines were generated and provided by Prof. Dr. Sunichi Takeda and Prof. Dr. Hiroyuki Sasanuma, Kyoto University, Japan





## List of publications

**Gothe HJ**, Bouwman BAM, Gusmao EG, Piccinno R, Petrosino G, Sayols S, Drechsel O, Minneker V, Josipovic N, Mizi A, Nielsen CF, Wagner EM, Takeda S, Sasanuma H, Hudson DF, Kindler T, Baranello L, Papantonis A, Crosetto N and Roukos V (2019) Spatial chromosome folding and active transcription drive DNA fragility and formation of oncogenic *MLL* translocations. *Molecular Cell*, Jul 25;75(2):267-283.e12

Bouwman BAM, Agostini F, Garnerone S, Petrosino G, **Gothe HJ**, Sayols S, Moor AE, Itzkovitz S, Bienko M, Roukos V & Crosetto N (2020) Genome-wide detection of DNA double-strand breaks by in-suspension BLISS. *Nature Protocols*, in revision

**Gothe HJ**, Minneker V and Roukos V (2018) Dynamics of double-strand breaks: Implications for the formation of chromosome translocations. *Advances in experimental medicine and biology*, 1044, 27–38.

Zirkel A, Nikolic M, Sofiadis K, Mallm J-P, Brackley CA, **Gothe H**, Drechsel O, Becker C, Altmüller J, Josipovic N, Georgomanolis T, Brant L, Franzen J, Koker M, Gusmao EG, Costa IG, Ullrich RT, Wagner W, Roukos V, Nürnberg P, Marenduzzo D, Rippe K and Papantonis A (2018) HMGB2 loss upon senescence entry disrupts genomic organization and induces CTCF clustering across cell types. *Molecular Cell*, 70:730–744.e6



## Chapter 1 Introduction

### 1.1 DNA damage and cancer

The integrity of cellular genetic material is constantly challenged by mutations of the primary hereditary information carrier: the genomic DNA. Spontaneous chemical reactions, enzymatic reactions and environmental influences provoke instability of DNA by the formation of lesions in the DNA structure (Hoeijmakers, 2009). To protect the genome from permanent damage resulting from mutations, cells have evolved elaborate mechanisms to cope with various kinds of DNA damage and to maintain their genomic integrity. Failure to repair DNA lesions can trigger programmed cellular senescence or apoptosis to avoid passing on of erroneous or damaged genetic material to descendant cells. However, the accumulation of damaged DNA has been associated with aging and with many genetic diseases, such as cancer (Ciccia and Elledge, 2010). Genomic instability is in fact one of the main features that is found in nearly all cancers and that promotes the accumulation of oncogenic driver mutations, which are required for malignant transformation (Hanahan and Weinberg, 2011).

Sources of DNA damage are multifaceted and include endogenous as well as exogenous sources. For example, essential cellular metabolic processes result in the formation of reactive chemical species that create alterations of the DNA and cause DNA breaks, or erroneous enzymatic reactions, such as errors or stalling of replication, can lead to mutations and DNA fragility. Exogenous DNA damage can be caused by irradiation, by chemical genotoxic agents, such as alkylating or crosslinking agents, or by inhibitors of factors involved in vital cellular processes such as replication or transcription (Ciccia and Elledge, 2010). More than 450 proteins have been identified to respond to the different kinds of DNA damage and the totality of responses triggered by DNA lesions is referred to as the DNA damage response (DDR) (O'Connor, 2015).

Most, if not all, cancers have lost parts of DDR pathways and are thereby impaired in their capacity to cope with some DNA lesions (Jackson and Bartek, 2009). The deregulation of proliferation leads to activation of the DDR in normal cells, which in turn limits cell cycle progression and can activate programmed cell death. It is believed that cancer cells require a partial loss of the DDR to thereby circumvent this proliferation barrier (Halazonetis et al., 2008). This ultimately leads to a higher dependency of cancers on the functional repair pathways that remain and can lead to the accumulation of genomic instability. One of the most well-known examples is the deficiency of tumor-suppressor genes *BRCA1* or *BRCA2*, that are

strongly associated with the development of breast and ovarian cancer. BRCA1 and BRCA2 are key players in the repair of DNA double-strand breaks (DSBs) by precise homologous recombination (HR) and their defects lead to greater dependency of DSB repair on more error-prone repair pathways, such as non-homologous end joining (NHEJ) (Prakash et al., 2015). Great efforts have been made to capitalize specific dependencies of cancers on repair pathways by exploiting synthetic lethality that arises upon combination of deficiencies (Lord et al., 2015). But the majority of cancer therapies today are general treatments, mainly radiotherapy and systemic chemotherapy, that function by inducing high levels of genomic instability to which cancer cells often respond sensitively. Radiotherapy using ionizing radiation induces various DNA lesions, including single-strand breaks (SSBs) and double-strand breaks (DSBs), and is proficient to cure many patients from their cancer (Lomax et al., 2013). Chemotherapies are more diverse and include DNA-damage inducing agents such as DNA cross-linking chemicals, alkylating agents or inhibitors of DNA topoisomerases that also cause the formation of SSB or DSBs. The most genotoxic type of lesions are DNA double-strand breaks (DSBs) that disrupt the integrity of DNA molecules and, if unrepaired, can lead to loss of genetic information during cell division. Erroneous repair of DSBs has furthermore been associated with the formation of single nucleotide variants (SNVs) and structural variations, such as inversions, deletions and chromosome translocations. Genomic instability and DNA damage are thus substantially linked to the induction and emergence of cancers, but also unfold treatment strategies for cancer therapies.

## 1.2 Genomic rearrangements as cancer drivers

The disease that is most frequently associated with somatic mutations is cancer. Genomic instability is a hallmark of human cancers and it mediates genetic changes scaling from individual nucleotides variations to the gain or loss of chromosomes leading to aneuploidy. Larger scale structural variations of chromosomes are frequently found in neoplasia and include deletions, insertions, inversions and translocations. Rearrangements can cause gain, loss or reshaping of genetic information, they vary in size and upon combination of several aberrations, they can lead to complex rearrangements (Mani and Chinnaiyan, 2010).

While the causality of chromosome aberrations for cancer development has already been postulated in the beginning of the 20<sup>th</sup> century by Theodor Boveri (Boveri, 1914), the first genetic abnormality that could be linked to a specific cancer was the Philadelphia chromosome that was described in 1960 and 1973 (Nowell and Hungerford, 1960; Rowley, 1973). This was the beginning of an era in which hundreds of new cancer-associated

structural variations were detected, driven by advances in cytogenetic analysis tools. In the 1980s, the first gene fusions resulting from translocations, which are commonly found in Burkitt lymphoma, were characterized. The oncogene *c-myc* on chromosome 8 was fused with parts of an immunoglobulin gene (*IGH*, *IGK*, *IGL*), thereby leading to unregulated expression of *c-myc* (Croce and Nowell, 1985; Leder et al., 1983). From then on, genomic rearrangements were considered important drivers of carcinogenesis.

The comprehension of structural variations in cancers has significantly improved by the advancement of next-generation sequencing (NGS) techniques and the collections of patient's whole-genome sequences (WGS) by consortia like The International Cancer Genome Consortium (ICGC) or The Cancer Genome Atlas (TCGA) (Yi and Ju, 2018). The most common SVs in cancers are deletions, followed by duplication events and chromosome translocations. Importantly, the prevalence and types of SVs differ considerably between cancer types and between patients, indicating selection mechanisms of rearrangements and showing the impact that genome rearrangements can have on cellular transformation (Li et al., 2020). This was underpinned by the observation that rearrangements resulting in gene fusion and mutations of these genes are often mutually exclusive, indicating the importance that SVs have on cancer emergence (Gao et al., 2018). Furthermore, certain SVs can be associated with specific genomic properties, for example deletions are prominently found in late-replicating regions whereas translocations commonly involve early-replicating regions in cancer patients (Li et al., 2020).

Among the most severe rearrangements are chromosome translocations that form by erroneous fusion of the arms of two different chromosomes. They can be balanced, e.g. they maintain all genetic information but in a different organization, or they might lead to loss of genomic DNA. The translocation of chromosomes can lead to outcomes with oncogenic potential in different ways. The disruption of genes or regulatory units can cause unbalancing of proliferation or differentiation. When breakpoints lie within genes, this leads to the formation of fusion genes that are comprised of parts of the two distinct genes (Roukos and Misteli, 2014). The most prominent example for a fusion gene resulting in an oncogenic chimeric protein is the *BCR-ABL1* fusion. It originates in a reciprocal translocation between chromosomes 9 and 22 (also known as Philadelphia chromosome) and is found in virtually all chronic myeloid leukemia (CML) patients. The gene fusion creates a constitutively activated tyrosine kinase *BCR-ABL1*, which deregulates oncogenic pathways such as the Ras/MAPK/ERK pathway and thereby enhances uncontrolled proliferation (Mitelman et al., 2007). In general, gene fusions drive approximately 16 % of cancers in combination with other mutations and were identified in more than 1 % of cancer cases as the sole driver (Gao

et al., 2018). To date, more than 32'000 unique gene fusions involving more than 14'000 distinct genes have been identified (Mitelman, Johansson and Mertens, 2020). The majority of these fusions are stochastic events that were only identified in a single patient. Fusions are a general feature of tumors and have been found in all tumor types, with most fusions detected in breast cancers and respiratory cancers (Mertens et al., 2015). However, certain gene fusions are recurrent and are associated with specific malignancies, such as *BCR-ABL1* with CML. Most recurrent gene fusions could be identified in prostate cancers and hematological disorders, where they mainly account for cases of acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) (Mertens et al., 2015; Mitelman et al., 2007). Overall, gene fusions that are found in patients show a bias towards overexpression of oncogenes and a downregulation of tumor suppressors, pointing out a widespread contribution that fusions can have on transforming gene expression changes and their potential role in oncogenesis (Gao et al., 2018). Thus, although genomic rearrangements are inherently rare events, their emergence is often linked to changes of cellular fitness and the onset of cancer. Diverse molecular mechanisms can contribute to the biogenesis of SVs and a deeper understanding of genomic properties and molecular processes that contribute to the formation of rearrangements is needed.

### 1.3 Therapy-related carcinogenesis

Cancer is increasingly understood as a chronic genetic disease because the survival of patients steadily advances by better diagnostics and improved therapies, but surviving patients are also constantly facing the risk of new cancer diagnosis. More than 5 % of U.S. Americans (17 out of ~330 million) have a cancer history and over the last decades, the 5-year survival rates of cancer patients steadily increased to now over 65 % (Miller et al., 2019). This has increased the awareness of secondary cancers, which are malignancies that are distinct from a patient's primary cancer, i.e. they do not originate from metastasis of a first neoplasm (Allan and Travis, 2005). Overall, ~18 % of cancer incidents are secondary cancers of which some have been identified to arise from side effects of cancer treatments (Travis et al., 2013). Secondary, therapy-related cancers are associated with specific treatments that patients received, such as radiotherapy or chemotherapy, which are moderately carcinogenic themselves. Clinicians and researchers thus became increasingly aware of the consequences that specific cancer therapies can have on patients.

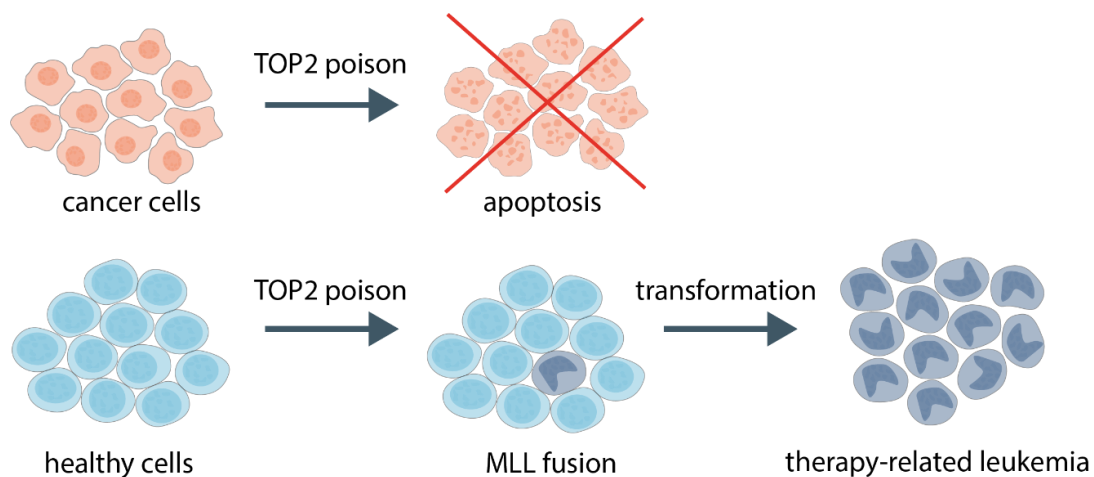
High doses of radiation such as used in radiotherapy have been linked to a significantly higher risk of developing second solid neoplasms or leukemias. Radiation therapy is used for various

types of neoplasms in approximately 50 % of cancer patients and several relationships between radiotherapy of primary cancers and the occurrence of second neoplasms were made (Kumar, 2012). For example, treatments of breast cancers by irradiation has been linked to a higher risk of developing a secondary contralateral breast, lung or esophagus cancers (Travis et al., 2012, 2013). The irradiation dose that primary Hodgkin lymphoma patients received was found to directly correlate to an elevated relative risk of developing secondary lung cancer. Radiotherapy-induced solid cancers are associated with a long latency period of more than 10 years, whereas leukemias usually develop within few years after treatments (Allan and Travis, 2005).

The most frequent therapy-related cancers are myeloid neoplasms, such as myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). They have been linked to treatments with chemotherapeutics, especially to alkylating agents and topoisomerase type II (TOP2) targeting drugs, generally occur with short latencies upon treatment, and patients exhibit poor prognosis. Around 70 % of therapy-induced myeloid neoplasms are MDS that often develops into AML and that are characterized by large genomic deletions resulting from the loss of parts of chromosome 5 or chromosome 7 (McNerney et al., 2017). These rearrangements are mostly seen after treatments with alkylating agents and sometimes after radiotherapy (Bhatia, 2013). Alternatively, therapy-related AMLs (t-AMLs) characterized by balanced chromosome translocations are seen in ~30 % of second myeloid leukemias (McNerney et al., 2017). The occurrence of translocation is directly linked to chemotherapeutic targeting of TOP2s and most often involves translocations of the *mixed lineage leukemia* gene (*MLL* or *KMT2A*) on chromosome 11 (Cowell and Austin, 2012) (Figure 1.1). Fusion proteins that arise from *MLL* translocation are often the driving mutation of secondary leukemias and few additional mutations are required to cause the onset of a second cancer (Figure 1.1) (Allan and Travis, 2005). Estimates for the overall risk of developing a therapy-related leukemia differ substantially but converge between 0.2 % to 10 % of cumulative risk (Bhatia, 2013; Ezoe, 2012; Gole and Wiesmüller, 2015; McNerney et al., 2017). Higher precision of therapies have certainly reduced the risk for secondary neoplasms within the last decades, but the treatment of a primary cancer is still a significant extrinsic driver of secondary neoplasms. It is thus important to monitor the risk that certain therapies pose and to understand their underlying mechanisms.

The link of chemotherapies to specific genomic rearrangements, such as deletions and translocations, is especially intriguing and calls for a molecular understanding of their biogenesis. The targeting of topoisomerases of type II by chemotherapeutics has directly been linked to the formation of oncogenic *MLL* translocations and the development of therapy-

related acute leukemias (Figure 1.1). Different classes of TOP2-targeting drugs are therapeutically used that either inhibit topoisomerase functions or cause the accumulation of TOP2s on genomic DNA, which are so-called topoisomerase poisons. These chemotherapeutics exploit the essential functions that topoisomerases have to protect the genome from topological stress.



**Figure 1.1 Chemotherapeutic TOP2 poisons can lead to the formation of therapy-related leukemia.** Chemotherapies that are targeting topoisomerases of type II, also known as TOP2 poisons, are highly potent therapeutics that lead to DNA damage and apoptosis of cancer cells. However, they have been associated with the formation of *MLL* translocations that can drive therapy-related leukemias, such as acute myeloid leukemia (AML).

## 1.4 Topoisomerases are a source of genomic instability

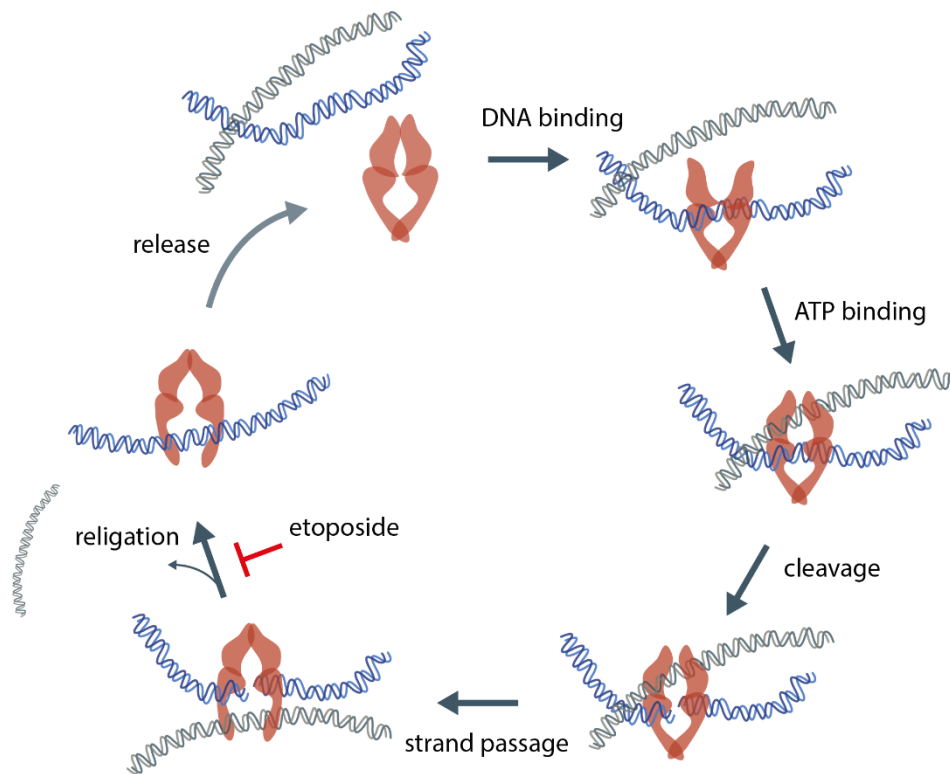
### 1.4.1 Topoisomerases relieve topological stress but are inherently risky

Genomic DNA in eukaryotes is structured into chromosomes in which the helical DNA duplex is wrapped around histone proteins and compacted at multiple layers. Molecular processes traversing genomic DNA, such as transcription, replication or chromatin structuring processes, lead to topological strain of the DNA helix. Overwound or underwound DNA is referred to as positive or negative supercoiling, respectively, and can lead to changes of DNA topology and blocking of DNA associated processes. Further topological obstacle are created by knotting of DNA, formation of catenated DNA or non-canonical DNA structures (Deweese et al., 2009). Human cells have evolved a set of six topoisomerase enzymes grouped into three subfamilies that resolve DNA supercoiling as well as entanglements and knots. Topoisomerases of type IA (TOP3A, TOP3B) and IB (TOP1, TOP1mt) resolve torsional stress by creating single-strand DNA breaks, and they allow the controlled rotation of the DNA



(TOP1 and TOP1mt) or passage of the other single DNA strand (TOP3A and TOP3B) (Cuya et al., 2017; Pommier et al., 2016).

Topoisomerases of type II, TOP2A and TOP2B, relieve positive and negative DNA supercoils as well as catenated and knotted DNA by creating a transient DNA double-strand break (DSB) and by forming a gate to pass through another DNA duplex (Pommier et al., 2016). The two isoforms TOP2A and TOP2B are encoded by individual genes, but they share around ~70 % amino acid homology and possess very similar enzymatic characteristics (Austin et al., 1993; Tan et al., 1992). Both TOP2s bind to strained DNA as homodimers, they create two opposing DNA nicks that are four bases apart and thereby form a double-strand break (Figure 1.2). Importantly, to stabilize the DNA backbone, TOP2s covalently link their active site tyrosine to the newly formed 5' end of the DNA duplex by a transesterification reaction. The stabilized homodimer forms a DNA gate to pass through another DNA segment. To complete the catalytic cycle, TOP2s religate the two DNA ends and release themselves from the DNA (Figure 1.2) (Deweese and Osheroff, 2009). While TOP2s are covalently bound to the 5'-DNA end, they are referred to as TOP2 cleavage complexes (TOP2ccs). TOP2ccs ensure the integrity of the DNA during the enzymatic reaction that, by the formation of DSBs, otherwise poses a threat to the genome integrity of cells. On the other hand, the formation of covalent TOP2-DNA complexes is also not without risk, as upon failure of resealing the DNA ends, stable protein-linked DNA breaks (PDBs) are formed (Ashour et al., 2015). TOP2ccs are usually short lived and only ~0.5 – 1 % of active TOP2s are present as cleavage complexes (Deweese and Osheroff, 2009). However, they can spontaneously be stabilized by nearby DNA lesions or natural products, and TOP2ccs are intentionally trapped by antineoplastic TOP2 poisons. Repair mechanisms for PDBs are evolutionally highly conserved and suggest that although topoisomerases are essential for cells to proliferate, they come with the inherent risk of DNA break formation (Ashour et al., 2015; Vos et al., 2011). Persistent TOP2-DNA complexes are eventually transformed into free DSBs that are no longer tethered by the TOP2 homodimer. Mechanisms of the transformation are not deeply understood but it is believed that collisions with replication forks or transcription machineries trigger a proteasome-dependent degradation of trapped TOP2s, whereby highly genotoxic DNA ends are formed.



**Figure 1.2 DNA topoisomerase type II catalytic cycle.** TOP2s relieve topological stress and DNA catenanes by binding to a DNA duplex strand, making a transient DSB that forms a gate to pass through a second DNA molecule and religate the DNA break. TOP2 poisons disrupt steps of the cycle that trap TOP2s on DNA while forming a covalent bond. Etoposide prevents the religation of the DSB thereby causing the accumulation of TOP2 cleavage complexes. Figure adapted from (Nitiss, 2009).

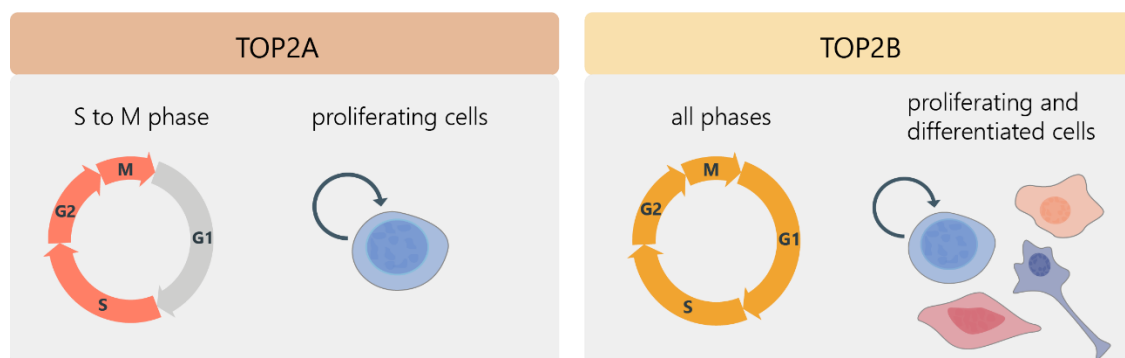
#### 1.4.2 Isoform specific functions of TOP2A and TOP2B

Topoisomerases are evolutionary conserved and are highly important for cellular fitness because most fundamental nuclear processes, such as transcription, replication, chromosome organization, condensation and segregation, generate changes of DNA topology (Forterre and Gadelle, 2009). It is believed that two vertebrate isoforms TOP2A and TOP2B originate from a common ancestor gene through a gene duplication event and have thereafter evolved different cellular expression patterns and specific cellular functions whereby they cannot substitute one another (Austin et al., 1993; Lang et al., 1998; Pendleton et al., 2014; Sng et al., 1999; Tan et al., 1992).

The activity of TOP2A is essential for the survival of proliferating cells, its absence is embryonic lethal, but it is not expressed in quiescent cells and fully differentiated tissues. TOP2A expression is cell cycle regulated with low protein levels in G1 and highest protein levels in G2/M-phase (Figure 1.3). TOP2A associates with replication forks and is implicated in resolving supercoiling and catenanes that arise during replication (Dykhuizen et al., 2013;

Pendleton et al., 2014). Catenation of chromatids in mitosis is primarily released by TOP2A to allow the segregation of chromatids (Pommier et al., 2016). Another function only attributed to TOP2A is its role in chromosome condensation and compaction during mitosis where loss of TOP2A leads to loss of hypercompaction of M-phase chromosomes, that can only partially be rescued by TOP2B (Farr et al., 2014; Shintomi et al., 2015). TOP2A is thus the isoform that was mainly associated with functions during replication and mitosis.

TOP2B, on the other hand, is expressed in all cell types and is not cell cycle regulated (Figure 1.3). TOP2B expression levels were linked to the status of cell differentiation with higher protein levels in fully differentiated mouse neurons than in embryonic stem cells (Tiwari et al., 2012). It is not required for cellular survival but is indispensable for neural development and its loss in mice models has been linked to death at birth (Pommier et al., 2016). Specific TOP2B only functions have mainly been linked to regulation of transcription during neurogenesis and of developmental genes (Ju et al., 2006; Lyu et al., 2006). Recruitment of TOP2B to gene promoters was found to be dependent on gene activity and was highest for highly active genes (Kouzine, Gupta, et al., 2013). Gene topology regulation resulting from transcription were thus mainly attributed to TOP2B activity. However, in pluripotent stem cells, recruitment of TOP2A to active genes and gene regulation by TOP2A was seen in a similar way as reported for TOP2B, indication that both TOP2 isoforms can play a role in the regulation of transcription (Thakurela et al., 2013). It is also the TOP2B isoform that has been associate with chromatin loop architecture points, whereas a potential role of TOP2A in managing DNA topology at chromatin architecture boundaries remains to be elucidated (Canela et al., 2017, 2019; Uusküla-Reimand et al., 2016).



**Figure 1.3 TOP2A and TOP2B isoforms expression patterns.** The two topoisomerase type II isoforms TOP2A and TOP2B are expressed in different cell types and cell cycle phases. TOP2A is mainly expressed in S, G2 and M phase of the cell cycle and in proliferating cells. TOP2B on the other hand is found in all cell types and throughout the cell cycle.

### 1.4.3 Roles of TOP2 isoforms in ETO-induced toxicity and DNA fragility

The various chemotherapeutics that poison TOP2s are associated with different TOP2 isoforms specificities. The TOP2 poison mitoxantrone for example is believed to predominantly target TOP2B whereas etoposide has been shown to efficiently trap both TOP2 isoforms (Austin et al., 2018; Lee et al., 2016). Various studies have however implicated the cytotoxicity of ETO to be mediated by poisoning of the TOP2A isoform whereas TOP2B was held responsible for the formation of DNA breaks and therapy-induced rearrangements (Cowell and Austin, 2012; Pendleton et al., 2014).

The formation of ETO-induced melanomas and DSBs was found to predominantly rely on TOP2B in a mouse skin cancer model with a skin-specific TOP2B knockout. No reduction in cytotoxicity could be observed upon TOP2B depletion, leading to the conclusion that TOP2A predominantly mediates ETO-induced cytotoxicity (Azarova et al., 2007). The view that TOP2A mediates cytotoxicity upon TOP2 poisoning was strengthened by the observation that the depletion of TOP2A by RNAi lead to doxorubicin (also a TOP2 poison) resistance in lymphoma cells (Burgess et al., 2008). In human B precursor NALM-6 TOP2B<sup>-/-</sup> cells, strongly reduced *MLL* breakage levels were observed, arguing that TOP2B is the main isoform leading to genomic instability but not to cytotoxicity (Cowell et al., 2012). Genome-wide dependencies of TOP2-lesions on TOP2B were seen in mouse embryonic fibroblasts (MEFs) extracted from wild-type or TOP2B<sup>-/-</sup> knockout mice, which showed reduced levels of TOP2-associated DSBs detected by DNA end profiling of the knockout MEF population (Canela et al., 2017). Nonetheless, trapped TOP2A cleavage complexes have also been detected within frequently translocating genome loci (Yu et al., 2017). Furthermore, induction of DNA damage signaling by etoposide was reported to mainly depend on TOP2A (de Campos-Nebel et al., 2010). The question of isoform specific roles mediating cell death and genomic instability is thus not clear yet and remains under investigation.

### 1.4.4 Chemotherapies targeting TOP2s are associated with secondary leukemia and oncogenic *MLL* translocations

The cytotoxic nature of trapped TOP2ccs is exploited in chemotherapy by trapping TOP2 homodimer as a covalent PDB, thereby inhibiting the religation of the DNA strands (Vos et al., 2011). These compounds are referred to as TOP2 poisons as they transform an essential enzymatic intermediate into a genotoxic DNA lesion. Etoposide was approved as a chemotherapeutic in the 1980s and has since been one of the most prescribed anticancer drugs in the world (Deweese and Osherooff, 2009). Various different systemic and solid malignancies are treated with etoposide, such as breast cancer, lung cancer, neuroblastomas

or leukemias (Pendleton et al., 2014). However, treatment with TOP2 poisons is also linked to the formation of secondary, therapy-related leukemias, mostly acute myeloid leukemia (t-AML) and at times acute lymphoblastic leukemia (t-ALL). An estimated 2 – 3 % of patients that receive TOP2 poisoning chemotherapy have been reported to develop secondary AML, which are commonly characterized by *MLL* translocations (Deweese and Osheroff, 2009).

*MLL* translocations are a general hallmark of acute leukemia and account for approx. 10 % of all leukemias. However, incidence rates vary among leukemic patient groups where *MLL* translocations are especially frequent in infant leukemia (70 % of infant ALL and up to 50 % of infant AML cases) and therapy-related leukemia (at least 70 % of cases after TOP2-targeting therapies) (Muntean and Hess, 2012; Winters and Bernt, 2017). In leukemic patients, more than 90 different genes that fuse with *MLL* were characterized (also referred to as *MLL* recombinome), showing a promiscuity of *MLL* fusions. Nevertheless, six specific fusion partner genes account for more than 85 % of all *MLL* translocations, namely *AF4*, *AF9*, *ENL*, *AF10*, *ELL*, *AF6* (Meyer et al., 2018). In therapy-induced leukemia, the most prominent fusions of *MLL* are with *AF9*, *AF4*, *ENL* and *ELL*, of which *MLL-AF4* fusions are characteristic for t-ALL cases whereas the other fusions are predominantly found in t-AML patients (Meyer et al., 2018).

Chromosomal breakpoints that result in fusion events of *MLL* were identified to lie in a region between exon 8 and exon 12 that is referred to as breakpoint cluster region (BCR) and spans a total of ~8 kb (Wright and Vaughan, 2014). In childhood and adult *de novo* leukemic patients, *MLL* breakpoints map to the whole BCR whereas for infant and therapy-induced leukemia, breakpoints cluster near exon 12 (Cowell and Austin, 2012; Zhang and Rowley, 2006). In fact, infant leukemia has also been linked to high consumption of naturally occurring TOP2 poisons such as genistein and some polyphenols (Deweese and Osheroff, 2009). An important question has thus been why the breakpoint within *MLL* is different between *de novo*, infant and therapy-related acute leukemias and if this bias is directly linked to TOP2 functions and isoforms. The BCR of *MLL* is characterized by an increased chromatin accessibility that was found by DNase I hypersensitivity testing (DNaseI-HS) (Strissel et al., 1998). A direct involvement of TOP2B in *MLL* breakage upon TOP2 poisoning has been indicated in human TOP2B-null pre-B leukemic cells, which show reduced *MLL* breakage compared to wild-type cells when analyzed by break-apart FISH. Furthermore, transcription was proposed to mediating proximity of *MLL* with potential fusion partner genes to promote illegitimate fusion events (Cowell et al., 2012). A more direct role of transcription in induction of *MLL* fragility has also been suspected, but remains under debate (Canela et al., 2017; Gómez-Herreros et al., 2017). To understand mechanisms of therapy-induced

rearrangements, it is vital to uncover molecular mechanisms that contribute to gene fragility upon TOP2 poisoning and specifically to *MLL* breakage that can result in oncogenic fusions.

### 1.5 Mechanisms of TOP2cc conversion to DNA breaks

#### 1.5.1 Etoposide induces genomic instability by conversion of TOP2ccs into DNA breaks

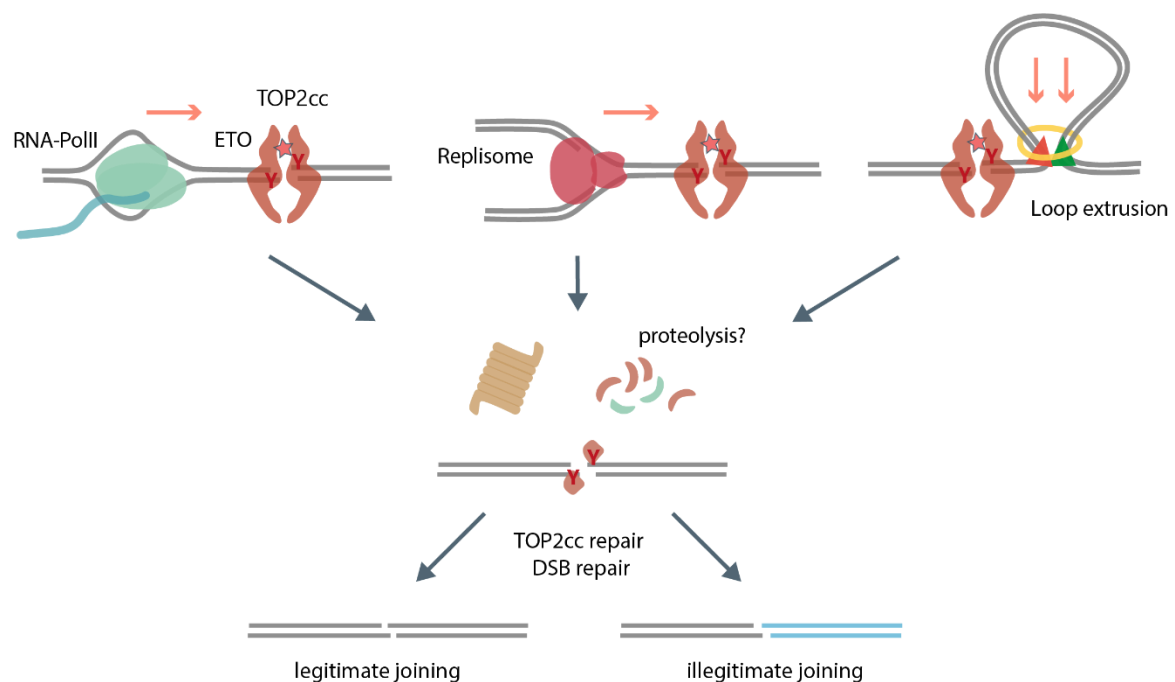
Trapping of TOP2s in chemotherapy has proven as an effective strategy to induce genotoxic stress and apoptosis of highly proliferating cells. The accumulation of TOP2-associated PDBs causes genomic instability by conversion into open, potentially deleterious DNA breaks or they trigger apoptosis (Ashour et al., 2015; Pendleton et al., 2014). The molecular mechanisms that contribute to the conversion of TOP2ccs into DNA breaks are, however, not fully understood. It is also important to note that stalled TOP2ccs by themselves would not directly cause genomic rearrangements, because the DNA ends are concealed by the TOP2 homodimers, which are bound to the 5' ends of the DNA (Deweese and Osheroff, 2009). The association of TOP2 homodimers is very strong and shields the DNA ends from illegitimate fusion events. The conversion of TOP2-bound DNA ends to protein-free breaks results in the formation of DSBs that are a prerequisite for chromosome translocations to occur. The underlying mechanisms of TOP2cc conversion have been a question of research for more than 20 years and several molecular mechanisms have been identified to contribute to DSB formation (D'Arpa, 1994; Yan et al., 2016).

#### 1.5.2 Transcription and replication contribute to TOP2-dependent DSB formation

Early on, proteasome-dependent degradation of TOP2s upon trapping of cleavage complexes by teniposide (also known as VM-26, an etoposide analogue) was reported for different human cell lines (leukemic, lung fibroblast and breast cancer cell lines), which was preferentially seen for TOP2B over TOP2A. Importantly, transcription inhibition but not replication inhibition could alleviate the degradation of TOP2 proteins upon teniposide treatment, which was reported to not be due to decrease TOP2cc formation (Mao et al., 2001). TOP2 protein degradation could also be seen upon inhibition of TOP2's ATP hydrolysis function by ICRF-193, a catalytic inhibitor that also stabilizes TOP2 non-covalently on DNA, showing that the degradation was not induced by unrepaired DSBs but by accumulation of TOP2 complexes on DNA (Xiao et al., 2003). More recent findings directly implicated transcription in the induction of DSBs upon TOP2 poisoning, leading to the current model of

TOP2ccs being converted into protein-free DSBs by collisions with the transcription machinery, followed by proteasomal degradation of the TOP2cc (Figure 1.4) (Gómez-Herreros et al., 2017; Yan et al., 2016).

In support of a role of transcription in TOP2cc conversion, mapping of TOP2A cleavage sites in human leukemic K562 cells revealed clustering of TOP2 activity in transcriptionally active genomic regions, open chromatin sites and enhancers (Yu et al., 2017). Furthermore, the dissipation of supercoiling at transcription start sites of highly expressed genes was reported to be TOP2B dependent (Kouzine et al. 2013). Similarly, in mouse liver cells, TOP2B binding was found in active genomic regions, specifically at promoters and was dependent on gene activity (Uusküla-Reimand et al., 2016). Genome-wide mapping of DSBs upon TOP2 trapping by Breaks Labeling, Enrichment on Streptavidin, and Sequencing (BLESS) suggested an accumulation of endogenous and ETO-induced DSBs at promoters (Yang et al., 2015).



**Figure 1.4 Molecular mechanisms that contribute to the formation of DSB induction upon TOP2 poisoning.** Upon stabilization of TOP2ccs by etoposide, DSBs are formed that have been linked to activities of transcription, replication or loop extrusion. Encounters of RNA-PolIII with TOP2ccs triggers proteasomal degradation of TOP2ccs, replication activities were linked to ETO-induced cytotoxicity and DNA damage signaling. Loop extrusion has been proposed to induce the activity of TOP2s at loop boundaries and to contribute to ETO-induced genomic instability. The conversion of TOP2ccs to protein-free DSBs relies on proteolysis or remodeling of TOP2s to induce TOP2cc repair and joining of DNA ends by DSB repair mechanisms.

DNA traversing by replication has been hypothesized to mediate cytotoxicity by induction of apoptosis by p53 phosphorylation rather than DSB formation itself upon ETO-treatment (Fan et al., 2008). On the other hand, replication protein A (RPA) foci, that likely represent 3' ssDNA

that results from resection of DSBs, are induced by etoposide mainly in S phase but cannot be seen in G1 phase cells (Yan et al., 2016). This argues for a role of replication in DSB formation, showing that the contribution of replication to TOP2-dependent DSB formation remains to be clarified (Figure 1.4).

### 1.5.3 Genome architecture as a new player in TOP2-induced genomic instability

The development of genome-wide TOP2 and DSB mapping techniques have extended the models of transcription and/or replication-dependent TOP2cc conversion and put forward a third potential mechanisms that relies on chromatin architecture activities.

Chromatin is organized within the nucleus into a multilevel 3D architecture that ranges from chromatin loops (enhancer – promoter interactions, architectural loops, polycomb-mediated loops) to topologically associated domains (TADs) to compartments (A and B) and chromosome territories (Bonev and Cavalli, 2016; Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014). Two factors stand out to have highly significant functions among many proteins that can modulate 3D genome organization in mammalian cells. CCCTC-binding factor (CTCF) is an essential protein that has tens of thousands of genomic binding sites and has been implicated in gene activity modulation. Chromatin loop anchors typically contain pairs of CTCF motifs of convergent orientation and are bound by CTCF and cohesin (Rao et al., 2014; de Wit et al., 2015). The loss of CTCF leads to loss of CTCF mediated interactions but does not disturb higher order compartmentalization (Kubo et al., 2017; Nora et al., 2017). It is described as a key regulator of chromatin loops by blocking extruding loops and thereby determining boundaries. The main protein complex to actively form chromatin loops is believed to be cohesin. Cohesin was first described to be essential for sister chromatid cohesion but has increasingly been recognized to have a dual function to also topologically entrap chromatin into loops (Bonev and Cavalli, 2016; Rowley and Corces, 2018). According to the prevalent loop extrusion model, cohesin complexes extrude chromatin until they are stopped by CTCF motifs bound by CTCF in a specific orientation (Davidson et al., 2019; Fudenberg et al., 2016; Kim et al., 2019; Sanborn et al., 2015). Cohesin is thereby found enriched at the 3' end of CTCF motifs and can be released by the cofactor WAPL (Rowley and Corces, 2018). Similar to loss of CTCF, depletion of the cohesin subunit RAD21 leads to the depletion of CTCF associated chromatin loops, though heterogeneous TAD-like structures are still present at the single-cell level (Bintu et al., 2018; Rao et al., 2017).

In mouse and human cells, TOP2B has been found enriched at CTCF and cohesin binding sites throughout the genome (Madabhushi et al., 2015; Manville et al., 2015; Uusküla-Reimand et al., 2016). Interactions between TOP2B with CTCF and cohesin complex proteins were



furthermore found in human HeLa cells, hinting at a role of TOP2s to regulate DNA topology at chromatin loop boundaries (Uusküla-Reimand et al., 2016). Mapping of TOP2-associated DSBs and the impact of genome organization on DSB formation was recently explored by the Nussenzweig group, that applied genome wide profiling of DNA ends by END-seq in mouse and human cells to map TOP2-bound DSBs upon ETO-treatment across the genome (Canela et al., 2017). ETO-induced DSBs were found at chromatin loop anchor sites and were largely transcription and replication-independent. However, due to degradation of TOP2s from DNA-ends by ExoVII treatments and very short ETO-treatments that do not facilitate the conversion of TOP2ccs into protein-free DSBs, rather TOP2ccs than free DSBs were mapped (Canela et al., 2017, 2019). Nevertheless, TOP2B-associated DSBs colocalized with around 50 % of CTCF and RAD21 binding sites, were enriched upon ETO-treatment and associated with loop anchor strength. Since TOP2Bccs were mostly found at the outside border of chromatin loops, the authors suggested that topological stress that is formed during loop extrusion required the activity of TOP2s and thereby pose loop anchors at risk of increases DSB formation (Figure 1.4) (Canela et al., 2017).

Various cellular systems and methods have been used to study mechanisms of TOP2cc conversion into DSBs, whereby a systematic understanding of the relative contributions of transcription, replication and genome architecture remain unclear. Furthermore, downstream repair processes of TOP2-DNA complexes and DSBs also contribute to susceptibility of cells to form illegitimate, oncogenic fusions.

## 1.6 Repair of TOP2-DNA covalent complexes

### 1.6.1 Repair of protein-bound DNA breaks

Chemotherapeutic TOP2 poisons rely on the cytotoxic effect that the accumulation of covalent TOP2ccs has on cells. Spontaneous stabilization of TOP2ccs can also occur under physiological conditions and the repair machinery for protein-linked DNA breaks (PDBs) is highly conserved (Ashour et al., 2015). Apart from TOP2-DNA lesions, cells have to cope with various kinds of protein-DNA crosslinks, such as TOP1 cleavage complexes linked to single-stranded DNA breaks, or DNA methyltransferase 1 (DNMT1) which can be covalently linked to DNA by the cytosine analogue 5-aza-dC (Maslov et al., 2012). If not reversed, covalent protein-DNA can cause genomic instability by their conversion into DNA single- or double-strand breaks (SSBs, DSBs) or by blocking molecular machineries that traverse DNA, which ultimately can also result in DNA breakage. Various enzymes have been detected that have

evolved to reverse or repair PDBs, which either target the protein entity, DNA entity or the covalent bond between the two (Stingele et al., 2017). For TOP2-DNA lesions, factors targeting each of these components have been studied (Figure 1.5) (Ashour et al., 2015). Since the removal of TOP2s from DNA by repair of the protein-crosslink leaves behind a DSB, the clean DNA break is channeled towards DSB repair mechanisms such as non-homologous end joining (NHEJ) or homologous recombination (HR) (Gómez-Herrerros, 2019).

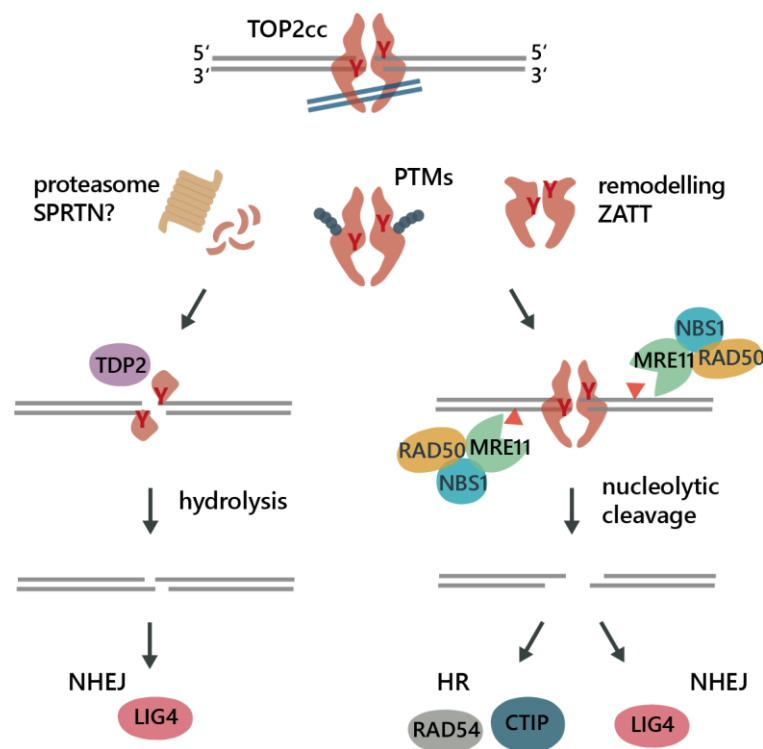
### 1.6.2 Reversal and removal of covalent TOP2-DNA complexes

Abortive TOP2-protein complexes are targeted by the 26S proteasome to be degraded and recycled (Mao et al., 2001; Zhang et al., 2006). However, the degradation is not complete and the proteasome cannot fully resolve the protein component of TOP2 PDBs. Instead, a peptide remnant, comprising the tyrosine that is linked to the 5'phosphate of the DNA backbone, is left behind which still presents a blocking DNA lesion. To facilitate DSB repair, the covalent bond can either be reversed by the Tyrosyl-DNA phosphodiesterase 2 (TDP2), or the DNA moiety linked to TOP2 or TOP2-remnants is nucleolytically removed by endonucleases such as Meiotic recombination 11 (MRE11, Figure 1.5) (Gómez-Herrerros, 2019).

The protease SPRTN has also been shown to facilitate replication-coupled repair of protein-DNA crosslinks (Lopez-Mosqueda et al., 2016; Stingele et al., 2016; Vaz et al., 2016). Proteolysis of protein-adducts by SPRTN can be initiated by stalling of the replication machinery, thereby causing SPRTN to degrade trapped DNA-linked proteins (Mórocz et al., 2017). This also leaves behind a peptide remnant, but it allows the passage of the replisome by translesion synthesis, or possibly recruits downstream repair factors to completely reverse the protein-DNA crosslink. SPRTN deficiency has been associated with etoposide hypersensitivity and accumulation of TOP2ccs (Lopez-Mosqueda et al., 2016; Vaz et al., 2016). Nonetheless, SPRTN has a wide variety of substrates and does not seem to specifically act on TOP2ccs but is probably a general first line of defense of cells to cope with protein-DNA complexes during replication (Stingele et al., 2017; Vaz et al., 2016).

The specific hydrolytic reversal of 5'-phosphotyrosine bond between TOP2 and the DNA end by TDP2 has been discovered by the group of Keith Caldecott (Ledesma et al., 2009). TDP is a phosphodiesterase that has high affinity to 5'-linked phosphotyrosine bonds (such as for TOP2ccs) but only weak activity at 3'-linked phosphotyrosine bonds (such as for TOP1ccs). TDP2 can however not resolve intact TOP2-DNA complexes because the TOP2 protein shields the covalent protein-DNA bond. Instead, activity of TDP2 is stimulated upon proteolytic degradation or remodeling of TOP2ccs (Gao et al., 2014). Modifications of TOP2ccs by the SUMO-ligase ZATT induces the removal of TOP2ccs by TDP2 in a proteasome-independent

pathway (Schellenberg et al., 2017). TDP2 produces clean DNA end by the removal of TOP2s, that are 5' phosphorylated and can directly be ligated (Ledesma et al., 2009). In fact, NHEJ was identified as the main pathway for the downstream DSB repair (Gómez-Herreros et al., 2013). By facilitating efficient PDB repair, TDP2 is thought to suppress transcription-dependent *MLL* breaks and chromosome rearrangements (Gómez-Herreros et al., 2017).



**Figure 1.5 Repair mechanisms of TOP2 cleavage complexes.** Stalled TOP2ccs are repaired by different repair mechanisms that are organized at several hierarchies. The covalently linked TOP2 protein can be proteolyzed by the proteasome or potentially by the SPRTN protease. TOP2ccs may be post-translationally modified (PTM) by ubiquitin or SUMO chains. Also remodeling of trapped TOP2 proteins by ZATT have been reported. The 5'-phosphotyrosine bond between the DNA end and TOP2 may be hydrolyzed by the phosphodiesterase TDP2, which requires accessibility to the Tyrosine (Y). Alternatively, nucleolytic cleavage by MRE11 can remove DNA strands that are linked to TOP2ccs. Downstream DSB repair is predominantly mediated by NHEJ but may also involve HR.

Endonucleolytic removal of abortive TOP2-DNA complexes by MRE11 constitutes a second mechanism that facilitates DSB repair by removal of the protein-linked DNA moiety (Hoa et al., 2016). Upon an initial endonucleolytic nick in a DNA strand, MRE11 resects the DNA in a 3' to 5' orientation, followed by a second endonucleolytic cut on the opposite strand (Stingele et al., 2017). The cleaned DNA ends are primarily joined by NHEJ, but they might also be repaired by HR (Figure 1.5) (Hoa et al., 2016). In the absence of the MRE11 endonuclease activity, TOP2ccs accumulate in cells treated with etoposide and increased levels of chromosome

aberrations can be observed. MRE11 has also more general roles in detecting and resecting DNA breaks. As part of the MRN complex (comprised of the nuclease MRE11, the ATPase RAD50 and regulatory factor NBS1), it is recruited to DSBs and activates the master regulatory kinase ataxia-telangiectasia mutated (ATM) (Syed and Tainer, 2018). It is furthermore involved in processing of DNA ends by resection to facilitate repair by HR and possibly microhomology-mediated end joining (MMEJ) (Paull, 2018).

### 1.6.3 Repair of TOP2-derived DSBs

Chromosome rearrangements are counteracted by faithful DNA break repair but are fueled by illegitimate joining of DNA ends. Downstream repair of processed TOP2-derived DNA ends can occur via HR, NHEJ or MMEJ pathways. The choice of HR for repair is limited to S/G2 cell cycle phases where homologs are available to allow error-free repair, whereas the end joining pathways are available throughout the cell cycle. In brief, during HR, recombination of sister chromatids is mediated to ensure accurate repair of DNA DSBs without loss of genetic information. Therefore, DNA ends are resected to form 3' ssDNA tails, they are bound by replication protein A (RPA) and form displacement loops by RAD51-mediated strand invasion. DNA polymerases then extend the invading DNA strand based on the copy of the sister chromatid and the break is finally resealed (Scully et al., 2019). Classical NHEJ is initiated by Ku70-Ku80 heterodimer binding to the DNA ends, which recruits NHEJ factors, such as DNA-dependent protein kinase catalytic subunit (DNA-PKcs). DNA ends can be processed to facilitate ligation, or are readily ligated by DNA ligase IV (LIG4) to connect the DNA ends. NHEJ is considered more error prone because end processing can lead to small deletions and DNA end affiliation is not controlled. HR-deficient cells that rely on end joining to repair DSBs generally exhibit higher levels of genomic rearrangements. In the absence of classical NHEJ, MMEJ involving ligase I or III can ensure joining of DNA ends independently of the availability of a homolog. MMEJ characteristically features short resection of DNA ends, alignment of internal microhomology stretches and is less processive than classical NHEJ (Lieber, 2010). The contributions of classical NHEJ and MMEJ to illegitimately fuse DNA ends and cause rearrangements are not yet clear and different dependencies were found in mouse and human cells (Biehs et al., 2017; Ghezraoui et al., 2014).

DSB repair mechanisms have also been studied in the context of TOP2-linked DSBs. Studies have shown that upon the removal of TOP2ccs from the DNA break by TDP2 or MRE11, DNA ends are predominantly channeled towards joining by NHEJ (Figure 1.5). Upon hydrolysis of 5'-phosphotyrosine bonds by TDP2, DNA ends have short cohesive overhangs that can be ligated in a preservative manner. Indeed, loss of KU70 has been shown to be epistatic over

TDP2 to cause etoposide hypersensitivity in avian cells, indicating that NHEJ acts downstream of TDP2 (Gómez-Herreros et al., 2013). DNA ends after MRE11 processing, however, exhibit resection of both ends that, upon direct ligation, are likely to cause the loss of short sequences. Loss of LIG4 or LIG4 and MRE11 in avian cells lead to similar levels of chromosome rearrangements, arguing that NHEJ might nevertheless be the main pathway to repair MRE11 derived DSBs (Hoa et al., 2016). Further evidence for the involvement of HR in TOP2-derived DSBs is scarce and requires additional research.

The multiple steps and complexity of pathways that are potentially involved in TOP2-linked DSB repair thus requires a systematic analysis to reveal the relevance of these pathways and the hierarchies that are potentially involved. On the one hand, methods to study the induction and kinetics of DSBs are required to estimate chromosome fragility upon ETO-treatments. On the other hand, the formation of oncogenic translocations are extremely rare events that require specific, sensitive methods to uncover mechanisms of oncogenic *MLL* fusion formation.

## 1.7 Methods to study genomic instability and the occurrence of oncogenic fusions

The formation of oncogenic fusions is a multistep process that requires the failure of cellular genome maintenance mechanisms at several stages. At the start, the formation of DNA DSBs is required within different chromosomes that are not repaired instantly and legitimately. Then, the pairing of different DNA ends has to be facilitated by physical proximity that favors fusion events. Finally, erroneous joining of DNA ends from the different chromosomes has to occur (Roukos, Burman, et al., 2013). Specific methodologies are required to mechanistically understand the steps and pathways in ETO-induced DNA fragility and *MLL* translocations formation.

### 1.7.1 Genome-wide profiling of DSBs

The formation of DSBs by ETO has been studied with various tools that either directly or indirectly detect DNA breaks. A widely accepted surrogate of DNA DSBs is the histone modification  $\gamma$ H2AX, which is the S139-phosphorylated form of the histone variant H2AX. Phosphorylation of H2AX is promptly and robustly induced in the vicinity of DSBs and it can easily be detected with specific antibodies (Kinner et al., 2008).  $\gamma$ H2AX is therefore used for the detection and quantification of DSBs within cells, for example by immunoblotting,

chromatin immunoprecipitation (ChIP) or immunofluorescence (IF) analysis by microscopy. Genome-wide analysis of  $\gamma$ H2AX by ChIP followed by sequencing (ChIP-seq) however revealed that large, up to megabase-sized  $\gamma$ H2AX domains are formed around endonuclease induced DSBs whereby the identification of genomic DSBs sites of unknown sequence is not possible (Iacovoni et al., 2010). The evaluation of repair markers to study DSB formation, such as  $\gamma$ H2AX levels, is thus a valuable tool to identify the overall response of cells to DNA breaks but remains indirect and might be susceptible to bias of unspecific events (Bouwman and Crosetto, 2018).

Since 2013, the development of techniques that directly detect DSBs within the genome by adaptor ligation to DNA ends has advanced (BLESS: Crosetto *et al.*, 2013; qDSB-Seq: Hoffman *et al.*, 2015; END-seq: Canela *et al.*, 2016; DSBCapture: Lensing *et al.*, 2016; BLISS: Yan *et al.*, 2017). The first method to directly capture and identify DSBs in cells was Breaks Labeling, Enrichment on Streptavidin, and Sequencing (BLESS) that has since then be refined into Breaks Labeling In Situ and Sequencing (BLISS) by the group of Nicola Crosetto (Crosetto et al., 2013; Yan et al., 2017). DSBs are captured in fixed cells by in situ blunting of DNA ends followed by dsDNA adapter ligation, which allow amplification of sequences by in-vitro transcription. The adapters furthermore contain sample barcodes and unique molecular identifiers (UMIs) that enable multiplexing and the quantification of DSBs upon sequencing (Yan et al., 2017). BLISS has been applied to identify endogenous and exogenous DSBs, for example for evaluation of on- and off-target effects of the RNA-guided Cas9 endonuclease (Bouwman and Crosetto, 2018). Another widely applied DSB profiling technique is END-seq, which has been developed by the group of André Nussenzweig (Canela et al., 2016). It avoids the fixation of cells by embedding cells in agarose plugs to modify DNA ends and ligate adapters. Sequences are thereafter amplified by PCR. END-seq has also be used for profiling of endonuclease induced DSBs (Canela et al., 2016). In addition, ETO-induced DSBs were profiled by END-seq in mouse cells and human cells. However, an exonucleolytic removal of TOP2 proteins from DNA ends was performed prior to adapter ligation whereby all DNA ends linked to TOP2ccs were mapped, resulting in profiling of TOP2 activity rather than free DNA DSBs (Canela et al., 2017, 2019). The removal of TOP2ccs prior to tagging of DNA ends has recently been omitted, allowing for the detection of real ETO-induced DSBs (Canela et al., 2019).

These DSB-mapping techniques now provide the possibility to obtain genome-wide profiles of DSBs upon induction of genomic instability by chemotherapeutics. They are ideal for the analysis of sites of DNA damage as well as for identifying genomic properties that could contribute to genome fragility. To date, they however provide limited information about

absolute numbers of DSBs per cell and are not able to detect DNA end joining events that lead to oncogenic fusions.

### 1.7.2 Methods to study rare chromosome translocations

To detect chromosome translocations and gene fusions, many sophisticated tools have been developed. The choice of method to detect translocations strongly depends on the research objective. When tumor samples are tested for fusions for diagnostic purposes, the penetrance of the fusion within the sample is usually high as most cancers have a clonal origin. Thus, low throughput and cost techniques that target specific fusions are often applied to identify cancer relevant fusions. On the other hand, unbiased fusion detections, as by deep-sequencing techniques, come with a higher cost but they provide information of genetic mutations at high resolution and can detect non-recurrent events in tumor samples. To study the biogenesis of chromosome translocations, techniques with high sensitivity and high throughput are required because the formation of translocations is extremely rare under physiological cellular conditions and translocations only gain higher penetration through clonal selection during tumorigenesis. Thus, methodologies with high sensitivity, broad specificity and ideally low cost are required to gain insights into molecular mechanisms contributing to the formation of oncogenic fusions.

### 1.7.3 Polymerase chain reaction to quantify specific chromosome translocations

Among the oldest techniques for translocation detection is the amplification of gene fusions by polymerase chain reaction (PCR). PCR is a simple yet highly specific tool to detect chromosome translocations on a molecular level. Primers are designed to lie within the two parts of the expected fusion junction, either within genomic DNA or within mRNAs to detect fusion transcripts (Figure 1.6). To estimate the frequency of fusion or translocation events, fusions can be detected by quantitative PCR methods. This has been done by the use of quantitative real-time PCR (RT-qPCR) or the serial dilution of input genomic DNA for PCRs (Brunet et al., 2009; Cohen et al., 2018; Piganeau et al., 2013). More precise quantifications of translocation formation in a cell population have been developed that use an absolute readout of qPCRs to mechanistically approach the biogenesis of translocations. One approach is to screen breakpoint junctions in a high throughput 96-well format based on SYBR green incorporation (Brunet et al., 2009). Alternatively the development of digital droplet PCRs allows for absolute quantification of fusion events in a cell population (Lund et al., 2016). For example, upon induction of DSBs within NPM1 and ALK using TAL endonucleases, the translocation frequency of NPM1 with ALK (t(2;5)(p23;q35)) was found at a levels of  $1 \cdot 10^{-3}$

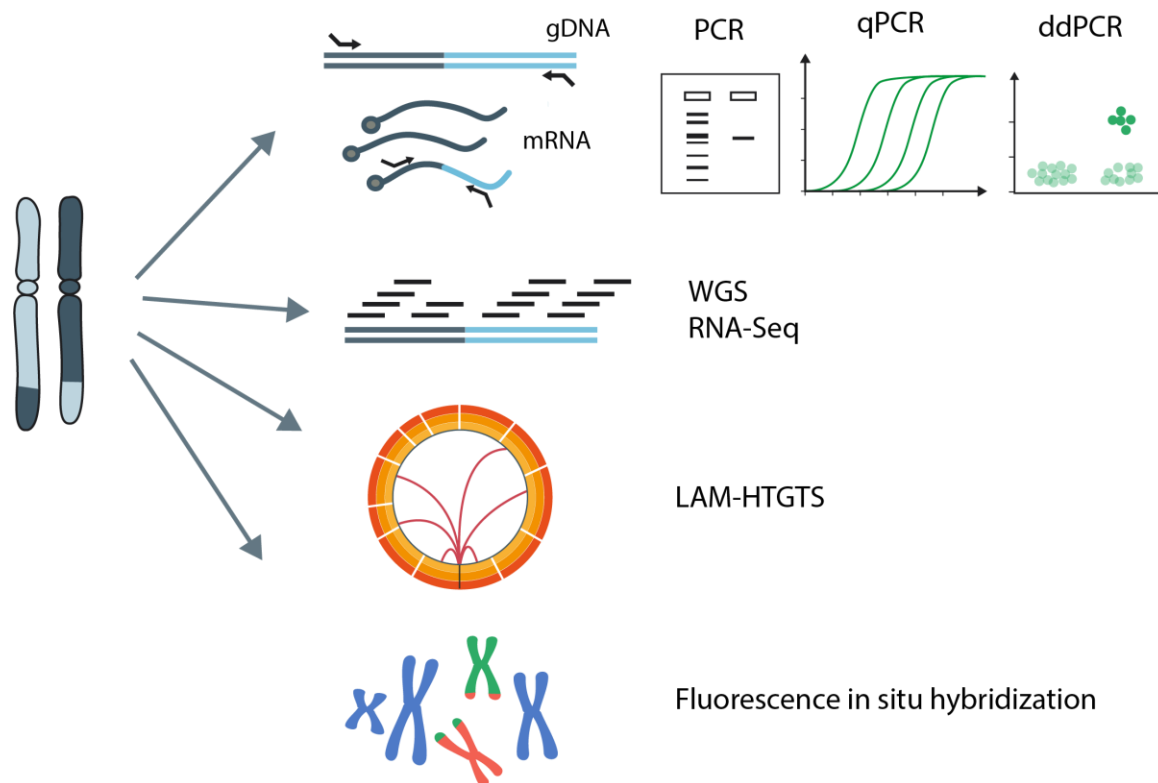
by 96-well fusion screening and by serial dilution PCRs of genomic DNA (Piganeau et al., 2013). PCR-based techniques are sensitive and precise but they require the exact knowledge of the fusion junction. For efficient amplification, primers need to be designed within a few hundred base pairs and the detection is influenced by resection at breakpoints and PCR artifacts. If one fusion site is known, this limitation can be overcome by a long-distance inverse PCR approach that detects fusion partners in an unbiased way. Inverse PCRs were used to detect many novel translocation partners of promiscuous fusion but they are not suitable for quantitative detections (Mertens et al., 2015).

### 1.7.4 Unbiased approaches by next-generation sequencing techniques

Another approach is the genome-wide detection of fusion partners based on linear amplification-mediated high-throughput genome-wide sequencing (LAM-HTGTS), that was developed by the groups of Frederick Alt and Richard Frock (Chiarle et al., 2011; Hu et al., 2016). PCR amplification of fusions with a *bait* genomic DNA site are combined with the detection of fusion partners by next generation sequencing (NGS). Upon induction of one specific *bait* DSB in mammalian cells, *prey* sequences that are fused with the bait are amplified, sequenced and identified throughout the genome (Figure 1.6). LAM-HTGTS has a very high sensitivity and can detect rare fusion but it is limited by the need of a specific *bait*. Furthermore, the formation of fusions is dependent on spatial proximity (Mani and Chinnaiyan, 2010), whereby the detection of inter-chromosomal fusions by LAM-HTGTS is much more pronounced than the detection of intra-chromosomal rearrangements (Wei et al., 2016). It is thus a semi-biased technique that can detect genome-wide fusion to a specific locus with high sensitivity.

Completely unbiased approaches where fusions are detected with neither one of the fusion partners known, became more prominent upon easier access to deep-sequencing techniques, such as whole genome sequencing (WGS) or RNA-Seq. For example, analysis of RNA-seq datasets from more than 4300 tumor samples collected by The Cancer Genome Atlas (TCGA) lead to the identification of nearly 8000 different fusion transcripts (Yoshihara et al., 2015). These were mostly novel fusions that were not reported previously by targeted techniques, such as PCR or fluorescence *in situ* hybridization (FISH). Interestingly, only 263 of the fusion transcripts were recurrent (detected at least twice), indicating a stochasticity of fusion events and the power of unbiased approaches to detect rare non-recurrent fusions. However, the analysis of fusion by transcriptomics only allows for the detection of events that eventually lead to fusions transcripts and strong changes in gene expression.





**Figure 1.6 Methodologies to detect gene fusions and chromosome translocations.** PCR based methods allow sensitive detection of known fusion that can be detected in genomic DNA (gDNA) or mRNA. Quantitative approaches include serial dilutions, RT-qPCR and digital droplet PCR (ddPCR). Completely biased of fusions and translocations can be achieved by next-generation sequencing methods such as whole genome sequencing (WGS) or RNA-Seq. Fusion are detected by changes in the alignment of reads to the genome or transcriptome. Linear amplification-mediated high-throughput genome-wide sequencing (LAM-HTGTS) is a semi-biased method that sensitively detects fusions of bait genomic sites to a prey DSB. Specific translocation can furthermore be fluorescently labelled by FISH and microscopically identified in metaphase chromosomes or interphase cells.

### 1.7.5 Microscopic detection of recurrent chromosome translocations by FISH

Recurrent fusions that are associated with specific diseases are commonly detected by labelling DNA loci of interest with fluorophores by FISH. It is a targeted approach where chromosome rearrangements are classically identified by visual inspection of condensed chromosomes in metaphase spreads and in which putative fusion partners are labelled (Figure 1.6) (Wolff et al., 2007). An advantage of FISH is that the probes can span larger genomic areas and thus allow detection of fusions in broader breakpoint cluster regions (BCRs). The technique is, however, limited by its low throughput and the necessity to choose specific probes of putative fusion. The need for condensed chromosomes furthermore causes a bias of fusion analysis towards cells cycling into M-phase, thereby neglecting arrested cells in interphase. Alternatively, FISH can be applied to interphase cells where fusion are

identified by colocalization of translocation partner probes, thereby circumventing the cell cycle bias. The power of fusion analysis by FISH lies in the single allele and single cell information that is given. It facilitates the quantification of rearrangements and yet provides a relative high flexibility by the choice of different genomic loci that can be sampled as well as its independence from the source of genomic instability to induced DSBs and fusion events. Advances to improve the sensitivity of FISH thus could yield a quantitative single cell method that would allow the mechanistic study of oncogenic translocation formation.

### 1.8 Aims

Millions of cancer survivors struggle with the long-term effects of the treatments they received. Among the most detrimental side effects of antineoplastic therapies are the formations of secondary, therapy-related cancers. Here, we aim at understanding the molecular mechanisms that contribute to the formation of oncogenic *MLL* translocations that are a hallmark of therapy-related acute leukemias. We want to uncover how TOP2-targeting drugs lead to genomic instability, *MLL* breakage and *MLL* translocations, as well as how TOP2-DNA repair and spatial proximity of potential fusion partners contribute to the formation of oncogenic fusions.

*How are ETO-induced TOP2ccs converted into DSBs across the genome?*

We first aimed to understand how TOP2 poisons, such as etoposide, cause genomic instability. Therefore, we strived to map DSBs throughout the genome in hematopoietic cells as well as primary, human CD34+ bone marrow progenitor cells; the cell population where leukemic transformation occurs. Therefore, a collaboration with Nicola Crosetto was established to apply genome-wide DSB profiling by BLISS upon ETO-treatments. We then asked which genomic context, such as the context of chromatin, gene activity and chromosome organization, contributes to TOP2-induced DSBs. We therefore performed correlation analysis with genomic features, datasets of chromatin accessibility or binding of proteins involved in transcription or chromatin organization gaining first insights into what makes genomic sites susceptible to ETO-induced DNA damage.

*How can we quantitatively model the formation of chromosome translocations to understand mechanisms of formation? At what frequencies do oncogenic MLL fusion occur upon ETO-treatment in human hematopoietic cells?*

We aimed to develop a FISH-based method that detects rare chromosome translocations under physiologically relevant conditions to study mechanisms of cancer-associated fusions. We thus develop a high-throughput multicolor FISH assay to study the breakage, spatial proximity and fusion formation in single cells and in single alleles, which we termed C-Fusion 3D. Initial testing of C-Fusion 3D verified its ability to detect rare fusion events with high sensitivity. We thus evaluated the frequency of breakage and fusions of *MLL* and common fusion genes. With this tool at hand, pathways and factors that contribute to the formation of gene fragility and oncogenic fusions could be studied.

*What is the contribution of replication, transcription and spatial genome organization to TOP2cc conversion and the formation of MLL translocations?*

The formation of DSBs upon poisoning of TOP2s is directly linked to the formation of therapy-related oncogenic translocations. Several studies have implicated transcription to be a main contributor to TOP2-associated DSB formation although direct evidence is still scarce. Alternatively, genome architecture has evolved as a potential driver of chromosome fragility. We wondered to what extent these molecular mechanisms contribute to the conversion of TOP2ccs into DSBs and how they influence the formation of *MLL* fusions. We thus aimed to elaborate on features of transcriptional activity, replication as well as chromatin organization that could be found at sites of ETO-induced DSBs profiles generated by BLISS. We also studied the characteristics of recurrent *MLL* fusion partners with regard to activity and nuclear positioning to understand why ETO-treatments are linked to specific genome rearrangements in secondary leukemias. To draw more direct links, we sought to quantify changes in gene fragility by C-Fusion 3D upon the inhibition of transcription and replication, or local perturbations, to uncover their relative contributions. Finally, we aimed to study if the contributions of these pathways are in relationships to each other and can cause TOP2-induced genome fragility through synergies that favor DSB formation.

*Does the spatial genome organization and proximity influence the frequency of the various MLL fusions?*

Next to the formation of DSBs, spatial proximity of broken loci is a prerequisite for illegitimate fusion events. We thus wanted to investigate if *MLL* is proximal to its common translocation partner genes and if this influences the amount of fusion that we can identify. This could give insights if the proximity of common fusion partners is reflected in the frequencies that these

fusion are found in t-AML patients, or if additional selection mechanisms are potentially involved to favor oncogenic fusions.

*How do TOP2 isoforms contribute to ETO-induced genomic instability and MLL translocations?*

The expression of TOP2A and TOP2B differs drastically between different cell types, as it depends on the degree of differentiation and the proliferation status of cells. The isoform specificity of TOP2-targeting chemotherapies is currently not taken into account, because the roles of TOP2A and TOP2B to genome fragility are not well understood. We thus aimed to systematically approach the roles of TOP2A and TOP2B to ETO-induced genomic instability, and specifically *MLL* fragility, in isogenic cell models. Since TOP2A is essential for proliferating cells in culture, we strived to employ inducible TOP2A depletion systems and combine them with the knockout of TOP2B. Finally, the dependency of the cell cycle was evaluated to estimate how cell states influence ETO-induced DNA fragility.

*Which repair factors contribute to the processing of TOP2ccs and how are protein-free DSBs repaired? Which pathway suppress or facilitate the formation of oncogenic MLL fusions?*

We aimed to estimate the contributions of MRE11 and TDP2 to ETO-induced TOP2cc repair as well as the contribution of DSB repair pathways NHEJ and HR to illegitimate joining of DNA ends to form oncogenic *MLL* fusions. We therefore sought to use a panel of isogenic cell lines with either gene disruptions or conditional depletion systems for the major candidates of TOP2cc repair. With these cell lines at hand we asked whether TOP2-dependent DNA damage is mainly suppressed by MRE11 or TDP2, and if they act epistatically. Furthermore, the occurrence of *MLL* translocations was studied in relation to DSB repair pathways to determine whether HR or NHEJ mediates repair of DSBs at the *MLL* translocation hotspot.

*Can we identify novel factors that modulate the susceptibility of MLL to etoposide and that potentially contribute to TOP2cc repair pathways?*

Given the multistep repair of TOP2-linked DNA breaks, we wondered if we can identify novel factors that are involved in the predisposition of *MLL* for gene fragility, in TOP2cc conversion or erroneous joining events leading to *MLL* fusions. We thus aimed to use C-Fusion 3D to screen for factors that modulate *MLL* fragility upon ETO-treatment and to uncover their contributions to TOP2-induced genome instability and the formation of oncogenic chromosome translocations.

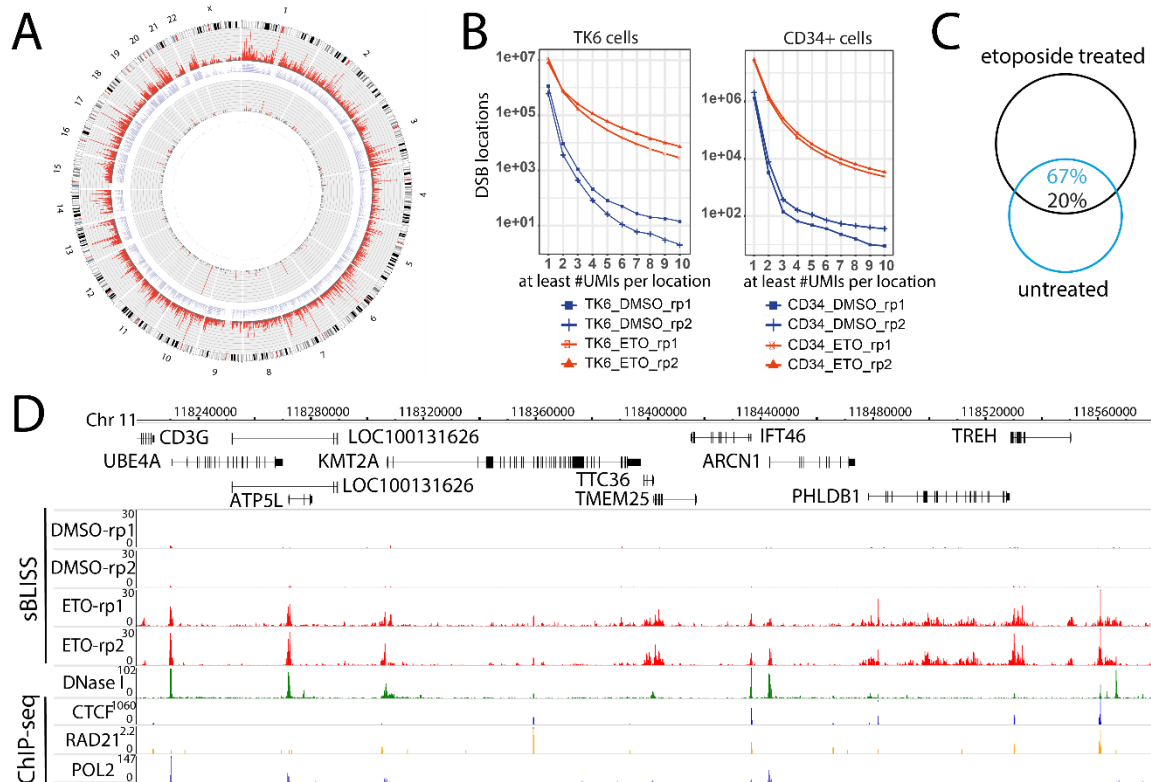
## Chapter 2 Results

### 2.1 Etoposide induces gene fragility and oncogenic translocations in hematopoietic cells

#### 2.1.1 Genome-wide profiling of etoposide-induced DSBs in human hematopoietic cells by sBLISS

To understand how TOP2cc conversion is linked to genomic instability upon etoposide treatment, we performed genome-wide profiling of DSBs by sBLISS upon ETO-treatment in human hematopoietic cells. sBLISS is an adaption of BLISS (see Chapter 1.7.1) that specializes on suspension cells and performs all steps in cell pellets (Bouwman et al., 2020; Yan et al., 2017). ETO-induced DNA breaks were detected in two hematopoietic cell lines, K562 and TK6, as well as in bone marrow CD34+ progenitor cells. sBLISS profiles indicated that DSBs upon ETO-treatment were not randomly distributed across the genome, but were found at recurrent genomic sites (Figure 2.1A). Clusters of DSBs, also named hotspots, were enriched in ETO-treated samples in comparison to DMSO mock samples and showed high reproducibility between replicates (Figure 2.1B, D). Interestingly, sites with recurrent DSBs were also identified in DMSO samples. These endogenous breaks overlapped with ETO-induced breaks by 67 % but only contributed a fraction to the overall breaks detected upon ETO-treatment (Figure 2.1C). Hotspots of DSBs were located in genic as well as intergenic regions and we observed site-specific overlaps with protein binding or chromatin features (Figure 2.1D). For example, in a gene dense regions on chromosome 11, a high overlap of sBLISS hotspots with DNaseI hypersensitivity (DNaseI-HS) was seen, which is a surrogate of open chromatin sites (green tracks, Figure 2.1D). Most of the hotspots in this regions seemed to accurately colocalize with CTCF, cohesin (RAD21) or RNA-PolII binding sites (blue, yellow, purple tracks, Figure 2.1D).

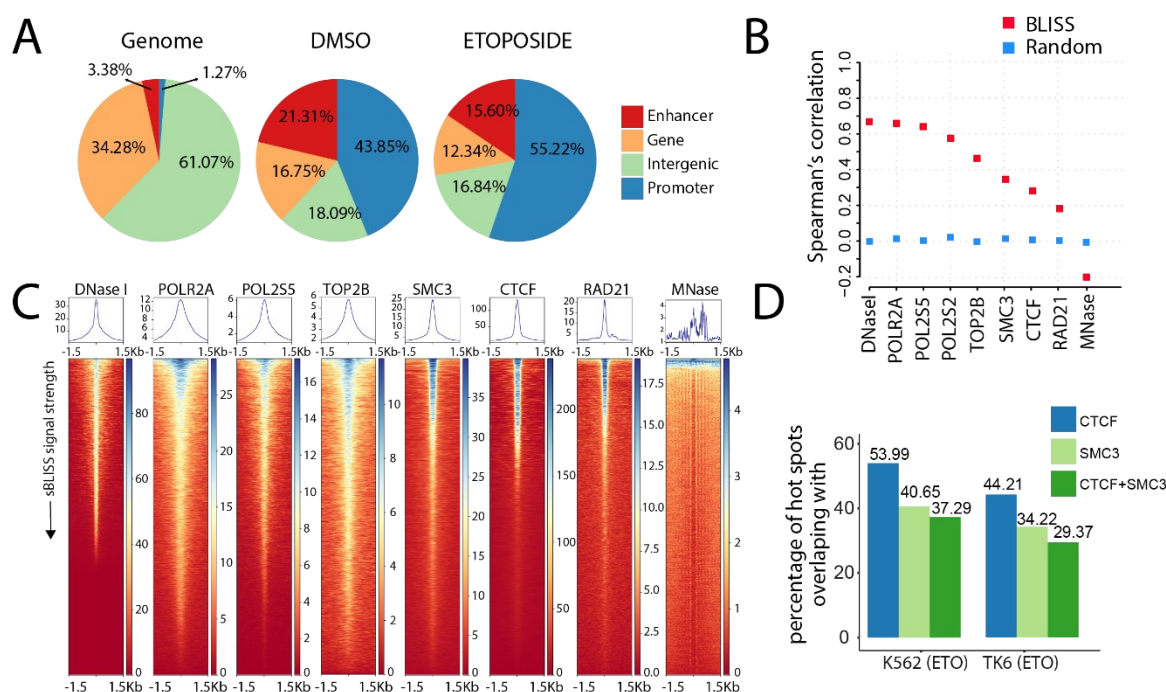
## Results



**Figure 2.1 sBLISS identifies DSBs across the genome in hematopoietic cells.** (A) Circus plot of DSBs detected by sBLISS in TK6 cells, either DMSO (inner circle) or ETO-treated (outer circle, 30  $\mu$ M for 4 h). Grey tracks mark hot spots with adjusted  $p < 0.05$ , red tracks with adjusted  $p < 1e-10$ . Coverage density of hotspots along chromosomes are marked light purple. (B) DSB locations detected in replicates of TK6 or CD34+ progenitor cells treated either with DMSO or ETO (TK6 cells: 30  $\mu$ M for 4 h, CD34+ cells: 60  $\mu$ M for 4 h), filtered by number of UMIs detected at location. (C) Venn diagram showing the overlap between spontaneous and ETO-induced (20  $\mu$ M, 6 h) DSB hot spots in K562 cells. (D) Representative genomic regions with sBLISS DSB profiles (red) in replicates of DMSO or ETO-treated TK6 cells. Overlay with DNaseI-HS tracks (green) and ChIP-seq data of CTCF (blue), RAD21 (yellow, cohesin subunit) and RNA-PolIII (purple). DNaseI-HS and ChIP-seq data were obtained from ENCODE (see Table 11).

To understand where DSBs are formed upon stabilization of TOP2ccs with etoposide, we analyzed the genome-wide distribution of breaks across these features. Although most of the genome is composed of intergenic sequences (Figure 2.2A, *Genome*), more than 80 % of DSBs in DMSO as well as ETO-treated K562 cells were found in promoters, enhancers and in genes (Figure 2.2A, DMSO and ETO). Genome-wide correlation analysis with DNA conformation data showed a high enrichment of breaks within DNaseI-HS regions (Figure 2.2B, C, Spearman's correlation,  $Sc = 0.67$ ,  $p < 10^{-5}$ ). In line with this observation, breaks were depleted from nucleosome-occupied regions, as was indicated by a negative correlation with MNase-Seq, which identifies nucleosome protected DNA (Figure 2.2B, C). ETO-induced DSBs were also highly enriched at sites of RNA-PolIII occupancy: strong correlations with the large subunit POLR2A ( $Sc = 0.66$ ,  $p < 10^{-5}$ ) as well as the activated forms harboring

phosphorylation of residues S2 or S5 were seen ( $pS2$ :  $Sc = 0.57$ ,  $p < 10^{-4}$ ;  $pS5$ :  $Sc = 0.64$ ,  $p < 10^{-5}$ ). Furthermore, we found a high correlation of DSBs with TOP2B ChIP-seq data ( $Sc = 0.45$ ,  $p < 10^{-4}$ ) indicating that sBLISS indeed identified DSBs that are dependent on poisoning of topoisomerases. Next, we investigated if DSB hotspots were found at binding sites of CTCF or cohesin, which are key complexes in the regulation of chromosome architecture and has been proposed to actively contribute to the formation of ETO-induced DSBs (Canela et al., 2017). DSBs hotspots moderately correlated with CTCF or cohesin binding (cohesin complex members RAD21 and SMC3 were analyzed) (Figure 2.2B, C). In total, around 40 % of detected DSBs hotspots coincided with binding of CTCF, RAD21 or both in TK6 and in K562 cell lines (Figure 2.2D), suggesting that localization at chromosome architecture sites may contribute to fragility of some, but not all, ETO-induced DSB hotspots.



**Figure 2.2 ETO-induced DSBs cluster in open chromatin regions and colocalize with RNA-PolIII and chromatin looping factor binding sites.** (A) Pie charts representing the distribution of genomic features in the genome (left) or distribution of endogenous (DMSO) or ETO-induced DSB hotspots across genomic features in K562 cells. (B) Spearman's correlation coefficient between ETO-induced DSBs (sBLISS of K562 cells, ETO 20  $\mu$ M for 6 h) and open chromatin (DNaseI-HS), nucleosome occupancy (MNase-seq) or RNA-PolIII, PolIII-pSer2, PolIII-pSer5, CTCF, RAD21, SMC3 binding (ChIP-seq, derived from ENCODE, RAD21 and SMC3 are both cohesin subunits). For a random control, correlations of one hundred randomized tests with DSBs within regions of these features are shown. (C) Aggregate plots and signal heatmaps of features (as in B) centered around ETO-induced DSB hotspots in K562 cells that were sorted by descending sBLISS signal and extended for a total of 3 Kbp from its summit. (D) Percentage of ETO-induced DSB hot spots in K562 and TK6 cells overlapping with ChIP-seq signals of CTCF (blue), SMC3 (light green) or signals of both (dark green).

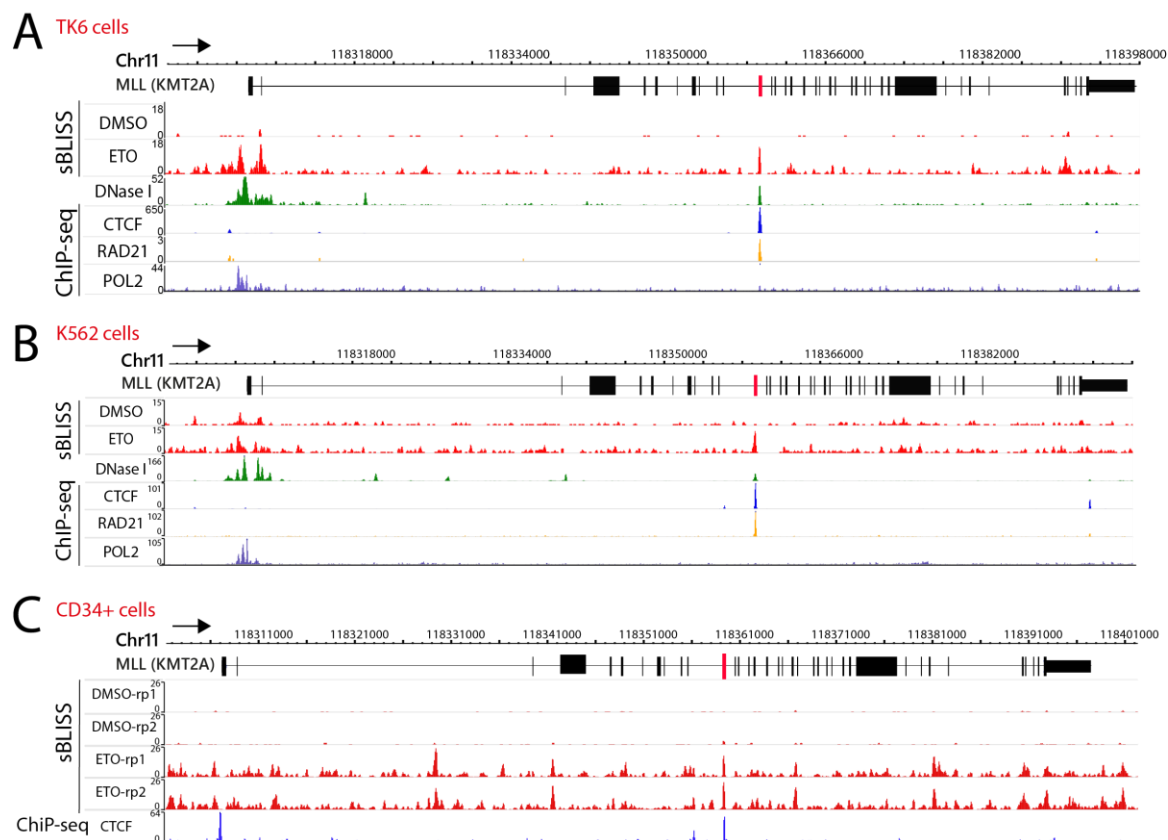
Our sBLISS data thus indicates that ETO induces DSBs in active, open chromatin regions that partially overlap with features of chromosome architecture. Endogenous DSBs detected by sBLISS often overlapped with locations of ETO-induced breaks, suggesting that a substantial fraction of spontaneously occurring breaks in human cells could possibly result from abortive TOP2 catalytic cycles.

### 2.1.2 sBLISS identifies DSB hotspots within *MLL* and recurrent fusion partners

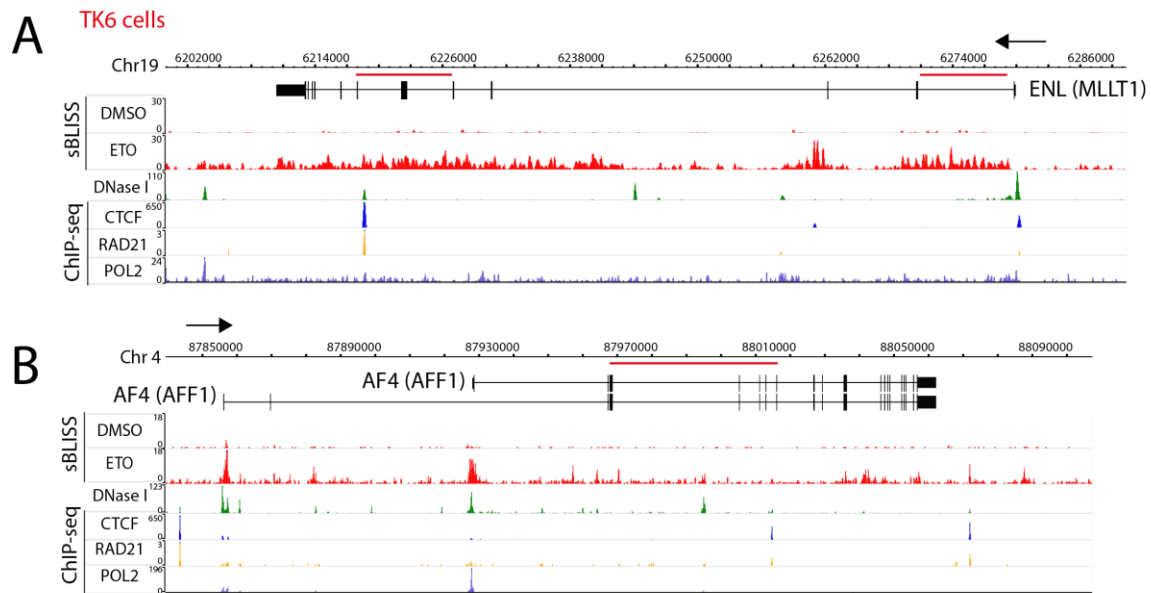
To understand how DSBs induced by etoposide relate to the occurrence of *MLL* translocations in patients, we specifically looked at sBLISS hotspots within the *MLL* gene (Figure 2.3). The translocation hotspot identified from patients is located within a BCR at *MLL* exon 12 (marked in red, Figure 2.3). We indeed found a hotspot of DSBs at *MLL* exon 12 upon ETO-treatment in TK6, K562 and CD34+ cells (Figure 2.3A, B, C). Importantly, this hotspot colocalized with an open chromatin region and binding sites of CTCF and RAD21 (TK6 and K562 cells) as has previously been reported (Canela et al., 2017; Cowell et al., 2012). As for most genes, recurrent breaks were also identified in the promoter region of *MLL* that coincides with DNaseI hypersensitivity and RNA-PolII binding. The primary CD34+ cells showed several additional hotspots in other exons and introns of *MLL*, which neither overlapped with CTCF binding nor DNaseI signals. In fact, an additional CTCF binding site was seen at the *MLL* promoter in CD34+ cells, which however did not colocalize with DSB clusters. The sBLISS data thus revealed that upon ETO-treatment, DSBs accumulate at the *MLL* promoter as well as the exon 12 translocation hotspot. The opportunity for oncogenic *MLL* fusions to appear could thus directly result from gene fragility of *MLL* at exon 12.

To form oncogenic translocations, illegitimate repair of *MLL* with other broken genomic sites has to take place. We thus reasoned that ETO-induced fragility might also occur within the most common fusion partner genes. Within *ENL* and *AF4*, enrichments of DSBs were found but breaks occurred in broader regions that only partially overlap with known translocation hotspots (marked as red bars, Figure 2.4). This suggests that selection of specific gene fusions with oncogenic potential plays an important role in the biogenesis of *MLL* translocations.





**Figure 2.3 DSBs in the *MLL* gene detected by sBLISS in TK6, K562 and CD34+ progenitor cells.** Tracks of DSBs detected by sBLISS within the *MLL (KMT2A)* gene on chromosome 11 in DMSO or ETO-treated cells: (A) TK6 cells treated with 30  $\mu$ M ETO for 4 h, (B) K562 cells treated with 20 $\mu$ M for 6 h or (C) CD34+ bone marrow progenitor cells treated with 60  $\mu$ M ETO for 4 h. The translocation hotspot at exon 12 is marked in red. Overlays with DNaseI-HS data and ChIP-seq data for CTCF, RAD21 and RNA-PolIII are also shown.



**Figure 2.4 DSBs in translocation partner genes *ENL* and *AF4* are partially found in translocation hotspot regions.** (A) sBLISS DSB profiles within the *ENL* (*MLLT1*) gene on chromosome 19 and (B) the *AF4* (*AFF1*) gene on chromosome 4 in DMSO or ETO-treated TK6 cells. Red lines indicate BCRs in these genes. Arrows represent the direction of transcription. ChIP-seq data is shown as in Figure 2.1.

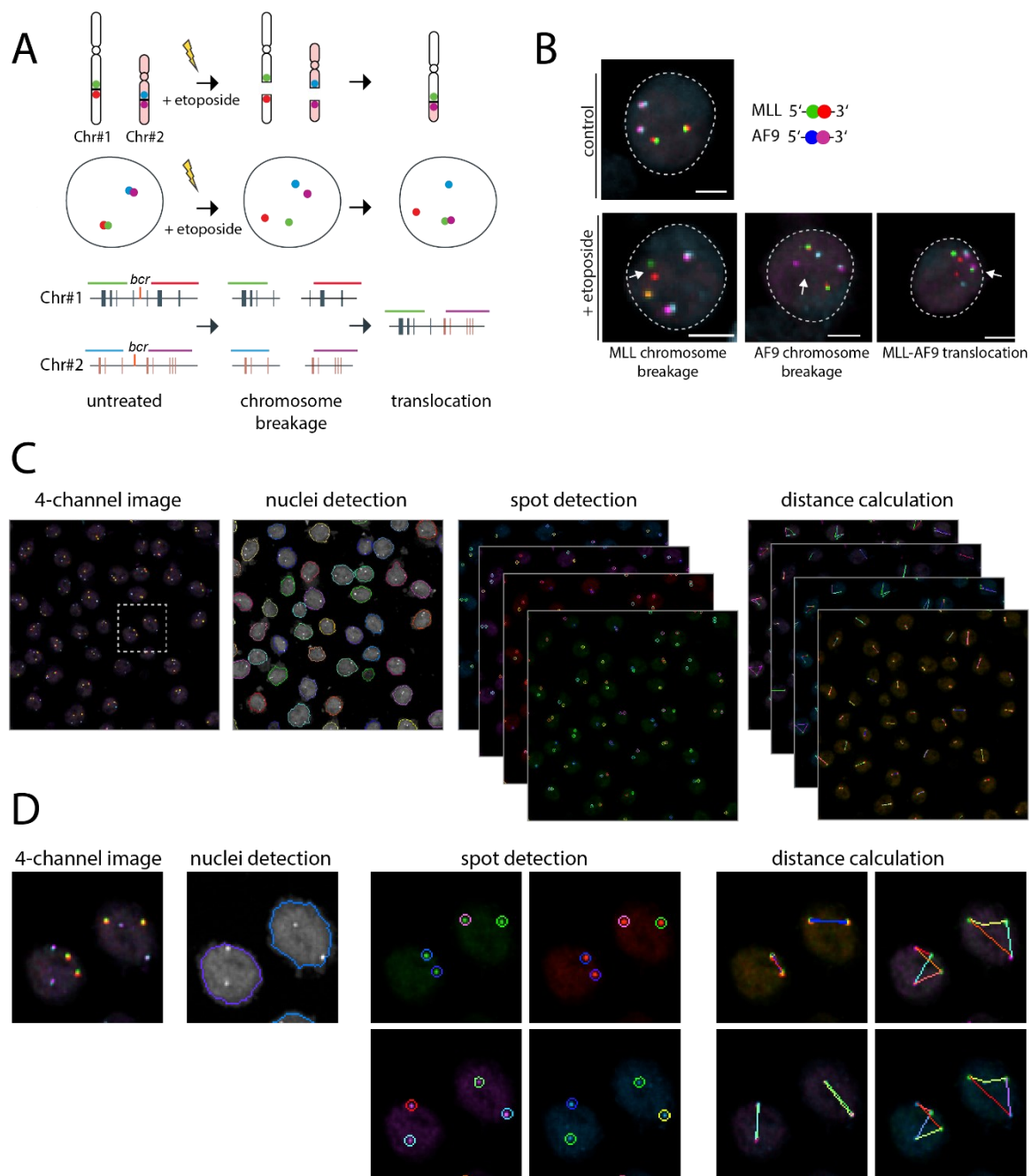
### 2.1.3 C-Fusion 3D: a novel tool to study nuclear positioning and rare chromosome rearrangements in 3D

Genome-wide profiling of DSBs by sBLISS revealed that ETO-treatment led to enrichment of DSB within fusion regions of *MLL* and common translocation partner genes. To quantify the frequencies of gene breakage and chromosome translocation events, we developed a novel high-throughput imaging method: C-Fusion 3D. C-Fusion 3D is based on fluorescence *in situ* hybridization (FISH) that labels up to four different genomic loci in interphase cells flanking chromosome breaks of genome rearrangement hotspots, combined with high-throughput microscopy and automated image analysis in 3D. It allows the detection chromosome breaks and fusion with very high sensitivity without exact knowledge of the breakpoints within genes. Genomic loci of interest are labelled with complementary, fluorescently labelled probes that can be detected in four different channels by confocal microscopy. The breakage of genes upon DNA-damage can be identified by separation of probes flanking breakpoint cluster regions (BCRs). The simultaneous fragility of two genes of interest can be probed by the usage of four colored probes in one sample (one flanked by *Green-Red* probes, one by *Blue-FarRed* probes, Figure 2.5A, B). This is an important attribute of 4-color C-Fusion 3D, since it facilitates the specific detection of chromosome translocations. One can identify the breakage of two translocation partner genes and in the same cell identify illegitimate fusion

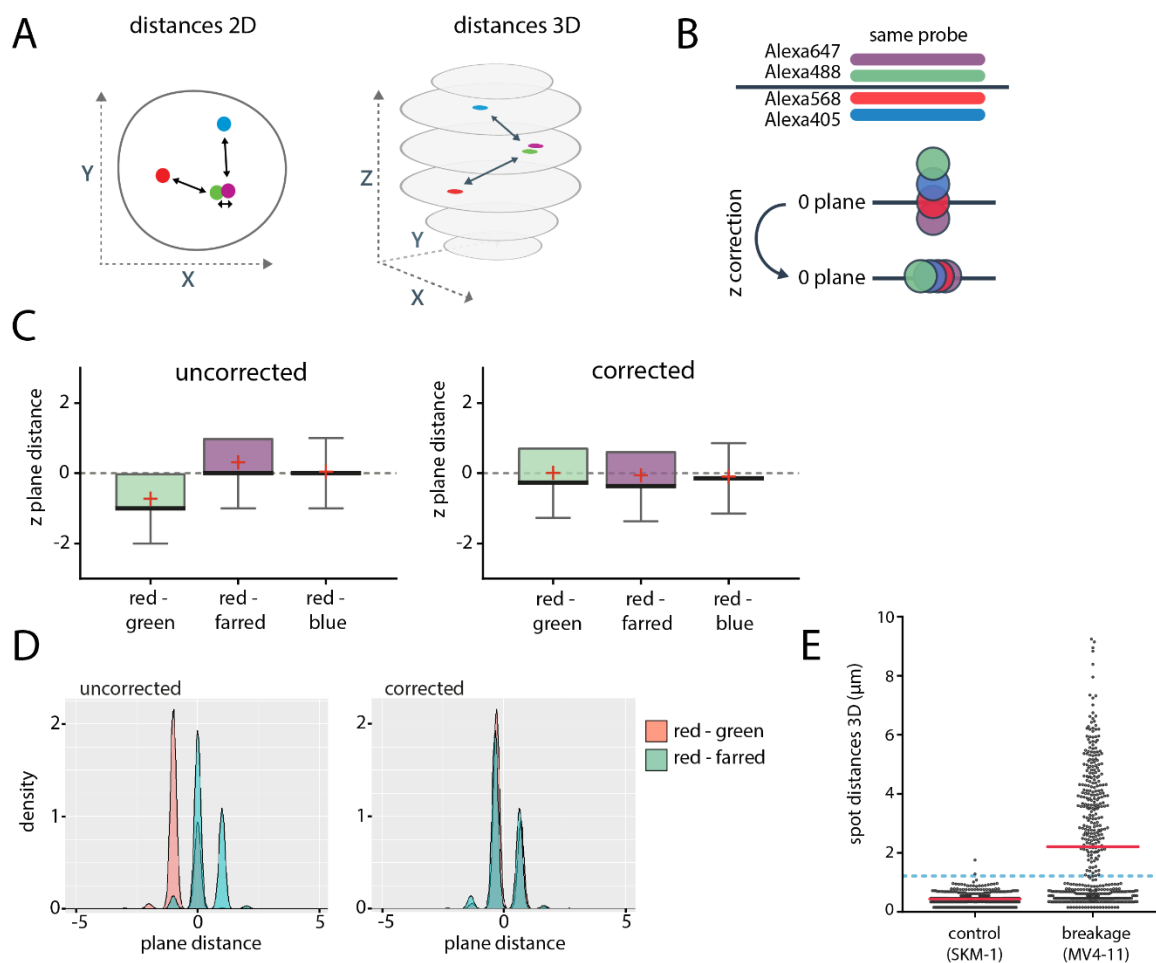
of DNA ends of different chromosomes by the colocalization of the two spots of two different chromosomes, e.g. the 5' end of chromosome #1 with the 3' end of chromosome #2 (Figure 2.5A). An automated image analysis pipeline faithfully determines inter-spot distances and detects separations and colocalizations of FISH probes in individual nuclei. Therefore, in a first step we segmented nuclei based on DNA staining (for two and three color FISH) or background signal of one of the fluorescent probes (for 4-color FISH, Figure 2.5C, D). Then, FISH spots were identified and distances between spot centers determined in flattened images (2D).

Nuclear distances of FISH probes are commonly detected in maximum projections of several z-stacks that slice the nucleus. Projected images reduce the amount of images that need to be acquired and simplify the analysis of large imaging datasets (Finn et al., 2017). Nonetheless, in light of the 3D nature of the mammalian nucleus, the detection of distances in 2D reduces the complexity of nuclear interactions. That is to say, distances in the third dimension  $z$  are not incorporated in the analysis of spatial separations whereby probes that lie in the same  $x$ - $y$  plane, but are separated in  $z$ , would be identified as colocalized. To achieve a resolution of nuclear distances in 3D, fine stacks that cover the whole cell nucleus need to be acquired. Here, we imaged between 10 and 16 z-stacks separated by 0.5  $\mu\text{m}$ , dependent on the heights of cellular nuclei that usually lie between 5 and 10  $\mu\text{m}$  in human cells (Figure 2.6A). A challenge of precise 3D distance determination is however the axial shift of light by chromatic aberrations that differs for different wavelengths (North, 2006). Since we span a broad range of wavelengths in four-color FISH, we sought to minimize this bias from chromatic shifts that can account for seeming separations of several hundreds of nanometers. We thus determined an axial-correction value for each color channel that normalizes  $z$ -plane values based on the *Red*/Alexa568 channel (Figure 2.6B). Therefore, the same FISH probe was labelled with the four different fluorescent colors (*Green*, *Red*, *FarRed*, *Blue*), the pairwise shifts in  $z$  were determined, and a correction value was calculated based on distance distributions (*Green*: -0.768, *FarRed*: 0.255, *Blue*: -0.146). The plane distances of *Green*, *FarRed* and *Blue* spots were then adjusted by the axial-correction value (Figure 2.6C, D) and Euclidean 3D distances between spot pairs were determined. We thus developed a high-throughput FISH based method that can accurately detect distances between spots pairs, such as different genomic loci, in 3D.

## Results



**Figure 2.5 Detection of chromosomal rearrangements by C-Fusion 3D.** (A) C-Fusion 3D utilizes a 3-color to 4-color fluorescence in situ hybridization (FISH) approach to label two chromosome break cluster regions (BCRs, red bars) of interest in cells, e.g. with red and green or far-red and blue fluorescent probes. FISH probes appear as small spots in interphase cells with overlapping probes of one BCR region. Upon breakage of chromosomes within the BCRs, spatial separation of spots are observed. Chromosome translocations are identified by breakage of both BCRs and the colocalization of two probes of two different chromosomes e.g. green and far-red. (B) Example images of TK6 cells with labelled *MLL* and *AF9* BCRs by 4-color FISH. Overlapping probes were seen in control cells (top). In cells treated with the DNA damage agent etoposide, breakage of both BCRs and *MLL-AF9* translocations were found. (C) Schematic of image analysis workflow. After high-throughput image acquisition, nuclei are segmented and defined as region of interest (ROI). Spots in all channels are detected and distances between spots in all channels are calculated. (D) Zoom-in of image analysis as shown in (C).



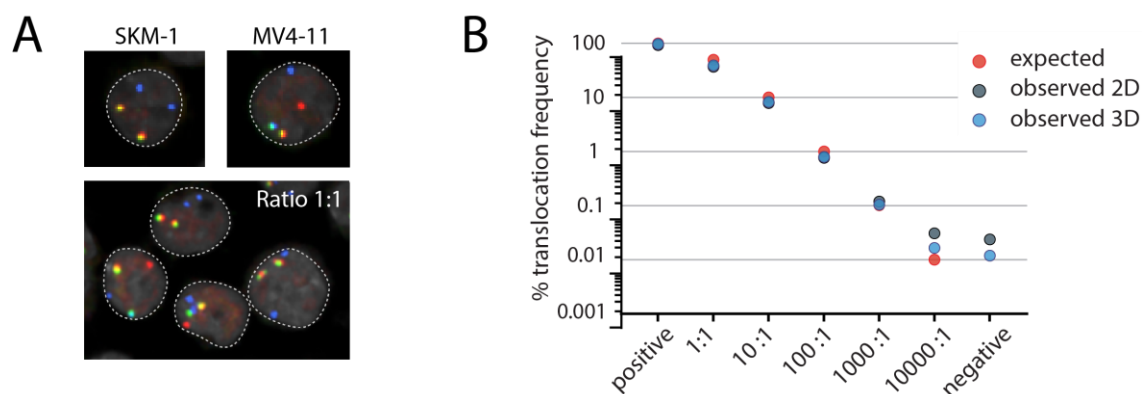
**Figure 2.6 Correction of chromatic shifts to optimize distance calculations in 3D.** (A) C-Fusion 3D calculates minimal distances between spots either in maximum projection (2D) or in individual plane of z-stacks (3D). (B) To correct z-shifts of fluorophores by chromatic aberrations, one genomic probe was labelled with the four different fluorophore-colors. The difference in the mean z-plane was calculated and subtracted from individual channels to align probes in z. (C) Z-plane distances between fluorophores of the same probe before and after correction. Red crosses represent the mean, black bars the median. (D) Density plot of uncorrected and corrected plane distances *Red-Green* and *Red-FarRed*. (E) Jitter plot of spot distances in 3D in  $\mu\text{m}$  between probes flanking the *MLL* BCR. In control cells (SKM-1), the *MLL* gene is intact whereas in MV4-11 cells one allele of *MLL* is involved in a chromosome translocation and thus permanently apart. Red bars indicate the mean, blue dotted line represent the 1.2  $\mu\text{m}$  distance threshold,  $n=500$ .

By the use of customized R scripts, the frequency of chromosome breaks, proximity of genomic loci in 3D and translocation rates can be calculated on the basis of the spot distances for individual nuclei from the same dataset. Based on distance distributions in control cells, in which the break-apart probes of a genomic locus were separated by the linear genomic distance of a few hundred kilobases, we defined a threshold of 1.2  $\mu\text{m}$  between spots in 3D that has to be exceeded to identify the separation of spots (Figure 2.6E). This was in agreement with previous studies that defined similar thresholds for colocalization and separation of fusion sites in live cells (Roukos, Voss, et al., 2013). When a *Green* and *Red* probe

flanking a BCR were separated by more than 1.2  $\mu\text{m}$ , the cell was thus counted to have a breakage. Only cells with equal spot numbers in all channels were taken into account for the analysis to ignore falsely identified spots or deletion events. The threshold thereby resulted in a false positive rate of  $\sim 0.3\%$  in control cells (SKM-1) whereas 98 % of cells containing a *MLL* translocation and thus harboring a permanent separation of one *MLL* allele (MV4-11 cells) were identified to have a *MLL* break. Nuclei with chromosome translocations were furthermore identified by the separation of probes flanking the BCRs and colocalization of the two probes of the two potential fusion partners. For example, for the detection of *MLL-AF9* translocations by C-Fusion 3D, nuclei were counted to harbor a chromosome translocation when *MLL* 5' and *AF9* 3' probes were overlapping and were separated from the *MLL* 3' and *AF9* 5' probes (Figure 2.5B). A more detailed description of selection parameters can be found in the Material and Methods Chapter 4.12.2.

### 2.1.4 C-Fusion 3D detects oncogenic *MLL* translocations with high accuracy and sensitivity

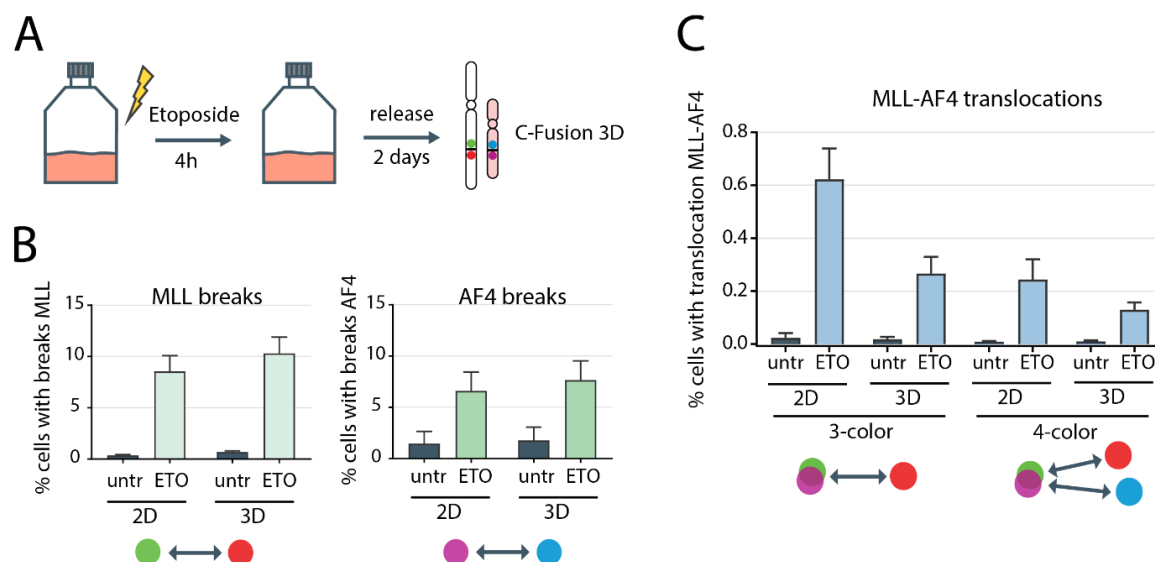
To test the sensitivity of C-Fusion 3D, we performed a proof of principle experiment with cells harboring the chromosome translocation *MLL-AF4* (MV4-11 cells) or having the wild-type intact chromosome conformation (SKM-1 cells) (Figure 2.7A). The specific breakpoint in *MLL* and *AF4* are not known for MV4-11 cells and we designed the FISH probes to flank the genes upstream and downstream (*MLL*: green and red, *AF4*: farred). High accuracy levels of spot detection were found with 99.7 % of spots being correctly identified and a maximum of 4 % of spots not being identified. To limit the impact of unidentified or falsely identified spots, translocation frequencies were only determined in cells harboring exactly two spots per probe (cells are diploid). Translocation frequencies were calculated by 2D and 3D analysis in samples where MV4-11 cells were spiked into SKM-1 cells at ratios 1:1 to 1:10'000 cells (Figure 2.7B). We found that C-Fusion 3D detects chromosome translocations above background at levels up to  $10^{-4}$  and could thus identify single rearrangements within thousands of cells. The observed frequencies by 3D analysis surpassed the accuracy of 2D analysis especially at translocation frequencies below  $10^{-3}$  and in negative controls.



**Figure 2.7 C-Fusion 3D detects chromosome translocations with high sensitivity at levels  $10^{-3}$  to  $10^{-4}$ .** (A) Spike-in experiment of MV4-11 cells harboring an *MLL-AF4* translocation into SKM-1 cells that do not have *MLL* rearrangements. *MLL* was labelled with red and green probes, *AF4* with a farred probe. (B) MV4-11 cells were mixed into SKM-1 cells at given ratios from 1:1 to 1:10000 cells and translocation frequencies were determined with C-Fusion 3D in 2D and 3D.

To compare the validity of the 2D and 3D analysis of C-Fusion 3D, we detected chromosome translocations under relevant conditions by sampling frequencies of *MLL-AF4* fusions that form in hematopoietic TK6 cells upon induction of DNA damage with etoposide. We thus treated lymphoblastoid TK6 cells with relevant doses of ETO for four hours and released cells to repair for two days (Figure 2.8A). Quantifications of separations of *MLL* or *AF4* flanking probes revealed specific chromosome breaks in 6 – 10 % of cells (Figure 2.8B). Calculations of distances in 3D increases the frequency of breaks that were detected over the 2D analysis by 15 – 20 %. Translocation frequencies were also determined in 2D and 3D by either 3-color or 4-color analysis, i.e. the separation of *AF4* probes was considered or not (Figure 2.8C). Analysis in 3D detected *MLL-AF4* fusions at approximately 50 % lower rates than in 2D, indicating that 3D distance calculations are crucial for the specificity of detection. The use of 4-color FISH further reduced translocation frequencies from 0.27 % to 0.13 % (Figure 2.8C), showing the benefit of the complex FISH assay.

C-Fusion 3D thus proves to be a new method to detect chromosome fusions with very high sensitivity at frequencies up to  $10^{-3}$  to  $10^{-4}$ . The complexity of the assay (break-apart only, 3-color or 4-color FISH) can be adjusted depending on the research objective. We thus applied C-Fusion 3D to study TOP2-induced gene fragility, *MLL* fusion formation and proximity of genes in 3D.



**Figure 2.8 Chromosome instability and translocation formation upon etoposide treatment can be studied by C-Fusion 3D.** (A) Lymphoblastoid TK6 cells were grown to medium density, acutely treated with etoposide for 4 h and released from treatment for 2 days. Then, chromosome breaks and translocations were quantified with C-Fusion 3D. (B) Chromosome breaks of *MLL* and *AF4* BCRs were detected in 2D and 3D. (C) Comparison of detected *MLL-AF4* translocations when utilizing 2D or 3D and 3-color or 4-color analysis of C-Fusion 3D.

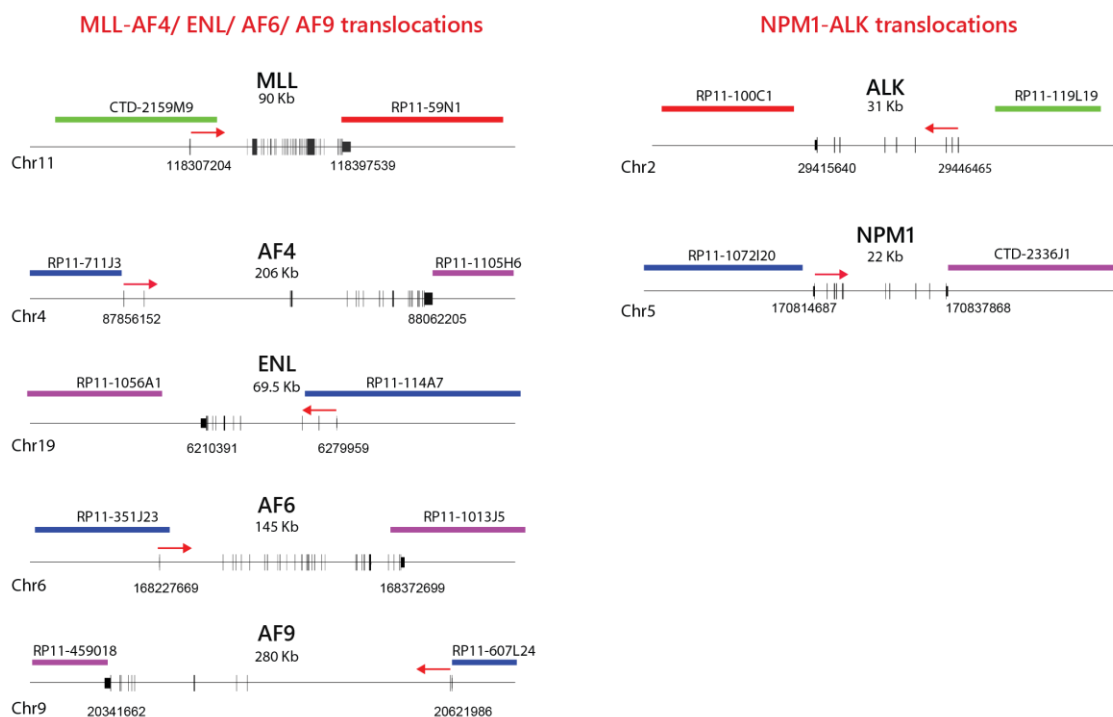
### 2.1.5 TOP2-induced *MLL* fusion are rare and occur at frequencies of $10^{-3}$ to $10^{-4}$

To estimate frequencies of *MLL* and common fusion partner breaks as well as fusion events, we employed C-Fusion 3D to quantify *MLL* rearrangements. For this project, we chose two FISH probes flanking the *MLL* gene colored in green and red as well as a combination of blue and far-red probes to simultaneously label one of the common fusion partners: *AF4*, *ENL*, *AF9* or *AF6* (Figure 2.9). This enabled the detection of DSBs in *MLL* and the fusion partner in the same experiment as well as quantification of their fusion rate. The well-studied translocation of *ALK-NPM1*, which is found in cases of anaplastic large cell lymphoma, served as a control.

Lymphoblastoid TK6 cells were acutely treated with etoposide and released into fresh medium for two days, to allow for repair of breaks and formation of fusions. Persistent DSBs in *MLL* as well as fusion partner genes were detected in a dose-dependent manner by C-Fusion 3D (Figure 2.10A). Breaks within *MLL* were most frequently found with up to 10 % of cells were identified with separations. The fusion partner genes showed lower fragility with breaks found in 7.2 % for *AF4*, 6.4 % for *ENL* and around 5 % for *AF9* and *AF6* of cells (Figure 2.10A). The fusion of two genes was only observed in approximately one out of thousand cells with *MLL* translocation rates between  $1 \times 10^{-3}$  for *MLL-AF4*, *MLL-AF6* and *MLL-AF9* and  $2 \times 10^{-3}$  for *MLL-ENL* (Figure 2.10B). The two times higher levels of *MLL-ENL* fusions could not merely



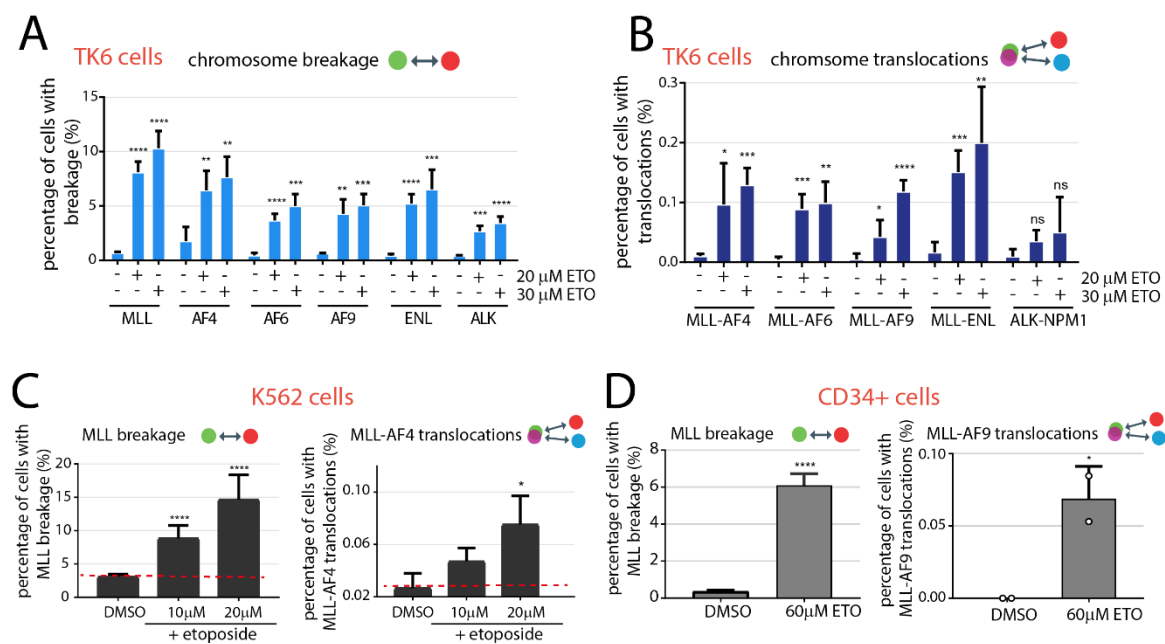
be explained by higher gene fragility as *ENL* showed similar break frequency as other fusion partners. Fusions of *ALK* with *NPM1* were only detected within background levels.



**Figure 2.9 Design of FISH probes to study *MLL* translocation by C-Fusion 3D.** Schematic of genes labelled with FISH probes to study *MLL* translocations with *AF4*, *AF9*, *AF6* or *ENL* and control *ALK-NPM1* translocations by C-Fusion 3D. Colors of probes represent fluorophores used for a 4-color FISH assay: green – Alexa488, red – Alexa568, blue – CF405, purple – Alexa647. Probes are labelled with their BAC IDs. Red arrows indicate the 5' to 3' orientation of genes.

In erythroleukemic K562 cells as well as primary CD34+ cells fragility of *MLL* was found at similar levels of magnitude, though at different etoposide concentrations (Figure 2.10C, D). A slightly higher frequency of cells with *MLL* breakage (~14 %) that was found in K562 cells could be explained by their multiploid genome since C-Fusion 3D quantifies the number of cells with chromosome breakage. Thus, the more copies of a gene are present in individual nuclei, the higher the chance of a gene breakage per cell. Importantly, we were able to detect *MLL-AF9* fusion in primary CD34+ progenitor cells, which are derived from the cell population where the formation of *MLL* translocations can lead to onset of t-AML (Figure 2.10D). The translocation frequency of  $7 \times 10^{-4}$  is in line with previous reports, indicating that C-Fusion 3D can be used to accurately model the formation of therapy-related translocations (Piganeau et al., 2013).

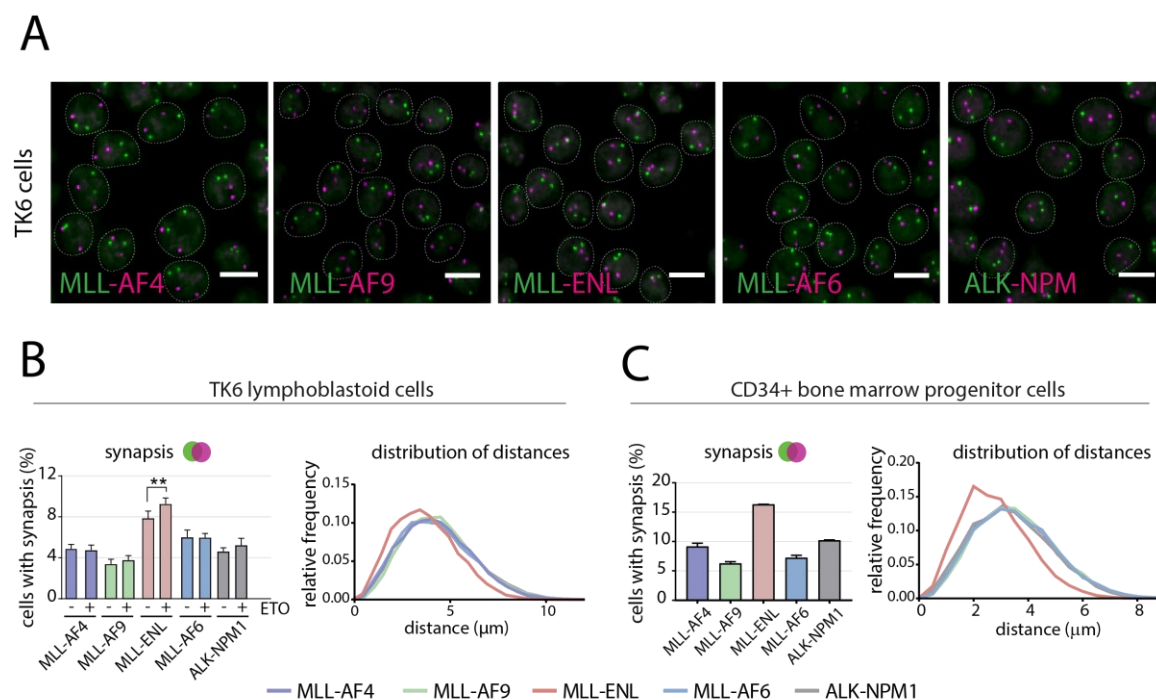
## Results



**Figure 2.10 C-Fusion 3D detects and quantifies ETO-induced gene breakage and *MLL* translocations in single hematopoietic cells.** (A) Chromosome breaks or (B) translocations of indicated genes were quantified in lymphoblastoid TK6 cells upon an acute ETO treatment (20 μM or 30 μM for 4 h) and a two-day release. (C) Breakage of *MLL* and *MLL-AF4* translocations in erythroleukemic K562 cells upon acute ETO-treatment (10 μM and 20 μM, 6 h) and a two-day release. (D) Breakage of *MLL* and *MLL-AF9* translocations in bone marrow CD34+ progenitor cells upon acute ETO-treatment (60 μM, 4 h) and a two-day release. In all cells, breaks and translocations were quantified by C-Fusion 3D with a 4-color FISH approach. Mean with SD shown, TK6 cells:  $n \geq 4$ , K562 cells:  $n \geq 3$ , CD34+ cells *MLL* breakage:  $n = 6$ , CD34+ cells *MLL-AF9* translocations:  $n = 2$ . \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$  by student's t-test compared to DMSO controls.

The formation of oncogenic gene fusions does however not only rely on gene fragility but also spatial proximity of DNA ends to facilitate incorrect DSB repair. We hypothesized that the more frequently observed *MLL-ENL* fusions could be a result of more frequent colocalization of *MLL* and *ENL* to promote fusion of DNA ends. To explore how spatial proximity contributes to formation of *MLL* translocations, we assessed the 3D Euclidean nuclear distances between *MLL* and common fusion partners. In TK6 and CD34+ progenitor cells, *MLL* was significantly closer to *ENL* than to other genes. This was observed in overall nuclear distances as well as in frequency of cells where genes colocalize (Figure 2.11A-C, synopsis TK6 cells *MLL-ENL* = 7.9 %, *MLL-AF4* = 5.1 %, *MLL-AF9* = 3.4 %, *MLL-AF6* = 6.0 %, *ALK-NPM* = 4.6 %). Intriguingly, in TK6 cells synopsis of *MLL* and *ENL* was significantly increased upon an acute treatment with etoposide, which was not observed for other fusion partners (Figure 2.11B,  $p = 0.004$ ). Higher levels of *MLL-ENL* synopsis could thus explain the two times higher levels of fusion seen for *MLL* with *ENL* (Figure 2.10B).

With the ETO-induced genome-wide DSB profiles in hematopoietic cells as well as C-Fusion 3D to quantify fragility and fusions of single alleles and in single cells, we next strived to understand how TOP2cc conversion into DSBs is mediated and which molecular processes contribute to the formation of *MLL* fragility and oncogenic fusions.



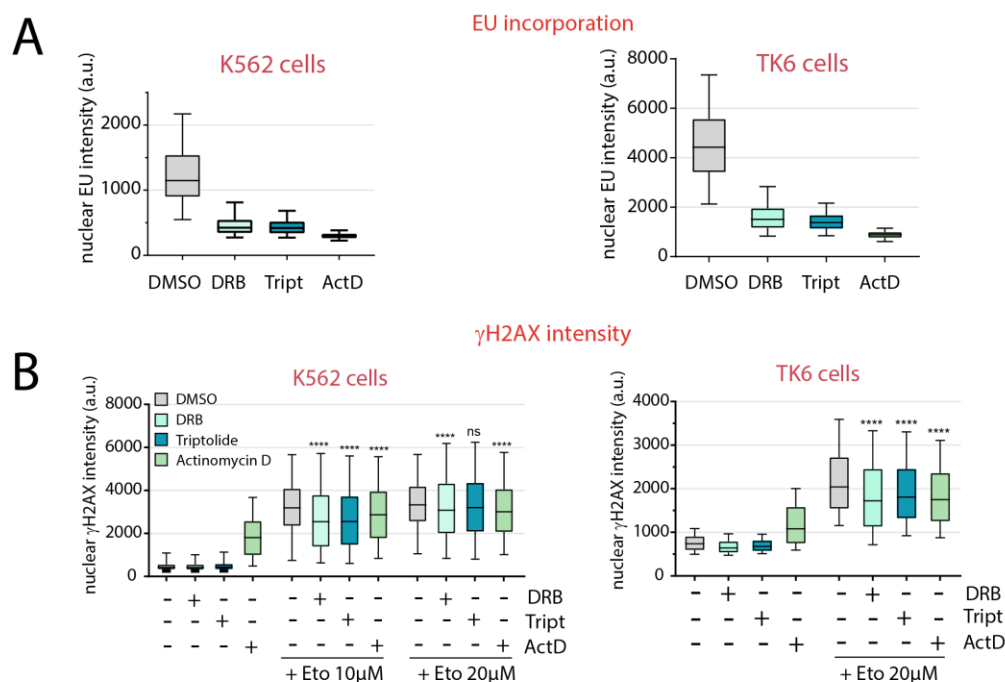
**Figure 2.11 Spatial proximity of *MLL* and *ENL* genes.** (A) Example images for spatial positioning of *MLL* (green) and fusion partner genes (magenta) in TK6 cells, labelled by FISH. The scale bar corresponds to 10  $\mu$ m. (B, C) Distribution of 3D Euclidean distances or percentage of cells with synapsis of *MLL* with fusion partner genes in TK6 and CD34+ bone marrow cells. For TK6 cells, synapsis rates were quantified in untreated and ETO-treated cells (20  $\mu$ M for 4 h, no release). Distances between *ALK* and *NPM* are shown as control. Values represent means  $\pm$  SD from at least two independent experiments (3700 – 6900 cells analyzed per sample). \*\* $p < 0.01$ ; student's t-test.

## 2.2 Transcription and spatial chromosome folding jointly induce genomic instability

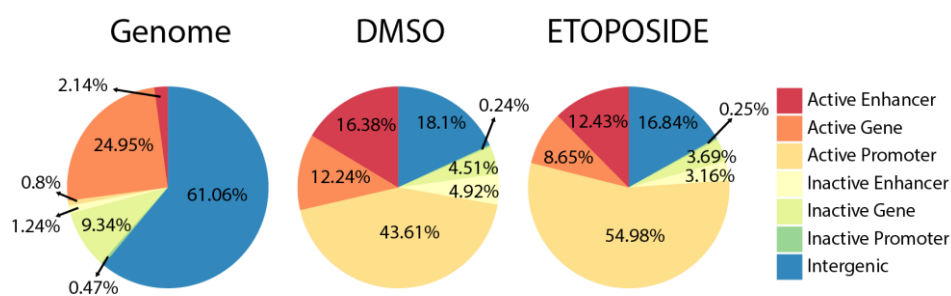
### 2.2.1 ETO-induced fragility is linked to transcriptional activity

Using sBLISS, we found that the majority of ETO-induced DSBs lie within promoters and highly overlap with RNA-PolII occupancy and open chromatin regions. To further study the role of transcription in the conversion of TOP2ccs to DSBs, we chemically inhibited transcription by different inhibitors and analyzed the ETO-induced DNA damage response in TK6 and K562 cells (Figure 2.12A). Triptolide was used to block initiation of RNA-PolII and DRB or Actinomycin D were used to inhibited transcription elongation (Bensaude, 2011). After a short pretreatment with the inhibitors, cells were acutely treated with etoposide and levels of the DNA damage response were quantified by evaluating levels of  $\gamma$ H2AX signaling. Reduced  $\gamma$ H2AX intensities were found by immunofluorescence analysis (IF) upon transcription inhibition, indicating that ETO-induced DNA damage was partially transcription-dependent (Figure 2.12B).

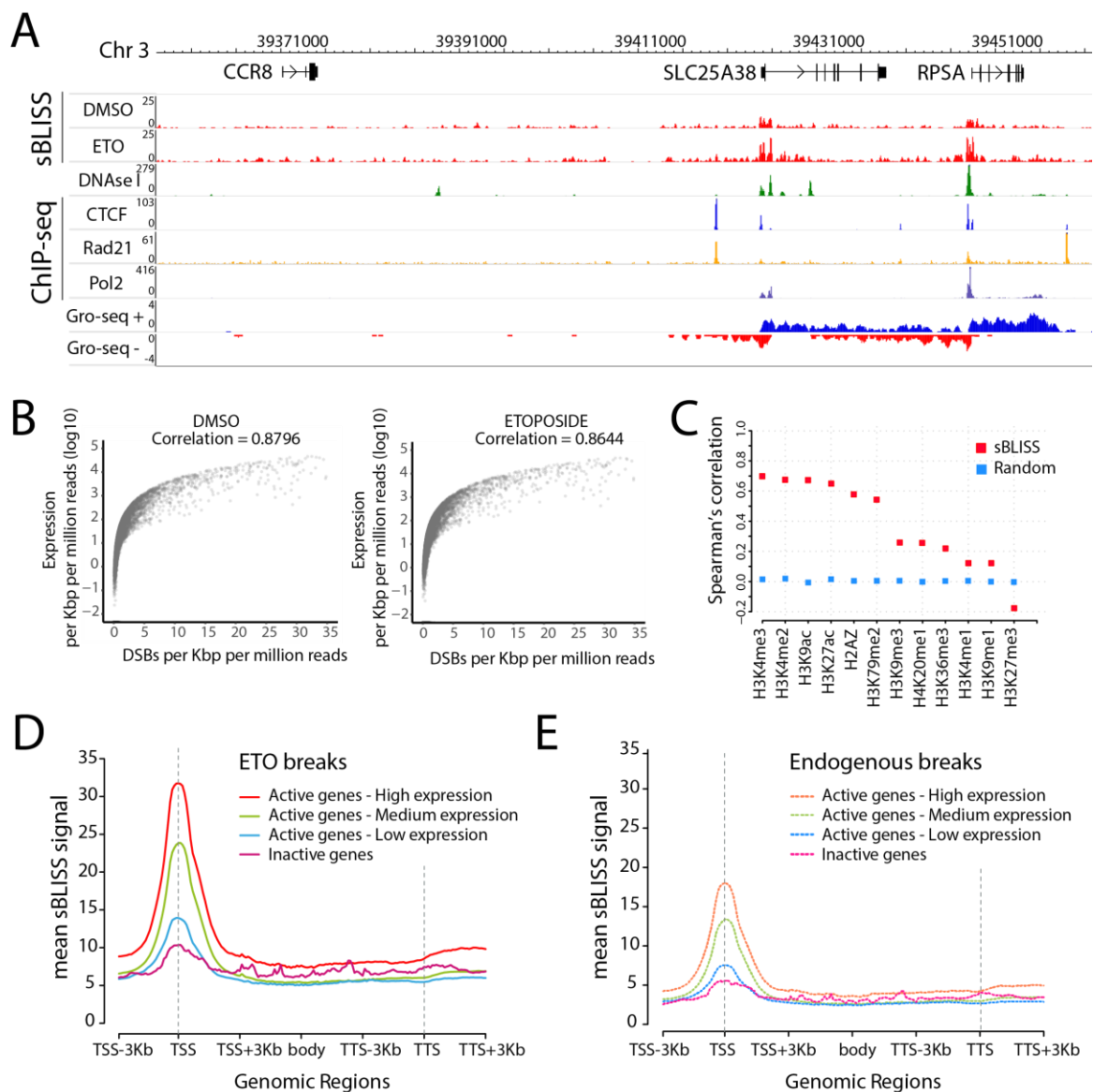
A further indication for a role of transcription in the generation of DSBs upon ETO was that more than 75 % of ETO-induced hotspots were found within active genomic regions, mainly within active promoters but also in active gene bodies and enhancers (Figure 2.13). Alongside, only 7 % of DSB hotspots were identified in inactive regions. Genic regions that produce nascent RNAs, and thus are actively transcribing, can be identified by GRO-Seq analysis. We found that active genes, marked by GRO-Seq signals, accumulated more DSBs than inactive genes (example genomic region in Figure 2.14A). When looking at a global connection of gene activity and fragility, we saw that endogenous as well as ETO-induced DSBs detected by sBLISS highly correlated with expression levels throughout the genome (Figure 2.14B,  $Sc > 0.86$ ,  $p < 10^{-5}$ ). This indicates that active genomic regions are naturally more prone to DNA fragility, and fragility can further be enhanced by etoposide. Moreover, fragile sites were highly enriched with histone modifications that mark active chromatin such as H3K4me3 methylation or H3K9 and H3K27 acetylation (Figure 2.14C,  $Sc = 0.69, 0.67, 0.64$ , respectively). Contrary, H3K27me3, which marks heterochromatin, was depleted from DSB hotspots. We then categorized genes into classes of transcriptional activity based on their GRO-Seq signals and for inactive genes the absence of RNA-PolII at promoters and lack of active histone marks (H3K4me3, H3K27ac). Genes of high transcriptional activity accumulated more DSBs spontaneously as well as upon ETO-treatment. Peaks of activity driven fragility were found at promoters but also within gene bodies and termination sites (Figure 2.14D, E).



**Figure 2.12 Transcription inhibition reduces ETO-induced DNA damage signaling.** (A) Nascent RNA expression was tested by 5-Ethynyl Uridine (EU) incorporation in K562 (left) or TK6 (right) cells pretreated with DRB (200  $\mu$ M), Triptolide (Tript, 1  $\mu$ M) or Actinomycin D (ActD, 200 nM) for 3 h. (B) Fluorescence intensities of  $\gamma$ H2AX were quantified in K562 or TK6 cells upon transcription inhibition by 3 h pretreatment (doses as in A) and damage induction with indicated doses of ETO (2 h in K562 cells and 4 h in TK6 cells). Approximately 1000 cells were quantified per condition by high-throughput microscopy and image analysis. \*\*\*\* $p$  < 0.0001 significance tested by One Way Anova and Tukey test, comparison to ETO controls.

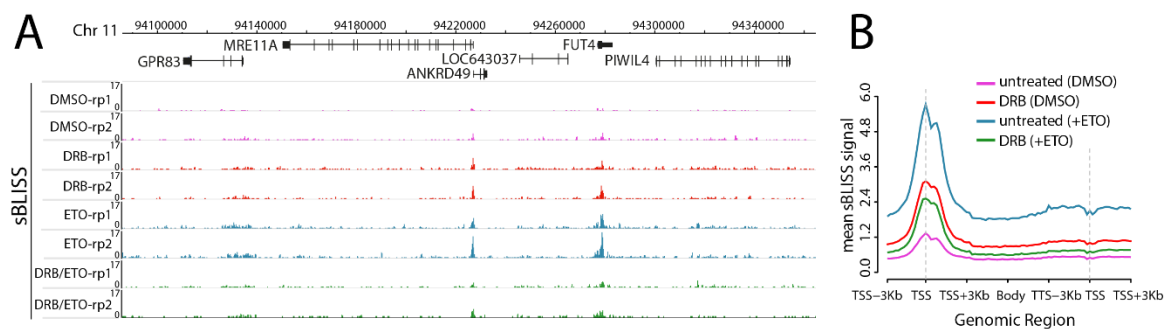


**Figure 2.13 ETO-induced DSBs cluster within active genomic regions.** Pie chart of the distribution of active, inactive or intergenic features within the genome (left) and distribution of spontaneous (DMSO) or ETO-induced sBLISS hotspots in these features in K562 cells (middle, right).



**Figure 2.14 Gene fragility correlates with gene activity.** (A) Representative genomic location of DSBs (sBLISS of K562 cells) in active or inactive features. Nascent RNA expression is indicated by GRO-seq signals, open chromatin is marked by DNaseI-HS, binding of CTCF, RAD21 or RNA-PolIII is represented by ChIP-seq tracks. (B) Correlation of gene activity (total GRO-seq reads RPKM per gene) plotted against DSBs (sBLISS reads in gene body  $\pm$ 2 kb promoter region) in K562 cells. Spearman's correlation coefficient is indicated for endogenous (DMSO, left) or ETO-induced DSBs (20  $\mu$ M for 6 h, right). (C) Spearman's correlation coefficient between ETO-induced DSBs in K562 cells and indicated histone marks. As a control, DSBs were correlated with of a set of randomly shuffled regions of histone modifications (blue). (D, E) Aggregate plots of (D) ETO-induced DSBs or (E) endogenous DSBs in K562 cells across genes grouped by expression. Genes were classified in highly expressed genes (GRO-seq RPKM count's percentile > 90 %), intermediately expressed genes (between 50 % and 90 %), lowly expressed genes (< 50 %) and inactive genes (< 50 % and absence of H3K4me3, H3K27ac and RNA-PolIII from promoter regions). Meta-gene plots from TSS + 3 Kbp to TTS - 3 Kbp with bins regularized to span the entire gene in equal portions are shown.

To show that transcription directly contributes to ETO-induced DSB formation, we performed sBLISS experiments in TK6 cells that were pretreated with DRB to inhibit transcription elongation, before ETO-treatments (Figure 2.15A). Indeed, transcription inhibition reduced the overall levels of DSBs detected by sBLISS upon ETO-treatment in TK6 cells (Figure 2.15B). Within genes, reductions in breakage were seen at promoters, gene bodies and around termination sites (Figure 2.15B). Together, our findings point out a strong correlation of gene activity with endogenous and ETO-induced gene fragility and establish transcription as a key driver of TOP2-induced DSBs.



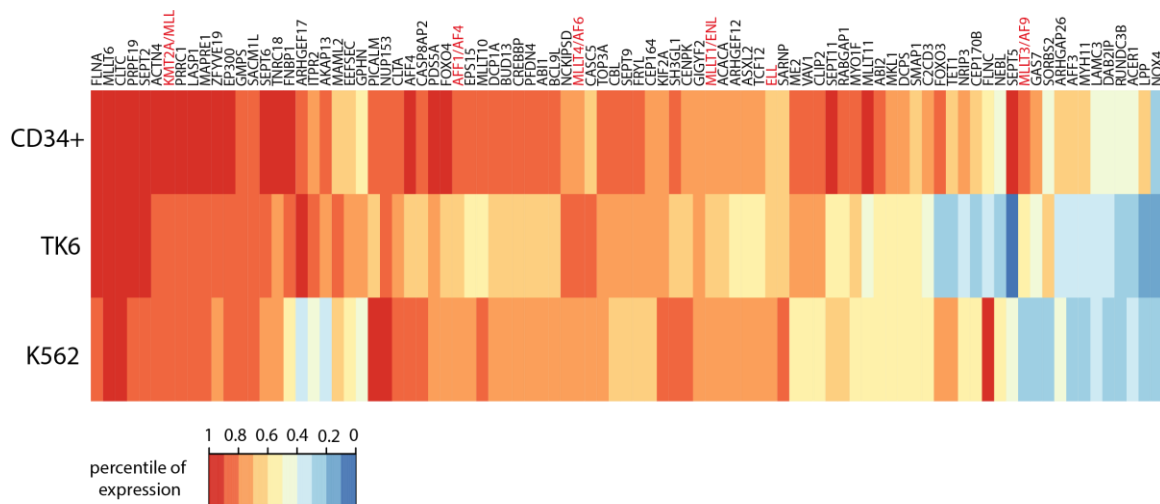
**Figure 2.15 Transcription inhibition reduces DSBs detected by sBLISS.** (A) Genomic location with sBLISS profiles of TK6 cells pretreated with DRB (200  $\mu$ M, 3 h) prior to ETO-treatment (30  $\mu$ M, 4 h). (B) Aggregate plot of the average DSBs per kilobase pair across a metagene TSS to TTS  $\pm$  3 kb of all genes in TK6 cells. Transcription inhibition with DRB and ETO-treatment as in (A).

### 2.2.2 *MLL* and frequent partner genes are highly active and enriched at chromatin loop boundaries

We sought to further investigate if DNA fragility mediated by transcriptional activity is also a feature of genes that are recurrently found fused in *MLL* translocations: the *MLL* recombinome (Meyer et al., 2013, 2018). Transcriptome analysis of TK6, K562 and CD34+ cells revealed that the majority of the recurrent translocation partner genes were among the most highly expressed genes in these hematopoietic cells (Figure 2.16). Similar expression patterns between the hematopoietic cell lines and progenitor cells were seen for many genes, indicating that their expression might be cell type specific. However, the most frequent *MLL* fusion partner genes that were identified from patients (such as *AF4*, *AF9* etc. highlighted in red, Figure 2.16) were not necessarily the most highly expressed genes, suggesting that high expression is not the only determinant for translocation partners choice.

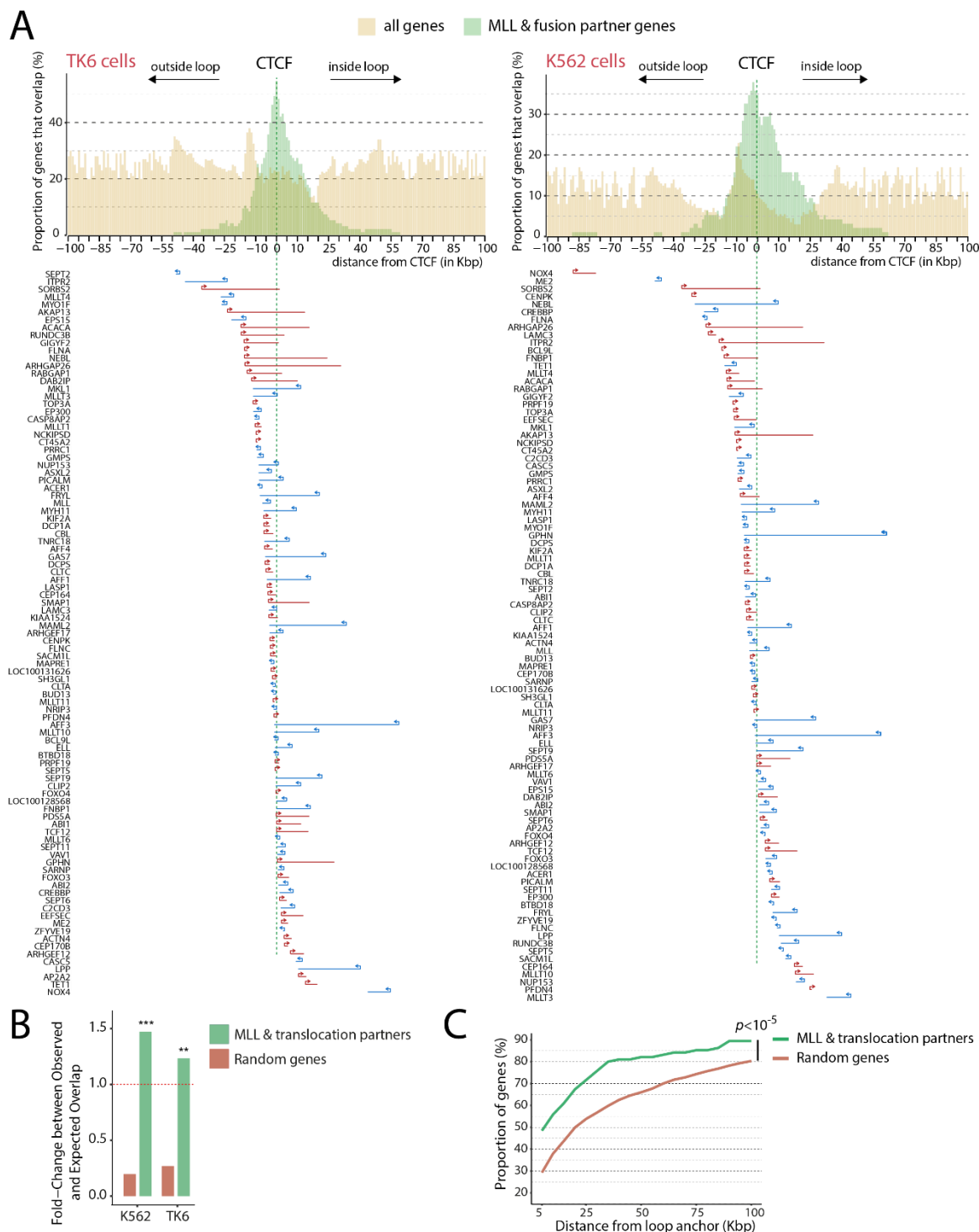
Since we initially found a substantial overlap of DSB hotspots with CTCF and cohesin binding sites (Figure 2.2), we further asked if frequent *MLL* fusion genes might also be biased towards

being at chromosome architecture features. Strikingly, *MLL* and most fusion partner genes were enriched around CTCF bound chromatin loop anchor sites (Figure 2.17A). We found that this is not a general feature of genes as the overlap of translocation partners genes with loop boundaries was much higher than of a random set of genes (Figure 2.17B). Furthermore, the proportion genes that were close to loop anchors was significantly higher for *MLL* fusion partners than for a control gene set, e.g. 50 % of recurrent fusion partners lay within 5kb around loop anchor sites whereas on average of 30 % of genes fall into this range (Figure 2.17C). We thus conclude that recurrently translocating genes that form oncogenic *MLL* fusions are highly expressed and are enriched at chromatin loop boundaries in hematopoietic cells.



**Figure 2.16 *MLL* and recurrent fusion partner genes are highly expressed in hematopoietic cell lines and CD34+ progenitor cells.** Heatmap of expression values of the *MLL* and fusion partners genes (Meyer et al., 2018), given as percentiles of the genome-wide expression in K562 cells (GRO-seq), TK6 cells (nascent RNA) and in progenitor CD34+ cells (RNA-seq).





**Figure 2.17 *MLL* and fusion partner genes are highly enriched at chromatin loop boundaries.**

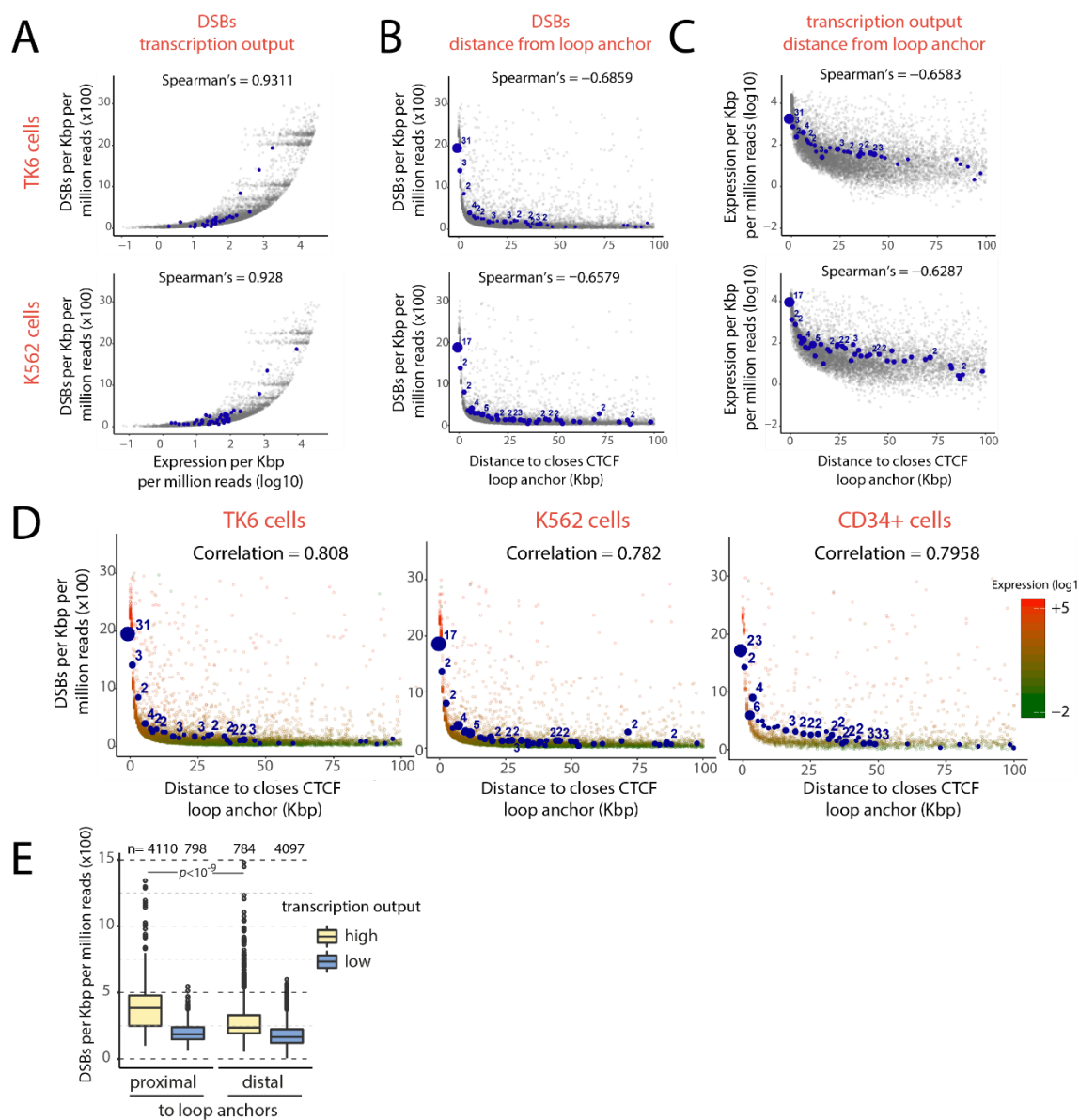
(A) Localization of *MLL* fusion genes and around CTCF marked loop anchors. Top: histograms of the proportion of overlapping gene bodies of control genes (all genes in the genome; yellow) or *MLL* and fusion partner genes (green) with loop anchor sites marked by CTCF, in TK6 and K562 cells. Distances were derived from genes' TSSs placed next to the closest loop anchor. Bottom: visual depiction of the orientation of *MLL* and fusion partners' genes around the centered CTCF position. Arrows on gene bodies show transcription orientation. (B) Fold-change between the expected and observed overlap of a random gene set (brown), or *MLL* partners gene set (green) with CTCF bound loop anchors in TK6 and K562 cells. [legend continued on next page]

(C) Proportion of genes of a random gene set (brown) or the *MLL* and translocation partners genes (green) found within the given distances (Kbp) from loop anchors. Significance was determined by Mann-Whitney-Wilcoxon test.

### 2.2.3 Transcriptional activity, localization at loop anchors and TOP2-induced breakage are interlinked

We intriguingly find that genes frequently involved in *MLL* fusions display high transcriptional activity and are likely positioned close to chromosome architecture features. To investigate if TOP2-induced gene fragility, expression and genome architecture are globally interlinked, we calculated genome-wide correlations between these features in TK6 and K562 cells. As previously observed (Figure 2.14E), we report a very strong correlation of DSBs with transcriptional output (Figure 2.18A). The higher the expression of genes, the more DSBs they tend to accumulate. Furthermore, we found a global inverse correlation of DSBs with the distance of genes from chromatin loop anchors, indicating that genes close to loop anchors are more fragile upon etoposide treatment (Figure 2.18B, TK6 cells:  $Sc = -0.69$ , K562 cells:  $Sc = -0.66$ ). We also noticed that transcription output and genes' distances from loop anchors inversely correlate with similar strength (Figure 2.18C, TK6 cells:  $Sc = -0.66$ , K562 cells:  $Sc = -0.63$ ). When combining these dependencies, we found a high triple correlation of TOP2-induced DSBs in genes with expression output and proximity to loop anchors in TK6, K562 and CD34+ progenitor cells (Figure 2.18D, TK6 cells:  $Sc = 0.81$ , K562 cells:  $Sc = 0.78$ , CD34+ cells:  $Sc = 0.80$ ). Many *MLL* fusion partner genes share all three features and exhibit high fragility upon poisoning of TOP2s (marked as blue spots, Figure 2.18D). These findings could suggest that high transcriptional activity at chromatin loop boundaries may require the activity of TOP2s, which in turn can result in higher gene fragility upon abortive TOP2cc formation. Because active genes generally localized closer to chromatin loop boundaries (Figure 2.18C), it could also be hypothesized that the positioning of genes at loop anchors does not actively contribute to the formation of DSBs, but that transcription is the driving force of TOP2-induced DSBs. To test this hypothesis, we compared TOP2-induced breakage in highly or lowly expressing genes that are either proximal or distal to loop anchor sites. Highly transcribed genes, regardless of their positioning, were more prone to accumulate DSBs than less active genes, showing the contribution of transcriptional activity to TOP2-induced DNA damage (Figure 2.18E). However, highly active genes that were additionally located proximal to loop anchors showed significantly more breaks than distal active genes ( $p < 10^{-9}$ ). Inactive genes at loop anchors, on the other hand, accumulated equally few breaks compared to inactive but distal genes. This argues that positioning of genes at loop anchors

contributes to TOP2-induced gene fragility, but predominantly in genes that are also transcriptionally active.

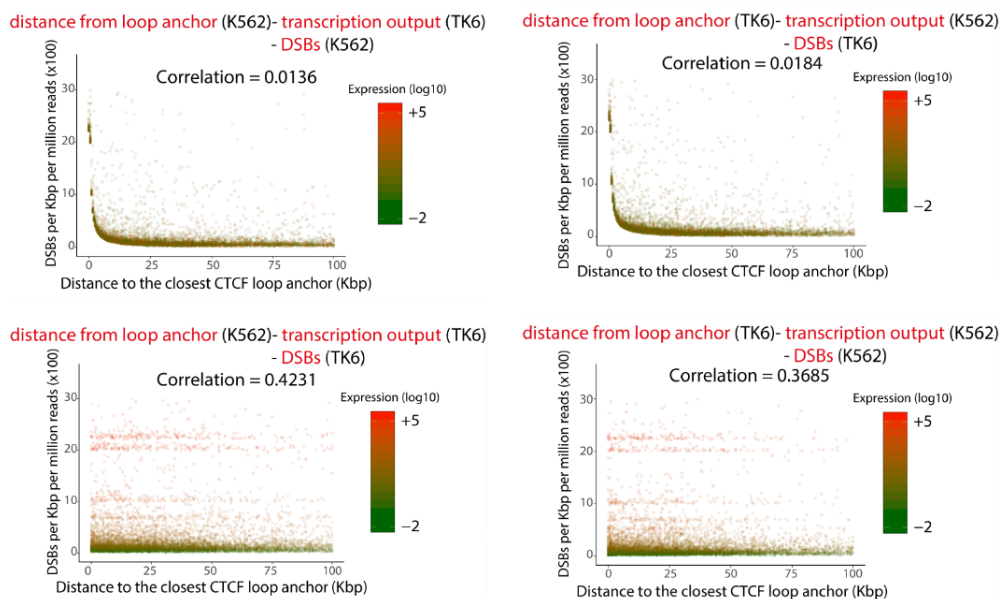


**Figure 2.18 Transcription output, proximity to loop anchors and gene fragility are interlinked.**

(A) Correlations between ETO-induced DSBs per gene (sBLISS) and transcription output of genes, (B) between DSBs and genes' distances from the closest loop anchor or (C) between transcription output of genes and distances from closest loop anchors, for TK6 (upper) and K562 cells (lower). Distances were determined from the closest gene end (TSS or TTS) to the middle of the closest loop anchor and rounded up to the nearest Kbp. *MLL* and fusion partner genes are shown as blue circles and the size and numbers correspond to the number of overlapping genes. Correlation coefficients are indicated above the graphs. (D) Triple correlations of ETO-induced DNA fragility of genes (y-axis) as a function of distance from their closest loop anchor (x-axis), and transcription output levels (color gradient), in TK6, K562 and CD34+ progenitor cells. Distances to loop anchors were determined as in (B, C). (E) DSB levels of genes that are either proximal or distal to loop anchors (less than or greater than 50<sup>th</sup> percentile of distances to the closes anchor) and that have a high or low transcription output (yellow or blue, less than or greater than 50<sup>th</sup> percentile of expression).

## Results

To test for a cell line specific linkage of gene fragility, expression and chromatin architecture, we performed cross-cell line correlations of the transcription output data, TOP2-induced DSBs and distances to loop anchors of TK6 and K562 cells. High triple correlations between these features were lost when for example sBLISS data and looping data of K562 cells were correlated with expression data from TK6 cells (Figure 2.19). We thus suggest that the linkage of expression, chromatin architecture and gene fragility are cell line specific and largely interdependent. We further hypothesize that transcription may be the central driving force in TOP2-induced breakage at loop anchors.

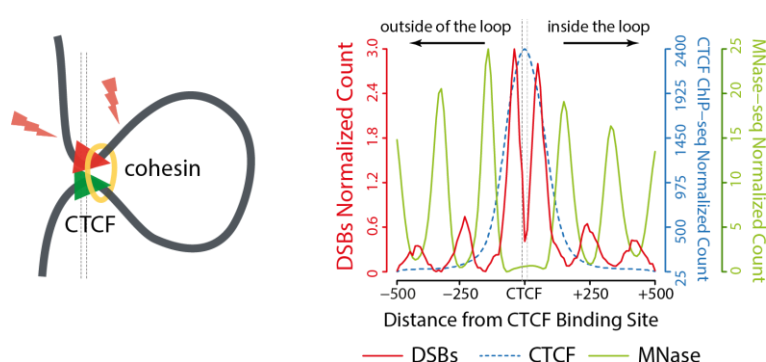


**Figure 2.19 Linkage of transcription and loop boundaries is cell line specific.** Cross-cell line correlation between genes' distances to the nearest CTCF loop anchor, transcription output and ETO-induced DSBs within genes (as determined in Figure 2.18 D). Data of these three features were interchanged between TK6 and K562 cells and correlations were determined as indicated.

### 2.2.4 DSBs at chromatin loop anchors are formed by transcriptional activity and directionality

We found an intriguing link between transcription and localization of genes at loop anchors that might predispose genes to being susceptible to DNA fragility upon TOP2cc formation. It was previously shown in mouse B-lymphocytes that stalled TOP2 complexes can be found around chromatin loop anchors and that TOP2ccs mainly localize before loop anchor sites, outside of the chromatin loop (Canela et al., 2017). We thus tested in the human hematopoietic cell line K562 if free DSBs, as detected by sBLISS, also show a bias towards

looping directionality. The zinc-finger protein CTCF marks boundaries of the majority of chromatin loops by binding to convergent DNA consensus motifs. We thus analyzed DSB signals genome-wide around G-rich, oriented CTCF binding sites, which allowed us to determine what is outside and what is inside of chromatin loops. As observed previously (Canela et al., 2017), DSBs were enriched in nucleosome free regions around the strongly positioned nucleosomes flanking CTCF binding sites (marked by MNase-Seq in green, Figure 2.20). Furthermore, DSBs peaked on both sides around CTCF motives at a mean distance of 45 nt from motif centers, independent of the looping directionality (Figure 2.20). We thus could not find a bias of free DSBs towards chromatin looping directionality.



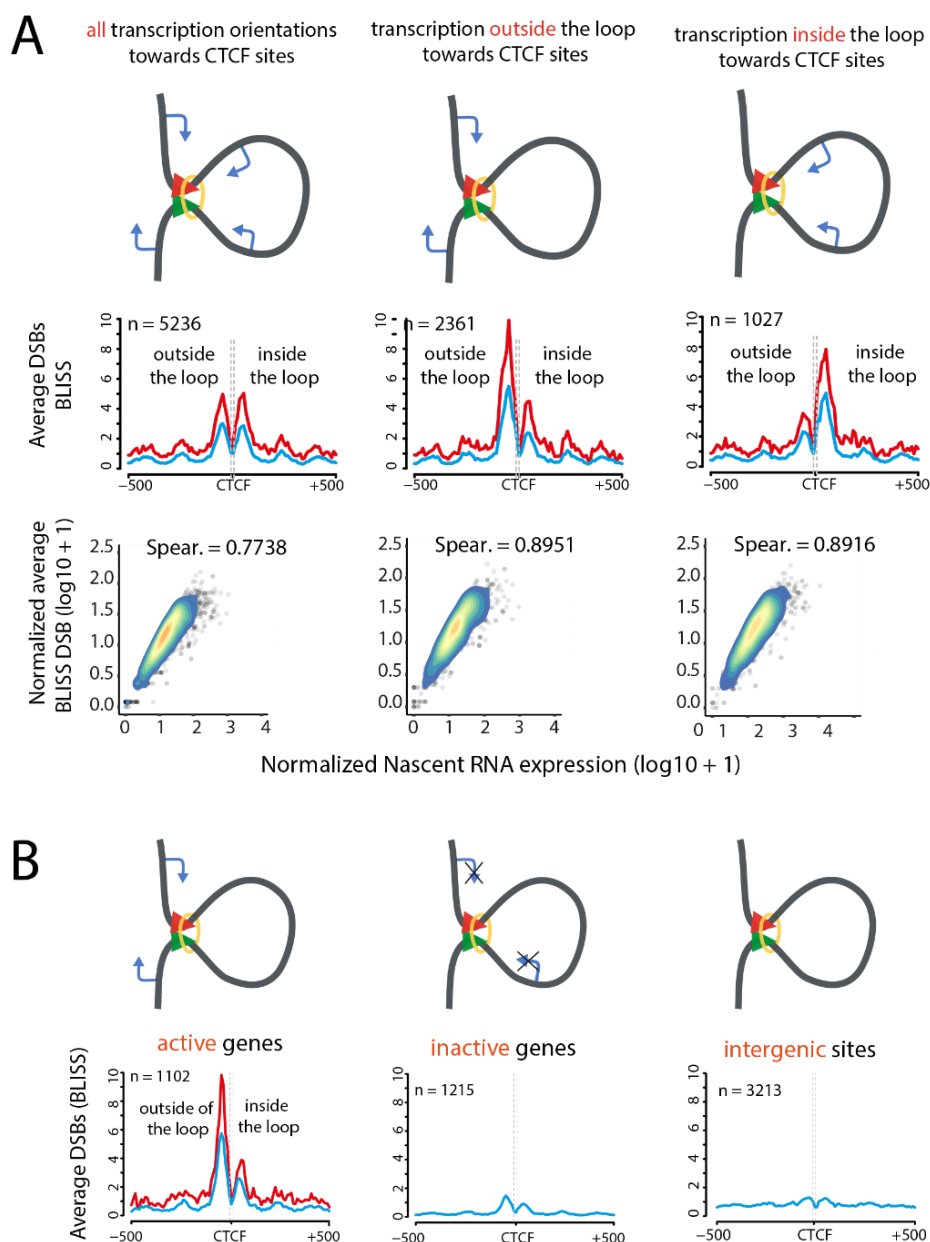
**Figure 2.20 ETO-induced DSBs localize in nucleosome free regions around chromatin loop boundaries.** Aggregate plot of ETO-induced DSBs in K562 cells (red) centered  $\pm 500$  bp around CTCF motives at chromatin loop boundaries (grey dashed lines), aligned on G-rich, oriented CTCF motifs to keep the directionality of chromatin loops. CTCF binding (ChIP-seq signal as blue dashed line) and nucleosome occupancy (MNase-seq, green) were also superimposed.

Next, we explored if the linkage of chromatin architecture with transcription to the formation of DSBs might also be important for positioning of DSB at loop anchors. We thus aligned intragenic regions around oriented CTCF motives at loop anchors as before but also took into account the genes' transcriptional activity and directionality. Within active genes, when transcription was oriented towards the loop anchors from inside and outside of the loop, DSBs localized on both sides of loop anchors (Figure 2.21A). However, for genes where transcription was solely coming from outside of loops towards loop anchors, DSBs mainly clustered outside of the chromatin loop, just 45 nt before the CTCF site. Vice versa, in genes where transcription was directed towards anchors from within loops, DSBs accumulated inside of loop anchors, again just 45 nt before the CTCF site (Figure 2.21A). This analysis indicates that the directionality of transcription determined the positioning of DSBs at loop anchors. The influence of transcription directionality on DSBs localization at loop anchors

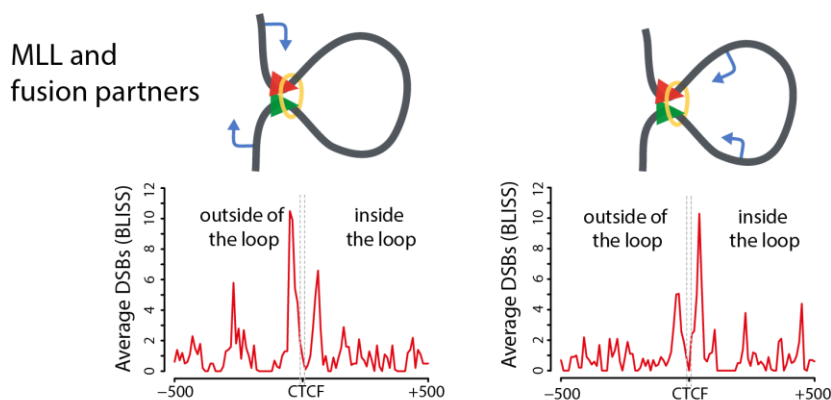
## Results

was also reflected in an increase in Spearman's correlation between gene expression and DSBs, once gene sets were ordered by their directionality of transcription (Figure 2.21A,  $S_c \sim 0.89$ ). Importantly, clustering of DSB at loop anchors was only observed in active genes. Inactive genes or intergenic sites that were located at loop boundaries only showed minimal enrichment of DSBs around CTCF sites (Figure 2.21B), showing that active transcription is required to evoke high levels of TOP2-induced DSBs at chromatin loop anchors and that the localization of break hotspots is linked to positioning of loop anchors in the vicinity of active genes.

We saw that most genes involved in *MLL* translocations had high transcriptional activity (Figure 2.16) and were found near chromatin loop boundaries (Figure 2.17). We thus asked if transcription directionality also plays a role in breakage of frequent *MLL* fusion partners. DSBs were also biased towards the side of active transcription of loop anchors in fusion partner genes, thereby influencing if breaks were found before or after a chromatin loop boundary (Figure 2.22). Hence, we hypothesize that torsional stress induced by transcription could be propagated towards loop anchors and that TOP2s dissipate these constraints. Thereby, DSBs are formed at loop anchor sites upon blocking the TOP2 catalytic cycle.



**Figure 2.21 Transcription orientation affects the localization of DSBs at chromatin loop anchors in active genes.** (A) Aggregate plots of ETO-induced DSBs in TK6 cells (middle panel, red and blue) centered  $\pm 500$  bp around CTCF-bound, oriented, intragenic chromatin loop boundaries (grey dashed lines) and grouped by transcription directionality towards loop anchors (left: from all orientations, center: from outside the chromatin loop, right: from inside the loop towards the CTCF site). Blue lines represent all aggregated DSBs of all genes and red lines only show signals of genes with expression higher than the 90<sup>th</sup> percentile. Blue arrows in schemes represent the direction of active transcription. Bottom: Scatter plots of DSBs against nascent RNA expression and Spearman's correlations are shown for sites with the respective transcription orientation. (B) Aggregate plots of ETO-induced DSBs at CTCF-bound loop anchors located within active genes (transcription from outside towards the anchor), inactive genes or intergenic sites in K562 cells.



**Figure 2.22 Transcription orientation also influences breakage of *MLL* fusion partners at loop boundaries.** Aggregate plots as in Figure 2.21 (A), but only for *MLL* and fusion partner genes located at loop anchor sites in ETO-treated K562 cells.

### 2.2.5 The *MLL* DSB hotspot is located in proximity to a chromatin loop

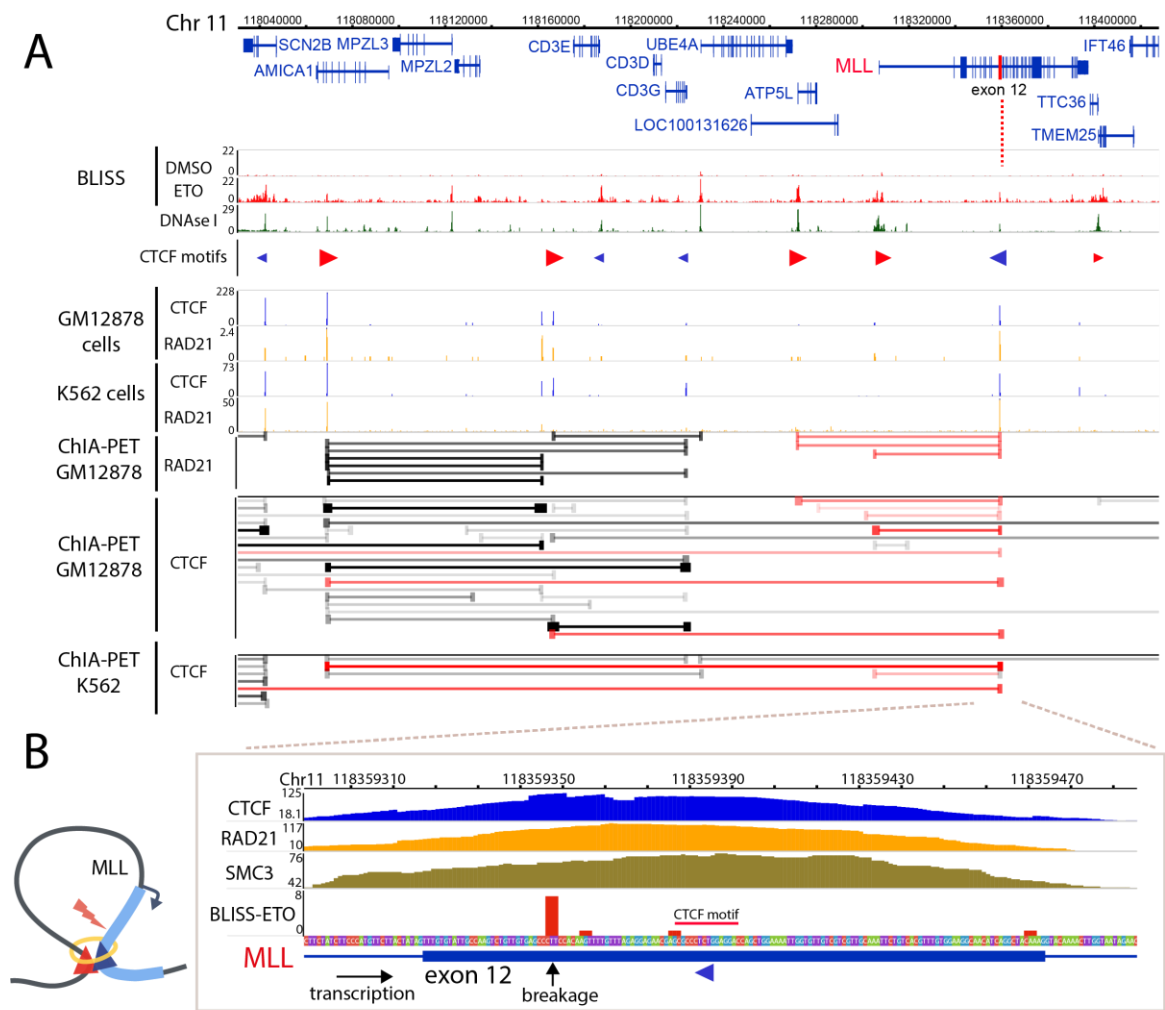
Previous analysis indicated that the *MLL* DSB hotspot located at the exon 12 translocation hotspot colocalizes with DNaseI hypersensitivity as well as CTCF and RAD21 binding (Figure 2.3). To investigate if the CTCF site within *MLL* exon 12 is marking a chromatin loop boundary, we analyzed chromatin loops that were identified by chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) of CTCF or cohesin/RAD21. ChIP-seq signals for CTCF and RAD21 in the hematopoietic cell lines K562 and GM12878 (lymphoblastoid, similar to TK6 cells) suggest that CTCF and cohesin are bound to chromatin around the *MLL* CTCF motif (Figure 2.23A, B). ChIA-PET interaction data showed that the *MLL* exon 12 was located at a chromatin loop boundary which extended into upstream genomic regions on chromosome 11 (Figure 2.23A). We then looked closer at the DSB hotspot at *MLL* exon 12 and found that DSBs localized just before the CTCF motif (Figure 2.23B). Hence, the breakage at the *MLL* exon 12 hotspot was mainly found within the chromatin loop towards the 5' end of *MLL* where transcription was coming from (Figure 2.23B). This finding further suggests that breakage at the *MLL* loop boundary might be mediated by the transcription activity of the gene.

To assess the contribution of the loop anchor within *MLL* to TOP2-induced *MLL* fragility, we engineered a cell line where the CTCF motif at exon 12 was mutated to prohibit CTCF binding (Figure 2.24A). We assumed that without CTCF binding, the formation of the chromatin loop would be disrupted. To generate an endogenous mutation within both *MLL* alleles, integration of a donor sequence via homology directed repair was utilized. First, in TK6 cells that were stably expressing Cas9, a DSB was generated at the end of the exon 12 CTCF motif (Figure 2.24B). Then, a short single-stranded homology donor with three point mutations in the CTCF

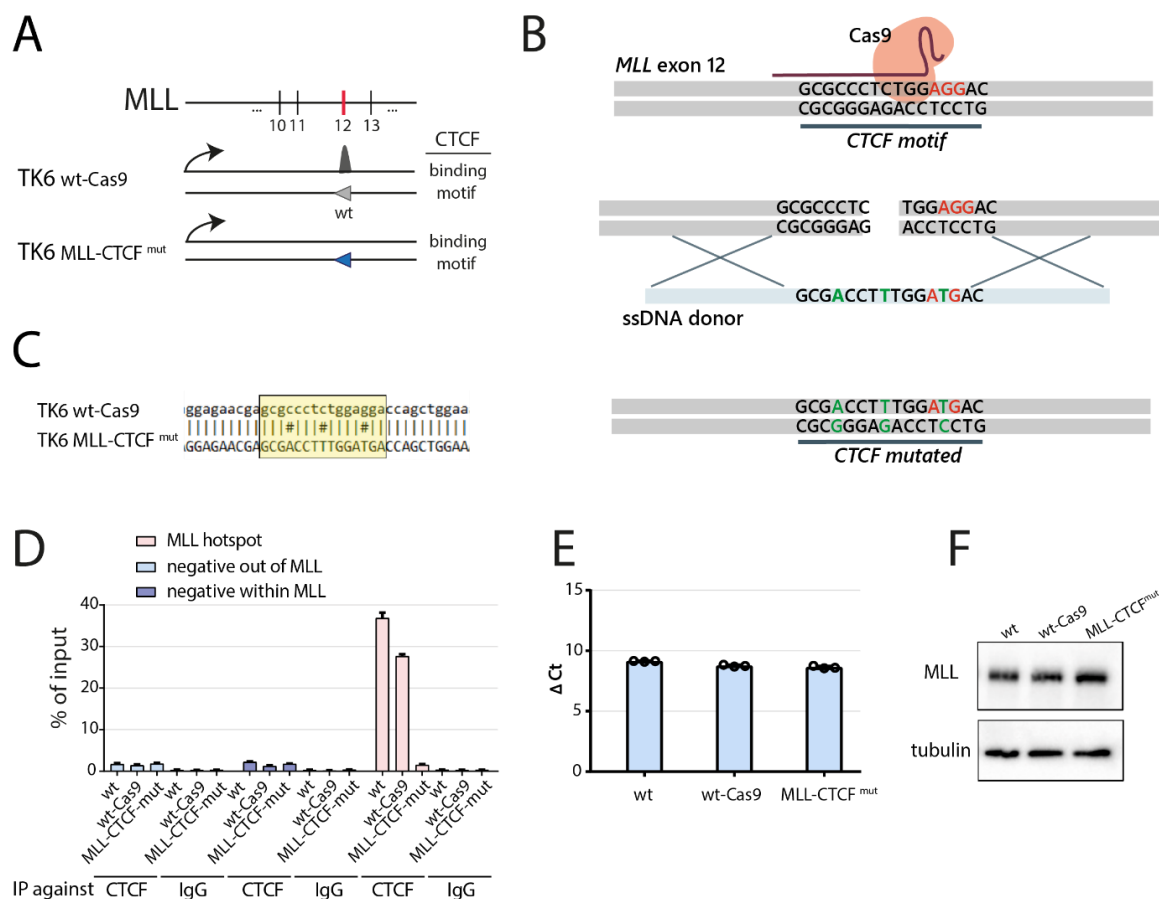


motif (green bases) served as a repair template. To prohibit Cas9 cutting of the donor oligo, the PAM sequence was also mutated in the donor (marked in red Figure 2.24B). All point mutations were designed as silent mutations in order to retain the amino acid sequence of *MLL*. One TK6 clone in which both alleles of *MLL* exon 12 were mutated was successfully generated, which we termed TK6 MLL-CTCF<sup>mut</sup> (Figure 2.24C). To validate the disruption of CTCF binding by the mutations, we performed a CHIP-qPCR assay for CTCF at *MLL* exon 12. In TK6 wt and TK6 Cas9 parental cells, CTCF binding was detected at the *MLL* hotspot at exon 12, but not in a control region within *MLL* or in an intergenic region (Figure 2.24D). For TK6 MLL-CTCF<sup>mut</sup> cells, no binding was detected at the *MLL* hotspot which confirmed that mutations of the motif prohibited CTCF binding (Figure 2.24D). We asked if the loss of CTCF at exon 12 influenced *MLL* expression and could thereby indirectly change the fragility of *MLL* by changing its transcriptional activity. To test this, we probed mRNA levels by RT-qPCR and *MLL* protein levels by western blot (Figure 2.24E, F). The TK6 MLL-CTCF<sup>mut</sup> cell line did however not show changes in expression when compared to TK6 wt or TK6 Cas9 parental cells.

Thus, the MLL-CTCF<sup>mut</sup> cell line provided us with a tool to understand the contribution of the chromatin loop within *MLL* to its TOP2-induced fragility. We probed the global DNA damage response in TK6 MLL-CTCF<sup>mut</sup> cells by IF of  $\gamma$ H2AX as well as local *MLL* breakage by C-Fusion 3D (Figure 2.25). As expected, we did not observe changes in  $\gamma$ H2AX levels after etoposide treatment in TK6 MLL-CTCF<sup>mut</sup> cells, indicating that the CTCF binding within *MLL* did not affect the overall response of cells to etoposide (Figure 2.25A). However, when probing *MLL* gene breakage by C-Fusion 3D after an acute ETO-treatment, we observed a decrease of *MLL* breakage by ~30 % (Figure 2.25B, from 4.1 % to 2.8 %,  $p < 0.01$ ). This suggested that the chromatin loop boundary at exon 12 contributes to TOP2-induced *MLL* gene fragility. On the other hand, transcription inhibition by pretreatments with DRB drastically reduced *MLL* breakage to ~1 % in parental and MLL-CTCF<sup>mut</sup> cells (Figure 2.25B,  $p < 0.0001$ ), showing that transcription is yet the predominant force to cause chromosome breaks in *MLL*.

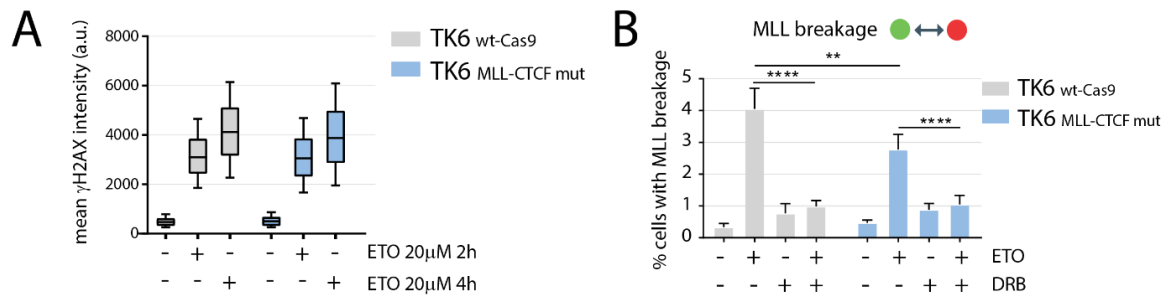


**Figure 2.23 The *MLL* exon 12 DSB hotspot is localized just before a chromatin loop boundary.** (A) The *MLL* exon 12 CTCF site (marked in red), located at the translocation hotspot, is part of a chromatin loop anchor. DSB profiles around the human *MLL* locus (TK6 cells, DMSO or ETO), CTCF motifs (G-rich, forward strand=red, reverse strand=blue) and DNaseI-HS profiles are shown. Binding sites of CTCF (blue) and Rad21 (yellow, ChIP-seq) are indicated. Chromatin loop interactions measured by CTCF or RAD21 ChIA-PET are represented by horizontal lines (pink lines indicate interaction emanating from the *MLL* hot spot with upstream genomic loci). (B) Magnification of the *MLL* translocation hot spot showing the localization of DSBs within exon 12, as well as signals for CTCF, RAD21 and SMC3 (ChIP-seq). Breakage was located between the CTCF motif (blue triangle) and the direction of *MLL* transcription (black arrow).



**Figure 2.24 Generation of a TK6 cell line with mutated CTCF motif within *MLL*.** (A) Generation of a TK6 MLL-CTCF<sup>mut</sup> cell lines in which both alleles of the *MLL* exon 12 CTCF motif were mutated to prohibit CTCF binding. (B) Scheme of endogenous mutation of the CTCF motif by homology directed repair. A DSB in the vicinity of the CTCF exon 12 motif was generated by CRISPR/Cas9 (PAM motif in red) for which a single-stranded DNA molecule (ssDNA donor) with short homology arms (~50 bp) was utilized for repair. The donor was designed to mutate the CTCF motif and the PAM sequence (green bases). (C) Sequence of a TK6 clones with the desired mutation. (D) ChIP-qPCR analysis for CTCF binding at the *MLL* hotspot (exon 12) or two control regions in TK6 wt, TK6 Cas9 parental cells and TK6 MLL-CTCF<sup>mut</sup> cells. The mean of triplicates is shown. (E) Quantitative reverse transcription PCR of *MLL* mRNA levels compared to GAPDH mRNA levels ( $\Delta$ Ct) in TK6 wt, Cas9 or MLL-CTCF<sup>mut</sup> cells, in triplicates. (F) Expression analysis of *MLL* by western blotting.

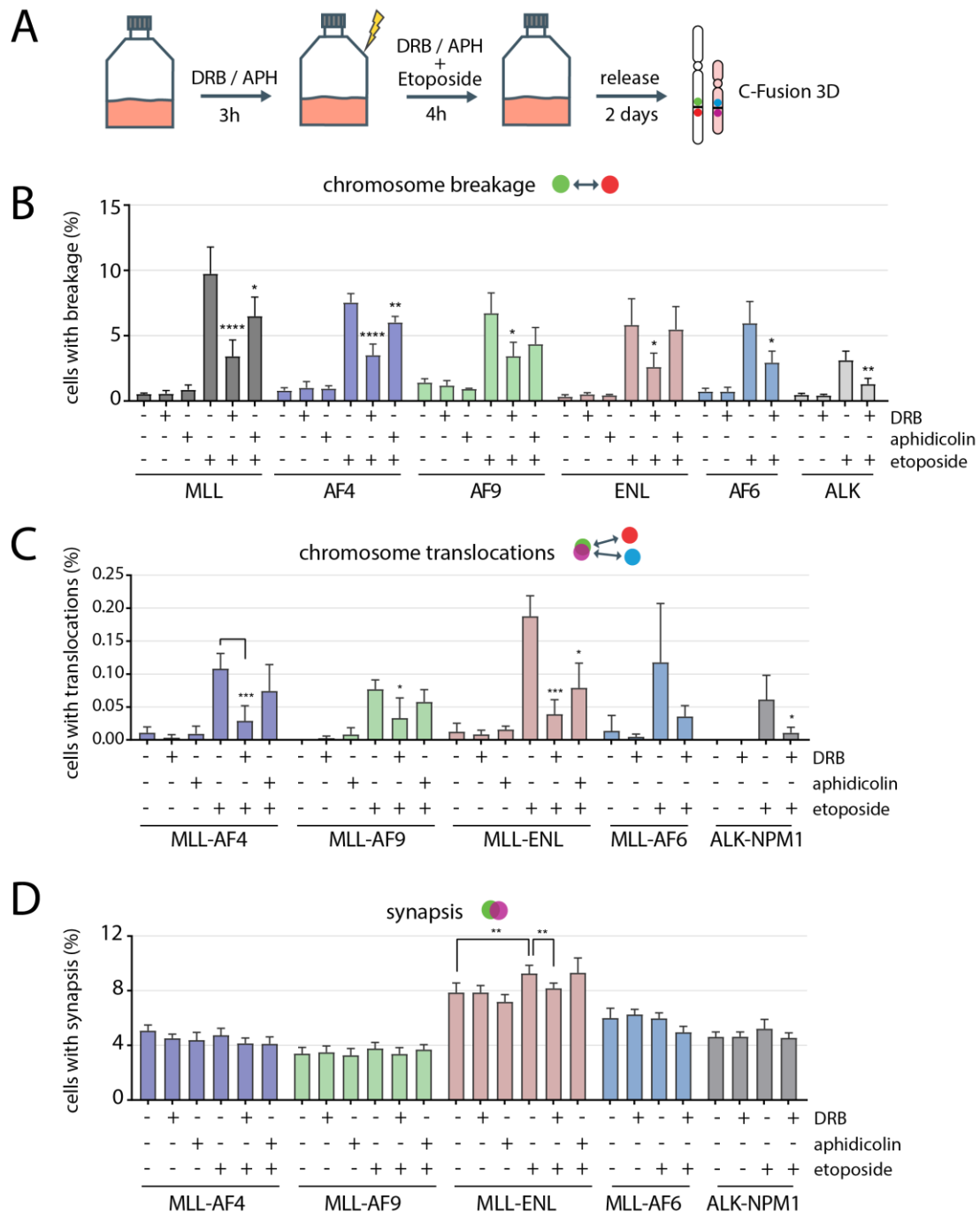
## Results



**Figure 2.25 TK6 MLL-CTCF<sup>mut</sup> cells have comparable DNA damage signaling but show reduced MLL fragility.** (A) Quantification of  $\gamma$ H2AX fluorescence intensities in TK6 Cas9 parental or MLL-CTCF<sup>mut</sup> cells upon acute treatments with 20  $\mu$ M ETO for 2 h or 4 h. (B) Percentage of TK6 cells with *MLL* breakage upon pretreatment with DRB (200  $\mu$ M for 3 h) and an acute ETO-treatment (20  $\mu$ M for 4 h), quantified by C-Fusion 3D. Mean of four independent experiments with SD is shown; 2100-11000 cells per sample were analyzed; \*\*p < 0.01 and \*\*\*\*p < 0.0001 in student's t-test.

### 2.2.6 Formation of *MLL* translocations is mainly transcription-dependent

To understand to which extent *MLL* breaks and translocations are transcription or replication-dependent, TK6 cells were treated with DRB to inhibit transcription elongation, or with aphidicolin (APH) to block replication. After a short pretreatment with the inhibitors and an acute ETO treatment in the presence of the inhibitors, cells were released for two days from all treatments and frequencies of rearrangements were tested by C-Fusion 3D (Figure 2.26A). We observed a highly significant decrease of *MLL* and common fusion partners (*AF4*, *AF9*, *ENL*, *AF6*) fragility upon transcriptional inhibition (Figure 2.26B, C, *MLL* from 9.7 % to 3.4 %, other genes by  $\sim$  two-fold). Upon replication inhibition, slight reductions in breakage were seen but were only significantly decreased for *MLL* and *AF4*. As observed previously (Figure 2.11), fusions of *MLL* and *ENL* occurred more frequently than fusions with the other fusion genes, which could again be explained by a higher synapsis rate, that was increased by ETO-treatment (Figure 2.26D). Colleagues had previously proposed that colocalization of *MLL* and fusion partners is partially mediated by the sharing of transcription factories (Cowell et al., 2012). To test this, we analyzed synapsis rates in cells upon transcription inhibition by DRB. However, we could not observe a reduction in gene synapsis upon transcriptional inhibition, showing that transcription is not a prerequisite for *MLL* and its recurrent fusion partners to colocalize in the nucleus (Figure 2.26D). Interestingly, the increase in *MLL* and *ENL* synapsis from 7.9 % to 9.3 % upon ETO-treatment was alleviated to control levels upon transcriptional inhibition, indicating that transcription-dependent DNA damage favors the spatial proximity of *MLL* and *ENL* (Figure 2.26D).

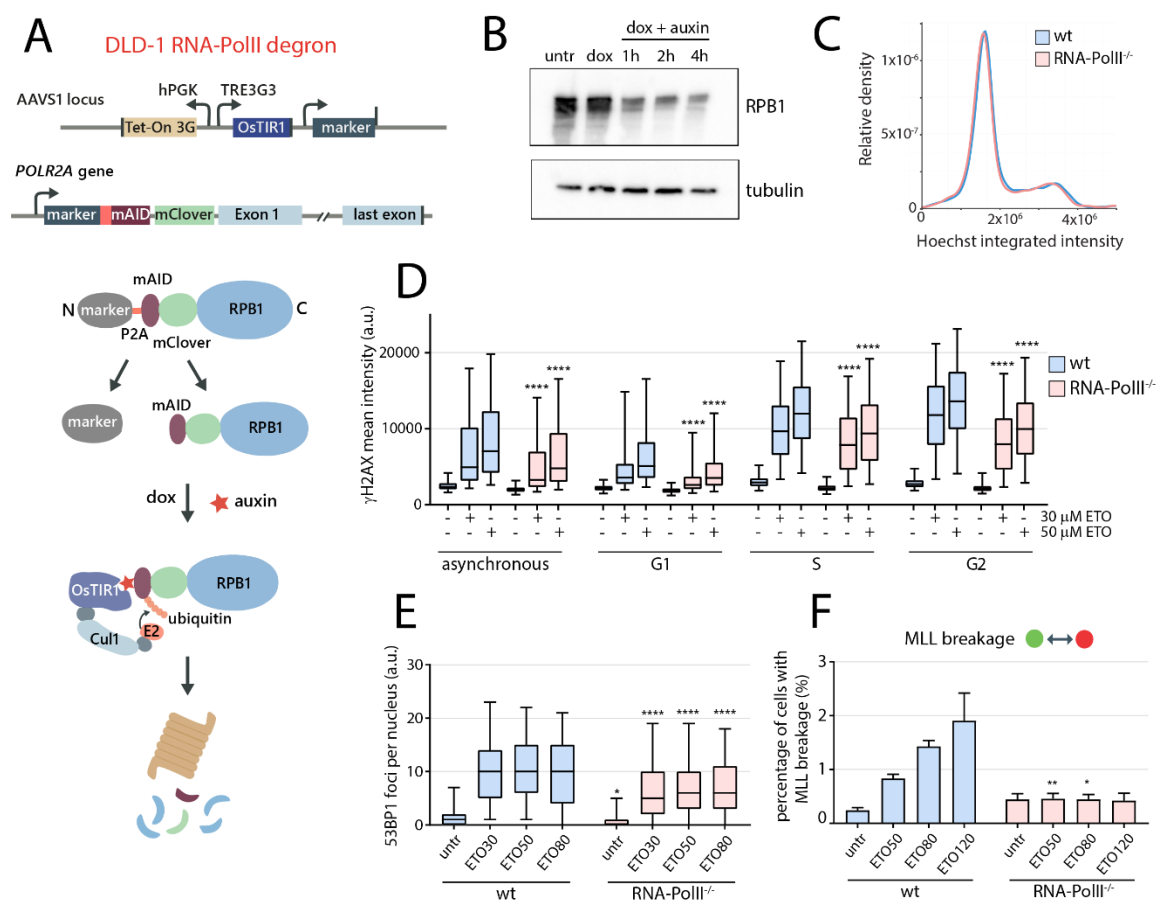


**Figure 2.26 MLL breakage and translocation formation is largely transcription-dependent.** (A) TK6 cells were pretreated with DRB (200  $\mu$ M) for transcription inhibition or APH (8  $\mu$ M) for replication inhibition for 3h and were subsequently treated with ETO (20  $\mu$ M for 4 h) and released for two day. (B) The percentage of cells with chromosome breakage, (C) chromosome translocations or (D) gene synapsis of indicated genes were assessed by 4-color C-Fusion 3D. Shown are means + SD of at least three independent experiments (4000 to 16500 cells per sample). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$  by student's t-test compared to respective ETO only sample or as indicated.

### 2.2.7 *MLL* gene fragility is directly linked to its expression

Chemical inhibition for transcription deactivation is a widely used tool to easily study the importance of transcription. However, inhibitors can affect several targets non-specifically or may have targets with multivalent cellular functions. The here used inhibitor DRB mainly targets the cyclin-dependent kinase 9 (CDK9) which is part of the Positive Transcription Elongation Factor (P-TEFb) complex and is required to release paused RNA-PolII to enable transcription elongation (Bensaude, 2011). Nevertheless, DRB also inhibits CDK7, which is required for transcriptional elongation, but has additional functions in cell cycle progression and phosphorylation of many targets, e.g. of p53 (Fisher, 2005).

We thus employed two strategies to exclude non-specific effects by chemical inhibition of transcription and to show that it is indeed the activity of transcription that facilitates TOP2-induced chromosome rearrangements. First, we utilized a recently developed RNA-PolII degron cell model in which the large subunit RPB1 can be degraded within hours (Nagashima et al., 2019). Secondly, we specially suppressed transcription of a single gene, here *MLL*, to circumvent pleiotropic effects of global transcription inhibition (Figure 2.29). To generate an RNA-PolII degron cell line, Ryosuke Nagashima and colleagues stably introduced the hormone receptor Transport inhibitor response 1 of the rice plant *Oryza sativa* (OsTIR1) into the AAVS1 locus in DLD-1 cells, under the control of a doxycycline inducible promoter (Figure 2.27A) (Nagashima et al., 2019). Then, the N-terminus of the large subunit of RPB1 (gene *POLR2A*) was endogenously tagged with a mAID-mClover cassette via homology directed repair. Upon addition of the plant hormone auxin, OsTIR1 facilitates polyubiquitination of RPB1 by the E3 ubiquitin ligase SCF complex and thereby targets it for degradation by the proteasome (Figure 2.27A) (Natsume et al., 2016; Yesbolatova et al., 2019). This is a rapid process where more than 80 % of RPB1 was degraded upon two hours of auxin treatment (Figure 2.27B). We first tested if the removal of the RNA-PolII large subunit would influence the cell cycle progression. After 4 h of auxin treatment, we could not observe cell cycle changes in DLD-1 cells (Figure 2.27C). To examine if depletion or RNA-PolII is sufficient to reduced TOP2-induced DNA damage, we analyzed levels of  $\gamma$ H2AX intensities and formation of 53BP1 foci in DLD-1 degron cells upon ETO-treatment. Degradation of RPB1 by auxin followed by ETO-treatment resulted in decreased  $\gamma$ H2AX intensities in all cell cycle phases (Figure 2.27D). Furthermore, a decrease in 53BP1 foci formation was observed (Figure 2.27E). This suggested that RNA-PolII is required for TOP2-induced DNA damage. To test the contribution of RNA-PolII activity to *MLL* fragility, *MLL* breakage was quantified in DLD-1 degron cells by C-Fusion 3D. Importantly, the depletion of RNA-PolII led to a complete loss of *MLL* breakage (Figure 2.27F).

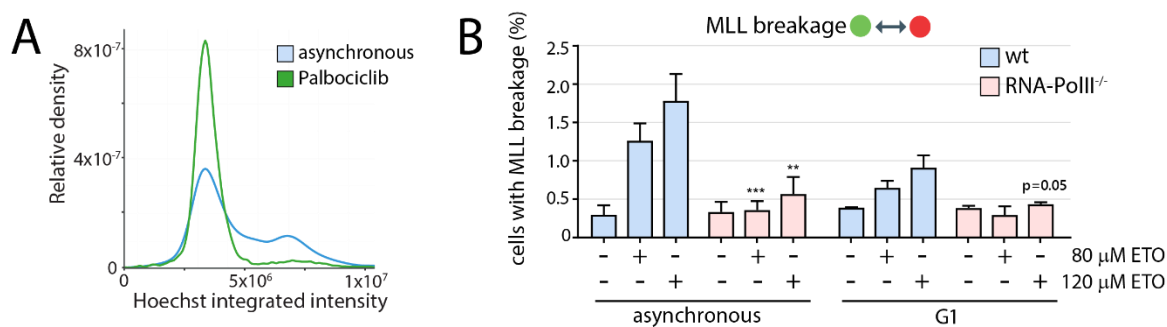


**Figure 2.27 RNA-PolIII is required for ETO-induced *MLL* fragility.** (A) Scheme of integrations for the generation of the DLD-1 RNA-PolIII degron cell line (Nagashima et al., 2019). The receptor OsTIR1 is under the control of a dox-inducible promoter. mAID-tagged RNA-PolIII large subunit (RBP1) is bound by auxin, recognized by OsTIR1 and thereby targeted for proteasomal degradation. (B) Western blotting for RNA-PolIII (RBP1) in DLD-1 degron cells upon induction of the TIR1 receptor with doxycycline (dox, 2  $\mu\text{g}/\text{mL}$  for 24 h) and degradation of RNA-PolIII by auxin (500  $\mu\text{M}$ , 1 h to 4 h). (C) Cell cycle distribution of DLD-1 cells with or without depletion of RNA-PolIII (for 4h), analyzed by measuring the DNA content by Hoechst intensities (Roukos et al., 2015). (D) Quantification of  $\gamma\text{H2AX}$  fluorescence intensities upon etoposide in DLD-1 cells deprived of RNA-PolIII, across the cell cycle. Cells were pretreated with auxin for 2 h and were then treated with ETO at 30  $\mu\text{M}$  or 50  $\mu\text{M}$  for 2 h. (E) Quantification of 53BP1 foci per nucleus upon etoposide (2 h) in DLD-1 cells deprived of RNA-PolIII. (F) Percentage of cells with *MLL* breakage upon acute ETO-treatment (2 h pretreatment with auxin, 4 h ETO), quantified by C-Fusion 3D in two independent experiments (1200 – 4600 cells per sample). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$  by One Way ANOVA test and Tukey test (IF analysis in D, E) or student's t-test (C-Fusion 3D in F), comparison of RNA-PolIII<sup>-/-</sup> to respective wild-type conditions.

These results suggest that the activity of RNA-PolIII is directly required for TOP2-induced DNA damage and chromosome breaks. However, we previously observe that replication contributed to ETO-induced DNA damage (Figure 2.26). Indeed, the two processes of transcription and replication may influence one another. For example, transcription can

## Results

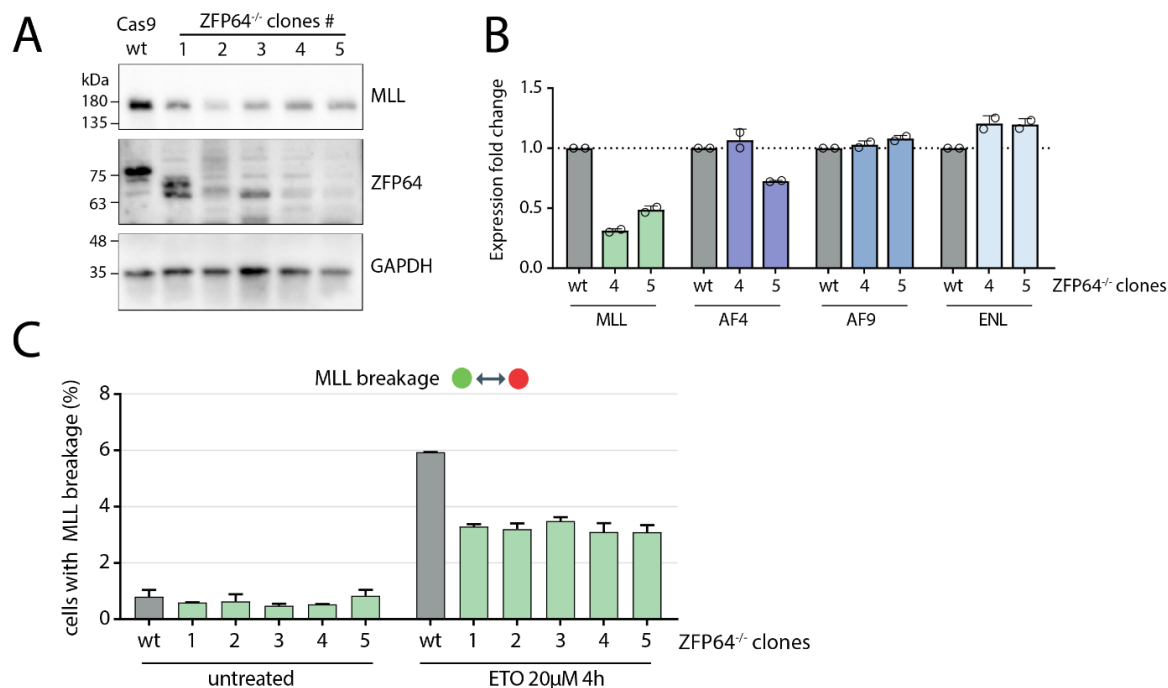
influence replication by its contribution to replication origin priming or transcription-replication conflicts (Prioleau and MacAlpine, 2016). Thus, to exclude indirect effects on replication by RNA-PolIII degradation, we probed *MLL* fragility in G1 arrested cells (Figure 2.28). DLD-1 cells were arrested in G1 with Palbociclib prior to RNA-PolIII degradation and ETO-treatments. We again observed reduced *MLL* breakage in RNA-PolIII depleted cells (Figure 2.28B,  $p = 0.05$ ), suggesting a direct role of active transcription in the generation of TOP2-induced DSBs.



**Figure 2.28 *MLL* breakage in G1 depends on active transcription.** (A) Cell cycle distribution of DLD-1 RNA-PolIII degron cells upon arrest in G1 by Palbociclib (5 μM for 24 h). (B) Percentage of cells with ETO-induced *MLL* breakage in asynchronous or G1 arrested DLD-1 RNA-PolIII degron cells. RNA-PolIII was degraded by 2 h pretreatment with auxin, breakage was induced by 80 μM or 120 μM ETO for 4 h. Means + SD of two experiments are shown (4100 to 10000 cells per sample) \*\* $p < 0.01$ , \*\*\* $p < 0.001$  by student's t-test compared to respective wild-type (wt) samples.

An alternative approach to show that active transcription generates chromosomal breaks was to specifically downregulate the expression of a single gene, here *MLL*, and to assess the effect on its fragility upon ETO-treatment. In order to downregulate *MLL* expression only, we took advantage of recent finding where the factor ZFP64 was described to be a specific transcriptional regulator of *MLL* (Lu et al., 2018). We thus generated several knockout clones of ZFP64 in TK6-Cas9 cells and found that upon disruption of *ZFP64* by CRISPR/Cas9, the expression of *MLL* was strongly reduced (Figure 2.29A). We further quantified mRNA levels of *MLL* but also AF4, AF9 and ENL by qPCR in two ZFP64<sup>-/-</sup> clones (#4 and #5, Figure 2.29B) to confirm that *MLL* expression was affected specifically by the knockout of ZFP64 and not transcription in general (Lu et al., 2018). Indeed, we found that mRNA levels of *MLL* were reduced by 50-70 % compared to wild-type cells, whereas expression of AF4, AF9 or ENL was not affected (Figure 2.29B). Finally, we probed for *MLL* breakage in ZFP64 knockout clones and TK6 wild-type cells by C-Fusion 3D. After an acute ETO-treatment, ~50 % less breaks of *MLL* were detected in ZFP64<sup>-/-</sup> clones than in wild-type TK6-Cas9 cells. This finding implies that the expression of *MLL* is directly linked to its susceptibility to TOP2-induced breakage, which is a prerequisite for *MLL* chromosome translocation to occur.





**Figure 2.29 Suppression of *MLL* expression by *ZFP64* knockout reduces *MLL* breakage.** (A) Western blotting of ZFP64 and MLL levels in TK6 Cas9 parental or ZFP64<sup>-/-</sup> clones. (B) Quantitative reverse transcription PCR of MLL, AF4, AF9 and ENL mRNA levels in ZFP64<sup>-/-</sup> clones #4 and #5 normalized to TK6 parental cells (in duplicates). (C) Percentage of cell with ETO-induced (20  $\mu$ M for 4 h, no release) *MLL* breakage in TK6 wild-type or ZFP64 knockout cells, quantified by C-Fusion 3D. Means + SD of two experiments are shown (1800 – 8600 cells analyzed per sample).

We thus conclude that the expression of genes is a key determinant of their ETO-induced fragility, which was eminent on a global as well as a local scale. The higher genes express, the more DSBs they tend to accumulate. Furthermore, TOP2-induced DDR signaling and breakage is highly transduced by RNA-PolIII, which strongly suggests that transcription is a key driving force of TOP2cc conversion to genotoxic DSBs. TOP2-induced DSB formation is further enhanced by an interplay of transcription and spatial chromosome folding, indicating that loop anchors are hotspots of genomic instability.

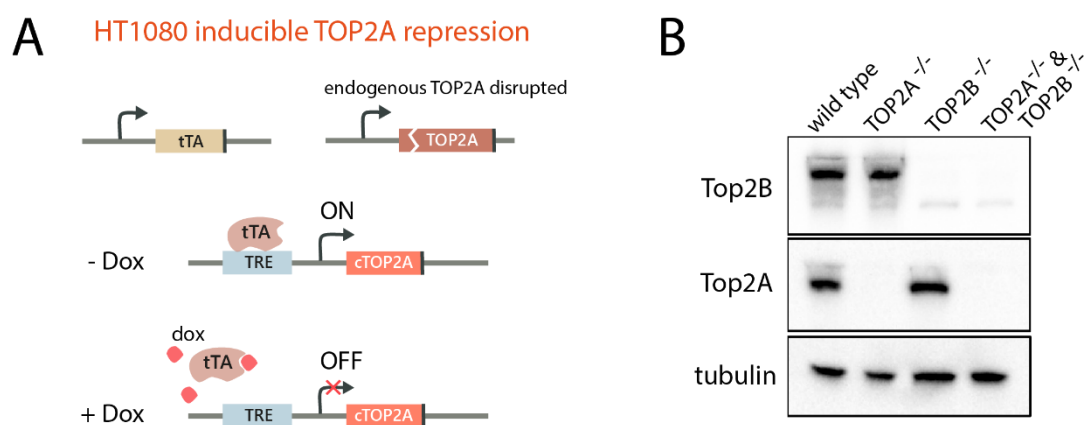
## 2.3 The TOP2 isoforms alpha and beta both contribute to *MLL* fragility

### 2.3.1 Cell models to study the contributions of TOP2 isoforms to ETO-induced DNA damage

Dissipation of torsional stress relies on the essential functions of TOP2s, which mitigate supercoiling and unfavorable DNA structures arising from transcription, replication, spatial chromosome organization and mitosis. We have found strong dependencies of ETO-induced gene fragility on transcription and partially on chromatin looping and replication. TOP2A and TOP2B have been described to mediate these processes to varying degrees and to be differentially expressed across cell types and through the cell cycle. The relative contributions of the TOP2 isoforms to ETO-induced DNA damage are not yet clear and have been under investigation by many researchers (Azarova et al., 2007; Canela et al., 2017; Cowell et al., 2012). A key challenge for systematic studies of both isoforms has on the one hand been the essentiality of TOP2A for cell survival and on the other hand the proliferation-dependent expression of TOP2A, which is mainly expressed in S, G2 and M phase.

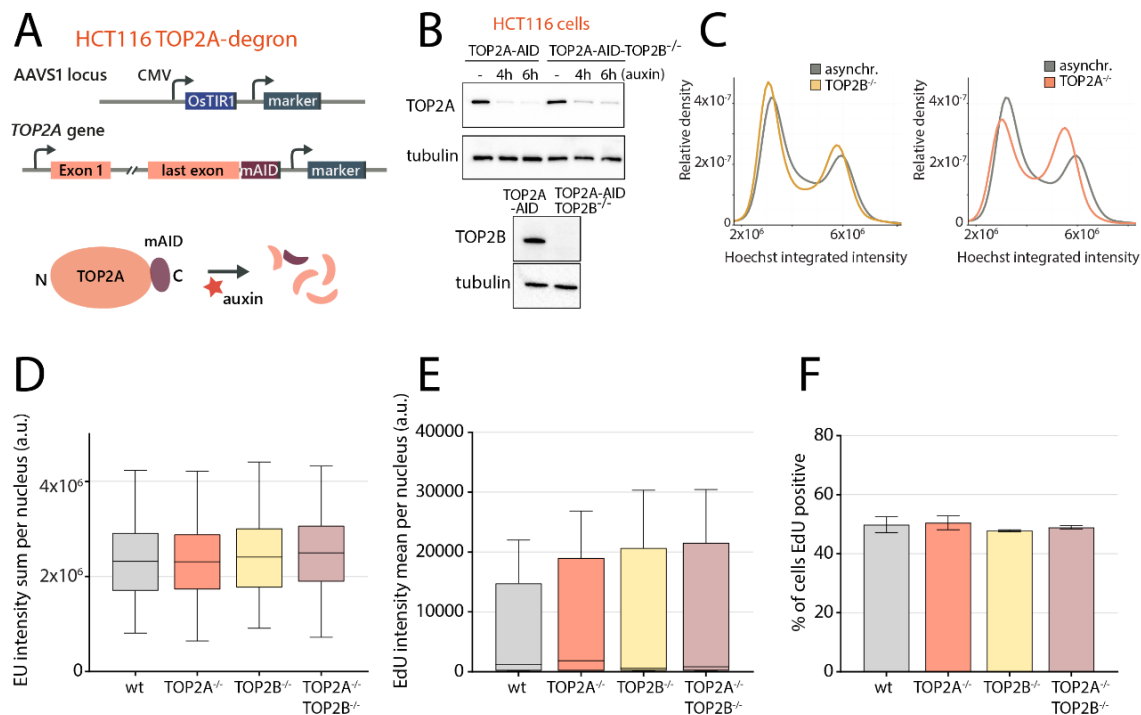
To systematically uncover the contributions of TOP2A and TOP2B to TOP2-induced DNA lesions, we utilized isogenic cell models, in which TOP2A can be depleted from cells inducibly, and combined these with *TOP2B* gene disruption by CRISPR/Cas9. One inducible TOP2A TET-Off cell system in HT1080 was previously developed by the group of Andrew Porter and has been used in the field to study the function of TOP2A during mitosis (also called HTETOP cells (Carpenter and Porter, 2004; Farr et al., 2014; Kozuki et al., 2017)). The cDNA of TOP2A (cTOP2A) was stably integrated into parental HT1080 cells under the control of a tetracycline response element (TRE) which is activated by a tetracycline-controlled transactivator protein (tTA) in the absence of tetracyclines (Figure 2.30A). Upon addition of a tetracycline, such as doxycycline (dox), expression of cTOP2A is repressed. The endogenous *TOP2A* gene was additionally disrupted to thereby create a dox-controlled TOP2A expression (Figure 2.30A) (Carpenter and Porter, 2004). We further modified this cell system by also disrupting the *TOP2B* gene, thereby creating TOP2B<sup>-/-</sup> cells that upon dox addition were depleted of both TOP2 isoforms (Figure 2.30B). We utilized the HT1080 TOP2 knockout cell model to study the contributions of TOP2A and TOP2B to ETO-induced DNA damage (see Figure 2.32 and Figure 2.33). However, the dox-activated depletion of TOP2A has to be induced for 48h, which caused severe cell cycle arrests of cells (Carpenter and Porter, 2004), mainly because of the importance of TOP2A for mitosis. Furthermore, the expression levels of the cTOP2A in HTETOP cells did not mirror endogenous protein levels that we observed for other human cell lines (Figure 2.32D).

To overcome the limitations of the dox-inducible TOP2A repression system, we sought to employ an auxin-inducible degron (AID) of TOP2A, which facilitates rapid and efficient degradation of the endogenous TOP2A protein within hours (Figure 2.31A, B). Christian Friberg Nielsen and Damien Francis Hudson generated this TOP2A degron cell lines in the colorectal cancer cell line HCT116 (Gothe et al., 2019). In HCT116 cells stably expressing OsTIR1, the endogenous TOP2A gene was C-terminally tagged with a mAID-tag. The addition of auxin triggers the ubiquitination of TOP2A and its degradation by the proteasome (Figure 2.31A, also see Figure 2.27). Indeed, when we treated the HCT116 TOP2A-mAID cells with auxin, up to 95 % of the protein was eliminated within four to six hours. We additionally created a *TOP2B* gene knockout in this cell system to study the role of TOP2B and create an inducible double knockout cell line for both TOP2 isoforms (Figure 2.31B). Disruption of *TOP2B* did not alter the cell cycle progression whereas depletion of TOP2A caused a partial G2/M arrest after 4 h of auxin (Figure 2.31C). Transcription and replication activities were, however, not affected by the depletion of TOP2A for several hours or the disruption of *TOP2B* (Figure 2.31D, E). We thus utilized the HCT116 cell model to study the contributions of the TOP2 isoforms to ETO-induced DNA damage and *MLL* fragility.



**Figure 2.30 Generation of a TOP2B knockout in a dox-inducible TOP2A repression HT1080 cell line.** (A) Inducible repression of TOP2A expression in HT1080 cells by a TET-Off system. The endogenous *TOP2A* gene was disrupted and a TET-regulated *TOP2A* stably integrated into HT1080 cells. The tetracycline-controlled transactivator protein (tTA) induces the expression of TOP2A only in the absence of doxycycline (Carpenter and Porter, 2004). (B) A constitutive knockout of TOP2B by CRISPR/ Cas9 was generated in dox-inducible HT1080 TOP2A knockout cells (HTETOP). Confirmation of the TOP2B knockout and inducible TOP2A depletion (10  $\mu$ g/mL dox for 48 h) was done by western blot.

## Results

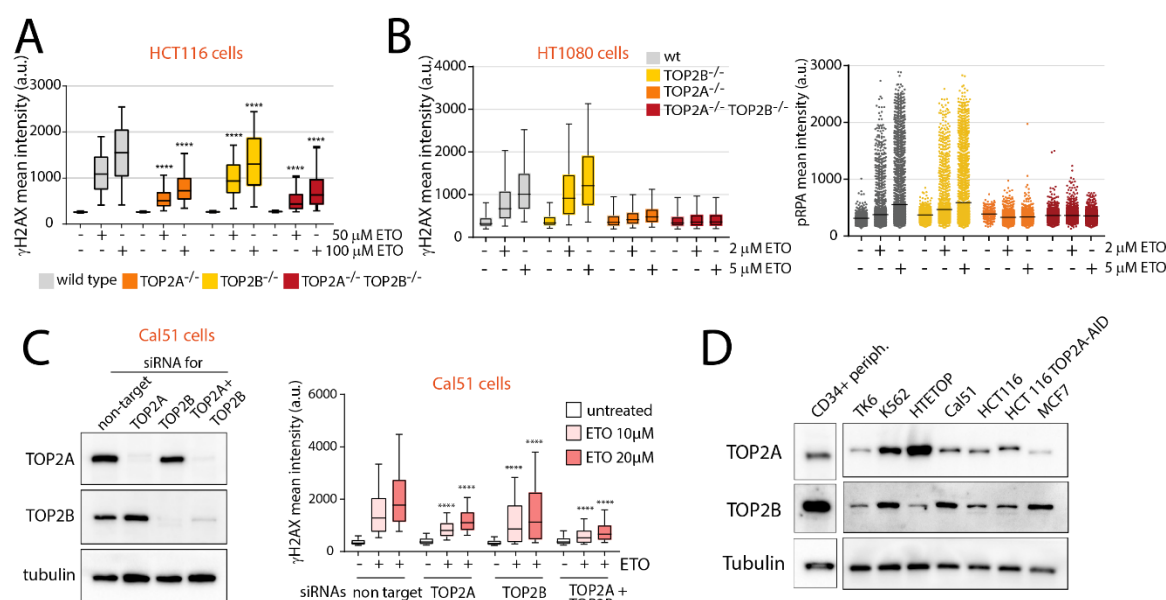


**Figure 2.31 Generation of a TOP2A auxin-inducible degron and TOP2B knockout in HCT116 cells.** (A) To generate an auxin-inducible degron cell line for TOP2A, the *TOP2A* gene in HCT116 was tagged by a mAID-tag, which leads to rapid proteasomal degradation of TOP2A upon addition of auxin (Gothe et al., 2019). (B) Western blot for TOP2A levels (top) which shows the rapid degradation of TOP2A in HCT116 degron cells upon addition of auxin (500  $\mu$ M). Additionally, a constitutive TOP2B knockout was generated by CRISPR/Cas9 (bottom) which allows for the double depletion of both TOP2 isoforms. (C) Cell cycle profiles determined by integrated nuclear Hoechst33342 intensities in TOP2A (auxin for 4 h) or TOP2B depleted cells. (D) Nascent RNA expression indicated by EU labelling in TOP2A, TOP2B or double deficient cells. (E, F) Active replication was estimated by EdU labelling, by quantification of nuclear intensities or the percentage of cells that were EdU positive and were thus engaged in replication.

### 2.3.2 Both TOP2A and TOP2B isoforms contribute to ETO-induced DNA damage and *MLL* translocations

To study the contribution of both TOP2 isoforms to ETO-induced DNA damage, we employed the inducible HCT116 and HT1080 cell models in which both TOP2 isoforms can be depleted separately or together (Figure 2.30, Figure 2.31). Quantifications of  $\gamma$ H2AX showed a strong reduction of ETO-induced DDR signaling upon the depletion of TOP2A or both isoforms in HCT116 and HT1080 cells (orange and red, Figure 2.32A, B). The knockout of TOP2B significantly decreased DNA damage signaling in HCT116 cells but did not lower the damage response in HT1080 cells (yellow). This was also mirrored in pRPA signaling in HT1080 cells (Figure 2.32B). To confirm the contributions of both isoforms, we silenced TOP2A and TOP2B by RNAi in another human cell line, the breast cancer cells Cal51. We found that in Cal51 cells, ETO-induced DNA damage signaling also depended on both TOP2A and TOP2B isoforms

(Figure 2.32C). To understand the different dependencies in different cell lines, we probed the relative protein levels of the two TOP2 isoforms in a panel of human cell lines (Figure 2.32D). Importantly, for the HT1080 cell model HTETOP, in which ETO-induced signaling was only TOP2A dependent, TOP2A was expressed at a much higher level than in other cell lines, and TOP2B was expressed at relatively low levels (Figure 2.32D). This could be explained by the design of the cell model in which the cTOP2A was stably, but randomly, introduced into HT1080 cells thereby allowing uncontrolled expression beyond endogenous levels. Our data thus suggest that both TOP2 isoforms contribute to ETO-induced DNA damage and that relative expression levels could be predictive of cell line specific dependencies on TOP2 isoforms.

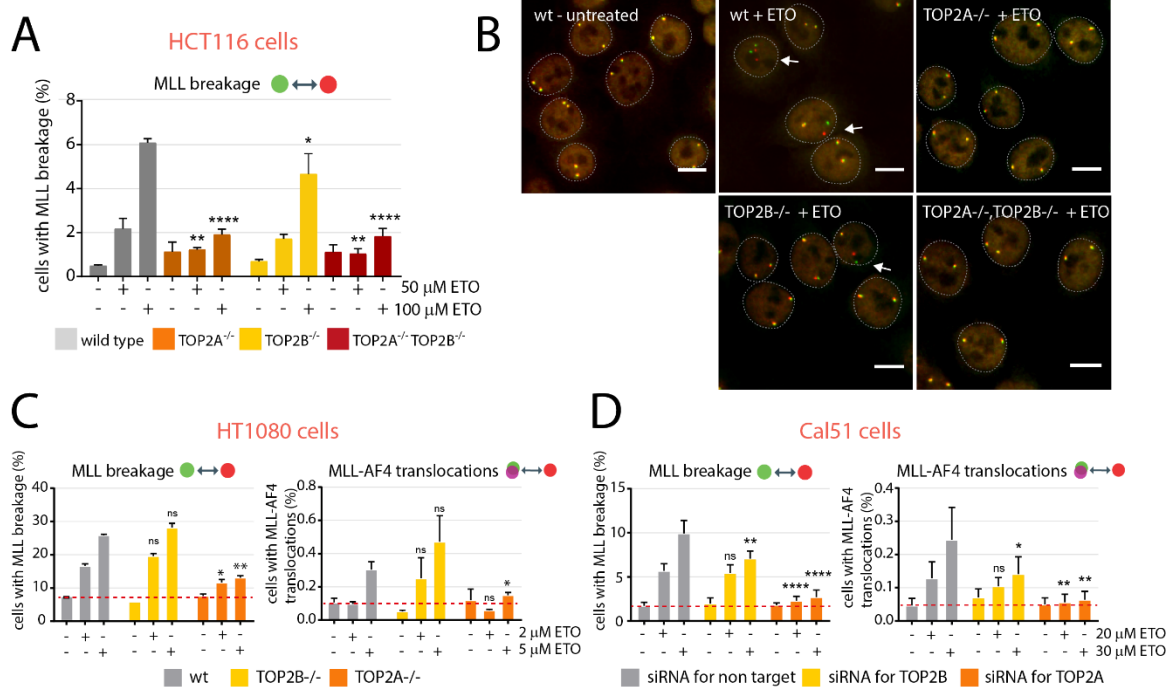


**Figure 2.32 Both TOP2 isoforms contribute to ETO-induced DNA damage signaling.** (A) Quantification of  $\gamma$ H2AX immunostaining upon ETO-induced DSB-signaling (50  $\mu$ M or 100  $\mu$ M for 3 h) in HCT116 cells in the absence of TOP2 isoforms (one representative out of three independent experiments is shown). (B) In HT1080 cells,  $\gamma$ H2AX and S4/S8 phospho-RPA (pRPA) signaling were quantified upon ETO-induced DNA damage (2  $\mu$ M or 5  $\mu$ M for 4 h), in the absence of TOP2 isoforms. (C) Verification of efficient TOP2A and TOP2B RNAi knockdown in Cal51 cells by immunoblotting, three days after siRNA transfection. Quantification of ETO-induced  $\gamma$ H2AX signaling (10  $\mu$ M or 20  $\mu$ M for 4 h) in RNAi silenced cells as indicated. (D) Expression levels of TOP2A and TOP2B in CD34+ peripheral cells and various human cell lines. \*\*\*\*p < 0.0001 by One Way ANOVA test.

To reveal the contributions of TOP2A and TOP2B to *MLL* breakage and oncogenic *MLL* translocations, we probed *MLL* fragility in HCT116, HT1080 and Cal51 cells upon depletion of the TOP2 isoforms by C-Fusion 3D. *MLL* breakage and the formation of *MLL-AF4* translocations were strongly dependent on TOP2A in all cell lines and also depended on

## Results

TOP2B in HCT116 and Cal51 cells (Figure 2.33A, B, C). This was reminiscent of the effect seen for DNA damage signaling (Figure 2.32). We thus conclude that trapping of either TOP2A or TOP2B isoforms can cause *MLL* DNA lesions, but that TOP2A has a more prominent role in ETO-induced *MLL* fragility and the formation of translocations.



**Figure 2.33 *MLL* breakage and translocations depend on both TOP2 isoforms.** (A) Analysis of *MLL* breakage in TOP2 deficient HCT116 cells by C-Fusion 3D. The TOP2A depletion was induced by 4 h of auxin pretreatment and cells were treated with ETO 50  $\mu$ M or 100  $\mu$ M for 3 h and were not released (1700 to 7300 cells were analyzed per sample in four independent experiments, means with SD). (B) Representative images of cells upon indicated conditions is shown (green probe is *MLL* 5', red probe is *MLL* 3', scale bar equal to 10  $\mu$ m). (C) Percentage of cells with *MLL* breakage or *MLL-AF4* translocations upon ETO (2  $\mu$ M or 5  $\mu$ M for 4 h, 24 h release) in HT1080 cells deficient for TOP2A or TOP2B assessed by 3-color C-Fusion 3D (1500 to 5600 cells were analyzed per sample, two independent experiments, means with SD). (D) *MLL* breakage or *MLL-AF4* translocations upon ETO (20  $\mu$ M or 30  $\mu$ M for 4 h, 24 h release) in Cal51 cells where TOP2A or TOP2B were silenced by RNAi (as in Figure 2.32C, 1500 to 8100 cells were analyzed per sample, three independent experiments, means with SD). \* $p$  < 0.05, \*\* $p$  < 0.01 and \*\*\*\* $p$  < 0.0001 by student's t-test.

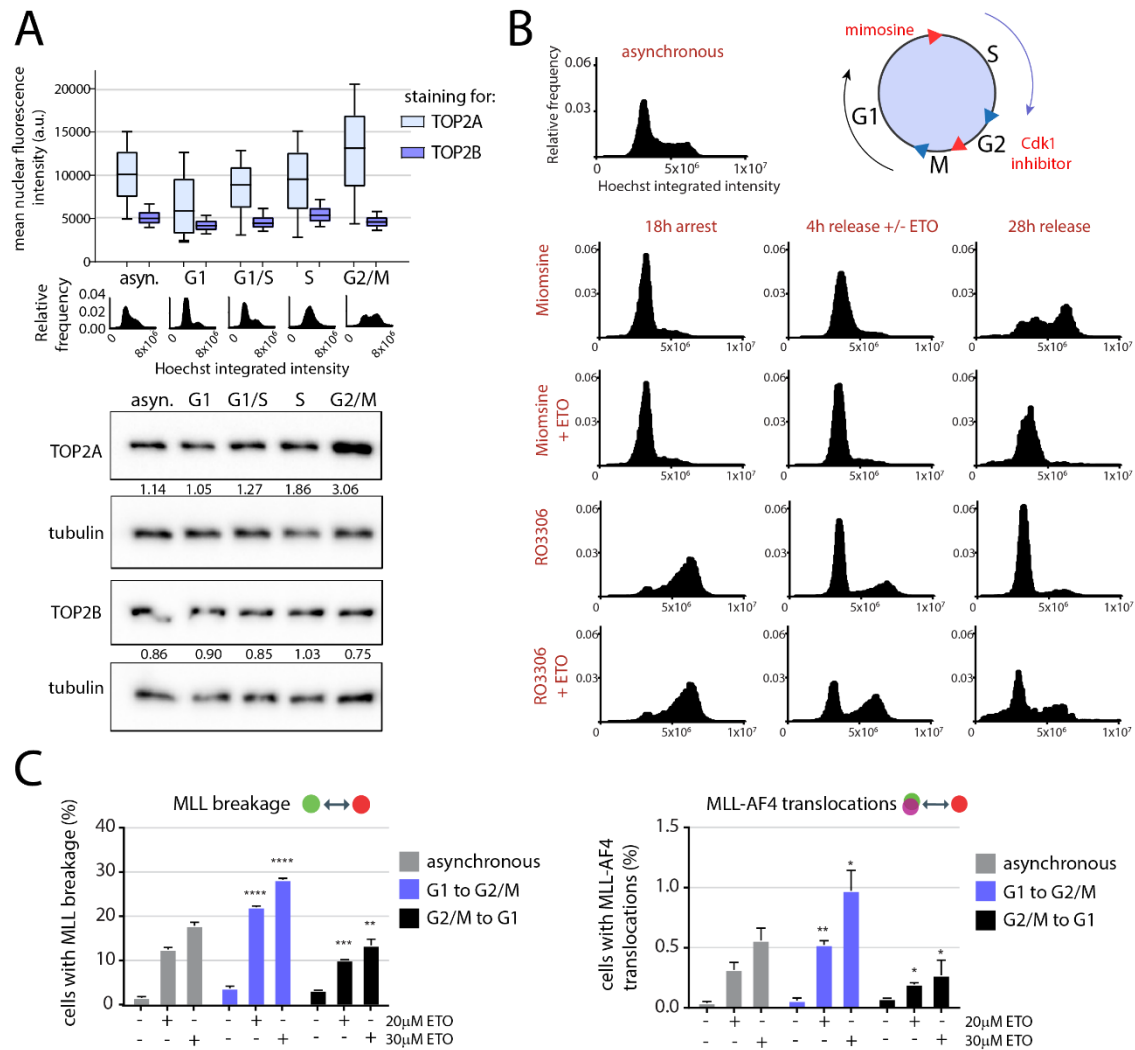
### 2.3.3 *MLL* fragility is cell cycle dependent

The expression of TOP2A and TOP2B has been reported to be differentially regulated by the cell cycle (Pommier et al., 2016). We confirmed that in TK6 cells TOP2B was expressed in all cell cycle stages whereas TOP2A expression increased in S-phase and peaked in G2/M (Figure 2.34A). To study if TOP2-induced *MLL* fragility was also cell cycle dependent, we arrest and

released TK6 cells into specific cell cycle phases by the use of inhibitors. Cells were either arrested in G1 by Mimosine and released into S and G2-phase, or they were arrested in late G2 by the Cdk1 inhibitor RO3306 and were released into G1-phase (Figure 2.34B). During the first 4 hours of release, cells were treated with etoposide. After 24 hours of release, *MLL* fragility was analyzed by C-Fusion 3D, which revealed more than two-fold higher levels of *MLL* breakage and *MLL-AF4* fusions in cells released into S/G2-phase than in cells released into G1-phase (Figure 2.34C). This was in line with the observation that TOP2A was mainly expressed in S to M-stages of the cell cycle (Figure 2.34A) and that *MLL* fragility strongly depends on the TOP2A isoform (Figure 2.33). It highlights the importance of a holistic view on ETO-induced DNA damage by studying TOP2-induced DNA damage in cycling cells through all cell cycle phases.

Our research indicates that both TOP2 isoforms contribute to ETO-induced DSBs and to the formation of *MLL* translocations, though to varying degrees among cell lines. This could be linked to their relative expression levels that differ among cell types *per se* and also throughout the cell cycle. Higher levels of breaks and translocations were found in cells cycling through S/G2 phase as cells cycling into G1 phase, where only TOP2B is expressed. Next, we sought to explore how TOP2-linked DNA breaks that contribute to *MLL* translocations are reversed and which pathways are involved in erroneous DSB repair.

## Results



**Figure 2.34 Higher *MLL* fragility occurs in S/G2 phase where both TOP2 isoforms are expressed.** (A) Expression of TOP2A and TOP2B throughout the cell cycle. TK6 cells were sampled in different cell cycle stages (middle panel) by blocking them with the inhibitors Mimosine (G1 arrest) or RO3306 (G2/M arrest) and releasing them. TOP2 levels were quantified by immunofluorescence analysis (top) and immunoblotting (bottom). Normalized intensities of TOP2A and TOP2B to tubulin are indicated for the immunoblot. (B) TK6 cells were arrested in G1 or G2/M phases with Mimosine or RO3306 and were released into the next cell cycle phase (S and G1, respectively) in the presence of the second inhibitor for 28 h. Within the first 4 h, cells were treated with ETO (20 or 30  $\mu$ M) and were afterwards release into the second inhibitor only for 24 h. Cell cycle distributions were derived from DNA staining by Hoechst33342. (C) Quantification of *MLL* breakage and *MLL-AF4* translocations upon ETO-treatment (4 h plus 24 h release) in arrested and release TK6 cells (as in B) quantified by 3-color C-Fusion 3D. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$  by student's t-test compared to control sample.



## 2.4 Repair of TOP2ccs is mediated by MRE11, TDP2 and NHEJ

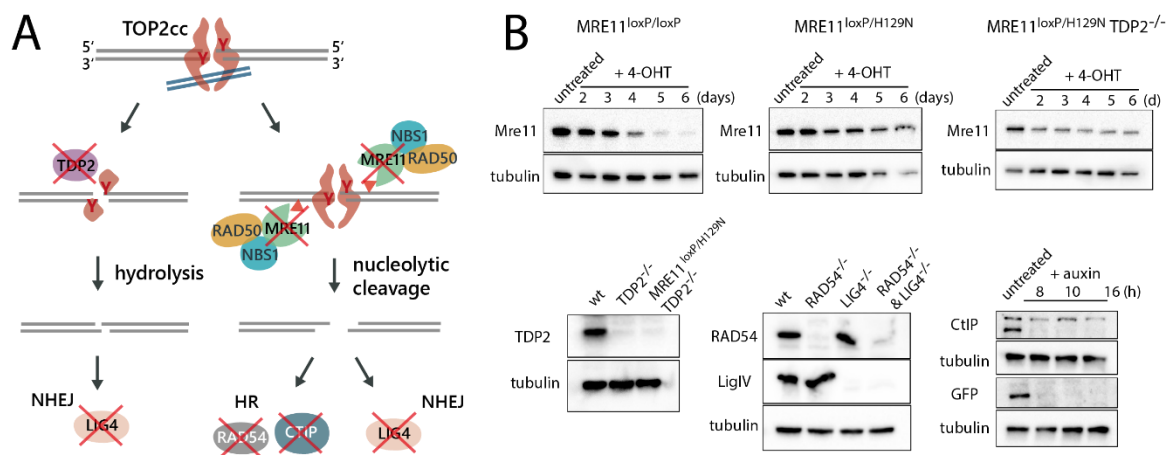
### 2.4.1 MRE11, TDP2 and NHEJ suppress ETO-induced DNA damage and the formation of *MLL* translocations

Covalent TOP2ccs can be removed in human cells by the phosphodiesterase activity of TDP2, which reverses the bond between the TOP2 tyrosine (Y) and the 5' phosphate of the DNA backbone (Gómez-Herreros et al., 2017; Ledesma et al., 2009) (Figure 2.35A). Alternatively, the DNA covalently linked to TOP2 can be removed by the nuclease activity of MRE11 (Hoa et al., 2016). The cleaned DSB can subsequently be repaired through NHEJ or HR (Figure 2.35A). To understand the contribution of these factors, we employed a panel of isogenic TK6 cells with gene knockouts (TDP2<sup>-/-</sup>, LIG4<sup>-/-</sup>, RAD54<sup>-/-</sup>), conditional depletion systems (MRE11<sup>loxP/loxP</sup>, MRE11<sup>loxP/H129N</sup>, CtIP-mAID) or combinations thereof (TDP2<sup>-/-</sup> MRE11<sup>loxP/H129N</sup>, RAD54<sup>-/-</sup> LIG4<sup>-/-</sup>) (Figure 2.35B). The MRE11<sup>loxP/H129N</sup> cell line is an inducible heterozygous knockout of MRE11, in which the second allele has a point mutation in the nuclease domain of MRE11 and thereby creates a nuclease-deficient MRE11<sup>-/nd</sup> (Hoa et al., 2015). End joining was studied by the knockout of LIG4, which abolishes classical NHEJ. Knockouts of RAD54 and inducible depletion of CtIP were used to study the effect of HR to ETO-induced damage induction. RAD54 is a motor protein that contributes to DNA strand invasion during HR and stimulates the HR recombinase RAD51 (Ceballos and Heyer, 2011). CtIP contributes to HR through resection of DNA ends (Hoa et al., 2015; Symington, 2016).

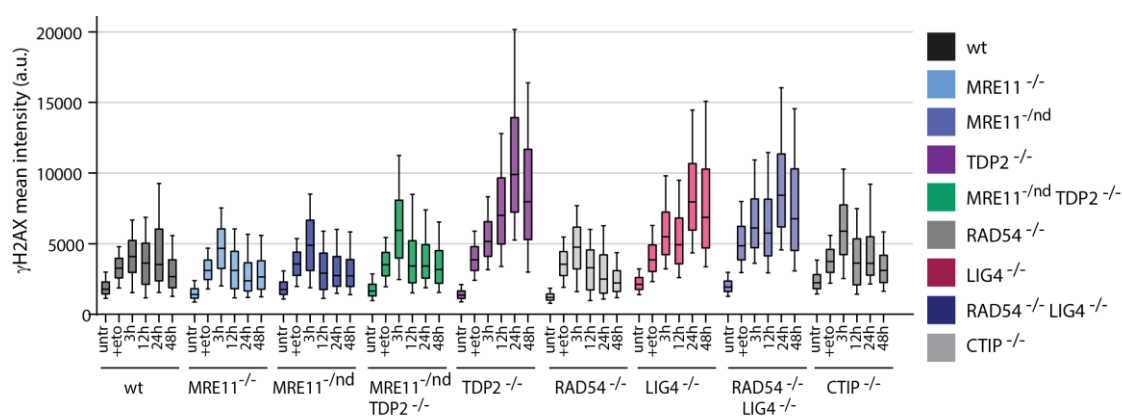
When challenging repair deficient TK6 cells with etoposide and following their DNA damage signaling for a release of two days, TDP2<sup>-/-</sup> cells showed strongly impaired repair kinetics (Figure 2.36). This is in line with previous reports where TDP2 has been implicated in protecting cells from ETO-induced DNA damage (Gómez-Herreros et al., 2013, 2017). Cells depleted of MRE11 or with nuclease-deficient MRE11, on the other hand, exhibited similar  $\gamma$ H2AX signaling as wild-type cells (Figure 2.36). Related repair kinetics were also seen in TDP2<sup>-/-</sup> MRE11<sup>-/nd</sup> cells, contrasting the delayed repair seen in only TDP2 deficient cells and previously reported dependencies of human cells on MRE11's nuclease activity for TOP2cc removal (Aparicio et al., 2016; Deshpande et al., 2016; Hoa et al., 2016). We thus reasoned that impaired DNA damage signaling capacities in cells deficient of MRE11 could possibly result from the role of MRE11 in activating the ATM kinase, which is a key kinase for the phosphorylation of H2AX (Lavin, 2007; Lavin et al., 2015). Delayed DNA repair was also observed for cells deficient of the NHEJ factor LIG4 (Figure 2.36). We could however not observe defects in ETO-induced DNA damage repair in cell lines missing the HR factors RAD54 or CtIP. Repair in cells deficient of both, LIG4 and RAD54, mirrored LIG4<sup>-/-</sup> kinetics,

## Results

further arguing that HR does not significantly contribute to repair of ETO-induced DNA damage. Instead, classical NHEJ seems to mediate the repair of ETO-induced DSBs.



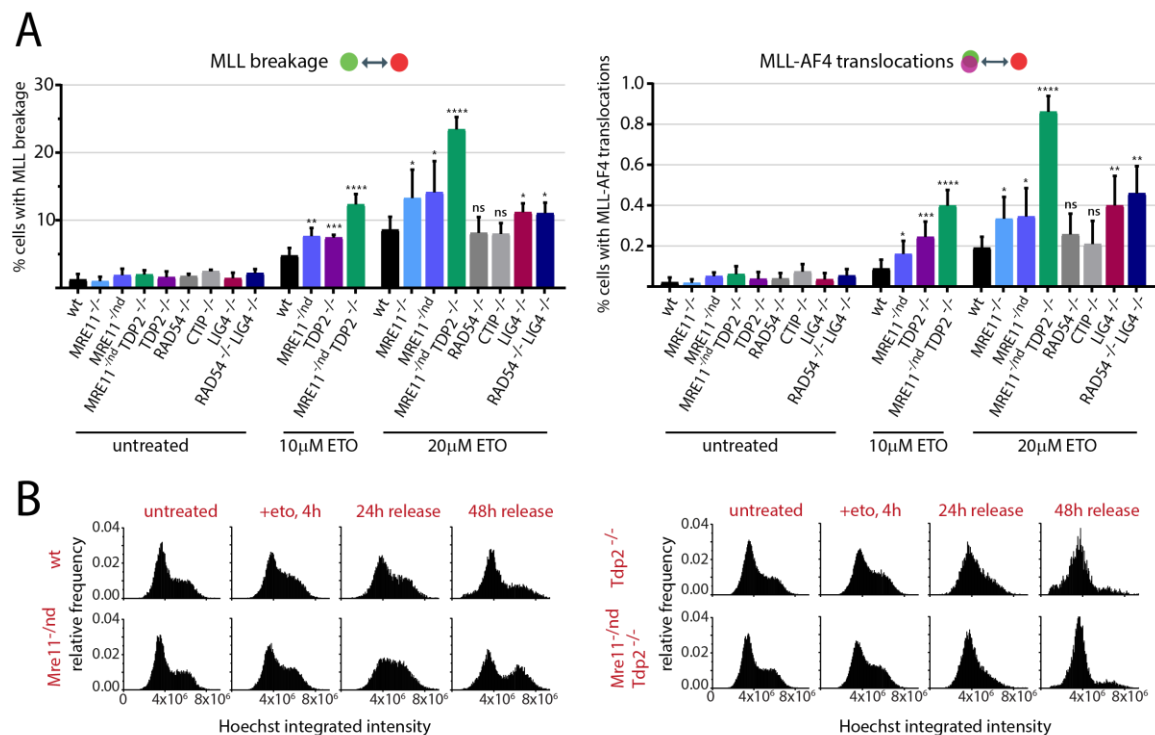
**Figure 2.35 TK6 cell lines deficient in TOP2cc removal or DSB repair pathways.** (A) TOP2 cleavage complexes can be removed by TDP2 reversing the covalent phosphodiester bond and providing a clean substrate for DSB repair by NHEJ (left). Alternatively, MRE11 may remove the DNA-TOP2cc complex by its endonuclease activity, giving rise to a resected DSB that may be repaired by either HR or NHEJ (Stingele et al., 2017). (B) Validation of a panel of TK6 cells deficient for factors involved in TOP2cc removal by western blotting. Knockouts of Mre11<sup>-/-</sup> and Mre11<sup>-/H129N</sup> were induced by treatment of cells with 4-OHT at 2  $\mu$ M, CtIP depletion was obtained by treatments with 500  $\mu$ M auxin.



**Figure 2.36 Kinetics of DSB repair in cell lines deficient for factors involved in TOP2cc removal or DNA-damage pathways.** TK6 cell lines deficient for repair factors were treated with 20  $\mu$ M ETO for 4 h and released for 3 h up to two days. Repair kinetics in the different mutant cell lines were assessed by quantifying levels of  $\gamma$ H2AX phosphorylation (750 to 35000 cells were analyzed per sample).

To address if TDP2, MRE11 and the choice of DSB repair pathway influence the formation of oncogenic *MLL* fusions, we tested ETO-induced *MLL* breakage and translocations in the deficient TK6 cell lines by C-Fusion 3D. Cells deficient of the HR factors RAD54 and CtIP showed similar *MLL* fragility as wild-type cells, arguing that HR does not contribute to the repair of ETO-induced *MLL* breakage (Figure 2.37A). The knockout of NHEJ factor LIG4 did however increase *MLL* susceptibility to persistent DSBs and fusions. An up to two-fold increase was seen in both LIG4<sup>-/-</sup> and RAD54<sup>-/-</sup> LIG4<sup>-/-</sup> cells, suggesting that NHEJ is required for efficient DSB repair and thereby may suppress oncogenic fusions (Figure 2.37A). Similarly, deficiency of MRE11, TDP2 or the nuclease-activity of MRE11 increased *MLL* breakage and translocations by ~two-fold. To assess *MLL* fragility in TDP2<sup>-/-</sup> cells, lower levels of ETO were used (only 10μM) since cells experienced higher cytotoxicity after ETO-treatment than other repair deficient cell lines, surprisingly also more than MRE11<sup>-/nd</sup> TDP2<sup>-/-</sup> double deficient cells. The cell lines MRE11<sup>-/nd</sup> and MRE11<sup>-/nd</sup> TDP2<sup>-/-</sup> were thus treated with both, 10μM and 20μM of etoposide to allow for the comparison of TDP2 and MRE11 repair under the same conditions. We observed that the single depletion of either TDP2 or MRE11 led to a similar accumulation of ETO-induced *MLL* fragility of approximately two-fold compared to wild-type cells. In MRE11<sup>-/nd</sup> TDP2<sup>-/-</sup> double mutant cells, additional *MLL* breakage and more than four times higher *MLL* translocation rates were seen (Figure 2.37A). The lack of epistasis caused by MRE11 and TDP2 double depletion indicates that TDP2 and MRE11 suppress *MLL* translocations through different pathways. Analysis of cell cycle profiles in TDP2 and MRE11 deficient cells confirmed that the increase in *MLL* fragility was not merely a consequence of changes in the cell cycle progression. Similar cell cycle distributions as in parental TK6 cells were seen in untreated and ETO-treated cell lines (Figure 2.37B).

## Results

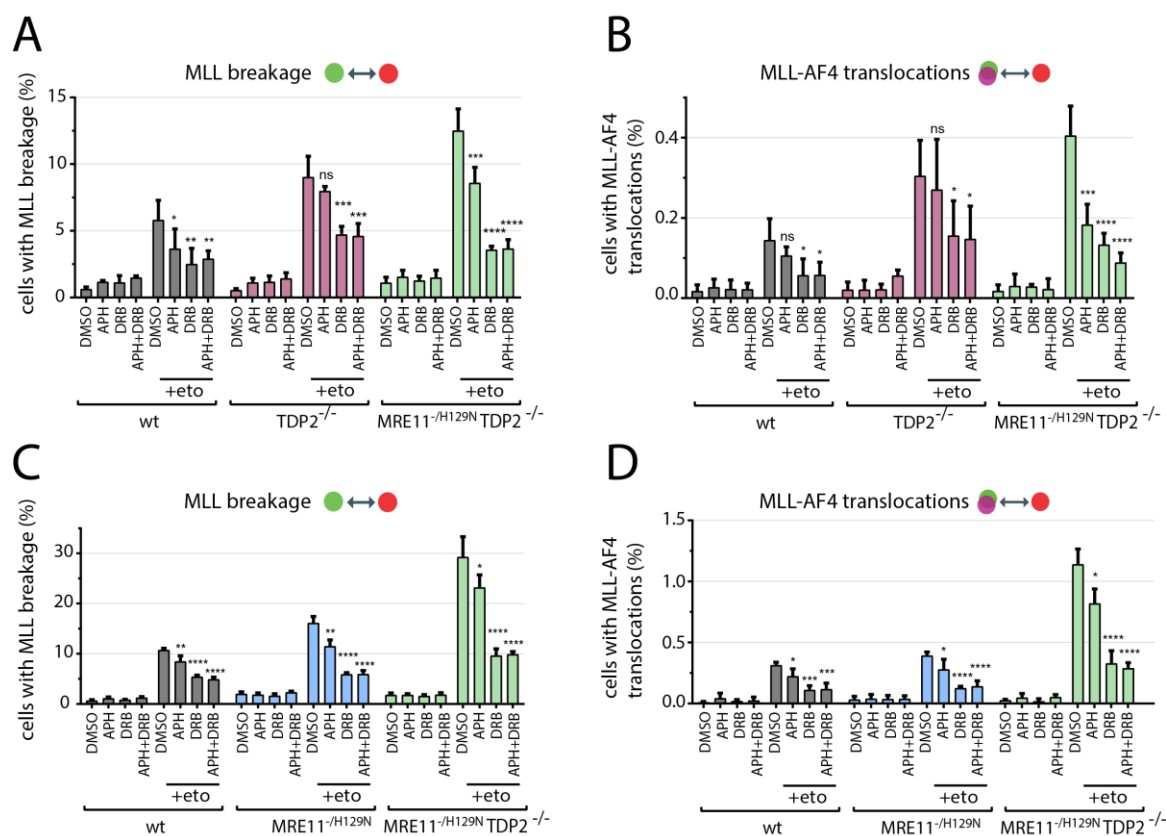


**Figure 2.37 TDP2, MRE11 and NHEJ, but not HR, suppress the formation of *MLL* translocations.** (A) Percentage of TK6 deficient of indicated genes with *MLL* breakage or *MLL-AF4* translocations upon 10  $\mu$ M or 20  $\mu$ M ETO for 4 h and released for two days. Analysis done by 3-color C-Fusion 3D, values represent means with SD from at least four independent experiments (1500 to 16,000 cells were analyzed per sample), \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$  by student's t-test compared to respective wt sample. (B) Cell cycle profiles of TK6 deficient cells upon treatment and release for up to 2 days from ETO treatment (10  $\mu$ M, 4 h).

### 2.4.2 Transcription and replication-dependent roles of MRE11 and TDP2 in the repair or TOP2ccs

We have found that etoposide can cause DSB formation within *MLL* through interference of TOP2ccs with active transcription or replication (Figure 2.26). MRE11 has been associated with active replication forks and to prevent DSBs during replication (Paull, 2018). TDP2 on the other hand is rather implied in transcription-dependent repair activities (Gómez-Herreros et al., 2017). We sought to connect the findings of TDP2 and MRE11 being suppressors of *MLL* fusions to the dependency of *MLL* fragility on transcription and replication. Therefore, we studied the effect of transcription and replication inhibition in TDP2, MRE11 and double mutant TK6 cells. Inhibition of transcription by DRB strongly decreased *MLL* fragility in all cell models, indicating that the MRE11 and TDP2 repair pathway act on transcription-induced TOP2 lesions (Figure 2.38A-D). Replication inhibition, however, could not decrease *MLL* fragility in TDP2<sup>-/-</sup> cells (Figure 2.38A, B). In contrast, MRE11<sup>-/nd</sup> and MRE11<sup>-/nd</sup> TDP2<sup>-/-</sup> cells showed significantly less *MLL* breakage and fusions upon

pretreatments with aphidicolin (Figure 2.38A-D), suggesting that MRE11 is active in suppressing TOP2-induced DNA damage within *MLL* that is mediated by transcription or replication. A transcription-specific role was seen for the activity of TDP2, which predominantly mediated repair of transcription-dependent chromosome instability upon etoposide (Gómez-Herreros et al., 2017).

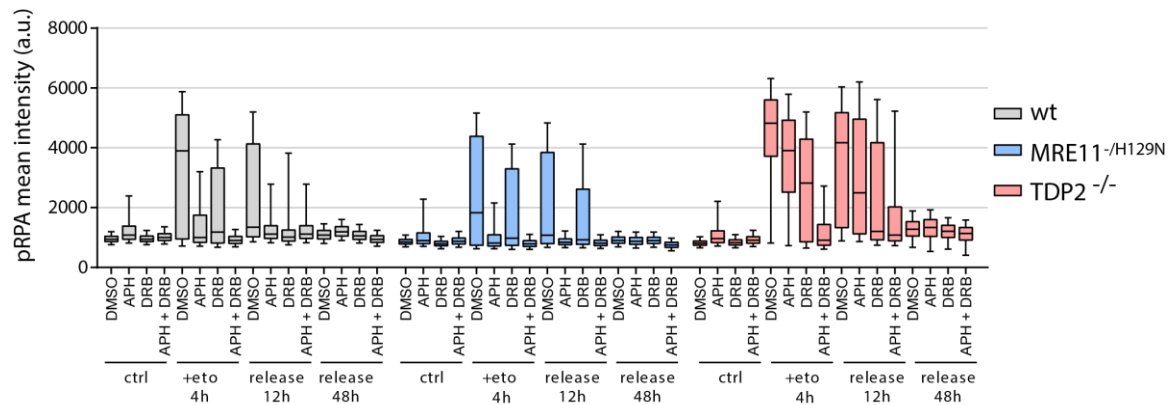


**Figure 2.38 Role of TDP2 in suppressing *MLL* translocations that are mainly mediated by transcription.** (A-D) TK6 wt, TDP2<sup>-/-</sup>, MRE11<sup>-/H129N</sup> or double mutant TDP2<sup>-/-</sup>/MRE11<sup>-/H129N</sup> cells were pretreated for 3 h with 200  $\mu$ M DRB, 8  $\mu$ M aphidicolin or both before ETO-treatments. (A, B) Cells were treated with 10  $\mu$ M ETO for 4 h and frequencies of *MLL* breakage (A) and *MLL-AF4* translocations (B) were assessed after two days release by 3-color C-Fusion 3D. (C, D) Cells were treated with 20  $\mu$ M ETO and *MLL* breakage and translocations were analyzed as in (A, B). Values represent means with SD from at least four independent experiments (1800 to 16,500 cells were analyzed per sample). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$  by student's t-test compared to respective DMSO control sample.

We additionally analyzed levels of S4/S8 RPA-phosphorylation (pRPA), which is a marker of ETO-induced DNA damage signaling and colocalizes with MRE11 upon damage induction (Liu et al., 2012). Compared to wild-type TK6 cells, MRE11 nuclease-deficient cells showed reduced levels of pRPA intensities that were highly sensitive to replication inhibition, but also transcription inhibition (Figure 2.39). Contrary, in TDP2<sup>-/-</sup> cells, highly elevated levels of pRPA

## Results

signaling were observed that could only effectively be abolished upon inhibition of both, transcription and replication, before etoposide treatments (Figure 2.39). This suggests that pRPA is indeed a marker of MRE11-associated TOP2cc repair and that MRE11 has a role in replication-induced DNA damage signaling. Our findings thus indicate that TDP2 and MRE11 are suppressors on TOP2-induced DNA damage with distinct roles in transcription and replication-mediated DNA instability upon etoposide.



**Figure 2.39 Mre11 causes a replication-associated stress response upon ETO.** TK6 wt, MRE11<sup>-/-H129N</sup> and TDP2<sup>-/-</sup> were treated as in Figure 2.38 (ETO 10  $\mu$ M for all cells) and S4/S8 pRPA was quantified by immunostaining and high-throughput imaging (2000 to 12,500 cells were analyzed per sample).

## 2.5 Screening for new factors influencing TOP2-induced *MLL* fragility

In the last years, many discoveries have been made that identified novel TOP2cc repair pathways and fragility mechanisms of *MLL*. For example, phosphodiesterase activity of TDP2 towards TOP2ccs was discovered in 2009 by the group of Keith Caldecott, ZATT was only recently indicated as a potential regulator of TOP2ccs and chromatin organization has been implied to contribute to TOP2-induced gene fragility (Canela et al., 2017; Ledesma et al., 2009; Schellenberg et al., 2017). Potential functions of more specific proteases like SPRTN, DNA helicases or signaling pathways involving ubiquitin and SUMO post-translational modification have been proposed, illustrating that we are not yet at the end of discoveries (Berti et al., 2013; Lopez-Mosqueda et al., 2016; Rao et al., 2016; Vaz et al., 2016).

Here, we aim at discovering new factors that regulate TOP2-induced gene fragility. Therefore, we probed candidate proteins that could have a potential function in DSB generation upon TOP2 poisoning by etoposide. RNA interference (RNAi) based screens have proved as a relatively easy and dynamic screening tool to identify factors of molecular pathways in human cells (Boutros and Ahringer, 2008). The endogenous RNAi pathway that targets messenger RNAs for degradation by short double-stranded RNAs (such as microRNAs) is utilized to thereby downregulate gene expression. Gene silencing of factors are accomplished by introducing exogenous short interfering RNAs (siRNAs) into cells that target the candidate mRNAs of interest. RNAi of targets can be done at all scales, from single candidates being silenced to genome-wide screens, and various readout assays have been used (Mohr et al., 2010). To probe TOP2-induced gene fragility, we developed C-Fusion 3D in a high-throughout 384-well format to probe gene specific breakage upon ETO-treatment. The success of RNAi screens is mostly based on a clear readout, a dynamic range that is verified by positive and negative controls and stringent validation of primary hits (Boutros and Ahringer, 2008). We aimed at performing a targeted screen to probe ~150 candidates factors involved in DNA damage repair, chromatin and transcription regulation or ubiquitin signaling. Primary candidates were subsequently verified and one candidate was followed up on in more detail.

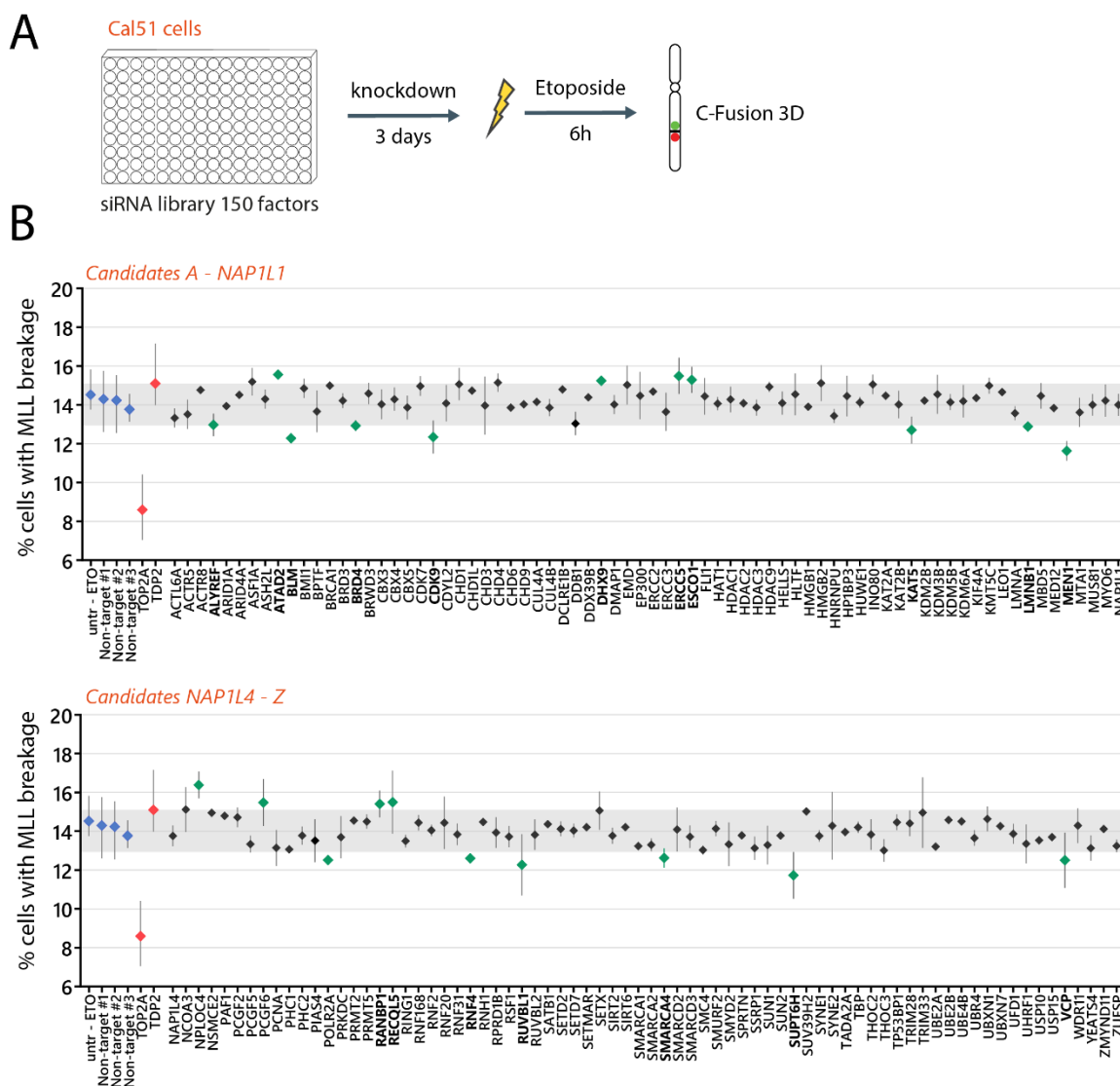
### 2.5.1 Loss-of-function screen identifies novel regulators of *MLL* fragility

We performed an RNAi knockdown screen to identify new factors contributing to or suppressing TOP2-induced instability of *MLL* upon treatment of cells with etoposide. A set of 150 cherry-picked candidates was select that have been associated with chromatin modifications, transcription regulation, ubiquitin signaling, DNA repair or were of unknown function. As a cell model, the near diploid breast cancer cell line Cal51 was chosen which had

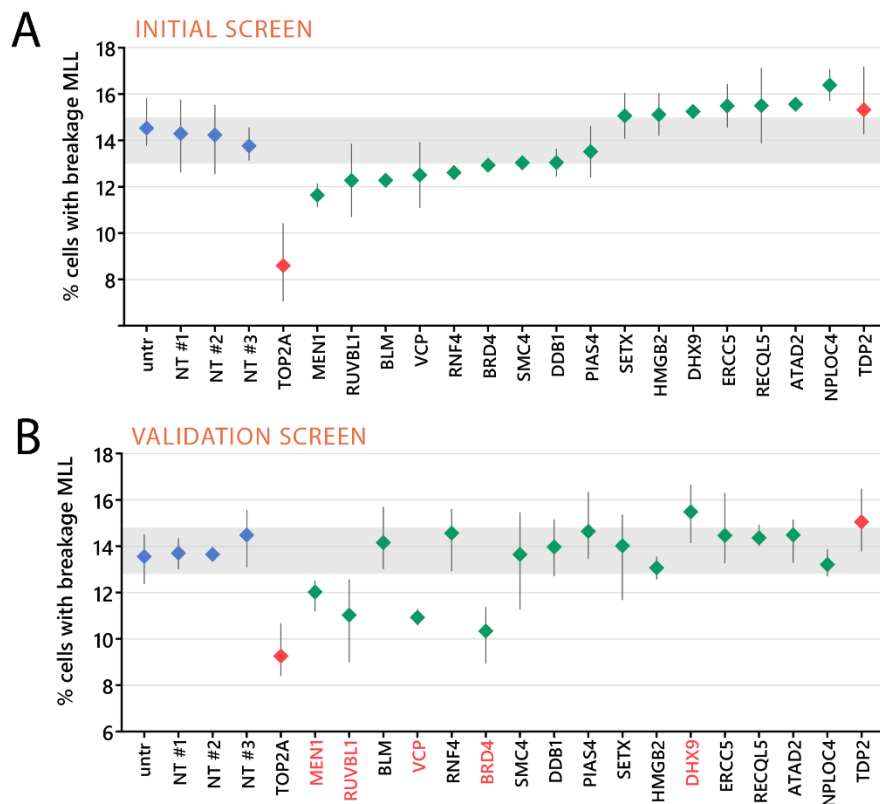
previously proved to mediate effective RNAi-mediated gene silencing, it responded to ETO at standard doses and had average expression of both TOP2 isoforms (Figure 2.32, Figure 2.33). *MLL* fragility was assessed in cells three days after the siRNA introduction and after an acute ETO-treatment of six hours with no release to limited cell death (Figure 2.40A). We performed *MLL* breakage analysis of two independent biological replicates by two-color C-Fusion 3D in 384-well plates whereby we were able to analyzed on average ~3500 cells per sample. The silencing of TOP2A and TDP2 served as controls for decreasing or increasing ETO-induced *MLL* breakage (Figure 2.33, Figure 2.37, Figure 2.40B). We identified 13 candidate factors that, upon knockdown, decreased *MLL* breakage from 14.1% in non-target controls to below 13 %, namely: ALYREF, BLM, BRD4, CDK9, KAT5, LMNB1, MEN1, POLR2A, RNF4, RUVBL1, SMARCA4, SUPT6H and VCP (Figure 2.40B). Inversely, an increase of *MLL* breakage above 15.2 % was observed upon knockdown of ATAD2, DHX9, ERCC5, ESCO1, NPLOC4, PCGF6, RANBP1, and RECQL5 (Figure 2.40B).

The knockdown of primary candidate factors was repeated for three additional independent experiments, to confirm a role of these 16 factors as regulators of *MLL* fragility. We could validate that MEN1, RUVBL, VCP or BRD4 mediate *MLL* fragility upon ETO-treatment whereas DHX9 acts as a suppressor of *MLL* fragility (Figure 2.41A, B). However, for some candidates a considerable variation between replicates was seen which is a typical caveat that is often experienced in RNAi screens (Boutros and Ahringer, 2008). Further testing and insights into these candidates' functions would be required to understand if changes in *MLL* breakage and translocations were mediated by candidates or were due to off-target effects. Nonetheless, a robust decrease in *MLL* breakage was observed upon the loss of *valosin*-containing protein (VCP: from ~14 % to 10.9 %, Figure 2.41B), which has previously been reported to play a role in remodeling of chromatin associated proteins (Torrecilla et al., 2017). We thus focused on understanding the role of VCP in TOP2cc removal and *MLL* fragility.





**Figure 2.40 RNAi screen identifies factors that modulate *MLL* fragility upon etoposide.** (A) Workflow of high-throughput RNAi screen for factors involved in ETO-induced *MLL* breakage in Cal51 cells. In 384-well plates, cells were reversely transfected with single siRNAs and after three days ETO-induced *MLL* breakage (60  $\mu$ M for 6 h) was quantified by C-Fusion 3D in two independent experiments. (B) Percentage of Cal51 cells with *MLL* breakage. Non-transfected and non-target samples are indicated in blue, positive and negative control knockdown samples, TOP2A and TDP2, are shown in red. RNAi samples for 150 factors are shown in black, except samples with *MLL* breakage below 13.0 % or above 15.2 % which are shown in green, in alphabetical order. For easier visualization, samples were split into two graphs with the controls shown twice (from two experiments, mean  $\pm$  range).



**Figure 2.41 Validation of factors modulating *MLL* fragility upon etoposide.** (A) Percentage of cells with *MLL* breakage (ETO 60  $\mu$ M for 6 h) in cells depleted of factors by RNAi in initial screen (Figure 2.40), ordered by increasing *MLL* fragility (controls shown in red). (B) Repeated analysis of *MLL* fragility upon depletion of factors by RNAi in three additional, independent replicates (mean  $\pm$  range). Names of factors that reproducibly modulated *MLL* breakage are highlighted in red. NT = non-target.

### 2.5.2 VCP and its cofactor PLAA might contribute to *MLL* fragility

VCP is an AAA+ ATPase, which is part of the ubiquitin-proteasome system and has broad cellular functions. Via ATP hydrolysis, VCP can unfold or segregate substrates from various cellular compartments, such as the endoplasmic reticulum, chromatin or protein complexes (Meyer et al., 2012; Torrecilla et al., 2017). Importantly, VCP recognizes substrates by their ubiquitin-modifications and targets them for proteasomal degradation or recycling by deubiquitination. Various adaptor proteins facilitate the specificity of VCP towards diverse cellular compartments and substrates. More than 30 of these cofactors have been described that, in various combinations, form sub-complexes with VCP to facilitate specific functions (Meyer et al., 2012; Ramadan et al., 2017). Nuclear activities of VCP have been linked to removal and degradation of chromatin-associated proteins and an emerging function in DNA repair is being reported (Torrecilla et al., 2017). For example, KU70/80 heterodimers, which

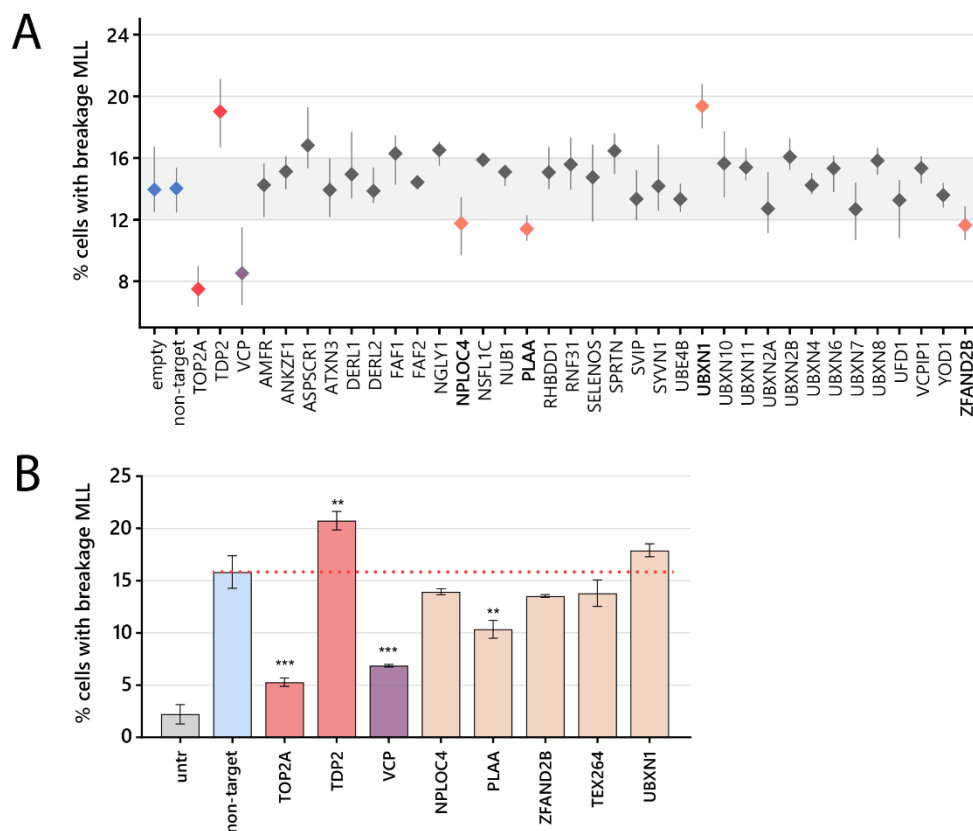
bind to DNA ends with very high affinity, rely on removal from chromatin mediated by VCP after ligation of DNA ends (van den Boom et al., 2016).

Removal of TOP2ccs by TDP2 through hydrolysis of TOP2-DNA bond is one of the main pathways to remove covalent TOP2 DPCs, but it requires remodeling or degradation of TOP2s to facilitate access to the TOP2-DNA bond (see Chapters 1.6.2 and 2.4). Previous studies have mainly focused on the pleiotropic roles of the 26S proteasome and the metalloprotease SPRTN for remodeling and removal of TOP2cc from DNA. SPRTN has been implicated in the processing of TOP1 as well as TOP2 adducts upon trapping of topoisomerases on chromatin, and it interestingly contains a ubiquitin-binding domain as well as a VCP interaction motif (Lopez-Mosqueda et al., 2016; Vaz et al., 2016). The role of VCP in SPRTN-associated TOP2cc removal has however not been studied, yet. The removal of TOP2ccs generally depends on modifications of the TOP2cc by SUMOylation or ubiquitination, which could also be the signal for VCP recruitment (Stingele et al., 2017). Proteolysis of TOP2ccs by the 26S proteasome has been described facilitate removal of TOP2ccs from DNA and inhibition of the proteasome leads to accumulation of trapped TOP2ccs (Lee et al., 2016). The mechanisms of recognition of the TOP2-DNA lesion by the proteasome is not well understood, but it might depend on transcriptional activity (Mao et al., 2001; Xiao et al., 2003). Next to the ubiquitous role of the proteasome for degradation of TOP2ccs, the SUMO ligase ZATT was identified to remodel TOP2 trapped on DNA and thereby allows access of TDP2 to the TOP2-DNA complex independently from proteasomal degradation (Schellenberg et al., 2017). We reason that VCP could be a novel modulator of TOP2ccs to facilitate TOP2 removal from chromatin by proteases and allow for downstream processing by TDP2 or MRE11.

To gain insights into how VCP could be targeted to trapped TOP2ccs, we performed a second RNAi screen for 33 known cofactors of VCP (Figure 2.42). The same setup as for the first screen was used to test *MLL* fragility, that is RNAi-mediated loss-of-function in Cal51 cells for three days followed by an acute ETO-treatment at 60  $\mu$ M and testing of *MLL* breakage by C-Fusion 3D. Contrary to the first screen, pools of four siRNAs for each target were used to improve the efficiency of knockdowns. The silencing of TOP2A, TDP2 and VCP were included as controls and showed strong reductions or increases in *MLL* breakage in comparison to the non-target control (Figure 2.42A, VCP: from 14 % to 8.5 %). From the candidate VCP cofactors, the removal of the three cofactors NPLOC4, PLAA and ZFAND2B reduced *MLL* breakage below 12 %, whereas highly elevated *MLL* fragility of 19.4 % was observed in UBXN1 RNAi samples (Figure 2.42A).

For verification of potential cofactors for VCP-dependent *MLL* breakage, another three independent experiments were performed for silencing of NPLOC4, PLAA, ZFAND2B, UBXN1

and TEX264. The factor TEX264 was included as a potential cofactor of VCP as it was reported by colleagues to mediate TOP1cc resolution at the nuclear periphery by recruiting VCP and SPRTN (Fielden et al., 2020). We confirmed a highly significant reduction in *MLL* fragility upon VCP depletion and also found a significant reduction for the cofactor PLAA (Figure 2.42B, from 15.8 % to VCP: 6.9 %, PLAA: 10.3 %,  $p < 0.01$ ). For the other candidate cofactors, significance was not reported for this experiment but further testing of the fragility of other genes would be required to get a more comprehensive understanding of cofactors' roles in TOP2-induced DSB formation. To this end, collaborative experiments with the research group of Petra Beli are underway to probe the interaction of VCP with TOP2s and to test if the interaction depends on the here identified cofactors.



**Figure 2.42 RNAi screen to identify VCP adaptor proteins that are required for *MLL* fragility upon etoposide.** (A) RNAi screen as in Figure 2.40, but with siRNA pools to target adaptors of VCP. Percentage of cells with *MLL* breakage was determined for non-transfected cells and a non-target pool (blue), for control TOP2A and TDP2 knockdowns (red), for VCP (purple) and VCP adaptors (grey) knockdowns. Three adaptors (NPLOC4, PLAA and ZFAND2B, orange) decreased *MLL* breakage below 12 %, whereas UBXN1 (orange) increased breakage to 19 % (three independent experiments, mean  $\pm$  range). (B) Validation of decrease *MLL* fragility upon knockdown of VCP and PLAA. Three independent experiments, mean  $\pm$  SD, significance by student's t-test compared to non-target control, \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

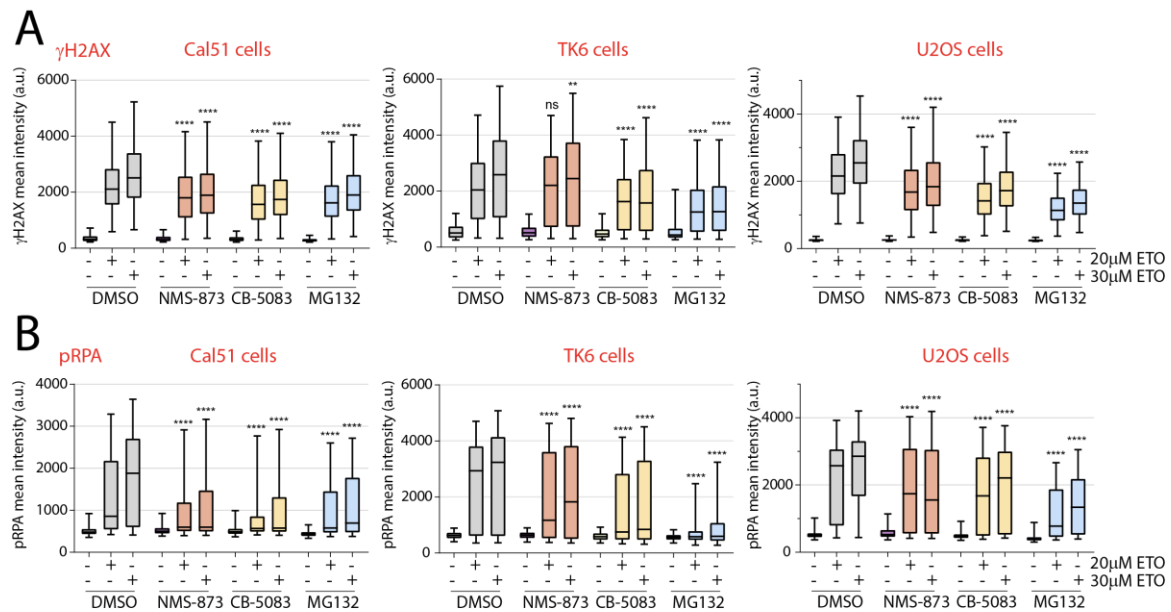
### 2.5.3 VCP mediates ETO-induced DNA damage and cytotoxicity

Our screen showed that depletion of VCP by RNAi suppressed TOP2-induced DSBs within *MLL*. We next sought to validate that this was not a cell type specific effect or an off-target effect by the siRNAs. To abrogate VCP functions fast and more efficiently than with RNAi knockdowns, we used two chemical inhibitors for VCP, NMS-873 and CB-5083, which have previously been shown to specifically and rapidly inhibit VCP functions (Anderson et al., 2015; Magnaghi et al., 2013; Tang et al., 2019).

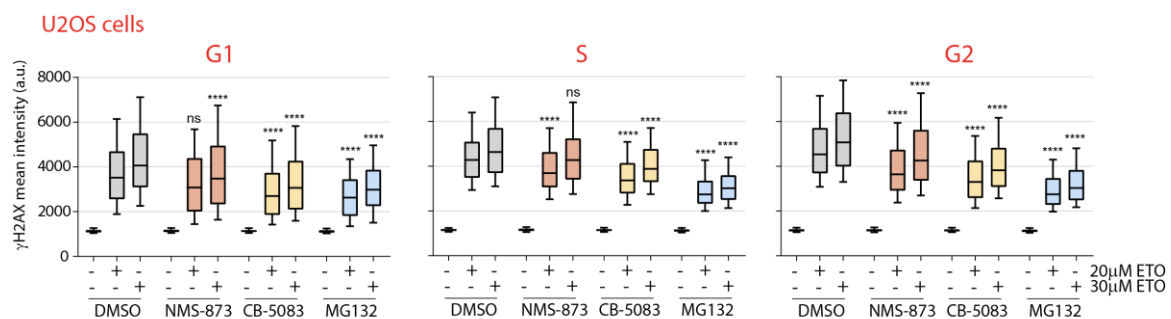
We tested DNA damage response levels by probing ETO-induced  $\gamma$ H2AX and pRPA intensities in three different cell lines upon VCP inhibition: the breast cancer cell line Cal51 (which was used for the RNAi screen), the osteosarcoma cell line U2OS and lymphoblastoid TK6 cells (Figure 2.43A, B). When VCP or the proteasome was inhibited before ETO-treatment by pretreatment with inhibitors, levels of DNA damage signaling were significantly lower in all cell lines. In U2OS cells, we furthermore quantified DNA damage signaling in different cell cycle stages and found reductions in ETO-induced  $\gamma$ H2AX signals in G1, S as well as G2-phase upon VCP or proteasome inhibition with the most prominent reductions seen in G1 and G2 phase (Figure 2.44).

Since we had previously seen the relevance of transcription and replication for ETO-induced DNA damage (see 0), we tested if VCP and proteasome inhibition affected transcription or replication activities or the cell cycle progression in different cell lines. Monitoring nascent RNA production by EU incorporation indicated that VCP, proteasome or replication inhibition did not inhibit active transcription to a relevant extend (DRB was used as a control for transcription inhibition) (Figure 2.45A). VCP inhibition by NMS-837 or transcription inhibition by DRB slightly reduced DNA synthesis by replication (EdU incorporation, replication inhibition control by APH) (Figure 2.45B). However, the cell cycle distribution was not affected by VCP inhibition in either TK6, Cal51 or U2OS cells (Figure 2.45C). We thus conclude that reductions in ETO-induced DNA damage were due to VCP specific functions and were not merely a consequence of alterations in transcription or replication activities.

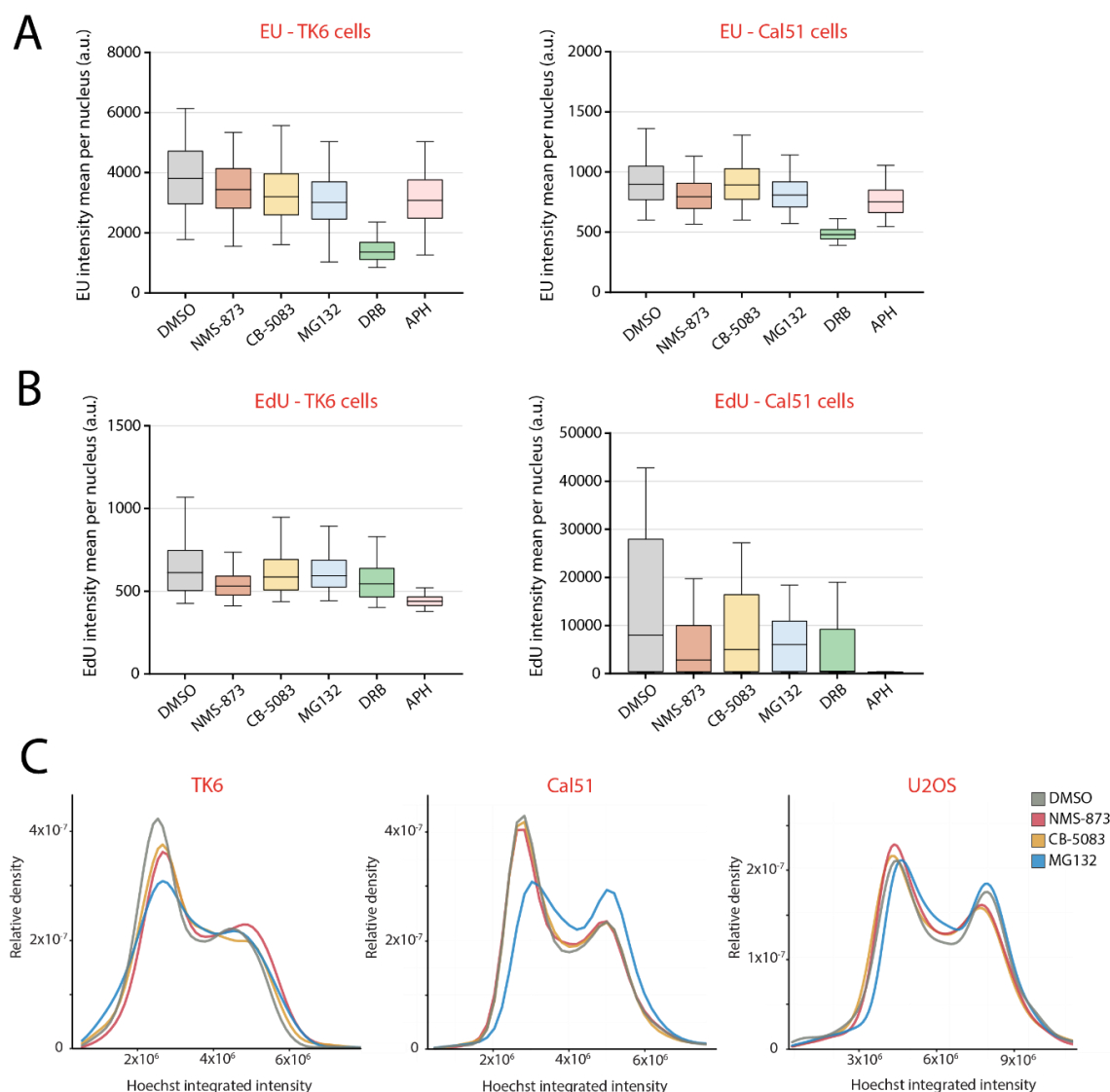
## Results



**Figure 2.43 Inhibition of VCP decreases ETO-induced DNA damage signaling.** (A, B) Cal51, TK6 or U2OS cells were pretreated with VCP inhibitors NMS-873 (5  $\mu$ M) or CB-5083 (1  $\mu$ M), proteasome inhibitor MG132 (1  $\mu$ M) or control DMSO for 3 h, and subsequently treated with 20  $\mu$ M or 30  $\mu$ M ETO for 3 h. DSB signaling was by quantifying levels of (A)  $\gamma$ H2AX phosphorylation or (B) S4/S8 phospho-RPA (pRPA). At least 1100 cells were analyzed per sample, \*\* $p < 0.01$ , \*\*\*\* $p < 0.0001$  by One Way ANOVA test compared to respective DMSO sample.



**Figure 2.44 VCP and proteasome inhibition decrease DNA damage signaling throughout the cell cycle.** Intensities of nuclear  $\gamma$ H2AX were determined in U2OS cells in G1, S and G2 cell cycle phases upon VCP or proteasome inhibition. Treatments were done as described in Figure 2.43.

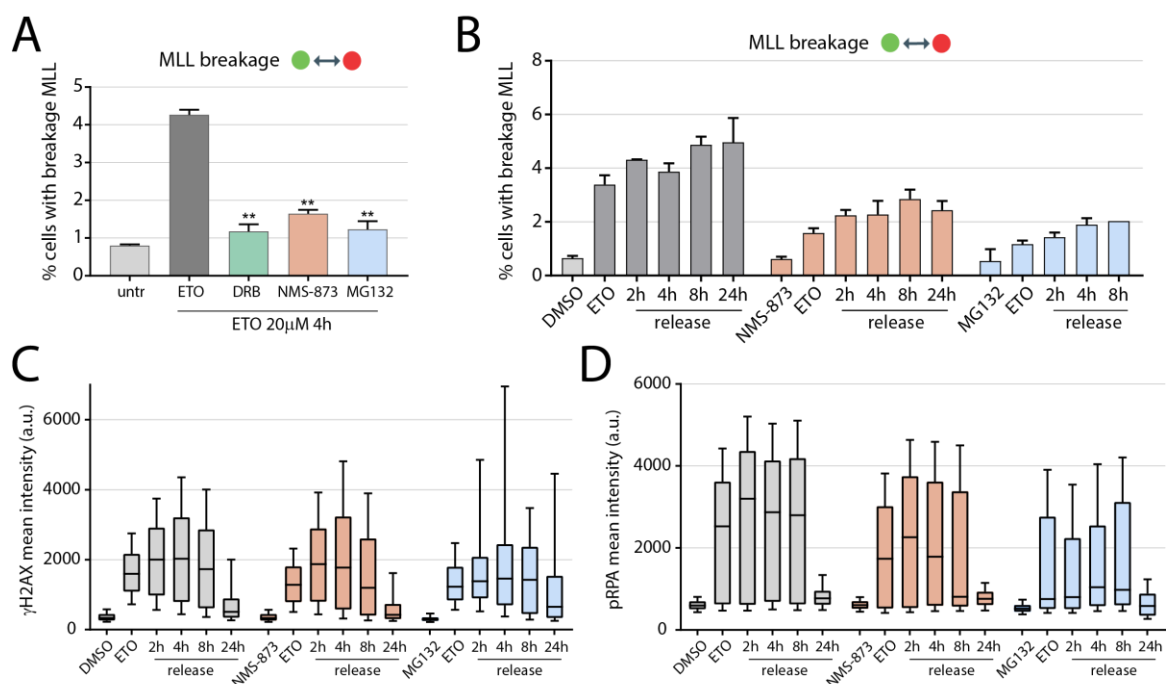


**Figure 2.45 Inhibition of VCP does not impede cellular processes.** (A) Levels of nascent RNA production indicated by EU labelling in TK6 or Cal51 cells upon inhibition of VCP (NMS-873 5  $\mu$ M, CB-5083 1  $\mu$ M), of the proteasome (MG132 1  $\mu$ M), of transcription (DRB 200  $\mu$ M) or replication (APH 8  $\mu$ M) for 3 h. (E, F) Active replication was estimated by EdU labelling by quantification of nuclear intensities in TK6 or Cal51 cells, treated as in (A). (C) Cell cycle profiles determined by integrated nuclear Hoechst33342 intensities in 6 h inhibited TK6, Cal51 or U2OS cells.

To estimate the relevance of VCP activity on *MLL* fragility, we quantified ETO-induced *MLL* breakage in TK6 cells upon inhibition of VCP (by NMS-873), of the proteasome (by MG132) or transcription (by DRB). We observed strong reductions in *MLL* fragility upon VCP and proteasome inhibition, at similar levels as upon transcription inhibition (Figure 2.46A, from 4.3 % to 1.6 % for NMS-873 and 1.2 % for MG132,  $p < 0.004$ ). We next probed ETO-induced *MLL* breakage as well as DSB-signaling for 24 h after release from etoposide and VCP or proteasome inhibitors, in order to understand if the formation of DSBs was suppressed altogether or was rather delayed upon inhibition. *MLL* fragility quantified by C-Fusion 3D

## Results

initially increased after release from ETO and plateaued after approximately eight hours of release (Figure 2.46B). In cells with VCP and proteasome inhibition, levels of *MLL* breakage were constantly lower than in control cells, during the ETO-treatment as well as during the release. Repair kinetics through signaling of ETO-induced DNA damage were observed with similar overall trends in VCP or proteasome inhibited cells compared to ETO-treated cell only, however also with lower overall levels of  $\gamma$ H2AX and pRPA intensities compared to control cells (Figure 2.46C, D).



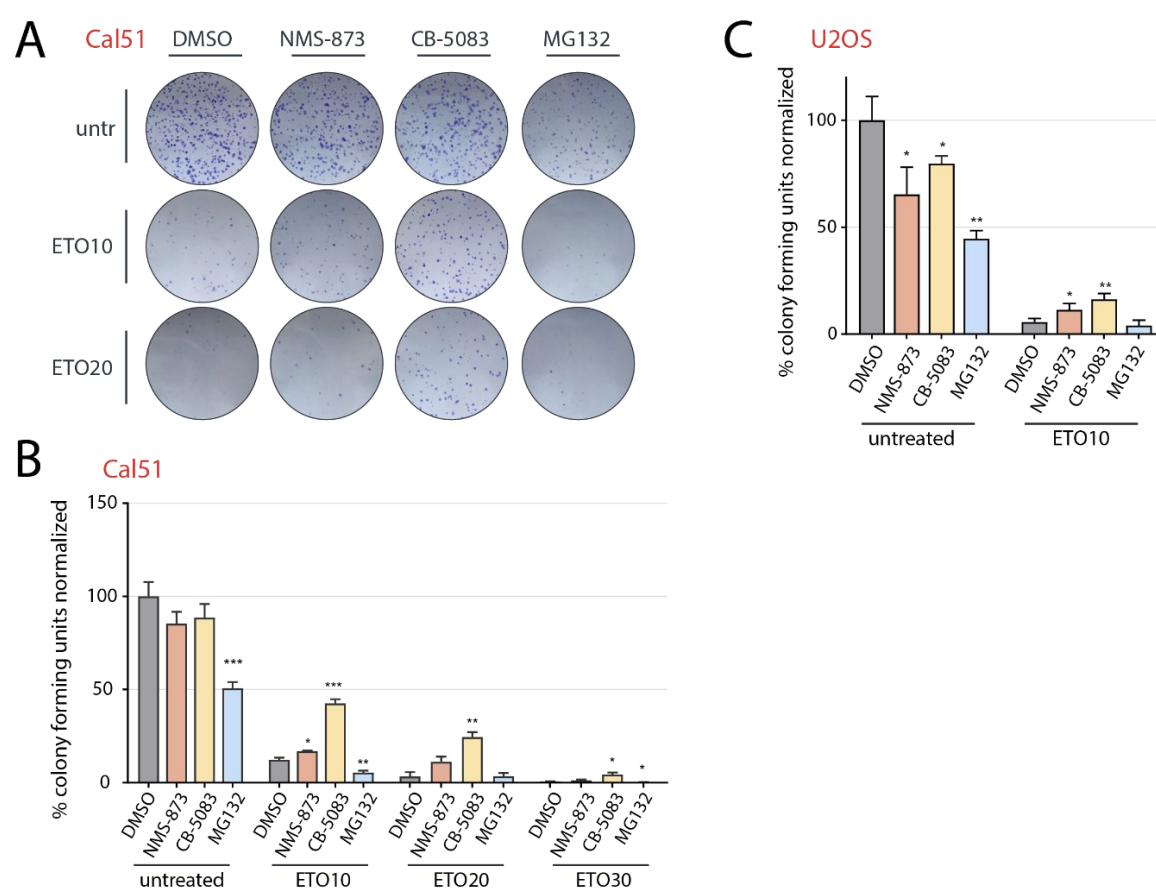
**Figure 2.46 VCP inhibition decreases ETO-induced DNA damage.** (A) Analysis of *MLL* breakage in TK6 cells by C-Fusion 3D, upon pretreatment with DRB (200  $\mu$ M), NMS-873 (5  $\mu$ M) or MG132 (1  $\mu$ M) for 3 h and followed by ETO-treatment of 20  $\mu$ M for 4 h (no release). Two independent experiments, mean  $\pm$  SD, significance by student's t-test compared to ETO control, \*\*p < 0.01. (B, C, D) Time course of *MLL* breakage,  $\gamma$ H2AX phosphorylation and S4/S8 phospho-RPA (pRPA) signaling in TK6 cells pretreated with inhibitors and etoposide as in (A) and sampled over treatment and release time points of up to 24 h release.

Finally, we investigated if the reduced DNA damage that we observed upon VCP inhibition also granted cells a survival advantage. Indeed, pretreatment of cells with NMS-873 and CB-5083 before ETO-treatments increased the colony formation capacity of Cal51 and U2OS cells (Figure 2.47A, B, C). Proteasome inhibition, on the other hand, did not improve viability of cells upon ETO-treatment, albeit its proficiency to reduce ETO-induced DNA damage. We reasoned that this could be attributed to the global importance of the proteasome pathway for survival as inhibition of the proteasome without induction of DNA damage by etoposide already decreased survival by 50 % (Figure 2.47 B, C). On the other hand, the inhibition of



VCP only mildly decreased survival in control cells, indicating a more restricted role of VCP for overall cell survival and an important role of VCP in TOP2cc processing.

Our findings thus propose that VCP could be an important factor for the conversion of TOP2ccs to free DNA DSBs. Upon removal of TOP2s from chromatin, the freed DSBs could account for gene fragility and potentially contribute to the formation of translocations. The RNAi screening approach in combination with high-throughput C-Fusion 3D was thus able to identify a novel factor in TOP2cc repair. We are thus confident that our screening approach could be applied to larger siRNA libraries and other imaging readouts, e.g. to detect genomic rearrangements or spatial positioning.



**Figure 2.47 Inhibition of VCP increases cell survival upon TOP2 poisoning with etoposide.** (A) Images of colony formation assay where Cal51 cell colonies were stained with crystal violet for visualization. Cells were seeded upon pretreatment with NMS-873 (5  $\mu$ M), CB-5083 (1  $\mu$ M) or MG132 (1  $\mu$ M) for 2 h, followed by ETO-treatment of 10  $\mu$ M or 20  $\mu$ M for 4 h and released into fresh medium. (B) Count of Cal51 colonies upon treatments as in (A). (C) Count of U2OS colonies upon treatments as in (A). The mean of three replicates is shown with SD, significance of samples compared to DMSO control, \* $p < 0.05$ , \*\* $p < 0.01$  and \*\*\* $p < 0.001$  by student's t-test.



## Chapter 3 Discussion

### 3.1 Etoposide induces DSB hotspots and leads to oncogenic *MLL* fusions

The mammalian cell nucleus is a dense environment that facilitates strong organization of the genome and accommodates various fundamental molecular processes, such as transcription, replication, splicing and epigenetic modulation. It requires a high degree of coordination and cooperation of pathways. We explored the delicate role that TOP2s have in containment of torsional stress that originates from processes, such as transcription, replication and spatial chromosome folding. It is an essential function that topoisomerases perform, but it can be highly risky for cells as it involves the transient formation of DNA DSBs. The antineoplastic effect of TOP2 poisons relies on the formation of covalent TOP2ccs and their transformation into toxic DNA DSBs. However, TOP2-targeting therapies have been associated with the occurrence of secondary acute leukemias that are driven by oncogenic *MLL* fusions. We aimed at understanding the molecular mechanisms that contribute to the formation of ETO-induced genomic instability and to the formation of *MLL* translocations.

To elucidate how TOP2 poisoning is linked to genomic instability, we combined genome-wide profiling of DSBs by sBLISS with quantitative detection of chromosome breakage and translocations by C-Fusion 3D. sBLISS in several human hematopoietic cell lines and bone marrow progenitor cells revealed that trapping of TOP2s by etoposide leads to the accumulation of DSB hotspots throughout the genome and that DSB hotspots strongly correlate with active chromatin regions, RNA-PolII binding and anchor sites of spatial chromosome folding. Intriguingly, localization of endogenous DSBs highly overlapped with breaks found upon TOP2-trapping by ETO and a similar distribution of endogenous and ETO-induced breaks across genomic features was seen. This suggests that a substantial fraction of endogenous DSBs could arise from abortive topoisomerase functions. Indeed, topoisomerases have been described as some of the most dangerous enzymes for cells as their activity involves formation of DNA breaks and the formation of transient but covalent protein-DNA complexes (Deweese and Osheroff, 2009). On the other hand, it has been proposed that activity-induced DSBs mediated by TOP2B are required for transcription (Ju et al., 2006; Madabhushi et al., 2015). The contribution of TOP2s to endogenous DNA damage is an interesting question for which sBLISS and C-Fusion 3D could be useful tools to understand the perks and perils of topoisomerase activities.

The *MLL* fusion hotspot at exon 12 accumulates DSBs upon ETO-treatment, whereas at other sites of the *MLL* BCR breaks were not frequently observed. Common fusion partner genes experienced broader distributions of DSBs within their gene bodies, arguing that specific fusion events of *MLL* and partner genes that are found in leukemic patients are selected over fusions that do not lead to cell transforming potentials. C-Fusion 3D analysis indicated persistent gene fragility in *MLL* as well as common fusion partner genes. *MLL* appeared most fragile with up to 10 % of cells showing a gene breakage. In contrast, *MLL* fusions were rare and occur at levels of 0.6 to  $2 \times 10^{-3}$  in lymphoblastoid TK6, erythroleukemic K562 and progenitor CD34+ cells. Hence, we find that TOP2 poisoning induces distinct hotspots of DSBs and that C-Fusion 3D can detect oncogenic *MLL* fusions upon formation in hematopoietic cell lines and progenitor cells.

### 3.2 The dangerous interplay of transcription, DNA topology and genome architecture

How the conversion of TOP2ccs to DSBs occurs has been a long standing question in the field and mechanisms that traverse through DNA, such as transcription or replication, were proposed to contribute to TOP2-linked genomic instability. In addition, the formation of chromatin architecture has been suggested to be a major force in the generation of genomic instability upon etoposide treatments. Here, we shown that transcription is the main driver of genome breakage upon TOP2-poisoning and is interlinked with spatial chromosome folding to generate genomic instability.

#### 3.2.1 Transcription drives TOP2-linked gene fragility

By correlating sBLISS DSBs profiles with genomic regions, we found that ETO-induced DSBs preferentially form in transcriptionally active genomic regions, mainly in promoters. Strong correlations with RNA-PolII occupancy, the S2-phosphorylated and S5-phosphorylated forms of RNA-PolII and open chromatin were seen, indicating a role of active transcription to DSB formation. Furthermore, the fragility of genes was directly linked to their activity. The higher genes express, the more DSBs were formed endogenously and upon ETO-treatment. Importantly, transcriptional inhibition led to reduced DNA-damage signaling and also led to less DSBs at promoters as well as gene bodies that were detected by sBLISS.

This is in line with studies that show TOP2 binding and cleavage sites in regions of open chromatin and regions that are transcriptionally active (Uusküla-Reimand et al., 2016; Yu et

al., 2017). Alongside, endogenous DSBs identified by BLESS were also found to accumulate at promoters (Bouwman and Crosetto, 2018; Yang et al., 2015). Furthermore, transcription inhibition was found to reduce levels of *MLL* breakage and the formation of chromosome rearrangements. However, only *MLL* breakage was studied and not genome-wide fragility (Gómez-Herreros et al., 2017). Here, we find direct evidence that the formation of protein-free DSBs upon ETO-treatment is indeed largely transcription-dependent throughout the genome. This was in contrast to a study by the Nussenzweig group where TOP2-induced breakage was reported to be transcription-independent. Interestingly, TOP2-linked breaks that were mapped by END-Seq (including protein-bound DNA breaks) were identified in more than 50 % of active promoters, but very few changes were observed upon inhibition of transcription elongation (Canela et al., 2017). More recent findings with the amended END-Seq to detect protein-free DSBs showed that the conversion of TOP2ccs to free DSBs and the formation of chromosome translocations upon etoposide are indeed transcription-dependent (Canela et al., 2019). These findings propose that TOP2 binding and the formation of TOP2-linked lesions are mainly correlating with CTCF and cohesin binding and are transcription-independent, whereas protein-free DSB are predominantly driven by transcriptional activity (Canela et al., 2017, 2019). However, loop-driven TOP2-lesions were studied in quiescent mouse B cells that were stimulated to become activated from the G<sub>0</sub>-phase of the cell cycle. In this cell system, transcription-related processes were not found but could have been underestimated by a relatively short activation of 12 hours and short ETO-treatments (30 min) (Canela et al., 2017; Kouzine, Wojtowicz, et al., 2013). Their recent findings indicate though that the formation of DSBs through transcription requires several hours to unfold (Canela et al., 2019). The contribution of transcription to protein-bound DSBs thus remains elusive. In our data, we also detected a subset of ~40 % ETO-induced DSBs that overlapped with either CTCF or cohesin binding sites. Anyhow, we did not study the distribution of TOP2 binding in hematopoietic cells, for example by ChIP-seq, to uncover what determines sites of TOP2 binding and TOP2-DNA lesions in our cell models. In conclusion, we find that transcription is the predominant force to drive the formation of free DNA breaks, which is a prerequisite for the formation of chromosome rearrangements.

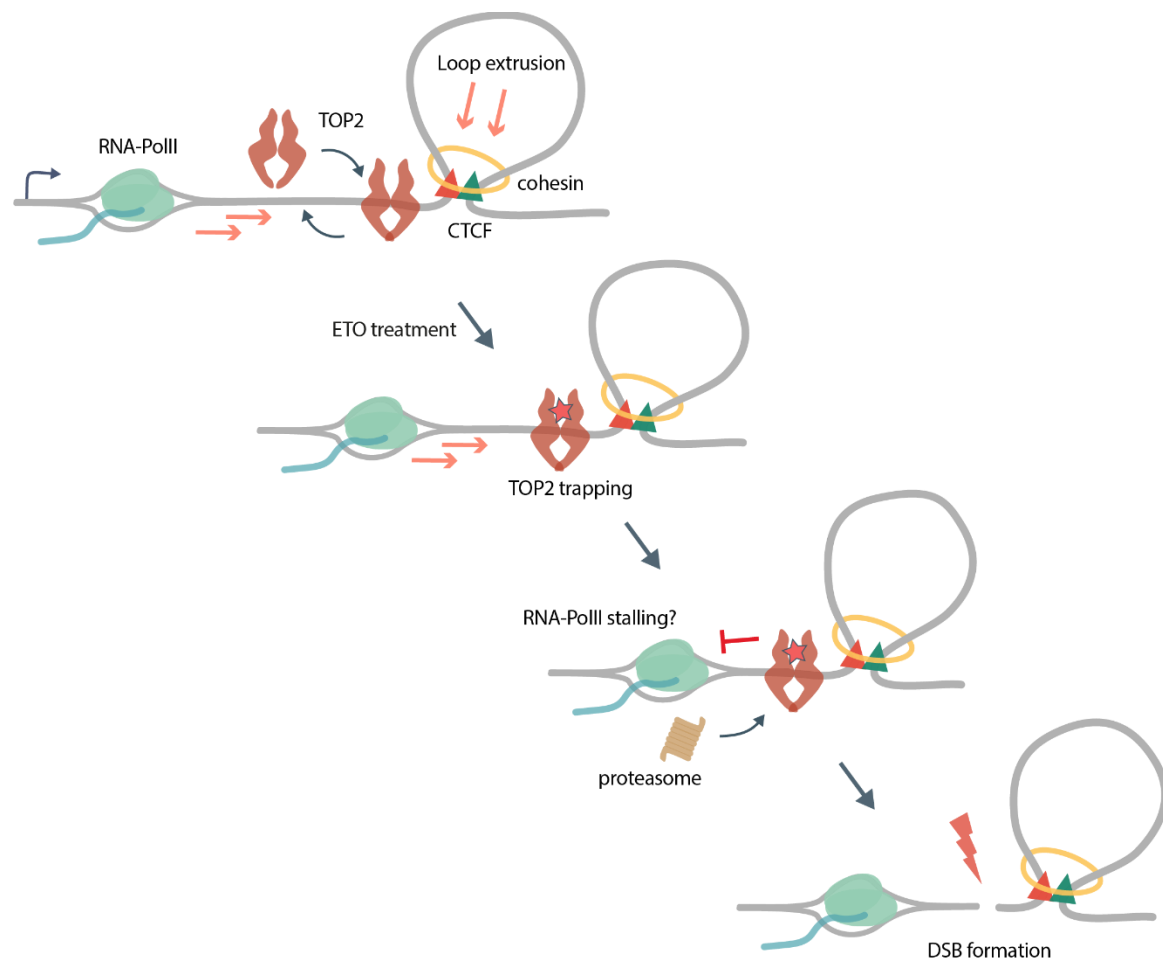
### 3.2.2 Transcription and chromosome organization are interlinked in generating DNA damage

We have seen that the binding and activity of TOP2s were proposed to correlate with chromatin loop anchor sites whereas transcription triggers the conversion of TOP2ccs into DSBs. Nevertheless, TOP2 binding has been found to also depend on gene activity (Kouzine,

Gupta, et al., 2013). Interestingly, we find that highly transcribed genes are more likely to be at or close to loop boundaries than silent genes. Thus, there is the possibility that TOP2 localizing to loop anchor sites and correlations with promoter activity could go hand-in-hand (Canela et al., 2017; Kouzine, Gupta, et al., 2013; Uusküla-Reimand et al., 2016; Yu et al., 2017). Indeed, a transcription-dependent genome organization has been proposed, where boundaries of DNA supercoiling arising from transcription are marked by CTCF, thereby controlling the propagation of DNA topology into neighboring domains. This concept would be supported by the finding that supercoiling at CTCF binding sites depends on TOP2 activity and is transcription-dependent (Uusküla-Reimand et al., 2016). We find that gene activity and localization at loop boundaries contribute to ETO-induced gene fragility synergistically (Figure 3.1). Many of the recurrently observed *MLL* fusion partner genes share the characteristics of being transcriptionally active, they localize at or close to loop boundaries and experience high gene fragility upon TOP2 poisoning. Importantly, proximity to loop anchors further increased the accumulation of DSBs in genes, but only in highly transcribed genes. Thus, the fragility of genes upon ETO-treatment is mainly driven by gene activity and increases upon an interplay with loop boundaries.

When looking closer at breakage around CTCF-bound loop anchors, we observed that DSBs at intragenic anchors were asymmetrically clustered towards the direction of transcription, both outside and within loops. The asymmetric fragility at loop boundaries was dependent on transcriptional activity and directionality, suggesting that loop boundaries are vulnerable to DNA breaks from abortive TOP2 complexes that relieve transcription induced topological stress. Loop anchors present a barrier to DNA supercoiling generated by transcription, which may require the cumulative action of TOP2. This is supported by the finding that CTCF binding flanks boundaries of transcriptionally active supercoiling domains, TOP2s interact with CTCF, and TOP2 activity at CTCF sites has been shown to regulate transcription-dependent supercoiling (Naughton et al., 2013; Uusküla-Reimand et al., 2016). Alternatively, loop extrusion might require the activity of TOP2 and a cohesin-dependent localization of TOP2B across the genome was found as well as the accumulation of TOP2-activity outside of loop boundaries (Canela et al., 2017, 2019). Our research could however not find a major contribution of looping directionality to the formation of DSBs upon ETO-treatments. It could thus be speculated that loop extrusion contributes to the positioning and activity of TOP2s but that encounters of the transcription machinery with TOP2ccs at loop boundaries after all trigger their conversion to DNA breaks. Nevertheless, break induction through loop extrusion might hold true for a specific subset of break hotspots, since we do find a few DSB hotspots at loop anchors that are distal to active genomic regions. We thus conclude that loop anchors

are hotspots of TOP2-induced DNA fragility, which predominantly depends on an interplay with transcription.



**Figure 3.1 The interplay of transcription with chromatin loop anchors leads to TOP2-induced DSB formation.** The traversing of transcription through DNA and loop extrusion lead to topological stress that requires the activity of TOP2s, at genes and loop anchors. Upon TOP2 trapping by etoposide, a transcription-blocking protein-DNA lesion is formed that might lead to RNA-PolII stalling and to transcription-dependent recruitment of the proteasome. The formation of DSBs is predominantly mediated by transcription and is enriched at CTCF bound chromatin loop anchors. The localization of breaks at loop boundaries relies on transcription activity and directionality.

The connection of transcription and spatial chromosome organization in DSB formation raises the question if and how transcription and chromatin looping coordinate their activities. Transcription has been suggested to play a direct role in chromosome architecture formation. Movement of cohesin along chromatin was proposed to happen through the movement of the transcription machinery (Szabo et al., 2019). Simulations of chromatin fiber topology have suggested a more indirect role of transcription where transcription-induced supercoiling could push cohesin and thereby contribute to chromatin loops (Racko et al., 2018). However,

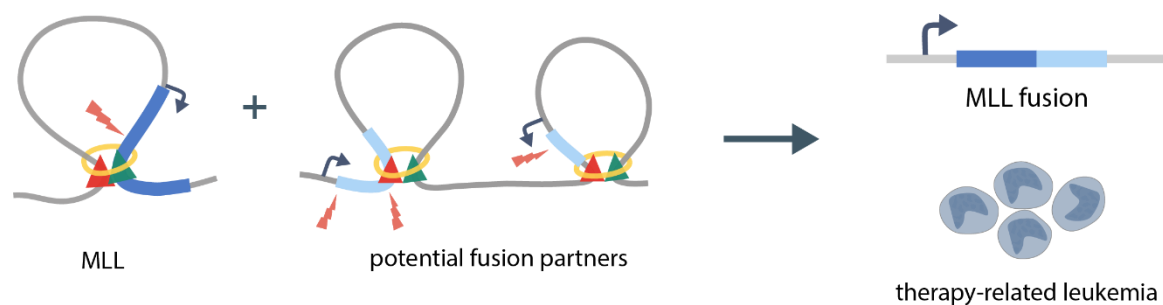
transcription has not been found to substantially contribute to chromatin loop formation upon establishment of loops. Modest changes in loop strength could be observed upon transcription inhibition as well as small reductions in colocalization of cohesin with loop anchor sites (Vian et al., 2018). Researchers thus argue that the establishment of loop domains is transcription independent. However, it has been proposed that the cohesin distribution could indeed be dependent on transcription as without CTCF and the cohesin release factor WAPL, cohesin is redistributed to the 3' ends of active genes (Busslinger et al., 2017). The translocation of cohesin by transcription could be a mechanism to prevent interference of chromatin loops with the opening of the DNA-duplex by RNA-PolII. A displacement rather than relocation of cohesin and CTCF upon transcription elongation could reconcile the above findings (Heinz et al., 2018). Torsional stress and entanglements generated by transcription might be transduced through chromatin up until stop sites of cohesin at CTCF motifs, and could then be released by the actions of TOP2s. We hypothesize that upon failure of TOP2s to complete their catalytic cycle, the accumulation of torsional stress or collisions of transcription with the loop boundaries would convert abortive TOP2cc into DSBs. New insights into how transcription, chromatin topology and gene fragility go hand-in-hand require correlations of DSB-mapping and supercoiling-mapping with genome-wide high-resolution TOP2 and RNA-PolII profiles as well as activity measures in human cells. So far, we conclude that interactions of transcription with chromosome architecture promote TOP2-induced genome fragility.

### 3.2.3 Breakage at the *MLL* DSB hotspot is enriched in proximity to a chromatin loop anchor and depends on gene activity

We find that genes that recurrently fuse with *MLL* are more frequently located at loop boundaries than random genes and they are often highly expressed. The *MLL* gene itself is among the top 10 % of expressed genes in hematopoietic cells and contains a CTCF motif close to the exon 12 DSB hotspot. Looping analysis revealed that *MLL* is indeed located at the end of a chromatin loop that reaches into upstream regions in chromosome 11. Importantly, the breakage hotspot is localized upstream of the CTCF binding site towards the TSS of *MLL* and thus towards the direction of *MLL* transcription (Figure 3.2). Mutations of the CTCF consensus motif in *MLL* prohibited CTCF binding and resulted in slightly decreased *MLL* breakage (by ~30 %). However, only the inhibition of transcription abolished the formation of *MLL* breaks upon etoposide. We thus find that the loop anchor within the *MLL* gene contributes to its gene fragility and that *MLL* breakage is predominantly transcription-dependent. Interestingly, binding sites of CTCF are often found in gene dense regions. Most CTCF binding sites occur in



intergenic regions (46 %) but a substantial fraction is found within promoters (20 %), exons (12 %) and introns (22 %) (Kim et al., 2007). One could hypothesize that active genes located at loop boundaries could be evolutionary more susceptible to rearrangements, such as seen here for *MLL*. Loop anchor points have been found to be enriched for genomic structural variants and were associated with evolutionary developments or mutation hotspots in cancers (Kaiser and Semple, 2018). To better understand the potentially dangerous interplay of transcription with CTCF binding in genes, further analysis to understand endogenous sources of DNA breaks are required that, as we report here, highly correlate with gene activity.



**Figure 3.2 *MLL* and common fusion partners localize at loop boundaries and are sensitive to transcription-dependent DSB formation.** The *MLL* exon 12 DSB hotspot colocalizes with CTCF binding at a loop anchor. *MLL* fragility depends on gene expression and breaks localize towards the direction of transcription. Many fusion partner genes are highly transcribed, close to chromatin loops and they accumulate high levels of TOP2-dependent DSBs. Upon rearrangements with *MLL*, oncogenic fusion proteins are thereby formed that can lead to the onset of therapy-related leukemias.

#### 3.2.4 Transcription- dependent *MLL* fusions and spatial proximity

We show that the fragility of recurrent *MLL* fusion partner genes and the formation of *MLL* fusions is predominantly transcription-dependent and to some extent replication-dependent. Loss of chromosome breaks upon RNA-PolIII degradation indicate that it is indeed the transcription machinery traversing through DNA that is required for TOP2-induced gene fragility. Furthermore, *MLL* transcription levels were directly linked to *MLL* breakage, showing that transcriptional activity on the gene levels is a driver of instability (Figure 3.2).

A prerequisite for the formation of recurrent chromosome translocations is the formation of DNA breaks, but spatial proximity is also required for the formation of illegitimate fusions (Hakim et al., 2012; Hu et al., 2016; Roukos, Voss, et al., 2013; Zhang et al., 2012). We find that common *MLL* fusion genes show similar break frequencies upon ETO-treatment, but we

observe that the *MLL-ENL* fusions arise more frequently than fusions with other partners. This could be explained by *ENL* being inherently more proximal to *MLL* than other common fusion genes in hematopoietic cell lines and CD34+ progenitor cells. Transcription has been speculated to favor recurrent fusions of *MLL* with *AF9* and *AF4* by enhancing spatial proximity of *MLL* and fusion partners through sharing of transcription factories (Cowell et al., 2012). The concept of proximity through activity has for example been reported to induce colocalization of IgH and MYC upon their activation, which recurrently fuse to induce Burkitt's lymphoma (Osborne et al., 2007). We could however not identify any changes in *MLL* and partner genes' spatial distances or synapsis rates upon perturbations of either transcription or replication under control conditions. Further experiments that locally modulate the transcriptional activity of potential *MLL* fusion genes would generate more precise insights about the influence of transcription factories on *MLL* translocation partner selection.

Moreover, we observe an increase in synapsis of *MLL* and *ENL* upon ETO-treatments, which is disrupted by transcription inhibition, indicating that transcription-dependent DNA damage favors the colocalization of already proximal genes. Clustering of transcriptionally-active genes with DSBs was previously shown to occur between different chromosomes (Aymard et al., 2017). This phenomenon was primarily found in G1-phase cells upon induction of breaks by endonucleases, irradiation or etoposide (Aten et al., 2004; Aymard et al., 2017; Caron et al., 2015). Nonetheless, the functions of DSB clustering are not clear as the chances of illegitimate repair of DNA ends to form chromosome rearrangements are thereby also increased. It has been hypothesized that the repair efficiency could be promoted by clustering of repair factors around several DSBs. However, the clustering of DSBs within active regions was rather linked to delayed repair kinetics (Aymard et al., 2017). It was hypothesized that a delay in repair facilitates a transition of cells through the G1/S checkpoint to promoted repair by more faithful homology-directed repair (Guénolé and Legube, 2017). For *MLL* and *ENL*, we did not study the influence of the cell cycle on colocalization probabilities and repair kinetics, yet. To study mechanism of synapsis more systematically, the distances between *MLL*, recurrent fusion genes and various loci with different genomic context could be studied with or without DNA damage and transcription inhibition, by implementing multiplexing of C-Fusion 3D (Gelali et al., 2019). Labelling of many gene combinations by, for example, Oligopaint FISH could generate new insights into how genes cluster upon DNA damage induction, the influence of clustering on repair kinetics and effects of cell cycle progression and if this contributes to the formation of oncogenic translocations.

Moreover, *MLL* fusion data from patients indicate that *MLL-AF9* and *MLL-AF4* are the most frequent fusion and are found in ~50 % of t-AML patients and t-ALL patients, respectively

(Meyer et al., 2018). Although we find higher fusion rates for *MLL* with *ENL* than with other recurrent partners, these fusions only occur in ~10 % of t-AML and t-ALL cases. This could result from varying transformation capacities of the many *MLL* fusions and thereby selective advantages of some fusions during cancer formation. Indeed, the different *MLL* fusions have been reported to confer varying transformation capacities which strongly depended on the cell lineage (Drynan et al., 2005; Krivtsov and Armstrong, 2007; Lin et al., 2017). We thus conclude that TOP2-induced translocation formation depends on transcription-dependent gene fragility, spatial positioning of fusion partners and the resulting transforming capacity that arises from gene fusions.

### 3.3 Implications of TOP2 isoform specificity

Topoisomerases of type II comprise two isoforms, TOP2A and TOP2B, which share enzymatic function but regulate DNA topology in different cellular contexts. The roles of the two isoforms in cellular mechanisms, such as transcription, replication, chromatin loop formation and mitosis is an active field of research. Here, we studied the contributions of the two isoforms to the induction of DNA damage and chromosome rearrangements upon TOP2 poisoning by etoposide.

#### 3.3.1 Both TOP2 isoforms contribute to ETO-induced DNA fragility

We find that TOP2A and TOP2B both contribute to ETO-induced DNA damage signaling as well as to *MLL* breakage and oncogenic *MLL* translocations. Using two engineered TOP2A depletion cell systems, we find that TOP2A has a more prominent role in causing DNA fragility upon ETO-treatment than TOP2B. This is contrary to the previously reported dominant role of TOP2B as the main isoform causing DNA breaks (Azarova et al., 2007; Canela et al., 2017; Cowell et al., 2012). On the other hand, the cytotoxicity of etoposide was previously mainly attribute to trapping of TOP2Accs and a dependency of the DNA damage marker  $\gamma$ H2AX upon TOP2A poisoning was found (Burgess et al., 2008; de Campos-Nebel et al., 2010; Pendleton et al., 2014). We see that DNA fragility in different cell lines does not depend on the two isoforms to the same extent. In engineered HTETOP cells, that express TOP2A at unusually high levels, only TOP2A-dependent DNA damage signaling and *MLL* breakage was found whereas in Cal51 cells, that express both isoforms at average levels, we observed TOP2A and TOP2B-dependent DNA damage. We find variable expression levels of the TOP2 isoforms in different transformed cell lines, which suggests that the expression status of TOP2s in cells is an important indicator of their isoform dependency. In non-transformed cells, it has been shown

that TOP2A expression is tightly linked to the proliferation status of cells whereas TOP2B expression is rather linked to the differentiation status of cells, but is expressed in all cell types (Pendleton et al., 2014; Tiwari et al., 2012). We find increased levels of ETO-induced DNA fragility in S/G2 phases of cycling human cells in which both TOP2 isoforms are expressed. This finding could be important in light of the differentiation state of cells where fully differentiated cells exit the proliferative cycle and thus dispense TOP2A activity.

### 3.3.2 Therapeutic implications of TOP2 isoform specificity

The development of isoform specific TOP2 poisons has been promoted and might help to target a specific cell type, which predominantly expresses one or the other isoform. In fact, several cancer types, such as breast cancer and colon cancer, were reported to commonly have elevated expression of TOP2A. The development of TOP2A-specific drugs could thus improve the efficacy of the antineoplastic effect in cancer cells by TOP2-poisoning (Pommier, 2013). However, we cannot report that the cytotoxicity of etoposide can be uncoupled from its genotoxic effect to induce DNA DSBs. We rather suggest that the cytotoxicity of ETO is caused by the induction of genomic instability and DNA breaks. To verify this, further experiments should be conducted that test the survival, apoptosis signaling, DNA damage signaling and DNA breakage for dependencies on the TOP2 isoforms. Initial experiments have indicated a linkage of genomic instability and cytotoxicity where cells lacking TOP2A, TOP2B or both, that show lower levels of DNA fragility also show an improved survival (data not shown). Thus, to prevent the formation of oncogenic *MLL* translocations that are related to TOP2 therapies, more targeted approaches, that only prohibit DNA damage within the *MLL* gene, could be a perspective. Since the formation of ETO-induced DSBs within *MLL* relies on active transcription, the modulation of *MLL* expression could thus potentially improve the otherwise very effective TOP2-targeting chemotherapies.

## 3.4 The multifaceted repair of TOP2-DNA complexes

The collective DNA damage response in human cells includes hundreds of proteins and numerous pathways to protect the genomic DNA from DNA damage. Of the many hazardous structures, crosslinks between proteins and the DNA are types of lesions that trigger responses from highly conserved repair enzymes, such as tyrosyl-DNA phosphodiesterase (Ashour et al., 2015). We here report two main axis of TOP2-DNA complex repair involving either TDP2 or MRE11 that act non-epistatically to revert TOP2ccs and facilitate DSB repair by NHEJ (Figure 3.3). While repair by TDP2 depends on active transcription, the endonuclease

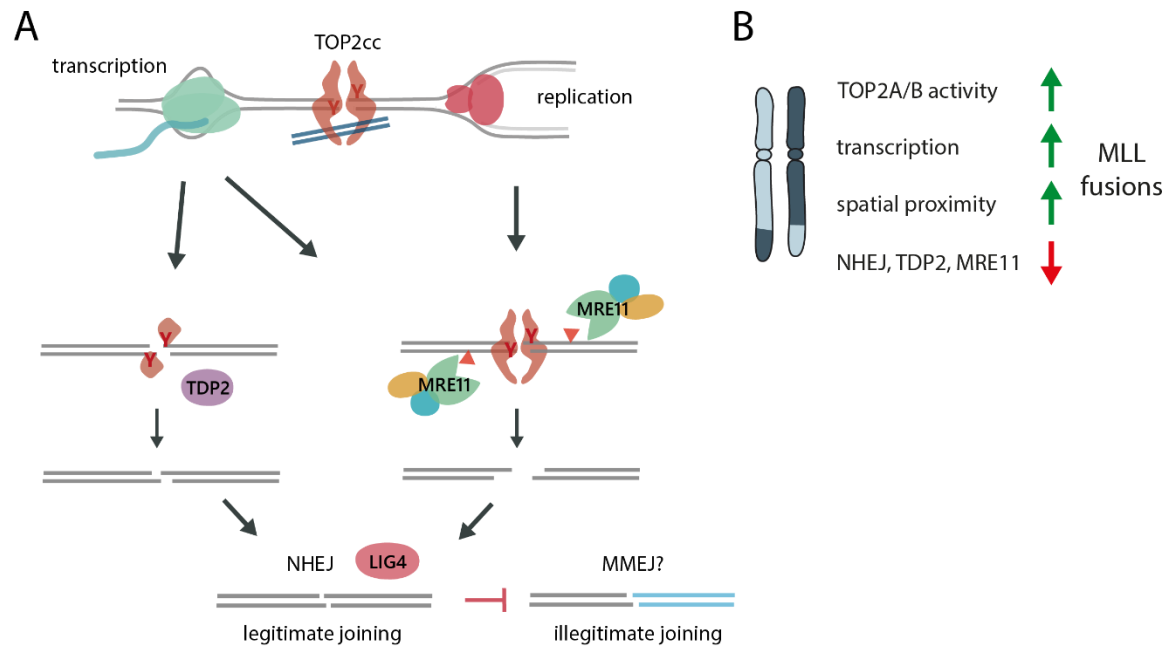
activity of MRE11 suppresses *MLL* breaks and translocations that arise during transcription and replication.

#### 3.4.1 MRE11 and TDP2 are suppressors of *MLL* fragility and fusions

We employed isogenic TK6 cells deficient for TOP2cc repair factors to find that endonucleolytic cleavage by MRE11 as well as the phosphodiesterase TDP2 suppress the formation of ETO-induced gene fragility. Deficiencies of these factors increased *MLL* breakage and translocations to similar extents, indicating that MRE11 and TDP2 both suppress *MLL* fragility and that they do not fully compensate one another. A non-epistatic relationship between the two pathways was found, indicated by high levels of additional *MLL* fragility in double-mutant MRE11<sup>-/nd</sup> TDP2<sup>-/-</sup> cells. This is in agreement with previous studies that have suggested two independent pathways for TDP2 and MRE11 to repair TOP2-DNA adducts (Hoa et al., 2016). Our findings show that TDP2 protects from TOP2-induced DNA damage and oncogenic *MLL* fusions that are predominantly arising during transcription activities, as has previously been indicated for *MLL* breakage (Gómez-Herreros et al., 2017). Contrary, nucleolytic suppression of TOP2-induced breaks by MRE11 depends on transcription and replication (Figure 3.3). MRE11 has previously been connected to replication, based on the association of MRE11 with replication forks and its broader activity to counteract replication instability (Syed and Tainer, 2018). Upon TOP2-induced DNA damage, the activity of MRE11 however goes beyond a role of TOP2cc removal during replication since we also observe a strong dependency on transcription activity. S-phase independent suppression of ETO-induced DSBs by MRE11 has previously been found in G1 cells, showing a complex and multifunctional role of MRE11 in PDB repair that is independent of HR (Hoa et al., 2016; Quennet et al., 2010). We also observed abrogated phosphorylation of H2AX in MRE11 knockout and endonuclease-deficient cell lines. This was somewhat surprising as we find highly elevated levels of gene fragility in MRE11-deficient cells, indicating that cells do encounter higher levels of DNA damage than wild-type cells, especially in MRE11<sup>-/nd</sup> TDP2<sup>-/-</sup> double mutant cells. MRE11 is important for the activation of the ATM and ATR kinases, which could explain reduced DDR signaling in MRE11 knockout cells that mainly depends on kinase activities of ATM but also ATR upon etoposide (Álvarez-Quilón et al., 2014; Sun et al., 2017). Nonetheless, the nuclease-deficient MRE11 only comprises a point mutation to inactivate its endonucleolytic function (Hoa et al., 2015). It thus remains to be tested if the MRE11 endonuclease deficiency affects the recruitment of MRE11 to sites of PDBs and how DDR kinase signaling could thereby be impaired upon TOP2-induced DNA damage.

### 3.4.2 NHEJ counteracts the formation of oncogenic *MLL* fusions

The formation of oncogenic fusions requires erroneous joining of DNA ends *in trans* by DSB repair pathways (Roukos and Misteli, 2014). We tested the contribution of HR and NHEJ to ETO-induced formation of *MLL* translocations and found that deficiency in NHEJ (*LIG4*<sup>-/-</sup>) increased persistent DSB and the formation of *MLL* fusions. Contrary, the depletion of HR factors RAD54 and CtIP did not influence TOP2-induced gene fragility. We thus conclude that NHEJ is the main pathway to repair ETO-induced DSBs and that classical end joining facilitates rapid intrachromosomal repair to suppress oncogenic fusions. NHEJ has however been described as an error-prone repair pathway that facilitates illegitimate joining of distal chromosome ends (Biehs et al., 2017; Ghezraoui et al., 2014). Interestingly, the depletion of NHEJ in combination with HR (*Lig4*<sup>-/-</sup> *RAD54*<sup>-/-</sup>) did not further increase DNA damage signaling or *MLL* fragility when compared to NHEJ-deficient cells only. Thus, our data suggests that in the absence of classical NHEJ, HR does also not substantially contribute to the repair of TOP2-induced DSBs. Alternative repair mechanisms, such as microhomology-mediated end joining (MMEJ), could take over DSB repair to ligate DNA breaks by activities of alternative ligases LIG I or III. MMEJ is associated with the use of microhomology to stabilize DNA ends, leading to small deletions or insertions (McVey and Lee, 2008). Hence, although NHEJ does not probe the legitimacy of DNA end fusions, its high processivity could prevent synapsis of several DSBs. In the absence of NHEJ, DNA ends could have more time to disconnect and colocalize with other DSBs. The search of microhomology for MMEJ furthermore requires resection of DNA ends and could increase the risk of DSB contacts *in trans*. Additional experiments with complementary techniques would further clarify which repair mechanisms contribute to the biogenesis of *MLL* translocations in the absence of NHEJ, since C-Fusion 3D does not provide absolute information about the physical ligation of DNA ends. The contributions of end joining pathways to chromosome rearrangements are currently investigated by Vera Minneker in the Roukos group.



**Figure 3.3 TDP2, MRE11 and NHEJ are suppressors of TOP2-induced DSBs and illegitimate fusions.** (A) TDP2 and MRE11 are required for processing of irreversible TOP2ccs into protein-free DSBs that can be repaired by NHEJ to facilitate fast and *in cis* DNA end joining. TDP2 suppresses transcription-dependent gene fragility whereas MRE11 has functions in processing transcription and replication-mediated TOP2-lesions. Protein-free DSBs are readily ligated by NHEJ, which facilitates legitimate joining. In the absence of NHEJ more fusion events are seen that might be mediated by MMEJ. A role of HR was not observed. (B) *MLL* fusion are mediated by TOP2A and TOP2B isoforms and they predominantly depend on transcriptional activities. Illegitimate joining is favored by spatial proximity of potential fusion partner genes that can be suppressed by TOP2cc processing by TDP2 and MRE11 or end joining by NHEJ.

### 3.4.3 VCP as a novel processing factor of TOP2ccs?

The removal of TOP2ccs from DNA is part of a multistep repair cascade that involves the detection of aberrant TOP2 complexes, signaling for proteolytic degradation, reversal or removal of TOP2-DNA bonds and downstream DSB repair. Efforts have been made to better understand the initial steps of TOP2cc recognition and induction of PDB repair mechanisms. It was only recently uncovered that modifications of TOP2-DNA complexes by SUMO and ubiquitin are linked to remodeling and degradation of TOP2ccs by ZATT and the proteasome (Schellenberg et al., 2017; Sun et al., 2019). In fact, an enhanced DNA damage response has been reported to depend on proteasomal activity that processes TOP2ccs into irreversible TOP2-DNA lesions (Sciascia et al., 2020).

We see similar dependencies of TOP2-induced DDR signaling and DSB formation on VCP activity, which we identified in a targeted screen to reduce *MLL* fragility upon ETO-treatment. VCP is a multifunctional protein that often recognizes ubiquitin-modified substrates and

targets them for degradation by the proteasome or for recycling by deubiquitination. Our findings propose that VCP contributes to the conversion of TOP2ccs to free DNA DSBs. Depletion of VCP by RNAi highly reduced *MLL* breakage upon ETO-treatment, indicating that VCP mediates TOP2-induced gene fragility. We could also observe a decrease in *MLL* fragility as well as reductions in global DNA damage signaling by using rapid chemical inhibition of VCP. Importantly, the effect was not due to perturbations of transcription, replication or cell cycle progression, hinting towards a direct role of VCP in TOP2-induced DNA damage. Furthermore, the knockdown of VCP adaptor proteins, especially PLAA, decreased *MLL* fragility. The adaptor PLAA and NPLOC4 are of interest in this context as they have previously been studied in the context of protein-DNA crosslink repair. Its yeast analog Doa1 has been shown to form a ternary complex with VCP (Cdc48) and SPRTN (Wss1) and to bind to SUMO substrates and potentially to TOP1 cleavage complexes (Balakirev et al., 2015). Similarly, SUMO-dependent substrate binding was reported for the yeast Cdc48/Ufd1/Npl4 complex (analog to VCP/UFD1/NPLOC4) that facilitates TOP1cc turnover and its loss in combination with Tdp1 loss causes hypersensitivity to TOP1 poisons (Nie et al., 2012). Potential roles of these cofactors in TOP2cc repair were, however, not evaluated. Here, we find mildly reduced levels of ETO-induced *MLL* breakage upon depletion of cofactors NPLOC4, ZFAND2B and TEX264. ZFAND2B has not previously been linked to protein-DNA crosslink repair, but TEX264 has just been reported to facilitate TOP1cc resolution by interactions with VCP and SPRTN (Fielden et al., 2020). Our findings thus suggest that VCP contributes to TOP2-induced gene fragility and involves the activity of adapter proteins. Here, we studied the contribution of VCP cofactors in a rather specific setup by looking only at *MLL* breakage. Thus, further experiments are required that estimate the contribution of VCP adaptors to breakage of other loci, to see their influence on TOP2cc accumulation and study the interplay of VCP and cofactors with other TOP2cc repair factors, such as TDP2 and MRE11 or SPRTN.

Our data shows that VCP inhibition leads to strong decreases in *MLL* breakage and moderate decreases in DNA damage signaling. Overall, we find similar responses of cells to VCP inhibition as to proteasome inhibition. Anyhow, the inhibition of the 26S proteasome has more global effects as it decreased cell viability and causes cell cycle changes. VCP inhibition on the other hand only mildly decreases cell viability, arguing for a more specific function in protein homeostasis. The reduction of *MLL* breaks upon VCP inhibition at first sight seemed a bit puzzling as we were anticipating that factors involved in TOP2cc repair rather suppress the formation of DSBs, as we have seen for MRE11 or TDP2. However, new insights into the multiple steps of TOP2cc repair were recently uncovered, indicating that around half of the poisoned TOP2ccs are reversible upon washout of etoposide and are only committed to TOP2cc repair upon proteolytic degradation (Canela et al., 2019; Sciascia et al., 2020).



Proteasome inhibition significantly reduces the formation of protein-free DSBs, suggesting that proteolysis of TOP2ccs commits protein-DNA complexes into an irreversible state, towards repair and the formation of free DSBs (Sciascia et al., 2020). Importantly, a substantial fraction of irreversible TOP2ccs was still observed upon proteasome inhibition, arguing that remodeling or degradation of TOP2ccs occurs that is independent of the proteasome (Canela et al., 2019; Sciascia et al., 2020). We hypothesize that VCP could be required to identify SUMO or ubiquitin-modified TOP2ccs to recruit proteases or potentially remodel TOP2ccs itself, thereby complementing 26S proteasome activity and licensing abrogated TOP2-DNA complexes for repair. Thus, inhibition of VCP would enhance TOP2cc reversibility by inhibiting TOP2cc metabolism and thereby DSB formation.

To further investigate the role of VCP in TOP2cc repair and ETO-induced DSBs, many additional experiments are important. First of all, an interaction of TOP2 with VCP upon TOP2 poisoning, or not, should be tested to see if VCP directly binds to TOP2ccs (currently done by the group of Petra Beli). Second, the accumulation of TOP2ccs could be analyzed upon inhibition or knockdown of VCP by biochemical assays, such as the *in vivo* complex of enzyme (ICE) assay. Alternatively, END-Seq and ExoT-END-Seq would reveal site of TOP2cc accumulation as well as of protein-free DSBs upon ETO-treatments with or without washout and VCP inhibition. Thereby the genome-wide impact of VCP on priming TOP2ccs for repair and on the formation of DSBs could be analyzed. Finally, it is important to study if VCP and the 26S proteasome act epistatically or provide different mechanisms of TOP2cc metabolism.

#### 3.4.4 Targeting of TOP2cc repair in cancer therapy

The investigation of TOP2cc repair pathways as suppressors or facilitators of persistent DSBs has important implications for cancer therapies. Chemical inhibition of MRE11, TDP2 or NHEJ have been proposed as a strategy to potentiate ETO-induced cytotoxicity and thereby cancer treatments. While the deficiencies of the repair factors TDP2, MRE11 and LIG4 indeed increased cell susceptibility to etoposide, it also presents an elevated risk of therapy-induced chromosome rearrangements by the inhibition of DSB repair efficiencies.

The block of NHEJ might prolong the time that protein free DNA ends remain unrepaired, thereby increase the chance of DSB pairing and chromosome fusions in *trans*. Thus, although NHEJ inhibitors are available and could potentiate ETO-treatment, this could possibly increase the side effects of oncogenic fusion in healthy cells to promote secondary cancers. Alongside, one could speculate that persistent protein-DNA complexes that cannot be converted into clean DSBs may not lead to gene fusions, since the ligation of DNA ends requires protein-free DSBs. Amplifying the accumulation of non-repaired but irreversible

PDBs by inhibition of MRE11 or TDP2 could therefore enhance cancer therapies with a potentially low risk of chromosome rearrangements. The targeting of TDP2 would be most specific and efforts are currently made to develop TDP2 inhibitors (Hornyak et al., 2016; Raouf et al., 2013). However, we and others do find higher levels of *MLL* translocations and other rearrangements in MRE11 and TDP2 deficient cells, arguing that the loss of one TOP2-DNA processing path rather increases the chance of illegitimate fusion events (Gómez-Herreros et al., 2013, 2017).

Combination therapies with proteasome inhibitors have also been proposed that target the initial TOP2 processing (Lee et al., 2016; Stingele et al., 2017). Nevertheless, one should be cautious in targeting the very first steps of PDB repair as we have shown that inhibiting PDB repair at different hierarchical levels leads to different outcomes. The inhibition of the proteasome or of VCP are in fact rather likely to antagonize the mechanism of actions of TOP2 poisons. Instead of maintaining TOP2ccs in an irreversible state, inhibition of the proteasome or VCP could facilitate the dissociation of TOP2ccs from DNA upon release from chemotherapeutic treatments. If pretreatments with such inhibitors indeed increase cell survival or potentiate cytotoxicity is however not clear yet (Lee et al., 2016; Sciascia et al., 2020). We found increased levels of survival upon ETO-treatments by inhibition of VCP but not by a proteasome inhibitor. Recent findings also suggested that inhibiting the proteasome can also increase cell survivability upon TOP2 poisoning (Sciascia et al., 2020). This shows that TOP2cc metabolism is complex and requires further studies before clinical consequences should be deduced.

Having the individual DDR backgrounds of cancers in mind, more targeted approaches that act on specific TOP2cc repair pathways could be more beneficial for the many different patients. An important outlook is to exploit synthetic lethality that can result from interplays and hierarchies of PDB repair. Here, we find that *MRE11* and *TDP2* are non-epistatic suppressors of gene fragility and that inhibition of both pathways substantially increases ETO-induced genomic instability. Patients that have cancers with mutations in one of the two pathways could benefit from inhibitors of the respective other pathway. For example, cancer cell lines with defects in mismatch repair often also harbor mutations in the *MRE11* gene (Stingele et al., 2017). Another approach would be to exploit the finding that the activity of genes is directly linked to their TOP2-induced fragility. *MLL* expression could be modulated by developing inhibitors for *MLL* specific transcription factors, such as for ZFP64. While this is an intriguing possibility, it requires a deeper understanding of *MLL* regulation. In summary, many promising approaches to enhance and improve TOP2-targeting therapies have been postulated but further research is required for an assessment of safe and efficient treatments.

### 3.5 C-Fusion 3D advances the detection of rare chromosome rearrangements and identifies candidate regulators of *MLL* fragility

Great breakthroughs in molecular biology have often been shaped by the innovation of novel methodologies. The development of genomics analysis has fundamentally changed our understanding of how genetic information is encoded, how the genome exerts its function and how vulnerable it can be to endogenous and exogenous mutagenic agents. A key to the understanding of rare genomic events, such as specific DSB formation and translocations, is the availability of methods with high sensitivity and specificity.

In this study, we combined genome-wide profiling of DSBs by sBLISS with high-throughput imaging by C-Fusion 3D to uncover and quantify genomic instability derived from abrogated TOP2 complexes. We identified thousands of individual DSB hotspots and revealed genome-wide correlations with gene activity and chromosome architecture. While sBLISS is a great tool to detect sites of DSBs, it can currently not provide absolute counts of DSBs per cell and is limited in comparisons between cell lines. Detection of genomic instability by C-Fusion 3D on the one hand provides an absolute quantification of gene breaks, synapsis or rearrangements of individual genomic loci and allows for precise comparisons between samples and cell lines. C-Fusion 3D enables a highly sensitive detection of chromosome rearrangements and thereby facilitates mechanistic studies of fusion events which is currently not facilitated by other methods.

#### 3.5.1 Quantification of rare chromosome translocations by C-Fusion 3D

The detection of rare chromosome rearrangements is important for diagnostic and research questions and remains a challenge until today. C-Fusion 3D is a novel, FISH-based technique that detects chromosome breaks and fusion with very high sensitivity. It is a targeted approach to detect specific rearrangements but is also versatile as C-Fusion 3D can detect numerous different chromosome fusions without exact knowledge of the breakpoints within BCRs. The FISH setup of C-Fusion 3D can furthermore be utilized for other research questions, for example to study other rearrangements, such as deletions or inversions or to quantify changes in nuclear distance of loci of interest upon chromatin changes.

For the detection of chromosome rearrangements, several spatial relationships need to be considered within the same cells. The advances in 3D distance determination for high-throughput imaging that we developed are especially important when very rare events at

ranges of  $10^{-3}$  are studied. Furthermore, based on the versatile FISH-setup of up to four colors, several readouts can be evaluated within the same dataset. For example, we can identify the percentage of breakage of two genes and their translocation rates, but also the distribution of distance between the two genes in control and treated samples or their radial positioning within the nucleus from the same dataset. We furthermore developed a high-throughput FISH protocol for sampling cells in 384-well plates, which facilitates screening experiments. It is currently still limited by the requirement of a specific probe design that impedes the study of numerous gene loci at a time. This could be improved by alternative labelling strategies, such as by Oligopaint methods that utilize arrays of labelled DNA oligos to improve resolution, expand color choice and allow paralleled high-throughput probe production (Beliveau et al., 2012; Gelali et al., 2019). This technical advance would further enhance the screening capacity of C-Fusion 3D, e.g. to study relationships of *MLL* with all known, recurrent translocation partner genes.

A drawback of C-Fusion 3D is that it can merely detect spatial positioning but it cannot visualize the ligation of DNA ends. We sample the colocalization of DNA ends at a steady state in fixed cells but we do not retain information about the mobility of the ends. This can be circumvented by live cell imaging systems to study translocations (Roukos, Voss, et al., 2013), but they are restricted to artificially integrated and highly repetitive DNA loci and do not allow the modelling of physiologically relevant rearrangements, such as therapy-related translocations. Nevertheless, we find ETO-induced translocation rates that are similar to previously reported endonuclease-induced fusions that were tested by PCR-based approaches (Piganeau et al., 2013). Moreover, translocation frequencies formed by DSBs induced at different genomic sites by CRISPR/Cas9 and analyzed by C-Fusion 3D or digital droplet PCR were comparable, further demonstrating that C-Fusion 3D accurately detects rare chromosome fusions (dissertation: Vera Minneker).

### 3.5.2 Additional candidate factors to modulate TOP2-dependent *MLL* gene fragility

Using C-Fusion 3D, we performed a screen that identified several primary candidates as novel regulators of TOP2-induced DSB inductions. We followed up on the contribution of VCP to ETO-induced DNA damage and *MLL* fragility but other interesting primary candidates also emerged from this screen.

Menin (multiple endocrine neoplasia type 1, encoded by *MEN1*) was among the factors that downregulate *MLL* breakage upon RNAi silencing. Menin is a highly unique protein that is involved in transcriptional regulation of various signaling pathways by forming complexes with various proteins, especially chromatin modifiers (Matkar et al., 2013). Interestingly, the

MLL1 protein (of the *MLL* gene) is among its interaction partners, which harbors a SET-domain and forms a histone methyltransferase complex that includes binding of Menin (Yokoyama et al., 2004). The Menin-MLL1 complex regulates gene expression by trimethylation of histone H3 at lysine 4 (H3K4me3) and has been directly associated with transcriptional regulation of leukemogenic factors, such as Meis1 and Hox-A9 (Matkar et al., 2013). Intriguingly, the knockout of Menin in mouse cells has been shown to also downregulate MLL1 protein levels. It was thus hypothesized that Menin and MLL1 could be co-expressed, though no direct evidence has been shown (Milne et al., 2005). This could explain reduced levels of *MLL* breakage upon MEN1 knockdown, because we have shown that the fragility of *MLL* is directly linked to its expression. Importantly, Menin inhibitors have been investigated as anticancer drugs that could potentially also downregulate *MLL* expression if the two genes are linked in their expression (Borkin et al., 2015). This is an intriguing possibility since the efficacy of ETO-treatments might not be altered globally but only locally for the *MLL* gene to prevent oncogenic *MLL* rearrangements. Menin thus emerges as a highly interesting candidate as a facilitator of *MLL* fragility that should further be investigated.

The candidate RUVBL1 is a AAA+ ATPase that has a multitude of cellular functions and binding partners. It is part of many chromatin remodeling complexes, mostly in conjunction with RUVBL2, such as the Ino80, SWR1 or TIP60/NuA4 complexes. Interestingly, RUVBL1 has also been proposed to interact with MLL1 as part of a MLL1 methyltransferase complex but its contribution has not yet been characterized (Dou et al., 2005). Since we only saw a downregulation of *MLL* fragility for RUVBL1 and not RUVBL2, this primary candidate should be handled with care and has to be validated further. On the other hand, some RUVBL1 specific functions have been reported, especially for the stabilization of the TIP60 histone acetyltransferase complex and its function is DNA damage signaling regulation (Jha and Dutta, 2009).

A potential suppressor of TOP2-induced breakage could be the helicase DHX9. DHX9 is a ubiquitously and highly expressed helicase that can unwind DNA, RNA duplexes or RNA/DNA heteroduplexes. It has furthermore been shown to act on non B-structures of DNA such as G-quadruplexes (G4) or H-DNA (Lee and Pelletier, 2016). It has been implicated in the regulation of transcription as well as in the cellular response to DNA damage. Recently, it was reported that DHX9 counteracts the formation of RNA/DNA hybrids, such as R-loops in vivo (Cristini et al., 2018). DHX9 could suppress the formation of R-loop associated DNA damage upon trapping of TOP1 complexes by camptothecin and its activity was dependent on transcription (Cristini et al., 2018). Since an interaction of DHX9 with TOP2A has also

previously been found (Zhou et al., 2003), one could speculate that DHX9 might counteract transcription associated DNA damage upon TOP2 poisoning, mediated by for example increased R-loop or G4-levels. This is however highly speculative and further investigations are required to understand a potential role of DHX9 to suppress TOP2-induced gene fragility.

### 3.6 Summary

We conclude that an interplay of transcriptional activity and spatial genome organization determines the formation of TOP2-induced gene fragility and the occurrence of *MLL* chromosome translocations. Insights into the interaction of molecular mechanisms are crucial for understanding the biogenesis of therapy-related genomic rearrangements, such as *MLL* translocations. Our findings suggest a mechanism in which transcription-associated changes of DNA topology at loop boundaries contributes to TOP2-induced genomic instability. The repair of TOP2-linked DNA breaks includes activities of various repair enzymes and the activities of TDP2, MRE11 and NHEJ counteract the formation of oncogenic fusions. Using C-Fusion 3D, we identified novel candidate factors involved in *MLL* breakage, broadening our understanding of the cellular response to ETO-induced genomic instability. We believe that understanding the underlying mechanisms of genomic instability in cancer therapies will remain an active field of research as it has a high impact on patients' well-being but also reveals fundamental mechanisms of how cells manage the constant threat of genomic instability by exogenous and endogenous sources.

## Chapter 4 Material and Methods

### 4.1 Maintenance of cell lines

TK6 cells were cultured in RPMI (Gibco 11875093) supplemented with 5 % horse serum (Gibco 11510516), 1 mM sodium pyruvate (Gibco 11360070), 100 U/mL penicillin-streptomycin (Gibco 15140122) and 2 mM L-glutamine (Gibco 25030081). K562, HCT116, SKM-1 and MV4-11 cells were maintained in RPMI supplemented with 10 % FBS (Gibco 10270106), 100 U/mL penicillin/streptomycin and 2 mM L-glutamine. HTETOP cells were grown in RPMI supplemented with 10 % FBS, 100 U/mL penicillin/streptomycin and 2 mM L-glutamine, 20 mM HEPES (Gibco 15630080) and 4 x MEM non-essential amino acids (Gibco 11140035). HEK293T, Cal51 and U2OS cells were cultured in DMEM (Gibco 41965062) supplemented with 10 % FBS, 100 U/mL penicillin/streptomycin and 2 mM L-glutamine. All cells were grown at 37°C under a humidified atmosphere with 5 % CO<sub>2</sub>.

Bone marrow CD34<sup>+</sup> hematopoietic stem and progenitor cells were purchased from Lonza (2M-101C) and maintained in StemSpan SFEM II (Stem Cell Technologies 09605) supplemented with SCF 100ng/mL, FLT3 50 ng/mL, TPO 50 ng/mL, IL3 20 ng/mL (Peprotech, 300-07, 300-19, 300-20, 200-03). Cells were thawed and cultured according to Lonza's protocols for handling of Poietics™ cells. Briefly, the tube containing ~1×10<sup>6</sup> frozen CD34<sup>+</sup> cells was thawed in a 37°C water bath for 2 min and cells were transferred into pre-warmed medium. After spinning at 300 × g for 10 min, the supernatant was removed, cells were resuspended by flicking and transferred into 3 mL of StemSpan SFEM II medium with cytokines. A cell density of 0.3 – 1.5×10<sup>6</sup> cells/ mL was maintained.

In HTETOP cells the conditional knockout of TOP2A was induced by 10 µg/mL doxycycline (D9891, Sigma-Aldrich) for 48 h prior to experiments. HCT116-TOP2A-mAID and HCT116-TOP2A-mAID-TOP2B<sup>-/-</sup> were maintained in 1µg/mL puromycin, 125 µg/mL hygromycin and 7.5 µg/mL blasticidin (Invivogen) and TOP2A depletion was triggered by the addition of 500 µM auxin (I5148, Sigma-Aldrich) for 4 h. CtIP depletion was obtained by treatment of TK6 degron cells with 1 mM auxin for 6 h prior to experiments.

A list of all cell lines and their source is indicated in Table 1.

**Table 1 Cell lines used in this study.**

Cell line	Source
Cal51	DSMZ, ACC 302
U2OS	ATCC, HTB-96
HEK293T	ATCC, CRL-3216
K562	ATCC, CCL-243
CD34+ bone marrow cells	Lonza, 2M-101C
SKM-1	Gift from Dr. Thomas Kindler
MV4-11	Gift from Dr. Thomas Kindler
HCT116-Top2A-mAID	Dr. Damien F. Hudson (Gothe et al., 2019)
HCT116-Top2A-mAID Top2B <sup>-/-</sup>	This work
HCT116 wt	Horizon, HD Par-082
HTETOP	Dr. Andrew Porter (Carpenter and Porter, 2004)
HTETOP Top2B <sup>-/-</sup>	This work
DLD-1 mAID-mClover-RPB1	Nagashima et al., 2019
TK6 wt (TSCER2)	Prof. Dr. Shunichi Takeda (Hoa et al., 2016)
TK6 Mre11 <sup>loxP/loxP</sup>	Prof. Dr. Shunichi Takeda (Hoa et al., 2016)
TK6 Mre11 <sup>loxP/H129N</sup>	Prof. Dr. Shunichi Takeda (Hoa et al., 2016)
TK6 Mre11 <sup>loxP/H129N</sup> Tdp2 <sup>-/-</sup>	Prof. Dr. Shunichi Takeda (Hoa et al., 2016)
TK6 Tdp2 <sup>-/-</sup>	Prof. Dr. Shunichi Takeda (Hoa et al., 2016)
TK6 Lig4 <sup>-/-</sup>	Prof. Dr. Shunichi Takeda (Hoa et al., 2016)
TK6 Rad54 <sup>-/-</sup>	Prof. Dr. Shunichi Takeda (Hoa et al., 2016)
TK6 Lig4 <sup>-/-</sup> Rad54 <sup>-/-</sup>	Prof. Dr. Shunichi Takeda (Hoa et al., 2016)
TK6 CtIP-AID	Prof. Dr. Shunichi Takeda (Hoa et al., 2015)
TK6 Cas9	This work
TK6 MLL-CTCF <sup>mut</sup>	This work
TK6 ZFP64 <sup>-/-</sup>	This work

## 4.2 Lentiviral transduction of human cell lines

For the generation of lentiviral particles, HEK293T cells were co-transfected with the second-generation packaging vectors psPax2 and pMD2.G (Addgene #12259 and #12260, gift from Didier Trono) and the lentiviral vector of interest (e.g. the Cas9 and TOP2B guideRNA expressing vectors). Therefore, HEK293T cells at medium confluency in a 10 cm dish were transfected with 2 µg psPax2, 1 µg pMD2.G and 4 µg of lentiviral transfer vector using X-tremeGENE HP DNA transfection reagent (Sigma-Aldrich, 06366236001). After 24 h, the medium was exchanged with 5 mL medium containing 30 % serum. Two and three days after transfection, the supernatant was collected and replaced with fresh medium. The virus containing supernatant was cleared from cells by filtering (0.45 µm) and was concentrated 10-100 × by centrifugation with the Lenti-X™ Concentrator, according to manufacturer's



instructions (Clontech, 631231). Around  $1 \times 10^6$  cells of interest were transduced with the concentrated virus and 5  $\mu\text{g}/\text{mL}$  polybrene (Merck, TR-1003-G) by spinfection at 800 g for 30 min at room temperature (RT). They were subsequently grown at 37°C for 24 h. Positively transduced cells were selected by antibiotics resistances or fluorescence introduced by the lentiviral integration.

### 4.3 Generation of knockout and engineered cell lines

The inducible TOP2A knockout cell lines HTETOP, which is derived from the fibrosarcoma human cell line HT1080, was a gift of Dr. Andrew Porter (Carpenter and Porter, 2004).

The HCT116-TOP2A-mAID auxin-inducible degron cell line was created by Dr. Christian F. Nielsen and Dr. Damien F. Hudson at the Murdoch Childrens Research Institute in Melbourne, Australia (Gothe et al., 2019).

$\text{Mre11}^{\text{loxP}/\text{loxP}}$ ,  $\text{Mre11}^{\text{loxP}/\text{H129N}}$ ,  $\text{TDP2}^{-/-}$ ,  $\text{Mre11}^{\text{loxP}/\text{H129N}} \text{TDP2}^{-/-}$ ,  $\text{LigIV}^{-/-}$ ,  $\text{RAD54}^{-/-}$ ,  $\text{LigIV}^{-/-}$   $\text{RAD54}^{-/-}$  and CtIP auxin-degron lymphoblastoid TK6 cells were created by and kindly made available by Prof. Dr. Hiroyuki Sasanuma and Prof. Dr. Shunichi Takeda from the Department of Radiation Genetic at the Graduate School of Medicine, Kyoto University, Japan (Hoa et al., 2015; Hoa et al., 2016).

#### 4.3.1 Generation of a stable TK6-Cas9 cell line

In TK6 cells, the stable expression of SpCas9 cells was achieved by lentiviral transduction of TK6 wt cells with pLentiCas9-Blast (Addgene #52962 (Sanjana et al., 2014)). Cells expressing Cas9 were enriched by selection with blasticidin at 8  $\mu\text{g}/\text{mL}$  and single clones were obtained by serial dilution. Single clones were tested for Cas9 expression by immunofluorescence and immunoblotting with a mouse monoclonal Cas9 antibody (#61577, Active Motif).

#### 4.3.2 Generation of HCT116-TOP2A-mAID-TOP2B<sup>-/-</sup> cells

For the generation of a TOP2B knockout in HCT116-TOP2A-mAID cells, cells were lentivirally transduced to have expression of SpCas9 and two guide RNAs (gRNAs) targeting the first two exons of TOP2B. SpCas9 was encoded in the lentiviral vector pL-CRISPR.EFS.GFP (Addgene #57818, (Heckl et al., 2014)). The two gRNA for the TOP2B exon 1 (5'-CGGCCGCAGCCACCCGACT-3') and TOP2B exon 2 (5'-CTTCGTCCTGATACATATAT-3') were cloned into a single pMuLE Lenti Dest Neo vector (Addgene #62178 (Albers et al., 2015)). The sequences of the two gRNAs were predicted using the CRISPR Design Tool from the Zhang lab

([crispr.mit.edu](http://crispr.mit.edu), accessed in 2016). In a first cloning step, the two gRNA sequences were introduced into pMuLE ENTR U6 stuffer sgRNA scaffold L1-R5 and pMuLE ENTR U6 stuffer sgRNA scaffold L5-L2 (Addgene #62127 and #62130 (Albers et al., 2015)), respectively. Therefore, oligos for the gRNAs with BfuAI overhangs (5'ACCG for the forward oligo and 5'AAAC for the reverse oligo) were annealed by 30 min at 37°C followed by ramp down from 95°C to 25°C at 5°C/min. Subsequently, they were phosphorylated by T4 PNK (NEB). The pMuLE ENTR vectors were digested by BfuAI (NEB), dephosphorylated by FastAP (ThermoFisher) and linear fragments were purified by gel extraction following manufacturer's instructions. Upon ligation (Quick Ligase (NEB), 30 min at RT), 3 µL of the ligation mix were transformed into chemically competent bacteria by standard heat shock transformation protocol. DNA was amplified in liquid cultures and purified by Miniprep (PureLink Quick Plasmid Miniprep Kit, ThermoFisher). The correct integration of guide RNA sequences was verified by sequencing. Then, both pMuLE ENTR vectors with TOP2B gRNAs were recombined with the pMuLE Lenti Dest Neo by LR recombination following manufacturer's instructions (Gateway™ LR Clonase™ II Enzyme Mix, ThermoFisher). Upon transformation, DNA was amplified and isolated as described above and the final pMuLE destination vector was verified by sequencing. The pMuLE-TOP2B-Neo vector was integrated into HCT116-TOP2A-mAID cells by co-infected with the pL-CRISPR.EFS.GFP vector. Positive clones were selected by resistance to G418 (700 µg/mL for 10 day) and screened for by GFP expression. The knockout of TOP2B in single clones was verified by western blot analysis using a monoclonal and polyclonal antibody (H-286 and A-12 from Santa Cruz).

### 4.3.3 Generation of HTETOP-TOP2B<sup>-/-</sup> cells

In HTETOP inducible TOP2A depletion cells, the CRISPR/Cas9 knockout of TOP2B was also done with two gRNAs for TOP2B exon 1 and exon 2 (as above). HTETOP cells were co-infection with lentiviral particles for the integration of two plasmids, each one expressing SpCas9 (plentiCRISPR v2, Addgene #52961 (Sanjana et al., 2014)) and each one a TOP2B gRNA. Cloning of gRNA sequences into the SpCas9 vectors was achieved by following instructions from the Zhang lab (*Target Guide Sequence Cloning Protocol*, (Sanjana et al., 2014; Shalem et al., 2014)). Individual clones were seeded and selected with 0.5 µg/mL puromycin for 10 days. The knockout of TOP2B was screened for and verified by immunofluorescence and western blot analysis (monoclonal and polyclonal antibodies H-286 and A-12 from Santa Cruz).

#### 4.3.4 Generation of TK6 ZFP64<sup>-/-</sup> cells

For the knockout of a ZFP64 in TK6 cells, TK6-Cas9 cells (see above) were lentivirally transduced with two guide RNAs (gRNAs) targeting exon 2 and exon 5 of ZFP64. The two gRNA ZFP64 e2 (5'-CTTGCAGATGCCGCAGATA-3') and gRNA ZFP64 e5.1 (5'-TGGAGTTGCGGCTGGCGTAG-3') sequences were derived from a previous publication of Bin Lu et al. (Lu et al., 2018). gRNAs were first cloned into pMuLE ENTR U6 stuffer sgRNA scaffold L1-R5 and pMuLE ENTR U6 stuffer sgRNA scaffold L5-L2, to then be recombined into pMuLE Lenti Dest Neo by LR recombination, as described above for the TOP2B gRNAs. The pMuLE-ZFP64-Neo vector was integrated into TK6-Cas9 cells via lentiviral transduction. Positive clones were enriched by resistance to G418 (700 µg/mL for 10 day) and single clones were thereafter generated by single cell dilutions into 96-well plates. Single clones were tested for the knockout of ZFP64 by western blot analysis (Proteintech 17187-1-AP). Five clones that showed disruption of ZFP64 by western blot analysis were also probed for MLL protein levels by western blot (Bethyl #A300-374A-M).

#### 4.3.5 Generation of the TK6 MLL-CTCF<sup>mut</sup> cell line

Mutation of the *MLL* exon 12 CTCF binding site in TK6 cells was mediated by homology directed repair. Therefore, 1×10<sup>6</sup> TK6 cells stably expressing Cas9 (see above) were electroporated with a gRNA targeting the *MLL* exon 12 and creating a Cas9 cut at the end of the CTCF motif (also see Figure 2.24). With the Neon Transfection System (ThermoFisher, 100 µL buffer R, 1350 V, 10 msec, 3 pulses), 8 µM of the gRNA (5'-GGAGAACGAGCGCCCTCTGG-3') and 1 µM single stranded DNA oligo with short homology arms were introduced into cells:

/5Phos/GCCAAGTCTGTTGTGAGCCCTTCCACAAGTTTTGTTTAGAGGAGAACGAGCGACCTTTG  
GATGACCAGCTGAAAATTGGTGTGTCGTCGTTGCAAATCTGTACGTTTGTGG (PAGE  
Ultramer Oligo, Integrated DNA Technologies). Four days after electroporation, TK6 cells were dispersed into 96-well plates at single cell dilutions. Around 125 individual clones were transferred into 24-well plate upon proliferation and genomic DNA (gDNA) was extracted using the DNeasy 96 Blood & Tissue Kit (#69581, Qiagen). The genomic region at *MLL* exon 12 was amplified by genomic PCR with the primers MLL-CTCF forward (AGGACAAACCAGACCTTACAAC) and MLL-CTCF reverse (GTAGCCTGATGTTGCCTTCCA) (PCR conditions are found in Table 2).

**Table 2 Genomic PCR components and cycling conditions for *MLL* exon 12.**

Component	25 $\mu$ l Reaction	Step	Temperature	Time
Nuclease-free water	15,5 $\mu$ L	Denaturation	98°C	30 sec
5X HF Phusion buffer	5 $\mu$ L	35 cycles	98°C	20 sec
10 mM dNTPs	0,5 $\mu$ L	Annealing	64°C	30 sec
10 $\mu$ M forward primer	1,25 $\mu$ L	Extension	72°C	40 sec
10 $\mu$ M reverse primer	1,25 $\mu$ L	Final Extension	72°C	10 min
HF DNA Pol (IMB CF)	0,5 $\mu$ L	Hold	4-10°C	$\infty$
Template DNA (50-100ng)	1 $\mu$ L			

The editing of the CTCF consensus sequence in individual clones was tested by HaeII (NEB) digestion of the PCR product and agarose gel electrophoresis. Clones with the expected bands were further tested by sequencing of gDNA. The editing of both *MLL* alleles in one clone, named TK6 *MLL*-CTCF<sup>mut</sup>, was confirmed by the pGEM-T Easy Vector Systems (Promega) and sequencing of eight transformants.

#### 4.4 Chemical inhibition of transcription and replication

To inhibit transcription, TK6 and K562 cells were treated with the indicated doses of 200nM Actinomycin D (Cayman Chemicals, 11421-5), 200 $\mu$ M DRB (Cayman Chemicals, 10010302-10) and 1 $\mu$ M Triptolide (Sigma-Aldrich, T3652) for 3h. The efficiency of inhibition was assessed by integration of 5-Ethynyl Uridine (EU) incorporation, a surrogate of nascent RNA production, using the Click-iT™ RNA Alexa Fluor™ 488 Imaging Kit (ThermoFisher, C10329). EU was added to the cells at the final concentration of 1 mM for 1 hour before fixation.

To inhibit replication, cells were treated with 8 $\mu$ M Aphidicolin (Sigma-Aldrich, A4487) for 3h. Efficiency of inhibition was assessed by monitoring 5-Ethynyl-2'-deoxyuridine (EdU) incorporation using the Click-iT™ EdU Alexa Fluor™ 488 Imaging Kit (ThermoFisher, C10337). EdU was added to the culture media at the final concentration of 10 mM for 2 hours prior to fixation.

#### 4.5 Cell cycle synchronization and release

To study DLD-1 RNA-PolII degron cells in G1 only (Figure 2.28), cells were seeded onto PLL coated coverslips and the next day either arrested in G1 using 5  $\mu$ M Palbociclib (Selleckchem, S1116) for 24 h or not. Asynchronous or arrested cells were pretreated with 500  $\mu$ M auxin or

DMSO for 2 h and treated with etoposide (Sigma-Aldrich, E1383) at indicated doses for 4 h. Cells were immediately fixed and analyzed by C-Fusion 3D.

TK6 cells were arrested in G1 or G2/M phase using 200  $\mu$ M Mimosine (Sigma-Aldrich, M0253) or 10  $\mu$ M RO-3306 (Selleckchem, S7747) for 18 h. For the release of cell into the next cell cycle phase, the inhibitors were washed out by one PBS wash and cells were released into fresh medium containing the other inhibitor, respectively. For the study of *MLL* fragility in different cell cycle phases (Figure 2.34), upon release of TK6 cells into S or M/G1 phase, they were treated with either DMSO or etoposide (20  $\mu$ M or 30  $\mu$ M, Sigma-Aldrich, E1383) for 4h. Then, cells were washed with PBS and maintained in fresh medium in the presence of the same cell cycle inhibitor as during the etoposide treatment for 24 h. Cells were subsequently spotted onto PLL coated coverslips, fixed in 4 % PFA for 20 minutes and analyzed by C-Fusion 3D.

The distribution of cells across cell cycle phases was analyzed by measuring the DNA content based on Hoechst intensity by fluorescence microscopy as previously shown (Roukos *et al.*, 2015). This was also utilized to analyze immunofluorescence signals in cells of different cell cycle phases without chemically arresting cells. Therefore, cells were gated into G1, S or G2 phase based on Hoechst intensities where gates were derived for each experiment from asynchronous cell histograms.

## 4.6 siRNA knockdown

### 4.6.1 RNAi of TOP2A and TOP2B in Cal51 cells

RNAi mediated knockdowns of TOP2A and TOP2B were performed by reverse transfection of Cal51 cells with siRNAs and Lipofectamine RNAiMAX (Invitrogen, 13778075). Therefore  $1 \times 10^5$  cells seeded and transfected with 30 nM final siRNA and 2  $\mu$ L RNAiMAX according to manufacturer's instructions. Medium was exchanged after 16 h and experiments were conducted upon 72 h of knockdown. The siRNAs were purchased from ThermoFischer: siRNA no target (Silencer Selected negative control, #4390843); siTOP2A (#4390824, siRNA ID #s14309); siTOP2B (#4390824, siRNA ID #s108). The knockdown efficiency was evaluated by western blotting.

### 4.6.2 RNAi screen for 150 factors in Cal51 cells (refers to Chapter 2.5.1)

RNAi mediated knockdowns of 150 factors and three non-target controls was done by reverse transfection of Cal51 cells. A cherry-picked library with single ON-TARGETplus siRNAs from

Dharmacon (see Table 3) of 0.1 nmol was resuspended in 50  $\mu$ L nuclease-free water to a concentration of 2  $\mu$ M and distributed onto four daughter plates to prevent multiple freeze-thaw cycles. Transfections of Cal51 cells was done in 384-imaging plates (CellCarrier-384 Ultra Microplates, PerkinElmer #6057300) that were PLL coated for 30 min and washed once with PBS before seeding of 2'500 cells in 70  $\mu$ L medium. siRNA dilutions in OptiMEM and complex formation with RNAiMAX were prepared in 384-well microplates and were pooled for replicates. Of this, 10  $\mu$ L were immediately added to each well with 70  $\mu$ L of cells upon seeding to reach a final concentration of 2  $\mu$ M siRNA and 0.13  $\mu$ L RNAiMAX per well in 80  $\mu$ L. After 72h incubation, cells were treated with 60  $\mu$ M etoposide (Sigma-Aldrich, E1383) by exchange of medium and fixed 6 h later with 4 % PFA for 20 min, followed by C-Fusion 3D analysis.

**Table 3 siRNA library for 150 factors to study ETO-induced *MLL* fragility.**

Gene	Catalog #	Sequence	Gene	Catalog #	Sequence
NT Control	D-001810-01	UGGUUUACAUGUCGACUAA	MUS81	J-016143-09	CAGCCUGGUGGAUCGAUA
NT Control	D-001810-02	UGGUUUACAUGUUGUGUGA	MYO6	J-006355-05	CCACACUGCUCAUAAUUAU
NT Control	D-001810-03	UGGUUUACAUGUUUCUGA	NAP1L1	J-017274-09	UAACCAUAGUUAUCGAAA
TOP2A	J-004239-06	CGAAAGGAAUGGUUAACUA	NAP1L4	J-012183-09	CCUUCAGCUACAUCGAAA
TDP2	J-017578-05	GUACAGCCCAGAUUGUGAUA	NCOA3	J-003759-10	GCAAUGAUGUCGUCAAGAA
PRKDC	J-005030-06	GGAAGAAGCUCAUUUGAUU	NPLOC4	J-020796-09	AGGAAAAGCAUUGGCGAUU
ACTL6A	J-008243-05	GAACGGAGGUUUAGCUCAU	NSMCE2	J-018070-05	GCUGUCAAUCUACAAUAA
ACTR5	J-018395-09	CCACUGUAUUCACGGCAA	PAF1	J-020349-09	CCACUGAGUUAACCGUUA
ACTR8	J-016441-09	GGUGAUACGGAGAACGGAA	PCGF2	J-006584-06	GUGACGUGCAGGUCCAUA
ALYREF	J-012078-05	UCUCAGACGCCGAUUAUCA	PCGF5	J-007089-05	GAAGAAUUAUUCGAUGU
ARID1A	J-017263-05	GAAUAGGGCCUGAGGGAAA	PCGF6	J-007084-05	GAGGAGGAGCGCCUGAUUA
ARID4A	J-003949-05	UAUAGAAGCUCAUAGUCUU	PCNA	J-003289-15	GCACGUUAUUGCCGAGAUC
ASF1A	J-020222-10	CGAUCAAGUUUUAGACUCU	PHC1	J-011850-05	UAGAGGAGGUGUACGAGUU
ASH2L	J-019831-05	AAAGAUGGCUAUCGGUUAU	PHC2	J-021410-05	GGGUGGACUUUGCCUUAUA
ATAD2	J-017603-05	CCACAGAUUAUACCACUA	PIAS4	J-006445-05	GAAUUAUGUCCACAGAACA
BLM	J-007287-06	CUAAAUCUGUGGAGGGUUA	POLR2A	J-011186-06	UAAUAGAGGUCAUCGAGAA
BMI1	J-005230-09	GAAAUUCAACCAACGGAAA	PRMT2	J-004033-05	GUGGGUACAUUCCGGCAA
BPTF	J-004025-05	GAACCGAAGACUCAUAAUA	PRMT5	J-015817-06	UGGCACAACUCCGGACUU
BRCA1	J-003461-09	CAACAUGCCACAGAUCAA	RANBP1	J-006627-06	GGGCAAAACUGUCCGAUU
BRD3	J-004936-05	AAUUGAACCGCCGGAUUA	RECQL5	J-019338-05	GAACCGUGGUGCAGAACGA
BRD4	J-004937-06	AAACCGAGAUCAUGAUAGU	RING1	J-006554-07	CGAGGUUAUGUGAAGACAAC
BRWD3	J-016759-09	GGUGAUAGAUUCCGCAGUA	RNF168	J-007152-05	GACACUUUCACACAGUAU
CBX3	J-010033-05	AGUACUAGAUCGACGUGUA	RNF2	J-006556-05	CGAGAUACAUAAAGACUUC
CBX4	J-008356-05	GCGAACUGCCUACCGUUA	RNF20	J-007027-05	GCUAAACAGUGGAGAUAAU
CBX5	J-004296-06	GGAUUGCCCUGAGCUAAUU	RNF31	J-021419-05	GCCGAGAUUGUCUGCGAUU
CDK7	J-003241-09	CAUACAAGGCUUAUUCUUA	RNF4	J-006557-07	GCUAAUACUUGCCCAACUU
CDK9	J-003243-09	GGCCAAACGUGGACAACUA	RNH1	J-016048-05	UCAAGGAGCUCACGGUUA
CDYL2	J-008755-09	CAUAAACGGAAGCGAAUUA	RPRD1B	J-013787-09	GGUACACACUAAACGUUUA
CHD1	J-008529-05	CACAAGAGCUGGAGGUCUA	RSF1	J-020374-09	GAAGAAAGAGCGUGCCGAA
CHD1L	J-014368-05	GGGAAGACCGCCAGACUA	RUVBL1	J-008977-05	AUAAGGUGGUGAACAAGUA
CHD3	J-023015-05	GAAUUAUCCUGAAUACGAA	RUVBL2	J-012299-05	UAACAAGGAUUGAGCGAAU
CHD4	J-009774-05	AAGAAGAUUCAGCCGAAA	SATB1	J-011771-09	GCUGAAAGAGACCGAAUUA
CHD6	J-017349-05	GUAGAUUAUGUCAGUCUUA	SETD2	J-012448-05	UAAAGGAGGUUAUUCGAAU
CHD9	J-031153-05	GAAAGUAUUUGAUGGAGUU	SETD7	J-014643-05	CAAACUGCAUCUACGAUUA

<b>CUL4A</b>	J-012610-05	GCACAGAUCCUUCGCUUUA	<b>SETMAR</b>	J-020013-23	CAAGUGUUCAAGACGCAUA
<b>CUL4B</b>	J-017965-05	UAAUAUACCUCCUUGAUGA	<b>SETX</b>	J-021420-05	GCACGUCAGUCAUGCGUAA
<b>DCLRE1B</b>	J-015780-05	GAUCCAUACUUAUACCUA	<b>SIRT2</b>	J-004826-06	CCGCUAAGCUGGAUGAAAG
<b>DDB1</b>	J-012890-06	CACUAGAUCGCGAUAAUAA	<b>SIRT6</b>	J-013306-05	CCAAGUGUAAGACGCAGUA
<b>DDX39B</b>	J-003805-05	GUAGAAGACUCGCCAUUU	<b>SMARCA1</b>	J-011392-05	UAACAUAGCUCGAGAGGUA
<b>DHX9</b>	J-009950-05	GGAUUAAACUGCAAUAUC	<b>SMARCA2</b>	J-017253-05	CAAAGCAGAUGAACGCUAU
<b>DMAP1</b>	J-006378-09	GCCCAAUUCGAGAAAGCGA	<b>SMARCA4</b>	J-010431-05	GCACACCGCUCGAGAACAA
<b>EMD</b>	J-011025-06	ACGCAGAUUUUCGGAUAC	<b>SMARCD2</b>	J-011393-05	GGGAGUACAUCAACUGCAA
<b>EP300</b>	J-003486-11	GGACUACCUAUCAAGUAA	<b>SMARCD3</b>	J-009459-06	GAAGAAGACGGCGUGCUAU
<b>ERCC2</b>	J-011027-06	CAUACUCCUUGCUCGUA	<b>SMC4</b>	J-006837-05	GUUAAACGCUUACACAAUA
<b>ERCC3</b>	J-011028-05	GAUCAAGGUUAUAGCUUCA	<b>SMURF2</b>	J-007194-05	GUUAAUGACUGGAAGGUAA
<b>ERCC5</b>	J-006626-05	GACUGAAGCCUUCGUA	<b>SMYD2</b>	J-020291-05	UCCGAGACAUGGUCAGUA
<b>ESCO1</b>	J-023413-09	GGAAAGAGCAAACGAGGUA	<b>SPRTN</b>	J-015442-19	CAUCAAAAGUCAAAAGCGAA
<b>FLI1</b>	J-003892-05	GUUCACUGCUGGCCUAUAA	<b>SSRP1</b>	J-011783-05	AAGAAGAACUAGCCAGUAC
<b>HAT1</b>	J-011490-05	CAAUGAAGUAUUGGCGUA	<b>SUN1</b>	J-025277-05	GGUAAUGCUGGGCAUUUA
<b>HDAC1</b>	J-003493-10	ACUAUGGUCUCUACCGAAA	<b>SUN2</b>	J-009959-09	CCAGAGACUCAUCGCCACA
<b>HDAC2</b>	J-003495-07	GCGGAUAGCUUGUGAUGAA	<b>SUPT6H</b>	J-010540-06	AUAAAGGCCUGUAGUCGAA
<b>HDAC3</b>	J-003496-09	AAAGCGAUGUGGAGAUUUA	<b>SUV39H2</b>	J-008512-05	GAAAGGUUCUGGAGAUUA
<b>HDAC9</b>	J-005241-12	UCUUGUAGCUGGUGGAGUU	<b>SYNE1</b>	J-014039-05	GAUAAUUGCCCAAGGAGAA
<b>HELLS</b>	J-017444-09	UAAUGAUGCUACUUCGUAA	<b>SYNE2</b>	J-019259-09	AGGAAUUUCUGCAAACCGA
<b>HLTF</b>	J-006448-06	CCAGAUGACUUCUAACUA	<b>TADA2A</b>	J-017516-05	GUGUUUCACUCGAGGCUUU
<b>HMGB1</b>	J-018981-05	CAAACUCAUUAUAGUCA	<b>TBP</b>	J-011790-05	GGUUUAAUCUACAGAAUGA
<b>HMGB2</b>	J-011689-05	GCAAAGAAUUGGGUGAAA	<b>THOC2</b>	J-025006-09	CAAACGAGAAUUGCGGAA
<b>HNRNPU</b>	J-013501-05	GUAGAACUCUCGUUAGCUA	<b>THOC3</b>	J-014879-09	GGUCUGUUCUCUCGGAGUU
<b>HP1BP3</b>	J-020867-09	GUGAAUUCUACCCGGGAAA	<b>TP53BP1</b>	J-003548-06	GAAGGACGGAGUACUAAUA
<b>HUWE1</b>	J-007185-07	GCUUUGGGCUGGCCUAAUA	<b>TRIM28</b>	J-005046-07	GAAAUGUGAGCGUGUACUG
<b>INO80</b>	J-004176-09	GGAAUUGAGUUUCGAUAGA	<b>TRIM33</b>	J-005392-08	GGACAAACCACAUUAGUAA
<b>KAT2A</b>	J-009722-17	AGGACAAUUGGUGCCCGA	<b>UBE2A</b>	J-009424-05	CUAUGCAGAUGGUAGUAUA
<b>KAT2B</b>	J-005055-05	GGUACUACGUGUCUAGAA	<b>UBE2B</b>	J-009930-05	GGAAUGCAGUUUAUUUUGG
<b>KAT5</b>	J-006301-08	CGUAAGAACAAGAGUUAAU	<b>UBE4B</b>	J-007202-06	GCAACUAGACACCGCGAAA
<b>KDM2B</b>	J-014930-05	GGGAGUCGAUGCUUAAUUA	<b>UBR4</b>	J-014021-09	GGGAACACCCUGACGUAAA
<b>KDM3B</b>	J-020378-09	GAUCAUAUGCAAACGACUA	<b>UBXN1</b>	J-008652-09	GGAGAAGGCAAACCCGCUU
<b>KDM5B</b>	J-009899-05	GGAGAUGCACUUCGAUUA	<b>UBXN7</b>	J-023533-17	GGUCAUUAUCAUGUAGUA
<b>KDM6A</b>	J-014140-10	GAGAGUAAUUCACGAAAGA	<b>UFD1L</b>	J-017918-05	GAGCGUCAACCUUCAAGUG
<b>KIF4A</b>	J-004961-09	AGGCGUACAUUCUCCUUA	<b>UHRF1</b>	J-006977-05	GCGGAACAGUCUUGUGAUC
<b>KMT5C</b>	J-018622-19	CGACAGAGUGACAGCACGA	<b>USP10</b>	J-006062-09	UGAGUUUGGUGUCGAUGAA
<b>LEO1</b>	J-016579-09	UGAUUUAGGAAACGACUUA	<b>USP15</b>	J-006066-05	GAUGAUACCAGGCAUUAUA
<b>LMNA</b>	J-004978-05	GAAGGAGGGUGACCUGAUA	<b>VCP</b>	J-008727-09	GCAUGUGGGUGUCGACUUA
<b>LMNB1</b>	J-005270-06	AAUAGAAGCUGUGCAAUUA	<b>WDR11</b>	J-012887-05	GCAGUCGUUUCAGAGUA
<b>MBD5</b>	J-027190-09	GGGAUAGACCUAGACAGA	<b>YEATS4</b>	J-005280-05	CAAGAGAAUGGCCGAAUUU
<b>MED12</b>	J-009092-05	UCGCGGUCUUCGAUAAA	<b>ZMYND11</b>	J-006936-05	GCAAAGAAAGGACGACGUA
<b>MEN1</b>	J-011082-05	CGCAAAGGCCUCUGAACUA	<b>ZUFSP</b>	J-015894-17	GAUUGGAGCAUGUGAAGUA
<b>MTA1</b>	J-004127-06	UCACGGACAUUCAGCAAGA			

#### 4.6.3 RNAi screen for 33 VCP adaptors in Cal51 cells (refers to Chapter 2.5.2).

RNAi mediated knockdowns of TOP2A, TDP2, VCP and 33 VCP adaptor was performed in Cal51 cells by reverse transfection as described above (Chapter 4.6.2). Instead of single siRNAs, a library with pooled SMARTpool ON-TARGETplus siRNAs from Dharmacon was used (see Table 4). Therefore, only one pooled non-target (NT) control was included. Again, 0.1 nmol of siRNA were resuspended in 50  $\mu$ L nuclease-free water to a concentration of 2  $\mu$ M and distributed onto daughter plates to prevent multiple freeze-thaw cycles. Transfections of 2'500 Cal51 cells was done in PLL coated 384-imaging plates as described before. Again, a

final concentration of 2  $\mu$ M siRNA and 0.13  $\mu$ L RNAiMAX per well was applied for reverse transfection. After 72h incubation, cells were treated with 60  $\mu$ M etoposide (Sigma-Aldrich, E1383) by exchange of medium and fixed 6 h later with 4 % PFA for 20 min, followed by C-Fusion 3D analysis.

The potential VCP cofactor TEX264 was included for validation experiments of VCP adaptors that might modulate *MLL* fragility upon etoposide. Here, a single pool of siRNAs for TEX264 was purchased from Dharmacon (SMARTpool ON-Targetplus #L-010662-01-0005).

**Table 4 siRNA library of four pooled siRNAs for 33 VCP cofactors.**

Gene	Catalog #	Sequence	Gene	Catalog #	Sequence
<b>NT Control</b>	D-001810-10	UGGUUUACAUGUUUCCUA	<b>VIMP</b>	L-019353-01	CGGAUGAGGCUAAGAAUCU
		UGGUUUACAUGUCGACUAA			CUGAAACGGAAUCGGACA
		UGGUUUACAUGUUGUGUGA			AAGAAGAACUAAAUGC GCA
		UGGUUUACAUGUUUCUGA			ACCUGAUGUUGUUGUAAA
<b>TOP2A</b>	L-004239-00	CGAAAGGAAUGGUUAACUA	<b>SPRTN</b>	L-015442-02	CAUCAAGUCAAAAAGCGAA
		GAUGAACUCUGCAGGCUAA			GGUUCUGCCUGUCGUGUAU
		GGAGAAGAUUAUCAUGUA			GGCUAUGUCAACGAGCUA
		GGUAAUCUCCUGAAAGUAA			AAGUGGAGCGUGCGAAUGA
<b>TDP2</b>	L-017578-00	GCAGAAGAGGGACACAUUA	<b>SVIP</b>	L-035054-01	CCUUAAGUGCAAUGCUAAA
		AAAGGGCUCUGAACUCCUA			GUAAGGAUCUUUGUAGUAA
		GUACAGCCCAGAUGUGAUA			ACUUUUUAGAGAACGAUCA
		UCUAAGGGAUCGAGAGGUU			CAAGUGGGCUUCAUCAAUU
<b>VCP</b>	L-008727-00	GCAUGUGGGUGCUGACUUA	<b>SYVN1</b>	L-007090-00	UCAUCAAGGUUCUGCUGUA
		CAAAUUGGCUGGUGAGUCU			GAGAAGAGAUGGUGACUGG
		CCUGAUUGCUCGAGCUGUA			CAACAUGAACCCUCUGUAU
		GUAAUCUCUUCGAGGUUAU			GGAAAGGCCUCCAGCUCCU
<b>AMFR</b>	L-006522-00	GCAAGGAUCGAUUUGAAUA	<b>UBE4B</b>	L-007202-00	GCAACUAGACACCCGCGAAA
		GGACGUAUGUCUAUUACAC			CAAGAACGCACGCGCAGAA
		GAAUUCGUCGGCACAAGAA			GGACUUGAUUGGCCAGAUU
		GUAAAUACCGCUUGCUGUG			GCUAAACACUGGCCUCAAU
<b>ANKZF1</b>	L-021152-00	GGGCUACAUIUGAGAAGUU	<b>UBXN1</b>	L-008652-01	GGAGAAGGCAAACCCGCUU
		GUACAAUACUGUUGCGUGC			CAAGAAGUAUGGUGGCAGU
		UCCAUAAGCUGACCACUUU			AGGAGUUAGCAGCCAGACA
		GAACAGAGGGAACAUUAUA			GGACGAGCCUUUAGAGACU
<b>ASPSR1</b>	L-006489-02	UCAGGGAGGCGCAGAUAAA	<b>UBXN10</b>	L-008657-02	GGAAAUUCAUCGUCCGAAC
		GCAGGACUCUUUCUGUUA			UUGUGUUUCUUGCGAGUAU
		GUGAAUAUGAUCGAAGUU			UGGAGGAACCAUCGGACCA
		GCUACGUCCUACAGGGCUU			GGUGUGCGCUCAUCAUUA
<b>ATXN3</b>	L-012013-00	GCAGGGCUAUUCAGCUAAG	<b>UBXN11</b>	L-016403-02	CGCUGAAGCUCUACCGGAA
		ACAGGAAGGUUAUUCUAUA			GGCAAGGCGAGGUGAUUGA
		GGACAGAGUUCACAUCCA			CCUCAGAGCUCACGCGACU
		GCACUAAGUCGCCAAGAAA			GUGAGUGACUUGCGCAAUC
<b>DERL1</b>	L-010733-02	GAGAGACCCUCAUACGCUA	<b>UBXN2A</b>	L-019140-02	ACGUUAAAGUUGAAGACAA
		CAACAAUCAUUAUCACGUA			GAAUAUGUUUGUCUACGA
		GGGCCAGGGCUUUCGACU			GCUCAGUAUAGUAUCAAGA
		CAAUUAUGUUGCAGCUACA			GCUAAUACGACCGGUACA
<b>DERL2</b>	L-010576-01	CUAUUCGUAUCUGUCGAA	<b>UBXN2B</b>	L-025945-01	GUGCCGUAAUAUAGAGGAA
		CUGCAGAUCCCACCGUCA			CUUCAAGAGCCCACGGACA
		GUCUAUGUGUGGAGCCGAA			GCAAAAGGUUAUACAGCCA
		AGGACAAAAUAAAACGGAA			UGACUUCAUUCGGAUAA



<b>FAF1</b>	L-009106-00	CAGAAAUGUUGAUGUGGUA	<b>UBXN4</b>	L-014184-01	UAGAGUGGACAUACGGAAA
		AAACAUUGACGAAGCUAUU			CAUAGAUAGUACUAGGUU
		GCAGAACAGUUUCGUUUGG			GAAAGUAGCUGGCGAGGUU
		GGAGAGACGUAACUCAACU			CAAGAAAUUUGUAUGCGUA
<b>FAF2</b>	L-010649-02	GGACCUAACUGACGAAUGA	<b>UBXN6</b>	L-008785-01	UCAUGAAGAUACACGUU
		AGGAUAACCUUACGUGAAA			AGUUCUACGUGCUGAGCGA
		UGUACAGGACAGAUUGAAU			GUGUUUCAGGAGCGAUUA
		CCUAAUGAUUCUCGAGUAG			AAAGGAACUUAAGCCGAA
<b>NGLY1</b>	L-016457-01	GCGAGUGGGCCAAUUGUUU	<b>UBXN7</b>	L-023533-02	GGAUCAAAUGUUAACGUUA
		CUGCCGAGCUGUAGGGUUU			CUGUGUUGUUUCAGGCAUA
		GGAUAAUUGUGGAGCUUGU			GACUGAAACUUAUUCGGCAA
		GGAGAAUGGCGUGUGGAAA			GGUCAAUAAUCAUGUAGUA
<b>NPLOC4</b>	L-020796-01	AGGAAAAGCAUUGGCGAUU	<b>UBXN8</b>	L-009648-01	CUUCAAGUCUUAUACGCA
		AUAAUGGCUUCUCGGUUUA			GAAAACUGGAGGAGCGCUU
		ACUCUGAAGUUCGCGCAUU			UUGACUGGAUGACGAGAAU
		CUGAAGUGGCGUGCAUUUA			AACUGAUGUUUGCGAUUUA
<b>NSFL1C</b>	L-017222-01	CCACUUGUUCACGAGAA	<b>UFD1L</b>	L-017918-00	GAGCGUCAACCUUCAAGUG
		UGACUACUUUCCGAACAA			GGUAAGAUAAUUAUCAUCA
		CCAGCAUGUUGUACGGAAA			UGACAUGAACGUGGACUUU
		GGAGAGACCAGUAAACCGA			CCCAAAGCCGUUUUAGAAA
<b>NUB1</b>	L-019158-00	GGAAAACAUUGACCGAUUG	<b>VCPIP1</b>	L-019137-00	GAGAAGCUCUGGUGAUUUAU
		GAAAUGACACCGUACUUG			GGGACAGACUUUAGUAAUA
		ACAGGCAGAUCAAUCAGAA			GGAGAUGGGUCUAUUGUGU
		CCUCAGAUUGUGGUGUUUA			CGACAGAAUACAUAAGAA
<b>PLAA</b>	L-016215-00	GCGAGUGUCUUGAAGUAUA	<b>YOD1</b>	L-027369-00	GAUCCAGACUUCUAUAGUG
		GUAAGUGAAUGCUGUAGAU			GACAGGCCAUACCAACUUU
		GGAAGUAGUACAAGACCUA			GCAUAGAGAUUACGAUUU
		UGUAAGAGGUUUGGCAUUU			CAUCCAUCUGGUGACAUG
<b>RHBDD1</b>	L-019378-00	CGGCAUACUACUUUAUA	<b>ZFAND2B</b>	L-016586-03	AGGCAGAAUACCAGCGCA
		CGAGGAAUACCAGAAUA			UGUGAACGUGUAGCCGAA
		GGGAUAAUACUGGACUUA			GCACAAGCACUGUCAGCCA
		GACAGCGGCUUCACAGAUU			CAGUGAUUGCUUUGCAGAA
<b>RNF31</b>	L-021419-00	CCUAGAACCUGAUCUUGCA			
		GCCGAGAUUGCUGCGAUU			
		GGCGUGGUGUCAAGUUUA			
		GCAGAAUACUCAUCCAAGA			

#### 4.7 Western blotting

Total cell lysates were either prepared by directly lysing cell pellets in SDS-page loading buffer (10'000 cells/  $\mu$ L buffer) or, for comparative analysis, prepared by lysis in RIPA buffer. Therefore, cells were washed once with cold PBS and lysed in cold RIPA buffer for 20 minutes on ice (for example 200  $\mu$ L for a 10 cm plate). Adherent cells were scraped after 5 min and transferred into microcentrifuge tubes. Samples were spun down at 21'000  $\times$  g for 15 min at 4°C. Supernatants were transferred into new tubes and protein concentrations were determined by colorimetric assay using Bradford reagent (Bio-Rad #5000205). Samples were adjusted to 1  $\mu$ g/ $\mu$ L and mixed with 4X NuPAGE™ LDS sample buffer (ThermoFisher #NP0007) and heated at 70°C for 10 min. Of this, 5-15  $\mu$ L were separated on pre-cast 4 % – 15 % gels (Bio-Rad 4561086, 4561085DC) and transferred to PVDF membranes

(Merck, IPFL00010) by wet transfer at 120 – 150 mA for 2-3 h (Mini Trans-Blot® Cell, Bio-Rad). Membranes were blocked in 5 % milk (Sigma-Aldrich, 70166) in PBS for 30 to 60 min and incubated with primary antibodies diluted in blocking solution over night at 4°C (Table 5). The following day, membranes were washed 3 times with PBST (PBS with 0,1 % Tween® 20) for 10 min and then incubated with secondary peroxidase-coupled antibodies in blocking solution at room temperature for 1 hour (Table 5). After another three washes with PBST, ECL-based chemiluminescence was detected with WesternBright Chemilumineszens Substrat Sirius (Biozym Biotech, 541021) following manufacturer’s instructions.

**Table 5 Antibodies and dilutions for western blotting.**

Protein target	Company	Identifier	Dilution
Anti-Top2A (G-6), mouse monoclonal	Santa Cruz	Cat# 166934	1:300
Anti-Top2A (C-15), goat polyclonal	Santa Cruz	Cat# 5346	1:200
Anti-Top2B (H-286), rabbit polyclonal	Santa Cruz	Cat# 13059	1:200
Anti-Top2B (A-12), mouse monoclonal	Santa Cruz	Cat# 365071	1:200
Anti-RNA-PolIII (RPB1), mouse monoclonal	BioLegend	Cat #904001	1:1000
Anti-ZFP64, rabbit polyclonal	Proteintech	Cat #17187-1-AP	1:500
Anti-MLL1, rabbit polyclonal	Bethyl	Cat# A300-374A	1:500
Anti-Mre11, rabbit polyclonal	Cell Signaling Technologies	Cat# 4895	1:1000
Anti-GFP, mouse monoclonal	Santa Cruz	Cat# 9996	1:1000
Anti-Rad54 (4E3/1), mouse monoclonal	Santa Cruz	Cat# 53433	1:1000
Anti-Lig4, rabbit monoclonal	Abcam	Cat# 193353	1:1000
Anti-Tdp2, rabbit polyclonal	Bethyl	Cat# A302-737A	1:1000
Anti-CtIP, rabbit polyclonal	Bethyl	Cat# A300-488A	1:1000
Anti-Cas9, mouse monoclonal	Active Motif	Cat# 61577	1:5000
Anti-alfa-tubulin, mouse monoclonal	Sigma Aldrich	Cat# T5168	1:2000
Anti-GAPDH, mouse monoclonal	Santa Cruz	Cat# 47724	1:1000
Anti-mouse IgG HRP-linked, horse polyclonal	Cell Signaling Technologies	Cat# 7076	1:1000
Anti-rabbit IgG HRP-linked, goat polyclonal	Cell Signaling Technologies	Cat# 7074	1:1000
Anti-goat IgG HRP-linked, mouse monoclonal	Santa Cruz	Cat# 2354	1:1000

#### 4.8 ChIP-qPCR of CTCF in TK6 cells

Chromatin immunoprecipitation (ChIP) analysis of CTCF was performed in TK6 cells with truChIP Chromatin Shearing Kit (Covaris, 520154) for chromatin preparation and ChIP-IT High Sensitivity Kit (Active Motif, 53040) for immunoprecipitation and DNA purification according to manufacturers’ instructions. Chromatin from  $2 \times 10^7$  cell was prepared by fixation in 1 % formaldehyde for 8 min and shearing of chromatin by 8 min pulsing (5 % duty factor,

intensity 4.0, 200 cpb) in a S2 Focused-ultrasonicator (Covaris). For the IP, 4 µg of antibodies for CTCF (Active Motif, 61311) or IgG (ThermoFisher, 31235) were used.

The presence of CTCF at *MLL* exon 12 or a negative control region in or outside of *MLL* was analyzed by comparison of DNA quantities in input samples and ChIP samples. For the qPCR, the SsoFast EvaGreen Supermix (Bio-Rad, 1725204) was used for 20 µL reactions run in a CFX96 Real-Time PCR Detection System (Bio-Rad). *MLL* exon 12 or control regions were amplified with primers shown in Table 6. The amplification was performed with denaturation at 95°C and annealing/extension at 60°C for 40 cycles.

**Table 6 Primers for qPCR amplification of immunoprecipitated chromatin of *MLL* exon 12 or control regions.**

Name	Sequence 5' to 3'
Neg_onMLL_fw	GCATAGCAGGCAACATACGA
Neg_onMLL_rev	GCATAGGGAGAATGCCCTCA
Neg_outMLL_fw	GTGGTTTGGGTGGCAGTTAT
Neg_outMLL_rev	TCCAGCTCTTCTCCCTCT
MLL_exon12_fw	TTGACATACTTCTATCTTCCCATGT
MLL_exon12_rev	GCCTCCACAAACGTGACAG

#### 4.9 Immunofluorescence

For immunofluorescence (IF), cells were fixed with 4 % paraformaldehyde (PFA) for 15 min in 96-well or 384-well cell-culture treated microscopy plates (PerkinElmer 6005550, 6057300). Therefore, adherent cells were directly grown in plates that were coated with poly-L-lysine (PLL, Sigma-Aldrich) and fixed with 4 % PFA/PBS. Suspension cells were generally grown in flasks and transferred into PLL-coated plates, were gently spun down at 400 × g for 15 sec pulsing and a same volume of 8 % PFA was added to reach a final concentration of 4 % PFA. Cells were washed three times with PBS after fixation to then be permeabilized with 0.3 % TritonX-100 (Sigma-Aldrich) in PBS for 5 min. Upon another three washes with PBS, cells were treated with 3 % BSA in PBS blocking buffer for 1 hour before incubation with primary antibodies overnight (Table 7). The following day, cells were washed three times with PBS and incubated with secondary, fluorescent antibodies for 1 h in the dark (Table 7). After three PBS washes, cells were stained with Hoechst33342 (Sigma-Aldrich) for 10 min which was then washed out with PBS.

Immunofluorescence samples were imaged with the *Opera Phenix High Content Screening System* (PerkinElmer) and were analyzed by automated *Harmony High Content Imaging and Analysis Software* (PerkinElmer) as indicated in Chapter 4.12.1.

**Table 7 Antibodies and dilutions for immunofluorescence.**

Protein target	Company	Identifier	Dilution
Anti-Top2A (G-6), mouse monoclonal	Santa Cruz	Cat# 166934	1:200
Anti-Top2B (H-286), rabbit polyclonal	Santa Cruz	Cat# 13059	1:100
Anti-phospho-Histone H2A.X (Ser139), mouse monoclonal (JBW301)	Millipore	Cat# 05-636	1:1000
Anti-Phospho RPA32(S4/S8), rabbit polyclonal	Bethyl	Cat# A300-245A	1:2000
Anti-Cas9, mouse monoclonal	Active Motif	Cat# 61577	1:1000
Anti-mouse IgG Alexa Fluor 568, donkey polyclonal	Invitrogen	Cat# A10037	1:1000
Anti-rabbit IgG Alexa Fluor 488, goat polyclonal	Invitrogen	Cat# A11034	1:1000
Anti-rabbit IgG Alexa Fluor 568, goat polyclonal	Invitrogen	Cat# A11011	1:1000
Anti-mouse IgG Alexa Fluor 488, donkey polyclonal	Invitrogen	Cat# A21202	1:1000
Anti-53BP1, rabbit polyclonal	Novus Biologicals	Cat# 100-304	1:1000

#### 4.10 Fluorescence *in situ* hybridization (FISH)

Fluorescently labelled FISH probes were generated from bacterial artificial chromosomes (BACs from BACPAC resources center CHORI or ThermoScientific) that were directly labelled by a nick translation enzymatic reaction that incorporated fluorescently labelled dUTPs of the color of choice. The correct identity of BACs were verified by PCR amplification of BAC DNA Minipreps (Zymo Research, D4048) with BAC specific primers. Correct BAC probes were prepared with the NucleoBond® Xtra BAC kit (MACHEREY-NAGEL, 740436). BAC DNA was labelled by nick translation following manufacturer's instructions (Abbott, 7J0001) using fluorescently labeled dUTPs (*Green*: Chromatide AlexaFluor 488-5-dUTP, *Red*: 568-5-dUTP, *FarRed*: AlexaFluor 647-AHA-dUTP from ThermoFisher Scientific (C11397, C11399, A32763) and *Blue*: CF405S-dUTP from Biotium (40004)). In Table 8, it is described which genomic locations was labelled with which BAC ID.

**Table 8 Bacterial artificial chromosomes (BACs) used for FISH labelling.**

BAC name	Source	Identifier
MLL-5'	ThermoFisher	CTD-2159M9
MLL-3'	BACPAC Resources CHORI	RP11-59N1
AF4-5'	BACPAC Resources CHORI	RP11-711J3
AF4-3'	BACPAC Resources CHORI	RP11-1105H6
AF6-5'	BACPAC Resources CHORI	RP11-351J23
AF6-3'	BACPAC Resources CHORI	RP11-1013J5
AF9-5'	BACPAC Resources CHORI	RP11-607L24
AF9-3'	BACPAC Resources CHORI	RP11-459O18
ENL-5'	BACPAC Resources CHORI	RP11-114A7
ENL-3'	BACPAC Resources CHORI	RP11-1056A10
ALK-5'	BACPAC Resources CHORI	RP11-119L19
ALK-3'	BACPAC Resources CHORI	RP11-100C1
NPM1-5'	BACPAC Resources CHORI	RP11-1072I20
NPM1-3'	ThermoFisher	CTD-2336J1

For 3D FISH, cells were either plated on glass poly-L-lysine (PLL) coated coverslips (13 mm in diameter, 170  $\mu$ m thick, Marienfeld Superior, 117530) in 24-well plates or in PLL-coated 384-well microscopy plates (CellCarrier-384 Ultra Microplates, PerkinElmer, 6057300). Adherent cells were directly grown on PLL-coated coverslips or plates for one to three days and fixed with 4 % PFA/PBS. Suspension cells were grown in cell culture flasks and were transferred into PLL-coated plates for fixation. Plates were centrifuged at 400 $\times$  g for 15 sec pulsing to spot suspension cells on plate bottoms and a same volume of 8 % PFA was added to reach a final concentration of 4 % PFA. After fixation in 4 % PFA for 20 min, cells were washed three times with PBS and permeabilized in 0.5 % saponin (Sigma Aldrich)/0.5 % Triton X-100 (Sigma Aldrich) /PBS) for 20 min. Upon three PBS washes, samples were incubated in 0.1 N HCl for 15 min and then washed with 2  $\times$  SSC for 10 min. Finally, samples were incubated in 50 % formamide/ 2  $\times$  SSC buffer for at least 30 min.

For each sample (one coverslip or one well of a 384-well plate), 80 ng of respective FISH probes from nick translations were mixed with 3  $\mu$ g COT1 DNA (GeneON, 3001) and 20  $\mu$ g yeast carrier RNA (ThermoFisher, AM7118) and ethanol precipitated. The probe preparation was pooled by upscaling for several samples and pellets were resuspended in 7  $\mu$ L (for coverslips) or 13 $\mu$ L (for 384-wells) hybridization buffer (10 % dextran sulfate, 50 % formamide, 2  $\times$  SSC, and 1 % Tween-20) per sample. Prepared probes were put onto cells and denatured at 85 $^{\circ}$ C for 5 min (for coverslips) or 10 min (for 384-well plates). Hybridization was achieved at 37 $^{\circ}$ C under humidified conditions overnight. The following day, excess probes were washed away with 2  $\times$  SSC followed by three 1  $\times$  SSC washes and three 0.1  $\times$  SSC washes, each for 5 min and at 45 $^{\circ}$ C. DNA staining was done for 2-color or 3-color FISH

samples but not for 4-color FISH. In 384-well plates, DNA was stained by Hoechst33342 (Sigma-Aldrich) for 10 min, washed out with PBS. Coverslips were mounted on glass slides (neoLab Migge) in Vectashield mounting medium (Vector Laboratories, H-1000 or H-1200), contained DAPI or not, and were sealed off using picodent twinsil (Picodent, 1300 1000).

### 4.11 High-Throughput Imaging

Images were acquired with a spinning disk Opera Phenix High Content Screening System (PerkinElmer), equipped with four laser lines (405 nm, 488 nm, 561 nm, 640 nm) and two 16 bit CMOS cameras (2160 by 2160 pixels).

For immunofluorescence samples, images were acquired in confocal mode using the 20× (NA 1.0) or 40× (NA 1.1) water objective lenses and a camera pixel binning of 2. Generally, three to five z-planes were imaged for all channels and analysis were done on maximally projected images.

Cell cycle analysis of fixed cells by fluorescence microscopy was performed as shown previously (Roukos *et al*, 2015). Three z-planes separated by 1 µm were obtained with the widefield mode of the 20× water objective lens (NA 1.0) of the Opera Phenix. To analyze cell cycle staging, an average of 10'000 cells for each condition were imaged by acquiring images of at least 30 randomly sampled fields.

Images of FISH experiments to calculate 3D distances were acquired in confocal mode using a 40× water objective lens (NA 1.1) with a camera pixel binning of 2 (corresponding to a pixel size of 300 nm). In general, between 9 and 16 z-planes separated by 0.5 µm were measured to span the different nuclear heights of the different cell lines. At least 40 and up to 200 randomly sampled fields were acquired per sample to cover an average of 10'000 cells per condition.

### 4.12 Image analysis

#### 4.12.1 Nuclei segmentation, intensity measurements for immunofluorescence analysis

Images were analyzed with the *Harmony High Content Imaging and Analysis Software* (version 4.4, PerkinElmer) using standard and customized building blocks. For IF analysis, nuclei of

cells were detected and segmented based on signals from DNA staining with Hoechst33342. For nuclear detection of suspension cell lines, the method A or B were used whereas adherent cells were detected based on method C of the Harmony software. Cells in the periphery of the image that were cut off were excluded from further analysis. Intensity variations by non-flat illumination were corrected by the flatfield correction algorithm *Advanced*, if available. The mean or sum of measured fluorescence intensities was determined in maximum projected images. Foci were detected by spot detection method B or D. Single-cell data were exported from Harmony for visualization and statistical testing by custom made R scripts (DistancePlotter, <https://github.com/imbforge/DistancePlotter.git>) and GraphPad Prism (version 7.02). The DistancePlotter was developed by the Oliver Drechsel of the Bioinformatics Core Facility (Institute of Molecular Biology, Mainz).

The thousands of individual data points for each IF sample were visualized as boxplots or jitter plots of the mean or sum of nuclear intensities or foci count per nucleus. For a simplified visualization, boxes of boxplots show the interquartile range with median and whiskers represent the 10-90 % or 5-95 % intervals. Thereby, some outliers are not graphically depicted but they are included in statistical testing. Statistical significance of changes in IF analysis was tested by the One-Way ANOVA and Tukey test. Visualization and analysis of cell cycle profiles derived from Hoechst/DAPI nuclear staining was performed as before (Roukos et al., 2015) and profiles are shown as density plots.

#### 4.12.2 C-Fusion 3D

##### *Nuclei segmentation and spot distance calculation in Harmony*

Images were analyzed with the *Harmony High Content Imaging and Analysis Software* (version 4.4, PerkinElmer) using standard and customized building blocks. For FISH analysis, nuclei of cells were detected and segmented based on signals from DNA staining with Hoechst33342 or DAPI or, in case of 4-color FISH analysis, the background signal of the *Red/AlexaFluor 568* or *FarRed/Alexa 647* staining. Nuclear detection was done as described above (Chapter 4.12.1). Spots of FISH probes were detected with the method C in all channels in maximal projected images. Distances between all different spots of all different channels were calculated in 2D by using custom-made building blocks. For allow for 3D calculation of distances, for each spot the z-plane with the maximum pixel intensity was identified. Single cell information was then exported from Harmony for analysis with custom made R scripts that were developed by Oliver Drechsel (previously in the IMB Bioinformatics Core Facility) and were based on previous versions from Gianluca Pegoraro (Burman et al., 2015).

### *Derivation of 3D distances between spot pairs*

To determine 3D distances between spots, the Euclidean distances of spot distances from maximum projections and distances between z-planes were calculated. Distances between z-planes of spots of different color were corrected for shifts due to chromatic aberration in z. This is due to the fact that fluorophore of different colors are shifted in the third dimension (Finn et al., 2017). To determine the z-shift, a genomic locus was marked with a BAC probe that was labeled with all different color fluorophores. From this, the pairwise shifts in z between the different channels were measured and a correction value was calculated (*Green*: -0.768, *FarRed*: 0.255, *Blue*: -0.146, see Figure 2.6). For 3D evaluations, the z-plane of spots that had the maximum pixel intensity was changed according to the defined correction value for each color-combination, whereby each spot in every channel was corrected according the identified shift between the different channels in z. The following formula was used to determine z-correction values:

$$\left(1 - \frac{\text{top.count}}{\text{top1} + 2.\text{count}}\right) * \Delta\text{distance} + \text{top.distance}$$

top.distance = the distance most often observed

top.count = the count of top.distance

top1+2.count = the count of the two most often observed distances

$\Delta$ difference = the difference of the top two distances

For the determination of proximity or separation of spot pairs, only the minimal distance of one spot to spots of other channels was considered. This is necessary as in in diploid cells for example each *Green* spot has two distances to both *Red* spots. Taking the smaller of the two distances secures with high probability that the distance to the spot on the same chromosome is analyzed.

### *Chromosome breakage and translocation analysis*

To discern chromosome loci that experienced damage and breakage from intact loci, we established a thresholds of physical separation and proximity (Burman et al., 2015). The distribution of intrachromosomal distances was evaluated in non-damaged control cells by evaluating *Green-Red* and *FarRed-Blue* distances. We defined a threshold of 1.2  $\mu\text{m}$  (corresponds to 4 pixels) for separation or proximity, where distances smaller or equal the threshold were considered intact and distances of more than 1.2  $\mu\text{m}$  were seen as separations. To eliminate cases of falsely identified or missed spots, only cells with the same



number of *Green* and *Red* spots, or *FarRed* and *Blue* spots respectively, were considered in the analysis. The number of spot pairs was defined based on the ploidy of cells. For TK6 cells, only cells with two signals in all channels were considered in the analysis whereas for K562 cells, for example, cells with more than two but equal spot number were also analyzed.

A cell was considered positive of a chromosome breakage event, when at least one *Green* FISH spot that had a minimum *Green-Red* distance of more than the 1.2  $\mu\text{m}$  threshold or when a *FarRed* FISH spot had a corresponding minimum *FarRed-Blue* distance of more than the threshold. Cells were assigned positive for synapsis when *Green-FarRed* FISH spot distances were less or equal than the threshold. For 4-color FISH, a cell was considered positive of a chromosome translocation when three conditions were met concomitantly in one cell:

- 1) a *Green* FISH spot had a *Green-Red* distance of more than the threshold
- 2) the same *Green* FISH spot had a minimum *Green-Blue* distance of more than the threshold
- 3) the same *Green* FISH spot had a minimum *Green-FarRed* distance of less than the threshold

In a 3-color C-Fusion 3D approach, a cell with a translocation was assigned when the first and third criteria were met, i.e. at least one *Green* FISH signal had a minimal *Green-Red* distance of more than the threshold and a concomitant *Green-FarRed* distance of fewer than the threshold. Frequencies of chromosome synapsis, breakage and translocations were calculated in R and exported as HTML files, plots were generated in GraphPad Prism. Values were visualized as bar plots of mean plus standard deviation and statistical significance was tested by two-tailed student's t test.

#### 4.13 Cell harvesting and fixation for sBLISS

Genome-wide mapping of DSB in human cell lines was done by sBLISS, which is an adaptation of the previously published BLISS protocol (Yan et al., 2017). In sBLISS, DSB ends are labeled in cell suspensions in microcentrifuge tubes. sBLISS experiments were collaboratively designed and generated by the groups of Dr. Vassilis Roukos and Dr. Nicola Crosetto. We designed experimental setups and grew, treated, and harvested cells by fixation for sBLISS. Britta Annika Maria Bouwman from the Crosetto group optimized the sBLISS protocol and processed all samples. For details of the sBLISS sample preparation, in vitro transcription, sequencing and data preparation, please refer to Supplementary methods on sBLISS in Chapter 4.14.

## Material and Methods

ETO-induced DSBs were mapped in the suspension cells TK6, K562 and CD34+ progenitor cells. To identify ETO-induced free DSBs, cycling cells at medium density were treated with ETO (Sigma-Aldrich, E1383) for 4 to 6 h at 20 to 60  $\mu\text{M}$ , dependent on the cell line. For comparison, cells from the same population were control treated with DMSO. To study the effect of transcription in TK6 cells, cells were pretreated with DRB at 200  $\mu\text{M}$  for 3 h or not and afterwards treated with ETO at 30  $\mu\text{M}$  for 4 h. For controls, equal amounts of DMSO as were used by DRB and ETO-treatment were added. All replicates and treatments are listed in Table 9. To verify ETO-induced DNA damage in sBLISS samples, some cells were fixed for immunofluorescence staining of DNA damage marker  $\gamma\text{H2AX}$  as described in Chapter 4.9.

To harvest suspension cells for sBLISS, cells were counted and pelleted by mild centrifugation (100-150 g for 10 min, dependent on the lowest speed necessary to pellet the cells). Cells were gently resuspended in prewarmed medium containing 10 % fetal bovine serum (K562, CD34+) or horse serum (TK6 cells) to yield a single-cell suspension of  $1.33 \times 10^6$  cells/mL. For each sample, 5 to 10 million cells were fixed by the addition of 16 % paraformaldehyde aqueous solution (Electron Microscopy Sciences, 15710, EM-grade, methanol-free) to a final concentration of 4 % and incubating for 10 min on a roller shaker at RT. The fixation was quenched by 2 M glycine added to a concentration of 125 mM and incubated for 5 min at RT, while gently rotating, followed by an additional 5 min on ice. Fixed cells were pelleted by centrifuging at 100-150 g for 10 min at 4°C and washed twice with ice-cold PBS. Final sBLISS pellets were kept in PBS at 4°C until sBLISS template preparation.

**Table 9 Datasets generated and published within this study.**

Cell line	Treatment	Replicate	Accession number
TK6	DMSO	1	GSM3444984
TK6	ETO 30 $\mu$ M for 4h	1	GSM3444985
TK6	DMSO	2	GSM3444986
TK6	ETO 30 $\mu$ M for 4h	2	GSM3444987
K562	DMSO	1	GSM3444988
K562	ETO 20 $\mu$ M for 6h	1	GSM3444989
CD34+	DMSO	1	GSM3687236
CD34+	ETO 60 $\mu$ M for 4h	1	GSM3687237
CD34+	DMSO	2	GSM3687238
CD34+	ETO 60 $\mu$ M for 4h	2	GSM3687239
TK6	DRB 200 $\mu$ M for 3h, DMSO	1	GSM3444990
TK6	DMSO	1	GSM3444991
TK6	DRB 200 $\mu$ M for 3h, ETO 30 $\mu$ M for 4h	1	GSM3444992
TK6	ETO 30 $\mu$ M for 4h	1	GSM3444993
TK6	DRB 200 $\mu$ M for 3h, DMSO	2	GSM3444994
TK6	DMSO	2	GSM3444995
TK6	DRB 200 $\mu$ M for 3h, ETO 30 $\mu$ M for 4h	2	GSM3444996
TK6	ETO 30 $\mu$ M for 4h	2	GSM3444997

#### 4.14 Supplementary methods: sBLISS

The following methods were contributed by Dr. Britta Annika Maria Bouwman and Dr. Nicola Crosetto. They are part of the collaborative publication about this research (Gothe et al., 2019) and are listed here to complement the overall methodologies that were important to generate the here presented data.

##### 4.14.1 sBLISS template preparation

To prepare single-nucleus suspensions for tagging of DSBs by in situ ligation of BLISS adapters,  $10^6$  fixed cells were lysed for 60 minutes on ice in a lysis buffer containing 10 mM Tris-HCl, 10 mM NaCl, 1 mM EDTA, and 0.2 % Triton X-100 (pH 8). We then pelleted the lysed cells at RT, removed the supernatant, and permeabilized the nuclei for 60 minutes at 37°C with a second pre-warmed lysis buffer containing 10 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, and 0.3 % SDS (pH 8). Then, nuclei were washed twice with pre-warmed 1× CutSmart Buffer (New England Biolabs (NEB) #B7204) supplemented with 0.1 % Triton X-100 (1× CS/TX100), and DNA Double Strand Break ends (DSB ends) were blunted with NEB's Quick Blunting Kit (NEB #E1201) according to the manufacturer's instructions in a final volume of 100 µl for 60 min at RT.

Blunted nuclei were washed twice with 1× CS/TX100 before proceeding with in situ ligation of sBLISS adapters (see below adapter preparation) to the blunted DSB ends. Adapter ligation was performed with 25 Weiss units T4 DNA Ligase (5 U/µl, ThermoFisher Scientific #EL0011) for 20-24 h at 16°C in reaction volumes of 100 µl, according to the manufacturer's instructions and supplemented with BSA (Thermo #AM2616) and ATP (Thermo #R0441). Per prep of  $10^6$  cells, 4 µl of a 10 µM working solution of the selected BLISS adapter was ligated. Prior to use, BLISS dsDNA adapters were prepared from two complementary HPLC-purified oligonucleotides ordered from Integrated DNA Technologies (IDT). Each dsDNA adapter contains the T7 promoter sequence for in vitro transcription (IVT), the RA5 Illumina RNA adapter sequence for downstream sequencing, an 8-nt Unique Molecular Identifier (UMI) sequence generated by random incorporation of the four dNTPs according to IDT's 'Machine mixing' strategy, and an 8-nt sample barcode for multiplexing of sBLISS libraries. First, sense oligos diluted to 10 µM in nuclease-free water were phosphorylated using NEB's T4 PNK system (NEB #M0201) supplemented with ATP, after which an equimolar amount of antisense oligo was added. We used a PCR thermocycler to anneal both oligos (5 min at 95°C, then cooling down to 25°C in cooling steps of 1.5°C per minute) to generate a 10 µM phosphorylated dsDNA adapter.

After ligation overnight, nuclei were washed twice with 1× CS/TX100. Then we reversed crosslinks and extracted gDNA by resuspending the nuclei in 100 µl DNA extraction buffer containing 10 mM Tris-HCl, 100 mM NaCl, 50 mM EDTA, and 1 % SDS (pH7.5), and supplemented with 10 µl Proteinase K (800 U/ml, NEB #P8107). Nuclei were incubated at 55°C for 14-18h while shaking at 800 rpm. When clumps were still present the next morning, we added an additional 10 µl Proteinase K for 1 hour. Afterwards, Proteinase K was heat-inactivated for 10 min at 95°C, followed by extraction using Phenol:Chloroform:Isoamyl Alcohol 25:24:1 with 10 mM Tris, pH 8.0, 1 mM EDTA (Sigma-Aldrich/Merck #P2069) and Chloroform (Merck #1024451000), and ethanol precipitation. We sonicated the purified gDNA in 100 µl TE using a BioRuptor Plus (Diagenode) with the following settings: 30 s ON, 60 s OFF, HIGH intensity, 30 cycles. Sonicated samples were concentrated with Agencourt AMPure XP beads (Beckman Coulter) and fragment sizes were assessed using a BioAnalyzer 2100 (Agilent Technologies) to range from 300 bp to 800 bp, with a peak around 400-600 bp. Sonicated and purified BLISS template was stored at -20°C until IVT and library preparation were initiated.

#### 4.14.2 sBLISS in vitro transcription (IVT) and NGS library preparation

Equal amounts (100 or 200 ng) of purified sonicated BLISS template of treated (Etoposide) and control (DMSO) samples were pooled into a single IVT reaction for linear amplification by T7-mediated transcription of the genomic ends of the BLISS-labeled DSBs. For IVT we used the MEGAscript T7 Transcription Kit (ThermoFisher #AMB13345) according to the manufacturer's prescriptions, with the exception that transcription was carried out for 14 hours at 37°C and that Ribosafe RNase Inhibitor (Bioline #BIO-65028) was added. Upon completion, gDNA was removed with 2 units DNase I (RNase-free) (ThermoFisher #AM2222) and amplified RNA (aRNA) was purified and concentrated using Agencourt RNAClean XP beads (Beckman Coulter). Next, oligos with the Illumina RA3 adapter sequence (Integrated DNA technologies) were ligated to the purified aRNA using T4 RNA Ligase 2, truncated (NEB #M0242) for two hours at 25°C. No RA5 adapter ligation was necessary, as this sequence was already present in each RNA molecule due to the composition of the BLISS adapter. Directly after RNA ligation, we performed reverse transcription with Reverse Transcription Primer (RTP) (Illumina sequence, ordered via IDT) and SuperScript IV Reverse Transcriptase (ThermoFisher #18090050). The manufacturer's prescriptions were followed with the exception of incubation time, which was extended to 50 min at 50°C followed by heat inactivation for 10 min at 80°C. To prevent RNase activity, we supplemented with RNaseOUT (ThermoFisher #10777019) during RA3 ligation and reverse transcription. After completion

of reverse transcription, libraries were amplified with NEBNext High-Fidelity 2× PCR Master Mix (NEB #M0541), the RP1 common primer, and a selected RPIX index primer (Illumina sequences, ordered through IDT). In total, 12 PCR cycles were completed following the manufacturer's protocol, after which the amplified libraries were subjected to clean-up according to the two-sided AMPure XP bead purification protocol aimed to retain library sizes ranging from roughly 300-800 bp. Final library profiles were assessed and quantified on a BioAnalyzer High Sensitivity DNA chip and using the Qubit dsDNA HS Assay kit (ThermoFisher #Q32851) and whenever necessary AMPure XP bead purification was repeated.

### 4.14.3 sBLISS sequencing and data processing

Sequencing was performed in-house on a NextSeq 500 with NextSeq 500/550 High Output Kit v2 chemistry for single-end (1×76) or paired-end (2×151) sequencing, with an additional 6 cycles for index sequencing when multiple libraries were pooled. Upon completion of the run, raw sequencing reads of pooled libraries were demultiplexed based on index sequences by Illumina's BaseSpace, after which the generated FASTQ files were downloaded. For libraries sequenced separately this step was omitted and FASTQ files were downloaded directly. As described previously (Yan et al., 2017), we applied a custom-built pipeline to keep only the reads containing the expected prefix of 8nt UMI and 8nt sample barcode using SAMtools (version 1.7) (Li et al., 2009) and scan for matches, allowing at most one mismatch in the barcode. The prefixes were clipped off and stored, and the trimmed reads per condition were aligned to the GRCh37/hg19 reference genome with BWA-MEM (version 0.7.17-r1188) (Li and Durbin, 2009). Only those reads with mapping quality scores  $\geq 30$  were retained. Next, we identified and removed PCR duplicates by searching for proximal reads (at most 30bp apart in the reference genome) with at most two mismatches in the UMI sequence. Finally, we generated BED files for downstream analyses, comprising a list of DSB end locations and a number of unique UMIs identified at these locations, which we refer to as 'UMI-DSB ends' or unique break ends.

### 4.14.4 sBLISS limitations on quantifying DSBs between different cell lines

Absolute comparison of DSB-counts between different cell lines by sBLISS (or any other NGS-based methods used to profile DSBs) is not currently quantitative. This is mainly due to different efficiencies in fixation, in situ blunting and/or adaptor ligation that likely exist between different cell lines. Different normalization strategies that can be used to make inter-

sample comparison more quantitative (e.g. spike-in experiments with a common DSB source etc) will help overcoming this limitation in the near future.

#### 4.15 Supplementary methods: Bioinformatics analyses

The following methods were contributed by researchers from the groups of Prof. Dr. Argyris Papantonis and the IMB Bioinformatics Core Facility. They following analysis were designed and generated with inputs from Dr. Vassilis Roukos and Prof. Dr. Argyris Papantonis.

The detection of sBLISS hotspots , ENCODE data analysis and genomic regions analysis were contributed by Dr. Giuseppe Petrosino and Dr. Sergi Sayols. Analysis of CTCF motifs and chromatin looping analysis were done by Dr. Eduardo Gade Gusmao. Natasa Josipovic and Dr. Athanasia Mizi performed and analyzed the nascent RNA-seq data in TK6 cells.

These methods are part of the collaborative publication about this research (Gothe et al., 2019) and are listed here to complement the overall methodologies that were important to generate the here presented data.

##### 4.15.1 Detection of sBLISS hotspots

We used Macs2 (version 2.1.1.20160309) (Zhang et al., 2008) to call peaks from the BED files of UMI-DSB. First, we created one entry per UMI-DSB in the BED file, and called Macs2 assuming no background model, no cross-correlation between strands around the hotspots, and requiring Macs2 to shift the reads by -100 bp and extending them 200 bp in order to recreate the original fragment, with its middle position pointing to the original UMI-DSB locus. What we call hotspots are the peaks identified by Macs2 with q-value < 0.05. Hot spots lists from different replicates were merged using bedtools (Quinlan, 2014) (version 2.25.0; default parameters).

##### 4.15.2 Definition of genomic regions

Gene body coordinates and transcription start sites were obtained from ENSEMBL (Release 75) (Zerbino et al., 2018) for the reference genome GRCh37 (hg19). Active genes were defined as overlapping H3K4me3 or H3K27ac histone modifications peaks or having a significant (Fisher's exact test,  $P < 0.05$ ) enrichment for RNA Polymerase II. Inactive genes were defined as the genes that did not match any of the previous criteria. Promoter regions were defined as a window of 3Kbp centered at the transcription start site of each gene. Active and inactive promoter annotations simply follow the same annotation from the gene in which it belongs

to. Enhancers were defined as following: we obtained K562- and TK6-specific enhancers from the Enhancer Atlas (Gao et al., 2016) and human enhancers from FANTOM5 (Andersson et al., 2014). These two enhancer datasets were merged together, forming a comprehensive enhancer list. Active enhancers were defined as the enhancers from the comprehensive list, which overlapped H3K4me1 and H3K27ac histone modifications peaks. The remaining enhancers from the comprehensive list were defined as the inactive/poised enhancers.

### 4.15.3 ENCODE data analysis

ChIP-seq (for transcription factors and histone modifications) and DNase-seq mapped reads (reference genome GRCh37; hg19, from 2016-03-06) data were downloaded from ENCODE (Davis et al., 2018) as BAM files (see Table S2). The reads from all BAM files were corrected for GC content and artificial genomic locus enrichment using BEADS (version 1.1)(Cheung et al., 2011). Furthermore, the reads from all BAM files were normalized using the reads-per-kilobase-per-million (RPKM) approach using the standard methodology (Conesa et al., 2016), when two or more datasets were being compared. When using the BAM files directly for ChIP-seq data, each read was extended downstream by 200 bp from its 5' end to reflect the global average fragment length from the ChIP procedure (Gusmao et al., 2016). When using DNase-seq data for graphical purposes, we used the “base-overlap” signal strategy. Briefly, all reads are modified to exhibit only +/-5 bp from the 5' end reads. For ChIP-seq data, we called peaks using Macs2 (version 2.1.1.20160309, parameters: --keep-dup auto --nomodel --nolambda --shift -100 --extsize 200 --call-summits). For DNase-seq data, all reads were first pre-processed by keeping only the 5' end bp, which reflects the here the DNase I enzyme digested the DNA (Gusmao et al., 2016), and, then, we called peaks using Macs2 (parameters: --nomodel --nolambda --call-summits). We further filtered all peaks, for each dataset, by the Macs2-based false positive rate corresponding to their 25<sup>th</sup> percentiles as recommended (Meyer and Liu, 2014).

The signal files (BIGWIG) for MNase-seq data were downloaded from ENCODE (Davis et al., 2018) and used for further analysis. DMSO or ETO sBLISS profiles from one replicate in K562 cells, were compared with data derived from ENCODE (K562 cells). sBLISS profiles from merged replicates (rep1+rep2) of DMSO or ETO samples in the TK6 lymphoblastoid cells, were compared with ENCODE data produced from the lymphoblastoid GM12878 cells (Table S2). Data were visualized with the WashU EpiGenome Browser. Merging between replicates in all TK6 or CD34+ sBLISS samples was performed by merging .bam files between the two biological replicates (DMSO or ETO; rep1 and rep2).



#### 4.15.4 CTCF motifs

The CTCF binding affinity motif was obtained in the Jaspas repository (Khan et al., 2018) and matched to the reference genome GRCh37 (hg19) using the package MOODS (version 1.9.3) (Korhonen et al., 2017). The motif hits were further filtered by a false positive rate of  $10^{-4}$  using the dynamic programming estimation of the distribution of bit-scores approach as described in (Wilczynski et al., 2009). CTCF motifs with higher likelihood to be bound by the CTCF protein (“active motifs”), were discovered using the HINT-BC tool (version 0.11.4) (Gusmao et al., 2016) using the following approach: (1) First, DNase-seq data for K562 and TK6 cell types was obtained from ENCODE (Davis et al., 2018); (2) DNase-seq hypersensitivity sites (peaks) were called using Macs2 version 2.1.1.20160309 (parameters: `--nomodel --nolambda --call-summits`); (3) The final set of hypersensitivity regions was obtained by extending  $\pm 100$  bp from the peak summit; (4) HINT-BC was executed using the DNase-seq data and the hypersensitivity regions as input; (5) Finally, we kept only the CTCF hits which overlapped by at least 1bp the resulting footprints from HINT-BC.

#### 4.15.5 Chromatin Loop analyses

We obtained the chromatin intrachromosomal contacts list for K562 cells and lymphoblastoid GM12878 cells (for comparisons with TK6 cells), including both upstream and downstream loop anchors and contact-mediating CTCF binding positions, from (Rao et al., 2014). We segregated all genes into two categories: (1) Genes overlapping loop anchors – if the gene overlapped a loop anchor by at least one base pair; and (2) Genes not overlapping loop anchors – if no overlap existed between the gene and loop anchors. For the genes overlapping loop anchors, we considered a gene to be “outside a chromatin loop” if the TSS of the gene was upstream of the anchor’s CTCF (for all upstream anchors) or downstream of the anchor’s CTCF (for all downstream anchors). Similarly, all genes whose TSS were upstream of the anchor’s CTCF (for all downstream anchors) or downstream of the anchor’s CTCF (for all upstream anchors) were regarded as “inside a chromatin loop”. Given this classification the genome was segregated into four categories: (1) Genes that were in the forward and reverse strand outside of a chromatin loop; (2) genes that were in the forward and reverse strand inside of a chromatin loop; (3) genes that were exclusively in the forward strand both inside and outside a chromatin loop; and (4) genes that were exclusively in the reverse strand both inside and outside a chromatin loop. These genes were further divided into active/inactive based on the same criteria defined in the STAR Methods section “Definition of genomic regions”. For clarity, we removed from these analyses all regions in which genes were overlapping in different strands. The criteria for a loop anchor to be annotated with a CTCF

were either: (1) the loop anchor was already annotated with a CTCF binding site in the original list obtained in (Rao et al., 2014); or (2) the loop anchor presented a DNase-seq footprint obtained by applying the software HINT-BC overlapping (by at least one base pair) a CTCF ChIP-seq peak obtained by applying the software Macs2 version 2.1.1.20160309 in ENCODE-derived CTCF ChIP-seq data (Davis et al., 2018). For more information, please refer to the Methods sections “CTCF motifs” and “Encode data analysis”. Finally, all loop anchors which did not present a gene at least  $\pm 100$ Kbp from the center of their loop anchors were defined as “intergenic sites”.

CTCF loops were called and annotated from CD34+ Hi-C data (GSM2861708) using HICCUPS as described in (Rao et al., 2014); only loops anchors with CTCF motifs and ChIP signal (GSM2861703) were considered in our downstream analysis.

### 4.15.6 RNA-seq data analysis

The raw data of RNA-seq primary CD34+ cells were downloaded from European Nucleotide Archive (ENA) (see Table S2). Reads were aligned on the *H. sapiens* genome assembly hg19 (GENCODE release 19) reference sequences using STAR (Dobin et al., 2013) (version 2.5.4b, parameters: `--outSJfilterReads Unique --outFilterMismatchNmax 2 --outFilterMultimapNmax 10 --alignIntronMin 21 --sjdbOverhang reads_length - 1`). The featureCounts program (Liao et al., 2014)(version 1.5.1, parameter: `-s 2`) was used to count the number of reads overlapping genes. The counts were normalized by library size and gene length, then, the gene expression levels were converted to percentile-rank values.

### 4.15.7 GRO-seq data analysis

K562 GRO-seq data were previously generated (GSM1480325). For our purposes, we realigned the raw reads to the Illumina’s iGenomes ([https://support.illumina.com/sequencing/sequencing\\_software/igenome.html](https://support.illumina.com/sequencing/sequencing_software/igenome.html)) UCSC GRCh37 (hg19) human reference from 2016-03-06, using Bowtie (Langmead et al., 2009) (using the `--best --strata -tryhard` flags, allowing up to 2 mismatches and discarding reads mapping to multiple positions), after removing reads that aligned to the Human ribosomal DNA complete repeating unit (GenBank: U13369.1). We then counted the fraction of reads that fall on genes, distinguishing between exons and introns, using featureCounts from the Rsubread (Lawrence et al., 2013) package and the accompanying GTF gene annotation of the iGenomes reference genome. The counts were normalized by library size and gene length, then, the gene expression levels were converted to percentile-rank values.

#### 4.15.8 ChIP-seq data analysis

The raw data of TOP2B and CTCF ChIP-seq for MCF10A and primary CD34+ cells, respectively, were downloaded from Gene Expression Omnibus (GEO) (see Table S2). Reads were aligned on the *H. sapiens* genome assembly (UCSC GRCh37/hg19) reference sequences using Bowtie for single-end (SE) reads (version 1.2.2, parameters: `--sam --best --strata --tryhard --chunkmbs 2000 -l 40 -m 1`) and Bowtie2 for paired-end (PE) reads (version 2.3.4.3, parameters: `--very-sensitive --end-to-end --maxins 1000`) (Langmead et al., 2009). SAM output files were converted in BAM and sorted using the view and sort sub-commands from SAMtools (Li et al., 2009) (version 1.5). For PE data, multiple-mapping and duplicated reads were removed using SAMtools with option `'-q 3'` and picard MarkDuplicates (<https://broadinstitute.github.io/picard/>) (version 2.7.0, parameters: `REMOVE_DUPLICATES=TRUE ASSUME_SORTED=TRUE`), respectively.

The track file (bigwig) for TOP2B was obtained with a combination of bedtools (Quinlan, 2014) (version 2.27.1) sub-commands: Bam file was converted in bed with `bamToBed` and extended to approximately 200 bp with `slopBed`; then `bamCoverage` from `deepTools` (Ramírez et al., 2016) (version 3.1.0, parameters: `--outFileFormat bigwig --smoothLength 60 --binSize 20 --normalizeUsing CPM`) was used to generate the bigwig file. The Peaks were called with `Macs2` (Zhang et al., 2008) (version: 2.1.2, parameters for SE data: `-g hs --bw 200 --keep-dup auto`, additional parameters for PE data: `--min-length 200 --format BAMPE`) comparing IP sample against the input sample.

#### 4.15.9 Nascent RNA-seq (transcription factories)

Nascent RNA-seq experiment in TK6 cells was performed as previously described (Melnik et al., 2016). Raw sequencing reads were mapped to human hg19 reference genome with STAR (Dobin et al., 2013) (version 2.5.3a) and gene level counting was performed with `iRNA-seq` pipeline using count mode (`iRNA-v1.1`) (Madsen et al., 2015). The resulting read counts were normalized by reads-per-kilobase-per-million (RPKM) using the standard methodology (Conesa et al., 2016).

#### 4.15.10 Overlap of genomic features

All overlap between two features (including genomic regions, BLISS hot spots, peaks called using MACS for transcription factors, histone modifications or open chromatin and CTCF motifs) were calculated using `bedtools` (Quinlan, 2014) (version 2.25.0, parameters: `-wa -u`).

### 4.15.11 Statistical tests and corrections

Details of statistical testing of imaging data including the number of cells quantified, biological replicates, standard deviation and statistical tests used are reported in the figure legends.

For bioinformatics analyses, to account for both linearity and non-linearity, all correlations between samples were calculated based on the Spearman's correlation coefficient. Furthermore, since none of the datasets could be fit to a known distribution, we used the non-parametric hypothesis test Mann-Whitney-Wilcoxon in all cases of distribution comparisons. All tests were two sided and the confidence level used was 99%. The resulting p-values from all hypothesis tests were corrected for multiple alignment using the Benjamini-Hochberg method. We further corrected the reported p-values by its distribution's false positive rate using the Longstaff and Colquhoun calculator available at <http://fpr-calc.ucl.ac.uk/> (Colquhoun, 2014). We assumed frequency equality as priors in all tests.

We evaluated the significance of overlap between MLL and fusion partner genes with the loop anchors by using the GAT tool (Heger et al., 2013), using the following parameters: “gat-run.py --ignore-segment-tracks --segments=[segments] --annotations=[annotations] --workspace=[workspace]”. On the first execution [segments] were the set of MLL and translocation partner genes and on the second execution random genes defined as ten times sampling without replacement of genes in the same expression percentile as the MLL and translocation partners. On both executions, [annotations] was set to represent the loop anchors and [workspace] represented the whole human genome.

### 4.15.12 Multiple correlation test

For correlations between more than two variables, we used the nonlinear weighted least-squares method calculated with the Gauss-Newton algorithm. The summary statistics of such method provides an adjusted  $r^2$  statistics. The  $r^2$  statistics reflects the fraction of variance explained by the model; thus, it is widely used as a correlation metric. In this work, we used the square root of the adjusted  $r^2$  statistics as our correlation metric between three or more variables.

## 4.16 Data availability and Software

Bioinformatics scripts developed by Dr. Eduardo Gade Gusmao, were published in the following Github repository: [https://github.com/eggduzao/Gothe\\_et\\_al](https://github.com/eggduzao/Gothe_et_al).

Raw data files of microscopy images and western blot membranes were deposited in Mendeley Data: <https://data.mendeley.com/datasets/bfhpc4cgrr/draft?a=c0a044f3-9e78-466b-b559-5db1395f4af2>

All genomics datasets generated within this study were publicly deposited and can be found under the accession number GSE121742.

Software and publicly available datasets that were used for the analysis of the data are listed in the tables below:

**Table 10 Software used within this study.**

Software	Resource	Link
GraphPad Prism (version 7.02)	GraphPad	<a href="https://www.graphpad.com/scientific-software/prism">https://www.graphpad.com/scientific-software/prism</a>
R (version 3.4.3)	R Development Core Team	<a href="https://www.r-project.org">https://www.r-project.org</a>
Harmony (version 4.4)	Perkin-Elmer	<a href="http://www.perkinelmer.de/product/harmony-4-8-office-hh17000001">http://www.perkinelmer.de/product/harmony-4-8-office-hh17000001</a>
SAMtools (version 1.3.1, 1.5 and 1.7)	Li et al., 2009	<a href="http://samtools.sourceforge.net">http://samtools.sourceforge.net</a>
BWA-MEM (version 0.7.17-r1188)	Li et al., 2009	<a href="http://bio-bwa.sourceforge.net">http://bio-bwa.sourceforge.net</a>
Macs2 (version 2.1.1.20160309)	Zhang et al., 2008	<a href="https://github.com/taoliu/MACS">https://github.com/taoliu/MACS</a>
BEADS (version 1.1)	Cheung et al., 2011	<a href="http://beads.sourceforge.net/">http://beads.sourceforge.net/</a>
MOODS (version 1.9.3)	Korhonen et al., 2009	<a href="https://www.cs.helsinki.fi/group/pss/mfind">https://www.cs.helsinki.fi/group/pss/mfind</a>
HINT-BC (version 0.11.4)	Gusmao et al., 2016	<a href="http://www.regulatory-genomics.org/hint">http://www.regulatory-genomics.org/hint</a>
STAR (version 2.5.3a and 2.5.4b)	Dobin et al., 2013	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
featureCounts (version 1.5.1)	Liao et al., 2014	<a href="http://bioinf.wehi.edu.au/featureCounts">http://bioinf.wehi.edu.au/featureCounts</a>
iRNA-seq (version 1.1)	Madsen et al., 2015	<a href="http://bioinformatik.sdu.dk/solexa/webshare/iRNA">http://bioinformatik.sdu.dk/solexa/webshare/iRNA</a>
bedtools (version 2.25.0)	Quinlan et al., 2015	<a href="https://bedtools.readthedocs.io/en/latest/content/installation.html">https://bedtools.readthedocs.io/en/latest/content/installation.html</a>
Longstaff and Colquhoun calculator	Colquhoun et al., 2014	<a href="http://fpr-calc.ucl.ac.uk">http://fpr-calc.ucl.ac.uk</a>
subread (version 1.5.1)	Liao et al., 2013	<a href="http://bioconductor.org/packages/release/bioc/html/Rsubread.html">http://bioconductor.org/packages/release/bioc/html/Rsubread.html</a>
Bowtie (version 1.1.2)	Langmead et al., 2009	<a href="http://bowtie-bio.sourceforge.net/index.shtml">http://bowtie-bio.sourceforge.net/index.shtml</a>
kentUtils (version 302)	ENCODE-DCC	<a href="https://github.com/ENCODE-DCC/kentUtils">https://github.com/ENCODE-DCC/kentUtils</a>
Python (version 2.7.12)	Python Software Foundation	<a href="https://www.python.org">https://www.python.org</a>
Circos (version 0.69-2)	Krzywinski et al., 2009	<a href="http://circos.ca/software">http://circos.ca/software</a>
Bioconductor (version 3.6)	Huber et al., 2015	<a href="https://www.bioconductor.org">https://www.bioconductor.org</a>
gplots (version 3.0.1)	Gregory Warnes	<a href="https://CRAN.R-project.org/package=gplots">https://CRAN.R-project.org/package=gplots</a>

**Table 11 Public datasets used within this study.**

Cell type	Assay title	Description	GEO accession	Format used
K562	GRO-seq	GRO-seq for chronic myelogenous leukemic cell line	GSM1480325	Fastq
K562	ChIA-PET	ChIA-PET for CTCF	GSM970216	BED
K562	Dnase-seq	Open chromatin regions (ENCODE)	GSM646567	BAM
K562	Mnase-seq	Nucleosome positions by Mnase-seq (ENCODE)	GSM920557	BIGWIG
K562	Hi-C	By Rao et al. using JUICER and HiCCUPS	GSE63525	HiCCUPS output
K562	ChIP-seq	ChIP-seq for PolR2A (ENCODE)	GSM935358	BAM
K562	ChIP-seq	ChIP-seq for Pol2S5 (ENCODE)	GSM803443	BAM
K562	ChIP-seq	ChIP-seq for Pol2S2 (ENCODE)	GSM935645	BAM
K562	ChIP-seq	ChIP-seq for Smc3 (ENCODE)	GSM935310	BAM
K562	ChIP-seq	ChIP-seq for CTCF (ENCODE)	GSM935407	BAM
K562	ChIP-seq	ChIP-seq for Rad21 (ENCODE)	GSM935319	BAM
K562	ChIP-seq	ChIP-seq for H2A.Z (ENCODE)	GSM733786	BAM
K562	ChIP-seq	ChIP-seq for H3K4me1 (ENCODE)	GSM733692	BAM
K562	ChIP-seq	ChIP-seq for H3K4me2 (ENCODE)	GSM733651	BAM
K562	ChIP-seq	ChIP-seq for H3K4me3 (ENCODE)	GSM733680	BAM
K562	ChIP-seq	ChIP-seq for H3K27ac (ENCODE)	GSM733656	BAM
K562	ChIP-seq	ChIP-seq for H3K27me3 (ENCODE)	GSM733658	BAM
K562	ChIP-seq	ChIP-seq for H3K9ac (ENCODE)	GSM733778	BAM
K562	ChIP-seq	ChIP-seq for H3K9me1 (ENCODE)	GSM733777	BAM
K562	ChIP-seq	ChIP-seq for H3K9me3 (ENCODE)	GSM733776	BAM
K562	ChIP-seq	ChIP-seq for H3K36me3 (ENCODE)	GSM733714	BAM
K562	ChIP-seq	ChIP-seq for H3K79me2 (ENCODE)	GSM733653	BAM
K562	ChIP-seq	ChIP-seq for H4K20me1 (ENCODE)	GSM733675	BAM
GM12878	Dnase-seq	Open chromatin regions (ENCODE)	GSM816665	BAM
GM12878	Mnase-seq	Nucleosome positions by Mnase-seq (ENCODE)	GSM920558	BIGWIG
GM12878	Hi-C	By Rao et al. using JUICER and HiCCUPS	GSE63525	HiCCUPS output
GM12878	ChIP-seq	ChIP-seq for Smc3 (ENCODE)	GSM935376	BAM
GM12878	ChIP-seq	ChIP-seq for CTCF (ENCODE)	GSM935611	BAM
GM12878	ChIP-seq	ChIP-seq for POL2RA (ENCODE)	GSM935386	BIGWIG
GM12878	ChIP-seq	ChIP-seq for RAD21 (ENCODE)	GSM803416	BIGWIG
GM12878	ChIA-PET	ChIA-PET for RAD21	GSM1436265	BED
GM12878	ChIA-PET	ChIA-PET for CTCF	GSM1872886	BED
CD34+	RNA-seq	RNA-seq for CD34 Primary Cells	GSM909310	Fastq
CD34+	ChIP-seq	ChIP-seq for CTCF	GSM2861703	Fastq
CD34+	Hi-C	Hi-C for human hematopoietic stem cells (CD34+/CD38-)	GSM2861708	HIC
MCF10A	ChIP-seq	ChIP-seq for TOP2B	GSM2442946	Fastq

## References

- Albers, J., Danzer, C., Rechsteiner, M., Lehmann, H., Brandt, L.P., Hejhal, T., Catalano, A., et al. (2015) A versatile modular vector system for rapid combinatorial mammalian genetics. *Journal of Clinical Investigation*, Vol. 125 No. 4, pp. 1603–1619.
- Allan, J.M. and Travis, L.B. (2005) Mechanisms of therapy-related carcinogenesis. *Nature Reviews. Cancer*, Vol. 5 No. 12, pp. 943–55.
- Álvarez-Quilón, A., Serrano-Benítez, A., Lieberman, J.A., Quintero, C., Sánchez-Gutiérrez, D., Escudero, L.M. and Cortés-Ledesma, F. (2014) ATM specifically mediates repair of double-strand breaks with blocked DNA ends. *Nature Communications*, Vol. 5 No. 1, pp. 1–10.
- Anderson, D.J., Le Moigne, R., Djakovic, S., Kumar, B., Rice, J., Wong, S., Wang, J., et al. (2015) Targeting the AAA ATPase p97 as an Approach to Treat Cancer through Disruption of Protein Homeostasis. *Cancer Cell*, Vol. 28 No. 5, pp. 653–665.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., et al. (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, Vol. 507 No. 7493, pp. 455–461.
- Aparicio, T., Baer, R., Gottesman, M. and Gautier, J. (2016) MRN, CtIP, and BRCA1 mediate repair of topoisomerase II-DNA adducts. *Journal of Cell Biology*, Vol. 212 No. 4, pp. 399–408.
- Ashour, M.E., Atteya, R. and El-Khamisy, S.F. (2015) Topoisomerase-mediated chromosomal break repair: an emerging player in many games. *Nature Reviews Cancer*, Vol. 15 No. 3, pp. 137–151.
- 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*, Vol. 303 No. 5654, pp. 92–95.
- Austin, C.A., Lee, K.C., Swan, R.L., Khazeem, M.M., Manville, C.M., Cridland, P., Treumann, A., et al. (2018) TOP2B: The first thirty years. *International Journal of Molecular Sciences*, 14 September.
- Austin, C.A., Sng, J.H., Patel, S. and Fisher, L.M. (1993) Novel HeLa topoisomerase II is the II $\beta$  isoform: complete coding sequence and homology with other type II topoisomerases. *BBA - Gene Structure and Expression*, Vol. 1172 No. 3, pp. 283–291.
- Aymard, F., Aguirrebengoa, M., Guillou, E., Javierre, B.M., Bugler, B., Arnould, C., Rocher, V., et al. (2017) Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nature Structural and Molecular Biology*, Vol. 24 No. 4, pp. 353–361.
- Azarova, A.M., Lyu, Y.L., Lin, C.P., Tsai, Y.C., Lau, J.Y.N., Wang, J.C. and Liu, L.F. (2007) Roles of DNA topoisomerase II isozymes in chemotherapy and secondary malignancies. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104 No. 26, pp. 11014–11019.
- Balakirev, M.Y., Mullally, J.E., Favier, A., Assard, N., Sulpice, E., Lindsey, D.F., Rulina, A. V, et al. (2015) Wss1 metalloprotease partners with Cdc48/Doa1 in processing genotoxic SUMO conjugates. *ELife*, Vol. 4 No. September2015.
- Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang, Y., et al. (2012) Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 109 No. 52, pp. 21301–21306.

- Bensaude, O. (2011) Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription*, Vol. 2 No. 3, pp. 103–108.
- Berti, M., Chaudhuri, A.R., Thangavel, S., Gomathinayagam, S., Kenig, S., Vujanovic, M., Odreman, F., et al. (2013) Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. *Nature Structural and Molecular Biology*, Vol. 20 No. 3, pp. 347–354.
- Bhatia, S. (2013) Therapy-related myelodysplasia and acute myeloid leukemia. *Seminars in Oncology*, Vol. 40 No. 6, pp. 666–675.
- Biehs, R., Steinlage, M., Barton, O., Juhász, S., Künzel, J., Spies, J., Shibata, A., et al. (2017) DNA Double-Strand Break Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during Homologous Recombination. *Molecular Cell*, Vol. 65 No. 4, pp. 671-684.e5.
- Bintu, B., Mateo, L.J., Su, J.H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., et al. (2018) Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, Vol. 362 No. 6413.
- Bonev, B. and Cavalli, G. (2016) Organization and function of the 3D genome. *Nature Reviews Genetics*, 14 October.
- van den Boom, J., Wolf, M., Weimann, L., Schulze, N., Li, F., Kaschani, F., Riemer, A., et al. (2016) VCP/p97 Extracts Sterically Trapped Ku70/80 Rings from DNA in Double-Strand Break Repair. *Molecular Cell*, Vol. 64 No. 1, pp. 189–198.
- Borkin, D., He, S., Miao, H., Kempinska, K., Pollock, J., Chase, J., Purohit, T., et al. (2015) Pharmacologic inhibition of the menin-MLL interaction blocks progression of MLL leukemia in vivo. *Cancer Cell*, Vol. 27 No. 4, pp. 589–602.
- Boutros, M. and Ahringer, J. (2008) The art and design of genetic screens: RNA interference. *Nature Reviews Genetics*, July.
- Bouwman, B.A.M., Agostini, F., Garnerone, S., Petrosino, G., Gothe, H.J., Sayols, S., Moor, A., et al. (2020) Genome-wide detection of DNA double-strand breaks by in-suspension BLISS. *Nature Protocols*, No. in revision.
- Bouwman, B.A.M. and Crosetto, N. (2018) Endogenous DNA double-strand breaks during DNA transactions: Emerging insights and methods for genome-wide profiling. *Genes*, 1 December.
- Boveri, T. (1914) Zur Frage der Entstehung maligner Tumoren.
- Brunet, E., Simsek, D., Tomishima, M., DeKever, R., Choi, V.M., Gregory, P., Urnov, F., et al. (2009) Chromosomal translocations induced at specified loci in human stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 106 No. 26, pp. 10620–10625.
- Burgess, D.J., Doles, J., Zender, L., Xue, W., Ma, B., McCombie, W.R., Hannon, G.J., et al. (2008) Topoisomerase levels determine chemotherapy response in vitro and in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105 No. 26, pp. 9053–9058.
- Burman, B., Misteli, T. and Pegoraro, G. (2015) Quantitative detection of rare interphase chromosome breaks and translocations by high-throughput imaging. *Genome Biology*, Vol. 16 No. 1.
- Busslinger, G.A., Stocsits, R.R., Van Der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N. and Peters, J.M. (2017) Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*, Vol. 544 No. 7651, pp. 503–507.
- de Campos-Nebel, M., Larripa, I. and González-Cid, M. (2010) Topoisomerase II-Mediated DNA Damage Is Differently Repaired during the Cell Cycle by Non-Homologous End Joining and Homologous Recombination. edited by Harris, R.S. *PLoS ONE*, Vol. 5 No. 9, p. e12541.



- Canela, A., Maman, Y., Huang, S., Yin N., Wutz, G., Tang, W., Zagnoli-Vieira, G., Callen, E., et al. (2019) Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. *Molecular Cell*, Vol. 75 No. 2, pp. 252-266.e8.
- Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., Kieffer-Kwon, K.-R., et al. (2017) Genome Organization Drives Chromosome Fragility. *Cell*, Vol. 170 No. 3, pp. 507-521.e18.
- Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P. and Nussenzweig, A. (2016) DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Molecular Cell*, Vol. 63 No. 5, pp. 898-911.
- Caron, P., Choudhary, J., Clouaire, T., Bugler, B., Daburon, V., Aguirrebengoa, M., Mangeat, T., et al. (2015) Non-redundant Functions of ATM and DNA-PKcs in Response to DNA Double-Strand Breaks. *Cell Reports*, Vol. 13 No. 8, pp. 1598-1609.
- Carpenter, A.J. and Porter, A.C.G. (2004) Construction, Characterization, and Complementation of a Conditional-Lethal DNA Topoisomerase II Mutant Human Cell Line. *Molecular Biology of the Cell*, Vol. 15 No. 12, pp. 5700-5711.
- Ceballos, S.J. and Heyer, W.D. (2011) Functions of the Snf2/Swi2 family Rad54 motor protein in homologous recombination. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1 September.
- Cheung, M.S., Down, T.A., Latorre, I. and Ahringer, J. (2011) Systematic bias in high-throughput sequencing data and its correction by BEADS. *Nucleic Acids Research*, Vol. 39 No. 15.
- Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., et al. (2011) Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell*, Vol. 147 No. 1, pp. 107-119.
- Ciccia, A. and Elledge, S.J. (2010) The DNA Damage Response: Making It Safe to Play with Knives. *Molecular Cell*, October.
- Cohen, S., Puget, N., Lin, Y.L., Clouaire, T., Aguirrebengoa, M., Rocher, V., Pasero, P., et al. (2018) Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nature Communications*, Vol. 9 No. 1.
- Colquhoun, D. (2014) An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science*, Vol. 1 No. 3.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., et al. (2016) A survey of best practices for RNA-seq data analysis. *Genome Biology*, 26 January.
- Cowell, I.G. and Austin, C.A. (2012) Mechanism of generation of therapy related leukemia in response to anti-topoisomerase II agents. *International Journal of Environmental Research and Public Health*, June.
- Cowell, I.G., Sondka, Z., Smith, K., Lee, K.C., Manville, C.M., Sidorczuk-Lesthuruge, M., Rance, H.A., et al. (2012) Model for MLL translocations in therapy-related leukemia involving topoisomerase II $\beta$ -mediated DNA strand breaks and gene proximity. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 109 No. 23, pp. 8989-94.
- Cristini, A., Groh, M., Kristiansen, M.S. and Gromak, N. (2018) RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Reports*, Vol. 23 No. 6, pp. 1891-1905.
- Croce, C. and Nowell, P. (1985) Molecular basis of human B cell neoplasia. *Blood*, Vol. 65 No. 1, pp. 1-7.
- Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., et al. (2013)

## References

- Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature Methods*, Vol. 10 No. 4, pp. 361–365.
- Cuya, S.M., Bjornsti, M.A. and van Waardenburg, R.C.A.M. (2017) DNA topoisomerase-targeting chemotherapeutics: what's new? *Cancer Chemotherapy and Pharmacology*, 1 July.
- D'Arpa, P. (1994) Determinants of Cellular Sensitivity to Topoisomerase-Targeting Antitumor Drugs. *Advances in Pharmacology*, Vol. 29 No. PB, pp. 127–143.
- Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G. and Peters, J.M. (2019) DNA loop extrusion by human cohesin. *Science*, Vol. 366 No. 6471, pp. 1338–1345.
- Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., et al. (2018) The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Research*, Vol. 46 No. D1, pp. D794–D801.
- Deshpande, R.A., Lee, J.H., Arora, S. and Paull, T.T. (2016) Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. *Molecular Cell*, Vol. 64 No. 3, pp. 593–606.
- Deweese, J.E., Osheroff, M.A. and Osheroff, N. (2009) DNA topology and topoisomerases: Teaching A “knotty” subject. *Biochemistry and Molecular Biology Education*, January.
- Deweese, J.E. and Osheroff, N. (2009) The DNA cleavage reaction of topoisomerase II: Wolf in sheep's clothing. *Nucleic Acids Research*, Vol. 37 No. 3, pp. 738–748.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, Vol. 485 No. 7398, pp. 376–380.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, Vol. 29 No. 1, pp. 15–21.
- Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., et al. (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell*, Vol. 121 No. 6, pp. 873–885.
- Drynan, L.F., Pannell, R., Forster, A., Chan, N.M.M., Cano, F., Daser, A. and Rabbitts, T.H. (2005) Mll fusions generated by Cre-loxP-mediated de novo translocations can induce lineage reassignment in tumorigenesis. *EMBO Journal*, Vol. 24 No. 17, pp. 3136–3146.
- Dykhuisen, E.C., Hargreaves, D.C., Miller, E.L., Cui, K., Korshunov, A., Kool, M., Pfister, S., et al. (2013) BAF complexes facilitate decatenation of DNA by topoisomerase II $\alpha$ . *Nature*, Vol. 497 No. 7451, pp. 624–627.
- Ezoe, S. (2012) Secondary leukemia associated with the anti-cancer agent, etoposide, a topoisomerase II inhibitor. *International Journal of Environmental Research and Public Health*, Vol. 9 No. 7, pp. 2444–2453.
- Fan, J.R., Peng, A.L., Chen, H.C., Lo, S.C., Huang, T.H. and Li, T.K. (2008) Cellular processing pathways contribute to the activation of etoposide-induced DNA damage responses. *DNA Repair*, Vol. 7 No. 3, pp. 452–463.
- Farr, C.J., Antoniou-Kourouniotti, M., Mimmack, M.L., Volkov, A. and Porter, A.C.G. (2014) The  $\alpha$  isoform of topoisomerase II is required for hypercompaction of mitotic chromosomes in human cells. *Nucleic Acids Research*, Vol. 42 No. 7, pp. 4414–4426.
- Fielden, J., Wiseman, K., Torrecilla, I., Li, S., Hume, S., Chiang, S.C., Ruggiano, A., et al. (2020) TEX264 coordinates p97- and SPRTN-mediated resolution of topoisomerase 1-DNA adducts. *Nature Communications*, Vol. 11 No. 1, pp. 1–16.
- Finn, E.H., Pegoraro, G., Shachar, S. and Misteli, T. (2017) Comparative analysis of 2D and 3D distance measurements to study spatial genome organization. *Methods*, Vol. 123, pp. 47–55.

- Fisher, R.P. (2005) Secrets of a double agent: CDK7 in cell-cycle control and transcription. *Journal of Cell Science*, Vol. 118 No. Pt 22, pp. 5171–80.
- Forterre, P. and Gadelle, D. (2009) Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Research*, Vol. 37 No. 3, pp. 679–692.
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N. and Mirny, L.A. (2016) Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, Vol. 15 No. 9, pp. 2038–2049.
- Gao, Q., Liang, W.W., Foltz, S.M., Mutharasu, G., Jayasinghe, R.G., Cao, S., Liao, W.W., et al. (2018) Driver Fusions and Their Implications in the Development and Treatment of Human Cancers. *Cell Reports*, Vol. 23 No. 1, pp. 227–238.e3.
- Gao, R., Schellenberg, M.J., Huang, S.Y.N., Abdelmalak, M., Marchand, C., Nitiss, K.C., Nitiss, J.L., et al. (2014) Proteolytic degradation of topoisomerase II (Top2) enables the processing of Top2·DNA and Top2·RNA covalent complexes by tyrosyl-DNA-phosphodiesterase 2 (TDP2). *Journal of Biological Chemistry*, Vol. 289 No. 26, pp. 17960–17969.
- Gao, T., He, B., Liu, S., Zhu, H., Tan, K. and Qian, J. (2016) EnhancerAtlas: A resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics*, Vol. 32 No. 23, pp. 3543–3551.
- Gelali, E., Girelli, G., Matsumoto, M., Wernersson, E., Custodio, J., Mota, A., Schweitzer, M., et al. (2019) iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. *Nature Communications*, Vol. 10 No. 1, pp. 1–15.
- Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., et al. (2014) Chromosomal Translocations in Human Cells Are Generated by Canonical Nonhomologous End-Joining. *Molecular Cell*, Vol. 55 No. 6, pp. 829–842.
- Gole, B. and Wiesmüller, L. (2015) Leukemogenic rearrangements at the mixed lineage leukemia gene (MLL)-multiple rather than a single mechanism. *Frontiers in Cell and Developmental Biology*, 25 June.
- Gómez-Herrerros, F. (2019) DNA Double Strand Breaks and Chromosomal Translocations Induced by DNA Topoisomerase II. *Frontiers in Molecular Biosciences*, 10 December.
- Gómez-Herrerros, F., Romero-Granados, R., Zeng, Z., Álvarez-Quilón, A., Quintero, C., Ju, L., Umans, L., et al. (2013) TDP2-Dependent Non-Homologous End-Joining Protects against Topoisomerase II-Induced DNA Breaks and Genome Instability in Cells and In Vivo. edited by Maizels, *N.PLoS Genetics*, Vol. 9 No. 3, p. e1003226.
- Gómez-Herrerros, F., Zagnoli-Vieira, G., Ntai, I., Martínez-Macías, M.I., Anderson, R.M., Herrero-Ruíz, A. and Caldecott, K.W. (2017) TDP2 suppresses chromosomal translocations induced by DNA topoisomerase II during gene transcription. *Nature Communications*, Vol. 8 No. 1.
- Gothe, H.J., Bouwman, B.A.M., Gusmao, E.G., Piccinno, R., Petrosino, G., Sayols, S., Drechsel, O., et al. (2019) Spatial Chromosome Folding and Active Transcription Drive DNA Fragility and Formation of Oncogenic MLL Translocations. *Molecular Cell*, Vol. 75 No. 2, pp. 267–283.e12.
- Guénolé, A. and Legube, G. (2017) A meeting at risk: Unrepaired DSBs go for broke. *Nucleus*, Vol. 8 No. 6, pp. 589–599.
- Gusmao, E.G., Allhoff, M., Zenke, M. and Costa, I.G. (2016) Analysis of computational footprinting methods for DNase sequencing experiments. *Nature Methods*, Vol. 13 No. 4, pp. 303–309.
- Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.R., Jankovic, M., Oliveira, T., et al. (2012) DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature*, Vol. 484 No. 7392, pp. 69–74.

## References

- Halazonetis, T.D., Gorgoulis, V.G. and Bartek, J. (2008) An oncogene-induced DNA damage model for cancer development. *Science*, 7 March.
- Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: The next generation. *Cell*, 4 March.
- Heckl, D., Kowalczyk, M.S., Yudovich, D., Belizaire, R., Puram, R. V., McConkey, M.E., Thielke, A., et al. (2014) Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nature Biotechnology*, Vol. 32 No. 9, pp. 941–946.
- Heger, A., Webber, C., Goodson, M., Ponting, C.P. and Lunter, G. (2013) GAT: A simulation framework for testing the association of genomic intervals. *Bioinformatics*, Vol. 29 No. 16, pp. 2046–2048.
- Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., et al. (2018) Transcription Elongation Can Affect Genome 3D Structure. *Cell*, Vol. 174 No. 6, pp. 1522–1536.e22.
- Ho, N.N., Akagawa, R., Yamasaki, T., Hirota, K., Sasa, K., Natsume, T., Kobayashi, J., et al. (2015) Relative contribution of four nucleases, CtIP, Dna2, Exo1 and Mre11, to the initial step of DNA double-strand break repair by homologous recombination in both the chicken DT40 and human TK6 cell lines. *Genes to Cells*, Vol. 20 No. 12, pp. 1059–1076.
- Ho, N.N., Shimizu, T., Zhou, Z.W., Wang, Z.-Q., Deshpande, R.A., Paull, T.T., Akter, S., et al. (2016) Mre11 Is Essential for the Removal of Lethal Topoisomerase 2 Covalent Cleavage Complexes. *Molecular Cell*, Vol. 64 No. 3, pp. 580–592.
- Hoeijmakers, J.H.J. (2009) DNA damage, aging, and cancer. *New England Journal of Medicine*, 8 October.
- Hoffman, E.A., McCulley, A., Haarer, B., Arnak, R. and Feng, W. (2015) Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome Research*, Vol. 25 No. 3, pp. 402–412.
- Hornyak, P., Askwith, T., Walker, S., Komulainen, E., Paradowski, M., Pennicott, L.E., Bartlett, E.J., et al. (2016) Mode of action of DNA-competitive small molecule inhibitors of tyrosyl DNA phosphodiesterase 2. *Biochemical Journal*, Vol. 473 No. 13, pp. 1869–1879.
- Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W. and Frock, R.L. (2016) Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nature Protocols*, Vol. 11 No. 5, pp. 853–71.
- Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D. and Legube, G. (2010) High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *The EMBO Journal*, Vol. 29 No. 8, pp. 1446–1457.
- Jackson, S.P. and Bartek, J. (2009) The DNA-damage response in human biology and disease. *Nature*, 22 October.
- Jha, S. and Dutta, A. (2009) RVB1/RVB2: Running Rings around Molecular Biology. *Molecular Cell*, 12 June.
- Ju, B.-G., Lunyak, V. V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K. and Rosenfeld, M.G. (2006) A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science*, Vol. 312 No. 5781, pp. 1798–802.
- Kaiser, V.B. and Semple, C.A. (2018) Chromatin loop anchors are associated with genome instability in cancer and recombination hotspots in the germline. *Genome Biology*, Vol. 19 No. 1.
- Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., Van Der Lee, R., Bessy, A., et al. (2018) JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research*, Vol. 46 No. D1, pp.

- D260–D266.
- Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D.D., Zhang, M.Q., et al. (2007) Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell*, Vol. 128 No. 6, pp. 1231–1245.
- Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J. and Yu, H. (2019) Human cohesin compacts DNA by loop extrusion. *Science*, Vol. 366 No. 6471, pp. 1345–1349.
- Kinner, A., Wu, W., Staudt, C. and Iliakis, G. (2008)  $\gamma$ -H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Research*, Vol. 36 No. 17, pp. 5678–5694.
- Korhonen, J.H., Palin, K., Taipale, J. and Ukkonen, E. (2017) Fast motif matching revisited: High-order PWMs, SNPs and indels. *Bioinformatics*, Vol. 33 No. 4, pp. 514–521.
- Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., et al. (2013) Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nature Structural and Molecular Biology*, Vol. 20 No. 3, pp. 396–403.
- Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K.R., Bandle, R., Nelson, S., et al. (2013) Global regulation of promoter melting in naive lymphocytes. *Cell*, Vol. 153 No. 5, p. 988.
- Kozuki, T., Chikamori, K., Surleac, M.D., Micluta, M.A., Petrescu, A.J., Norris, E.J., Elson, P., et al. (2017) Roles of the C-terminal domains of topoisomerase II $\alpha$ ' and topoisomerase II $\beta$  in regulation of the decatenation checkpoint. *Nucleic Acids Research*, Vol. 45 No. 10, pp. 5995–6010.
- Krivtsov, A. V. and Armstrong, S.A. (2007) MLL translocations, histone modifications and leukaemia stem-cell development. *Nature Reviews Cancer*, Vol. 7 No. 11, pp. 823–833.
- Kubo, N., Ishii, H., Gorkin, D., Meitinger, F., Xiong, X., Fang, R., Liu, T., et al. (2017) Preservation of Chromatin Organization after Acute Loss of CTCF in Mouse Embryonic Stem Cells. *BioRxiv*, p. 118737.
- Kumar, S. (2012) Second malignant neoplasms following radiotherapy. *International Journal of Environmental Research and Public Health*, December.
- Lang, A.J., Mirski, S.E.L., Cummings, H.J., Yu, Q., Gerlach, J.H. and Cole, S.P.C. (1998) Structural organization of the human TOP2A and TOP2B genes. *Gene*, Vol. 221 No. 2, pp. 255–266.
- Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, Vol. 10 No. 3.
- Lavin, M.F. (2007) ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. *Oncogene*, 10 December.
- Lavin, M.F., Kozlov, S., Gatei, M. and Kijas, A.W. (2015) ATM-dependent phosphorylation of all three members of the MRN complex: From sensor to adaptor. *Biomolecules*.
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., et al. (2013) Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, Vol. 9 No. 8.
- Leder, P., Battey, J., Lenoir, G., Moulding, C., Murphy, W., Potter, H., Stewart, T., et al. (1983) Translocations among antibody genes in human cancer. *Science*, Vol. 222 No. 4625, pp. 765–771.
- Ledesma, F.C., El Khamisy, S.F., Zuma, M.C., Osborn, K. and Caldecott, K.W. (2009) A human 5'-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature*, Vol. 461 No. 7264, pp. 674–678.
- Lee, K.C., Bramley, R.L., Cowell, I.G., Jackson, G.H. and Austin, C.A. (2016) Proteasomal inhibition potentiates drugs targeting DNA topoisomerase II. *Biochemical Pharmacology*, Vol. 103, pp. 29–39.

## References

- Lee, T. and Pelletier, J. (2016) The biology of DHX9 and its potential as a therapeutic target. *Oncotarget*.
- Lensing, S. V., Marsico, G., Hänsel-Hertsch, R., Lam, E.Y., Tannahill, D. and Balasubramanian, S. (2016) DSBCapture: In situ capture and sequencing of DNA breaks. *Nature Methods*, Vol. 13 No. 10, pp. 855–857.
- Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, Vol. 25 No. 14, pp. 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, Vol. 25 No. 16, pp. 2078–2079.
- Li, Y., Roberts, N.D., Wala, J.A., Shapira, O., Schumacher, S.E., Kumar, K., Khurana, E., et al. (2020) Patterns of somatic structural variation in human cancer genomes. *Nature*, Vol. 578 No. 7793, pp. 112–121.
- Liao, Y., Smyth, G.K. and Shi, W. (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, Vol. 30 No. 7, pp. 923–930.
- Lieber, M.R. (2010) The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry*, Vol. 79 No. 1, pp. 181–211.
- Lin, S., Luo, R.T., Shrestha, M., Thirman, M.J. and Mulloy, J.C. (2017) The full transforming capacity of MLL-Af4 is interlinked with lymphoid lineage commitment. *Blood*, Vol. 130 No. 7, pp. 903–907.
- Liu, S., Opiyo, S.O., Manthey, K., Glanzer, J.G., Ashley, A.K., Amerin, C., Troksa, K., et al. (2012) Distinct roles for DNA-PK, ATM and ATR in RPA phosphorylation and checkpoint activation in response to replication stress. *Nucleic Acids Research*, Vol. 40 No. 21, pp. 10780–10794.
- Lomax, M.E., Folkes, L.K. and O'Neill, P. (2013) Biological consequences of radiation-induced DNA damage: Relevance to radiotherapy. *Clinical Oncology*, Vol. 25 No. 10, pp. 578–585.
- Lopez-Mosqueda, J., Maddi, K., Prgomet, S., Kalayil, S., Marinovic-Terzic, I., Terzic, J. and Dikic, I. (2016) SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. *ELife*, Vol. 5 No. NOVEMBER2016.
- Lord, C.J., Tutt, A.N.J. and Ashworth, A. (2015) Synthetic Lethality and Cancer Therapy: Lessons Learned from the Development of PARP Inhibitors. *Annual Review of Medicine*, Vol. 66 No. 1, pp. 455–470.
- Lu, B., Klingbeil, O., Tarumoto, Y., Somerville, T.D.D., Huang, Y.H., Wei, Y., Wai, D.C., et al. (2018) A Transcription Factor Addiction in Leukemia Imposed by the MLL Promoter Sequence. *Cancer Cell*, Vol. 34 No. 6, pp. 970–981.e8.
- Lund, H.L., Hughesman, C.B., Fakhfakh, K., McNeil, K., Clemens, S., Hocken, K., Pettersson, R., et al. (2016) Initial Diagnosis of ALK-Positive Non-Small-Cell Lung Cancer Based on Analysis of ALK Status Utilizing Droplet Digital PCR. *Analytical Chemistry*, Vol. 88 No. 9, pp. 4879–4885.
- Lyu, Y.L., Lin, C.-P., Azarova, A.M., Cai, L., Wang, J.C. and Liu, L.F. (2006) Role of Topoisomerase II in the Expression of Developmentally Regulated Genes. *Molecular and Cellular Biology*, Vol. 26 No. 21, pp. 7929–7941.
- Madabhushi, R., Gao, F., Pfenning, A.R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., et al. (2015) Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell*, Vol. 161 No. 7, pp. 1592–1605.
- Madsen, J.G.S., Schmidt, S.F., Larsen, B.D., Loft, A., Nielsen, R. and Mandrup, S. (2015) iRNA-seq: Computational method for genome-wide assessment of acute transcriptional

- regulation from total RNA-seq data. *Nucleic Acids Research*, Vol. 43 No. 6, p. e40.
- Magnaghi, P., D'Alessio, R., Valsasina, B., Avanzi, N., Rizzi, S., Asa, D., Gasparri, F., et al. (2013) Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. *Nature Chemical Biology*, Vol. 9 No. 9, pp. 548–559.
- Mani, R.S. and Chinnaiyan, A.M. (2010) Triggers for genomic rearrangements: Insights into genomic, cellular and environmental influences. *Nature Reviews Genetics*, December.
- Manville, C.M., Smith, K., Sondka, Z., Rance, H., Cockell, S., Cowell, I.G., Lee, K.C., et al. (2015) Genome-wide ChIP-seq analysis of human TOP2B occupancy in MCF7 breast cancer epithelial cells. *Biology Open*, Vol. 4 No. 11, pp. 1436–1447.
- Mao, Y., Desai, S.D., Ting, C.Y., Hwang, J. and Liu, L.F. (2001) 26 S proteasome-mediated degradation of topoisomerase II cleavable complexes. *The Journal of Biological Chemistry*, Vol. 276 No. 44, pp. 40652–8.
- Maslov, A.Y., Lee, M., Gundry, M., Gravina, S., Strogonova, N., Tazearslan, C., Bendebury, A., et al. (2012) 5-Aza-2'-deoxycytidine-induced genome rearrangements are mediated by DNMT1. *Oncogene*, Vol. 31 No. 50, pp. 5172–5179.
- Matkar, S., Thiel, A. and Hua, X. (2013) Menin: A scaffold protein that controls gene expression and cell signaling. *Trends in Biochemical Sciences*, August.
- McNerney, M.E., Godley, L.A. and Le Beau, M.M. (2017) Therapy-related myeloid neoplasms: When genetics and environment collide. *Nature Reviews Cancer*, 1 September.
- McVey, M. and Lee, S.E. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in Genetics*, November.
- Mertens, F., Johansson, B., Fioretos, T. and Mitelman, F. (2015) The emerging complexity of gene fusions in cancer. *Nature Reviews Cancer*, 22 May.
- Meyer, C., Burmeister, T., Gröger, D., Tsauro, G., Fehina, L., Renneville, A., Sutton, R., et al. (2018) The MLL recombinome of acute leukemias in 2017. *Leukemia*, Vol. 32 No. 2, pp. 273–284.
- Meyer, C., Hofmann, J., Burmeister, T., Gröger, D., Park, T.S., Emerenciano, M., Pombo de Oliveira, M., et al. (2013) The MLL recombinome of acute leukemias in 2013. *Leukemia*, Vol. 27 No. 11, pp. 2165–76.
- Meyer, C.A. and Liu, X.S. (2014) Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nature Reviews Genetics*, 25 November.
- Meyer, H., Bug, M. and Bremer, S. (2012) Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nature Cell Biology*, February.
- Miller, K.D., Nogueira, L., Mariotto, A.B., Rowland, J.H., Yabroff, K.R., Alfano, C.M., Jemal, A., et al. (2019) Cancer treatment and survivorship statistics, 2019. *CA: A Cancer Journal for Clinicians*, Vol. 69 No. 5, pp. 363–385.
- Milne, T.A., Hughes, C.M., Lloyd, R., Yang, Z., Rozenblatt-Rosen, O., Dou, Y., Schnepf, R.W., et al. (2005) Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102 No. 3, pp. 749–754.
- Mitelman, F., Johansson, B. and Mertens, F. (2007) The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*, April.
- Mitelman, F., Johansson, B. and Mertens, F. (n.d.). Mitelman Database Chromosome Aberrations and Gene Fusions in Cancer. *Mitelman Database Chromosome Aberrations and Gene Fusions in Cancer*.
- Mohr, S., Bakal, C. and Perrimon, N. (2010) Genomic Screening with RNAi: Results and Challenges. *Annual Review of Biochemistry*, Vol. 79 No. 1, pp. 37–64.
- Mórocz, M., Zsigmond, E., Tóth, R., Enyedi, M.Z., Pintér, L. and Haracska, L. (2017) DNA-

## References

- dependent protease activity of human Spartan facilitates replication of DNA-protein crosslink-containing DNA. *Nucleic Acids Research*, Vol. 45 No. 6, pp. 3172–3188.
- Muntean, A.G. and Hess, J.L. (2012) The Pathogenesis of Mixed-Lineage Leukemia. *Annual Review of Pathology: Mechanisms of Disease*, Vol. 7 No. 1, pp. 283–301.
- Nagashima, R., Hibino, K., Ashwin, S.S., Babokhov, M., Fujishiro, S., Imai, R., Nozaki, T., et al. (2019) Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II. *Journal of Cell Biology*, Vol. 218 No. 5, pp. 1511–1530.
- Natsume, T., Kiyomitsu, T., Saga, Y. and Kanemaki, M.T. (2016) Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Reports*, Vol. 15 No. 1, pp. 210–218.
- Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., et al. (2013) Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nature Structural and Molecular Biology*, Vol. 20 No. 3, pp. 387–395.
- Nie, M., Aslanian, A., Prudden, J., Heideker, J., Vashisht, A.A., Wohlschlegel, J.A., Yates, J.R., et al. (2012) Dual recruitment of Cdc48 (p97)-Ufd1-Npl4 ubiquitin-selective segregase by small ubiquitin-like modifier protein (SUMO) and ubiquitin in SUMO-targeted ubiquitin ligase-mediated genome stability functions. *Journal of Biological Chemistry*, Vol. 287 No. 35, pp. 29610–29619.
- Nitiss, J.L. (2009) DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews Cancer*, May.
- Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., et al. (2017) Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, Vol. 169 No. 5, pp. 930–944.e22.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., et al. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, Vol. 485 No. 7398, pp. 381–385.
- North, A.J. (2006) Seeing is believing? A beginners' guide to practical pitfalls in image acquisition. *Journal of Cell Biology*, 1 February.
- Nowell, P.C. and Hungerford, D.A. (1960) Chromosome studies on normal and leukemic human leukocytes. *Journal of the National Cancer Institute*, Vol. 25 No. 1, pp. 85–109.
- O'Connor, M.J. (2015) Targeting the DNA Damage Response in Cancer. *Molecular Cell*, 19 November.
- Osborne, C.S., Chakalova, L., Mitchell, J.A., Horton, A., Wood, A.L., Bolland, D.J., Corcoran, A.E., et al. (2007) Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biology*, Vol. 5 No. 8, pp. 1763–1772.
- Paull, T.T. (2018) 20 Years of Mre11 Biology: No End in Sight. *Molecular Cell*, 2 August.
- Pendleton, M., Lindsey, R.H., Felix, C.A., Grimwade, D. and Osheroff, N. (2014) Topoisomerase II and leukemia. *Annals of the New York Academy of Sciences*, Vol. 1310, pp. 98–110.
- Piganeau, M., Ghezraoui, H., De Cian, A., Guittat, L., Tomishima, M., Perrouault, L., René, O., et al. (2013) Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome Research*, Vol. 23 No. 7, pp. 1182–1193.
- Pommier, Y. (2013) Drugging topoisomerases: Lessons and Challenges. *ACS Chemical Biology*, 18 January.
- Pommier, Y., Sun, Y., Huang, S.-Y.N. and Nitiss, J.L. (2016) Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nature Reviews Molecular Cell Biology*.
- Prakash, R., Zhang, Y., Feng, W. and Jasin, M. (2015) Homologous recombination and human health: The roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harbor*



- Perspectives in Biology*, Vol. 7 No. 4.
- Prioleau, M.-N. and MacAlpine, D.M. (2016) DNA replication origins-where do we begin? *Genes & Development*, Vol. 30 No. 15, pp. 1683–97.
- Quennet, V., Beucher, A., Barton, O., Takeda, S. and Löbrich, M. (2010) CtIP and MRN promote non-homologous end-joining of etoposide-induced DNA double-strand breaks in G1. *Nucleic Acids Research*, Vol. 39 No. 6, pp. 2144–2152.
- Quinlan, A.R. (2014) BEDTools: The Swiss-Army tool for genome feature analysis. *Current Protocols in Bioinformatics*, Vol. 2014, pp. 11.12.1-11.12.34.
- Racko, D., Benedetti, F., Dorier, J. and Stasiak, A. (2018) Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes. *Nucleic Acids Research*, Vol. 46 No. 4, pp. 1648–1660.
- Ramadan, K., Halder, S., Wiseman, K. and Vaz, B. (2017) Strategic role of the ubiquitin-dependent segregase p97 (VCP or Cdc48) in DNA replication. *Chromosoma*, 1 February.
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., et al. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, Vol. 44 No. W1, pp. W160–W165.
- Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., et al. (2017) Cohesin Loss Eliminates All Loop Domains. *Cell*, Vol. 171 No. 2, pp. 305-320.e24.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, Vol. 159 No. 7, pp. 1665–1680.
- Rao, T., Gao, R., Takada, S., Al Abo, M., Chen, X., Walters, K.J., Pommier, Y., et al. (2016) Novel TDP2-ubiquitin interactions and their importance for the repair of topoisomerase II-mediated DNA damage. *Nucleic Acids Research*, Vol. 44 No. 21, pp. 10201–10215.
- Raouf, A., Depledge, P., Hamilton, N.M., Hamilton, N.S., Hitchin, J.R., Hopkins, G. V., Jordan, A.M., et al. (2013) Toxoflavins and deazaflavins as the first reported selective small molecule inhibitors of tyrosyl-DNA phosphodiesterase II. *Journal of Medicinal Chemistry*, Vol. 56 No. 16, pp. 6352–6370.
- Roukos, V., Burman, B. and Misteli, T. (2013) The cellular etiology of chromosome translocations. *Current Opinion in Cell Biology*, June.
- Roukos, V. and Misteli, T. (2014) The biogenesis of chromosome translocations. *Nature Cell Biology*, Vol. 16 No. 4, pp. 293–300.
- Roukos, V., Pegoraro, G., Voss, T.C. and Misteli, T. (2015) Cell cycle staging of individual cells by fluorescence microscopy. *Nature Protocols*, Vol. 10 No. 2, pp. 334–348.
- 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*, Vol. 341 No. 6146, pp. 660–664.
- Rowley, J.D. (1973) A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*, Vol. 243 No. 5405, pp. 290–293.
- Rowley, M.J. and Corces, V.G. (2018) Organizational principles of 3D genome architecture. *Nature Reviews Genetics*, Vol. 19 No. 12, pp. 789–800.
- Sanborn, A.L., Rao, S.S.P., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., et al. (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 112 No. 47, pp. E6456–E6465.
- Sanjana, N.E., Shalem, O. and Zhang, F. (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods*.

## References

- Schellenberg, M.J., Lieberman, J.A., Herrero-Ruiz, A., Butler, L.R., Williams, J.G., Muñoz-Cabello, A.M., Mueller, G.A., et al. (2017) ZATT (ZNF451)-mediated resolution of topoisomerase 2 DNA-protein cross-links. *Science*, Vol. 357 No. 6358, pp. 1412–1416.
- Sciascia, N., Wu, W., Zong, D., Sun, Y., Wong, N., John, S., Wangsa, D., et al. (2020) Suppressing proteasome mediated processing of Topoisomerase II DNA-protein complexes preserves genome integrity. *ELife*, Vol. 9.
- Scully, R., Panday, A., Elango, R. and Willis, N.A. (2019) DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews Molecular Cell Biology*, 1 November.
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, Vol. 343 No. 6166, pp. 84–87.
- Shintomi, K., Takahashi, T.S. and Hirano, T. (2015) Reconstitution of mitotic chromatids with a minimum set of purified factors. *Nature Cell Biology*, Vol. 17 No. 8, pp. 1014–1023.
- Sng, J.H., Heaton, V.J., Bell, M., Maini, P., Austin, C.A. and Fisher, L.M. (1999) Molecular cloning and characterization of the human topoisomerase II $\alpha$  and II $\beta$  genes: Evidence for isoform evolution through gene duplication. *Biochimica et Biophysica Acta - Gene Structure and Expression*, Vol. 1444 No. 3, pp. 395–406.
- Stinglee, J., Bellelli, R., Alte, F., Hewitt, G., Sarek, G., Maslen, S.L., Tsutakawa, S.E., et al. (2016) Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Molecular Cell*, Vol. 64 No. 4, pp. 688–703.
- Stinglee, J., Bellelli, R. and Boulton, S.J. (2017) Mechanisms of DNA-protein crosslink repair. *Nature Reviews Molecular Cell Biology*, 1 September.
- Strissel, P.L., Strick, R., Rowley, J.D. and Zeleznik-Le, N.J. (1998) An in vivo topoisomerase II cleavage site and a DNase I hypersensitive site colocalize near exon 9 in the MLL breakpoint cluster region. *Blood*, Vol. 92 No. 10, pp. 3793–803.
- Sun, B., Ross, S.M., Rowley, S., Adeleye, Y. and Clewell, R.A. (2017) Contribution of ATM and ATR kinase pathways to p53-mediated response in etoposide and methyl methanesulfonate induced DNA damage. *Environmental and Molecular Mutagenesis*, Vol. 58 No. 2, pp. 72–83.
- Sun, Y., Jenkins, L.M.M., Su, Y.P., Nitiss, K.C., Nitiss, J.L. and Pommier, Y. (2019) A conserved SUMO-Ubiquitin pathway directed by RNF4/SLX5-SLX8 and PIAS4/SIZ1 drives proteasomal degradation of topoisomerase DNA-protein crosslinks. *BioRxiv*, p. 707661.
- Syed, A. and Tainer, J.A. (2018) The MRE11–RAD50–NBS1 Complex Conducts the Orchestration of Damage Signaling and Outcomes to Stress in DNA Replication and Repair. *Annual Review of Biochemistry*, Vol. 87 No. 1, pp. 263–294.
- Symington, L.S. (2016) Mechanism and regulation of DNA end resection in eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology*, 3 May.
- Szabo, Q., Bantignies, F. and Cavalli, G. (2019) Principles of genome folding into topologically associating domains. *Science Advances*, 1 April.
- Tan, K.B., Dorman, T.E., Falls, K.M., Chung, Thomas D. Y. Mirabelli, Christopher K. Crooke, S.T. and Mao, J. (1992) Topoisomerase II $\alpha$  and Topoisomerase II $\beta$  Genes: Characterization and Mapping to Human Chromosomes 17 and 3, Respectively. *Cancer Research*, Vol. 52 No. 14, pp. 3831–3837.
- Tang, W.K., Odzorig, T., Jin, W. and Xia, D. (2019) Structural basis of p97 inhibition by the site-selective anticancer compound CB-5083. *Molecular Pharmacology*, Vol. 95 No. 3, pp. 286–293.
- Thakurela, S., Garding, A., Jung, J., Schübeler, D., Burger, L. and Tiwari, V.K. (2013) Gene regulation and priming by topoisomerase II $\alpha$  in embryonic stem cells. *Nature*

- Communications*, Vol. 4 No. 1, pp. 1–13.
- Tiwari, V.K., Burger, L., Nikolettou, V., Deogracias, R., Thakurela, S., Wirbelauer, C., Kaut, J., et al. (2012) Target genes of Topoisomerase II $\beta$  regulate neuronal survival and are defined by their chromatin state. *Proceedings of the National Academy of Sciences*, Vol. 109 No. 16, pp. E934–43.
- Torrecilla, I., Oehler, J. and Ramadan, K. (2017) The role of ubiquitin-dependent segregase p97 (VCP or Cdc48) in chromatin dynamics after DNA double strand breaks. *Philosophical Transactions of the Royal Society B: Biological Sciences*.
- Travis, L.B., Ng, A.K., Allan, J.M., Pui, C.-H., Kennedy, A.R., Xu, X.G., Purdy, J.A., et al. (2012) Second Malignant Neoplasms and Cardiovascular Disease Following Radiotherapy. *JNCI: Journal of the National Cancer Institute*, Vol. 104 No. 5, pp. 357–370.
- Travis, L.B., Wahnefried, W.D., Allan, J.M., Wood, M.E. and Ng, A.K. (2013) Aetiology, genetics and prevention of secondary neoplasms in adult cancer survivors. *Nature Reviews Clinical Oncology*, 26 May.
- Uusküla-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M.V., Liang, M., Medina-Rivera, A., Mohammed, H., et al. (2016) Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biology* 2016 17:1, Vol. 17 No. 1, p. 182.
- Vaz, B., Popovic, M., Newman, J.A., Fielden, J., Aitkenhead, H., Halder, S., Singh, A.N., et al. (2016) Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair. *Molecular Cell*, Vol. 64 No. 4, pp. 704–719.
- Vian, L., Pękowska, A., Rao, S.S.P., Kieffer-Kwon, K.R., Jung, S., Baranello, L., Huang, S.C., et al. (2018) The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*, Vol. 173 No. 5, pp. 1165–1178.e20.
- Vos, S.M., Tretter, E.M., Schmidt, B.H. and Berger, J.M. (2011) All tangled up: How cells direct, manage and exploit topoisomerase function. *Nature Reviews Molecular Cell Biology*, December.
- Wei, P.C., Chang, A.N., Kao, J., Du, Z., Meyers, R.M., Alt, F.W. and Schwer, B. (2016) Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell*, Vol. 164 No. 4, pp. 644–655.
- Wilczynski, B., Dojer, N., Patelak, M. and Tiuryn, J. (2009) Finding evolutionarily conserved cis-regulatory modules with a universal set of motifs. *BMC Bioinformatics*, Vol. 10 No. 1, p. 82.
- Winters, A.C. and Bernt, K.M. (2017) MLL-rearranged leukemias- An update on science and clinical approaches. *Frontiers in Pediatrics*, 9 February.
- de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., et al. (2015) CTCF Binding Polarity Determines Chromatin Looping. *Molecular Cell*, Vol. 60 No. 4, pp. 676–684.
- Wolff, D.J., Bagg, A., Cooley, L.D., Dewald, G.W., Hirsch, B.A., Jacky, P.B., Rao, K.W., et al. (2007) Guidance for fluorescence in situ hybridization testing in hematologic disorders. *Journal of Molecular Diagnostics*, Vol. 9 No. 2, pp. 134–143.
- Wright, R.L. and Vaughan, A.T.M. (2014) A systematic description of MLL fusion gene formation. *Critical Reviews in Oncology/Hematology*, Vol. 91 No. 3, pp. 283–91.
- Xiao, H., Mao, Y., Desai, S.D., Zhou, N., Ting, C.Y., Hwang, J. and Liu, L.F. (2003) The topoisomerase II $\beta$  circular clamp arrests transcription and signals a 26S proteasome pathway. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 100 No. 6, pp. 3239–3244.
- Yan, H., Tammaro, M. and Liao, S. (2016) Collision of trapped topoisomerase 2 with transcription and replication: Generation and repair of DNA double-strand breaks with 5' adducts. *Genes*, 1 July.

## References

- Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., Custodio, J., et al. (2017) BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nature Communications*, Vol. 8 No. 1, pp. 1–9.
- Yang, F., Kemp, C.J. and Henikoff, S. (2015) Anthracyclines induce double-strand DNA breaks at active gene promoters. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, Vol. 773, pp. 9–15.
- Yesbolatova, A., Natsume, T., Hayashi, K. ichiro and Kanemaki, M.T. (2019) Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. *Methods*, Vol. 164–165, pp. 73–80.
- Yi, K. and Ju, Y.S. (2018) Patterns and mechanisms of structural variations in human cancer. *Experimental and Molecular Medicine*, 1 August.
- Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W., et al. (2004) Leukemia Proto-Oncoprotein MLL Forms a SET1-Like Histone Methyltransferase Complex with Menin To Regulate Hox Gene Expression. *Molecular and Cellular Biology*, Vol. 24 No. 13, pp. 5639–5649.
- Yoshihara, K., Wang, Q., Torres-Garcia, W., Zheng, S., Vegesna, R., Kim, H. and Verhaak, R.G.W. (2015) The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*, Vol. 34 No. 37, pp. 4845–4854.
- Yu, X., Davenport, J.W., Urtishak, K.A., Carillo, M.L., Gosai, S.J., Kolaris, C.P., Byl, J.A.W., et al. (2017) Genome-wide TOP2A DNA cleavage is biased toward translocated and highly transcribed loci. *Genome Research*, Vol. 27 No. 7, pp. 1238–1249.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., et al. (2018) Ensembl 2018. *Nucleic Acids Research*, Vol. 46 No. D1, pp. D754–D761.
- Zhang, A., Lyu, Y.L., Lin, C.P., Zhou, N., Azarova, A.M., Wood, L.M. and Liu, L.F. (2006) A protease pathway for the repair of topoisomerase II-DNA covalent complexes. *Journal of Biological Chemistry*, Vol. 281 No. 47, pp. 35997–36003.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., et al. (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, Vol. 9 No. 9, p. R137.
- Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., et al. (2012) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, Vol. 148 No. 5, pp. 908–921.
- Zhang, Y. and Rowley, J.D. (2006) Chromatin structural elements and chromosomal translocations in leukemia. *DNA Repair*, Vol. 5 No. 9–10, pp. 1282–1297.
- Zhou, K., Choe, K., Zaidi, Z., Wang, Q., Mathews, M.B. and Lee, C. (2003) RNA helicase A interacts with dsDNA and topoisomerase II $\alpha$ . *Nucleic Acids Research*, Vol. 31 No. 9, pp. 2253–2260.

## List of Abbreviations

Abbreviation	Full name
<b>AID</b>	auxin-inducible degron
<b>ALL</b>	acute lymphoblastic leukemia
<b>AML</b>	acute myeloid leukemia
<b>APH</b>	Aphidicolin
<b>BAC</b>	bacterial artificial chromosomes
<b>BCR</b>	breakpoint cluster region
<b>BLESS</b>	Breaks Labeling, Enrichment on Streptavidin, and Sequencing
<b>BLISS</b>	Breaks Labeling In Situ and Sequencing
<b>CDK</b>	cyclin-dependent kinase
<b>CDK9</b>	cyclin-dependent kinase 9
<b>ChIA-PET</b>	chromatin interaction analysis by paired-end tag sequencing
<b>ChIP</b>	chromatin immunoprecipitation
<b>CTCF</b>	CCCTC-binding factor
<b>DDR</b>	DNA damage response
<b>DNA-PKcs</b>	DNA- dependent protein kinase catalytic subunit
<b>DNaseI-HS</b>	DNaseI hypersensitivity
<b>DNMT1</b>	DNA methyltransferase 1
<b>dox</b>	doxycycline
<b>DPC</b>	DNA-protein crosslink
<b>DRB</b>	5,6-Dichloro-1- $\beta$ -d-ribofuranosylbenzimidazole
<b>DSB</b>	double-strand break
<b>EdU</b>	5-Ethynyl-2'-deoxyuridine
<b>ETO</b>	etoposide
<b>EU</b>	5-Ethynyl Uridine
<b>FISH</b>	fluorescence in situ hybridization
<b>G4</b>	G-quadruplexe
<b>gDNA</b>	genomic DNA
<b>gRNA</b>	guide RNA
<b>H3K4me3</b>	trimethylation to histone H3 at lysine 4
<b>HR</b>	homologous recombination
<b>ICE</b>	in vivo complex of enzyme
<b>IF</b>	immunofluorescence
<b>LAM-HTGTS</b>	linear amplification-mediated high-throughput genome-wide sequencing
<b>MEF</b>	mouse embryonic fibroblast
<b>MLL</b>	<i>mixed lineage leukemia</i> gene
<b>MMEJ</b>	microhomology-mediated end joining
<b>MRE11</b>	Meiotic recombination 11
<b>NGS</b>	next generation sequencing
<b>NHEJ</b>	non-homologous end joining

## List of Abbreviations

<b>NT</b>	non-target
<b>OsTIR1</b>	<i>Oryza sativa</i> Transport inhibitor response 1
<b>PBS</b>	phosphate-buffered saline
<b>PCR</b>	polymerase chain reaction
<b>PDB</b>	protein-linked DNA break
<b>PFA</b>	paraformaldehyde
<b>PLL</b>	Poly-L-lysine
<b>pRPA</b>	phospho-RPA
<b>P-TEFb</b>	Positive Transcription Elongation Factor
<b>PTM</b>	post-translational modification
<b>RNAi</b>	RNA interference
<b>RNA-PolIII</b>	RNA polymerase II
<b>ROI</b>	region of interest
<b>RPA</b>	replication protein A
<b>RPKM</b>	reads-per-kilobase-per-million
<b>RT</b>	room temperature
<b>RT-qPCR</b>	quantitative real-time PCR
<b>Sc</b>	Spearman's correlation coefficient
<b>SD</b>	standart deviation
<b>siRNAs</b>	small interfering RNA
<b>SMC</b>	structural maintenance of chromosomes protein
<b>SSB</b>	single-strand break
<b>ssDNA</b>	single-stranded DNA molecule
<b>TAD</b>	topologically associated domains
<b>t-ALL</b>	therapy-related acute lymphoblastic leukemia
<b>t-AML</b>	therapy-related acute myeloid leukemia
<b>TDP2</b>	Tyrosyl-DNA phosphodiesterase 2
<b>TOP2</b>	topoisomerases type II
<b>TOP2cc</b>	TOP2 cleavage complex
<b>TSS</b>	transcription start site
<b>TTS</b>	transcription termination site
<b>VCP</b>	valosin-containing protein
<b>WGS</b>	whole genome sequencing
<b>wt</b>	wild-type

## List of Figures

### Chapter 1: Introduction

Figure 1.1 Chemotherapeutic TOP2 poisons can lead to the formation of therapy-related leukemia.	6
Figure 1.2 DNA topoisomerase type II catalytic cycle.	8
Figure 1.3 TOP2A and TOP2B isoforms expression patterns.	9
Figure 1.4 Molecular mechanisms that contribute to the formation of DSB induction upon TOP2 poisoning.	13
Figure 1.5 Repair mechanisms of TOP2 cleavage complexes.	17
Figure 1.6 Methodologies to detect gene fusions and chromosome translocations.	23

### Chapter 2: Results

Figure 2.1 sBLISS identifies DSBs across the genome in hematopoietic cells.	28
Figure 2.2 ETO-induced DSBs cluster in open chromatin regions and colocalize with RNA-PolII and chromatin looping factor binding sites.	29
Figure 2.3 DSBs in the <i>MLL</i> gene detected by sBLISS in TK6, K562 and CD34+ progenitor cells.	31
Figure 2.4 DSBs in translocation partner genes <i>ENL</i> and <i>AF4</i> are partially found in translocation hotspot regions.	32
Figure 2.5 Detection of chromosomal rearrangements by C-Fusion 3D.	34
Figure 2.6 Correction of chromatic shifts to optimize distance calculations in 3D.	35
Figure 2.7 C-Fusion 3D detects chromosome translocations with high sensitivity at levels $10^{-3}$ to $10^{-4}$ .	37
Figure 2.8 Chromosome instability and translocation formation upon etoposide treatment can be studied by C-Fusion 3D.	38
Figure 2.9 Design of FISH probes to study <i>MLL</i> translocation by C-Fusion 3D.	39
Figure 2.10 C-Fusion 3D detects and quantifies ETO-induced gene breakage and <i>MLL</i> translocations in single hematopoietic cells.	40
Figure 2.11 Spatial proximity of <i>MLL</i> and <i>ENL</i> genes.	41
Figure 2.12 Transcription inhibition reduces ETO-induced DNA damage signaling.	43
Figure 2.13 ETO-induced DSBs cluster within active genomic regions.	43
Figure 2.14 Gene fragility correlates with gene activity.	44
Figure 2.15 Transcription inhibition reduces DSBs detected by sBLISS.	45
Figure 2.16 <i>MLL</i> and recurrent fusion partner genes are highly expressed in hematopoietic cell lines and CD34+ progenitor cells.	46

## List of Figures

Figure 2.17 <i>MLL</i> and recurrent fusion partner genes are highly enriched at chromatin loop boundaries.	47
Figure 2.18 Transcription output, proximity to loop anchors and gene fragility are interlinked.	49
Figure 2.19 Linkage of transcription and loop boundaries is cell line specific.	50
Figure 2.20 ETO-induced DSBs localize in nucleosome free regions around chromatin loop boundaries.	51
Figure 2.21 Transcription orientation affects the localization of DSBs at chromatin loop anchors in active genes.	53
Figure 2.22 Transcription orientation also influences breakage of <i>MLL</i> fusion partners at loop boundaries.	54
Figure 2.23 The <i>MLL</i> exon 12 DSB hotspot is localized just before a chromatin loop boundary.	56
Figure 2.24 Generation of a TK6 cell line with mutated CTCF motif within <i>MLL</i> .	57
Figure 2.25 TK6 <i>MLL</i> -CTCF <sup>mut</sup> cells have comparable DNA damage signaling but show reduced <i>MLL</i> fragility.	58
Figure 2.26 <i>MLL</i> breakage and translocation formation is largely transcription-dependent.	59
Figure 2.27 RNA-PolIII is required for ETO-induced <i>MLL</i> fragility.	61
Figure 2.28 <i>MLL</i> breakage in G1 depends on active transcription.	62
Figure 2.29 Suppression of <i>MLL</i> expression by <i>ZFP64</i> knockout reduces <i>MLL</i> breakage.	63
Figure 2.30 Generation of a TOP2B knockout in a dox-inducible TOP2A repression HT1080 cell line.	65
Figure 2.31 Generation of a TOP2A auxin-inducible degron and TOP2B knockout in HCT116 cells.	66
Figure 2.32 Both TOP2 isoforms contribute to ETO-induced DNA damage signaling.	67
Figure 2.33 <i>MLL</i> breakage and translocations depend on both TOP2 isoforms.	68
Figure 2.34 Higher <i>MLL</i> fragility occurs in S/G2 phase where both TOP2 isoforms are expressed.	70
Figure 2.35 TK6 cell lines deficient in TOP2cc removal or DSB repair pathways.	72
Figure 2.36 Kinetics of DSB repair in cell lines deficient for factors involved in TOP2cc removal or DNA-damage pathways.	72
Figure 2.37 TDP2, MRE11 and NHEJ, but not HR, suppress the formation of <i>MLL</i> translocations.	74
Figure 2.38 Role of TDP2 in suppressing <i>MLL</i> translocations that are mainly mediated by transcription.	75
Figure 2.39 Mre11 causes a replication-associated stress response upon ETO.	76
Figure 2.40 RNAi screen identifies factors that modulate <i>MLL</i> fragility upon etoposide.	79
Figure 2.41 Validation of factors modulating <i>MLL</i> fragility upon etoposide.	80



Figure 2.42 RNAi screen to identify VCP adaptor proteins that are required for <i>MLL</i> fragility upon etoposide.	82
Figure 2.43 Inhibition of VCP decreases ETO-induced DNA damage signaling.	84
Figure 2.44 VCP and proteasome inhibition decrease DNA damage signaling throughout the cell cycle.	84
Figure 2.45 Inhibition of VCP does not impede cellular processes.	85
Figure 2.46 VCP inhibition decreases ETO-induced DNA damage.	86
Figure 2.47 Inhibition of VCP increases cell survival upon TOP2 poisoning with etoposide.	87
<b>Chapter 3: Discussion</b>	
Figure 3.1 The interplay of transcription with chromatin loop anchors leads to TOP2-induced DSB formation.	93
Figure 3.2 <i>MLL</i> and common fusion partners localize at loop boundaries and are sensitive to transcription-dependent DSB formation.	95
Figure 3.3 TDP2, MRE11 and NHEJ are suppressors of TOP2-induced DSBs and illegitimate fusions.	101

## List of Figures

## List of Tables

Table 1 Cell lines used in this study.	110
Table 2 Genomic PCR components and cycling conditions for <i>MLL</i> exon 12.	114
Table 3 siRNA library for 150 factors to study ETO-induced <i>MLL</i> fragility.	116
Table 4 siRNA library of four pooled siRNAs for 33 VCP cofactors.	118
Table 5 Antibodies and dilutions for western blotting.	120
Table 6 Primers for qPCR amplification of immunoprecipitated chromatin of <i>MLL</i> exon 12 or control regions.	121
Table 7 Antibodies and dilutions for immunofluorescence.	122
Table 8 Bacterial artificial chromosomes (BACs) used for FISH labelling.	123
Table 9 Datasets generated and published within this study.	129
Table 10 Software used within this study.	139
Table 11 Public datasets used within this study.	140

## List of Tables

## **Acknowledgements**

Curriculum Vitae

## **Curriculum Vitae**

