The molecular principles of chromosomal translocations

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 14
Biochemie, Chemie und Pharmazie
Der Johann Wolfgang Goethe - Universität
In Frankfurt am Main

von Patrick Streb
aus Nidderau

Frankfurt 2023
D30

vom Fachbereich 14 Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe – Universität
als Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

1. Gutachter: Prof. Dr. Rolf Marschalek
2. Gutachter: Prof. Dr. Robert Fürst

Datum der Disputation:

Table of Contents

Table of figures

List of tables

List of abbreviations

ABA *abscisic acid*
ABI1 *abscisic acid insensitive protein I*
ALL *acute lymphoblastic leukemia*
AML *acute myeloid leukemia*
AML-MRC *AML with myelodysplasia-related changes*
BD *bromo domain*
CIP *chemically induced proximity*
circRNA *circular RNA*
CLL *chronic lymphocytic leukemia*
CLOuD9 *chromatin loop reorganization using CRISPR-dCas9*
CML *chronic myeloid leukemia*
CO_2 *carbon dioxide*
CREB *cAMP responsive element-binding protein*
CRISPR *clustered regularly interspaced short palindromic repeats*
crRNA *CRISPR-RNA*
dCas9 *dead Cas9*
deSpCas9 *dead enhanced Cas9 from Streptococcus pyogenes*
dSaCas9 *dead Cas9 from Staphylococcus aureus*
DSB *DNA double strand break*
FAB *French-American-British*
FCS *fetal calf serum*
FISH *fluorescence in situ hybridization*
HR *homologous recombination*
HSC *hematopoietic stem cell*
KMT2A-PTD *KMT2A-partial tandem duplication*
MDS *myelodysplastic syndromes*
mRNA *messenger RNA*
ncRNA *non coding RNA*
NGEFT *non-genomically encoded fusion transcript*
NHEJ *non-homologous-end-joining*
NOS *AML not otherwise specified*
O_2 *oxygen*
PAM *protospacer adjacent motif*
PBMC *peripheral blood mononuclear cell*
Ph *Philadelphia*
PHD *plant homeodomain*
PIKK *phosphatidylinositol 3-kinase-like kinase*
PTD *partial tandem duplication*
PTMs *pre-mRNA trans-splicing molecules*
PTTs *premature terminated transcripts*
PYL *PYR1-like component*
PYR *pyrabactin resistance component*
qRT-PCR *quantitative reverse transcription PCR*
RCAR *regulatory component of the ABA receptor*
SAR *scaffold attachment region*
SD *splice-donor site*
SET *Su(var)3-9,Enhancer-of-zeste and Trithorax*
sgRNA *single guide RNA*

SL RNA *spliced-leader RNA*

SMaRT *spliceosome-mediated RNA trans-splicing*

TAD *transactivation domain*

t-AML *therapy related acute myeloid leukemia*

TPGs *translocation partner genes*

tracrRNA *trans-activating RNA*

WHO *World Health Organization*

Deutschsprachige Zusammenfassung

Chromosomale Translokationen sind die Ursache für viele verschiedene Krebsarten, darunter auch Leukämien. *KMT2A*-Rearrangements sind dabei für ungefähr 5 – 10% der chromosomalen Translokationen bei Kindern und Erwachsenen mit akuter lymphatischer Leukämie, akuter myeloischer Leukämie und biphänotypischen Leukämien verantwortlich [1]. Wegen der schlechten Therapierbarkeit und einer allgemein schlechten Prognose wird die t(4;11)-Leukämie als eine Hochrisiko-Leukämie eingestuft [2, 3]. Die häufigste Aberration des *KMT2A*-Gens ist, mit ungefähr 80%, eine balancierte chromosomale Translokation und das häufigste Partnergen bei *KMT2A*-Rearrangements ist mit 36% aller Fälle das *AFF1*-Gen [4]. Diese balancierte Translokation führt zur reziproken Translokation t(4;11) mit den beiden Derivatchromosomen der(4) und der(11). Die Häufigkeit des *KMT2A* Partnergens ist dabei abhängig vom Alter und von der Art der Leukämie. Während *AFF1* das häufigste Partnergen bei akuter lymphatischer Leukämie ist, ist bei akuter myeloischer Leukämie *MLLT3* das häufigste Partnergen [4]. Die Derivatchromosomen führen im Anschluss zu den Fusionstranskripten *KMT2A::AFF1* und *AFF1::KMT2A*, diese werden in zwei Fusionsproteine translatiert (KMT2A-AFF1 und AFF1-KMT2A), die beide ein onkogenes Potential aufweisen. Neben den Fusionstranskripten, die von den Derivatchromosomen abgeschrieben werden, konnte das *KMT2A-ex*9*::AFF1-ex4* Transkript jedoch auch in den peripheren mononukleären Zellen von gesunden Spendern gefunden werden. Bei RT-PCR Versuchen mit peripherem Blut von gesunden Spendern wurde das Transkript in den Proben von 9 der 10 Spender gefunden, das reziproke Fusionstranskript *AFF1::KMT2A* konnte dabei in keinem der Spender nachgewiesen werden [5]. Interessanterweise wurde dies nicht nur für *KMT2A* und *AFF1*, sondern auch für andere Gene gefunden, die regelmäßig an chromosomalen Translokationen beteiligt sind. Beispiele hierfür sind *BCR::ABL* [6, 7], *TEL::AML1* [8, 9], *AML1::ETO* [9], *PML::RARα* [10], *NPM::ALK* [11, 12] und *ATIC::ALK* [12]. Da diese Transkripte experimentell nachgewiesen werden können, ohne dass sie direkt im Genom kodiert werden, werden sie als „*non-genomically encoded fusion transcripts*“ (NGEFTs) oder „nicht-genomisch kodierte Fusionstranskripte“ bezeichnet. Nach den vorhandenen experimentellen Daten entstehen sie mit größter Wahrscheinlichkeit als Ergebnis von alternativem Spleißen, wie beispielsweise intra- oder intergenisches *trans*-Spleißen [13-17]. Beim

intragenischen *trans*-Spleißen haben die generierten mRNAs meist duplizierte Exons oder Wiederholungen mehrerer Exons. Dies geschieht, wenn Teile verschiedener Transkripte desselben Gens zusammenspleißen [18-20]. Entscheidend für die nicht-genomisch kodierten Fusionstranskripte ist jedoch das intergenische *trans*-Spleißen. Es beschreibt das Zusammenspleißen von Transkripten zweier Gene, die zusammen in einer Transkriptionsfabrik transkribiert werden, und führt zur Bildung chimärer mRNAs. Dies geschieht im Fall von *KMT2A* und *AFF1* und auch im Fall der anderen Fusionstranskripte, die trotz fehlender chromosomaler Translokation nachweisbar sind [21-24]. Ein genauer Mechanismus, wie es zu den nicht-genomisch kodierten Fusionstranskripten kommt, ist bislang nicht aufgeklärt, jedoch wurden neben ihnen auch andere RNA Spezies gefunden, die von vielen Genen exprimiert werden, die an chromosomalen Translokationen beteiligt sind. Bei diesen RNA Spezies handelt es sich um die so genannten „premature-terminated transcript“ (PTTs), RNA Transkripte die vorzeitig terminieren und nicht komplett abgeschrieben werden [25]. Diese PTTs terminieren interessanterweise meist in den Bruchpunktsregionen der jeweiligen Gene, d.h. in der Region, in der die Gene sonst brechen und es dann zu einer Translokation kommt [26]. Durch die vorzeitige Terminierung haben PTTs eine Besonderheit: sie kodieren am letzten Exon für eine ungesättigte Spleiß-Donor Stelle (Figure 22). Aufgrund dieser freien Spleiß-Donor Stelle ist es möglich, dass PTTs zu anderen, neu gebildeten Transkripten in ihrer Nähe *trans*-spleißen. Warum das genau mit den Transkripten dieser Gene passiert, kann durch die so genannten „Chromosomenterritorien“ und „Transkriptionsfabriken“ erklärt werden. Im Interphase Zellkern sind die Chromosomen in so genannten Chromosomenterritorien angeordnet [27], jedes Chromosom nimmt dabei einen fest definierten Raum ein (Figure 6). An den Kontaktflächen zweier Chromosomen vermischen sich die Chromatinfasern einzelner Chromosomen und schaffen so Areale mit einer niedrigeren Chromatindichte [28, 29]. An diesen Stellen befinden auch vermehrt die Transkriptionsfabriken. Diese Transkriptionsfabriken sind eine Ansammlung von Transkriptionskomplexen, die dadurch in der Lage sind, mehrere Gene gleichzeitig abschreiben zu können. Dabei können diese Gene vom selben oder von verschiedenen Chromosomen kommen (Figure 7) [30-32]. In Versuchen konnte dabei gezeigt werden, dass Gene, die häufig an chromosomalen Translokationen beteiligt sind, zusammen in diesen Transkriptions-fabriken transkribiert werden, wie beispielsweise die Genpaare *KMT2A*/*AFF1* [22-24],

BCR/*ABL* [33, 34] und *cMYC*/*IgH* [35, 36]. Durch die gemeinsame Transkription in einer Transkriptionsfabrik liegen die Gene räumlich sehr nah beieinander, sodass es im Falle eines zeitgleichen DNA Schadens in beiden Genen zu einer Translokation kommen kann. Für die fehlerhafte Reparatur gibt es dabei zwei Hypothesen. Zum einen die „contact first" Hypothese, die besagt, dass sich die Translokationspartner schon vor dem DNA Schaden in räumlicher Nähe befanden und die „*breakage first*" Hypothese, nach der die gebrochenen DNA Enden sich frei bewegen können und nach einer Kollision mit dem falschen Ende falsch repariert werden [37, 38]. Die Reparatur findet dabei durch das so genannte „*non-homologous end joining*" (NHEJ) statt [39-42], hierbei werden DNA Enden ohne Vorlage oder Homologie miteinander verbunden. In Versuchen konnte gezeigt werden, dass RNA Transkripte, die vor dem DNA Schaden gebildet wurden, als Vorlage für die Reparatur mittels des NHEJ Weges dienen können [43-45]. Um die Entstehung chromosomaler Translokationen aufzuklären, wurde der Einfluss dieser einzelnen Bausteine untersucht, der der räumlichen Nähe der beteiligten Gene, sowie die Anwesenheit eines NGEFT, bevor es zu einem DNA Schaden kommt.

Um den Einfluss der räumlichen Nähe und den der NGEFTs auf die Entstehung einer t(4;11) zu untersuchen, wurde ein Zellsystem entwickelt, in dem die Nähe zweier Gene chemisch induziert werden kann. Als Zelllinie wurden HEK293T Zellen genutzt, da *KMT2A* und *AFF1* dort nicht in direkter räumlicher Nähe liegen. Das chemisch induzierbare Dimerisierungssystem bestand aus Proteindomänen des Abscisinsäure Weges aus Pflanzen. Die Dimerisierungspartner des Abscisinsäure Weges wurden modifiziert und nur die wesentlichen, für die Dimerisierung notwendigen Domänen für ein chemisch induzierbares System verwendet. Dabei handelt es sich um die regulatorische Domäne des ABA Rezeptors (PYL1) und das *Abscisic Insensitive Protein I* ohne Phosphatasedomäne (ABI1) [46]. An die beiden Proteinendomänen wurden enzymatisch inaktive dCas9 Proteine aus verschiedenen Spezies fusioniert, so kann das Dimerisierungssystem die DNA binden und nach der Zugabe von Abscisinsäure die beiden Gene in räumliche Nähe bringen. Die Zielsequenzen für die dCas9 Proteine lagen in *KMT2A* Intron 9 und in *AFF1* Intron 3, als Kontrollgene wurden *KMT2B* und *CCND3* gewählt, da diese weder ein NGEFT noch ein PTT exprimieren (Figure 9). Hier lagen die Zielsequenzen in *KMT2B* Intron 13 und *CCND3* Intron 2, zusätzlich wurde auch ein „non-targeting" System genutzt, zweie dCas9 Proteine mit sgRNAs, die nirgends im humanen Genom binden können (Negativkontrolle). Nach einer Inkubation mit Abscisinsäure über einen

Zeitraum von 14 Tagen, konnte in den Zellen mit den induzierbaren Systemen für *KMT2A* und *AFF1* und *KMT2B* und *CCND3* eine räumliche Nähe zwischen den Genen mittels Chromosome Conformation Capture gezeigt werden (Figure 11). Im Gegensatz dazu konnte für keines der Genpaare die Nähe in HEK293T Zellen mit dem non-targeting System gezeigt werden, was verdeutlicht, dass die Gene in Wildtyp HEK293T zu weit voneinander entfernt sind. Nachdem bestätigt werden konnte, dass die Gene durch das induzierbare System in räumliche Nähe gebracht werden können, wurden HEK293T Zellen mit den Systemen für die verschiedenen Zielgene für 14 Tage mit Abscisinsäure inkubiert. Anschließend wurde das Auftreten von NGEFTs in unbehandelten Zellen mit den 14 Tage behandelten Zellen vergleichen. Dabei konnte gezeigt werden, dass nach 14-tägiger, induzierter Nähe nur zwischen *KMT2A* und *AFF1* das *KMT2A-ex9::AFF1-ex4* Fusionstranskript gebildet wurde, jedoch kein reziprokes *AFF1::KMT2A* Transkript nachweisbar war (Figure 13). Obwohl zwischen *KMT2B* und *CCND3* auch räumliche Nähe induziert wurde, konnten keine Fusionstranskripte gefunden werden, ebenfalls nicht für die Zelllinie mit dem non-targeting System. Dies zeigt, dass die Bildung von NGEFTs zwar von der räumlichen Nähe abhängt, dies jedoch nicht der alleinig ausschlaggebende Faktor ist. Um den Einfluss auf chromosomale Translokationen zu untersuchen, musste in beiden beteiligten Genen ein Doppelstrangbruch ausgelöst werden. Hierfür wurde enzymatisch aktives Cas9 Protein genutzt, das vor der Transfektion mit den entsprechenden sgRNAs beladen wurde. In jeder Zelllinie wurden so zwei „zeitgleiche“ Schnitte im Genom gesetzt, entweder in *KMT2A* Intron 11 und *AFF1* Intron 3 oder in *KMT2B* Intron 13 und *CCND3* Intron 2. Durch die gezielten Schnitte sollte der zufällige und gleichzeitige DNA Schaden in beiden beteiligten Genen nachgestellt werden. Nach der Inkubation mit Abscisinsäure über 14 Tage und der anschließenden Induktion eines DNA Schadens, konnte zwischen *KMT2A* und *AFF1* eine balancierte Translokation (Figure 16) und die Fusionstranskripte für *KMT2A-ex11::AFF1-ex4* und *AFF1-ex3::KMT2A-ex12* (Figure 15) abundant nachgewiesen werden. Im Gegensatz dazu konnte für *KMT2B* und *CCND3*, trotz räumlicher Nähe, keine Translokation nachgewiesen werden. In den Zellen mit non-targeting System konnte keine der Translokationen nachgewiesen werden, hier waren die Gene weder in räumlicher Nähe, noch war zum Zeitpunkt des Schadens ein NGEFT nachweisbar. Da die räumliche Nähe für eine Translokation zweier Gene eine logische Notwendigkeit darstellt, sollte letztendlich der Einfluss der *fusions-RNA* untersucht werden. Hierfür

wurde zusätzlich zum DNA Schaden auch eine artifizielle RNA in die Zellen mit induzierter Nähe zwischen *KMT2B* und *CCND3* transfiziert. Die RNA bestand dabei aus *KMT2B* Exon 9 – 13 und *CCND3* Exon 3 – 5. In Anwesenheit der artifiziellen *fusions-RNA* führte der DNA Schaden auch bei *KMT2B* und *CCND3* zu einer balancierten Translokation (Figure 19), eine t(6;19), die noch nirgends in der Natur gefunden oder in der Literatur beschrieben wurde. Wichtig ist dabei auch, dass in den Zellen mit non-targeting System wiederum keine Translokation stattgefunden hat. Da jeder Versuch eine heterogene Verteilung von verschiedenen Bruchpunkten ergibt (auf der Nukleotidebene), wurden die Regionen der neuen Genfusionen molekular analysiert (Figure 17, Figure 20). Hier konnte festgestellt werden, dass - genau wie bei Leukämie-Patienten – es im Bereich des induzierten DNA Bruchs zur Deletion weniger Basen, dem Vorhandensein von sogenannter „*Filler DNA*", *mini-direct repeats* und generellen Fehlern bei der DNA Reparatur gekommen war.

Zusammenfassend konnten in dieser Arbeit die notwendigen Voraussetzungen für die Entstehung einer chromosomalen Translokation identifiziert und molekular nachgestellt werden. Diese Ergebnisse sind nicht nur auf die Gene *KMT2A* und *AFF1* anwendbar, sondern vermutlich ausschlaggebend für die Entstehung vieler bis aller chromosomaler Translokationen. Es konnten drei kritische Punkte identifiziert und bewiesen werden. Zum einen die räumliche Nähe der beiden beteiligten Gene (1), dies ging aus den Versuchen mit der non-targeting Kontrolle hervor, die in keinem Szenario eine chromosomale Translokation entwickelte. Entscheidend war zweitens der gleichzeitige DNA Schaden (2), der in der Natur nur mit einer sehr geringen Wahrscheinlichkeit gleichzeitig auftritt. Als dritten Punkt konnte die Eigenschaft PTTs zu exprimieren (3) als kritisch identifiziert werden. Die PTTs führen zu alternativ gespleißten Fusionstranskripten, die im Anschluss als Vorlage für die DNA Reparatur genutzt werden können. Dies konnte durch die artifizielle *KMT2B::CCND3 fusions-RNA* und der entstandenen t(6;19) gezeigt werden. Alle drei Punkte (1-3) waren für die Ausbildung der Translokation ausschlaggebend, sobald eine der Voraussetzungen fehlte, kam es nicht zu einer Translokation. Diese Erkenntnisse zeigen nicht nur die Voraussetzungen für die Entstehung einer Translokation, sondern identifizieren auch Gene, die PTTs exprimieren als mögliches Risiko und zeigen, dass genomische RNAs an Reparaturprozessen beteiligt sind.

1. Introduction

1.1. The hematopoietic system

The blood is one of the most highly regenerative tissues. The human hematopoietic system produces about one trillion new cells every day [47]. There is a wide variety of cellular morphologies corresponding to cells of various lineages and stages of differentiation. A Russian biologist postulated that hematopoiesis is organized as a hierarchy derived from a common precursor, the hematopoietic stem cell (HSC) [48].

The formation and differentiation of new blood cells is called the hematopoiesis. HSCs differentiate to multi- and oligopotent precursor cells and then mature into blood cells. The lymphopoiesis is the generation of B-cells, T-cells and NK-cells from lymphatic stem cells. B-cell maturation takes place in the bone marrow, where they are negative selected for autoantigen binding, just 1 out of about 200 cells survives this negative selection. T-cells on the other hand mature within the thymus, here just 1 out of 100 cells survives the negative selection for the binding activity to autoantigens. The naive B- and T-cells derived from the primary hematological organs, bone marrow and thymus, differentiate into B- and T-effector cells in the secondary lymphatic organs after their activation.

The myelopoiesis takes place exclusively in the red bone marrow. It is the differentiation of progenitor cells into erythrocytes, monocytes, megakaryocytes, and granulocytes (eosinophils, neutrophils and basophils). Myelopoiesis mainly results in the most abundant population of erythrocytes which are the main cellular component of the blood. Erythrocytes are nucleus free cells which transport oxygen (O_2) from the lung into the whole body and carbon dioxide (CO_2) back to the lung by binding both gases to tetrameric hemoglobins.

Platelets are formed and released by cells called megakaryocytes that reside within the bone marrow. Thousands of platelets can be released from a single megakaryocyte by a series of remodelling events [49]. Platelets only circulate for about 7 – 10 days, their main task is to stop hemorrhage following vascular injuries. In case of a vascular injury platelets adhere to the extracellular matrix and aggregate to stop bleeding [50].

Monocytes are immature macrophages and part of the innate immune system. They are both critical effectors and regulators of inflammation [51]. Monocytes produce inflammatory cytokines and take up cells and toxic molecules. They can also migrate

into tissues and differentiate into inflammatory dendritic cells or macrophages, which is determined by the inflammatory milieu [52].

Granulocytes can be classified as neutrophils, eosinophils and basophils, depending on their staining characteristics using hematoxylin and eosin histological preparations [53]. Neutrophils are regarded as the first defence of the innate immune system and represent the highest population of leukocytes. They destroy invading microorganisms through phagocytosis or release of granules [54]. Eosinophils are crucial for the control of parasitic infections [55]. Basophils are also able to fight parasitic infections and are also recruited to tissues under allergic disease conditions, but represent the smallest population of peripheral blood leukocytes [56].

The hematopoiesis is a very strictly regulated mechanisms, the deregulation can lead to a variety of different diseases.

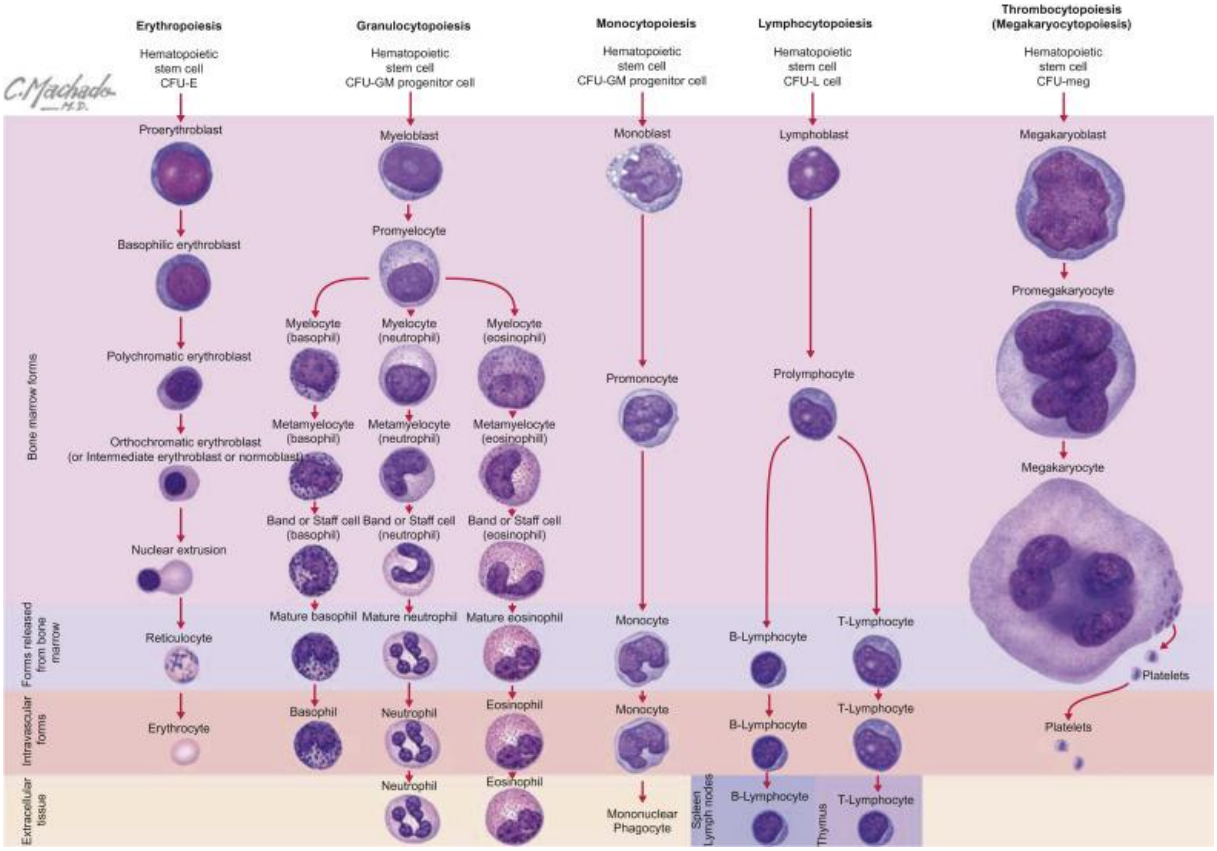

Figure 1. Scheme of the human hematopoiesis. [57] The hemopoietic and lymphatic tissues consist of erythrocytes, leukocytes and platelets. The hematopoietic tissue arises from hematopoietic stem cells (HSC) and produces the naïve lymphocytes and myeloid cells in the bone marrow, thymus and peripheral blood, while the lymphoid tissues make up the lymphatic system and also include the secondary lymphoid tissues of lymph nodes, spleen and mucosa-

associated tissues [57]. In the bone marrow, hematopoietic stem cells differentiate into myeloid and lymphatic precursor cells. Those cells further differentiate into the different cell types of the hematopoietic system. B-cells, T-cells and natural killer cells belong to the lymphatic lineage, platelets, erythrocytes, monocytes and granulocytes (neutrophils, basophils, eosinophils) belong to the myeloid lineage.

1.2. Leukemia

Leukemia is a malignant disease of the primary hematopoietic systems (bone marrow, thymus). Under malignant conditions mutant cells in these organs produce an excess of immature or abnormal leukocytes. This suppresses the production of normal blood cells and results in syndromes like cytopenia (erythrocytes, thrombocytes) with all consequences (diagnostic criteria). The malignant transformation usually occurs at the pluripotent stem cell level.

Leukemia is currently classified based on the 2016 World Health Organization (WHO) system. This classification is based on a combination of morphological, immuno-phenotypic, clinical and genetic features. Less commonly used classification systems include the French-American-British (FAB) system, which is based on the morphology of the abnormal leukocytes. Leukemias are also commonly categorized as acute or chronic, depending on the percentage of blasts or leukemic cells in the bone marrow or blood and as myeloid or lymphoid, based on the dominant lineage of the malignant cells. Acute leukemias consist of immature, poorly differentiated cells, while chronic leukemias have more mature cells and usually manifest as leukocytosis with or without cytopenia. Factors that increase the risk of developing leukemia include exposure to ionizing radiation or to chemicals. Prior treatment with certain antineoplastic drugs, like alkylating agents, topoisomerase II inhibitors and maintenance lenalidomide can lead to therapy-related acute leukemia (t-AML, t-ALL). Also, infections with viruses, for example human T lymphotropic virus 1 and 2 and Epstein Barr virus can rarely cause certain forms of ALL. Hematologic disorders like myelodysplastic syndromes and myeloproliferative neoplasm or pre-existing genetic conditions can also lead to acute myeloid leukemias.

The four most common types of leukemia are the acute myeloid leukemia (AML), the acute lymphoblastic leukemia (ALL), the chronic myeloid leukemia (CML) and the chronic lymphocytic leukemia (CLL). The acute lymphoblastic leukemia and the acute

myeloid leukemia are often caused by a series of genetic aberrations, leading to a malignant transformation. ALL is further categorized as B-lymphoblastic leukemia or as T-lymphoblastic leukemia, depending on the lineage. Malignant transformation that occurs in secondary hematological organs (lymph nodes, spleen, etc.) are defined as B-cell or T-cell lymphomas. The classification of acute myeloid leukemia is more complex, the subtypes are distinguished by morphology, immunophenotype, cytochemistry and genetic abnormalities. There are seven classes described in the WHO classification: AML with recurrent genetic abnormalities, AML with myelodysplasia-related changes (AML-MRC), therapy-related AML (t-AML), AML not otherwise specified (AML-NOS), myeloid sarcoma, myeloid proliferations related to down syndrome, blastic plasmacytoid dendritic cell neoplasm.

The chronic lymphocytic leukemia is the most common type of leukemia in the western world. In CLL CD5+ B-cells undergo malignant transformation. The cells become continuously activated by the acquisition of mutations, further accumulation of genetic abnormalities and the subsequent oncogenic transformation leads to CLL. CML is often caused by a t(9;22) chromosomal translocation leading to a fusion oncogene. The resulting chromosomes are called Philadelphia (Ph) chromosomes and they are present in 90 to 95% of CML cases. The oncoprotein BCR-ABL has uncontrolled tyrosine kinase activity, which deregulates the cellular proliferation and protects leukemic cells from normal programmed cell death.

All types of leukemia are differentially diagnosed by white blood cell (WBC) counts, peripheral blood smears, bone marrow examinations, cytogenetics and immunophenotyping. [58]

1.3. *KMT2A*-rearrangements

The *KMT2A* gene (in the past also named *MLL*, *ALL-1*, *HRX*, *hTRX*) is located at 11q23 and rearrangements of this gene account for 5 - 10% of chromosomal rearrangements in children and adults with acute lymphoblastic leukemia (ALL), acute myeloblastic leukaemia (AML), poorly differentiated or biphenotypic leukemias and myelodysplastic syndromes (MDS). The majority of acute and secondary or treatment-related leukemias in infants involve 11q23 rearrangements [1] with an overall proportion of about 80% of

cases [59]. These 11q23 rearrangements are strongly predictive for poor clinical outcome [1].

Chromosomal aberrations concerning the chromosome band 11q23, have been observed in ALL and especially AML, so in 1991 the gene on 11q23 was found and named *KMT2A* gene (myeloid/lymphoid, or mixed-linage leukemia) [60]. It has been shown that age was an especially important factor for the outcome of *KMT2A* abnormalities. Infants younger than 1 year fared significantly worse, than patients that were 1 year or older [2]. Taken to the fact that *KMT2A* gene rearrangements in infant (<1 year) acute lymphoblastic leukemia occurs in about 75% of cases with bad prognosis [61]. Furthermore, for infants any abnormality in 11q23 conferred a poor outcome, whereas in older patients t(4;11) and t(9;11) were associated with a worse outcome than other changes [2]. Taken together this shows that patients with rearrangements in the *KMT2A* gene causing leukemia have a poor prognosis depending on the age.

1.4. *KMT2A*-protein

The KMT2A protein with its 3969/3972/4002/4005 amino acid residues (4 different variants due to alternative splice events) has a molecular weight of ~435 kDa and at least 6 functional protein domains. These domains include three DNA-binding AT hooks, a cysteine-rich CXXC domain with a homology to DNA methyltransferases, plant homeodomain (PHD) finger motifs, a bromo domain (BD), a transactivation domain (TAD), a nuclear receptor interaction motif (NR box), a WDR5 interaction motif and a C-terminal Su(var)3-9,Enhancer-of-zeste and Trithorax (SET)-domain [62]. The AT-hooks are binding to the minor groove in AT-rich DNA sequences [63]. The CXXC domain is essential for target promotor recognition. The domain binds selectively to unmethylated CpG DNA, therefore targeting *KMT2A* to active genes, because in vertebrates the promotors of active genes are generally hypomethylated [62]. The crescent-like shape of the CXXC domain coordinates two zinc ions using the two conserved CGXCXXC motifs. Mutation of the cysteine residues involved in zinc ion coordination results in protein unfolding, showing that the zinc ions are essential for the structural integrity [64]. The interaction of the TAD-domain with the CREB-binding protein (CREBBP) is important for KMT2A-mediated transcriptional activation [65]. The highly conserved SET domain

is found in a number of chromatin-associated proteins with diverse transcriptional activities [66]. KMT2A belongs to the SET1 family of SET domain proteins, that regulate H3K4 methylation levels and are found in conserved multi subunit complexes [67]. It has been shown, that the WD40 repeat protein WDR5 is a conserved component of SET family complexes and important for the H3K4 methylation and HOXA gene expression in hematopoiesis and development [68]. WDR5 interacts directly with KMT2A and bridges interactions between KMT2A and other components if the KMT2A core complex. This functional region in the KMT2A protein is called WDR5 interaction motif [69].

1.5. *KMT2A*-gen

The *KMT2A*-gene is approximately 92 kb large and consists of 37 exons which decode ~12 kb of coding sequence. The breakpoint cluster region, where most breakpoints in *KMT2A* rearranged leukemia occurs, is localized in an 8,331 bp BamHI-fragment between exon 8 and exon 14 [70]. There are several motifs that could lead to this breakpoint cluster region. This includes Alu-sequences [70], a Topoisomerase II binding site [63], a DNase I hypersensitive site [71] and a gene internal transcription initiation site [72]. Additionally, there is a scaffold attachment region (SAR) which spans the break point cluster region, while most of the Topoisomerase II binding sites are in the telomeric region of the break point cluster region and overlap with the SAR [63]. Interestingly, in *AFF1* and *MLLT3*, the most common translocation partners of *KMT2A* in ALL and AML, there are SARs and Topoisomerase II binding sites within the break point cluster region too [71, 73].

1.6. *KMT2A*-fusions

Chromosomal or genetic fusions can happen in different ways. There are deletions, inversions, repeats, insertions and translocations. This also applies for the *KMT2A* gene, where different types of fusions are known. The so called interstitial 11q deletion is the deletion of a DNA fragment caused by 2 breaks in the q arm of chromosome 11, at least one within the *KMT2A* gene. An inversion would be present if the fragment is inserted the wrong way around. The *KMT2A-partial tandem duplication* (*KMT2A*-PTD) is a duplication of a *KMT2A* fragment, normally fragments that code *KMT2A* exon 3 to exon 9. An insertion is an additional fragment of DNA from a translocation partner gene in the

KMT2A gene or a part of the *KMT2A* gene inserted into a translocation partner gene. The most frequent aberration regarding the *KMT2A* gene is, with about 80%, a balanced reciprocal translocation. In this case a double strand break in two chromosomes occurs and a fragment of chromosome 11 is exchanged with a fragment of a partner chromosome. The most common reciprocal translocation of the *KMT2A* gene is the t(4;11), leading to the chromosome derivatives der(4) and der(11).

Not all fusions cause functional proteins, for example because of a phase shift. Exons do not always end with a complete triplet of bases, so only introns with the same phase can be fused to maintain the open reading frame [74]. In exceptional cases *spliced fusions* with a destroyed open reading frame were found, that formed functional transcripts by splicing. There were also splice variants observed with genes next to the fusion partner gene [4, 75].

Until now 135 different *KMT2A*-rearrangements have been identified, of which 95 translocation partner genes are characterized at the molecular level. The most frequent *KMT2A* fusion gene partners are *AFF1* (36%), *MLLT3* (19%), *MLLT1* (13%), *MLLT10* (8%) and *ELL*, *AFDN* and *KMT2A partial tandem repeats* (4% each). The frequency of the *KMT2A* fusion genes is age and disease type specific (Figure 2). While *AFF1* is the most common *KMT2A* fusion partner gene for acute lymphoblastic leukemia, *MLLT3* is the most frequent for acute myeloid leukemia. This also varies in the different age groups, infants, children and adults [4].

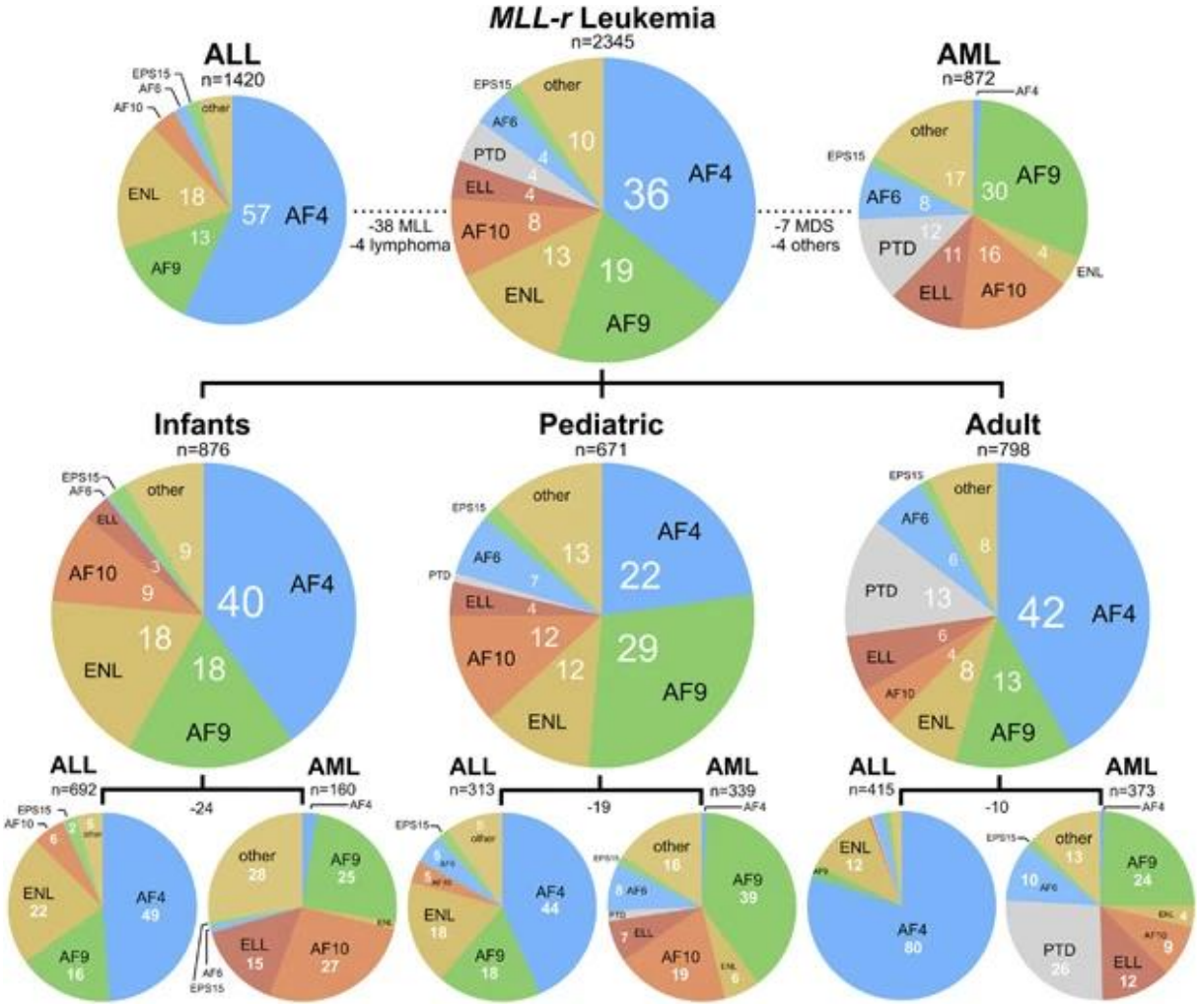

Figure 2. Classification of patients with *KMT2A-r* according to age and disease. Top: frequency of most frequent translocation partner genes (TPGs) in the investigated patient cohort (n=2345) and divided into ALL and AML cases. Middle: TPG frequency for different age groups. Bottom: subdivision of all three age groups into ALL and AML patients [4].

1.7. Inter- and intragenic *trans*-splicing

Splicing either happens as *cis*- or *trans*-splicing. *Cis*-splicing describes the normal splice event to eliminate intronic sequences and to fuse all exons in order to generate a continuous reading frame. On the other side, *trans*-splicing describes splicing events where two different pre-mRNAs are spliced together.

Trans-splicing is a very common process in eukaryotes and higher eukaryotes and seems to be enhanced by sequence complimentary in the intronic regions of the two pre-mRNAs [76]. In *Caenorhabditis elegans* it has been found that the *spliced-leader* RNA (SL RNA) is *trans*-spliced to many RNAs [77], for example when the polycistronic pre-

mRNAs are processed to monocistronic mRNAs. To date, *trans*-splicing has been observed in various species, including *Homo sapiens* [13-17]. There are two different subtypes of *trans*-splicing, inter- and intragenic *trans*-splicing.

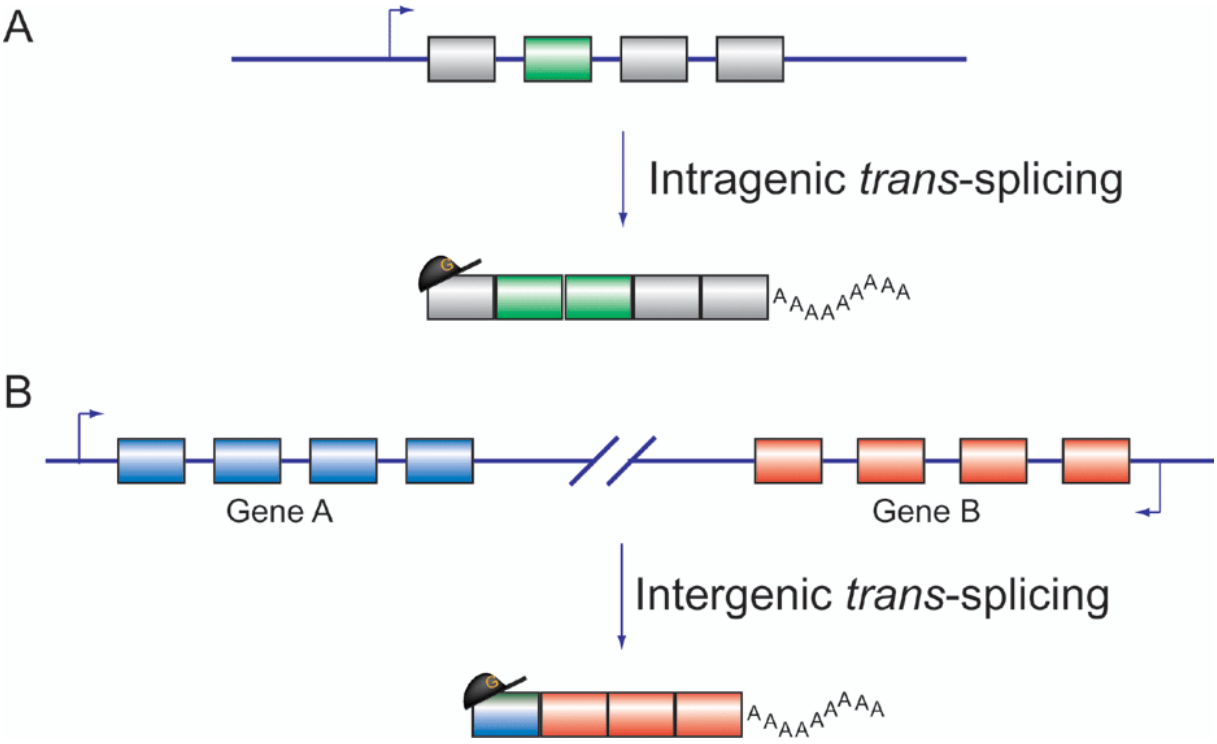

Figure 3. Alternative *trans*-splicing in mammals. A) mRNAs containing tandem duplications are generated by intragenic *trans*-splicing. B) Chimeric mRNAs are generated by intergenic *trans*-splicing of the two different genes, Gene A and Gene B. [78]

1.7.1. Intragenic *trans*-splicing

Intragenic *trans*-splicing leads to mRNAs that have duplicated exons. Carnitine octanoyltransferase and a medium-chain acyl-CoA- synthetase, which produce relatively large amounts of mRNAs containing duplications of exon sequences, were the first examples of intragenic *trans*-splicing found in rats [14, 79, 80]. Subsequently, exon repetitions were also reported in other loci [18-20]. These *trans*-spliced RNAs of carnitine octanyltransferase and medium-chain acyl-CoA synthetase have been detected only in particular rat strains and are produced in an allele-specific manner [79]. It is not clear whether mRNAs with duplicated exons have any function or significance.

Intragenic *trans*-splicing may occur through exon repetition, sense-antisense fusion and exon scrambling. In *Drosophila* there are chimeric RNAs of the *mod* and *lola* genes that originate from intragenic *trans*-splicing [81-83]. Although the specific function of most

intragenic *trans*-spliced RNAs remains elusive, there are some human intragenic *trans*-spliced RNAs that were confirmed to be expressed across primates and rodents [84, 85]. Studies on the human *SP1* gene [86] have shown that long introns flanking the exon promote *trans*-splicing, this finding was also characterized by other groups [87]. They proposed that long flanking introns increase the time taken for the synthesis of the introns by RNA polymerase II, and thereby increasing the chance that introns from separate pre-mRNA molecules interact in an RNA-RNA secondary structure [87, 88]. There is evidence that the *trans*-splicing takes place between two RNA molecules while they are attached to the same gene, because the exon repetition happens in *cis*, meaning that transcripts from two alleles do not splice to each other [79]. Previous studies have shown that the introduction of complementary intronic sequences can increase *trans*-splicing between pre-mRNA molecules *in vitro* by forming small RNA duplexes between the separate molecules [89].

From the late 90s researchers have utilized their knowledge about the *trans*-splicing process as a tool for gene therapy for genetic defects such as cystic fibrosis or to reprogram gene expression [90, 91].

Another rather new example of intragenic *trans*-splicing are the circular RNAs. They are generated by a special form of alternative splicing called back splicing. Precursor mRNAs are back spliced by joining the 3´-splice site of a downstream exon to the 5´-splice site of an upstream exon creating an RNA circle. The majority of circRNAs derive from exons in the middle of the gene and contain two to three exons [92-94]. However, circRNAs do not always come from one gene, there were also circRNAs found deriving from intergenic splicing or from intronic regions [95]. These circular RNAs were shown to have a variety of functions within the cell including miRNA sponging [96], protein sponging [97, 98], protein scaffolding [99] and protein coding [100, 101].

Taken together there is a variety of intragenic *trans*-splicing in different species, including humans and other vertebrates. Some functions of the product from alternative splicing have already been observed, some are still to discover.

1.7.2. Intergenic *trans*-splicing

Intergenic *trans*-splicing generates chimeric mRNAs between pre-mRNAs from two different genes. One early example is the acyl-CoA:cholesterol acyltransferase 1 gene

which produces a mRNA that contains exons derived from two different chromosomes resulting in a protein with an N-terminal extension [21]. This protein was functional when transfected into cultured cells and presented in monocyte-derived macrophages, but the function of this protein remains unknown [102]. There are also other examples [103] like the human liver, that is capable to produce a variety of hybrid CYP3A (cytochrome P450 3A) mRNA molecules from four different genes sharing a high degree of similarity, CYP3A4, CYP3A5, CYP3A7 and CYP3A43 [104].

There are also therapeutic strategies involving intergenic *trans*-splicing. Either using *trans*-splicing to repair mutations at the mRNA level and second inducing cell death by *trans*-splicing constructs encoding toxic factors. Spliceosome-mediated RNA *trans*-splicing (SMaRT), ribozyme- mediated *trans*-splicing and split intein-mediated *trans*-splicing, where the first therapeutic techniques used. SMaRT exploits the cell´s splicing machinery to *trans*-splice therapeutic RNA into endogenous pre-mRNA delivered by artificially engineered pre-mRNA *trans*-splicing molecules (PTMs), containing a splicing and a hybridization domain [90]. Another technique is to use engineered ribozymes to cleave a target pre-mRNA transcript upstream of the sequence that needs to be replaced and catalyze the *trans*-splicing of a therapeutic sequence to the 3´-end, so this technique is limited to 3´-end replacement [105]. For the protein *trans*-splicing techniques, split inteins are used to separately deliver gene fragments that *trans*-splice themselves together to reconstitute the full-length protein within the target cell [106, 107].

In contrast to non-functional, functional or therapeutic intergenic *trans*-splicing there are also chimeric RNAs that mimic the existence of a fusion transcript from a DNA fusion in the absence of the actual translocation, this will be discussed in more detail in the next paragraphs.

1.8. Premature terminated transcripts (PTT)

Premature terminated transcripts (PTTs) are pre-mRNAs that are terminated and polyadenylated upstream of their actual stop codon. This phenomenon has mainly been observed during the transcriptional initiation when polymerase II is poised at nucleotide +50 and awaits 5'-capping. But it also happens within the pre-mRNA, mainly in intronic sequences, making PTTs a general feature for many human genes, that also become poly-adenylated at cryptic polyadenylation sites encoded by these introns [25].

There are also PTTs where the polyadenylation site is within an exon but these transcripts without stop codon get typically degraded rapidly through the nonstop decay pathway [108]. In very rare cases they can also result in truncated proteins [109]. It has been shown that about 40% of murine genes [110] and 16% of genes in human immune cells [111] express isoforms with internal polyadenylation. Furthermore, intronic polyadenylation has been shown to have diverse roles in immune cells [112], inactivation of tumour suppressor genes [113] and regulation of DNA repair [114]. An analysis of intronic polyadenylation in human tissues, primary immune cells and multiple myeloma cells has shown nearly 5,000 high-confidence intronic polyadenylation events [112]. Short transcripts appear to undergo very quick degradation [115-118], while stable PTTs are mainly polyadenylated and contribute to the intronic polyadenylated transcriptome. These PTTs can be further divided into ncRNA and protein-coding mRNAs, although some non-coding PTTs contain an open reading frame. The PTT of *ASCC3* is a stable ncRNA and critical for the transcription after UV damage [119]. These ncRNAs generated by intronic polyadenylation also interact with RNA-binding proteins and could thereby regulate other RNAs [120]. Protein coding intronic polyadenylated isoforms can generate proteins with physiologically distinct functions or truncated dominant negative proteins. In B-cells for example, the cellular activation can cause a switch from full length immunoglobulin M heavy chain mRNA to an intronic polyadenylated isoform resulting in a change from membrane-bound to secreted forms of the antibody [121-123]. It has been shown that 499 genes encoding transmembrane proteins undergo intronic polyadenylation and in 152 cases this leads to the loss of the transmembrane domain [120]. Similarly, intronic polyadenylation seems to occur preferentially on genes encoding transcriptional regulators [120]. It remains unclear how the different mechanisms lead to or prevent the formation of PTT and which of the factors enhancing or preventing the formation are doing it directly.

There are genes that are frequently involved in chromosomal translocations. Genes like *KMT2A*, *AFF1*, *MLLT3*, *ELL* and *MLLT1* display this feature and produce premature terminated transcripts (PTTs). In qRT-PCR experiments it has been shown that these genes frequently involved in chromosomal translocations exhibit intronic polyadenylated PTTs in a relative abundance of 10–20% compared to the full-length polyadenylated transcripts [26]. Interestingly, the termination of the PTTs always occurred in introns that are part of the break point cluster region of the corresponding gene, in

KMT2A intron 8 and 9, *AFF1* intron 3, *MLLT3* intron 5, *ELL* intron 2 and *MLLT1* intron 3. In other studies, there were also PTTs from other genes that were terminated within the breakpoint cluster region, *ETV6* intron 5, *EWSR1* intron 7, *NUP98* intron 12 and intron 13 and RUNX1 intron 6 , while several other genes, like *GAPDH*, *ACTB*, *HSPCB*, *CCND3, RPL3* or *KMT2D*, did not display PTTs [124]. In General, these PTTs are of interest because they contain an unsaturated splice-donor site (SD) in their 3´-region, which are normally all saturated by splicing during the pre-mRNA maturation. It is convincible that PTTs with their unsaturated splice-donor site could be a source for exon repetitions, found in some transcripts and also intergenic *trans*-spliced RNAs, forming non-genomically encoded fusion transcripts (NGEFTs) [26, 124].

1.9. Non-genomically encoded fusion transcripts

Non-genomically encoded fusion transcripts (NGEFTs) are *trans*-spliced fusion mRNAs that are not directly encoded by the genomic DNA and are usually identified in low abundance. A first hint for these NGEFTs came from experimental results where peripheral blood or biopsied bone marrow cells of healthy individuals were investigated. With these investigations different research groups were able to demonstrate the occurrence of known oncogenic fusion transcripts without any genomic rearrangement of the corresponding loci. Many of these fusion transcripts are normally expressed in different types of cancers when there is a chromosomal translocation, like *KMT2A::AFF1* fusion transcripts or *KMT2A*-partial tandem duplications (PTDs) [22-24] or fusion transcripts for *BCR::ABL* [6, 7], *TEL::AML1* [8, 9], *AML1::ETO* [9], *PML::RARα* [10], *NPM::ALK* [11, 12] and *ATIC::ALK* [12]. These findings were controversially discussed in the literature and were first either classified as PCR artefacts or as rarity, although different laboratories provided conclusive evidence for the existence of these NGEFTs. NGEFTs are of very low abundance but they could be translated into oncogenic fusion proteins, without triggering a disease phenotype because it was not diagnosed in the investigated healthy individuals. It is assumed that the expression of these oncofusion proteins does not sufficiently reach a level high enough to reveal their oncogenic potential or they are expressed in a non-stem cell compartment. An example is the *JAZF1::JJAZ1* fusion RNA, that is normally expressed in endometrial stromal sarcomas harbouring a t(7;17)(p15;q21), which can be readily detected in normal endometrial stromal cells

[17, 125]. Moreover, this low abundant *JAZF1::JJAZ1* fusion mRNA produces a chimeric protein that exhibits anti-apoptotic effects providing a gain-of-survival function in this tissue. Similar findings were reported for the *trans*-spliced *SLC45A3::ELK4* fusion transcript identified in normal and prostate cancer cells [126].

In additional experiments further attempts were made to enlighten the cellular mechanisms that could explain the high frequency of *KMT2A* rearrangements. In these experiments *KMT2A::AFF1* fusion transcripts could be observed in PBMCs from 9 out of 10 healthy individuals, but the reciprocal *AFF1::KMT2A* was absent. The same healthy individuals also displayed a *NPM::ALK* fusion transcripts usually found in cells harbouring a t(2;5)(p23;q35) [26].

The *trans*-splicing events that are producing the NGEFTs can only be explained when the genes that splice together are transcribed in the same transcription factory or from corresponding chromatin loops in close proximity. This has been demonstrated by RNA- and DNA-FISH experiments for *PML* and *RARA* [127], *IgH* and *cMYC [35, 36]* or the genes *BRC* and *ABL* [34]. But there is a missing link between the presence of NGEFTs and their encoded oncofusion proteins because their presence alone does not enforce any oncogenic potential. So, it was hypothesized that the missing link might be a DNA repair process using the NGEFTs as a template and thereby driving the repair process into an aberrant direction [28, 128]. But there is no experimental evidence proving it until today.

1.10. The DNA double strand break response

In case of a DNA double strand break in eukaryotic cells, there are basically two different repair mechanisms, one is the non-homologous end joining (NHEJ) and the other is the homologous recombination (HR). HR normally results in a perfect repair, while defects in proteins involved in HR lead to genomic instability [129-131]. Because NHEJ joins two strands without the usage of a template it is error prone and can lead to duplications, deletions, insertions, translocations and other chromosomal aberrations. Interestingly, all chromosomal translocations associated with leukemia do not share homologous sequences in their breakpoint cluster regions. *KMT2A*-translocations normally arise through NHEJ and therefore have no specific sequence motif at their breakpoints [39-42]. This is not true for *KMT2A* partial tandem duplications because they undergo Alu-Alu mediated recombination in most cases [132]. The NHEJ DNA repair mechanism is

not limited to a specific phase of the cell cycle but mostly occurs in G0/G1. This repair type is not capable of accurately repairing a DNA double strand break in the case that small parts of the sequence at the break point are missing [133]. In contrast to that, the HR is limited to the late S- and the G2-phase of the cell cycle [134], because the mechanism needs the sister chromatid as a template to repair the DNA double strand break and this is only available during these phases [135].

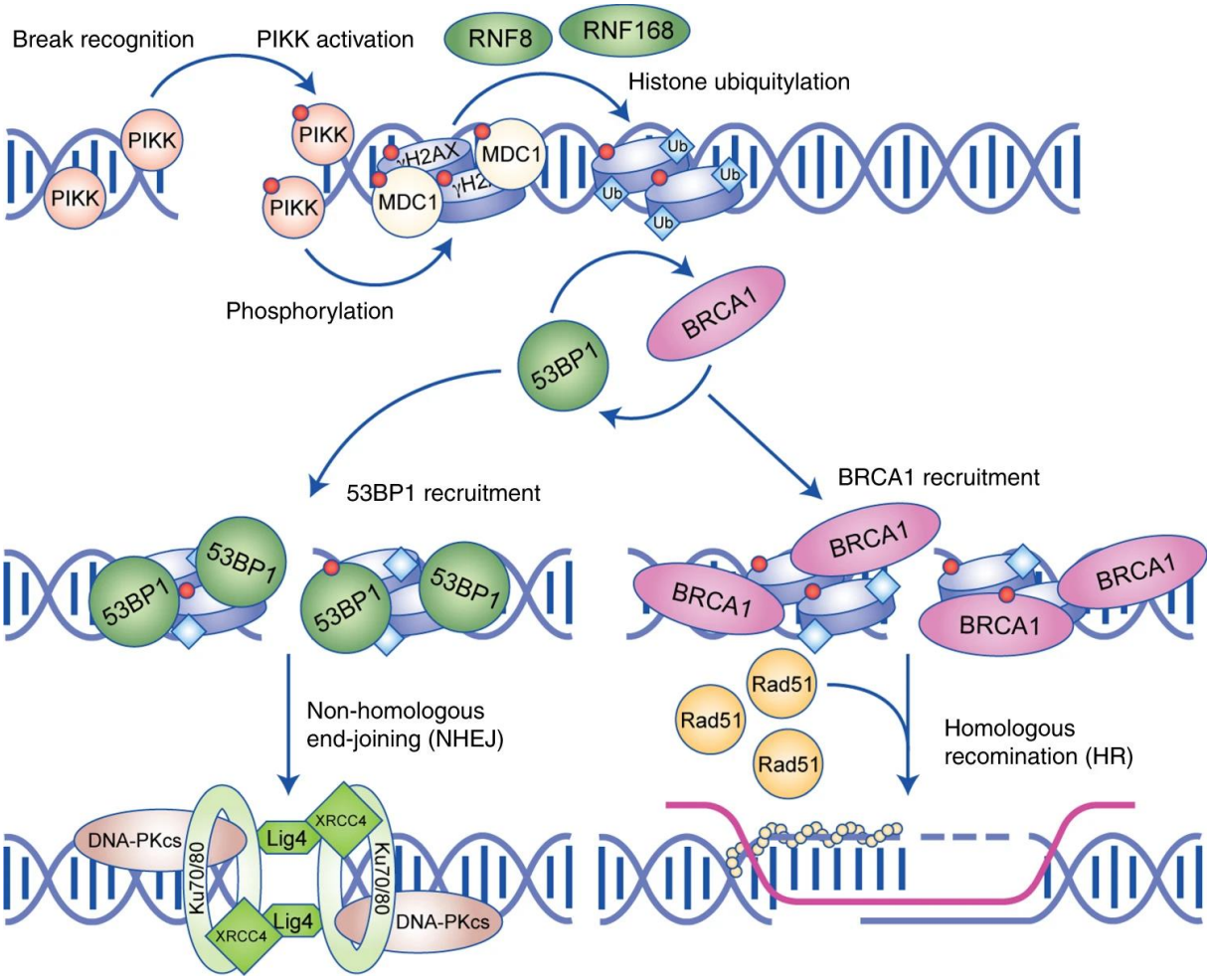

Figure 4. Diagram of the DNA double strand break (DSB) repair cascade. The recognition of the DSB recruits and activates phosphatidylinositol 3-kinase-like kinases (PIKKs). The activated PIKKs then phosphorylate the early cascade proteins H2AX and MDC1. The ubiquitin ligase RNF8 is recruited to phosphorylated MCD1 and initiates the ubiquitylation of histone H1, recruiting RNF168 which is ubiquitylating H2A. The modified chromatin acts as a marker for the recruitment of 53BP1 or BRCA1, either facilitating NHEJ or HR. NHEJ is dependent on the recruitment of the DNA-PK complex (Ku70/80 bound to DNA-PKcs) and end-processing factors such as XRCC4 and DNA Lig4. HR is dependent on resection of the broken DNA to produce a

single-stranded DNA overhang that gets coated in a RAD51 filament and invades the sister chromatid to facilitate templated repair. [136]

1.10.1. The role of RNA in DNA repair

The evidence suggesting that RNA plays a significant role in the repair of DNA damage through currently unresolved mechanisms is growing and gets increasingly important [137-144]. It is becoming more and more obvious that transcription, RNA-interacting proteins, RNA-processing proteins and RNA itself plays a big role in the repair of DNA damage [145-148] and that understanding the contribution of RNA in the DNA damage response will provide a new level of insight into genome maintenance. A role of nuclear RNAs could be shown in experiments where RNase A treatment 20 min after irradiation significantly impaired 53BP1 formation, while the formation of H2AX foci was unaffected. In a rescue experiment this could be reverted by the addition of nuclear RNA extracted from other cells [139, 149], suggesting a direct role of RNA in the recruitment of repair factors. It is proposed that RNA precursors are processed into an active form, called DNA-damage response RNA, which then has a role in the subsequent repair mechanism [139, 140]. Further experiments using transcription inhibitors have shown that the inhibition of transcription in short periods before damage induction reduces the recruitment of NHEJ- and HR specific factors, suggesting a direct role for transcription in their recruitment [139, 150].

Because there is a strong interaction between RNA-processing enzymes and R-loops, while DNA-RNA hybrids connect RNA- and DNA-related processes, the role of RNAs contributing to the DNA repair process is extensively investigated [143, 144, 151, 152]. Actually, R-loops were in the past considered to be a source of genomic instability [153-155], but were recently shown not to be the source of instability [156]. This is why they are now increasingly implicated in the repair of DNA damage and preservation of genome stability [143, 150, 157]. It has been shown that in close proximity to DNA breaks R-loops were generated in response to the damage dependent on the transcriptional activity of the locus [143, 144]. Core components of the DNA damage response have been shown to interact with R-loops which implicates R-loops being part of the canonical DNA repair process rather than simply being a part of a distinct RNA-dependent DNA repair [150, 152, 158]. There are two theories for the origin of RNA

involved in DNA repair, it is either the result of transcription that occurs after the break induction and produces RNA of the broken strand or the utilisation of a transcript that was produced prior to break induction and remains in the vicinity of the break. Experiments using highly expressed reporter systems and inducible breaks could show the bi-directional transcription of damage-induced long non-coding RNAs [138, 140, 146, 159-161] which could not be verified by experiments using endogenous break sites in mammalian cells [140, 143, 160]. The usage of RNA transcripts produced prior to damage induction could be demonstrated to occur in an NHEJ-dependent manner in yeast [43-45] and the theory has gained popularity as transcripts prior to DNA damage have the potential to serve as template for the repair. Additionally, this mechanism would be similar to HR but would be independent of the cell cycle because the sister-chromatid is not needed. RNA-templated DNA repair has also been proposed for terminally differentiated cells, for example neurons which are not rapidly replaced and therefore need a mechanism of high-fidelity DNA damage repair [45, 150]. Taken together, recruitment of RNA-processing enzymes, repair factors and R-loop formation seem to be transcription level dependent and RNA-templated DNA repair could provide a high-fidelity mechanism to repair DNA damage independent from the cell cycle. Therefore, this whole process can function as a safeguarding mechanism for the most active regions of the genome.

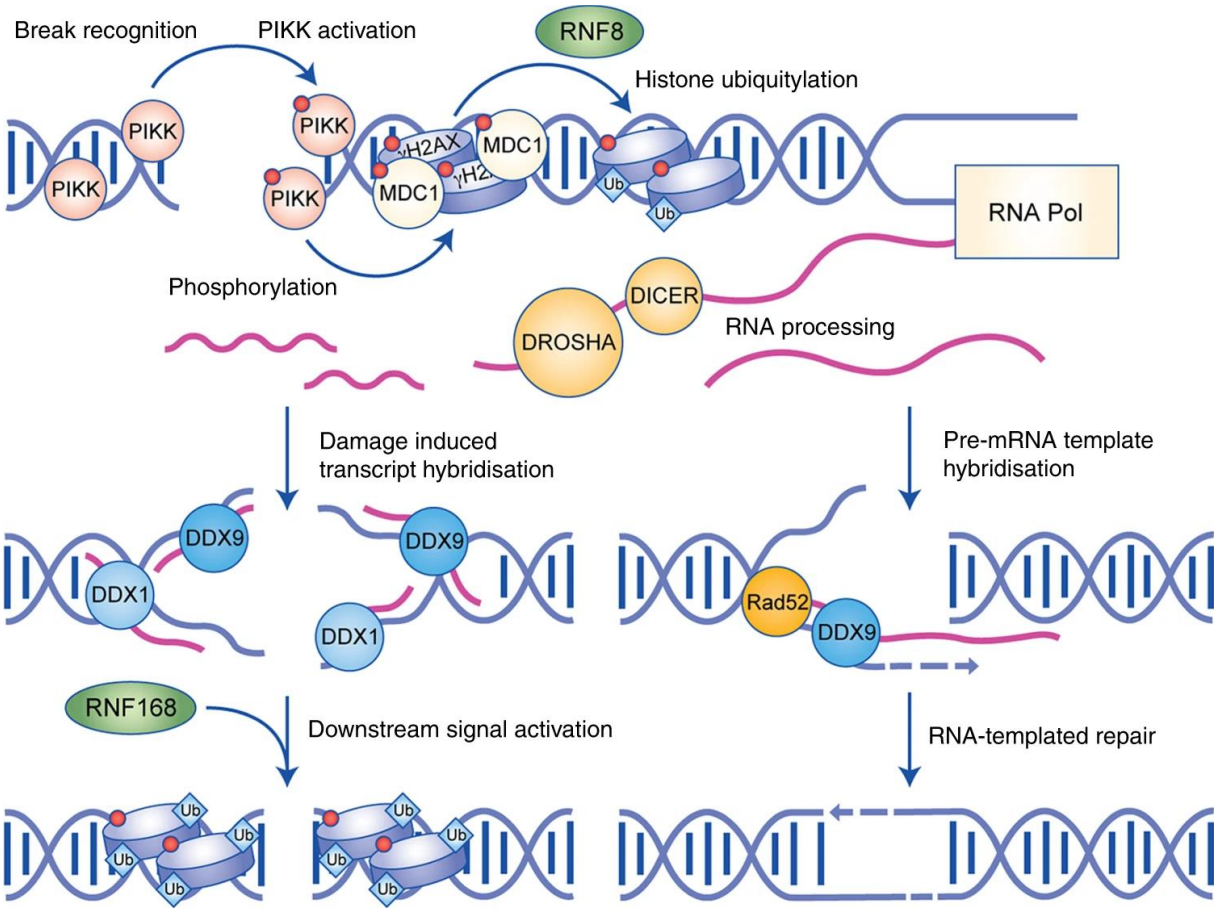

Figure 5. Scheme of RNA-dependent DNA repair. The DNA damage repair is initiated in the same way as the canonical double strand break repair but at the point of the ubiquitylation cascade, DNA-damage response RNAs are produced. These DNA-damage response RNAs could either hybridise as small RNAs with the broken DNA via the aid of helicases such as *DDX1* and *DHX9* or as long RNA molecules bridging the break by hybridising to the DNA via a *RAD52*-dependent strand invasion. Either the small RNA acts as a signal for canonical double strand break repair pathways or the long RNA molecule is used for a templated high fidelity mechanism [136].

1.11. Nuclear structure, transcription factories and translocations

Although specific recurring chromosome aberrations are often associated with a particular type of leukemia, lymphoma or other type of cancer, the underlying molecular mechanism of a translocation event are generally unknown [162]. Most chromosomal translocations are the result of the exchange of large chromosomal segments after DNA double strand breaks occurred. This could lead to chimeric genes at the fusion points

that code for hybrid proteins with altered functions, like in the case of *KMT2A::AFF1*. Microscopic approaches have documented different subnuclear functions, such as morphologically recognizable blocks of heterochromatin, the existence of subnuclear bodies and the organization of DNA into chromosome territories [163, 164]. Further experiments based on chemical cross-linking of chromatin have demonstrated the existence of chromatin loops and domains as ubiquitous feature in the organization of the genome [165, 166]. These experiments showed that during the interphase each chromosome occupies a defined region known as chromatin territory [27] and that this position is non-random [167]. More recent studies point out that many genome features are characterized by a high degree of freedom and heterogeneity. For example, differences in the three-dimensional position, functions of the two alleles and that many chromatin-chromatin interactions only occur in low frequencies within a small subpopulation of cells [168-170]. This seems to be paradox, because territories seem to be non-random pointing to the existence of regulatory mechanisms, but the high variability suggests that these features are not hard-wired or required for correct genome function [171].

Chromatin fibers are made up of DNA wrapped around nucleosomes that consist of octamers of core histone proteins [172]. The chromatin fiber then self-interacts to form loops, these chromatin loops are a very common feature of the genome in most organisms and they exist in sizes from kbs to Mbs [173]. The next level of genome organization are chromatin domains, first described by *in situ* hybridization [29] and confirmed by cross-linking approaches [174-176]. These individual chromatin domains assemble into higher-order chromatin compartments, based on interaction and imaging studies (see Figure 6) [177-179]. Chromosomes represent the major unit of genome organization which, during the interphase, occupy a relatively compact, spatially restricted territory of the nucleus [29]. The interior of these territories is thereby permeated by a network of channels to facilitate the access of regulatory factors. Furthermore, the chromatin fibers of neighbouring chromosomes often intermingle at their periphery, creating a chromatin with lower density at the chromosome interfaces [28, 29]. The last level of genome organization is the location of the chromosomes in the three-dimensional space of the nucleus. This was carried out by measurements of chromosomes and genes relative to the nuclear center or periphery and the position relative to other chromosomes or loci [180, 181]. It has been shown that inactive genes

often locate near the nuclear edge and associate with peripheral heterochromatin, while active genes are frequently located in the nuclear interior [180]. As demonstrated by live-cell imaging, a chromatin locus undergoes diffusional motion [182] and also very rare long-range directed motion of chromatin loci has been observed, dependent on nuclear actin and myosin [183, 184].

Although most of the organization shows non-randomness and the organization patterns are tissue and cell-type specific, the precise role of the three-dimensional positioning remains unclear [185-187].

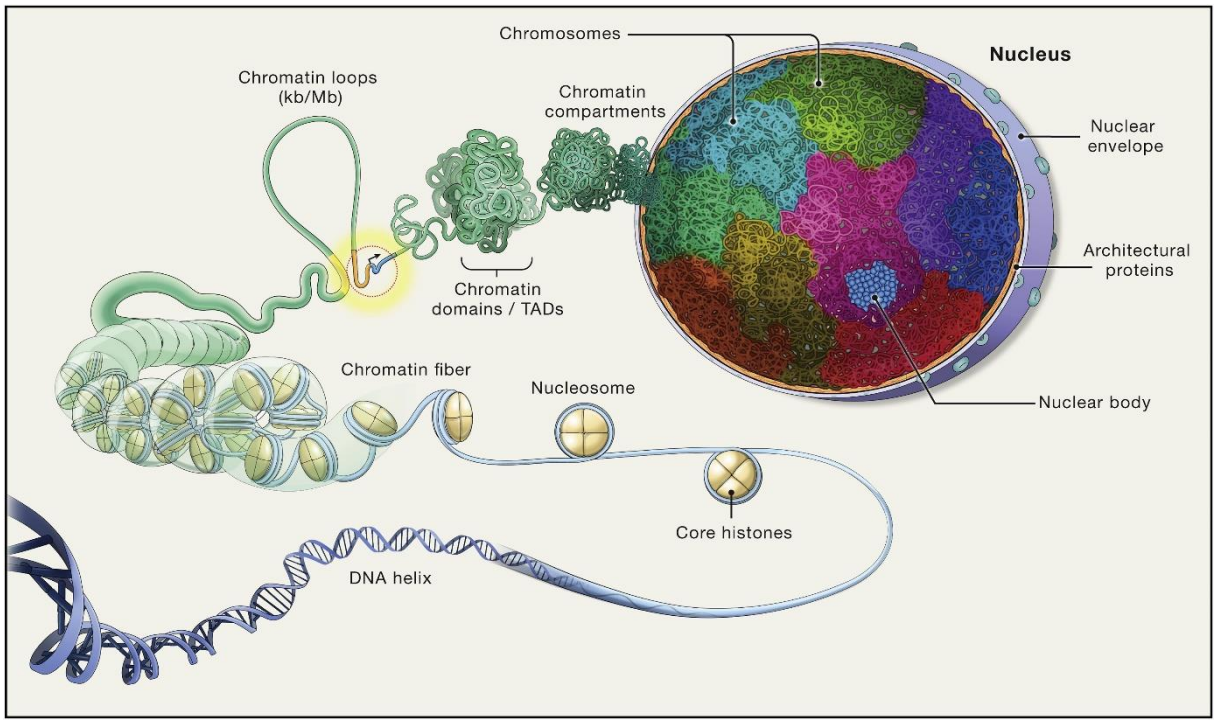

Figure 6. Organization of the eukaryotic genome. The DNA is wrapped around the nucleosome, consisting of eight core-histones, forming the chromatin fiber. Chromatin fibers fold into loops, often bringing regulatory elements (yellow), like enhancers, into proximity of promotor genes (gold/blue), controlling their transcription (black arrow). The fibers fold into chromatin domains, which associate with each other and create chromatin compartments. Each chromosome occupies its territory within the nucleus, generating non-random patterns of chromosomes. In the DNA-free space, the nucleus also contains RNA and protein aggregates which form nuclear bodies (blue). [171]

This non-randomized organization of the genome but also the high level of variability plays an important role in the formation of chromosomal translocations. The spatial arrangement influences the proximity of genes in specific cell lines and the variability in

the periphery enables the contact of two gene from different chromosomes, which could lead to illegitimate repair of two DNA double-strand breaks. There are two main hypothesis that try to explain the proximity of damaged loci resulting in translocations. The “contact first” hypothesis proposes that the translocation partners are in close proximity, which is key, before the event leading to the double-strand break [37]. This proximity between two loci can either be permanent [188] or temporary [35]. This colocalization may occur in specific compartments, such as replication foci, transcription factories or others [31, 32, 189-193]. The hypothesis states that broken chromosome ends have a reduced mobility, maximizing the probability to translocate with neighbouring loci. The “breakage first” hypothesis on the other hand postulates that chromosome ends can freely move within the nuclear space after a double-strand break and this may lead to translocations when colliding with other damaged loci. The probability of a translocation is increased with the area of movement, either by directed motion or diffusion [38]. Published data on mobility of double-strand breaks are quite controversial in higher eukaryotes. Mobility of double-strand breaks did not differ from that of intact loci [194-196], very restricted movement of damaged loci was observed, while large-scale movement was a minority. But there were also experiments showing the opposite, where movement could already be seen five minutes after damage induction with more large-scale movement [197].

The so called transcription factories could be an indication for the “contact first” hypothesis. These transcription factories where first found in 1996, when Cook and colleagues noticed, that after incorporating labelled nucleotides the resulting labelled nascent transcripts were not distributed throughout the nucleus, but were concentrated in discrete focal sites detectable by electron and light microscopy [31]. They can also be made visible by immunostaining phosphorylated RNA polymerase II [30] and the spatial colocalization of actively transcribing gene loci and foci of Ser2- and Ser5-phosphorylated RNA polymerase II can be confirmed by immuno-RNA FISH [198-200]. The amount of transcription factories per cell is very different depending on the cell type, differentiation state and method from approximately 100 – 8000 per nucleus, but among these different cells the density of 8 - 9 factories/μm^3, an approximate distance of < 1 µm between two factories, remains almost constant [201]. By BrUTP incorporation ~1500 factories could be observed in human erythroblasts [198], whereas in mouse erythroid cells just a few hundred factories could be observed via Ser5P-RNA polymerase II

immunostaining [200]. As a consequence of these immobilized transcription factories, the transcriptional machinery is not recruited to the gene loci moving along the template, the gene loci move to the immobilized RNA polymerase II already present in a factory (Figure 7) [31, 32]. The fact that transcription factories can bring two genes into spatial proximity also has an influence on genome stability, for example in the case of double strand breaks and the subsequent NHEJ during DNA repair, where different chromosomes could be joined together resulting in a chromosomal translocation. It has been shown for many different genes associated with chromosomal translocations in different kinds of cancer, that they are transcribed in the same transcription factory, thereby creating the spatial proximity for an illegitimate repair. That they are transcribed together in the same transcription factory has been shown for genes like *KMT2A* and *AFF1* [22-24], *BCR* and *ABL* [33, 34] and *cMYC* and *IgH* [35, 36].

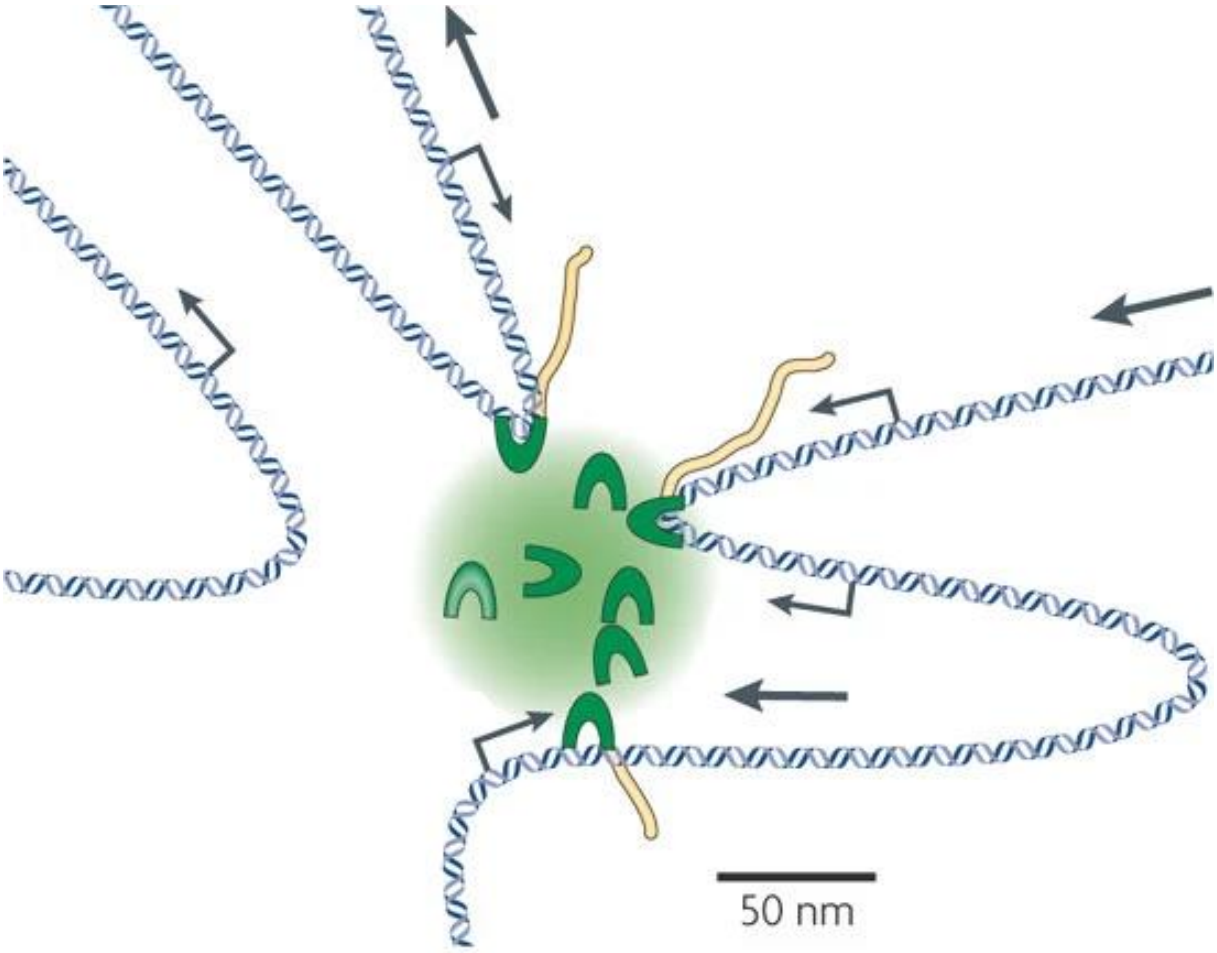

Figure 7.Transcription factory. The transcription factory has a diameter of ~70 nm and contains eight RNA polymerase II (green crescents). The genes are reeled through the

polymerases in the direction of the arrows as they are transcribed into nascent RNA (yellow). Genes from the same of from different chromosome may associate with the same transcription factory. [202]

1.12. Cas9 protein

The Cas9 protein is a bacterial immune protein, protecting the cell from invading nucleic acids. Investigations of the bacterial adaptive immune system, also called "clustered regularly interspaced short palindromic repeats" (CRISPR) system, led to the discovery of the RNA-guided DNA nuclease Cas9 [203, 204]. Cas9, in its biological context, is part of a type II CRISPR interference system that degrades pathogenic phage or plasmid DNA. The targeting of the Cas9 protein is enabled by CRISPR-RNAs (crRNAs), which recognize 20 bp of complementary DNA. The Cas9 protein itself also binds a short DNA sequence adjacent and opposite the protospacer called the protospacer adjacent motif (PAM). There is a significant variation in PAM sequences depending on the host bacteria, the most used Cas9 from *Streptococcus pyogenes* (SpCas9) recognizes the PAM 5´-NGG-3´ [205]. Upon target recognition, the two nuclease domains, RuvC and HNH, engage and cleave the separated strands of DNA between 3 and 4 bp upstream of the PAM site [206]. A second RNA, called *trans*-activating RNA (tracrRNA), is also needed for Cas9 maturation and activity. In experiments it could be shown that crRNA and tracrRNA can be fuse together to form a single guide RNA (sgRNA) [206]. The easy programmability of the Cas9 protein by the expression of a sgRNA has driven the rapid success in Cas9-based genomic engineering, as it can target any DNA locus by simply changing the sgRNA sequence. The interest in genetic engineering by Cas9 has led to a number of different methods to change or improve Cas9 functions. Different point mutations in the RuvC and HNH nuclease domain transformed the Cas9 endonuclease activity into a nickase, inducing single strand breaks, or a catalytically dead mutant (dCas9), that can be used for DNA binding and for example function as transcription inhibitor (CRISPRi) [207, 208]. Dead mutant Cas9 (dCas9) has also been fused to other factors, to recruit alternative factors to specific DNA loci, for the improvement of genome editing [209-211]. Computational analysis comparing different co-variant models based on the sequence and secondary structure of their tracrRNA showed seven distinct clusters,

indicating the possibility to use different Cas9 orthologues within the same system, by using different tracrRNA fragments for the corresponding sgRNA [212].

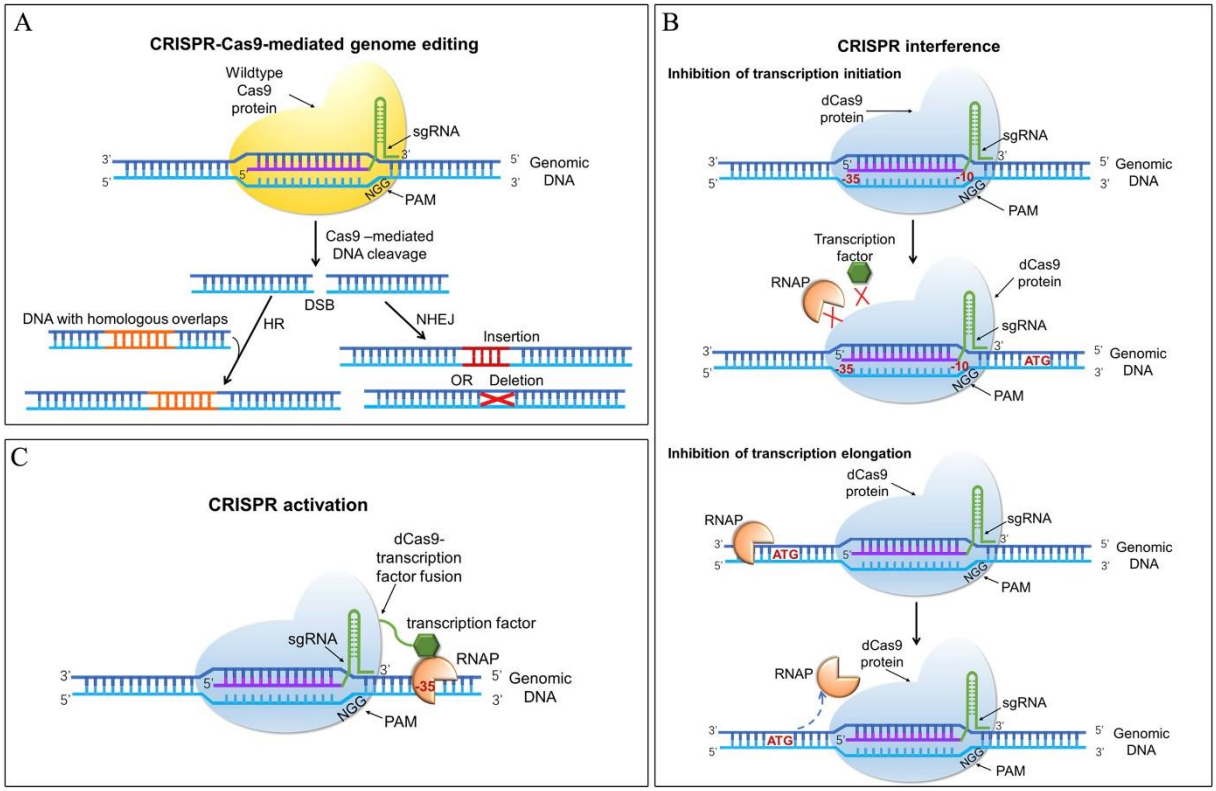

Figure 8. Applications of the CRISPR/Cas9 system in genetic engineering. A) Cas9 protein complex with sgRNA binds the target DNA sequence, creating a double-strand break that is repaired by NHEJ or HR. By providing a donor sequence sharing a homology to the DNA ends, specific sequences can be integrated into the genomic DNA by HR. NHEJ repair leads to random insertions or deletions. B) A dead Cas9-sgRNA complex can be used to target the promotor or enhancer sequence to block RNA polymerase or transcription factors, thereby inhibiting the transcription initiation. The dCas9 can also be used to target the gene sequence to prevent the transcription elongation. C) dCas9 can also be fused to transcription factors, to deliver the transcription factor to the promotor to enhance transcription. [213]

1.13. Dimerization-system using dCas9 and the abscisic acid pathway

To induce gene proximity between genes that are not in vicinity in HEK293T cells, a DNA-binding and dimerizing system was used and established. This dimerization system was planned for reversible chromatin loop reorganization and called chromatin loop reorganization using CRISPR-dCas9 (CLOuD9) system [214]. The system uses two dead

Cas9 proteins from different species so they recognise different scaffold RNAs and can be used together in the same cell. One dCas9 from *Staphylococcus aureus* (dSaCas9) and one from *Streptococcus pyogenes* (deSpCas9). These dead Cas9 variants can bind the DNA without cutting and by redesigning the corresponding sgRNA they can bind anywhere in the genome. For the dimerization the abscisic acid pathway from plants was used and engineered [46]. The chemically induced proximity (CIP) is a tool for controlling and studying biological processes like receptor function, signalling, protein secretion, protein splicing and others. CIP uses membrane-permeable, small-molecule inducers to control dimerization between proteins of interest that are fused to the inducer-binding proteins.

To get a more broadly useful CIP system, some focused on the plant phytohormone *S*-(+)-abscisic acid (ABA). The ABA signalling pathway in plants is a response for abiotic stress and developmental decisions in plants [215]. Thereby ABA binds to the pyrabactin resistance (PYR)/PYR1-like (PYL)/regulatory component of the ABA receptor (RCAR) and the ABA-PYL1 complex then dimerizes with the abscisic acid insensitive protein I (ABI1). This pathway was engineered to have a proximity inducing tool by the addition of the non-toxic ABA. In the engineered CIP system PYL1 is the regulatory domain of the ABA receptor and ABI1 is the abscisic acid insensitive I protein without its phosphatase domain [46]. This ABA-CIP system can then be fused to various other proteins in order to induce proximity. In order to bind the DNA, the ABA-system was fused to the two dCas9 variants, to get a tool for inducing chromatin proximity. In first experiments researchers used the CLOuD9 system to re-establish the β-globin gene expression in K562 cells by targeting the promotor of the β-globin locus and the HS2 region of the locus control region. They could observe that the dimerization induced a conformational change in the chromatin that was even reversible by ABA washout. Additionally, they could see , that β-globin expression was significantly increasing in the presence of ABA.

To transfer the system in the context of *KMT2A* and *AFF1* or *KMT2B* and *CCND3* proximity, the sgRNAs needed to be adapted. In order to do that, two different systems were set up, one targeting *KMT2A* intron 9 and *AFF1* intron 3 and the other targeting *CCND3* intron 2 and *KMT2B* intron 13 (Figure 9) and a non-targeting control system with two sgRNA that are not binding in the human genome. The systems were constitutively expressed in HEK293T cells, binding to their target sequence (Figure 9 A). Proximity is

induced by ABA treatment over 14 days, leading to the dimerization of the ABA-CIP and remodelling of the chromatin (Figure 9 B).

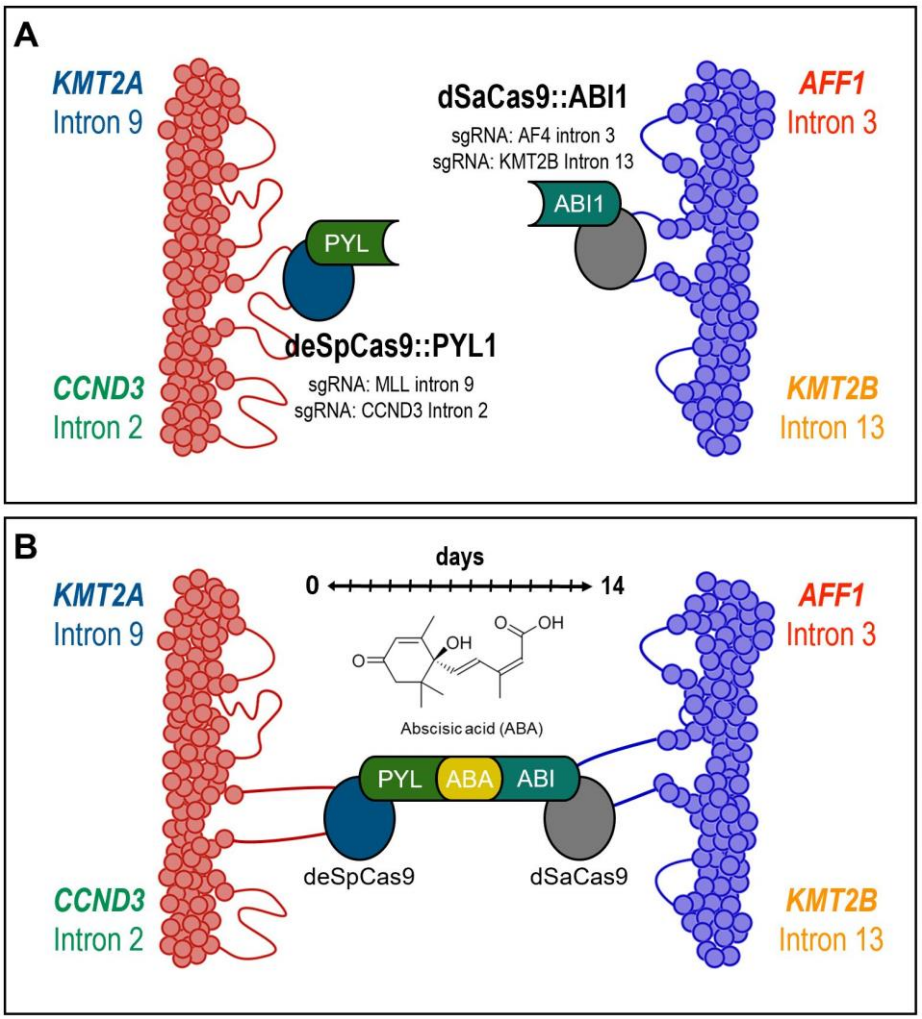

Figure 9. Overview of the proximity induction. A) The contact area between two different chromosomes is packed with chromatin loops. The ABA-CIP system is binding to the chromatin, deSpCas9 fused to PYL1 (deSpCas9::PYL1) either targets *KMT2A* intron 9 or *CCND3* intron 2 and dSaCas9 fused to ABI1 (dSaCas9::ABI1) either targets *AFF1* intron 3 or *KMT2B* intron 13. B) In the presence of abscisic acid, the fusion proteins dimerize pulling *KMT2A* intron 9 and *AFF1* intron 3 or *CCND3* intron 2 and *KMT2B* intron 13 into proximity.

1.14. Aim of this work

KMT2A-rearrangements account for about 5–10 % of all ALL and AML cases in children and adults caused by chromosomal translocations [1]. Due to its bad prognosis and treatability, t(4;11)-leukemia is classified as a high risk leukemia [2, 3]. Further, the balanced t(4;11) leads to the derivative chromosomes der(4) and der(11) which express

the two fusion transcripts *KMT2A::AFF1* and *AFF1::KMT2A*, leading to the oncogenic fusion proteins. It has been shown that in the blood of healthy individuals a *KMT2A-ex9::AFF1-ex4* fusion transcript can be found, in the absence of the reciprocal transcript and the underlying chromosomal translocation [5]. This phenomenon has also been found for fusion transcripts of other frequent translocations like *BCR::ABL* [6, 7], *TEL::AML1* [8, 9], *AML1::ETO* [9], *PML::RARα* [10], *NPM::ALK* [11, 12] and *ATIC::ALK* [12]. It has been hypothesized that these non-genomically encoded fusion transcripts (NGEFTs) are formed by *trans*-splicing of the so called PTTs that are expressed by many genes undergoing chromosomal translocation [26] and could be a reason for the illegitimate DNA repair resulting in a translocation. In addition to that many genes undergoing translocations have been shown to be in close proximity in the interphase nucleus or are even transcribed in the same transcription factory [22-24].

The PhD project termed "The molecular principles of chromosomal translocations" aimed to investigate the prerequisites that are necessary to form a chromosomal translocation after a DNA damage. To achieve that, the influence of gene proximity, DNA damage and the *KMT2A-ex9::AFF1-ex4* fusion transcript on the formation of chromosomal translocations should be investigated. Additionally, the influence of an artificial fusion transcript should also show whether the RNA is a necessary component of the repair process leading to a chromosomal translocation.

2. Materials

2.1. Antibiotics

Table 1. List of used antibiotics

Antibiotics	**Company**
Ampicillin	Roth
Blasticidin S	AppliChem
Doxycyclin	Sigma-Aldrich
Hygromycin B	Invitrogen
Kanamycin	Roth
Penicillin/Streptomycin	GE Healthcare
Puromycin	PAA

2.2. Bacteria

Table 2. Bacteria

Bacteria	**Genotype**
E. coli DH5α	F$^-$φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdkR17(rκ$^-$, mκ$^+$) phoA supE44 λ$^-$thi-1 gyrA96 relA1
E. coli NEB® Stable	F' *pro*A+B+ *lac*I^q Δ(*lac*Z)M15 zzf::Tn10 (Tet^R)/Δ(*ara-leu*) 7697 *ara*D139 *fhu*A Δ*lac*X74 *gal*K16 *gal*E15 e14- Φ80*dlac*ZΔM15 *rec*A1 *rel*A1 *end*A1 *nup*G *rps*L (Str^R) *rph spo*T1 Δ(*mrr-hsd*RMS-*mcr*BC)
E. coli Top 10 F´	F'[*lac*I^q, *Tn*10(Tet^R)] *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L(Str^R) *end*A1 *nup*G

E. coli XL 10-Gold	Tet^{r}Δ(*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 *end*A1 *sup*E44 thi-1 *rec*A1 *gyr*A96 *rel*A1 *lac*Hte[F' *pro*AB *lac*I^{q}ZΔM15 Tn*10*(Tet^r) Amy Cam^r]

2.3. Buffer

Table 3. Buffers and Solutions

Buffers and Solutions	**Composition**	
1x PBS (pH7.4)	137 mM	NaCl
	2,7 mM	KCl
	10 mM	Na_2HPO_4
	1,76 mM	KH_2PO_4
		in dd H_2O
1x SDS sample buffer	125 mM	Tris-HCl (pH 6.8)
	20 %	Glycerol
	4 %	SDS
	0.8 %	DTT
	0.01 %	Pyronin Y
		in dd H_2O
5x SDS sample buffer	312.5 mM	Tris-HCl (pH 6.8)
	50 %	Glycerol
	5 %	SDS
	2 %	DTT
	0.025 %	Pyronin Y
		in dd H_2O
10x TBE	1 M	Tris-HCl
	500 mM	H_3BO_3
	20 mM	EDTA
10x TBS (pH 7.5)	100 mM	Tris-HCl
	150 mM	NaCl
		in dd H_2O

Buffers and Solutions	**Composition**	
Buffer 1 (Mini-Prep.)	50 mM	Tris-HCl (pH 8.0)
	10 mM	EDTA
Buffer 2 (Mini-Prep.)	1 %	SDS
	200 mM	NaOH
Buffer 3 (Mini-Prep.)	3 M	Potassium acetate (pH 4.8)
10x Cas9 Reaction Buffer	200 mM	HEPES (pH 6.5)
	1 M	NaCl
	50 mM	$MgCl_2$
	1 mM	EDTA
Cell culture Medium HEK293T cells		DMEM Low-Glucose
	10 %	FBS
	1 %	Penicillin/Streptomycin
	1 %	L-Glutamine
DNA loading dye	50 %	Saccharose
	100 mM	EDTA
	0.1 %	Bromophenol blue
	0.1 %	Xylene-cyanol-FF
Ethidium bromide solution	10 mg/ml	Ethidium bromide
Cell culture freezing solution	90 %	FBS
	10 %	DMSO
IPTG	100 mM	IPTG
LB agar		LB-Medium
	15 g/l	Agar
LB medium (pH 7.0)	10 g/l	Bacto tryptone
	5 g/l	Yeast extract
	5 g/l	NaCl
LB/Ampicillin medium		LB-medium
	100 µg/ml	Ampicillin
LB/Ampicillin/IPTG-plates		LB-Medium
	100 µg/ml	Ampicillin
	500 nM	IPTG

Buffers and Solutions	**Composition**	
Low-TE	10 mM	Tris-HCl (pH 8.0)
	100 nM	EDTA
Modified RIPA lysis buffer	4 mM	cOmplete Mini
	1 mM	NaF
	0.3 mM	Na_3VO_4
	1 mM	PMSF
	3 mM	β-glycerophosphate
	10 mM	Sodium pyrophosphate
	0.12 %	H_2O_2
		RIPA buffer (freshly added)
RIPA buffer	50 mM	Tris-HCl (pH 7.5)
	150 mM	NaCl
	0.1 %	SDS
	1 %	Nonidet P-40
	0.25 %	Sodium deoxycholat
RIPA lysis buffer	4 mM	cOmplete Mini
	1 mM	NaF
	0.3 mM	Na_3VO_3
	1 mM	PMSF
		RIPA buffer (freshly added)
RLN buffer	50 mM	Tris-HCl (pH 8.0)
	140 mM	$NaCl_2$
	1.5 mM	$MgCl_2$
	0.5 %	Nonidet P-40
RSB buffer	10 mM	Tris-HCl (pH 7.4)
	10 mM	NaCl
	3 mM	$MgCl_2$
	(0.5 %	Triton X-100)
ST Buffer (Mini-Prep.)	40 g	Saccharose

Buffers and Solutions	Composition	
	25 ml	Triton X-100
	50 ml	EDTA (0.5 M)
	25 ml	Tris-HCl (pH 8.0; 1 M)
	ad 500 ml	VE H_2O
TBS-T (pH 7.5)	100 mM	Tris-HCl
	150 mM	NaCl
	0.1 %	Tween 20
		ddH_2O
TE	10 mM	Tris-HCl (pH8.0)
	1 mM	EDTA
TENS buffer	100 mM	Tris-HCl (pH 8.0)
	40 mM	EDTA
	200 mM	NaCl
	1 %	SDS
X-Gal	100 mg	X-Gal
	2 ml	Dimethylformamide (DMF)
Abscisic Acid	2 µM	Abscisic Acid in Methanol

2.4. Cell lines

Table 4. Cells

Cell Line	Cell type	DSMZ number
HEK 293 T	highly transfectable derivative of the human primary embryonal kidney cell line 293 (ACC 305) carrying a plasmid containing the temperature sensitive mutant of SV-40 large T-	ACC 635

	antigen (tsA1609); classified as risk group 1 according to the German Central Commission for Biological Safety (ZKBS)

2.5. Chemicals

Table 5. Chemicals and reagents

Chemicals and reagents	**Company**
β-Mercaptoethanol	Serva
λ-DNA cut with ClaI	MBI Fermentas
Abscisic acid	Sigma-Aldrich
Acetone	Roth
Agarose	Invitrogen
Bacto tryptone	Oxoid
Boracic acid	AppliChem
Bromophenol blue	Sigma-Aldrich
Cycloheximide	Merck
Disodium hydrogen phosphate	Merck
Dimethylformamide	Merck
Dimethyl sulfoxide	Sigma-Aldrich
Deoxyribonucleoside triphosphate	Roche
Dithiothreitol	Invitrogen
Ethylenediaminetetraacetic acid	Roth
Ethanol	Roth
Ethidium bromide	Sigma-Aldrich
Glycine	Roth
Hi-Di Formamide	Applied Biosystems
Isopropyl-β-D-thiogalactopyranoside	Fermentas
Isopropanol	Roth

Chemicals and reagents	Company
Magnesium chloride	Merck
Metafectene Pro	Biontex
Methanol	Sigma-Aldrich
NP-40	AppliChem
PB-Buffer	Qiagen
Potassium chloride	Merck
Potassium dihydrogen phosphate	Merck
RNAse-free water	Qiagen, New England Biolabs, Zymogen
Saccharose	Roth
Sodium dodecyl sulphate	Roth
Sodium acetate	Merck
Sodium chloride	Roth
Sodium hydroxide	Merck
Tris	Roth
Triton X-100	Sigma-Aldrich
X-Gal	Roth
Xylen-Cyanol-FF	Merck
Yeast extract	Oxoid
z-VAD-FMK	Enzo Lifescience

2.6. Enzymes

Table 6. Enzymes

Enzyme	Company
Cas9 NLS	PNA Bio
DNase	Qiagen
GoTaq Polymerase	Promega
HiScribe T7	New England Biolabs
KAPA HiFi HS RM	Roche
Lysozyme	Sigma-Aldrich

Proteinase K	Roth
RNase A	Roth
RNase-Free DNase kit	Qiagen
rRNasin	Promega
SuperScript II	Invitrogen
T4-DNA Ligase	New England Biolabs

2.7. Equipment and consumables

Table 7. Equipment

Equipment	**Company**
1 kb marker	New England Biolabs GmbH
100 bp marker	New England Biolabs GmbH
6-well plate	Sigma-Aldrich
10 cm Dish, TC dish 100 standard	Sarstedt
24-well cell culture plate Cellstar	Greiner bio-one
48-well cell culture plate Cellstar	Greiner bio-one
96-well cell culture plate Cellstar	Greiner bio-one
Abi Prism 310 Genetic Analyzer	Applied Biosystems
Adhesive sealing film	Nerbe Plus
Analytical Balance TE124S-OCE	Sartorius
Balance PM4800 DeltaRange	Mettler
Centrifuges	Eppendorf (5424, 5702 R) Beckman (J2-HS, J6-HC) VWR (Galaxy MiniStar)
ChemiDOC XRS+	BioRad
Co_2-Incubator HERAcell 240	Thermo Scientific
Cryo. S.Vials.	Greiner bio-one
Freezer Profi Line	Liebherr
Fridge BluPerformance	Liebherr
Fluorescence Microscope Observer Z1	Carl Zeiss

Equipment	Company
Gel chamber Easycast™ B2 Owl	Thermo Scientific
Get Documentation c200	Azure Biosystems
GeneAmp PCR System 9700	Applied Biosystems
Flow Bench	Waldner
Ice Machine	Ziegra
Incubator B6060	Heraeus
Laminar Flow HLB 2472	Heraeus
Mignetic Stirrer, RCTbasic	IKA Labortechnik
Microcentrifuge vials (1.5 ml)	Eppendorf, Greiner bio-one
Microwave Kor-6305	Daewoo
Multiplate PCR Plates, 96-well	Bio-Rad
N_2-Tank VHC 35	Taylor-Wharton
Nanophotometer P330	Implen
Parafilm M	Pechiney
PCR vials	Greiner bio-one
Petri dish	Greiner bio-one
pH-Meter	Inolab, Schott
Plate reader Infinite 200 Pro	Tecan Group Ltd.
Pipetteboy accu-jet pro	Brand
Pipettes 10, 100, 1000 µl	Eppendorf
Powersupply E865	Consort
Reaction tubes 1.5, 2.5, 5, 15, 50 ml	Greiner bio-one
Sapphire PCR tubes	Greiner bio-one
SeqStudio, genetic Analyzer	Applied Biosystems
Sequencing tubes, Tube-310-NC	Axygen Scientific
Shaker Incubator CH-4103	Infors AG
StepOnePlus Real-Time PCR System	Applied Biosystems
Sterile pipettes (5, 10, 25 ml)	Greiner bio-one
Sterile pipette tips	Thermo Scientific Sarstedt

Equipment	Company
	Greiner bio-one
TC10 automated cell counter	Bio-Rad
TC10 System counting slides	Bio-Rad
Thermomixer comfort	Eppendorf
Microscope	Nikon
Tubes (15, 50 ml), Cellstar	Greiner bio-one
Ultra-low temperature freezer, U725 innova	New Brunswick Scientific
UV-desk, TFP-35M	Fröbel Laborgeräte
Vacuum pump VAC-MAN™	Promega
Vortex-Genie 2	Scientific Industries
Water bath SW-20C	Julabo

2.8. Kits

Table 8. Kits

Kit	Company
AllPrep DNA/RNA mini	Qiagen
BigDye Terminator v3.1	Applied Biosystems
Clarity ECL Western Substrate	Bio-Rad
DNA Clean & Concentrator Kit	Zymo Research
dNTP-Set Long Range	Peqlab
GoTaq Long PCR Master Mix	Promega
ORA qPCR Green Rox H Mix	highQu
PCR Purification Kit	Qiagen
pGEM-T Vector System	Promega
PureYield Plasmid Midiprep	Promega
QIAamp DNA Blood Mini Kit	Qiagen
QIAEX II Gel Extraction Kit	Qiagen
QIAshredder	Qiagen

RNA Clean & Concentrator Kit	Zymo Research
RNeasy Mini Kit	Qiagen
TOPO blunt-end Kit	Invitrogen
TA Cloning Kit	Invitrogen

2.9. Media and supplements

Table 9. Media and supplements

Media and supplements	**Company**
Accutase	GE Healthcare
DMEM Low Glucose	GE Healthcare
Fetal Bovine Serum	GE Healthcare
L-Glutamine (200 mM)	GE Healthcare
Phosphate buffered saline	GE Healthcare
Penicillin/Streptomycin	GE Healthcare
Trypan blue (0.4 %)	Gibco

2.10. Software and online-tools

Table 10. Software and online-tools

Software	**Company or Publisher**
ABI Prism 310 Genetic Analyzer Data Collection 3.1.0	Applied Biosystems
BLAST, Basic Local Alignment Search Tool	National Center for Biotechnology Information, U.S. National Library of Medicine
BLAST, Basic Local Alignment Search Tool	Ensemble
EndNote	Alfasoft GmbH
ImageJ 1.48v	National Institute of Health, USA
Image Lab Software 3.0.1	Bio-Rad
Micro-Manager 1.4.22	University of California, San Francisco
NEBioCalculator	New England Biolabs GmbH

Microsoft Office	Microsoft
SeqStudio Plate Manager 1.1.0	Applied Biosystems
Sequencing Analysis Software 6	Applied Biosystems
StepOne Software v2.3	Applied Biosystems
Tecan i-control 2.0	Tecan Group Ltd.

2.11. Primers

Table 11. Oligonucleotides.

Oligonucleotides	
MLL-I9	5'-CACCGTACTCTTGAGAAGCTGAGGCAG-3'
AF4-I3	5'-CACCGTCTAATAAATGGCTTGACTTG-3'
CCND3-I2	5'-CCCTTGGTAGGATATCAGGG-3'
KMT2B-I13	5'-GGGGCTCTCAGGAGGAGCAG-3'
NT-I	5'-CACCGGTAATCGCTAGCTTGTCGACT-3'
NT-II	5'-CACCGGTAATCGCTAGCTTGTCGACT-3'
MLL.I9.CCC.2.F	5'-TTGACCCCAACATCCTTTAGCAA-3'
AF4.I3.CCC.Guide2.2.R	5'-TTCCAGGGCTGAGAGGGGAAA-3'
CCND3.CCC.1.HindIII.F	5'-ATAGCCTGGGGTGGGGTCAT-3'
KMT2B.CCC.1.HindIII.R	5'-GTGTATGCAGGACAGCGAGGG-3'
AF4.3	5'-GTTGCAATGCAGCAGAAGCC-3'
AF4.5	5'-ACTGTCACTGTCCTCACTGTCA-3'
8.3	5'-CCCAAAACCACTCCTAGTGAG-3'
13.5	5'-CAGGGTGATAGCTGTTTCGG-3'
CCND3.E2.2.F	5'-TTCCCCCTGGCCATGAACTAC-3'
CCND3.E3.1.R	5'-AGGTCCCACTTGAGCTTCCC-3'
KMT2B.E8.1.F	5'-ATTCGGATGACTCGGAGCCC-3'
KMT2B.E15.1.R	5'-CATGCACCCAGTGATCGCAC-3'
CCND3.E4.2.R	5'-GCGGGTACATGGCAAAGGTAT-3'
AF4.cut.guide.F	5'-taatacgactcactataGGGGGGGCGTGATTCCTACCgttttagagctagaaatagc-3'

Oligonucleotides	
MLL.cut.guide.F	5’-taatacgactcactataGGAGCTCCTTATAGATGAAGgttttagagctagaaatagc-3’
CCND3.cut.guide3.F	5’-taatacgactcactataGGGAGTAGGAGGCCAGAGTTgttttagagctagaaatagc-3’
KMT2B.cut.guide6.F	5’-taatacgactcactataGGAGACGGGCGAGGAGGAATgttttagagctagaaatagc-3’
Universal Reverse Primer	5’-AGCACCGACTCGGTGCCACT-3’
KMT2B.E9.T7.F	5’-TAATACGACTCACTATAAGCTGCCACTGCCAGAACCTGAGGAGC-3’
KMT2B.CCND3.fRNA.OE.R	5’-CCTAGGACCAGCACCTCCCAGTCCCAGTGGCGCCGTTTGCGC-3’
KMT2B.CCND3.fRNA.OE.F	5’-CGCAAACGGCGCCACTGGGACTGGGAGGTGCTGGTCCTAGGG-3’
CCND3.E5.fRNA.R	5’-CTACAGGTGTATGGCTGTGACATCTGTAGGAGTGCTGG-3’
MLL.I11.BsaI.F	5’-AAACAGAGACCATTTAGCAGGTACTATTCCCTGT-3’
MLL.I11.SapI.R	5’-CCAAGAAGAGCTAACTAATCTATATTTCTTTCCACCAACA-3’
AF4.I3.BsaI.F	5’-AAACAGAGACCTAGAGCATACATTCTGTCACTGAAATT-3’
AF4.I3.SapI.R	5’-CCAAGAAGAGCGTGCATTAAATCTTCTTAAGTTAGGATCA-3’
KMT2B.E13.F	5’-GCCATGCATACCACCCGGCCTGTCT-3’
KMT2B.E14.R	5’-ATCTCCAGACCACTCGACGTCCCAGTTCTT-3’
CCND3.I2.F	5’-CAGGCCCTGATTGCTTTAAGGAAGGGCTGA-3’
CCND3.I2.3.R	5’-TCAGTGCGATGCTGAGGCAG-3
AF4.cont.F	5’-TCCGAAGGTACTGACCGACAGCCTTAACTA-3’
MLL.con.R1	5’-AGCCAAAGAGGTATCTGCCAGGAATTTAAGAA-3’
MLL.cont.F	5’-CAGAATCAGGTGAGTGAGGAGGGCAAGAAG-3’
AF4.cont.R	5’-TTTCCCATGAGGTGAAAGGAAAGGACAGGG-3’
KMT2B.LR.1.F	5’-CGGTCCCGGCGGGGAAAGGT-3’
CCND3.LR.1.R	5’-GGCTGGCCGGGCCCCTTAGTG-3’
CCND3-KMT2B.LR.1.F	5’- TGCAGCGGGAGATCAAGCCGCACA-3’

Oligonucleotides	
CCND3-KMT2B.LR.1.R	5’- GCGCCCGCGCCCTCACCATCTG-3’
CCND3-KMT2B.LR.3.F	5’- TGGCGGGGAAGGCGATGGGGGTG-3’
CCND3-KMT2B.LR.3.R	5’- CCAGCGTGGCCCCTGCATGCCC-3’

2.12. Vectors

Vectors are described after usage; the corresponding maps can be found in 7.1 Vector maps.

2.12.1. Cloning Vector Kits

Cloning vectors are specifically designed vectors for easy and direct cloning of PCR fragments amplified with polymerases producing a 3´-A overhang.

2.12.1.1. pGEM®-T

The pGEM®-T vector system from Promega is a pre-linearized vector with 3´-T overhangs to directly insert PCR fragments amplified with polymerases producing 3´-A overhangs. It has a blue/white selection marker to select for positive clones when supplemented with X-Gal.

2.12.1.2. TA Cloning™ Kit with pCR™2.1 Vector

The TA Cloning™ Kit with pCR™2.1 vector provides a one-step cloning strategy for directly inserting a PCR product with 3´-A overhangs. The Kit uses the pCR™2.1 vector and ExpressLink™ T4 DNA Ligase to generate a ligation product in a short time. It has a blue/white selection marker to select for positive clones when supplemented with X-Gal.

2.12.2. Transposase vector

For the establishment of stably transfected cell lines the pcGlobin SB100$_{XCO}$ was used. It encodes the engineered and optimized version of the *sleeping beauty* transposase, the SB100$_{XCO}$. The vector was kindly provided by Prof. Dr. Zoltán Ivics.

2.12.3. Template vectors

Table 12. Recombinant DNA.

Recombinant DNA		
pSBbi-RB, sleeping beauty backbone with blasticidin and RFP	Kowarz et al., 2015	AddGene #60522
pSBbi-GP, sleeping beauty backbone with puromycin and GFP	Kowarz et al., 2015	AddGene #60511
pX603-AAV-CMV::NLS-dSaCas9(D10A,N580A)-NLS-3xHA-bGHpA, dead SaCas9	Ran et al., 2015	AddGene #61594
pX330-Flag-deSpCas9, dead SpCas9	Kulcsar et al., 2017	AddGene #92114
pBW2286_pCAG-PV1-FlpO-N396-L1-ABI-NLS-BGHpA, Abscisic acid binding domain ABI	Weinberg et al., 2017	AddGene #87559
pBW2287_pCAG-PV1-PYL-L1-FlpO-397C-NLS-BGHpA, Abscisic acid binding domain PYL	Weinberg et al., 2017	AddGene #87560
pX330-U6-Chimeric_BB-CBh-hSpCas9, U6 promotor and cassette for SpCas9 sgRNA	Cong et al., 2013	AddGene #42230
pX552, U6 promotor and cassette for SaCas9 sgRNA	Swiech et al., 2014	AddGene #60958
pSpCas9(BB)-2A-GFP (PX458), template plasmid for the sgRNAs	Ran et al., 2013	AddGene #48138
DNA template for the artificial KMT2B::CCND3 fusion RNA	This paper	N/A
Sleeping beauty transposase plasmid pcGLobin-SB100$_{XCO}$	Moudgil et al., 2020	AddGene #154887

2.12.4. *Sleeping Beauty* Vectors

The pSBbi-vectors are expression vectors for the stable integration into eukaryotic cells. They constitutively express the gene of interest that can easily be introduces by two specific SfiI restriction sites. Additionally, an U6 promotor with corresponding sgRNAs was introduced in each vector, to express the guide RNA needs for the dead Cas9 variant. The vector maps can be found in 7.1 Vector maps.

Table 13.pSBbi-Vectors

Vector	Transgene	Resistance	Colour
pSBbi RB::dSaCas9-ABI1/U6::KMT2A-I9	dSaCas9-ABI1/ U6::KMT2A-I9	Ampicillin, Blasticidin	dTomato (red)
pSBbi GP::deSpCas9-PYL1/U6::AFF1-I3	deSpCas9-PYL1/ U6::AFF1-I3	Ampicillin, Puromycin	GFP (green)
pSBbi RB::dSaCas9-ABI1/U6::CCND3-I2	dSaCas9-ABI1/ U6::CCND3-I2	Ampicillin, Blasticidin	dTomato (red)
pSBbi GP::deSpCas9-PYL1/U6::KMT2B-I13	deSpCas9-PYL1/ U6::KMT2B-I13	Ampicillin, Puromycin	GFP (green)
pSBbi RB::dSaCas9-ABI1/U6::NT-I	dSaCas9-ABI1/ U6::NT-I	Ampicillin, Blasticidin	dTomato (red)
pSBbi GP::deSpCas9-PYL1/U6::NT-II	deSpCas9-PYL1/ U6::NT-II	Ampicillin, Puromycin	GFP (green)

2.12.5. AddGene Plasmids

Table 14. AddGene Plasmids

Plasmid	Source	Number
pSBbi-RB,	Kowarz et al., 2015	#60522
pSBbi-GP	Kowarz et al., 2015	#60511
pX603-AAV-CMV::NLS-	Ran et al., 2015	#61594

Plasmid	Source	Number
dSaCas9(D10A,N580A)-NLS-3xHA-bGHpA		
pX330-Flag-deSpCas9	Kulcsar et al., 2017	#92114
pBW2286_pCAG-PV1-FlpO-N396-L1-ABI-NLS-BGHpA	Weinberg et al., 2017	#87559
pBW2287_pCAG-PV1-PYL-L1-FlpO-397C-NLS-BGHpA	Weinberg et al., 2017	#87560
pX330-U6-Chimeric_BB-CBh-hSpCas9	Cong et al., 2013	#42230
pX552	Swiech et al., 2014	#60958
pSpCas9(BB)-2A-GFP (PX458)	Ran et al., 2013	#48138

3. Methods

3.1. Standard methods

Methods performed after standard protocols (Wiley & Son`s current protocols) are listed in Table 15.

Table 15. Methods after standard protocol

Methods after standard protocols
Agarose-gel electrophoresis
Alcohol precipitation of nucleic acids
DNA-ligation
DNA-restriction
Insertion of restriction sites by modified primers
Generation of fusion fragments using modified primers
Isolation of cytosolic RNA from eukaryotic cells
Isolation of genomic DNA from eukaryotic cells
PCR analysis
Pei-transfection of HEK293T cells
Phenol/Chloroform extraction of nucleic acids
Photometric determination of DNA and RNA concentration
Preparation of plasmid DNA after alkali lysis
Transformation of *E. coli*

3.2. Methods according to the manufacturer

Table 16. Methods according to the manufacturer

Methods according to the manufacturer
All prep DNA/RNA extraction
Big Dye Terminator Sequencing Kit (ABI Prism)
Gel Extraction (Quiagen)
Gel Extraction (Zymoclean™)

Cloning of PCR-fragments (Promega)
Long Range PCR (Promega)
Midi preparation (Promega)
Mini preparation (peqLab)
PCR fragment purification
RNA extraction (Quiagen)
RT-PCR (Invitrogen, NEB, Quiagen)
TOPO TA cloning (Invitrogen)

3.3. Cultivation of adherent cells

All the cell culture work was performed on a sterile workbench in order to avoid contamination of the human cells. The cultivation of HEK293T cells was carried out in cell culture dishes in a CO_2-incubator with 5% CO_2, a relative humidity of 95% and 37°C. The cells were held in DMEM low glucose medium supplemented with 10% (v/v) FCS, 2 mM L-Glutamine and 100 U/ml Penicillin/Streptomycin. Cells were passaged every two to three days to stay below 80% confluency. The medium was discarded and the cells were washed with half the volume PBS. Up to 1 ml of Accutase® was pipetted onto the cells and incubated for about 5 min in the CO_2-incubator at 37°C. The cells were separated by pipetting up and down with a 1000 µl pipette and afterwards 1/10 of the cell suspension was transferred into a new cell culture dish and filled up with fresh medium.

3.4. DNA-transfection of HEK293T cells using Metafectene Pro®

Transfection describes the process of introducing purified nucleic acids into eukaryotic cells. In this work mainly stable integration using the *Sleeping Beauty* transposon system was used. The result of a stably integrated gene of interested is a cell line which carries the new DNA as an integral part.

For the stable transfection our *Sleeping Beauty* plasmid with the gene of interest was transfected together with a plasmid carrying the information of the *Sleeping Beauty* transposase (pcGoblin $SB100_{XCO}$). As a transfection solution Metafectene Pro® was used.

The transfection was carried out with HEK293T cells at about 70 – 80% confluency in 10 cm cell culture dishes. For the transfection a solution A was prepared consisting of 2 µg plasmid 1 from the proximity system, 2 µg plasmid 2 from the proximity system and 0.02 µg pcGlobin SB100$_{XCO}$ and filled up with PBS to 700 µl. Solution B consisted of 18 µl Metafectene Pro® and 682 µl PBS, afterwards solution A was pipetted into solution B and then incubated for 20 min at room temperature. The transfection solution was pipetted onto the cells and incubated in the CO_2-incubator for 24h before the medium was changed. In order to select the cells for double positive transfections, the transfected plasmids contained an antibiotic resistance and a colour tag for the selection. The cells were incubated with 2 µg/ml puromycin and 20 µg/ml blasticidin for 1 to 2 days and the selection process was observed via fluorescence microscopy. This was repeated until only double positive cells were left.

3.5. Sequence PCR

The PCR bands were extracted after the agarose gel electrophoresis and either sequenced directly or amplified with nested primers and sequenced afterwards. The sequence PCR was performed with the BigDye™ Terminator Cycle Sequencing kit. Therefore 2 µl purified PCR product were mixed with 1 µl of the correct primer and 2 µl BigDye™ and afterwards amplified. The PCR protocol used was 96°C for 30 s for the initial denaturation, 30 cycles of 96°C for 10 s, 58°C for 10 s and 60°C for 4 min.

3.6. DNA precipitation

After the sequence PCR the reaction mix was filled up with water to a total volume of 100 µl. 250 µl of 96% ethanol with 10 µl 3 M sodium acetate (pH = 4,8) were added. After 20 min of centrifugation at 14000 g the supernatant was removed and 250 µl of 70% ethanol were added and centrifuged for 5 min at 14000 x g. The supernatant was removed carefully with a pipette and the pellet was dried at room temperature and resuspended in 10 µl HI-DI by vortexing.

3.7. Chromosome Conformation Capture [216]

3.7.1. Single cell preparations from adherent cells

The medium was removed and the cells were washed with PBS twice. The cells were treated with 1 ml Accutase for 5 – 10 min at 37°C. Afterwards the cells were resuspended in 9 ml of 10 % (v/v) of FCS/PBS and centrifuged for 1 min at 400g at room temperature. The supernatant was discarded and the pellet resuspended in 500 µl of 10% (v/v) FCS/PBS per 1×10^7 cells. The suspension was filtered through a 40 µm cell strainer to make a single-cell suspension.

3.7.2. Formaldehyde crosslinking

9.5 ml of 2% formaldehyde/10% FCS/PBS per 1×10^7 cells were added and incubated for 10 min at room temperature (18 – 22 °C), while tumbling. The tube was transferred to ice and 1.425 ml of 1 M glycine (ice cold) were added to quench the cross-linking reaction. Afterwards the cells were centrifuged for 8 min at 225g at 4°C and the supernatant was carefully removed.

3.7.3. Cell lysis

The pellet was resuspended in 5 ml ice cold lysis buffer and incubated for 10 min on ice. Afterwards it was centrifuged for 5 min at 400g at 4°C and the supernatant was removed.

3.7.4. DNA hydrolysis

The nuclei were taken up in 0.5 ml of 1.2× restriction enzyme buffer and transferred to a safe-lock tube. The tube was placed at 37°C and 7.5 µl 20% (w/v) SDS (final: 0.3% SDS) is added. The sample is then incubated at 37°C for 1 h while shaking at 900 rpm. Afterwards 50 µl of 20% (v/v) Triton X-100 (final: 2% Triton X-100) is added. A 5 µl aliquot of the sample is stored as undigested genomic DNA control. The sample is stored at -20°C until it is needed. Then 400 U of the restriction enzyme is added to the remaining sample and incubated over night at 37°C or the optimal temperature over

night while shaking at 900 rpm. Another sample is taken out as a digested genomic DNA control.

3.7.5. Ligation

In the first step of the ligation, 40 µl of 20% (w/v) SDS (final 1.6%) were added to the remaining sample and incubated for 20 – 25 min at 65°C while shaking at 900 rpm. After the incubation the solution was transferred to a 50 ml falcon tube. First 6.125 ml of 1.15 x ligation buffer and then 375 µl of 20% (v/v) Triton X-100 (final 1% Triton X-100) were added and the sample incubated for 1h at 37°c while shaking gently. After that 5 µl ligase (100 U total) were added and incubated for 4 h at 16°C followed by 30 min at room temperature (18 – 22°C). In the end 30 µl of 10 mg/ml Proteinase K (300 µg final) were added and the sample was incubated at 65°C over night to de-crosslink the DNA.

3.7.6. DNA purification

For the DNA purification, 30 µl of 10 mg/ml RNase were added and incubated for 30 – 45 min at 37°C. 7 ml of phenol-chloroform were added and mixed powerfully. Afterwards the sample was centrifuged for 15 min at 2,200g at room temperature (18 – 22°C) and the supernatant transferred into a new 50 ml falcon tube. 7 ml of distilled water, 1.5 ml of 2 M sodium acetate pH 6.5 followed by 35 ml of ethanol were added. The sample was mixed and placed at -80°C for approximately 1h and then centrifuged for 45 min at 2,200g at 4°C. The supernatant was removed and 10 ml of 70% (v/v) ethanol added and centrifuged for 15 min at 2,200g and 4°C. In the last step the supernatant was removed carefully and the pellet dried at room temperature (18 – 22°C). The pellet was dissolved in 150 µl of 10 mM Tris pH 7.5 and was ready for PCR analysis.

3.7.7. Amplification of 3C template

In order to analyse whether the induction of proximity worked, the formed DNA hybrids had to be investigated. To see whether the DNA hybrids were formed PCRs were performed spanning the used restriction site between both gene fragments. The PCR for *KMT2A* and *AFF1* was performed using the primers MLL.I9.CCC.2.F and

AF4.I3.CCC.Guide2.2.R and for *KMT2B* and *CCND3* the primers CCND3.CCC.1.HindIII.F and KMT2B.CCC.1.HindIII.R.

3.8. cDNA synthesis

For the cDNA synthesis 1 µg of RNA was filled up to 7 µl with RNase-free water and mixed with 1 µl N6 primer (100 pmol). The sample was incubated for 10 min at 70 °C and then cooled down on ice for 2 min. For the reaction, 1 µl RNasin, 4 µl 5x First Strand, 4 µl 2.5 mM dNTPs, 2 µl 100 mM DTT and 1 µl SuperScript™ II Reverse Transcriptase were added, incubated for 10 min at RT and then for 1 h at 42 °C. To stop the reaction, 30 µl RNase-free water was added and the sample incubated for 10 min at 70 °C.

3.9. RT-PCR for NGEFTs

The RNAs were extracted with the Qiagen RNeasy Mini Kit according to the manufacturer's protocol. The synthesis of cDNA was performed after 3.8. To monitor trans-spliced fusion mRNAs, the cDNAs from the different cell lines transfected with the dimerization systems for *KMT2A* and *AFF1* or *CCND3* and *KMT2B* were analyzed at day 0 and day 14 of abscisic acid treatment via RT-PCR. The endogenous *AFF1* gene transcript was amplified with oligonucleotides AF4.3 and AF4.5, while the endogenous *KMT2A* gene transcript with oligonucleotides 8.3 and 13.5, respectively. The fusion transcript *AFF1::KMT2A* was amplified with oligonucleotides AF4.3 and 13.5, and the *KMT2A::AFF1* fusion transcript was analyzed with oligonucleotides 8.3 and AF4.5. The *CCND3* gene transcript was amplified with CCND3.E2.2.F and CCND3.E3.1.R and the *KMT2B* gene transcript was analyzed with KMT2B.E8.1.F and KMT2B.E15.1.R. The *CCND3::KMT2B* fusion transcript was analyzed with CCND3.E2.2.F and KMT2B.E15.1.R, while the *KMT2B::CCND3* fusion transcript was analyzed with KMT2B.E8.1.F and CCND3.E4.2.R. The PCR protocol was 94°C for 3 min for the initial denaturation, 35 cycles of 94°C for 30 s, 63°C for 30 s and 72°C for 90 s, followed by 72°C for 2 min for the final elongation. RT-PCR products were analyzed on a 2% agarose gel.

3.10. Method for CRISPR/Cas9 DNA damage experiments

3.10.1. Production of sgRNAs

The oligonucleotides needed for the target gene are generated on crisprscan.org. The guide RNAs are sorted by highest CRISPRscan score and lowest number of mismatches. The oligonucleotides are then generated including the sequence for the T7 promotor, the region complementary to the target region, the scaffold overlap and an additional "ATAGC" to increase the primer specificity for the PCRs.

3.10.2. Amplification of the DNA template for the sgRNA

The cloning oligos for each target gene were designed using the CRISPRscan server and only oligos with a high CRISPRscore and as few as possible off-targets were chosen. CRISPRscan can directly design primers including T7 promotor, the target complementary region and a scaffold overlap. For better primer binding a 'ATAGC' was added to the end of the primer. With these primers PCRs were performed to produce the template DNA for the sgRNA production. For this, the Kapa HiFi HotStart ReadyMix was used. The reaction mix consisted of 2 µl 10 µM sgRNA primer; for *AFF1* AF4.cut.guide.F, for *KMT2A* MLL.cut.guide.F, for *CCND3* CCND3.cut.guide3.F and for *KMT2B* KMT2B.cut.guide6.F in combination with 1 µl 10 µM universal reverse Primer, 1 µl 2-4 ng/µl of the PX458 plasmid, 10 µl KAPA HiFi Mix 2x and 5 µl water. The PCR protocol used was 95°C for 3 min for the initial denaturation, 30 cycles of 98°C for 3 s, 60°C for 5 s and 72°C for 10 s, followed by 72°C for 1 min for the final elongation. The DNA was separated on a 2% agarose gel and the corresponding bands were cut out and recovered using the Zymoclean Gel DNA Recovery Kit following the manufacturer's protocol.

3.10.3. DNA template for the artificial fusion RNA

For the synthesis of a DNA template for a synthetic and artificial fusion RNA, a PCR for *KMT2B* was performed with KMT2B.E9.T7.F and KMT2B.CCND3.fRNA.OE.R, and for *CCND3* with KMT2B.CCND3.fRNA.OE.F and CCND3.E5.fRNA.R, respectively. The two fragments were separated on a 2% agarose gel, cut out and extracted using the Zymoclean Gel DNA Recovery Kit following the manufacturer's protocol. An overlap-extension PCR was performed using the two PCR fragments in combination with the

outer primers KMT2B.E9.T7.F and CCND3.E5.fRNA.R. The PCR fragment was separated on a 1% agarose gel and recovered again using the Zymoclean Gel DNA Recovery Kit following the manufacturer's protocol. The primer KMT2B.E9.T7.F introduces a T7 promotor in order to use the template for an *in vitro* transcription.

3.10.4. *In vitro* transcription of sgRNA

The *in vitro* transcription of sgRNAs was performed using the HiScribe T7 High Yield RNA Synthesis Kit. For the transcription reaction, 400 – 1000 ng of template DNA were mixed with 8 µl dNTPs (ATP, GTP, CTP, UTP), 2 µl Reaction Buffer (10x), 2 µl T7 RNA Polymerase Enzyme Mix and 0 – 4 µl water to adjust to a final volume of 20 µl. Transcription was performed for a minimum of 5h at 37°C. Following the amplification step, the sgRNAs were purified using Zymo RNA Clean & Concentrator Kit according to the manufacturer. Sample volume was made up to 50 µl with nuclease-free water and 100 µl RNA Binding Buffer added and mixed. 150 µl ethanol was added and mixed. The sample was transferred onto the Zymo-Spin™ IICR Column and centrifuged for 30 s at 10,000 x g. For the on-column DNA digestion using the RNase-Free DNase Set, 10 µl DNase-Stock-Solution in 70 µl RDD Buffer were added to the middle of the membrane and incubated for 1 h at RT. 400 µl RNA Prep Buffer was added and the tube centrifuged for 30 s at 10,000 x g, the flow through was discarded. 700 µl RNA Wash Buffer were added, the tube centrifuged and the flow through discarded. 400 µl RNA Wash Buffer were added to the column, the column centrifuged for 1 min at 10,000 x g and transferred into a new reaction tube. The sgRNAs were eluted in 25 µl nuclease-free water that were incubated for 7 min on the column before centrifugation for 1 min at 10,000 x g. Concentration was measured using a Nanodrop, and the final concentration was diluted to 1 µg/µl and stored at -80°C.

3.10.5. *In vitro* cleavage assay

Fragments for *KMT2A* were synthesized using MLL.I11.BsaI.F and MLL.I11.SapI.R, *AFF1* with AF4.I3.BsaI.F and AF4.I3.SapI.R, *KMT2B* using KMT2B.E13.F and KMT2B.E14.R and *CCND3* with CCND3.I2.F and CCND3.I2.3.R. For the cleavage assay, 400 ng of Cas9 protein (1 µg/µl) were mixed with 400 ng sgRNA (1 µg/µl), 2 µl Cas9 reaction buffer and filled up with water to a final volume of 20 µl together with the dsDNA and incubated at

37°C for 10 min. 200 ng of the amplified DNA fragment was added to the reaction and incubated for 3 to 5 h at 37°C. The reaction mix was analysed on a 2% agarose gel to examine the cleavage.

3.10.6. Transfection of Cas9 preloaded with sgRNA

After 2 µM ABA induction for 14 days, the double strand break was induced by Cas9 preloaded with the sgRNA. 1 µg of Cas9 protein and 1 µg of sgRNA were mixed, filled up to 10 µl with PBS and incubated at 37°C for 20 min. For the transfection 18 µl Metafectene Pro® and 282 µl of PBS were mixed. In a different reaction tube, 10 µl of each Cas9 preincubated with the sgRNA were mixed and filled up to 300 µl with PBS. For the experiments with the artificial fusion RNA, 1 µg of RNA was added to the mixture. The solution with the Cas9 proteins (and fusion RNA) was then pipetted into the Metafectene Pro® mix and incubated for 20 min at room temperature. The mixture was then pipetted onto the cell culture dish and mixed by tilting the dish. After 4 h the medium was exchanged with fresh DMEM (10% FCS, 1% glutamine, 1% PenStrep) supplemented with ABA.

3.10.7. RT-PCR after DNA damage

To analyse the RNA and DNA from the same culture dish the AllPrep RNA/DNA Mini Kit was used according to the manufacturer's instructions. The synthesis of cDNA was performed after 3.8. *AFF1* wild-type transcript was amplified with AF4.3 and AF4.5, *KMT2A* wild-type transcript with 8.3 and 13.5, *AFF1::KMT2A* transcript with AF4.3 and 13.5 and *KMT2A::AFF1* transcript was analyzed with 8.3 and AF4.5. The *CCND3* wild-type gene was amplified with CCND3.E2.2.F and CCND3.E3.1.R, the *KMT2B* wild-type gene with KMT2B.E8.1.F and KMT2B.E15.1.R, the *CCND3::KMT2B* fusion transcript with CCND3.E2.2.F and KMT2B.E15.1.R, and the *KMT2B::CCND3* fusion transcript was analyzed with KMT2B.E8.1.F and CCND3.E4.2.R. PCR protocol used was 94°C for 3 min for the initial denaturation, 35 cycles of 94°C for 30 s, 63°C for 30 s and 72°C for 90 s, followed by 72°C for 2 min for the final elongation. RT-PCR products were analyzed on a 2% agarose gel.

3.10.8. Long Range PCR after DNA damage

The chromosomal translocations were analyzed using long-range PCR in combination with a touch-down PCR protocol. The initial denaturation was 93°C for 3 min, the touch-down was 10 cycles starting with 93°C for 15 s, 68°C for 30 s and 68°C for 10 min while reducing the annealing temperature by 1°C every cycle, followed by 32 cycles, starting with 93°C for 15 s, 63°C for 30 s and 68°C for 10 min, while increasing the elongation time by 20 s in every cycle, followed by 68°C for 10 min as final elongation. The *AFF1::KMT2A* translocation was analyzed with AF4.cont.F and MLL.con.R1 and the *KMT2A::AFF1* translocation with MLL.cont.F and AF4.cont.R. The *KMT2B::CCND3* translocation was analyzed with KMT2B.LR.1.F and CCND3.LR.1.R and the *CCND3::KMT2B* translocation with CCND3-KMT2B.LR.1.F and CCND3-KMT2B.LR.1.R. Because there were many off target bands in the PCR for *CCND3::KMT2B* a nested PCR with CCND3-KMT2B.LR.3.F and CCND3-KMT2B.LR.3.R was performed. To differentiate the sequences of one translocation from distinct cells, smaller pieces of the bands were amplified to clone them into pCR™2.1-TOPO™ according to the manufacturer's manual. For *AFF1::KMT2A* the primers AF4.I3.BsaI.F and MLL.I11.SpaI.R and for *KMT2A::AFF1* the primers MLL.I11.BsaI.F and AF4.I3.SapI.R were used. For *KMT2B::CCND3* the primers KMT2B.E13.F and CCND3.I2.3.R and for *CCND3::KMT2B* the primers CCND3.I2.F and KMT2B.E14.R were used. The primers were also used or sequencing.

4. Results

4.1. Plasmid constructs and stable cell lines

In order to induce the proximity of *KMT2A* and *AFF1* or *CCND3* and *KMT2B* the corresponding *Sleeping Beauty* vectors were used. Two constitutive expression vectors, pSBbi-RP and pSBbi-GP were used to clone the two fusion genes, namely the coding region for a dead SaCas9 or a dead enhanced SpCas9 fused to either one of the abscisic acid binding domains ABI1 or PYL1, respectively (dSaCas9-ABI1 and deSpCas9-PYL; Figure 10). Additionally, an U6 promotor driven cassette was cloned into each of the 2 vectors that exhibit the different sgRNA sequences (AFF1-I3 and KMT2A-I9) by using the two self-deleting BbsI or BsaI cleavage sites. HEK293T cells were transfected with either one of the vector pairs, pSBbi-RB::dSaCas9-ABI1/U6::KMT2A-I9 and pSBbi-GP::deSpCas9-PYL1/U6::AFF1-I3 to target *KMT2A* and *AFF1*, pSBbi-RB::dSaCas9-ABI1/U6::CCND3-I2 and pSBbi-GP::deSpCas9-PYL1/U6::KMT2B-I13 to target *KMT2B* and *CCND3*, pSBbi-RB::dSaCas9-ABI1/U6::NT-I and pSBbi-GP::deSpCas9-PYL1/U6::NT-II without targets in the genome as control cell line. The plasmid with the corresponding AddGene numbers used for the different constructs can be found in 2.12.5 AddGene Plasmids. The sgRNAs with the corresponding target sequence are summarized in Table 17 and an overview of the different plasmids used to generate stable cell lines is shown in Figure 10.

Table 17. Characterizations of sgRNAs

sgRNA	Genomic target	Scaffold	Spacer/targeting sequence
KMT2A-I9	KMT2A intron 9	SaCas9	5'-CACCGTACTCTTGAGAAGCTGAGGCAG-3'
AFF1-I3	AFF1 intron 3	SpCas9	5'-CACCGTCTAATAAATGGCTTGACTTG-3'
CCND3-I2	CCND3 intron 2	SaCas9	5'-CCCTTGGTAGGATATCAGGG-3'
KMT2B-I13	KMT2B intron 13	SpCas9	5'-GGGGCTCTCAGGAGGAGCAG-3'
NT-I	no target	SaCas9	5'-CACCGGTAATCGCTAGCTTGTCGACT-3'
NT-II	no target	SpCas9	5'-CACCGGTAATCGCTAGCTTGTCGACT-3'

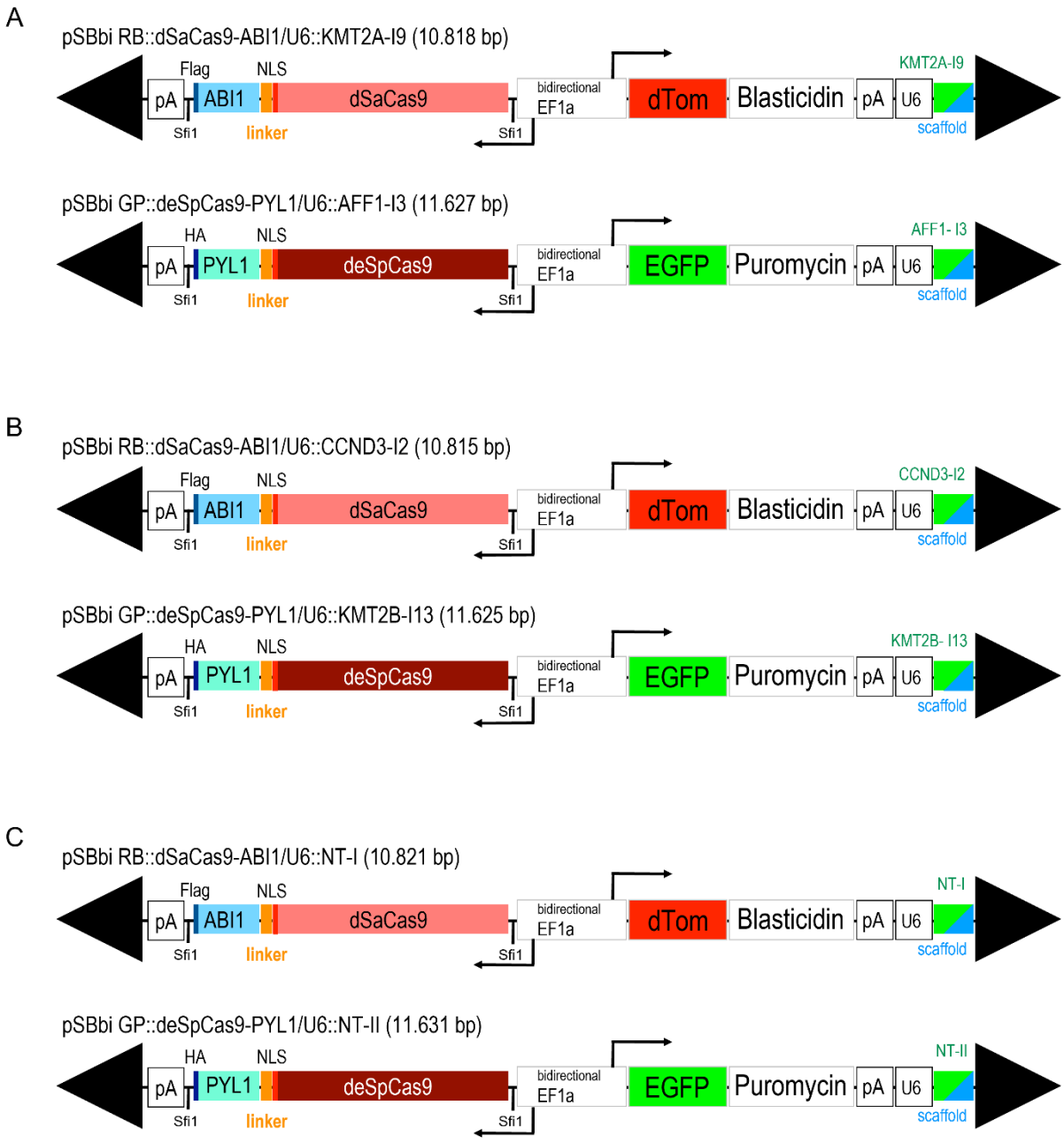

Figure 10. Overview of the plasmids used for stable HEK293T cells. A, B and C each represent one stable cell line. The plasmids contain a dead SaCas9 (dSaCas9) and a dead enhanced SpCas9 (deSpCas9) fused to either PYL1 or ABI1. The constructs either contain a dTomato (dTom) for red or an enhanced GFP (EGFP) for green fluorescence and a resistance for puromycin or blasticidin as selection markers. The three vector pairs only differ in their encoded sgRNA sequence, targeting different regions of the genome. A) Encoding the target sequence with the corresponding scaffold either for *KMT2A*-I9 or *AFF1*-I3. B) Encoding the target sequence with the corresponding scaffold either for *CCND3*-I2 or *KMT2B*-I13. C) Encoding two non-targeting sequences.

4.2. Induction of gene Proximity

After establishing these stable cell lines, it was necessary to validate whether the administration of ABA induces a close proximity of each pair of target genes. Therefore, the different cell lines were incubated with ABA over a period of 14 days. Subsequently, the cells were harvested and chromosome conformation capture (3C) experiments were performed. The 3C experiment allows to investigate gene or chromosome proximity by cross-linking, digesting and ligating the corresponding regions on different chromosomes. There will only be a ligation product of each gene pair if the above mentioned ABA treatment worked successfully. These 3C templates were then used for subsequent PCR experiments in order to validate the proximity of either *KMT2A* and *AFF1* or *CCND3* and *KMT2B*.

The performed PCR experiments revealed that the ABA treatment for 14 days was sufficient to successfully bring the coding chromatin loops for both gene pairs, *KMT2A* and *AFF1* as well *KMT2B* and *CCND3*, into close proximity (Figure 11). As expected, no PCR amplification product was obtained in the non-targeting control cell line. These negative results validated that both gene pairs are not close enough in the 3-dimensional nuclear space of wild type HEK293T cells and no proximity for these 4 genes could be induced by the non-targeting control plasmids.

In order to determine the identity of the PCR bands, a sequence analysis was performed. The sequences show the PCR fragments amplified from the 3C template either with a part of *KMT2A* intron 9 and *AFF1* intron 3 with the AvrII restriction site or a part of *CCND3* intron 4 and *KMT2B* upstream region and the HindIII restriction site (Figure 12).

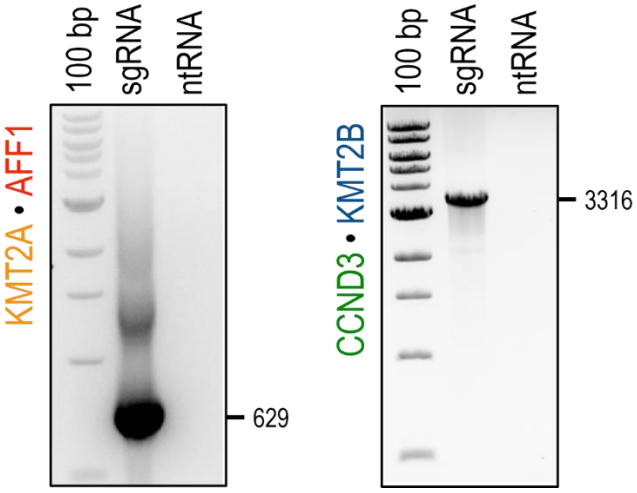

Figure 11. Chromosome conformation capture (3C). Chromosome conformation capture was performed to validate the proximity of *KMT2A* and *AFF1* and *CCND3* and *KMT2B* after 14 days of ABA treatment. The 3C hybrid DNA, created by ligation if the genes were close enough for cross-linking, was used for subsequent PCR analysis. The numbers mark the base pairs of the positive PCR products. For the cell line with targeting constructs a PCR fragment is visible, while there is no band for both gene pairs in the non-targeting control cell line.

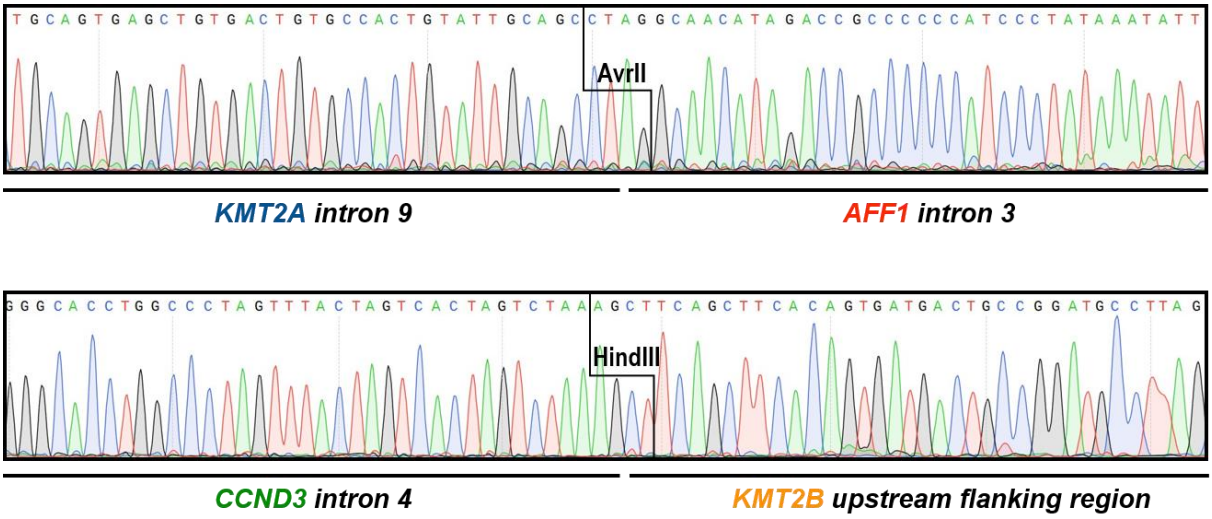

Figure 12. Sequence analysis of the 3C PCR bands. The sequence of the bands amplified from the 3C templates were analysed to validate that the 3C bands are no artefacts.

These results clearly validated our assumption that the ABA treatment for 14 days was sufficient to "rearrange" the chromatin territories inside the nucleus in a way, that distinct genes from different chromosomes were now placed in close proximity within the 3-dimensional space of the cell's nuclei. Moreover, this kind of "chromosome replacement" should have occurred in nearly all cells (in a quantitative fashion).

4.3. Analysis of non-genomically encoded fusion transcripts

In RT-PCR experiments with PBMCs from healthy individuals the non-genomically encoded fusion transcript of *KMT2A-ex9::AFF1-ex4* could always be identified in very low amounts, they were only visible in nested RT-PCR experiments. But a reciprocal *AFF1::KMT2A* fusion transcript has never been found. In a previous publication of our laboratory this fusion RNA was present in 9 out of 10 of the healthy individuals in the absence of a chromosomal translocation (Figure 21). To investigate whether the induced proximity of two genes is crucial for the formation of the above mentioned fusion transcript, RT-PCR experiments were performed with RNA obtained from the 0 (negative control) and 14 days ABA treated cells. In particular, HEK293T cells with the proximity system for *KMT2A* and *AFF1* (Figure 13 left panels) or *KMT2B* and *CCND3* (Figure 13 right panels) were investigated. At day 0, the cell line with the *KMT2A/AFF1* proximity system only expressed the wildtype transcripts for *KMT2A* and *AFF1*. Similar results were obtained in the corresponding cell line with the KMT2B/CCND3 proximity system, identical to the results obtained in the non-targeting control cell line. By contrast, a highly abundant fusion transcript for *KMT2A::AFF1* - but again not the reciprocal *AFF1::KMT2A* - was observed at day 14 when using the *KMT2A/AFF1* proximity system. No such *KMT2B::CCND3* or *CCND3::KMT2B* fusion transcripts were observed in the *KMT2B*/*CCND3* proximity system (Figure 13). This might be due to the fact that neither *KMT2B* nor *CCND3* exhibit PTT formation capacity [26] which was already published before as the causal reason and genetic prerequisite for generating these non-genomically encoded fusion transcripts by *trans*-splicing events. Finally, we validated the identity of the observed PCR band by sequence analysis and demonstrated that this novel PCR band was exactly the expected *KMT2A-ex9::AFF1-ex4* fusion RNA (Figure 13). The amount of this *trans*-spliced fusion mRNA was equally strong as the wild-type mRNA transcripts produced from the endogenous *KMT2A* and *AFF1* genes, respectively.

This demonstrated again, that both proximity systems had rearranged complete chromosomes in a quantitative fashion, because the *trans*-spliced fusion RNA between *KMT2A* and *AFF1* (MA4) was as strong as both wildtype transcripts.

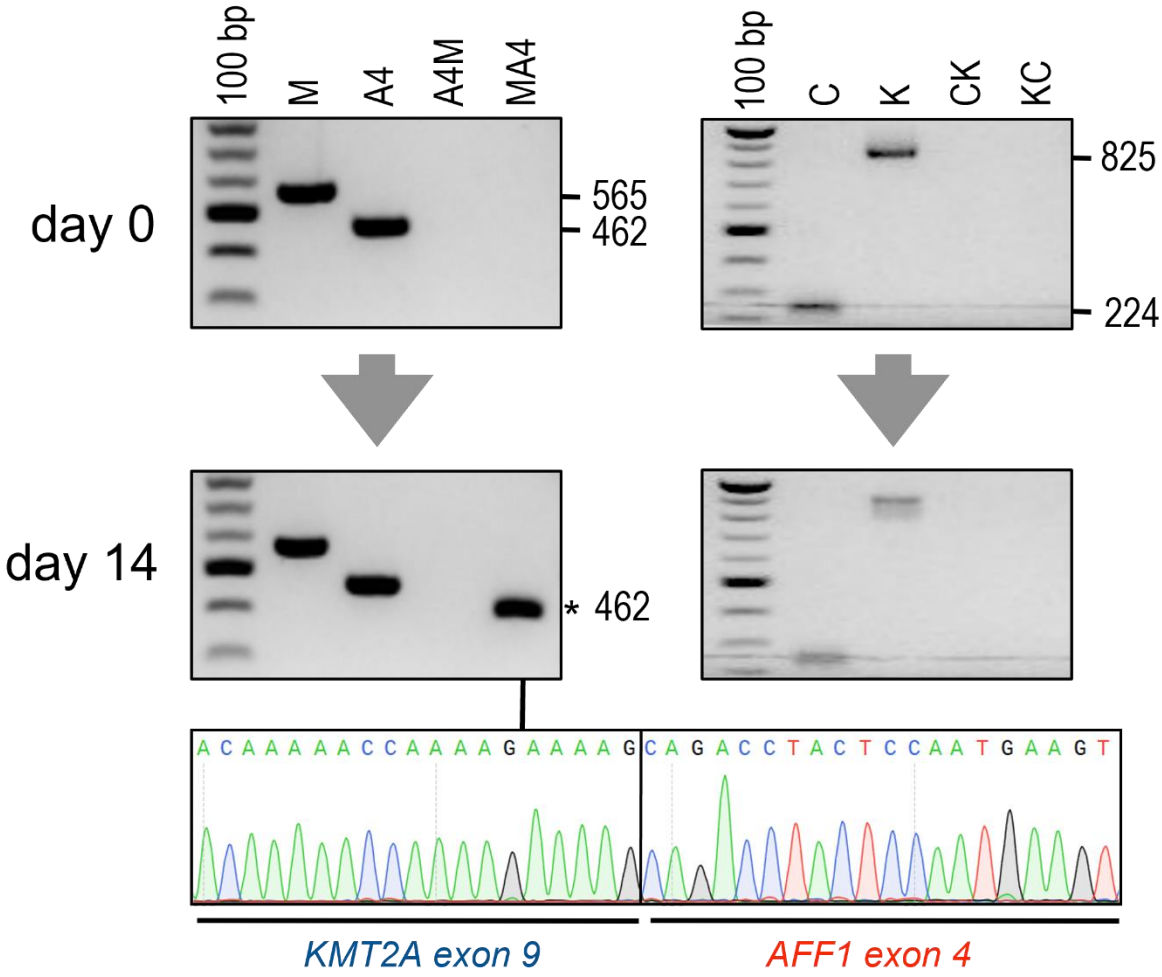

Figure 13. *KMT2A::AFF1* fusion RNA after induction of gene proximity. HEK293T cells with the *KMT2A/AFF1* and the *KMT2B/CCND3* proximity system were harvested at day 0 or 14 of ABA treatment. While all endogenous wild-type genes were expressed, the trans-spliced and non-genomically encoded *KMT2A::AFF1* fusion RNA was only obtained after 14 days of ABA treatment. The novel PCR band was validated by sequence analysis. M: *MLL/KMT2A*; A: *AF4/AFF1*; A4M: *AF4/AFF1::MLL/KMT2A*; MA4: *MLL/KMT2A::AF4/AFF1*; C: *CCND3*; K: *KMT2B*; CK: *CCND3::KMT2B*; KC: *KMT2B::CCND3*.

4.4. *In vitro* digestion of target gene fragment

In order to use this gene proximity system as a source to study the onset of real chromosomal translocations, we aimed to use an additional active CRISPR/Cas9 system to induce specific double strand breaks in the genomic region of these genes. In addition to the scaffold, the spacer sequence is important for the protein's cleavage performance. There are several factors that influence the spacer sequence, the best known being the PAM motif. In order to find sgRNAs, that induce a double strand break in the gene of interest, spacer sequences were generated using the online tool CRISPRscan. To test the activity with the generated sgRNAs, an *in vitro* cleavage of the PCR amplified target region was performed. For the cleavage assay, the Cas9 protein was pre-incubated with the corresponding sgRNA and afterwards the DNA amplicon of the target region was added and incubated. The template is indicated by the opposing arrows with the line under each gene (Figure 14 B). About six different sgRNAs for each gene were evaluated until a single sgRNA was identified that provided us with a satisfying cleavage activity (see Table 18). After the incubation with the pre-loaded Cas9, the template DNA was cut into corresponding subfragments (indicated by a star in Figure 14 A). After approximately 5h of incubation the efficiency was around 100% because there was no full-length template DNA visible anymore on the gel.

Table 18. Spacer sequences to induce DNA double strand breaks.

Name	**Target sequence**
AF4.cut.guide	5´-GGGGGGGCGTGATTCCTACC-3´
MLL.cut.guide	5´-GGAGCTCCTTATAGATGAAG-3´
CCND3.cut.guide	5´-GGGAGTAGGAGGCCAGAGTT-3´
KMT2B.cut.guide	5´-GGAGACGGGCGAGGAGGAAT-3´

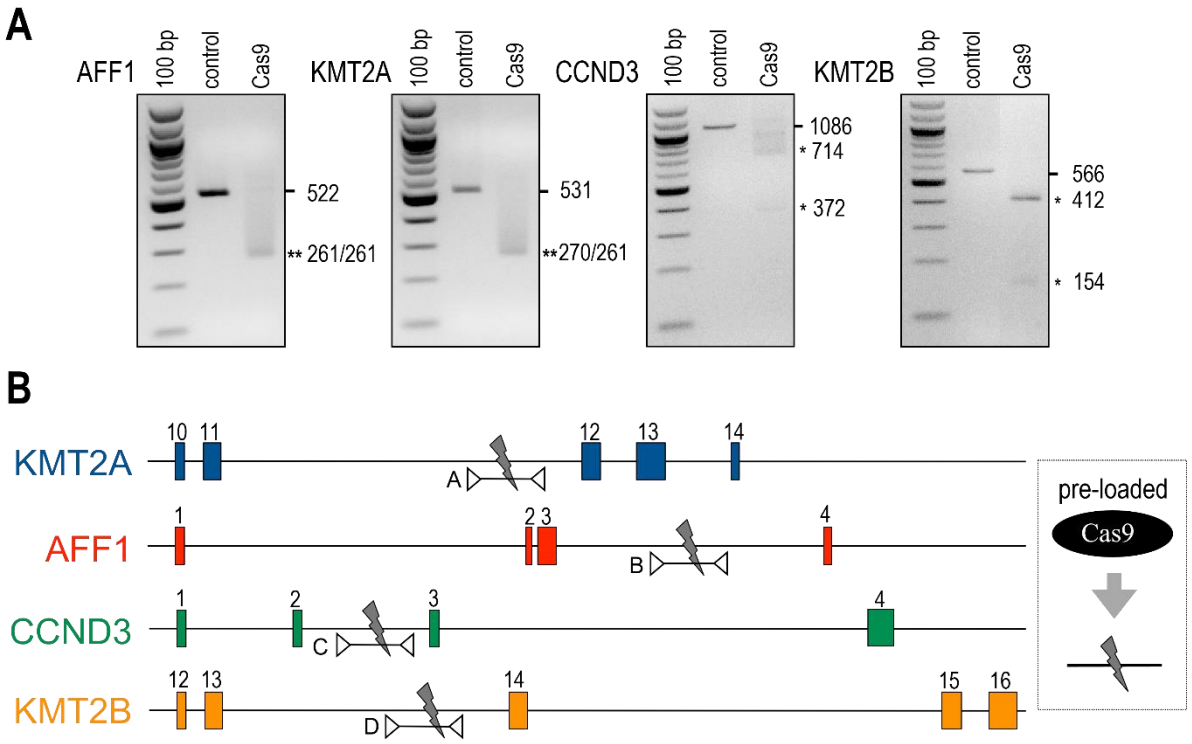

Figure 14. *In vitro* cleavage assay of Cas9. DNA templates of *KMT2A*, *AFF1*, *CCND3* and *KMT2B* were amplified by PCR. Cas9 protein was pre-incubated with the corresponding sgRNA and afterwards the DNA PCR amplicon of the target area was added. A) *In vitro* cleavage assay of the template DNA for the corresponding genes *AFF1*, *KMT2A*, *CCND3* and *KMT2A*. The numbers indicate the size, the template fragment is indicted by a line and the fragments after cutting with a stars. B) Overview of the Cas9 target region in each gene. The template fragment for the *in vitro* cleavage assay is indicated by the opposing arrows with the line.

4.5. Analysis of fusion transcripts after the induction of DNA double strand breaks

Non-genomically encoded fusion transcripts are present and detectable in healthy cells in the absence of any chromosomal translocation. However, it was unclear whether they are necessary for the onset of chromosomal translocations or not. To investigate this important question, two simultaneous DNA double strand breaks were introduced into *KMT2A* and *AFF1* or *CCND3* and *KMT2B* in 14 days ABA treated cells with the gene proximity systems and additionally, in the HEK293T cells with the non-targeting system

(negative control). The cells were harvested 3 – 5 days after induction of the DNA damage and RNA was isolated.

In HEK293T cells with the *CCND3*/*KMT2B* proximity system as well as in the non-targeting control, only transcripts deriving from wildtype *CCND3* and *KMT2B* genes were identified in the RT-PCR experiments. No fusion transcripts for *CCND3::KMT2B* or *KMT2B::CCND3* were detectable (Figure 15 A). By contrast, the cells with the *KMT2A*/*AFF1* proximity system displayed - besides the trans-spliced *KMT2A-ex9::AFF1-ex4* fusion transcript - PCR bands in the expected size of fusion transcripts for *KMT2A-ex11::AFF1-ex4* and the reciprocal *AFF1-ex3::KMT2A-ex12* (Figure 15 B). These novel bands were sequenced to validate that these 2 PCR bands were indeed deriving from a chromosomal translocation t(4;11) that was induced by the DNA damage situation. In fact, both PCR bands turned out to be *bona fide KMT2A::AFF1* and *AFF1::KMT2A* fusion transcripts (Figure 15 C). To this end, a chromosomal translocation was easily inducible by three prerequisites: (1) gene proximity, (2) the presence of a specific RNA (NGEFT), and (3) a simultaneous DNA damage situation within both genes.

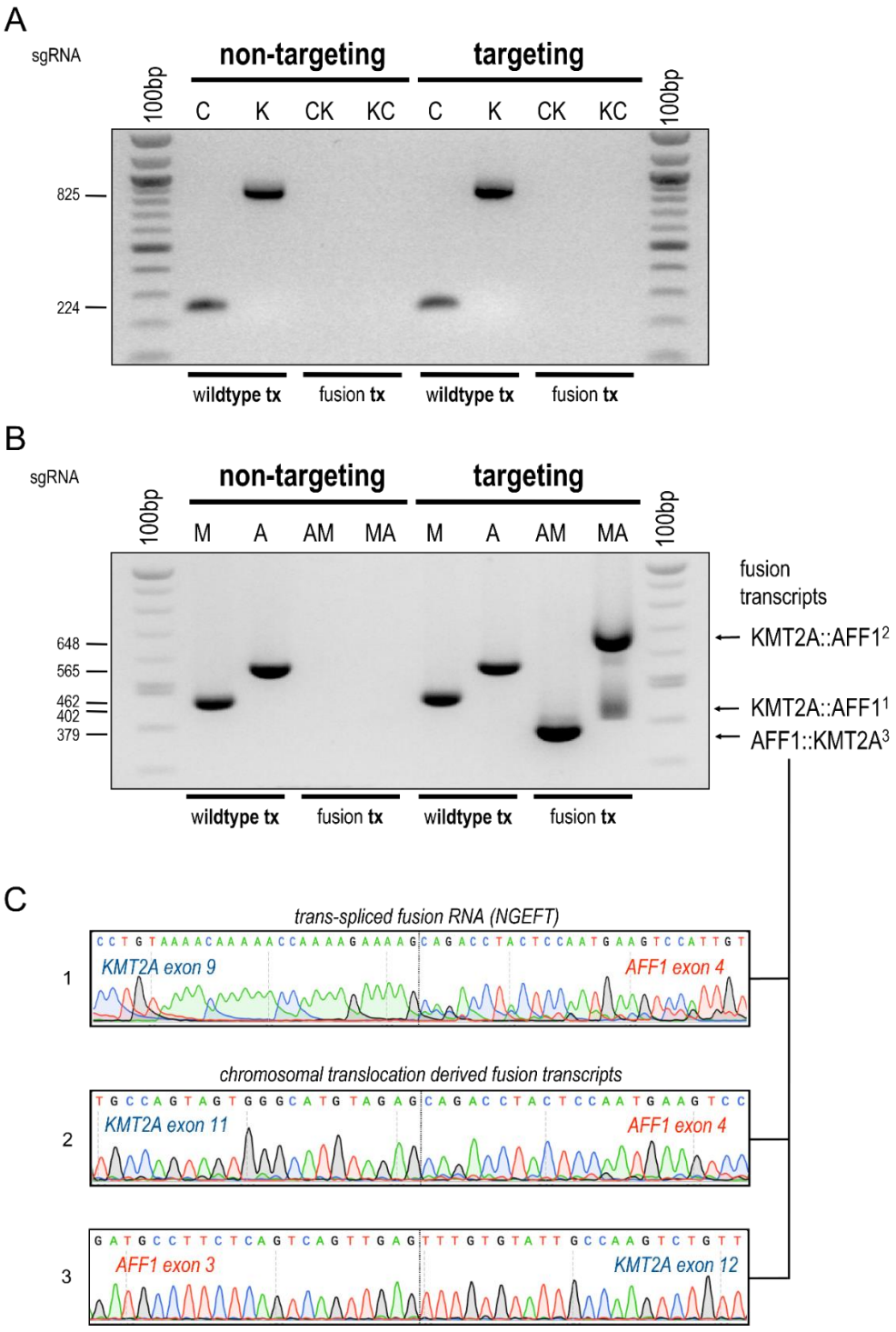

Figure 15. Fusion transcripts after the induction of double strand breaks. HEK293T cells with the *KMT2A/AFF1* or the *CCND3/KMT2B* proximity system or cells with the non-targeting system were treated with ABA for 14 days. DNA damage in both genes was introduced using the Cas9 pre-incubated with the corresponding sgRNAs. Cells were harvested after 3 – 5 days and the RNA extracted to perform cDNA synthesis with subsequent RT-PCR. A) RT-PCR experiments for *CCND3* (C), *KMT2B* (K), *CCND3::KMT2B* (CK) or *KMT2B::CCND3* (KC) using the RNA from

HEK293T cells with the non-targeting proximity system (non-targeting) or the *KMT2B*/*CCND3* proximity system (targeting). Lanes are divided into wildtype transcripts (wildtype tx) and fusion transcripts (fusion tx) and the numbers indicate the band size. B) RT-PCR experiments for *AFF1* (A), *KMT2A* (M), *AFF1::KMT2A* (AM) or *KMT2A::AFF1* (AM) using the RNA from HEK293T cells with the non-targeting proximity system (non-targeting) or the *KMT2A*/*AFF1* proximity system (targeting). Lanes are divided into wildtype transcripts (wildtype tx) and fusion transcripts (fusion tx) and the numbers indicate the band size. C) Results of the sequence analysis for the bands of *KMT2A::AFF1* and *AFF1::KMT2A*.

4.6. Analysis of fusion genes after the induction of DNA double strand breaks

In order to further validate whether these fusion transcripts arised from a chromosomal translocation, the genomic regions of both genes where the CRISPR/Cas9 cut had occurred were investigated. Therefore, the genomic DNA of these cells was extracted to analyze the presence of individual translocations by using an established long range PCR technology. Since all translocations have occurred in single cells, we expected to find slightly different chromosomal fusions due to the error-prone NHEJ DNA repair pathway.

As shown in Figure 16 (upper panel), the long range PCR band of both rearranged alleles (*KMT2A::AFF1* and *AFF1::KMT2A*) was cut out from the gene, and subcloned into a vector for DNA sequencing. Since no fusion alleles could be obtained for the *KMT2B*/*CCND3* proximity system, they were excluded from this analysis (Figure 16 lower panel).

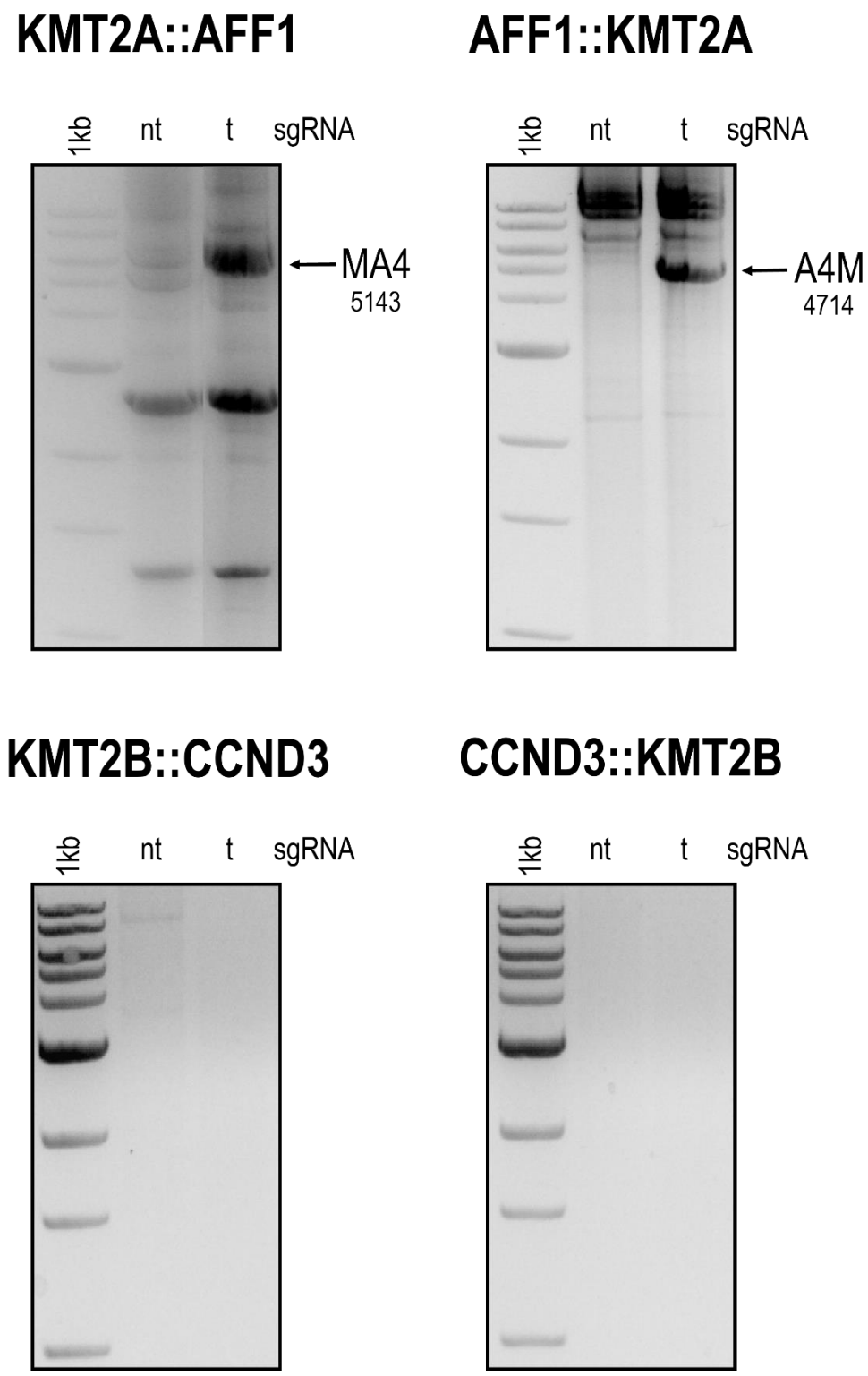

Figure 16. Chromosomal translocation after induced DNA double strand breaks. HEK293T cells with the *KMT2A/AFF1* or the *CCND3/KMT2B* proximity system or the non-targeting control cells were treated with ABA for 14 days. DNA damage in both genes was introduced using the Cas9 pre-incubated with the corresponding sgRNAs. The cells were harvested after 3 – 5 days and the genomic DNA was isolated. HEK293T cells with the proximity system for the

corresponding genes (t) are compared to HEK293T cells with the non-targeting system (nt). Positive bands are indicated by arrows and the band size is written below the band of *KMT2A::AFF1* (MA4) and *AFF1::KMT2A* (A4M).

The obtained sequencing data from single colonies, representing individual chromosomal translocations, for *KMT2A::AFF1* and *AFF1::KMT2A* indeed showed a large variety of different break points (Figure 17). Based on these data, small deletions, the presence of mini-direct repeats (MDR) and also the presence of filler DNA (FD) clearly argued for a NHEJ-mediated DNA repair process at the basis of these specific genetic rearrangements.

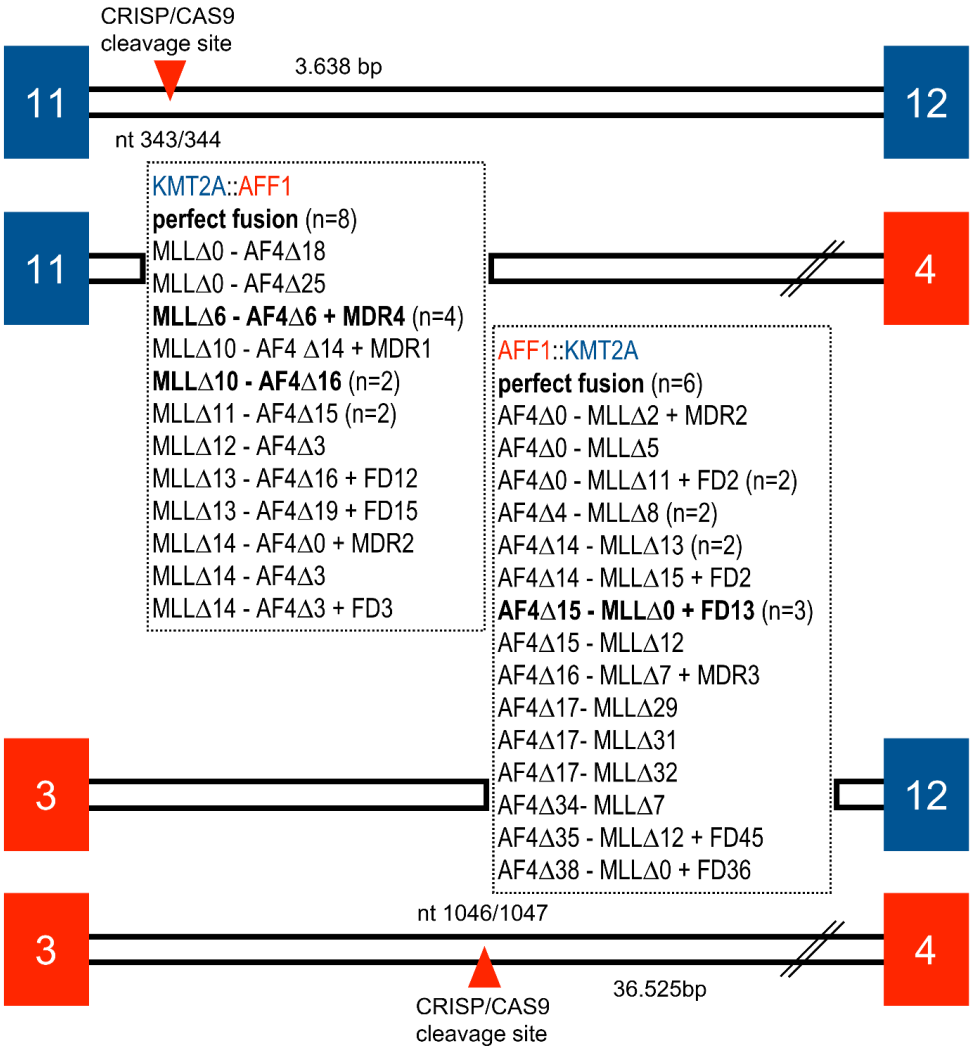

Figure 17. Analysis of different sequences from the artificial t4;11. PCR bands of *KMT2A::AFF1* and *AFF1::KMT2A* were analysed to distinguish between different break points

within *KMT2A* and *AFF1*. The different sequences for *KMT2A::AFF1* and *AFF1::KMT2A* are listed in the corresponding boxes, together with additional information about missing bases (indicated by Δ), filler DNA (FD) and mini-direct repeats (MDR), all followed by the number of bases.

4.7. The influence of an artificial fusion RNA after the induction of a DNA double strand break

Since we had obtained the chromosomal translocation t(4;11) only in the gene proximity system where also a NGEFT was visible, we wanted to find out whether the presence of this particular *KMT2A-ex9::AFF1-ex4* fusion RNA was an essential key component for the onset of the observed chromosomal translocation, or just produced because of the known PTT capability of *KMT2A*. Therefore, a final experiment was designed where we were able to test the importance of these NGEFTs for the onset of chromosomal translocations. For this purpose, a synthetic *KMT2B::CCDN3* fusion was generated and T7-*in vitro* transcribed into an RNA of ~1000 nucleotides. This fusion RNA was simultaneously introduced into the cell together with pre-loaded Cas9 protein to induce the two DNA double strand breaks. Again, the cells were harvested 3 – 5 days after the DNA damage was induced and the RNA and DNA was extracted from the treated cells. To analyse the presence of a possible chromosomal translocation, RT-PCR experiments were conducted (as described above) in order to find the resulting fusion transcripts.

The induction of double strand breaks in *KMT2B* and *CCND3* in the non-targeting control system only revealed the two wild type transcripts , although these cells were co-transfected with the synthetic *KMT2B::CCND3* fusion RNA (Figure 18). By contrast, when the *KMT2B*/*CCND3* proximity system was used, two appropriate fusion transcripts were detected besides the two wildtype transcripts (Figure 18). A subsequently performed sequence analysis revealed that these additional PCR bands were indeed fusion transcripts for *KMT2B-ex13::CCND3-ex3* and *CCND3-ex2::KMT2B-ex14* , respectively (Figure 18).

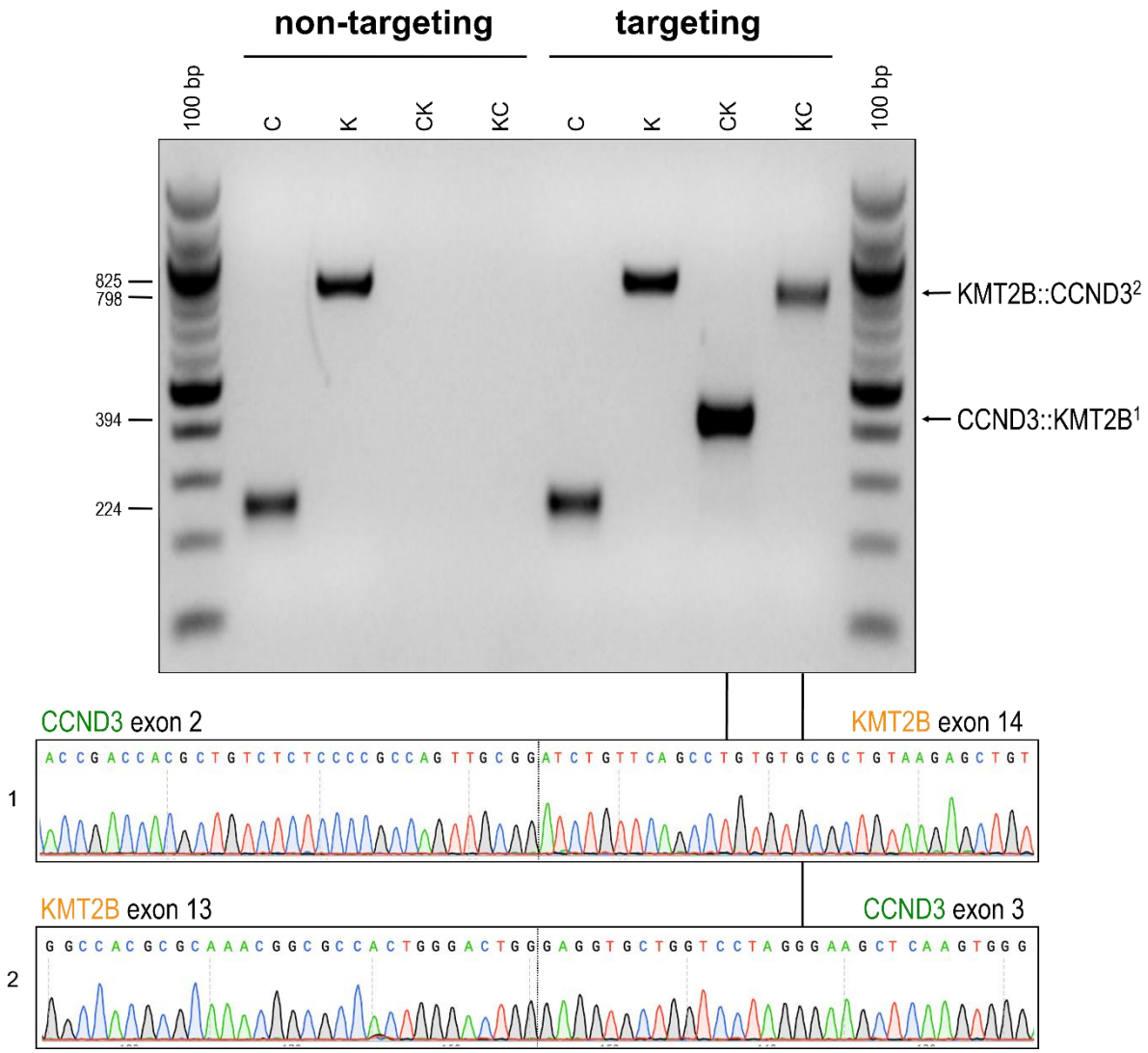

Figure 18. Induction of DNA double strand breaks in the presence of an artificial fusion RNA. HEK293T cells with the *KMT2B*/*CCND3* proximity system and non-targeting control cells were transfected with Cas9 preloaded with sgRNA for *KMT2B* and *CCND3* together with the artificial *KMT2B::CCND3* fusion RNA. The cells were harvested 3 – 5 days after the DNA damage was induced and the RNA and DNA from the same dish was extracted. To analyse a possible chromosomal translocation, RT-PCR experiments were performed to find resulting fusion transcripts. The band sizes are indicated on the left side, the sequence files show the identity of the two fusion transcripts found in the HEK293T cells with proximity between *KMT2B* and *CCND3*. *CCND3* (C), *KMT2B* (K), *KMT2B::CCND3* (KC), *CCND3::KMT2B* (CK).

Because the fusion transcripts indicated the presence of a chromosomal translocation, the genomic DNA was analysed for potential fusion alleles between *KMT2B* and *CCND3*. The subsequently performed genomic PCR experiments revealed the chromosomal

translocation at the genomic DNA level. PCR fragments spanning the genomic fusion sites were readily produced (Figure 19).

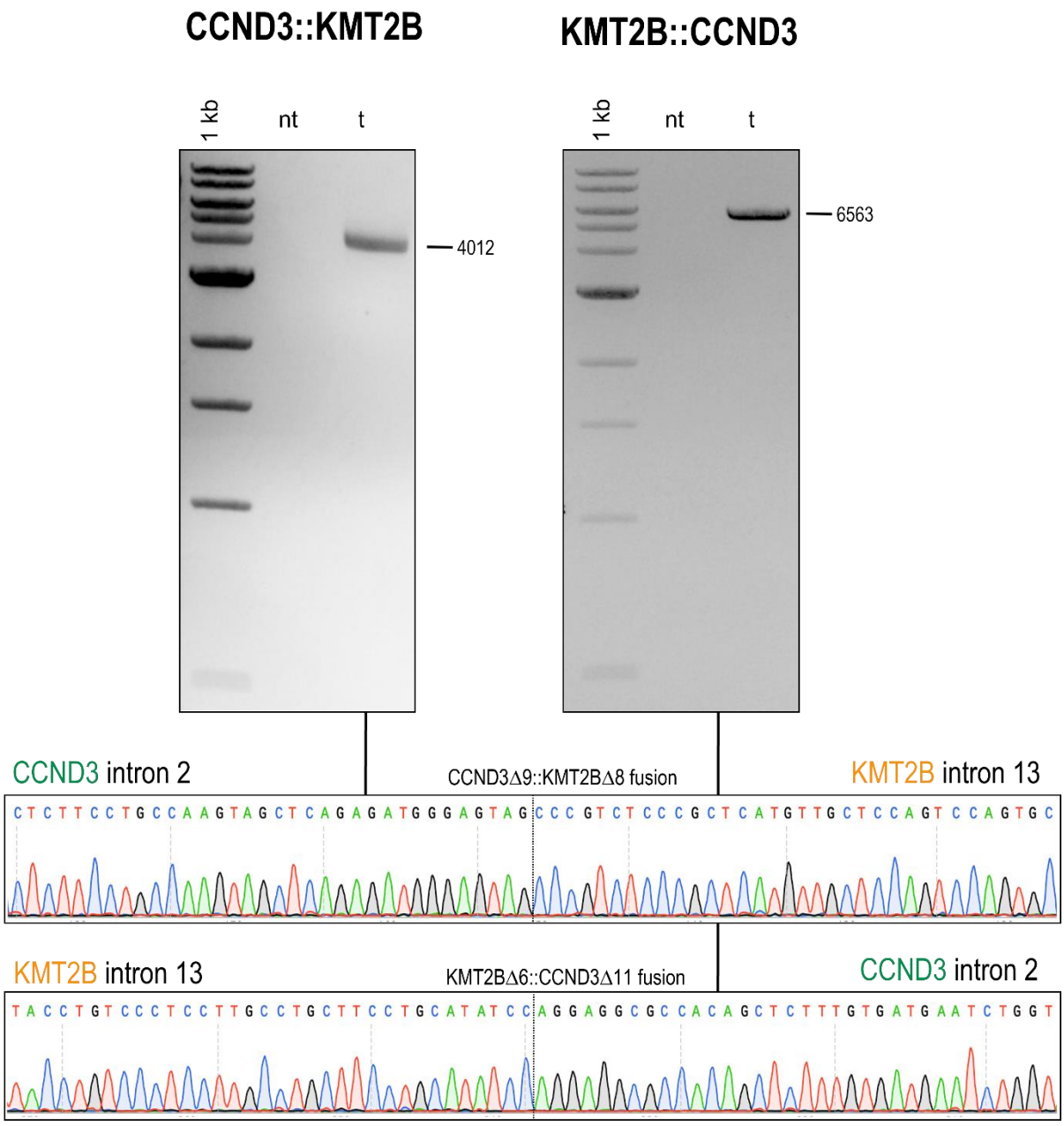

Figure 19. Chromosomal translocation of *CCND3* and *KMT2B*. HEK293T cells with the *KMT2B*/*CCND3* proximity system and non-targeting control cells were transfected with Cas9 preloaded with sgRNA for *KMT2B* and *CCND3* together with the artificial *KMT2B::CCND3* fusion RNA. The cells were harvested 3 – 5 days after the DNA damage was induced and the genomic DNA was extracted. For the non-targeting (nt) control no translocation is detectable, while the cell line with the targeting system (t) shows two bands for a balanced translocation. The sequences of the bands were analysed to validate the chromosomal translocation. The bottom panel shows one example of a sequence.

In order to differentiate between translocation events in distinct cells, the PCR bands were subcloned and individual clones were picked for DNA preparation and sequence analysis. As summarized in Figure 20, different fusion alleles were sequenced with the same hallmarks as for the t(4;11) translocation alleles. Namely, small deletion, the presence of MDR's and FD at the junctions of the fused chromosomes. This important experiments revealed that there are indeed 3 pre-requisites for the onset of chromosomal translocations: (1) gene proximity, (2) the presence of a specific RNA (NGEFT), and (3) a simultaneous DNA damage situation within both genes.

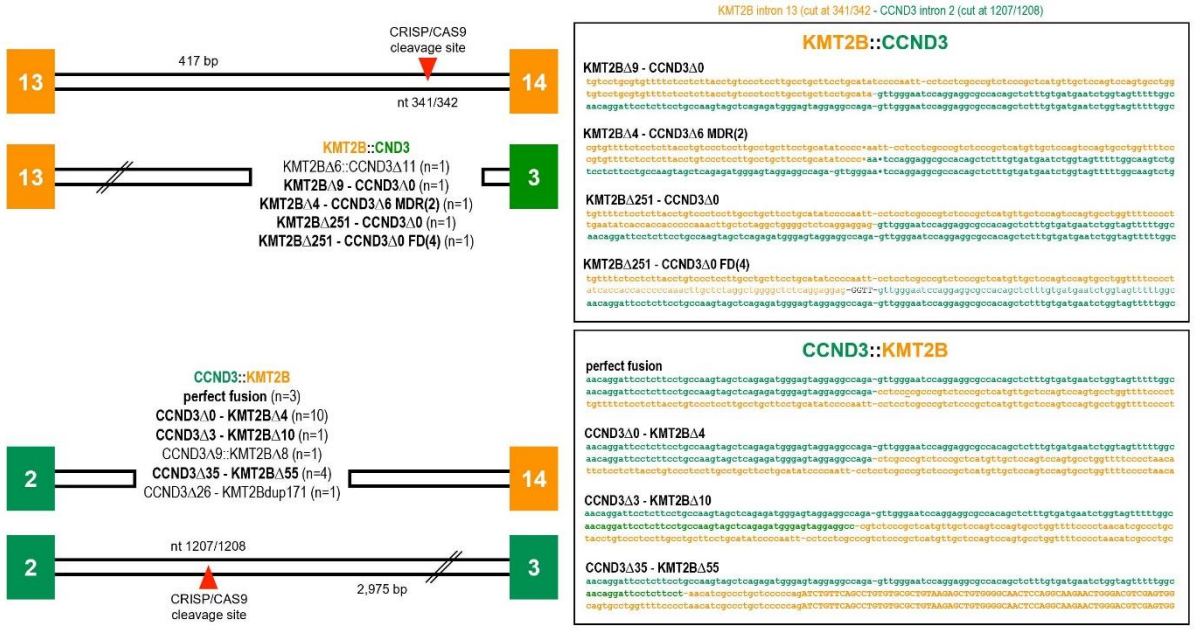

Figure 20. Overview of *KMT2B* and *CCND3* fusion. The bands of *KMT2B::CCND3* and *CCND3::KMT2B* were subcloned and the sequence was analysed. NHEJ mediated DNA repair errors are indicated by missing bases (Δ), mini-direct repeats (MDR) and filler DNA (FD).

5. Discussion

5.1. Genetic prerequisites for chromosomal translocations

It was demonstrated that chromosomal translocations arise due to DNA double strand breaks and subsequent DNA repair via the non-homologous end joining pathway (or the sub-classes C-NHEJ, A-EJ, MMEJ) [39, 42]. DNA double strand breaks may occur anywhere in the genome, but it seems that specific genes may exhibit a certain genetic instability [63]. The underlying mechanism was investigated by different laboratories proposing SAR/MAR chromatin loop structures [41, 73], a DNase I hypersensitive site, a Topoisomerase II site [71], an internal transcription initiation site [72] or site-specific DNA cleavage due to early apoptotic processes [217-219]. Such illegitimate recombination events between two genes will always lead to reciprocal fusion genes. Since the DNA lesion mostly occurs in introns, transcription of the two resulting fusion genes will lead to pre-mRNA in which exons of both genes are spliced together, thereby creating two novel fusion mRNAs that encode chimeric fusion proteins with a high oncogenic potential. That the formation of chromosomal translocations is a general cancer-causing mechanism is underlined by the fact that until today ~800 different chromosomal translocations have been described in hematological malignancies, while ~80 in solid tumors [220].

Although the occurrence of chromosomal translocations was investigated from many different groups and in different types of cancer, essential steps are still unknown. Why do cells fuse the open ends of two different chromosomes instead of the matching ends and why are particular genes recurrently affected in different individuals. To address all these questions, the hypothesized potential reasons leading to a translocation were imitated in a cellular test system. This test system should include the possibility to experimentally induce gene proximity, transcription within the same transcription factory, the formation of PTTs and resulting NGEFTs and the possibility to induce DNA damage situations.

5.2. Gene proximity, DNA damage and PTTs as drivers for non-genomically encoded fusion transcripts

The interphase nucleus is organized in the so called chromosome territories. Each chromosome occupies a defined, non-random region within the 3-dimensional space of the cell´s nucleus [163]. This spatial arrangement influences the proximity of genes and is specific for different tissues and cell types [185-187]. If the proximity of different genes plays a role in the illegitimate repair of two broken chromosomes, a cell type-specific chromosomal arrangement could be in part a rational explanation for the fact, that most known chromosomal translocations occur in a cell type-specific manner. Another interesting finding of the past were fusion transcripts of known chromosomal translocation partner genes in isolated blood cells of healthy individuals, in the absence of specific chromosomal translocations at the genomic level [26]. There are plenty of publications on this topic, but we had investigated this in the past as well: PBMCs from peripheral blood of 10 healthy individuals were tested for fusion transcripts of *KMT2A::AFF1* and *AFF1::KMT2A*, in 9 out of 10 samples there was a fusion transcript for *KMT2A-ex9*::*AFF1-ex4* while no fusion transcript for the reciprocal *AFF1::KMT2A* was present (Figure 21). Such fusion transcripts in healthy individuals were also identified for chromosomal translocations like *BCR::ABL* [6, 7], *TEL::AML1* [8, 9], *AML1::ETO* [9], *PML::RARα* [10], *NPM::ALK* [11, 12] and *ATIC::ALK* [12]. Based on the fact that they were not deriving from rearranged chromosomes, these transcripts were termed "non-genomically encoded fusion transcripts" (NGEFTs) since they are produced in the absence of an underlying chromosomal translocation. In the past we have shown that these NGEFTs are deriving from *trans*-splicing events due to gene proximity within the 3-dimensional space of the cell´s nucleus.

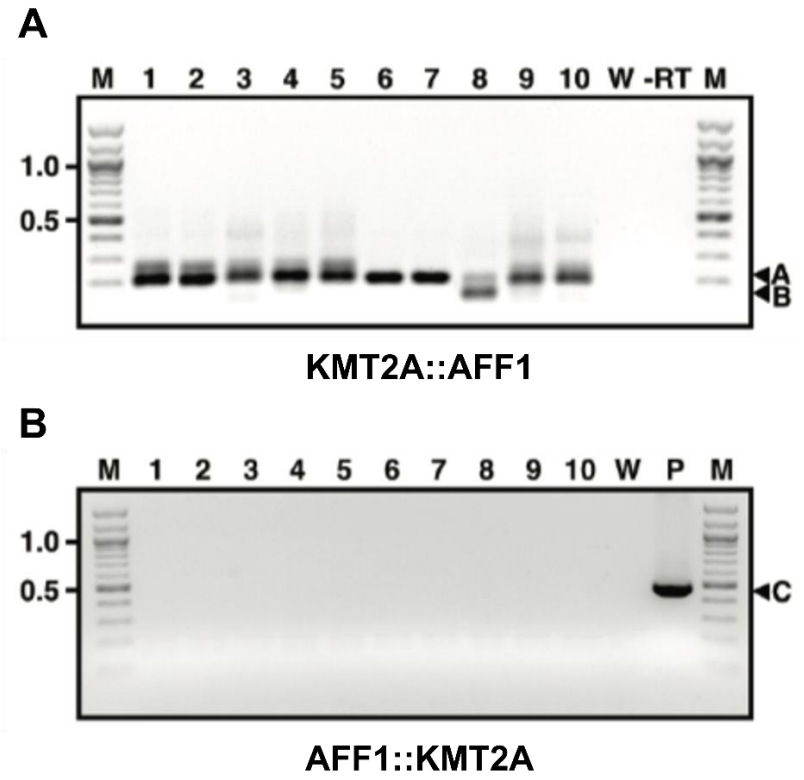

Figure 21. *KMT2A::AFF1* fusion transcripts in healthy individuals. A) Nested RT-PCR experiments of 10 different healthy blood donors (1 – 10). Sequence analysis of the PCR amplicons revealed a *KMT2A-ex9::AFF1-ex4* fusion transcript (A) in 9 out of 10 individuals, while in patient 8 *KMT2A* exon 8 was fused to *AFF1* exon 4 (B). W: water control, -RT: minus RT control, 100 bp ladder. B) Nested RT-PCR experiments revealed no *AFF1::KMT2A* fusion transcript in neither of the same volunteers (1 – 10). W: water control, P: positive control, 100 bp ladder. [26]

In order to investigate the influence of gene proximity on the formation of *KMT2A* and *AFF1* derived fusion transcripts, a dead Cas9 based gene proximity system has been developed which allows to induce gene proximity in mammalian cells in a controlled fashion. We have chosen the cell line HEK293T, because in this cell line none of the tested genes (*AFF1, CCDN3, KMT2A and KMT2B*) were in close proximity. To target two different regions in the genome, we used available dead Cas9 (dCas9) variants from two different species, one from *Streptococcus pyogenes* (SpCas9) and one from *Staphylococcus aureus* (SaCas9). Both mutant Cas9 proteins need different scaffold sequences, and thus, can be used simultaneously within the same cell. The different dCas9 were both fused to a protein domain that enabled chemical dimerization (either PYL1 or

ABI1) via abscisic acid (ABA). Both full-length proteins are used in plants to enable their dimerization in the presence of the plant phytohormone ABA.

The idea was to use specific guide RNAs that either bind *KMT2A* intron 9 or *AFF1* intron 3 in order to rearrange the chromosomes 4 and 11 inside the space of the nucleus by the administration of ABA. To validate the successful rearrangement of intact chromosomes after ABA treatment for 14 days, chromosome conformation capture (3C) experiments should be performed to validate gene proximity of *KMT2A* and *AFF1* or *KMT2B* and *CCND3*, respectively. In fact, after 14 days of ABA treatment *KMT2A* and *AFF1* and *KMT2B* and *CCND3* were in close proximity, shown by the PCR amplicons in Figure 11. No such amplicon could be found in the control cell line that expressed the non-targeting guide RNAs.

A subsequently performed RT-PCR experiment revealed the presence of an *KMT2A-ex9::AFF1-ex4* fusion RNA that was not present in the original HEK293T cell line or the non-targeting control cells (Figure 13). This highlighted that gene proximity between the two genes is one of the essential pre-conditions for the formation of *trans*-spliced fusion transcripts. The second prerequisite to generate such NGEFTs is the ability of certain genes in our genome to produce PTTs. PTTs all carry a non-saturated splice donor site which allows such transcripts to perform either intragenic or intergenic splice reactions. Since transcriptional processes are taking place in transcription factories, intergenic *trans*-splicing is a valid option if two genes are transcribed in the same transcription factory. As a matter of fact, it has been shown in the past that *KMT2A* and *AFF1* are transcribed within the same transcription factory [22-24] and similarly, other genes that are frequently involved in chromosomal translocations (e.g. *BCR* and *ABL* [33, 34], *cMYC* and *IgH* [35, 36]. Another important fact in these situations is that only one fusion transcript can be monitored, while the reciprocal fusion transcript is always missing. To this end, gene proximity allows the production of NGEFTs in a tissue-specific manner, if one of the two involved genes has the capability to produce PTTs.

Additionally, when doing the same experiment with *KMT2B* and *CCND3* , no such NGEFT could be shown. This is due to the fact that we have shown for a series of human genes that they do not produce PTTs and KMT2B and CCND3 were two of these genes. Therefore, this gene pair was chosen as internal control for all our experiments. Although gene proximity between *KMT2B* and *CCND3* was induced by our dCas9-mediated dimerization system (Figure 11), no fusion transcripts (*KMT2B::CCND3* nor a

CCND3::KMT2B) were detectable in RT-PCR experiments (Figure 13). The only alternative explanation for our findings would be that *KMT2A* and *AFF1* - but not *KMT2B* and *CCND3* - are transcribed within the same transcription factory, making their *trans*-splicing possible (Figure 7). But this was assumed to be unlikely.

Interestingly, these premature terminated transcripts mostly terminate somewhere in the breakpoint cluster region (BCR) (Figure 22 A). With an amount of about 10 to 20 % relative to the full length transcripts the formation of these premature terminated transcripts is not a rare event (Figure 22 B). The PTTs harbour an unsaturated splice donor site which could be the reason for alternative splicing events like intra- or intergenic *trans*-splicing (Figure 22 C). All this confirms the suspicion that in addition to gene proximity, the ability to generate premature terminated transcripts is essential for the formation of non-genomically encoded fusion transcripts. Although the comparison of a pair of genes generating PTTs with a pair of genes not generating PTTs is a confirmation for their importance in the formation of NGEFTs, a sure proof would be a negative control experiment. The naturally occurring PTTs of *KMT2A* and *AFF1* should be eliminated to see, whether a NGEFT is still forming after the induction of proximity. Because this experiment would be very challenging, due to the fact that these PTTs are naturally occurring and are produced by any cell, the experiments performed in this PhD thesis are clearly supporting the notion about NGEFT production.

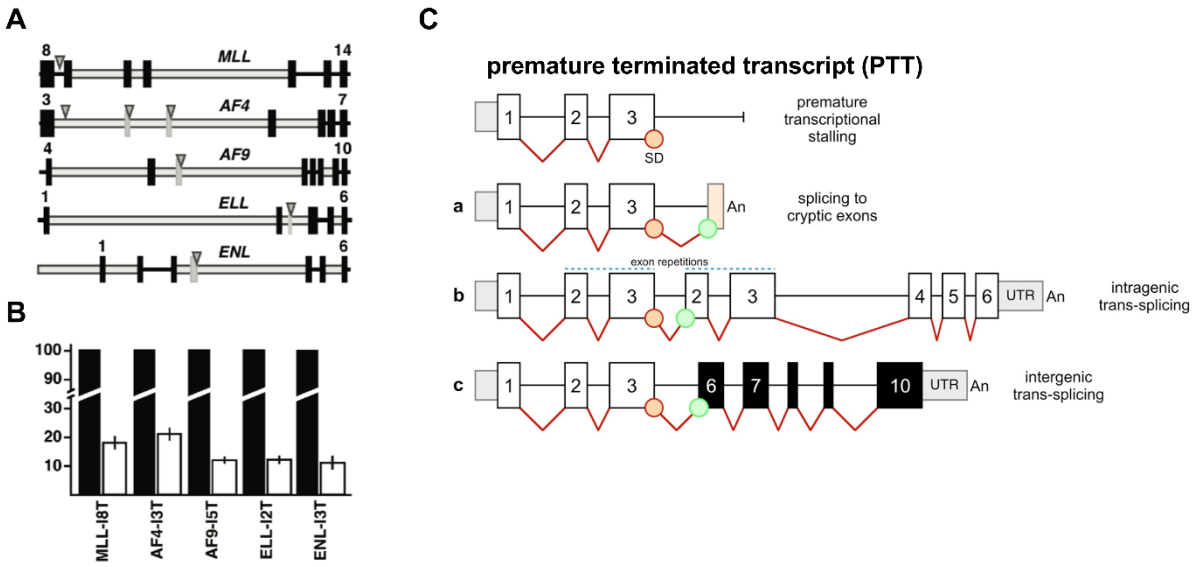

Figure 22. Premature terminated transcripts of genes undergoing chromosomal translocations. A) Breakpoint cluster regions (BCR) of the investigated genes *MLL, AF4, MLLT3,*

ELL and *MLLT1.* The black boxes show the exons separated by introns (black lines). The light grey bars mark the known breakpoint cluster regions and the numbers indicate exons flanking the individual BCRs. The grey vertical boxes show cryptic exons. The triangles show cryptic poly-A sites in early-terminated transcripts. B) Quantification of premature terminated transcripts (white bars) in relative comparison to full length transcripts (black bars). C) Premature transcriptional stalling leads to premature terminated transcripts (PTTs). With their unsaturated splice donor site, they can either splice to cryptic exons, to other exons from a transcript of the same gene (intragenic *trans*-splicing) or to an exon from the transcript of another gene (intergenic *trans*-splicing) leading to NGEFTs. [26, 124]

In order to replicate the situation leading to chromosomal translocations in patients, we used HEK293T cells to investigate the influence of gene proximity. Before, the investigated genes (*AFF1*, *CCND3*, *KMT2A* and *KMT2B*) were shown to not be in proximity in the nuclei of this cell line. This has been experimentally shown by chromosome conformation capture with a non-targeting control HEK293T cell line (Figure 11). After inducing gene proximity, demonstrating that this causes the NGEFT *KMT2A-ex9::AFF1-ex4,* we finally had to induce a DNA damage situation as the final prerequisite for the formation of a chromosomal translocation. The necessary DNA double strand breaks were introduced within the BCR (known for KMT2A and AFF1, unknown for KMT2B and CCND1) by CRISPR/Cas9 mediated DNA cleavage. The introns in KMT2B and CCND3 were chosen that a potentially induced chromosomal translocation would result in functional fusion proteins.

The introduction of a DNA double strand break in *KMT2A* and *AFF1* was not sufficient to induce a chromosomal translocation when gene proximity is not given, as shown by the non-targeting control cell line (Figure 16). With induced gene proximity between *KMT2A* and *AFF1,* the DNA damage led to a chromosomal translocation and the expected fusion transcripts for *KMT2A::AFF1* and *AFF1::KMT2A* were readily detectable (Figure 15, Figure 16). Interestingly, no translocation or fusion transcripts were detectable after the induction of DNA double strand breaks in HEK293T cells with the KMT2B/CCND3 gene proximity system (Figure 13). This is due to the fact that gene proximity between those 2 genes was not resulting in a *trans*-spliced fusion mRNA due to their missing PTT capability.

This leads to the conclusion that the NGEFT is also essential for the formation of a chromosomal translocation. This finding would mean that the ability of a gene to

express PTTs could be a necessary property for a gene to undergo illegitimate DNA repair and thereby generate a chromosomal translocation. But because there are also genes involved in translocations that do not exhibit PTT formation, this could only explain the illegitimate repair for some of them. Furthermore, this would mean, finding genes that express PTTs could identify certain genes that have a higher risk of being a driver for chromosomal translocations.

To validate this assumption, we created an artificial fusion RNA consisting of *KMT2B* exon 9 – 13 and *CCND3* exon 3 – 5 which was co-transfected together with preloaded Cas9 protein into the *KMT2B/CCND3* gene proximity system to cause two DNA double strand breaks and mediate the subsequent DNA repair. In contrast to the first experiment, this time a chromosomal translocation occurred (Figure 19), although this translocation has never been described in the literature or was found in any kind of disease.

This final experiment demonstrated that gene proximity, the presence of a specific NGEFT, and the occurrence of a DNA damage situation was sufficient to induce a chromosomal translocation (Figure 19), resulting in the production of both reciprocal fusion transcripts (Figure 18). With this final experiments the 3 prerequisites essential for the formation of a chromosomal translocation were classified. These three prerequisites are 1) gene proximity, 2) PTTs and the resulting NGEFTs and 3) a DNA double strand break in both involved genes. When either one of the three pre-conditions was missing no chromosomal translocation was formed. Without the induced DNA damage, no illegitimate repair could happen, without induced proximity no NGEFT is forming and the DNA ends are probably too far away for an illegitimate repair and without the NGEFTs no chromosomal translocation occurred, although DNA damage and gene proximity were induced.

5.3. NGEFTs as drivers for illegitimate repair

That gene proximity and DNA damage are the essential key elements for the formation of a chromosomal translocation has been posed in the past by several groups. It is pretty clear that without a DNA damage situation in 2 genes no chromosomal translocation may occur. It is also clear that chromosomal translocations can only occur when 2 genes are in close proximity. Here, we show that the third prerequisite, namely the presence of

an NGEFT is bridging the two different chromosomes to get a repair process performed that leads to a chromosomal translocation. Over ten years ago our group proposed a mechanism involving the *KMT2A-ex9::AFF1-ex4* fusion transcript in the repair process of the DNA damage (Figure 23). In case of a DNA damage in both breakpoint cluster regions, the fusion RNA is annealing to both DNA strands, aligning them for the DNA damage response machinery. The single strands are then repaired by DNA Ligase IV which is a part of the NHEJ system. The other two DNA strands are repaired by NHEJ in a second repair process, resulting in a balanced t(4;11). The idea that RNA plays a role in the DNA damage response through currently unresolved mechanisms was growing over the years and got increasingly important [137, 139-144]. Different groups could show the direct recruitment of repair factors by nuclear RNAs [146]. There are mainly two theories about the origin of RNAs that are involved in the DNA repair process. Either a bi-directional transcription after damage induction, which could be shown in reporter systems [140, 146, 159] but could not be verified in using endogenous breaks in mammalian cells [140, 143]. The other theory says that the RNA transcripts were produced prior to the DNA double strand break and remain in vicinity of the break. The usage of RNA transcripts produced prior to the DNA damage could be demonstrated to occur in a NHEJ-dependent manner in yeast [43-45]. This theory gained popularity because an RNA transcript that was produced prior to a DNA damage could serve as a template for the repair. This mechanism would be similar to HR but it would be independent from the cell cycle. This theory would also fit for the illegitimate repair of *KMT2A* and *AFF1* after DNA damage (Figure 23). The experiment could show the essential role of the RNA fusion transcript that has been found in the blood of healthy blood donors (Figure 21). The combination of DNA damage and proximity in the presence of the artificial fusion RNA led to the formation of a t(6;19), a chromosomal translocation that has never been observed in nature yet. The artificial *KMT2B::CCND3* fusion RNA was able to redirect an "ordinary DNA repair process" into a "wrong DNA repair process" (Figure 19). This would not only have an impact on *KMT2A* rearrangements but also on other rearrangement, where these fusion transcripts have been found in the samples of healthy donors like *BCR::ABL* [6, 7], *TEL::AML1* [8, 9], *AML1::ETO* [9], *PML::RARα* [10], *NPM::ALK* [11, 12] and *ATIC::ALK* [12]. The fact that these NGEFTs have only been found for genes that frequently undergo chromosomal translocations could thereby be a key driver for their occurrence and explain why a pure selection hypothesis

was not sufficient to explain the frequency and recurrence of these genetic alterations. The involvement of the nuclear RNA has effects on the overall function of nuclear RNA. They could be part of a mechanism for genetic stability where nuclear RNAs act as template für NHEJ repair in a cell cycle independent fashion. This would increase genetic stability because the error prone NHEJ pathway would be supported by nuclear RNA templates to decrease the probability of errors. But unfortunately, this mechanism for genetic stability could also be the reason for the formation of chromosomal trans-locations in an “accidental fashion”. These results point out that the capability to produce PTT's is defining a set of "critical genes" in our genome that harbour the risk for genetic rearrangements. Therefore, it would be of great interest to analyse RNA-Seq data to identify those genes by their ability to generate NGEFTs in a tissue-specific manner. Such genes are by definition potential proto-oncogenes and can be used for diagnostic purposes to detect ovel cancer genes when rearranged or otherwise genetically modified.

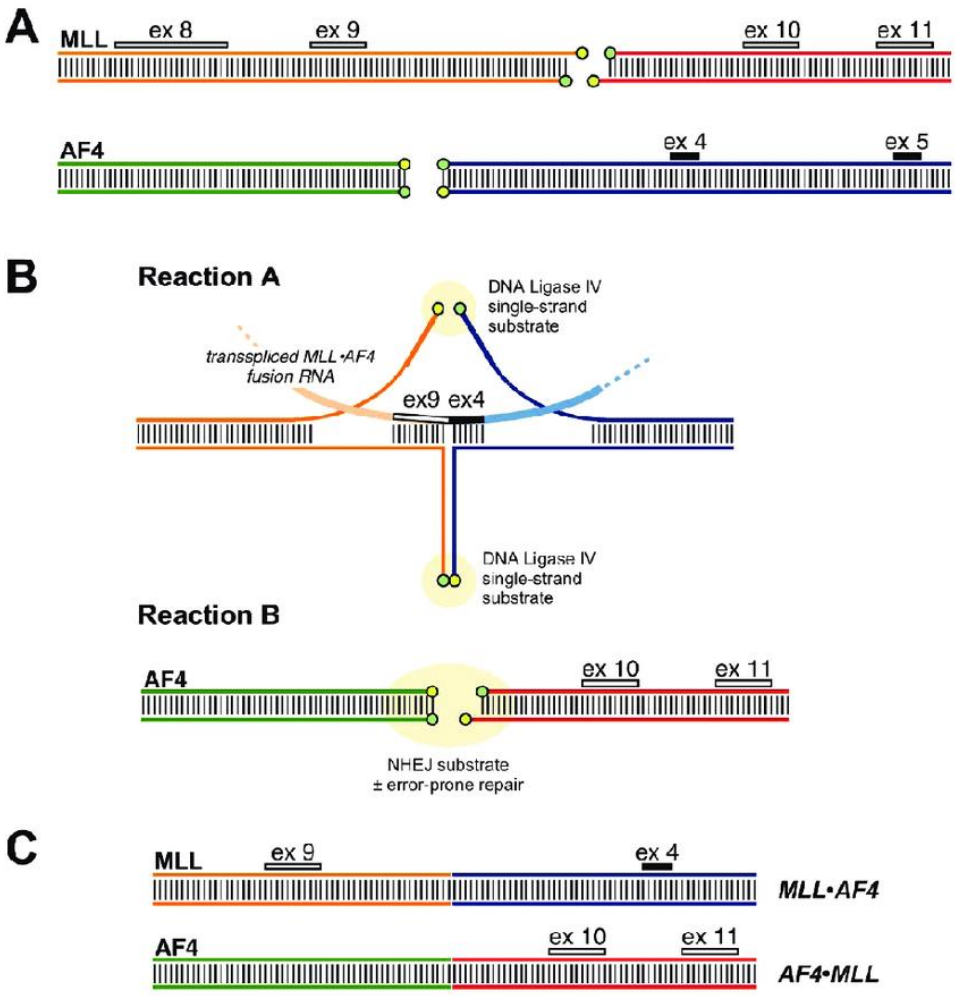

Figure 23. Proposed mechanism for the RNA-guided *KMT2A* and *AFF1* repair. A) DNA double strand break in *KMT2A* and *AFF1*. B) The non-genomically encoded fusion transcript aligns the two chromosomes via *KMT2A* exon 9 and *AFF1* exon 4. The single stranded ends of both chromosomes are ligated by DNA Ligase IV, a part of the NHEJ machinery. The repair is resulting in the *KMT2A::AFF1* fusion, the reciprocal fusion *AFF1::KMT2A* is also carried out the NHEJ repair system. C) The final product is a balanced t(4;11) translocation as diagnosed in most t(4;11) patients. [26]

5.4. Evaluation of the test system

In order to investigate the formation of chromosomal translocations the here used test system was set up to replicate the natural situation and make it comparable to another cell types by using non-targeting controls. With this system the influence of proximity, DNA damage and NGEFTs could be observed, making it an interesting tool to investigate different genes frequently undergoing translocations. By simply exchanging the sgRNAs

this system could be applied to many different genes. A good test system simulates the natural situation as best as it can. Because naturally occurring chromosomal translocations are formed by illegitimate repair via the error prone NHEJ pathway, there are differences at the break points. This includes mutations like insertions with filler DNA, inversions, deletion and mini-direct repeats. Although, the illegitimate repair takes place in an RNA templated NHEJ mechanism, the errors fit to the hypothesis, because in this case the RNA is annealing to exons and the DNA ends are finding each other, not having a direct template at the point of repair. In order to see, whether the chromosomal translocations produced with the induction of proximity and the NGEFTs reproduce the natural situation, the bands from the long range PCRs (Figure 16, Figure 19) were subcloned to get single sequences. In this mixture of different translocation events within different cells, there was a variety of different mutations at the break point (Figure 17). This was not only the case for the t(4;11) but also for the t(6;19) that has never been described in nature before (Figure 20). In the different sequences there were filler DNAs of different lengths, different numbers of deleted bases and also mini-direct repeats (Figure 17, Figure 20). When every single cell with a translocation event is seen as a patient, this experiment would show the naturally occurring pattern of break point mutations for chromosomal translocations. This shows that the test system is leading to results that are quite similar to the natural process of translocations. After the induction of proximity with subsequent DNA damage for *KMT2A* and *AFF1* there is a complete population of different break points, with different mutations around the break points. That this is also the case for *KMT2B* and *CCND3* with the addition of an artificial fusion RNA shows that the illegitimate repair may work after a fixed principle when all prerequisites are present.

6. Literature

1. DiMartino, J.F. and M.L. Cleary, *MLL rearrangements in haematological malignancies: lessons from clinical and biological studies.* British Journal of Haematology, 1999. **106**(3): p. 614-626.
2. Pui, C.H., et al., *Outcome of treatment in childhood acute lymphoblastic leukaemia with rearrangements of the 11q23 chromosomal region.* Lancet, 2002. **359**(9321): p. 1909-15.
3. Pui, C.H., et al., *Biology, risk stratification, and therapy of pediatric acute leukemias: an update.* J Clin Oncol, 2011. **29**(5): p. 551-65.
4. Meyer, C. and T. Burmeister, *The MLL recombinome of acute leukemias in 2017.* 2018. **32**(2): p. 273-284.
5. Kowarz, E., et al., *Premature transcript termination, trans-splicing and DNA repair: A vicious path to cancer.* American journal of blood research, 2011. **1**: p. 1-12.
6. Biernaux, C., et al., *Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals.* Blood, 1995. **86**(8): p. 3118-22.
7. Bose, S., et al., *The presence of typical and atypical BCR-ABL fusion genes in leukocytes of normal individuals: biologic significance and implications for the assessment of minimal residual disease.* Blood, 1998. **92**(9): p. 3362-7.
8. Eguchi-Ishimae, M., et al., *Breakage and fusion of the TEL (ETV6) gene in immature B lymphocytes induced by apoptogenic signals.* Blood, 2001. **97**(3): p. 737-743.
9. Mori, H., et al., *Chromosome translocations and covert leukemic clones are generated during normal fetal development.* Proc Natl Acad Sci U S A, 2002. **99**(12): p. 8242-7.
10. Quina, A.S., et al., *PML-RARA fusion transcripts in irradiated and normal hematopoietic cells.* Genes, Chromosomes and Cancer, 2000. **29**(3): p. 266-275.
11. Beylot-Barry, M., et al., *Characterization of t(2;5) Reciprocal Transcripts and Genomic Breakpoints in CD30+ Cutaneous Lymphoproliferations.* Blood, 1998. **91**(12): p. 4668-4676.
12. Maes, B., et al., *The NPM-ALK and the ATIC-ALK fusion genes can be detected in non-neoplastic cells.* Am J Pathol, 2001. **158**(6): p. 2185-93.
13. Vellard, M., et al., *C-myb proto-oncogene: evidence for intermolecular recombination of coding sequences.* Oncogene, 1991. **6**(4): p. 505-14.
14. Caudevilla, C., et al., *Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12185-90.
15. Hirano, M. and T. Noda, *Genomic organization of the mouse Msh4 gene producing bicistronic, chimeric and antisense mRNA.* Gene, 2004. **342**(1): p. 165-77.
16. Ma, L., et al., *Identification and analysis of pig chimeric mRNAs using RNA sequencing data.* BMC Genomics, 2012. **13**: p. 429.
17. Li, H., et al., *A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells.* Science, 2008. **321**(5894): p. 1357-61.
18. Takahara, T., et al., *Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of homotypic trans-splicing.* J Biol Chem, 2000. **275**(48): p. 38067-72.
19. Flouriot, G., et al., *Natural trans-spliced mRNAs are generated from the human estrogen receptor-alpha (hER alpha) gene.* J Biol Chem, 2002. **277**(29): p. 26244-51.
20. Akopian, A.N., et al., *Trans-splicing of a voltage-gated sodium channel is regulated by nerve growth factor.* FEBS Lett, 1999. **445**(1): p. 177-82.
21. Li, B.L., et al., *Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes.* J Biol Chem, 1999. **274**(16): p. 11060-71.
22. Caldas, C., et al., *Exon scrambling of MLL transcripts occur commonly and mimic partial genomic duplication of the gene.* Gene, 1998. **208**(2): p. 167-176.

23. Marcucci, G., et al., *Detection of Unique ALL1 (MLL) Fusion Transcripts in Normal Human Bone Marrow and Blood: Distinct Origin of Normal versus Leukemic ALL1 Fusion Transcripts1.* Cancer Research, 1998. **58**(4): p. 790-793.
24. Uckun, F.M., et al., *Clinical Significance of MLL-AF4 Fusion Transcript Expression in the Absence of a Cytogenetically Detectable t(4;11)(q21;q23) Chromosomal Translocation.* Blood, 1998. **92**(3): p. 810-821.
25. Kamieniarz-Gdula, K. and N.J. Proudfoot, *Transcriptional Control by Premature Termination: A Forgotten Mechanism.* Trends in Genetics, 2019. **35**(8): p. 553-564.
26. Kowarz, E., et al., *Premature transcript termination, trans-splicing and DNA repair: a vicious path to cancer.* Am J Blood Res, 2011. **1**(1): p. 1-12.
27. Misteli, T., *Self-organization in the genome.* Proceedings of the National Academy of Sciences, 2009. **106**(17): p. 6885-6886.
28. Branco, M.R. and A. Pombo, *Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations.* PLoS Biol, 2006. **4**(5): p. e138.
29. Cremer, T. and M. Cremer, *Chromosome territories.* Cold Spring Harb Perspect Biol, 2010. **2**(3): p. a003889.
30. Phatnani, H.P. and A.L. Greenleaf, *Phosphorylation and functions of the RNA polymerase II CTD.* Genes & development, 2006. **20**(21): p. 2922-2936.
31. Iborra, F.J., et al., *Active RNA polymerases are localized within discrete transcription "factories' in human nuclei.* Journal of cell science, 1996. **109**(6): p. 1427-1436.
32. Martin, S. and A. Pombo, *Transcription factories: quantitative studies of nanostructures in the mammalian nucleus.* Chromosome Research, 2003. **11**(5): p. 461-470.
33. Biernaux, C., et al., *Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals.* 1995.
34. Kozubek, S., et al., *The topological organization of chromosomes 9 and 22 in cell nuclei has a determinative role in the induction of t(9,22) translocations and in the pathogenesis of t(9,22) leukemias.* Chromosoma, 1999. **108**(7): p. 426-435.
35. Osborne, C.S., et al., *Myc dynamically and preferentially relocates to a transcription factory occupied by Igh.* PLoS Biol, 2007. **5**(8): p. e192.
36. Osborne, C.S., et al., *Active genes dynamically colocalize to shared sites of ongoing transcription.* Nature Genetics, 2004. **36**(10): p. 1065-1071.
37. Meaburn, K.J., T. Misteli, and E. Soutoglou, *Spatial genome organization in the formation of chromosomal translocations.* Semin Cancer Biol, 2007. **17**(1): p. 80-90.
38. Glukhov, S.I., et al., *The broken MLL gene is frequently located outside the inherent chromosome territory in human lymphoid cells treated with DNA topoisomerase II poison etoposide.* PLoS One, 2013. **8**(9): p. e75871.
39. Reichel, M., et al., *Fine structure of translocation breakpoints in leukemic blasts with chromosomal translocation t(4;11): the DNA damage-repair model of translocation.* Oncogene, 1998. **17**(23): p. 3035-44.
40. Reichel, M., et al., *Rapid Isolation of Chromosomal Breakpoints from Patients with t(4;11) Acute Lymphoblastic Leukemia: Implications for Basic and Clinical Research1.* Cancer Research, 1999. **59**(14): p. 3357-3362.
41. Hensel, J.P., et al., *Breakpoints of t(4;11) translocations in the human MLL and AF4 genes in ALL patients are preferentially clustered outside of high-affinity matrix attachment regions.* J Cell Biochem, 2001. **82**(2): p. 299-309.
42. Gillert, E., et al., *A DNA damage repair mechanism is involved in the origin of chromosomal translocations t(4;11) in primary leukemic cells.* Oncogene, 1999. **18**(33): p. 4663-71.
43. Chakraborty, A., et al., *Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes.* Nat Commun, 2016. **7**: p. 13049.
44. Keskin, H., et al., *Transcript-RNA-templated DNA recombination and repair.* Nature, 2014. **515**(7527): p. 436-9.

45. Wei, L., et al., *DNA damage during the G0/G1 phase triggers RNA-templated, Cockayne syndrome B-dependent homologous recombination.* Proceedings of the National Academy of Sciences, 2015. **112**(27): p. E3495-E3504.
46. Liang, F.-S., W.Q. Ho, and G.R. Crabtree, *Engineering the ABA Plant Stress Pathway for Regulation of Induced Proximity.* Science Signaling, 2011. **4**(164): p. rs2-rs2.
47. Doulatov, S., et al., *Hematopoiesis: A Human Perspective.* Cell Stem Cell, 2012. **10**(2): p. 120-136.
48. Maximow, A.A., *Der Lymphozyt als gemeinsame Stammzelle der verschiedenen Blutelemente in der embryonalen Entwicklung und im postfetalen Leben der Säugetiere.* Cellular Therapy and Transplantation, 2009. **1**(3): p. 9-13.
49. Patel, S.R., J.H. Hartwig, and J.E. Italiano, Jr., *The biogenesis of platelets from megakaryocyte proplatelets.* J Clin Invest, 2005. **115**(12): p. 3348-54.
50. Periayah, M.H., A.S. Halim, and A.Z. Mat Saad, *Mechanism Action of Platelets and Crucial Blood Coagulation Pathways in Hemostasis.* Int J Hematol Oncol Stem Cell Res, 2017. **11**(4): p. 319-327.
51. Auffray, C., M.H. Sieweke, and F. Geissmann, *Blood monocytes: development, heterogeneity, and relationship with dendritic cells.* Annu Rev Immunol, 2009. **27**: p. 669-92.
52. Serbina, N.V., et al., *Monocyte-mediated defense against microbial pathogens.* Annu Rev Immunol, 2008. **26**: p. 421-52.
53. Bleyer, M., et al., *Morphology and staining behavior of neutrophilic and eosinophilic granulocytes of the common marmoset (Callithrix jacchus).* Exp Toxicol Pathol, 2016. **68**(6): p. 335-43.
54. Mayadas, T.N., X. Cullere, and C.A. Lowell, *The multifaceted functions of neutrophils.* Annu Rev Pathol, 2014. **9**: p. 181-218.
55. Valent, P., et al., *Contemporary consensus proposal on criteria and classification of eosinophilic disorders and related syndromes.* J Allergy Clin Immunol, 2012. **130**(3): p. 607-612.e9.
56. Knol, E., et al., *The role of basophils in allergic disease.* The European Respiratory journal. Supplement, 1996. **22**: p. 126s-131s.
57. Ward, J.M., S. Cherian, and M.A. Linden, *19 - Hematopoietic and Lymphoid Tissues*, in *Comparative Anatomy and Histology (Second Edition)*, P.M. Treuting, S.M. Dintzis, and K.S. Montine, Editors. 2018, Academic Press: San Diego. p. 365-401.
58. Emadi, A. and J. York Law, *Overview of Leukemia.* 2022.
59. SORENSEN, M.N., *Can We Talk? A Multiskills Approach to Communication.* TESOL Quarterly, 1994. **28**(1): p. 216-217.
60. Ziemin-van der Poel, S., et al., *Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(23): p. 10735-10739.
61. Pui, C.H., et al., *Clinical heterogeneity in childhood acute lymphoblastic leukemia with 11q23 rearrangements.* Leukemia, 2003. **17**(4): p. 700-706.
62. Cosgrove, M.S. and A. Patel, *Mixed lineage leukemia: a structure–function perspective of the MLL1 protein.* The FEBS Journal, 2010. **277**(8): p. 1832-1842.
63. Broeker, P.L., et al., *The mixed lineage leukemia (MLL) protein involved in 11q23 translocations contains a domain that binds cruciform DNA and scaffold attachment region (SAR) DNA.* Curr Top Microbiol Immunol, 1996. **211**: p. 259-68.
64. Allen, M.D., et al., *Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase.* Embo j, 2006. **25**(19): p. 4503-12.
65. Ernst, P., et al., *MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein.* Mol Cell Biol, 2001. **21**(7): p. 2249-58.
66. Jenuwein, T., et al., *SET domain proteins modulate chromatin domains in eu- and heterochromatin.* Cell Mol Life Sci, 1998. **54**(1): p. 80-93.

67. Malik, S. and S.R. Bhaumik, *Mixed lineage leukemia: histone H3 lysine 4 methyltransferases from yeast to human.* Febs j, 2010. **277**(8): p. 1805-21.
68. Wysocka, J., et al., *WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development.* Cell, 2005. **121**(6): p. 859-72.
69. Patel, A., et al., *A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex.* J Biol Chem, 2008. **283**(47): p. 32162-75.
70. Gu, Y., et al., *Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved in acute leukemia.* Cancer Res, 1994. **54**(9): p. 2327-30.
71. Strissel, P.L., et al., *An in vivo topoisomerase II cleavage site and a DNase I hypersensitive site colocalize near exon 9 in the MLL breakpoint cluster region.* Blood, 1998. **92**(10): p. 3793-803.
72. Scharf, S., et al., *Transcription linked to recombination: a gene-internal promoter coincides with the recombination hot spot II of the human MLL gene.* Oncogene, 2007. **26**(10): p. 1361-1371.
73. Strick, R., et al., *Common chromatin structures at breakpoint cluster regions may lead to chromosomal translocations found in chronic and acute leukemias.* Hum Genet, 2006. **119**(5): p. 479-95.
74. Marschalek, R., et al., *The structure of the human ALL-1/MLL/HRX gene.* Leuk Lymphoma, 1997. **27**(5-6): p. 417-28.
75. Meyer, C., et al., *Genomic DNA of leukemic patients: target for clinical diagnosis of MLL rearrangements.* Biotechnol J, 2006. **1**(6): p. 656-63.
76. Zaphiropoulos, P.G., *Trans-splicing in Higher Eukaryotes: Implications for Cancer Development?* Frontiers in genetics, 2011. **2**: p. 92-92.
77. Krause, M. and D. Hirsh, *A trans-spliced leader sequence on actin mRNA in C. elegans.* Cell, 1987. **49**(6): p. 753-61.
78. Horiuchi, T. and T. Aigaki, *Alternative trans-splicing: a novel mode of pre-mRNA processing.* Biology of the Cell, 2006. **98**(2): p. 135-140.
79. Rigatti, R., et al., *Exon repetition: a major pathway for processing mRNA of some genes is allele-specific.* Nucleic Acids Res, 2004. **32**(2): p. 441-6.
80. Frantz, S.A., et al., *Exon repetition in mRNA.* Proc Natl Acad Sci U S A, 1999. **96**(10): p. 5400-5.
81. Dorn, R., G. Reuter, and A. Loewendorf, *Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in Drosophila.* Proc Natl Acad Sci U S A, 2001. **98**(17): p. 9724-9.
82. Horiuchi, T., E. Giniger, and T. Aigaki, *Alternative trans-splicing of constant and variable exons of a Drosophila axon guidance gene, lola.* Genes Dev, 2003. **17**(20): p. 2496-501.
83. McManus, C.J., et al., *Global analysis of trans-splicing in Drosophila.* Proc Natl Acad Sci U S A, 2010. **107**(29): p. 12975-9.
84. Takahara, T., et al., *The trans-spliced variants of Sp1 mRNA in rat.* Biochem Biophys Res Commun, 2002. **298**(1): p. 156-62.
85. Yu, S., et al., *Brahma regulates a specific trans-splicing event at the mod(mdg4) locus of Drosophila melanogaster.* RNA Biology, 2014. **11**(2): p. 134-145.
86. Takahara, T., et al., *Delay in synthesis of the 3′ splice site promotes trans-splicing of the preceding 5′ splice site.* Mol Cell, 2005. **18**(2): p. 245-51.
87. Dixon, R.J., I.C. Eperon, and N.J. Samani, *Complementary intron sequence motifs associated with human exon repetition: a role for intragenic, inter-transcript interactions in gene expression.* Bioinformatics, 2007. **23**(2): p. 150-5.
88. Proudfoot, N.J., A. Furger, and M.J. Dye, *Integrating mRNA processing with transcription.* Cell, 2002. **108**(4): p. 501-12.
89. Konarska, M.M., R.A. Padgett, and P.A. Sharp, *Trans splicing of mRNA precursors in vitro.* Cell, 1985. **42**(1): p. 165-71.

90. Puttaraju, M., et al., *Spliceosome-mediated RNA trans-splicing as a tool for gene therapy.* Nat Biotechnol, 1999. **17**(3): p. 246-52.
91. Liu, X., et al., *Partial correction of endogenous DeltaF508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing.* Nat Biotechnol, 2002. **20**(1): p. 47-52.
92. Salzman, J., et al., *Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types.* PLoS One, 2012. **7**(2): p. e30733.
93. Jeck, W.R., et al., *Circular RNAs are abundant, conserved, and associated with ALU repeats.* Rna, 2013. **19**(2): p. 141-57.
94. Memczak, S., et al., *Circular RNAs are a large class of animal RNAs with regulatory potency.* Nature, 2013. **495**(7441): p. 333-8.
95. Zhang, X.O., et al., *Complementary sequence-mediated exon circularization.* Cell, 2014. **159**(1): p. 134-147.
96. Kristensen, L.S., et al., *The biogenesis, biology and characterization of circular RNAs.* 2019. **20**(11): p. 675-691.
97. Dudekula, D.B., et al., *CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs.* RNA Biol, 2016. **13**(1): p. 34-42.
98. Abdelmohsen, K., et al., *Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1.* RNA Biol, 2017. **14**(3): p. 361-369.
99. Du, W.W., et al., *Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2.* Nucleic Acids Res, 2016. **44**(6): p. 2846-58.
100. Pamudurti, N.R., et al., *Translation of CircRNAs.* Mol Cell, 2017. **66**(1): p. 9-21.e7.
101. Legnini, I., et al., *Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis.* Mol Cell, 2017. **66**(1): p. 22-37.e9.
102. Yang, L., et al., *Human acyl-coenzyme A:cholesterol acyltransferase 1 (acat1) sequences located in two different chromosomes (7 and 1) are required to produce a novel ACAT1 isoenzyme with additional sequence at the N terminus.* J Biol Chem, 2004. **279**(44): p. 46253-62.
103. Chatterjee, T.K. and R.A. Fisher, *Novel alternative splicing and nuclear localization of human RGS12 gene products.* J Biol Chem, 2000. **275**(38): p. 29660-71.
104. Finta, C. and P.G. Zaphiropoulos, *Intergenic mRNA molecules resulting from trans-splicing.* J Biol Chem, 2002. **277**(8): p. 5882-90.
105. Han, S.R., et al., *Targeted suicide gene therapy for liver cancer based on ribozyme-mediated RNA replacement through post-transcriptional regulation.* Molecular Therapy - Nucleic Acids, 2021. **23**: p. 154-168.
106. Sarmiento, C. and J.A. Camarero, *Biotechnological applications of protein splicing.* Current Protein and Peptide Science, 2019. **20**(5): p. 408-424.
107. Li, Y., *Split-inteins and their bioapplications.* Biotechnology letters, 2015. **37**(11): p. 2121-2137.
108. Vasudevan, S., S.W. Peltz, and C.J. Wilusz, *Non-stop decay—a new mRNA surveillance pathway.* BioEssays, 2002. **24**(9): p. 785-788.
109. Yao, P., et al., *Coding Region Polyadenylation Generates a Truncated tRNA Synthetase that Counters Translation Repression.* Cell, 2012. **149**(1): p. 88-100.
110. Hoque, M., et al., *Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing.* Nature Methods, 2013. **10**(2): p. 133-139.
111. Lianoglou, S., et al., *Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression.* Genes Dev, 2013. **27**(21): p. 2380-96.
112. Singh, I., et al., *Widespread intronic polyadenylation diversifies immune cell transcriptomes.* Nature Communications, 2018. **9**(1): p. 1716.
113. Lee, S.-H., et al., *Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia.* Nature, 2018. **561**(7721): p. 127-131.
114. Dubbury, S.J., P.L. Boutz, and P.A. Sharp, *CDK12 regulates DNA repair genes by suppressing intronic polyadenylation.* Nature, 2018. **564**(7734): p. 141-145.

115. Almada, A.E., et al., *Promoter directionality is controlled by U1 snRNP and polyadenylation signals.* Nature, 2013. **499**(7458): p. 360-3.
116. Preker, P., et al., *RNA exosome depletion reveals transcription upstream of active human promoters.* Science, 2008. **322**(5909): p. 1851-4.
117. Seila, A.C., et al., *Divergent transcription from active promoters.* Science, 2008. **322**(5909): p. 1849-51.
118. Ntini, E., et al., *Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality.* Nat Struct Mol Biol, 2013. **20**(8): p. 923-8.
119. Williamson, L., et al., *UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene.* Cell, 2017. **168**(5): p. 843-855.e13.
120. Singh, I., et al., *Widespread intronic polyadenylation diversifies immune cell transcriptomes.* 2018. **9**(1): p. 1716.
121. Alt, F.W., et al., *Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends.* Cell, 1980. **20**(2): p. 293-301.
122. Early, P., et al., *Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways.* Cell, 1980. **20**(2): p. 313-9.
123. Rogers, J., et al., *Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin mu chain.* Cell, 1980. **20**(2): p. 303-12.
124. Kowarz, E., T. Dingermann, and R. Marschalek, *Do non-genomically encoded fusion transcripts cause recurrent chromosomal translocations?* Cancers (Basel), 2012. **4**(4): p. 1036-49.
125. Li, H., et al., *Gene fusions and RNA trans-splicing in normal and neoplastic human cells.* Cell Cycle, 2009. **8**(2): p. 218-22.
126. Rickman, D.S., et al., *SLC45A3-ELK4 is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer.* Cancer Res, 2009. **69**(7): p. 2734-8.
127. Neves, H., et al., *The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation.* Blood, 1999. **93**(4): p. 1197-207.
128. Gingeras, T.R., *Implications of chimaeric non-co-linear transcripts.* Nature, 2009. **461**(7261): p. 206-11.
129. Chen, C. and R.D. Kolodner, *Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants.* Nat Genet, 1999. **23**(1): p. 81-5.
130. Brenneman, M.A., et al., *XRCC3 controls the fidelity of homologous recombination: roles for XRCC3 in late stages of recombination.* Mol Cell, 2002. **10**(2): p. 387-95.
131. Lio, Y.C., et al., *Human Rad51C deficiency destabilizes XRCC3, impairs recombination, and radiosensitizes S/G2-phase cells.* J Biol Chem, 2004. **279**(40): p. 42313-20.
132. Strout, M.P., et al., *The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia.* Proc Natl Acad Sci U S A, 1998. **95**(5): p. 2390-5.
133. Davis, A.J. and D.J. Chen, *DNA double strand break repair via non-homologous end-joining.* Transl Cancer Res, 2013. **2**(3): p. 130-143.
134. Lieber, M.R., *The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.* Annu Rev Biochem, 2010. **79**: p. 181-211.
135. Brandsma, I. and D.C. Gent, *Pathway choice in DNA double strand break repair: observations of a balancing act.* Genome Integr, 2012. **3**(1): p. 9.
136. Bader, A.S., et al., *The roles of RNA in DNA double-strand break repair.* British Journal of Cancer, 2020. **122**(5): p. 613-623.
137. Li, L., E.A. Monckton, and R. Godbout, *A role for DEAD box 1 at DNA double-strand breaks.* Mol Cell Biol, 2008. **28**(20): p. 6413-25.
138. Wei, W., et al., *A role for small RNAs in DNA double-strand break repair.* Cell, 2012. **149**(1): p. 101-12.

139. Francia, S., et al., *Site-specific DICER and DROSHA RNA products control the DNA-damage response.* Nature, 2012. **488**(7410): p. 231-5.
140. Michalik, K.M., R. Böttcher, and K. Förstemann, *A small RNA response at DNA ends in Drosophila.* Nucleic Acids Res, 2012. **40**(19): p. 9596-603.
141. Jain, A., et al., *DHX9 helicase is involved in preventing genomic instability induced by alternatively structured DNA in human cells.* Nucleic Acids Res, 2013. **41**(22): p. 10345-57.
142. Marin-Vicente, C., et al., *RRP6/EXOSC10 is required for the repair of DNA double-strand breaks by homologous recombination.* J Cell Sci, 2015. **128**(6): p. 1097-107.
143. Lu, W.-T., et al., *Drosha drives the formation of DNA: RNA hybrids around DNA break sites to facilitate DNA repair.* Nature communications, 2018. **9**(1): p. 1-13.
144. Cohen, S., et al., *Senataxin resolves RNA: DNA hybrids forming at DNA double-strand breaks to prevent translocations.* Nature communications, 2018. **9**(1): p. 1-14.
145. Hawley, B.R. and W.T. Lu, *The emerging role of RNAs in DNA damage repair.* 2017. **24**(4): p. 580-587.
146. Michelini, F., et al., *From "Cellular" RNA to "Smart" RNA: Multiple Roles of RNA in Genome Stability and Beyond.* 2018. **118**(8): p. 4365-4403.
147. Adamson, B., et al., *A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response.* Nat Cell Biol, 2012. **14**(3): p. 318-28.
148. Paulsen, R.D., et al., *A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability.* Mol Cell, 2009. **35**(2): p. 228-39.
149. Pryde, F., et al., *53BP1 exchanges slowly at the sites of DNA damage and appears to require RNA for its association with chromatin.* Journal of Cell Science, 2005. **118**(9): p. 2043-2055.
150. Welty, S., et al., *RAD52 is required for RNA-templated recombination repair in post-mitotic neurons.* J Biol Chem, 2018. **293**(4): p. 1353-1362.
151. Cristini, A., et al., *RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage.* Cell Rep, 2018. **23**(6): p. 1891-1905.
152. Mazina, O.M., et al., *Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair.* Mol Cell, 2017. **67**(1): p. 19-29.e3.
153. Amon, J.D. and D. Koshland, *RNase H enables efficient repair of R-loop induced DNA damage.* 2016. **5**.
154. Gan, W., et al., *R-loop-mediated genomic instability is caused by impairment of replication fork progression.* Genes Dev, 2011. **25**(19): p. 2041-56.
155. Yang, Z., et al., *RNase H1 Cooperates with DNA Gyrases to Restrict R-Loops and Maintain Genome Integrity in Arabidopsis Chloroplasts.* 2017. **29**(10): p. 2478-2497.
156. García-Pichardo, D., et al., *Histone mutants separate R loop formation from genome instability induction.* Molecular cell, 2017. **66**(5): p. 597-609. e5.
157. Ohle, C., et al., *Transient RNA-DNA hybrids are required for efficient double-strand break repair.* Cell, 2016. **167**(4): p. 1001-1013. e7.
158. Yasuhara, S., et al., *Mitochondria, endoplasmic reticulum, and alternative pathways of cell death in critical illness.* Crit Care Med, 2007. **35**(9 Suppl): p. S488-95.
159. Lee, H.C., et al., *qiRNA is a new type of small interfering RNA induced by DNA damage.* Nature, 2009. **459**(7244): p. 274-7.
160. Bonath, F., et al., *Next-generation sequencing reveals two populations of damage-induced small RNAs at endogenous DNA double-strand breaks.* Nucleic Acids Res, 2018. **46**(22): p. 11869-11882.
161. Miki, D., et al., *Efficient Generation of diRNAs Requires Components in the Posttranscriptional Gene Silencing Pathway.* 2017. **7**(1): p. 301.
162. Murmann, A.E., et al., *Local gene density predicts the spatial position of genetic loci in the interphase nucleus.* Experimental Cell Research, 2005. **311**(1): p. 14-26.

163. Misteli, T., *Beyond the sequence: cellular organization of genome function.* Cell, 2007. **128**(4): p. 787-800.
164. Rowley, M.J. and V.G. Corces, *Organizational principles of 3D genome architecture.* 2018. **19**(12): p. 789-800.
165. Dekker, J. and L. Mirny, *The 3D Genome as Moderator of Chromosomal Communication.* Cell, 2016. **164**(6): p. 1110-1121.
166. Pombo, A. and N. Dillon, *Three-dimensional genome architecture: players and mechanisms.* Nat Rev Mol Cell Biol, 2015. **16**(4): p. 245-57.
167. Lin, C., L. Yang, and M.G. Rosenfeld, *Molecular Logic Underlying Chromosomal Translocations, Random or Non-Random?*, in *Advances in Cancer Research*, K.D. Tew and P.B. Fisher, Editors. 2012, Academic Press. p. 241-279.
168. Finn, E. and T. Misteli, *Molecular basis and biological function of variability in spatial genome organization. Science 365 eaaw9498*. 2019.
169. Cattoni, D.I. and A.M. Cardozo Gizzi, *Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions.* 2017. **8**(1): p. 1753.
170. Rodriguez, J., et al., *Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity.* Cell, 2019. **176**(1-2): p. 213-226.e18.
171. Misteli, T., *The Self-Organizing Genome: Principles of Genome Architecture and Function.* Cell, 2020. **183**(1): p. 28-45.
172. Felsenfeld, G. and M. Groudine, *Controlling the double helix.* Nature, 2003. **421**(6921): p. 448-53.
173. Dekker, J. and T. Misteli, *Long-Range Chromatin Interactions.* Cold Spring Harb Perspect Biol, 2015. **7**(10): p. a019356.
174. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.
175. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.
176. Sexton, T., et al., *Three-dimensional folding and functional organization principles of the Drosophila genome.* Cell, 2012. **148**(3): p. 458-72.
177. Fraser, J., et al., *Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation.* Molecular systems biology, 2015. **11**(12): p. 852.
178. Shopland, L.S., et al., *Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence.* J Cell Biol, 2006. **174**(1): p. 27-38.
179. Wang, S., et al., *Spatial organization of chromatin domains and compartments in single chromosomes.* Science, 2016. **353**(6299): p. 598-602.
180. Croft, J.A., et al., *Differences in the localization and morphology of chromosomes in the human nucleus.* J Cell Biol, 1999. **145**(6): p. 1119-31.
181. Takizawa, T., K.J. Meaburn, and T. Misteli, *The meaning of gene positioning.* Cell, 2008. **135**(1): p. 9-13.
182. Marshall, W.F., et al., *Interphase chromosomes undergo constrained diffusional motion in living cells.* Curr Biol, 1997. **7**(12): p. 930-9.
183. Caridi, C.P., M. Plessner, and R. Grosse, *Nuclear actin filaments in DNA repair dynamics.* 2019. **21**(9): p. 1068-1077.
184. Chuang, C.-H., et al., *Long-range directional movement of an interphase chromosome site.* Current Biology, 2006. **16**(8): p. 825-831.
185. Parada, L.A., P.G. McQueen, and T. Misteli, *Tissue-specific spatial organization of genomes.* Genome Biol, 2004. **5**(7): p. R44.
186. Crosetto, N. and M. Bienko, *Radial Organization in the Mammalian Nucleus.* Front Genet, 2020. **11**: p. 33.

187. Shachar, S. and T. Misteli, *Causes and consequences of nuclear gene positioning.* 2017. **130**(9): p. 1501-1508.
188. Nikiforova, M.N., et al., *Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells.* Science, 2000. **290**(5489): p. 138-41.
189. Bartek, J. and J. Lukas, *DNA damage checkpoints: from initiation to recovery or adaptation.* Current Opinion in Cell Biology, 2007. **19**(2): p. 238-245.
190. Sandell, L.L. and V.A. Zakian, *Loss of a yeast telomere: Arrest, recovery, and chromosome loss.* Cell, 1993. **75**(4): p. 729-739.
191. Hozák, P., et al., *Site of transcription of ribosomal RNA and intranucleolar structure in HeLa cells.* Journal of cell science, 1994. **107**(2): p. 639-648.
192. Haeusler, R.A., et al., *Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes.* Genes & development, 2008. **22**(16): p. 2204-2214.
193. Pombo, A., et al., *Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III.* The EMBO journal, 1999. **18**(8): p. 2241-2253.
194. Falk, M., et al., *Chromatin dynamics during DSB repair.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2007. **1773**(10): p. 1534-1545.
195. Jakob, B., et al., *Live cell microscopy analysis of radiation-induced DNA double-strand break motion.* Proceedings of the National Academy of Sciences, 2009. **106**(9): p. 3172-3177.
196. Soutoglou, E., et al., *Positional stability of single double-strand breaks in mammalian cells.* Nature cell biology, 2007. **9**(6): p. 675-682.
197. Aten, J.A., et al., *Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains.* Science, 2004. **303**(5654): p. 92-95.
198. Brown, J.M., et al., *Association between active genes occurs at nuclear speckles and is modulated by chromatin environment.* The Journal of cell biology, 2008. **182**(6): p. 1083-1097.
199. Morey, C., C. Kress, and W.A. Bickmore, *Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression.* Genome research, 2009. **19**(7): p. 1184-1194.
200. Osborne, C.S., et al., *Active genes dynamically colocalize to shared sites of ongoing transcription.* Nature genetics, 2004. **36**(10): p. 1065-1071.
201. Faro-Trindade, I. and P.R. Cook, *Transcription factories: structures conserved during differentiation and evolution.* Biochemical Society Transactions, 2006. **34**(6): p. 1133-1137.
202. Sutherland, H. and W.A. Bickmore, *Transcription factories: gene expression in unions?* Nature Reviews Genetics, 2009. **10**(7): p. 457-466.
203. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
204. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna, *RNA-guided genetic silencing systems in bacteria and archaea.* Nature, 2012. **482**(7385): p. 331-8.
205. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-7.
206. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
207. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.
208. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.
209. Bikard, D., et al., *Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system.* Nucleic Acids Res, 2013. **41**(15): p. 7429-37.
210. Guilinger, J.P., D.B. Thompson, and D.R. Liu, *Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.* Nat Biotechnol, 2014. **32**(6): p. 577-582.

211. Tsai, S.Q., et al., *Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing.* Nat Biotechnol, 2014. **32**(6): p. 569-76.
212. Gasiunas, G. and J.K. Young, *A catalogue of biochemically diverse CRISPR-Cas9 orthologs.* 2020. **11**(1): p. 5512.
213. Tian, P., et al., *Fundamental CRISPR-Cas9 tools and current applications in microbial systems.* Synthetic and Systems Biotechnology, 2017. **2**(3): p. 219-225.
214. Morgan, S.L., et al., *Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping.* Nature Communications, 2017. **8**(1): p. 15993.
215. Sah, S.K., K.R. Reddy, and J. Li, *Abscisic Acid and Abiotic Stress Tolerance in Crop Plants.* Frontiers in Plant Science, 2016. **7**.
216. Hagège, H., et al., *Quantitative analysis of chromosome conformation capture assays (3C-qPCR).* Nature Protocols, 2007. **2**(7): p. 1722-1733.
217. Stanulla, M., et al., *DNA cleavage within the MLL breakpoint cluster region is a specific event which occurs as part of higher-order chromatin fragmentation during the initial stages of apoptosis.* Mol Cell Biol, 1997. **17**(7): p. 4070-9.
218. Betti, C.J., et al., *Apoptotic triggers initiate translocations within the MLL gene involving the nonhomologous end joining repair system.* Cancer research, 2001. **61**(11): p. 4550-4555.
219. Stanulla, M., et al., *Mechanisms of MLL gene rearrangement: site-specific DNA cleavage within the breakpoint cluster region is independent of chromosomal context.* Human molecular genetics, 2001. **10**(22): p. 2481-2491.
220. Mitelman, F., B. Johansson, and F. Mertens, *The impact of translocations and gene fusions on cancer causation.* Nature Reviews Cancer, 2007. **7**(4): p. 233-245.

7. Appendix

7.1. Vector maps

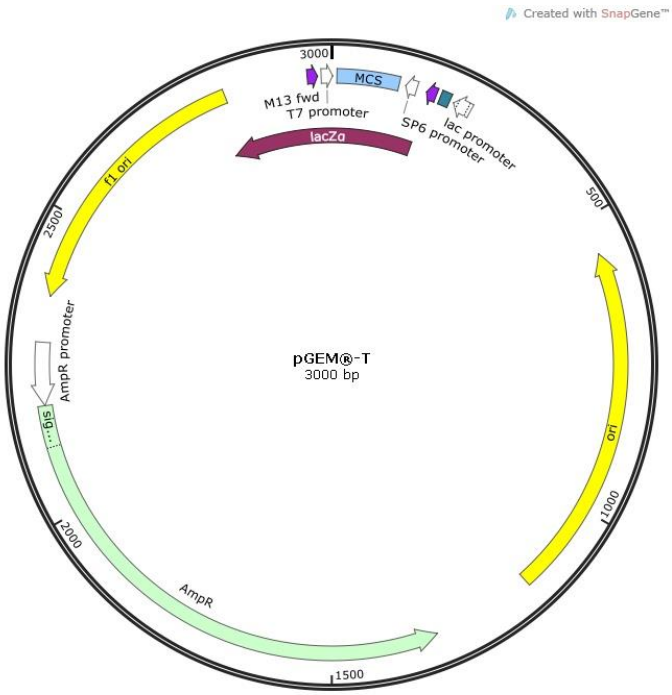

Figure 24. pGEM-T

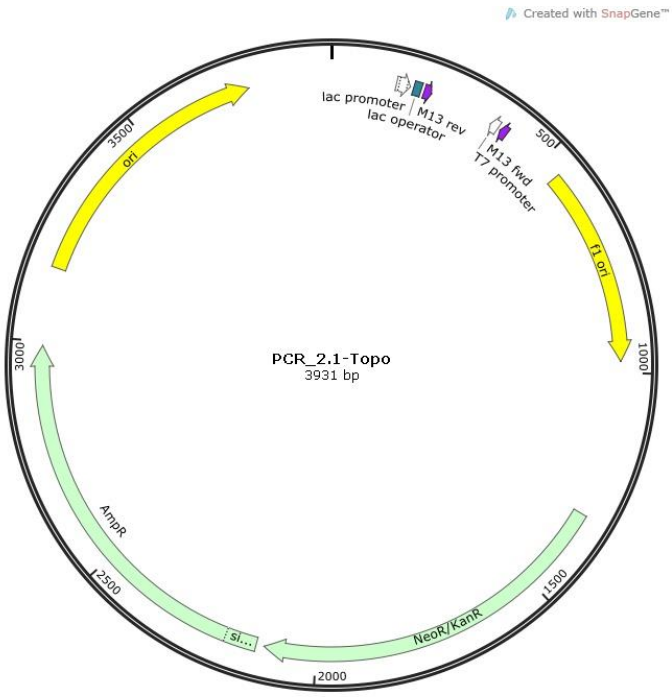

Figure 25. pCR_2.1-TOPO

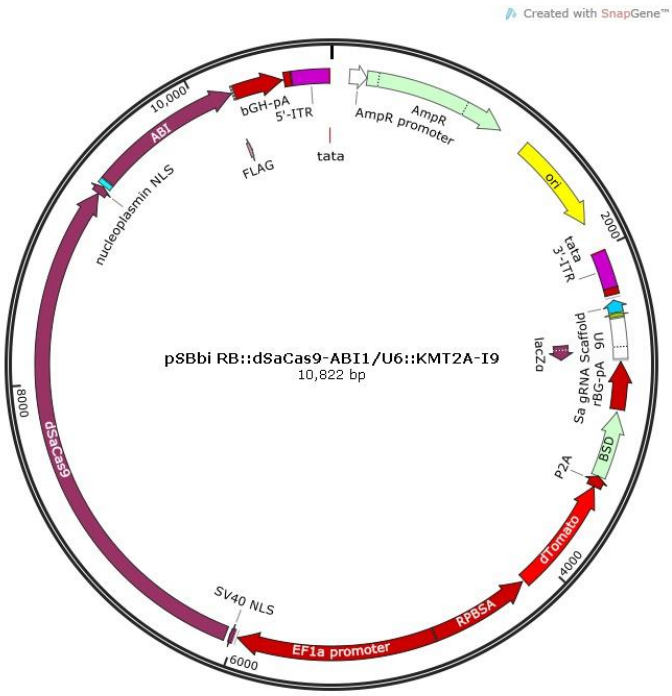

Figure 26. pSBbi RB::dSaCas9-ABI1/U6::KMT2A-I9

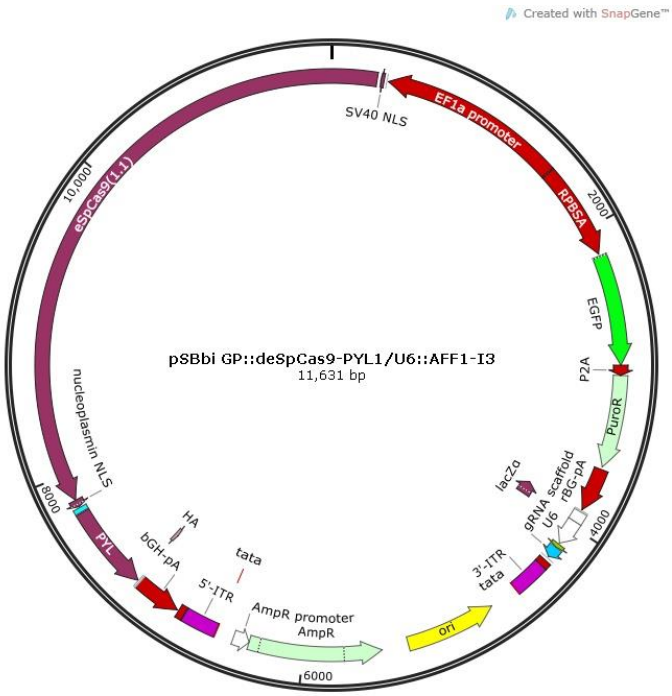

Figure 27. pSBbi GP::deSpCas9-PYL1/U6::AFF1-I3

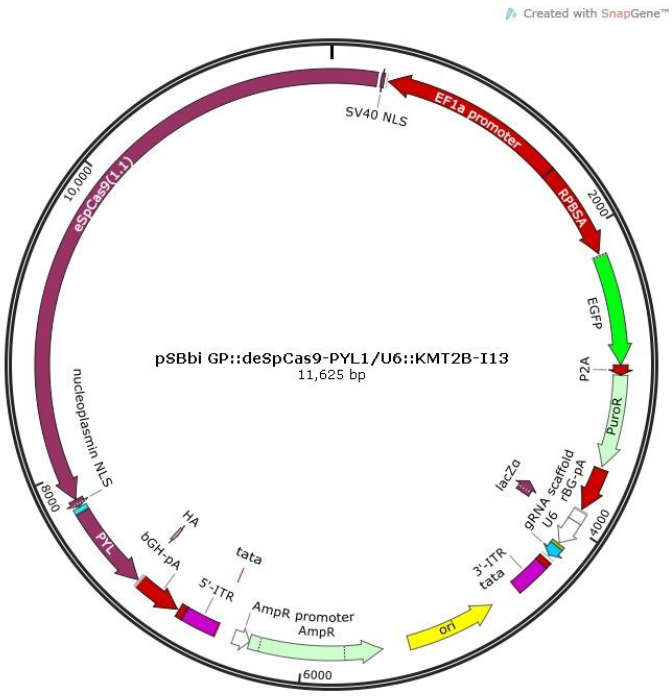

Figure 28. pSBbi GP::deSpCas9-PYL1/U6::KMT2B-I13

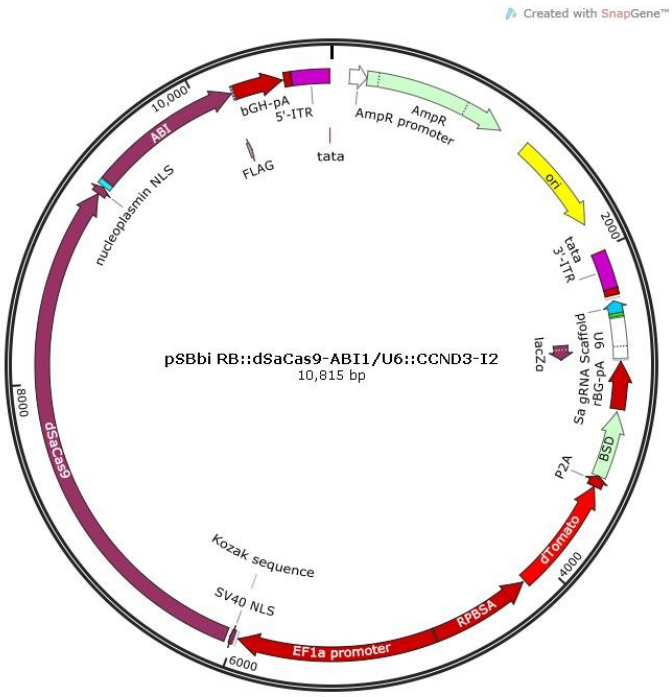

Figure 29. pSBbi RB::dSaCas9-ABI1/U6::CCND3-I2

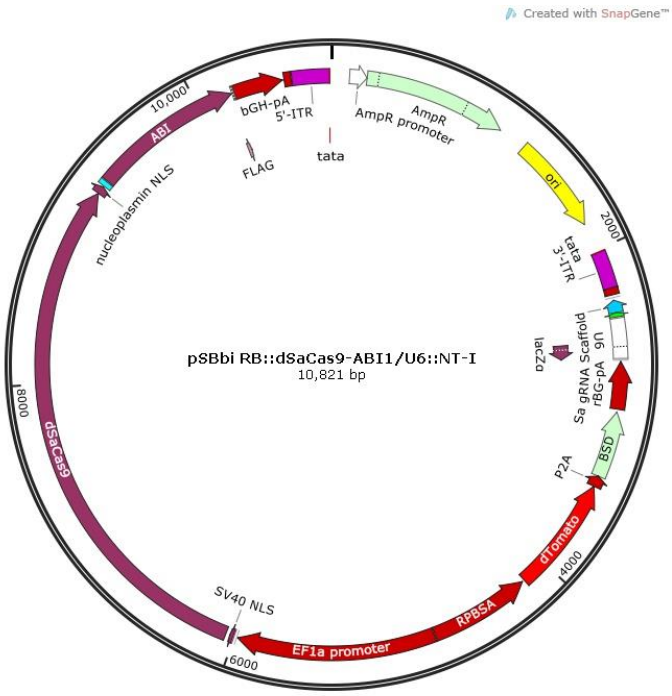

Figure 30. pSBbi RB::dSaCas9-ABI1/U6::NT-I

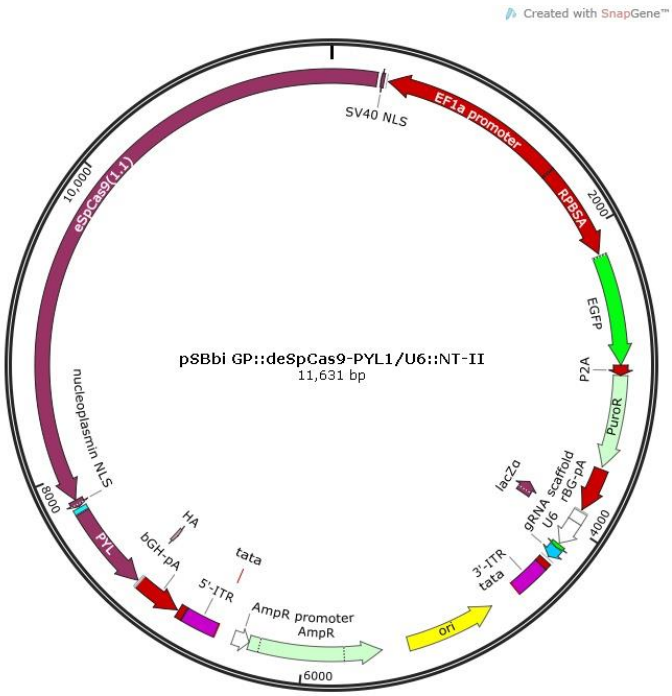

Figure 31. pSBbi GP::deSpCas9-PYL1/U6::NT-II

8. Acknowledgements

I would like to thank all people inside and outside of the laboratory who supported and motivated me over the last years. Some times were tough and the mental support and the good conversations helped me to grow personally and scientifically.

First, I would like to thank Prof. Dr. Rolf Marschalek for the opportunity to work in his lab and both, Prof. Dr. Marschalek and Dr. Eric Kowarz, for the interesting project and the support that turned into a very nice result and insight into the understanding of chromosomal translocations.
Furthermore, I want to thank Prof. Dr. Robert Fürst for the second supervision of the PhD thesis.

I want to thank Silvia Bracharz for the assistance in the cell culture and especially Jenny Merkens for all the help with the cloning and sequencing, the personal support and all the funny times during the work in the lab, you really made things better.

I especially want to thank Dr. Vanessa Luciano, Dr. Anna Lena Siemund, Dr. Marius Külp and Dr. Claus Meyer, the best working environment is, when colleagues become friends and I will definitely miss the conversations and the things outside of work.

I also want to thank all other colleagues and interns, you were not less important, Robin, Patrizia, Alissa, Tamara, Thomas, Lisa, Freia, Arpita, Valerie, Bruno, Alex and the members of the Fürst group.

I would like to thank Elsa, Silke, Mitja, Guido, Joscha and my whole family from the bottom of my heart, for supporting me and for enduring my stress and sometimes bad mood. Your assistance has been a big reason that made this possible.

9. Declaration

I herewith declare that I have produced my doctoral dissertation on the topic of

"The molecular principles of chromosomal translocations"

independently and using only the tools indicated therein. In particular, all references borrowed from external sources are clearly acknowledged and identified.

I confirm that I have respected the principles of good scientific practice and have not made use of the services of any commercial agency in respect of my doctorate.

Frankfurt am Main, _____

_____