

Inaugural dissertation
for
obtaining the doctoral degree
of the
Combined Faculty of Mathematics, Engineering and Natural Sciences
of the
Ruprecht-Karls-University
Heidelberg

Presented by

Vivien-Sandra Ionasz, M.Sc.
born in: Oberwart, Austria

Oral examination: October 27th, 2023

Elucidating the role of transcription in recurrent DNA break formation

Referees:

Prof. Dr. Ingrid Lohmann
Prof. Dr. Michaela Frye

Pentru Maia

Summary

Mild replication stress in neural stem and progenitor cells leads to the formation of recurrent DNA break clusters (RDC). Genes containing these RDCs (RDC-genes) play important roles in brain functions such as synaptogenesis and cell-cell-adhesion. Most RDC-genes do not harbor break clusters in cell types in which the RDC-genes are not being actively transcribed, but the link between transcription activity and the formation of DNA breaks in RDC-genes has not been investigated before.

To examine whether transcription is the licensing factor for RDC formation, the promoters of two robust RDC-genes, Catenin Alpha 2 (*Ctnna2*) and Neurexin 1 (*Nrxn1*), have been independently successfully deleted in multiple ESC-NPC cell lines *in vitro*. In these cell lines, the transcription of *Ctnna2* and *Nrxn1* genes was abolished. Moreover, the number of DNA double-strand breaks was reduced in the gene-of-interest while the amount of DNA breaks in other RDC-genes remained unaffected. Additionally, also the replication timing did not change significantly when comparing the cells. Using the same unbiased genome-wide nucleotide resolution assay to detect the recurrent DNA break clusters, I was able to detect the movement of stalled/collapsed replication forks across increasing levels of replication stress and identify the observed breaks as single-ended double-strand breaks. Taking these findings together with transcription and replication directionality, it is evident that there is a bias towards head-on collision (40% more DNA double-strand breaks) versus co-directional collision of the replication fork and transcription machinery. These data fit to the hypothesized transcription/replication conflict, which is believed to play an essential role in the formation of the recurrent DNA break clusters. All things considered, my project illuminates the relationship between transcription, replication, and DNA damage in the form of double strand breaks in replication stress conditions, which can be translated to influences on the developing brain and its genome.

Zusammenfassung

Leichter Replikationsstress in neuronalen Stamm- und Vorläuferzellen führt zur Bildung wiederkehrender DNA-Bruch-Cluster (RDC). Gene, die diese RDCs enthalten (RDC-Gene), spielen eine wichtige Rolle bei Gehirnfunktionen wie Synaptogenese und Zell-Zell-Adhäsion. Die meisten RDC-Gene beherbergen keine Bruchcluster in Zelltypen, in denen die RDC-Gene nicht aktiv transkribiert werden, aber der Zusammenhang zwischen Transkriptionsaktivität und der Bildung von DNA-Brüchen in RDC-Genen wurde bisher nicht untersucht.

Um zu untersuchen, ob die Transkription der Lizenzfaktor für die RDC-Bildung ist, wurden die Promotoren zweier robuster RDC-Gene, Catenin Alpha 2 (*Ctnna2*) und Neurexin 1 (*Nrxn1*), unabhängig voneinander erfolgreich in mehreren ESC-NPC-Zelllinien *in vitro* deletiert. In diesen Zelllinien wurden die Gene *Ctnna2* und *Nrxn1* nicht mehr transkribiert. Darüber hinaus wurde die Anzahl der DNA-Doppelstrangbrüche im deaktivierten Gen reduziert, während die Anzahl der DNA-Brüche in anderen RDC-Genen unverändert blieb. Darüber hinaus änderte sich auch der Replikationszeitpunkt beim Vergleich der Zellen nicht wesentlich. Mithilfe der gleichen unverzerrten, genomweiten Nukleotidauflösungsmethode zur Detektion wiederkehrender DNA-Bruchcluster konnte ich die Bewegung blockierter/kollabierender Replikationsgabeln über zunehmende Replikationsstressniveaus hinweg erkennen und die beobachteten Brüche als einseitige Doppelstrangbrüche identifizieren. Betrachtet man diese Ergebnisse zusammen mit der Transkriptions- und Replikationsrichtung, so ist es offensichtlich, dass eine Tendenz zur Frontalkollision (40 % mehr DNA-Doppelstrangbrüche) gegenüber der gleichgerichteten Kollision der Replikationsgabel und der Transkriptionsmaschinerie besteht. Diese Daten passen zum angenommenen Transkriptions-/Replikationskonflikt, der vermutlich eine wesentliche Rolle bei der Bildung der wiederkehrenden DNA-Bruchcluster spielt. Alles in allem beleuchtet mein Projekt den Zusammenhang zwischen Transkription, Replikation und DNA-Schäden in Form von Doppelstrangbrüchen unter Replikationsstressbedingungen, die sich auf Einflüsse auf das sich entwickelnde Gehirn und sein Genom übertragen lassen.

Table of Contents

Summary	I
Zusammenfassung	III
Table of Contents	V
1 Introduction	1
1.1 The causes of replication stress	2
1.1.1 Oncogene activation and cell cycle	2
1.1.2 Nucleotide pool imbalance	3
1.1.3 Interstrand cross-links	4
1.1.4 DNA secondary structures and sequences at risk of fork stalling	4
1.1.5 Transcription-associated replication stress	5
1.2 Transcription/replication collision.....	7
1.2.1 Transcription.....	7
1.2.2 Replication.....	7
1.3 DNA damage	8
1.3.1 DNA double-strand breaks	8
1.3.2 Common Fragile Sites.....	9
1.3.3 Recurrent DNA break clusters.....	9
1.3.4 Aphidicolin	10
1.4 DNA repair mechanisms.....	11
1.4.1 Non-homologous end joining.....	11
1.4.2 Alternative end joining	12
1.4.3 Homologous recombination.....	13
1.4.4 Tumor Protein 53.....	14
1.5 Genome integrity in health and disease	15
1.5.1 Health	15
1.5.2 Cancer.....	15
1.5.3 Neurological disorders.....	16
2 Aims	17
3 Rationale	19
4 Results	23
4.1 RDCs are transcription-dependent.	23
4.1.1 <i>Catenin Alpha 2 (Ctnna2)</i>	25
4.1.2 <i>Neurexin 1 (Nrxn1)</i>	41
4.2 RDCs are single-ended DSBs.	56
4.3 Head-on collision leads to more DNA breaks.....	60

5	Discussion	63
5.1	Transcription activity is required for <i>Ctnna2</i> and <i>Nrxn1</i> RDC formation.	63
5.2	LAM-HTGTS and RDC detection.	64
5.3	Replication timing changes.	65
5.4	Head-on collision in this study.	66
5.5	CRISPR off-target during genome editing.	67
5.6	Differentiation into neural progenitor cells.	68
5.7	Deficiency of p53 and XRCC4.	69
5.8	Cell cycle dependence of replication.	69
5.9	Mechanisms of Common Fragile Site Formation versus Recurrent DNA Break Cluster Formation.	70
5.10	The effect of transcription on the genome.	71
5.11	Epigenetic changes and how they translate to replication timing.	74
5.12	Future perspectives.	76
6	Conclusion	79
7	Materials and Methods	81
7.1	Materials.	81
7.1.1	Cell culture	81
7.1.2	Molecular biology lab work	88
7.2	Methods	97
7.2.1	Cell culture	97
7.2.2	Molecular biology lab work	103
7.2.3	Analyses of the sequencing data	115
8	References	117
9	Appendix	131
9.1	Abbreviations	131
9.2	List of Figures	137
9.3	List of Tables	145
9.4	Supplementary Figures	146
9.5	Supplementary Tables	151
10	Acknowledgements	155

1 Introduction

Error-free replication of the genomic DNA is necessary to ensure genomic integrity in eukaryotic cells. It is a highly regulated process during the cell cycle to duplicate the chromosomes in preparation for the cell's division¹. Nonetheless, there are some issues that could arise, such as physical barriers for the replication fork. Replication stress is known as the slowing down and consequently the stalling of the progressing replication fork in response to unusual DNA structures, DNA damage, encounter with the transcription machinery, or nucleotide pool imbalance. The majority of stalled replication forks are able to resume their process shortly after the blockage, nevertheless, some stalled replication forks also collapse and dissociate from their DNA template². The collapsed forks are usually handled by topoisomerases as well as endonuclease complexes, which then eventually leads to the generation of double-strand breaks³.

In the following pages, I summarized the current understanding in the causes of replication stress, transcription/replication collision, consequences of replication stress on the genome, how the genome is repaired, as well as biological consequences of replication stress in diseases.

1.1 The causes of replication stress

1.1.1 Oncogene activation and cell cycle

The overexpression of different oncogenes leads to re-replication of the genome⁴. Proto-oncogenes are proteins that regulate the cell cycle in terms of cell growth, differentiation, and apoptosis. Point mutations in these genes can lead to a constitutive expression of the gene, also referred to as oncogene activation. Cells affected by this activation have a growth advantage and increased proliferatory potential. However, this mechanism renders the cells prone to genome instability⁵. One consequence is observable during replication origin firing by the deregulation of cyclin-dependent kinase (CDK) activity as well as the overexpression of origin licensing factors causing an excessive firing of replication origins. This can lead to a depletion of available dNTPs and histones generating under-replicated DNA as well as ssDNA. Additionally, these changes are also affecting the transcriptional activity in the cells, increasing the risk of transcription/replication collisions which also further accelerates the replication stress.

The Cyclins (Cyc), which are essential complexes for the cell cycle activate the CDKs to allow continuation of the cell cycle. Cyclin E is required for the progression through the G1 phase, while the Cyclin D complex initiates the S phase⁶. Overexpression of Cyc E causes alterations in DNA replication by increased origin firing and increased DNA synthesis. The consequence of that is the depletion of the nucleotide pool, which leads to replication stress and a deregulation of the cell cycle; therefore, these two cyclins are considered proto-oncogenes⁷.

Similarly, an overexpression of the MYC protein is perceptible in different tumor types including MYCN in neuroblastomas and c-MYC in lymphomas⁸. MYC activates genes that promote DNA synthesis and initiation of replication. Furthermore, it also leads to the deregulation of the normal cell cycle through its function as a transcription factor while it also recruiting chromatin modifying co-factors⁹. MYC also inhibits the ATR checkpoint, allowing the cells to continue with their cell cycle even though the cell cycle should be arrested⁸. A study in transformed cells with a Rb-E2F activation showed that MYC overexpression leads to an increase of endogenous deoxynucleotides, which has the same effect as an

exogenous supply of nucleotides, and is able to rescue DNA replication stress caused by nucleotide pool imbalance¹⁰.

Oncogenic RAS activates the G1 checkpoint of the cell cycle, allowing the cells to enter the S phase without a previous stimulation using growth factors. The cells undergo replication without the cellular components required for accurate DNA synthesis¹¹. Another source of replication stress are the genome-wide high expression rates due to RAS activation and an increase in the production of reactive oxygen species (ROS)^{12,13}. Oncogenic RAS also induce replication stress by enhancing Cdc6 levels. This can result in a rise of origin firing, which consequently causes a higher need of nucleotides for the DNA synthesis. This leads to a stalled replication fork and, if not resolved, to DNA damage¹⁴.

1.1.2 Nucleotide pool imbalance

It has been proposed that improper control of replication initiation leads to the excessive origin firing, which can deplete nucleotide pools and slows down replication fork speed^{15,16}. The consequences of this imbalance are also the decoupling of cell growth from cell proliferation. The imbalance can only be sensed during S phase by replication stress signaling; therefore, cells will undergo disproportionate cell growth due to continued biomass production while the cell division remains inhibited¹⁷. Normally, dNTP pools are tightly regulated, guaranteeing cell viability and prevents elevated mutagenesis rates¹⁸. Usually, replication initiation is tightly controlled through the cell cycle checkpoint before entering S phase as well as a strictly guarded licensing and firing process^{19,20}. The dNTP pools are upregulated during the transition from G1 to the S phase and the level of nucleotides remains high until the end of S phase²¹. At the same time, origin firing is inhibited, replication forks are stabilized, and DNA repair is modulated, to ensure faithful DNA synthesis²².

Studies have demonstrated that both in yeast²³, and in mammalian cells¹⁰, imbalanced nucleotide levels result in genomic instability. Therefore, targeting of the nucleotide metabolism became a common therapeutic strategy for cancer, and other immune disorders²⁴. Imbalances also play a role in antibody gene

diversification; critical processes like V(D)J recombination are affected. All these processes are predominantly restricted to the G1 phase of the cell cycle, which is an indicator that the diversification of antibody genes has evolved to be functional with low concentrations of dNTPs. Consequently, the accumulation of purine nucleotides during antibody gene diversification can cause immunodeficiencies by inhibiting the diversification itself²⁵.

1.1.3 Interstrand cross-links

Covalently linked strands of double-stranded DNA are also known as interstrand cross-links (ICLs), which lead to a block of DNA replication caused by the converging of the two replication forks at the crosslink, as well as transcription²⁶. ICLs can arise due to different conditions, such as treatment with chemotherapeutic drugs like cisplatin, exposure to environmental toxins, UV radiation, ionizing radiation, or endogenous metabolism by the generation of aldehydes. The blockage of replication and transcription leads to mutations or cell death²⁷.

1.1.4 DNA secondary structures and sequences at risk of fork stalling

Certain DNA sequences can form DNA secondary structures such as G-quadruplexes, Z-form, triplex DNA, and hairpins. They all are sufficient to form a physical barrier which can slow down the progressive replication fork or even terminate replication²⁸.

G-quadruplexes are four-stranded helical structures that assemble when guanine-rich DNA sequences stack on top of each other and form stable hydrogen bonds between their base pairs. G-quadruplexes play significant roles during replication, transcription, and translation. This structure is also known for stabilizing telomeres to protect the ends of the chromosomes from degradation²⁹.

Z-form DNA is a left-handed version of double-stranded DNA that forms a zig-zag pattern, but only forms under special conditions like high salt concentrations or alternating purine-pyrimidine sequences. This unique conformation has been linked to human diseases. There are proteins that can specifically bind to this

conformation, but Z-DNA can also influence transcription, induce genome instability, or even elicit immune response³⁰.

Triplex DNA is a structure that involves three strands of DNA instead of two. A third strand binds to the major groove of a double-stranded DNA molecule by Hoogsteen base pairing. Naturally they occur as an intramolecular H-DNA structure in areas with repeated sequences or inverted repeats and are essential for DNA metabolism as well as gene function. Additionally, intermolecular triplexes can also form, together with a triplex-forming oligonucleotide (TFO). Intermolecular triplex DNA has the potential to be used for cancer or antiviral therapy to alter gene expression, to stimulate DNA repair or to obstruct replication³¹.

Hairpin loops can form when sequences with inverted repeats are present. These structures can be folded either from single-stranded DNA during transcription, replication, as well as DNA repair but also from double-stranded DNA in the form of a cruciform. Hairpins in the DNA can be recognized by proteins, alter gene expression levels, and are involved in RNA processing and stability³².

These structures have essential biological roles in genome organization, gene expression and DNA repair but pose the risk to hinder the cell's replication prior to cell division leading to DNA damage or mutations²⁸.

1.1.5 Transcription-associated replication stress

Transcription-associated replication stress can occur when DNA replication encounters obstacles caused by ongoing transcription. In this case, the RNA polymerase collides with the replication fork; consequently, stalling and collapse of the replication machinery is observable^{33,34}. The speed and direction of DNA replication as well as the activity in selected cell cycle phases are tightly controlled. However, transformed cells can alter their transcription activity, leading to more potential for interference of transcription and replication³⁵.

1.1.5.1 R-loops

R-loops are DNA-RNA hybrids with a length of 100 to 2,000bp which usually form during transcription when newly transcribed RNA hybridizes with the template strand displacing a single-stranded DNA loop³⁶⁻³⁸. The R-loop level is increased upon head-on replication-transcription conflicts³⁹. These accumulations result in additional cellular stress by the disruption of the fork progression during S phase leading to DNA damage and genome instability⁴⁰. There are various mechanisms that either prevent or promote the formation of R-loops. Helicases, like senataxin, and nucleases, like RNases H, are amongst the enzymes that can remove R-loops once they have formed^{41,42}. Additionally, the degradation of RNA or nucleotide excision repair could also remove R-loops^{43,44}. At the same time, processes such as increased negative supercoiling, nicks in the non-template strand and guanine-rich DNA sequences foster R-loop formation^{40,45}. R-loops serve various biological functions, including the regulation of transcription, DNA replication, and maintenance of genome stability. R-loops can also facilitate transcription initiation and elongation by recruiting transcription factors and other regulatory proteins⁴³.

1.1.5.2 Hypertranscription

Cells can regulate their transcription levels and increase them in response to stimuli like growth factors, hormones, cytokines, or stress signaling. This upregulation of specific genes or global increase is known as a hypertranscription. There are multiple mechanisms involved: hypertranscription can occur through the recruitment of additional transcription factors and coactivators, modifications to chromatin structure, activation of signaling pathways, or the regulation of RNA processing and stability⁴⁶. This has important implications depending on the context and the duration of high transcription levels. They can be important for cell differentiation and development but are also acting as a response to stress through DNA damage or viral infection. Persistent or dysregulated hypertranscription leads to overexpression of oncogenes, activates pro-inflammatory pathways, and disrupts cellular homeostasis. This can lead to the development of diseases like cancer, chronic inflammation, or neurodegenerative disorders⁴⁷.

1.2 Transcription/replication collision

Encounters of replication with transcription can lead to deletions as well as mutations in the genome^{48,49}. This is due to a collision of the RNA polymerase (RNAP) with the replication fork, which often leads to a collapsed replication fork⁵⁰. Studies have shown that the effect is different based on the collision's directionality^{49,51}. Replication has been found to be inhibited by head-on transcription in yeast⁵². Additionally, the gene expression level of the head-on transcription correlates with the severity of the stalled replication fork. Through experimentation with *B. subtilis*, it was discovered that inverting highly transcribed genes, such as ribosomal operons, results in a considerable replication delay compared to their typical co-directional orientation⁴⁹. Furthermore, in all known bacteria, such genes are found to be co-directionally oriented to their replication⁵³. In higher eukaryotes genes that have high levels of transcription are also mainly found in the vicinity of replication origins and pointed away from origins to reduce the possibilities of head-on collisions⁵⁴. However, head-on collision affects genome stability independently of the level of transcription^{48,49}.

1.2.1 Transcription

Modulators that can remove RNA polymerase from the DNA template are essential to resolve collisions. Additionally, they affect the rate, efficiency, and accuracy of transcription. Reduced viability is observed in cells without these modulators after replication fork stalling⁵⁵.

1.2.2 Replication

The conflict can also be resolved from the replication perspective: auxiliary helicases are traveling together with the replication fork and dislodge the transcription machinery in collaboration with the replicative helicase. Upon removal of the transcription block, the replication fork is able to restart³⁴.

1.3 DNA damage

DNA damage describes alterations or breaks in the structure of the DNA that can prevent the replication mechanism from functioning. DNA damage stems from sources that can be environmental such as exposition to UV radiation or chemicals, or of endogenous nature, for example by replication stress or reactive oxygen species. Different types of DNA damage can be observed: they include single-strand breaks, double-strand breaks, base damage, and crosslinking⁵⁶. The cell cycle checkpoints, and DNA repair enzymes are mechanisms for detecting and repairing DNA damage. However, these DNA lesions can accumulate over time and lead to cancer and aging⁵⁷. Human cells are estimated to experience around 70,000 DNA lesions per day. It is estimated that there are 10 to 50 of these lesions classified as DNA double-strand breaks (DSBs). Even though, DSBs occur less frequently than other types of DNA lesions; they are affecting the cells more negatively and are a threat to the genomic integrity of the cell⁵⁸. Especially if these DSBs remain unrepaired, they can result in mutations, chromosomal translocations, or rearrangements, which have been shown to be associated with cancer and other disorders⁵⁹. However, irradiation as well as some DNA damage-inducing chemicals are also being used as therapy to induce cell death in cancer cells. Since these treatment options do not target cancer cells only, they have many side effects, which include hair loss or gastrointestinal problems⁶⁰.

1.3.1 DNA double-strand breaks

DNA double-strand breaks (DSBs) are the most severe form of DNA damage, where both strands of the DNA break apart at the same location⁶¹. DSBs are usually generated by endogenous processes, such as the cellular metabolism or the interaction of DNA replication and transcription; while exogenous sources of DSBs can be ionizing radiation or chemicals⁶². Two main repair pathways are used for the repair of DSBs: non-homologous end-joining (NHEJ) and homologous recombination (HR). NHEJ ligates two broken DNA ends with each other, while HR is more complex and uses an intact homologous copy of the damaged DNA as a template for the repair⁶³. If the cell cannot repair the DSB appropriately, this can lead to cell death or chromosomal abnormalities, with the latter possibly leading to cancer development⁶⁴.

1.3.2 Common Fragile Sites

Common fragile sites (CFS) are regions of the genome that frequently break upon replication stress. They have first been identified as gaps in metaphase chromosomes and can be found in all human chromosomes. Breakage or rearrangement at the CFS can lead to genomic instability through rearrangements that are associated with copy number variations, such as deletions in tumor suppressor genes or amplification of oncogenes, which also links them to cancer⁶⁵. Characteristics of CFS include that they are often found in repetitive or AT-rich sequences due to the possibility of secondary structure formation. Additionally, they are often situated inside large genes longer than 1 Mb with large transcription units⁶⁶. Furthermore, CFS might also play a role in human evolution and diversity of the brain structure. Common fragile sites have been found near genes that are involved in brain function. Large CFS genes that have been associated with brain development were found to be mutated in patients with neurological disorders⁶⁷.

1.3.3 Recurrent DNA break clusters

Neural stem and progenitor cells are multipotent cells that develop into neurons, astrocytes, and oligodendrocytes, populating the central nervous system⁶⁸. Through the high proliferation potential of NSPCs during embryonic neurogenesis, they could accumulate genetic alterations which might transmit to successor cells⁶⁹. RDCs are occurring in gene bodies or across multiple genes that are transcribed. They are also present and detectable in human neural progenitor cells, while in astrocyte progenitors or glia progenitors, their number is reduced⁷⁰. About 100 recurrent DNA break clusters (RDCs) have been found in murine NSPCs^{71,72}.

The great majority of RDC-containing genes (RDC-genes) were observed upon the treatment of aphidicolin (APH), a DNA polymerase inhibitor that is used to induce common fragile sites. Over 70% of RDC-genes are very long (> 300 kb) and late-replicating, two features that are also often observed at CFS loci. In addition, most RDC-genes encode proteins that are regulating synaptogenesis and synapse function. The disruption of RDC-genes has been linked to multiple cancers as well as neuropsychiatric disorders⁷¹. These findings all support the notion that recurrent

DNA breaks could occur in NSPCs. However, whether RDCs play a direct role in brain disease development awaits to be investigated.

1.3.4 Aphidicolin

Aphidicolin is a potent inhibitor of the DNA polymerase alpha that prevents the incorporation of dNTPs into the newly replicated DNA strand⁷³. This leads to replication fork stalling and delay of replication timing. The compound is often used in experiments studying DNA replication and repair mechanisms⁷⁴. It was also in clinical trials as a chemotherapeutic agent⁷⁵; however, it displayed low water solubility and fast liver clearance, making it an unfavorable choice⁷⁶.

1.4 DNA repair mechanisms

The DNA damage response is highly regulated on both a genetic and epigenetic level; however, it is also dependent on the 3D conformation of the chromatin⁶². Depending on the cell type, cells have the ability to coordinate various DNA repair pathways through signaling cascades. These cascades either regulate the cell cycle to allow repair or induce apoptosis in the event of irreversible DNA damage⁷⁷. In general, there are many different mechanisms based on which type of DNA damage the cell suffered. It is possible to differentiate between repair mechanisms that target base mismatches, ssDNA breaks, DNA adducts as well as double-strand DNA breaks⁷⁸. Based on the scope of my project, I will only focus on the repair of DSBs via non-homologous end joining and homologous recombination in this chapter.

1.4.1 Non-homologous end joining

Non-homologous end joining (NHEJ) functions in both dividing and non-dividing cells, independent of cell cycle phase⁷⁹. NHEJ is most active during the G1 phase of the cell cycle and repairs DNA damage by ligating two open ends of the DNA strands to each other without synthesizing any additional DNA⁸⁰. NHEJ repairs DNA lesions without depending on a template. Due to the end processing, the resulting DNA often experienced partial loss of genomic information since this process frequently results in deletions or insertions⁸¹. Nonetheless, NHEJ generally suppresses translocation formation⁸².

Mechanistically, the Ku complex, made up of a heterodimer from Ku70 and Ku86, binds to the DNA ends to protect the ends and prevent excessive resection. Subsequently, the catalytic subunit of the DNA-dependent protein kinase (DNA-PKcs) and the endonuclease Artemis are recruited to the open ends. The Ku complex shifts inwards and the DNA-PKcs builds a bridge between the two DNA ends while Artemis is phosphorylated by DNA-PKcs. This allows the further recruitment of Ligase IV/XRCC4 and PNK. The resection capacity of Artemis, together with a gap-filling DNA polymerase from the X family (Pol μ and Pol λ), allows DNA ligase IV to seal the nick and repair the DSB⁸³. The resection, usually

no longer than 20 nucleotides, exposes or generates small regions of microhomology with approximately 0 to 4 bp in length⁸⁴.

DSBs are usually not introduced into the cell deliberately and NHEJ repairs DSBs introduced into cells by ionizing radiation or radiomimetic agents. However, an application of NHEJ in a physiological context is the V(D)J recombination during which immunoglobulin or T-cell receptor gene segments are combined to generate genes for a vast spectrum of different immune responses. Here, the recombination-activating genes, RAG1 and RAG2, introduce DSBs at specific sites, the recombination signal sequences⁸⁵.

1.4.1.1 X-ray Repair Cross-Complementing protein 4

The X-ray repair cross-complementing protein 4 (XRCC4) is a key player in non-homologous end joining and forms a complex together with DNA ligase IV, to ligate the open DNA ends and repair the DSB⁸⁶. In an attempt to knock out XRCC4 in mice, this genotype leads to embryonic lethality, defects in cellular proliferation as well as substantial apoptosis of neuronal cells⁸⁷. Mice that accumulated DSBs, were not able to perform NHEJ and repair their DNA. Therefore, cells were arrested in their cell cycle or went into apoptosis. This phenotype was rescued by inactivating the *p53* gene as well. XRCC4 is crucial for NHEJ, especially for maintaining genomic stability and seems to also play a role in brain development⁸⁸. In a subsequent study, mice with an *XRCC4*- and *p53*-deficient background in their neural cells continued to develop early-onset medulloblastoma due to chromosomal alterations⁸⁹.

1.4.2 Alternative end joining

Alternative end joining (a-EJ) likely is a backup mechanism to remove DSBs from the DNA, which is normally utilized when one or more of the NHEJ key factors are lacking and the mechanism is compromised⁹⁰. It is also known as microhomology and Pol θ -mediated end joining. The resections at the DSBs that are happening during a-EJ are larger than the ones during NHEJ, and it also requires a microhomology of 2 to 20 bp⁸⁴. Factors necessary for a-EJ are PARP1, CtIP, MRN as well as Pol θ . PARP1 is an enzyme that is involved in the sensing of DNA

damage, which then allows the binding of phosphorylated CtIP to the MRN complex. The MRN complex then resects the DNA ends, leaving a small region of microhomology, prior to the gap-filling DNA synthesis by Pol θ ⁹¹. It has been observed that a-EJ is able to repair DNA affected by replication fork-induced breaks⁹². However, in comparison to NHEJ, a-EJ leads to more chromosomal translocations. This might be due to the slower kinetics during the end resections, which makes it more probable for unrelated ends to interact with each other⁹³.

1.4.3 Homologous recombination

DNA repair mediated by homologous recombination (HR) is active during the S and the G2 phase of the cell cycle. The process employs homologous regions of the DNA, such as the sister chromatid or homologous chromosome, as a template. It has the capability to perform DNA repair in a high-fidelity manner, leading to a seamless and scar-free restoration of the damaged DNA⁸¹. The cells utilize either of the two major DSB repair mechanisms NHEJ and HR mainly based on their cell cycle phase. In dividing cells, during the S phase, homologous recombination is most active, while it is virtually absent in G1⁹⁴. In addition to the repair of regular double-ended DSBs, HR is also active at telomeres and is involved in the recovery of collapsed replication forks, which are single-ended DSBs⁹⁵.

Mechanistically speaking, homologous recombination works as follows: After the formation of the DSB, DNA is resected by the DNA2, EXO1 and MRE11 endo- and exonucleases that form multiple kilobase long, single-stranded 3' overhangs⁹⁶. These ssDNA ends are bound by replication protein A (RPA) in eukaryotes to protect the ends from degradation and resolve any secondary structures. Proteins like Rad52 and BRCA2 then stimulate the replacement of RPA by the Rad51 recombinase. With the help of Rad54 and Rhd54, the partner chromosome is scanned for the homologous sequence prior to strand invasion, which develops into a D-loop structure. Then the gap of the invaded strand will be synthesized with the homologous sequence serving as the template. After this step, there are two different strategies for the DSB to be resolved: one of these pathways is synthesis-dependent strand annealing (SDSA), while the other is known as double-strand break repair (DSBR)⁹⁷.

In the case of SDSA, the D-loop is unwound, and the newly synthesized DNA can anneal to single-strand overhang. Subsequently, the second strand is also synthesized to fill the gap and the ends are ligated together. This pathway only leads to the formation of non-crossover products⁹⁵.

For the DSBR pathway, the second DSB end is captured at the homologous chromosome as well and forms an intermediate, which then leads to a double Holliday junction. Following this, the DNA gaps between the two open ends are synthesized and ligated. The resolution of the homologous recombination products is then performed by nicking endonucleases, which cut one of the DNA strands. Consequently, as the Holliday junctions are resolved, chromosomal crossover frequently occurs, influenced by the orientation in which the endonuclease makes its cuts. In this case, the DNA sequence around the DSB area of the two homologous chromosomes is exchanged. Non-crossover products are also able to form through the DSBR pathway, however, they are less common⁹⁵.

1.4.4 Tumor Protein 53

Tumor protein 53 (TP53) is a tumor suppressor protein encoded by the p53 gene and inactivated in up to 50% of human cancers. It is also known as the guardian of the genome as it is preventing cells from proliferating in the case of DNA damage⁹⁸. TP53 mainly functions as a transcription factor; hereby, it can regulate hundreds of target genes which are involved in apoptosis, cell cycle arrest, senescence, differentiation, and DNA repair, making it one of the most important factors in a complex network of interactions and signaling cascades⁹⁹.

Studies have found that TP53 leads to the downregulation of genes involved in telomere maintenance, DNA repair, as well as centromere structure, which firstly seems counterintuitive but may help to prevent tumor formation by arresting cells and activating senescence or apoptosis¹⁰⁰. In the case of DNA DSBs, the cell cycle arrests are crucial, since they give the cell more time to perform NHEJ or HR, allowing the cell to repair the DNA prior to replication, which is important for the genome's integrity¹⁰¹.

1.5 Genome integrity in health and disease

Since cells and their function are vulnerable to DNA damage, genome integrity is essential for cell homeostasis. Therefore, the DNA repair mechanisms presented in the previous chapter are important. Genomic alterations such as mutations, deletions or other chromosome rearrangements are accumulating in cells, which subsequently predisposes them to diseases such as cancer. Not only is genomic instability observed in the majority of human cancers, it can also help the cancer cells to adapt and build resistance to therapy¹⁰².

1.5.1 Health

Embryonic stem cells (ESCs), which have the potential to self-renew and are pluripotent, express more proteins related to homologous recombination and have a more potent cell cycle control in comparison to differentiated cells. The elevated cell cycle control is characterized by high proliferation rates and a prolonged S phase. This is essential since early mutations would translate and reproduce in all the daughter cells. The G1 phase is shortened in ESCs, which helps to keep them pluripotent by reducing the possibility of stimulating differentiation-inducing signals. Since HR has higher fidelity compared to other DNA repair mechanisms, both the elevated levels of HR-related proteins, as well as prolonged S phase are major factors that help ESCs repair DNA lesions and recover replication stress in an error-free manner¹⁰³.

1.5.2 Cancer

Based on the genomic information that can be extracted from cancer cells, it is possible to learn about the DNA damage that the cell underwent, which can originate from either endogenous physiological processes or from external sources as discussed earlier. For example, cancer cells with aberrant homologous recombination pathways, as would be the case with deficiencies in BRCA1 and BRCA2, would present with SNVs, indels, structural variations as well as CNV. This knowledge could elucidate why the tumor formed and could then be used as a biomarker to improve treatment options, especially since cells already suffering from genomic instability are more prone to develop therapy-resistant clones¹⁰⁴.

Presently, cancer cells resulting from mutations in their DNA repair pathways have become susceptible to targeted treatment through synthetic lethality. For the concept of synthetic lethality, it is important to exploit the cancer's mutations and find a second hit, with which the cell is not able to survive since it will not be able to compensate for its absence. In the case of the previously mentioned BRCA1/2-deficiencies, one would commonly use PARP inhibitors. By inhibiting PARP, single-stranded DNA breaks would accumulate, which would contribute to DSBs at replication forks. Normally, these lesions would then be repaired via HR. However, in BRCA1/2-deficient cells, the cells would go into apoptosis due to the accumulation of DNA damage. This type of therapy can, in comparison to widely used radiation therapy and genotoxic drugs, specifically target the cancer cells that are deficient in homologous recombination while sparing the healthy cells. All in all, this strategy is potent and seems to also tackle the problem of cytotoxicity; nonetheless, the development of resistances remains an issue¹⁰⁵.

1.5.3 Neurological disorders

Due to the magnitude of control that the brain has over the rest of the body and its function, DNA damage is expected to have a fatal effect on the nervous system. A multitude of neuropathologies like brain tumors, microcephaly, neuroinflammation as well as neurodegeneration are associated with defects in the DNA damage response. During neurodevelopment, DNA lesions are mainly replication-associated due to the rapid expansion of the nervous system, while in the adult brain high oxygen consumption leads to an increased level of ROS and free radicals that result in DNA breakage. Additionally, neurons can also have increased levels of transcription, which can increase the levels of DNA damage that the cells are experiencing¹⁰⁶. Gaining a deeper understanding of the causes and mechanisms responsible for DNA damage in the neuronal lineage is crucial to fully comprehend the significance of DNA damage in the context of neurological diseases.

2 Aims

This project included different aims regarding the the role of transcription in recurrent DNA breaks and identification how these DNA break clusters are formed.

1. Elucidating the role of transcription in recurrent DNA break cluster formation.
 - 1.1. Examining whether transcription is the licensing factor for RDC formation.
 - 1.2. Investigating the replication timing upon deletion of regulatory elements.

2. Identifying recurrent DNA break clusters as single-ended DSBs.
 - 2.1. Exploring fork stalling in response to aphidicolin treatment.
 - 2.2. Interpreting the link between break formation and collision directionality.

3 Rationale

Since RDCs were observed upon aphidicolin treatment *in vitro*^{71,72,107}, I hypothesize that RDCs were generated through transcription and replication collision. It has been shown that when the DNA replication fork and RNA transcription machinery collide head-on, the stalling and dissociating of the replisome may lead to DNA double strand breaks by enzymatic cleavage due to run-off¹⁰⁸. Based on the fatal consequences of replication stress, it has been proposed as an additional hallmark of cancer⁹. The observable replication stress-induced DSBs could be a source of oncogenic rearrangement^{109,110}.

Although it has been already hypothesized that transcription plays a crucial part in the formation of recurrent DNA break clusters by Wei *et al.*⁷¹ in 2018, the role of transcription in this replication stress-mediated setting still remained elusive. The importance here is to clarify the mechanism of this phenomenon, since it could have a critical impact for the understanding of genome instability in cancer and neurological diseases.

Transcription

Based on the previously mentioned characteristics, the core hypothesis of this project is the requirement of active transcription for the formation of the RDCs. It has been shown that in B lymphocytes, where RDC loci are not transcribed, RDC-genes do not contain DNA break clusters¹⁰⁷. Additionally, DNA break levels are increased in embryonic stem cell (ESC)-derived neural progenitor cells (ESC-NPC), in which RDC-genes are actively transcribed, in contrast to the ES cells with no active transcription. However, these previous findings have to be interpreted more cautiously since they compare different cell types using different DNA repair pathways, as B lymphocytes utilize NHEJ for the V(D)J recombination¹¹¹ while neural progenitor cells can utilize NHEJ as well as homologous recombination^{112,113}. I hypothesize that RDCs are induced by DNA replication and transcription collision (see Figure 1). The collision of the replisomes and the RNA polymerases would lead to replication fork pausing, and long-term pausing could lead to fork collapse which subsequently would result in DNA breaks^{39,114}.

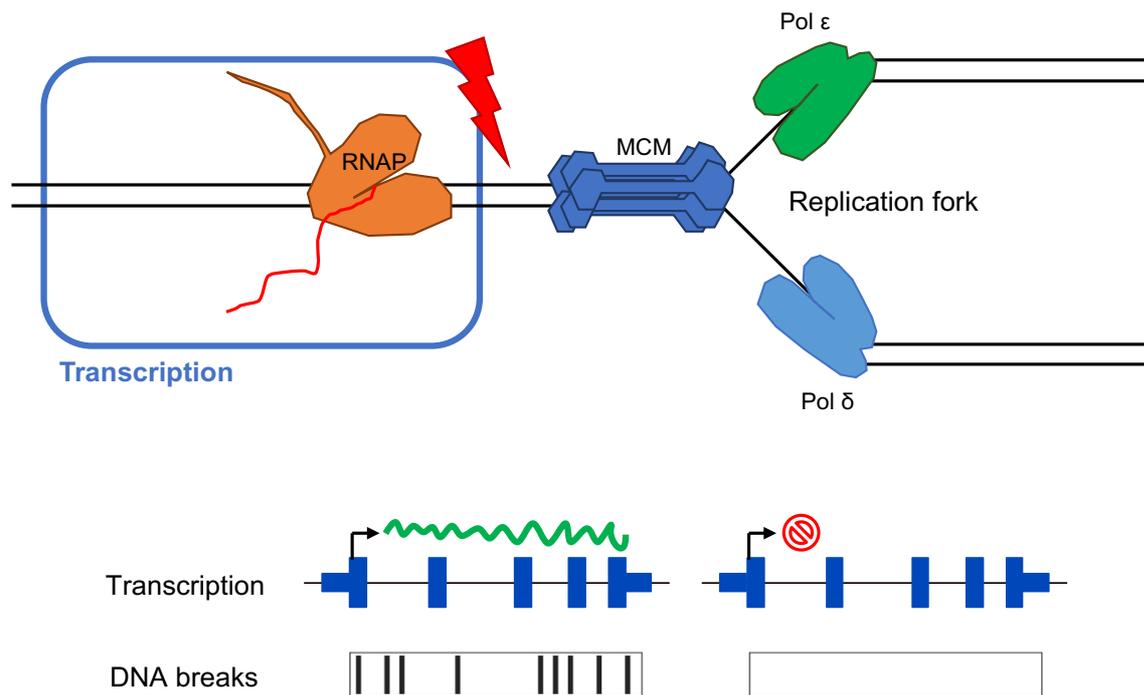


Figure 1: **Transcription-replication collision and the fundamental hypothesis.** Top: Schematic representation of a head-on collision between RNA polymerase II (RNAP - orange) and the replication fork (Pol ϵ - green, Pol δ - light blue and MCM - dark blue). The red lightning bolt represents replication stress and a potentially resulting DNA double-strand break. Bottom: Illustrative gene body shown in dark blue, with active transcription (RNA - green) on the left and abolished transcription on the right. This change in transcription is expected to also translate to the level of DNA breaks. For the transcribed gene in the lower left panel (DNA breaks), black bars are visible, these indicate double-strand breaks at the corresponding location of the gene body above. For the non-transcribed gene, the expectation is to have less DNA double-strand breaks or even eradicate the formation of a cluster altogether.

I proposed deletion of the promoter and proximal enhancer region of a target gene to examine whether transcription is essential for RDCs. I chose to target two RDC-genes: *Catenin Alpha 2 (Ctnna2)*, and *Neurexin 1 (Nrxn1)* independently because they are not essential for neural progenitor cell development or cell survival. Besides, they are among the most robustly detected RDCs in ESC-NPCs and NSPCs. The experiments were carried out in mouse embryonic stem (ES) cell-derived neural progenitor cells to allow easy genome editing.

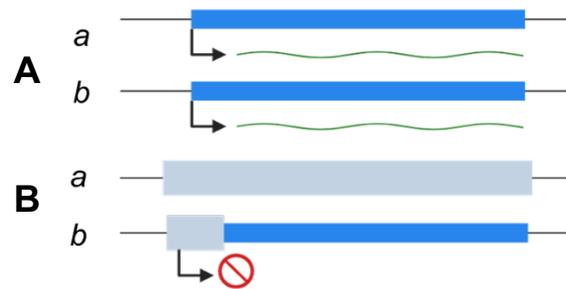


Figure 2: **Experimental approach for Aim 1.** [A] Wild-type expression is depicted by a green transcript along the total gene length. [B] Promoter and enhancer deletion using CRISPR/Cas9 leads to the abolishment of transcription. Gene bodies are depicted in blue, deleted areas are shown in gray, transcripts are shown in green, and the red crossed out circle illustrates transcriptional inactivation upon deletion of the proximal regulatory elements on the b allele.

4 Results

4.1 RDCs are transcription-dependent.

The core research question was to examine whether transcription is the licensing factor for RDC formation. To assess if RDCs are transcription-dependent, it was necessary to abolish transcription. For that, individual embryonic stem cell lines with the deletion of the promoter and proximal enhancer (“p/e clones”) have been generated. This was performed for two separate RDC genes, namely *Ctnna2* and *Nrxn1*, which have been observed to be very robust in the formation of DNA breaks.

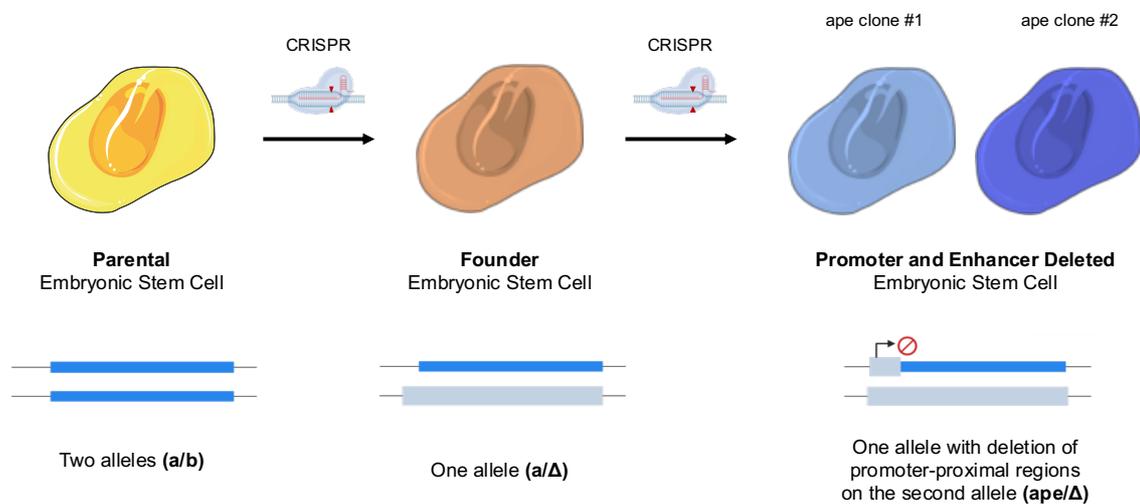


Figure 3: Experimental system for the deactivation of transcription. Embryonic stem cell lines with *p53*- and *Xrcc4*-deficiency have been used as starting material for genome editing. These parental cell lines (yellow) are derived from mice and therefore have two alleles (a/b) of the RDC-genes that have been targeted. In the first step, one allele of the RDC-gene of interest has been deleted completely by CRISPR/Cas9. This intermediary cell line is going to be referred to as the founder cell line (a/Δ). The experimental promoter and enhancer deleted clones (ape/Δ) have been established by another round of CRISPR/Cas9 genome editing by long-range deletion of the promoter-proximal regulatory elements of the gene-of-interest on the remaining second allele.

Since both the targeted RDC-genes (*Ctnna2* and *Nrxn1*) are not being expressed in embryonic stem cells, neural progenitor cells derived from these cell lines have been used for all experiments. For this purpose, a well-established two-week NPC induction protocol has been used. Once the cells generated neural progenitor cells, they were mainly used for two experiments: to isolate nuclei for sequencing of nascent RNA via GRO-seq (Global Run-On sequencing) as well as bait

nucleofection for LAM-HTGTS (Linear Amplification-Mediated High-Throughput Genome-wide Translocation Sequencing), which allows observation of the DNA break dynamics. In the case of *Ctnna2*, also RNA has additionally been isolated for a test run to assess the transcription levels in the promoter and enhancer-deleted clones via RT-qPCR. This experiment was performed as a control to compare total RNA levels with nascent RNA from GRO-seq data.

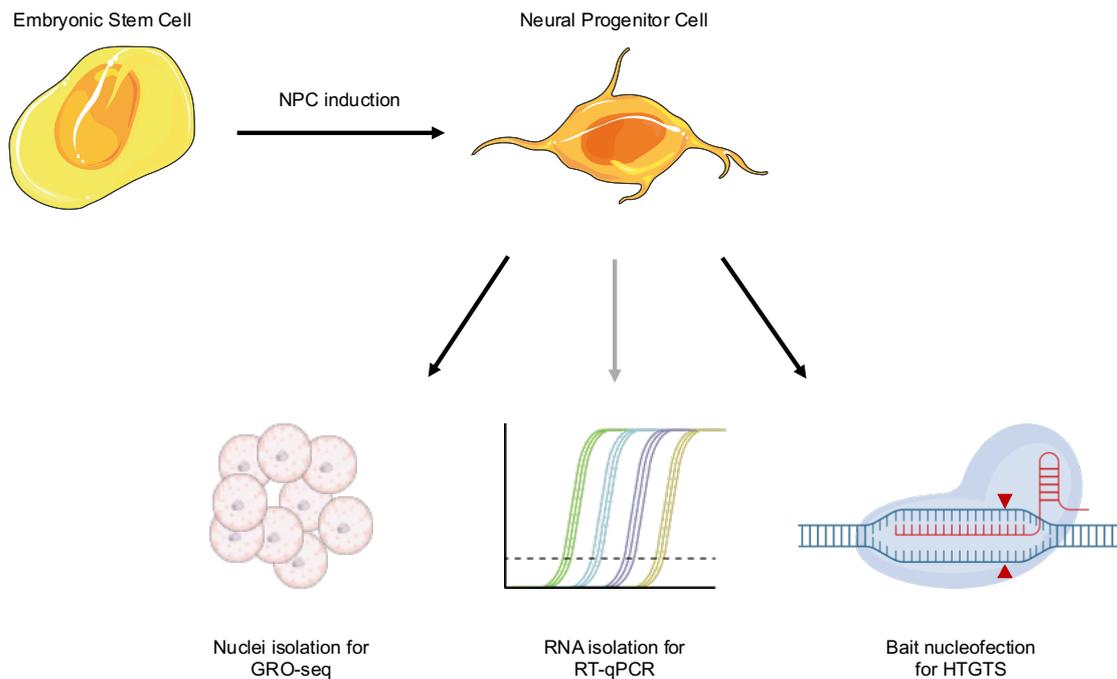


Figure 4: Workflow of the experiments. After induction of the embryonic stem cells to neural progenitor cells, experiments were able to be conducted. For the *Ctnna2* as well as the *Nrxn1* cell lines nuclei have been harvested for GRO-seq (Global Run On-sequencing) and cells were nucleofected with CRISPR/Cas9 to induce the bait break that is necessary for HTGTS (High-Throughput Genome-wide Translocation Sequencing), the method applied to assess break dynamics by sequencing translocations. In a test run, also RNA has been isolated from the *Ctnna2* cell lines to assess if RT-qPCR and GRO-seq are both able to confirm the deactivation of transcription.

4.1.1 Catenin Alpha 2 (*Ctnna2*)

4.1.1.1 Generation of the *Ctnna2* founder cell line.

The whole locus of one of the two alleles of *Ctnna2* has been deleted through genome engineering via CRISPR/Cas9. For this purpose, the gene locus of *Ctnna2* was closely investigated to decide where to design the targeting sgRNAs. The reference genome mm10 was examined in the UCSC Genome Browser¹¹⁵ and the ENCODE Regulation Tracks for histone modifications were used to assess regulatory elements along the gene body. The screenshot of the UCSC Genome Browser view of the *Ctnna2* locus and the sgRNA positions can be found in Supplementary Figure 1.

The parental embryonic stem cell line used for the *Ctnna2* experiments is NXP010. To generate the founder cell line with a one-allelic deletion, NXP010 has been nucleofected with two sgRNAs targeting areas downstream and upstream of *Ctnna2*.

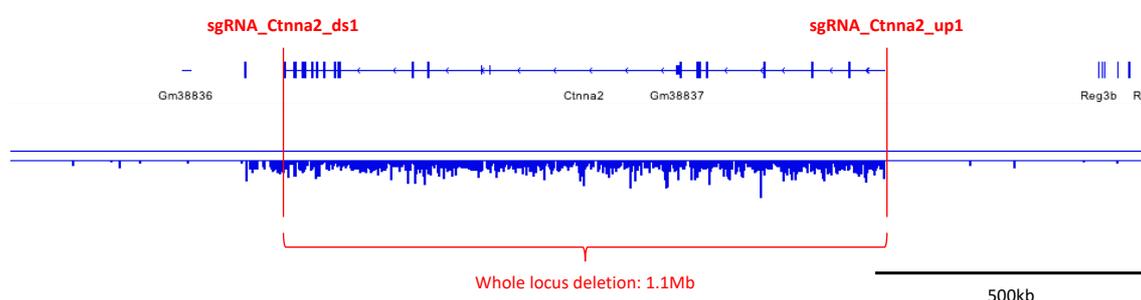


Figure 5: **Whole locus deletion of *Ctnna2*.** Gene locus of *Ctnna2* (blue - gene body at the top) ± 0.5 Mb (chr6:76,379,637-78,481,703) with corresponding transcription signal (blue - at the bottom) from nascent RNA in the parental cell line NXP010. Red vertical lines indicate the location of the two sgRNAs used to target *Ctnna2* and delete the whole locus of one allele. sgRNA_ *Ctnna2*_ds1 is located at chr6:76,879,005-76,879,024 and sgRNA_ *Ctnna2*_up1 at chr6:77,981,693-77,981,712. The whole locus deletion spans across 1.1 Mb. The scale bar at the bottom right indicates 500 kb.

This led to a 1.1 Mb deletion of one of the alleles and was validated after colony picking by a PCR screen of 96 colonies. Positive clones have been additionally analyzed with Sanger sequencing. One clone with the shortest end resections after Cas9 cutting has been selected to be the founder cell line to produce the experimental cell lines, which are the promoter and enhancer deleted clones.

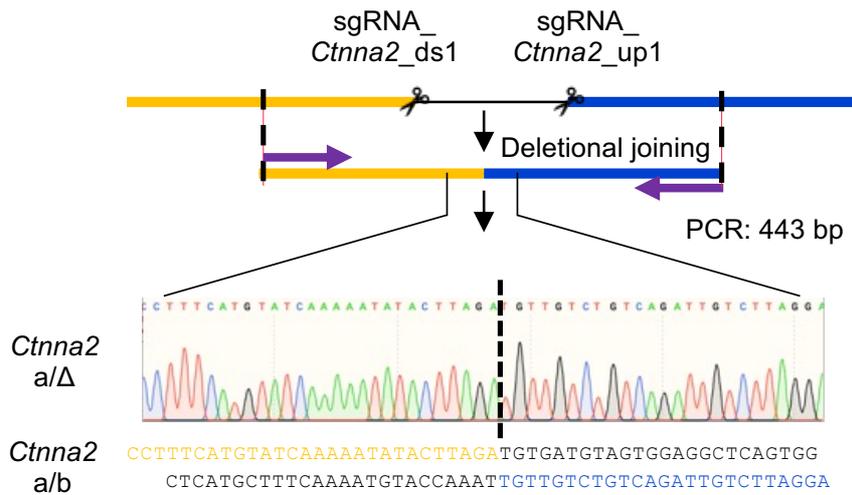


Figure 6: **Workflow for PCR validation and Sanger sequencing of positive clones after CRISPR/Cas9 cutting.** The *Ctnna2* locus is here represented in black. sgRNAs are cutting downstream and upstream of the gene body (scissors). After deletional joining the yellow and blue part of the genome are merged. The deletion can be validated by using the PCR primers indicated as purple arrows, which produce a PCR product with a length of 443 bp. Positive clones have been additionally analyzed by Sanger sequencing. The result of the selected founder cell line (*a/Δ*) is shown at the bottom as sequencing peaks and maintained sequence (yellow and blue) compared to the deleted original sequence (black) from the parental cell line (*a/b*).

4.1.1.2 Generation of the *Ctnna2* promoter and enhancer deleted cell lines.

For the generation of cell lines that do not express *Ctnna2* at all, promoter-proximal regulatory elements have been deleted in the founder cell line that has only one allele of the *Ctnna2* gene. The same approach as in 4.1.1.1 has been applied: The gene body has been scanned in mm10 using the UCSC Genome Browser with the ENCODE Regulation Tracks of histone modifications turned on (see Supplementary Figure 2). Hereby, it was possible to assess which parts are likely to be involved in the regulation of transcription initiation and would need to be deleted to achieve transcription abolishment in the promoter and enhancer-deleted clones. Finally, two sgRNAs have been designed to target the areas upstream and downstream of the *Ctnna2* promoter.

To establish the promoter and enhancer deleted cell lines, another genome editing step has been executed. The validated founder embryonic stem cell line with a one-allelic deletion (*a/Δ*) has been nucleofected with two sgRNAs targeting areas downstream and upstream of *Ctnna2*'s promoter to delete this area on the remaining second allele.

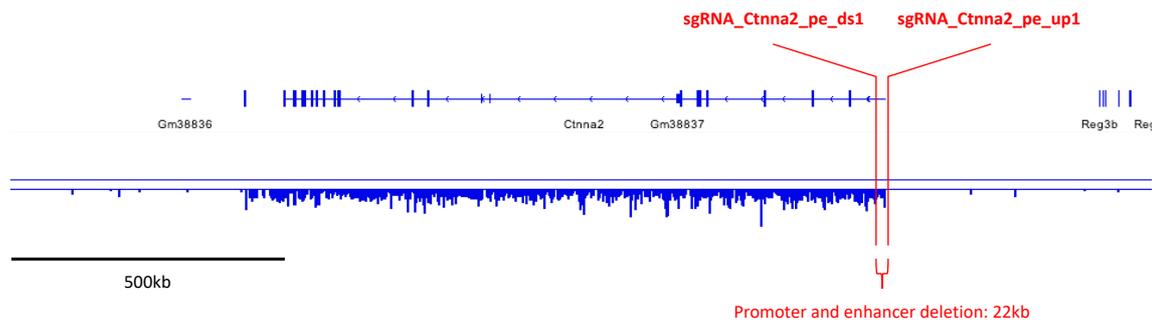


Figure 7: **Promoter-proximal deletion of *Ctnna2*'s promoter.** Gene locus of *Ctnna2* (blue - gene body at the top) ± 0.5 Mb (chr6:76,379,637-78,481,703) with corresponding transcription signal (blue - at the bottom) from nascent RNA in the parental cell line NXP010. Red vertical lines indicate the location of the two sgRNAs used to target the promoter area of *Ctnna2* and delete the regulatory elements. sgRNA_Ctnna2_pe_ds1 is located at chr6:77,962,891-77,962,910 and sgRNA_Ctnna2_pe_up1 at chr6:77,984,629-77,984,648. The promoter and enhancer deletion spans across 22 kb. Scale bar at the bottom left indicates 500 kb.

This led to a 22 kb deletion of the second allele and was validated after colony picking by a PCR screen of 96 colonies as well. Positive clones have been additionally analyzed using Sanger sequencing. Two unique clones with the shortest end resections after Cas9 cutting have been selected as the promoter and enhancer deleted clones (ape/ Δ) of the experimental cell lines.

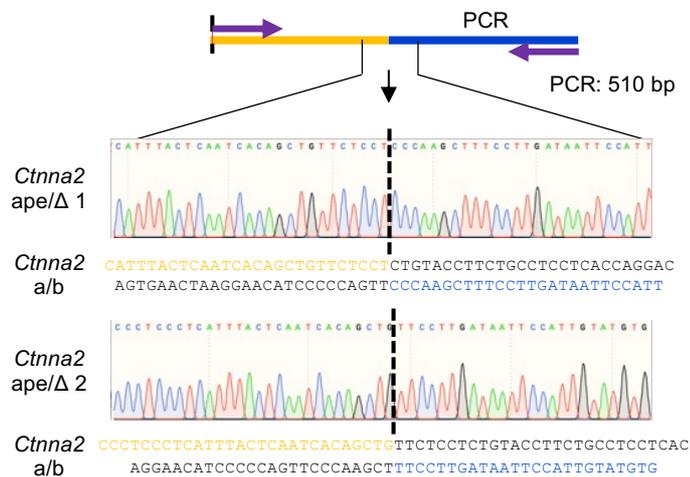


Figure 8: **Workflow for PCR validation and Sanger sequencing of positive clones after the second CRISPR/Cas9 cutting.** After deletional joining the yellow and blue part of the genome were merged and the proximal promoter area of *Ctnna2* was removed. The deletion was validated by using the PCR primers indicated as purple arrows, which produces a PCR product with a length of 510 bp. Positive clones have been additionally analyzed using Sanger sequencing. The result of the selected promoter and enhancer deleted clones (ape/ Δ 1 and ape/ Δ 2) are shown at the bottom as sequencing peaks and maintained sequence (yellow and blue) compared to the deleted original sequence (black) from the parental cell line (a/b).

4.1.1.3 Expected alleles are present in the *Ctnna2* deleted cell lines.

To validate not only the deletion, but the actual number of alleles that are present, TaqMan probes have been designed to target the whole locus as well as the promoter proximal area. This allowed checking the cell lines that have been used in the further experiments and to make sure that I was working with the correct material. The parental cell line (a/b) has two alleles of the whole locus (blue) as well as the promoter proximal area (orange). In the founder cell line (a/ Δ) with the one allelic deletion of the whole locus, only one allele can be detected with both probes, confirming the generated cell line. Also, for the two promoter and enhancer deleted clones (ape/ Δ 1 and ape/ Δ 2) the probes validate the correct number of alleles. While the area around the middle of the gene body of *Ctnna2* can still be detected and shows an allele density of one (correspondent to the founder cell line), the promoter proximal area is no longer detectable (n.d. = not detectable) by a TaqMan probe validating the deletion of this area on the remaining allele.

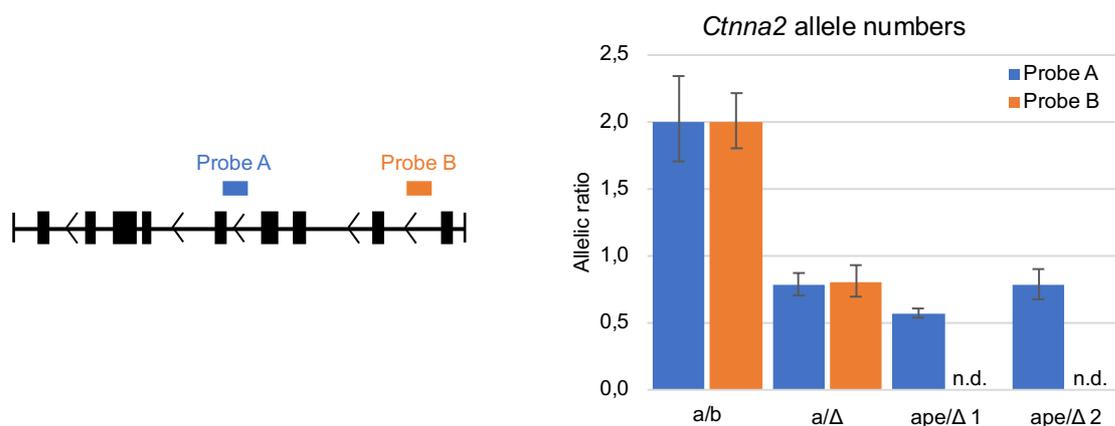


Figure 9: **TaqMan-qPCR validates the deletion of the first allele in the a/ Δ cell line and the promoter proximal area on the second allele in the ape/ Δ clones.** Left: Schematic representation of the *Ctnna2* gene body (black) with the transcriptional direction from right to left (Transcription start site on the right and transcription termination on the left). Probe A (blue) is represented in the middle of the gene body and detects the presence of the locus. Probe B (orange) is shown around the transcription start site and detects the area that is deleted during the generation of the promoter and enhancer deleted cell lines. Right: Allelic ratio of *Ctnna2* (on the y-axis) in the experimental cell lines (on the x-axis) based on the data generated with the TaqMan probes. Alleles of the whole locus (blue - Probe A) as well as the promoter proximal area (orange - Probe B) are detected. Error bars correspond to the standard deviation of the values. n.d. = not detectable. n=2.

4.1.1.4 Transcription is abolished in the *Ctnna2* promoter and enhancer deleted cell lines.

After the generation and validation of the experimental cell lines, the first goal was to assess the levels of transcription across the different cell lines. For these experiments, as stated before, the embryonic stem cell lines have been induced to neural progenitor cells since *Ctnna2* is not transcribed in ESCs. Subsequently, cells were harvested, and nuclei were isolated for Global Run On-sequencing (GRO-seq) to check transcription initiation events via nascent RNA. At the *Ctnna2* locus, the parental cell line (a/b) shows active transcription across the whole gene body in both replicates, while the founder cell line with the one allelic deletion (a/ Δ) already shows a decrease in transcriptional activity. Nonetheless, in the promoter and enhancer deleted clones (ape/ Δ 1 and ape/ Δ 2), transcriptional activity is eradicated completely at the *Ctnna2* locus as anticipated after deletion of the promoter proximal area.

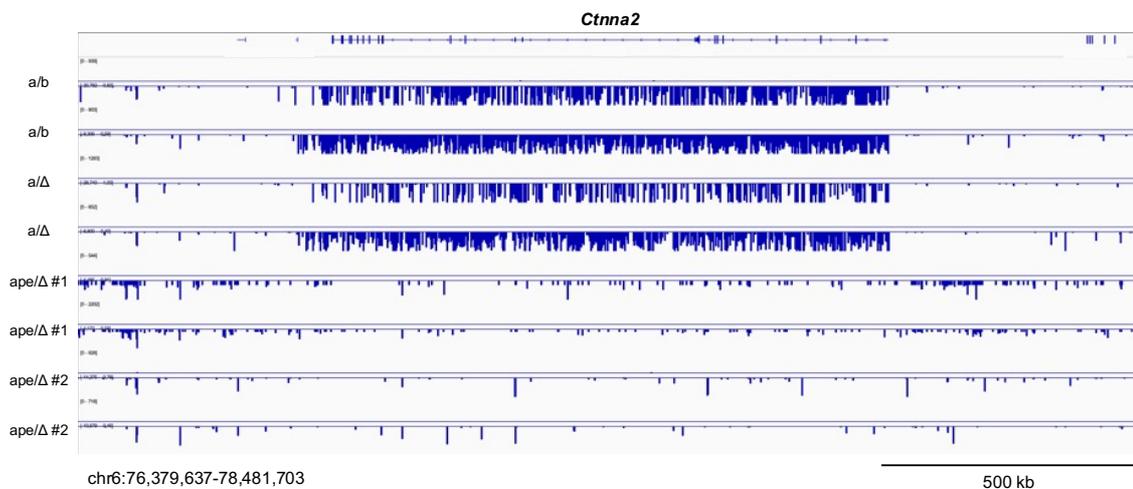


Figure 10: Transcription is abolished upon deletion of the promoter proximal area. The locus of *Ctnna2* \pm 0.5 Mb (chr6:76,379,637-78,481,703) with *Ctnna2*'s gene body is shown on the top. GRO-seq signal corresponding to active transcription is shown in dark blue on the minus strand for the different samples. From top to bottom: Two replicates of the parental cell line (a/b) shown at the top. Underneath, two replicates from the founder cell lines (a/ Δ) are shown. The lower samples are two replicates for each promoter and enhancer deleted cell line with ape/ Δ 1 and ape/ Δ 2 for the last two rows. Scale bar at the bottom right indicates 500 kb. n=2.

4.1.1.5 RT-qPCR also validates the transcriptional inhibition of *Ctnna2*

expression in the Ctnna2 promoter and enhancer deleted cell lines.

As an additional step to the analysis of nascent RNA levels via GRO-seq, total RNA levels were assessed via RT-qPCR. For this, multiple primers across the gene body of *Ctnna2* have been used to detect transcripts. After primer testing (data not shown), five primers have been selected. The primer locations across the *Ctnna2* locus are shown in Supplementary Figure 5. Cells were grown and induced into neural progenitor cells prior to RNA isolation and cDNA synthesis. cDNA of founder cells (*a/Δ*) was run together with cDNA from the promoter and enhancer deleted clones (*ape/Δ* 1 and *ape/Δ* 2), and expression fold change was compared after standardization with the house keeping gene *HPRT*. The expression of the founder cells (*a/Δ*) has been set to 1 and the experimental cell lines (*ape/Δ* 1 and *ape/Δ* 2) were normalized accordingly. The values for the expression of *Ctnna2* in *ape/Δ* 1 and *ape/Δ* 2 are between 0.000 to 0.009, and 0.001 to 0.017, respectively. This indicates a decrease in gene expression based on total RNA levels of 99.1 to 100% for *ape/Δ* 1 and a decrease of 98.3 to 99.9% *ape/Δ* 2, consistent to the data from GRO-seq.

Name	Sequences (5' to 3')	Exons
Ctnna2_17_1_1_F	GAGGAGGAGGCGAGAACTC	EX1-2
Ctnna2_17_1_1_R	CCTGGCACTGGAGCTATGAG	EX1-2
Ctnna2_13_1_2_F	ATGTAGCAGCAAGACGGCAG	EX6-7
Ctnna2_13_1_2_R	TGTACAGCATGGTGGCATTTC	EX6-7
Ctnna2_12_1_2_F	TGCAGCCCTGAATGAGTTTG	EX7-8
Ctnna2_12_1_2_R	GTCTCTCCTCCAGGGATGGT	EX7-8
Ctnna2_10_1_1_F	TGCGATTGACAAGATGACCA	EX9-10
Ctnna2_10_1_1_R	AGGGGACATTGGTTTCCAAG	EX9-10
Ctnna2_6_1_2_F	CAGAGAAAGTGCTGGAAGCC	EX13-14
Ctnna2_6_1_2_R	TCAACTTGTTCAAGCAAAGCG	EX13-14

Table 1: Primers used to target *Ctnna2*'s gene body for RT-qPCR. Name refers to the official qPCR primer names from the UCSC Genome Browser track "Mouse (mm10) Whole Transcriptome qPCR Primers" based on Zeisel A et al., *Bioinformatics* (2013).¹¹⁶. Sequences are specified for all forward and reverse primer pairs. Additionally, the primers are all designed to be intron spanning to ensure the detection of mRNA; therefore, all primers detect areas spanning from one exon to another. The exact targeted exons of *Ctnna2* are indicated here for each primer pair.



Figure 11: Reduced expression of *Ctnna2* in the experimental promoter and enhancer deleted clones. Cell lines and primer locations are illustrated on the x-axis and the expression fold change of *Ctnna2* is on the y-axis. Each bar is labeled with its value and standard deviation shown as error bars. The two clones with the deletion at the promoter proximal area show a significant reduction in terms of gene expression of *Ctnna2*.

4.1.1.6 Analysis shows lower break formation at Ctnna2 after aphidicolin treatment in Ctnna2 promoter and enhancer deleted cell lines.

Upon validating the transcriptional level of *Ctnna2*, the DNA double-strand break dynamics were analyzed in detail. For that the parental, founder, and promoter and enhancer deleted cell lines were induced to neural progenitor cells (NPCs) and nucleofected with Cas9 together with a sgRNA targeting chromosome 6, on which *Ctnna2* is located. By additional treatment with aphidicolin, a DNA polymerase alpha inhibitor, in order to induce replication stress, it is possible to stimulate DNA double-strand breaks that will translocate to each other due to proximity. Genomic DNA is harvested from these cells after 96 hours and HTGTS libraries are prepared for sequencing.

In the supplement, in Supplementary Table 1 and Supplementary Table 2 all prepared libraries that were used for analysis in this chapter are presented and compared to each other in terms of read and resulting junction numbers.

Firstly, the locus of *Cttna2* has been further investigated. Therefore, the libraries were aligned to the mm10 reference genome after sequencing, demultiplexed and run through the HTGTS pipeline (see 7.2.3.1) to yield the sequences which joined the induced CRISPR/Cas9 cut. Hereby, it is possible to extract the exact position of the joining event from translocation. In qualitative plots, these break positions are represented as individual black vertical bars across the genome. In Figure 12 below shows the break distribution around *Cttna2* is shown. The gene body is in the middle of the plot and has 500 kb spanning it on the left and right side. The area inside *Cttna2*'s gene body is highlighted with a gray background for the break distribution. Starting with the parental cell line which has two alleles (a/b), it is possible to see a strong increase in break frequency upon aphidicolin treatment (APH +) in comparison to the DMSO control (APH -). Already in the founder cell line with the one allelic deletion (a/ Δ), a reduction is observable in terms of break frequency; however, the recurrent DNA break cluster is still visible in the aphidicolin treated cells (break number decreased to circa 44%). In the two experimental cell lines with the deletion of the promoter proximal area (ape/ Δ 1 and ape/ Δ 2), the aphidicolin treatment no longer induces the same break cluster formation seen in the cells with one or two transcribable alleles (break number reduced to 3.8% for ape/ Δ 1 and 1.2% for ape/ Δ 2 compared to parental cells treated with aphidicolin).

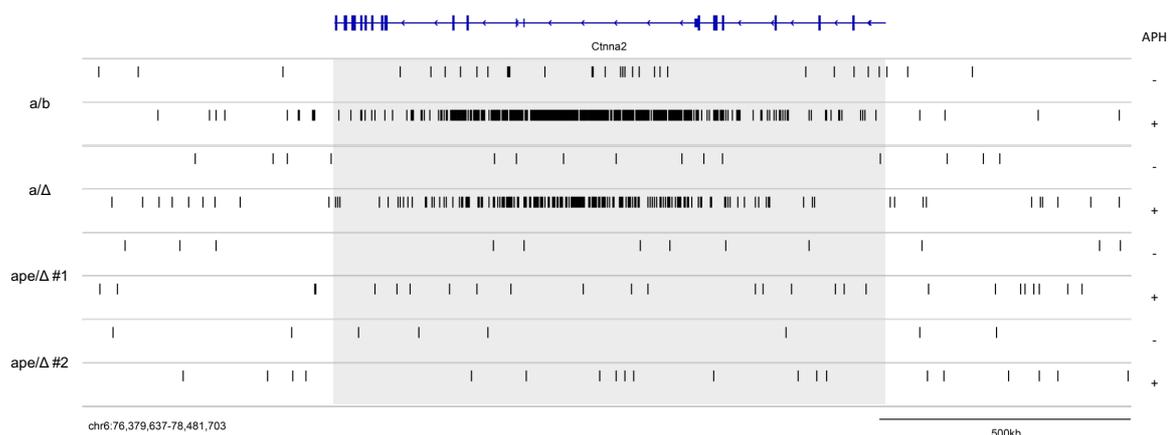


Figure 12: Break distribution around the locus of *Cttna2*. The gene body of *Cttna2* is displayed at the top in blue. On the left the breaks are assigned to the cell line labels. On the right, the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of *Cttna2*'s gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 10,000 junctions from three replicates for each sample group. $n=3$.

Since Figure 12 only shows the results from a qualitative point of view, the data of all available cell lines and replicates have been explored further. To compare different samples under different conditions with each other, junctions were normalized. First, junctions inside the gene body of *Ctnna2* were counted as well as the junctions around the induced bait break (± 10 kb) and the total chromosome 6. Bait breaks were subtracted from the total number of breaks on chromosome 6, yielding an adjusted value for aphidicolin induced breaks on chromosome 6. The number of breaks inside *Ctnna2* was then divided by the adjusted value for chromosome 6 breaks and divided by 1,000. This new value represents “Junctions per thousand”, which cannot yet be used to compare break density between different genes. Therefore, the length of *Ctnna2* has been also considered for the calculation by dividing the “Junctions per thousand” by the length of *Ctnna2* converted into megabases. The resulting “Junctions per thousand per Megabase” was then used to compare the break density between different conditions.

Similar to the qualitative view of the *Ctnna2* locus, the parental and founder cell lines with two and one transcribable alleles, respectively, show a significantly increased level of break density after aphidicolin treatment in comparison to the DMSO vehicle control samples. The mean break density for parental cells increases from approximately 23.9 to 187.0 Junctions per thousand per Megabase in response to aphidicolin treatment, while the founder cells' increase is from 8.0 to 94.5 Junctions per thousand per Megabase (both increases are statistically significant with $p < 0.0001$). After the deletion of promoter and enhancer elements on the remaining second allele, which leads to the abolishment of transcription, the break density also remains low after addition of aphidicolin to the cells for 96 hours. The two cell lines behave as following: *ape*/ Δ 1 shows an insignificant increase in mean break density from 4.3 to 4.9 Junctions per thousand per Megabase in response to APH, while for *ape*/ Δ 2 the mean break density is also increased insignificantly from 4.7 to 5.0 Junctions per thousand per Megabase.

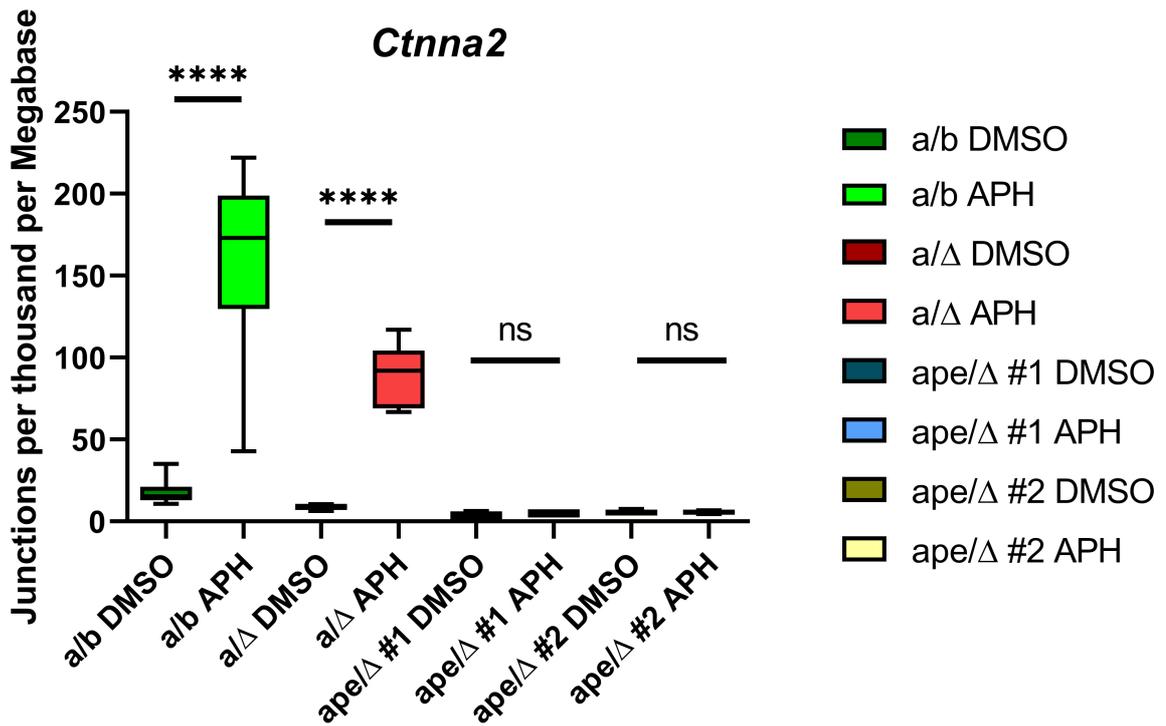


Figure 13: **Break density analysis at Ctnna2 across the experimental cell lines.** On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in “Junctions per thousand per Megabase” is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (a/b), the founder cells (a/Δ), the first p/e clone (ape/Δ #1) and the second p/e clone (ape/Δ #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. **** represents $p < 0.0001$. ns = not significant. $n \geq 3$.

4.1.1.7 Analysis shows no change in break formation at *Ccser1* after aphidicolin treatment in *Ctnna2* promoter and enhancer deleted cell lines.

As a control, further RDC-containing genes (RDC-genes) on the same chromosome (chr6) have been observed. Therefore, the RDC-genes *Ccser1* and *Grid2* have been evaluated more closely. Firstly, *Ccser1* has been viewed and assessed qualitatively. All experimental cell lines formed a visible cluster of DNA double-strand breaks at the *Ccser1* locus only after aphidicolin treatment when matching aphidicolin-treated samples with their non-treated counterparts. In comparison to the behavior at the *Ctnna2* locus (seen in Figure 12) no change in break distribution is observable when comparing the different experimental cell lines when treated with aphidicolin (approximately 95% increase at the *Ccser1* locus in the number of detectable breaks after APH treatment).

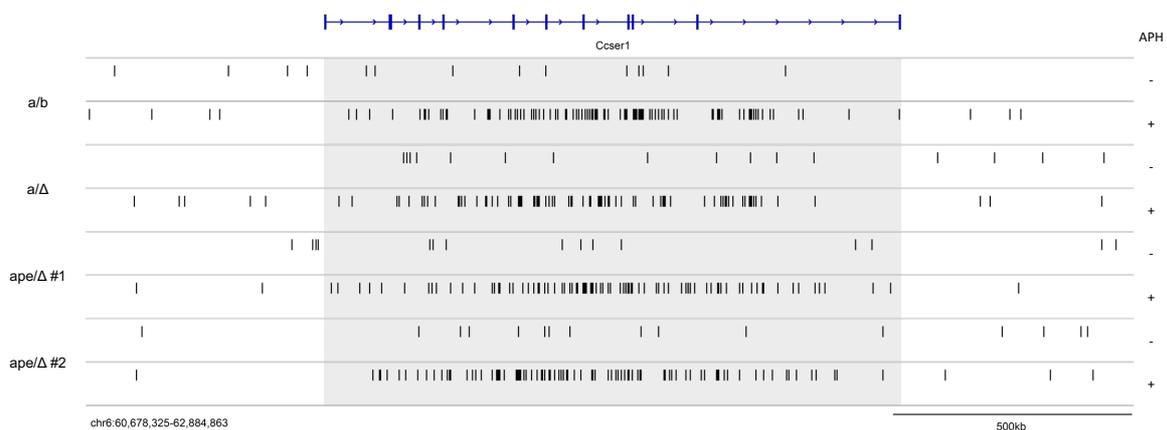


Figure 14: **Break distribution around the locus of *Ccser1*.** The gene body of *Ccser1* is illustrated at the top in blue. On the left the breaks are assigned to the cell line labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of *Ccser1*'s gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 10,000 junctions from three replicates for each sample group. $n=3$.

As before, the break density at the *Ccser1* locus has also been analyzed and quantified to compare the different conditions. While the increase of break density in “Junctions per thousand per Megabase” is significant in all cell lines across the gene body of *Ccser1*, regardless of the number of *Ctnna2* alleles and transcription, the value seems to be more variable at *Ccser1* than it is at *Ctnna2*. This could be due to the lower number of breaks at this control locus; a smaller number of breaks

at *Ccser1* could lead to more variation in terms of the calculated resulting value across the different libraries. Nonetheless, *Ccser1*, which has not been altered in any of the cell lines used in the experiments, exhibits a trend to show cluster formation under aphidicolin treatment, but not when treated with DMSO, and this trend is visible in the qualitative evaluation and significant in the quantitative break density analysis. The parental cell line shows a significant increase from a mean of 5.9 to 32.1 Junctions per thousand per Megabase and the founder cell line also shows a significant increase in the mean break density from 5.9 to 27.8 Junctions per Megabase in response to aphidicolin (both $p < 0.0001$). The increases are also statistically significant with *ape/Δ* 1 showing increases in mean break density from 4.4 to 31.7 Junctions per thousand per Megabase ($p = 0.0072$) and for *ape/Δ* 2 7.0 to 38.9 Junctions per thousand per Megabase ($p = 0.0153$).

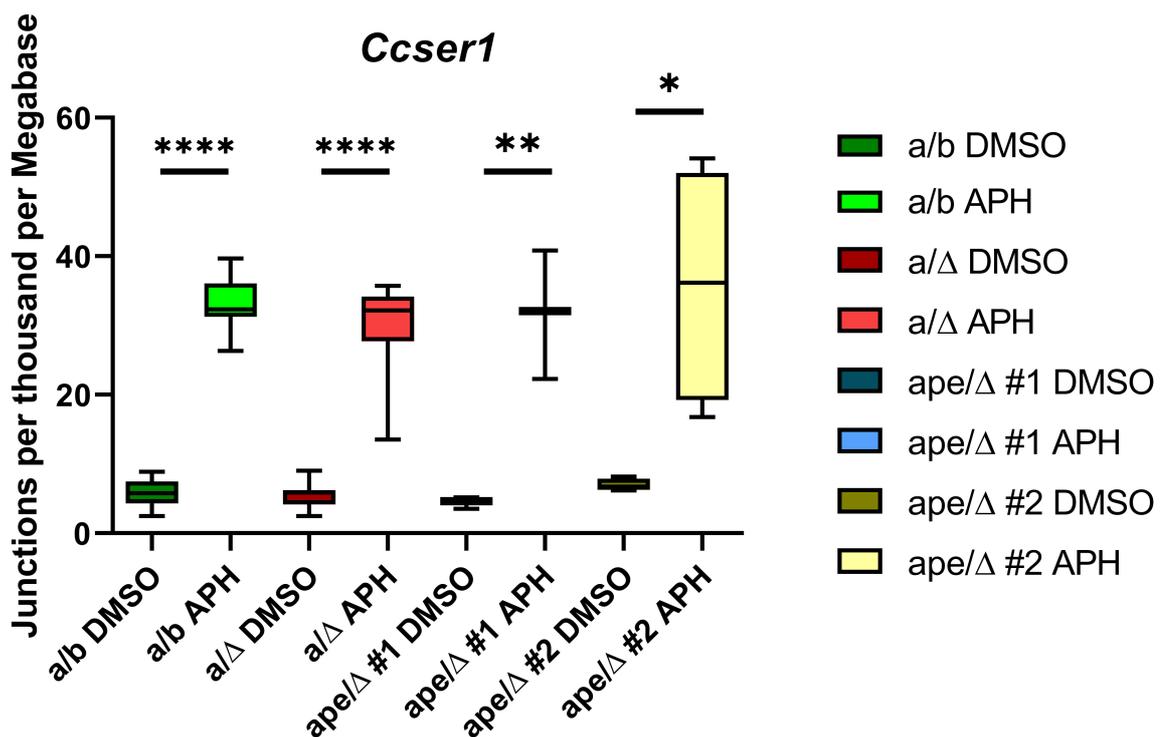


Figure 15: **Break density analysis at *Ccser1* across the experimental cell lines.** On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in “Junctions per thousand per Megabase” is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (*a/b*), the founder cells (*a/Δ*), the first *p/e* clone (*ape/Δ* #1) and the second *p/e* clone (*ape/Δ* #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. **** represents $p < 0.0001$, ** indicates $p = 0.0072$ and * represents $p = 0.0153$. ns = not significant. $n \geq 3$.

4.1.1.8 Analysis shows no change in break formation at *Grid2* after aphidicolin treatment in *Cttna2* promoter and enhancer deleted cell lines.

Another control locus for *Cttna2* on chromosome 6 is the RDC-gene *Grid2*. The break distribution around *Grid2* has been compared between samples treated with aphidicolin or left untreated. In correspondence with the prior control locus, *Ccser1*, the break clusters are still forming after aphidicolin treatment in the experimental cell lines with the proximal promoter deletion. All cell lines show an increase of around 94% in detected break junctions when comparing APH- with DMSO-treated libraries. This shows that the deletions at *Cttna2* are not affecting adjacent RDC-genes or altering their break formations.

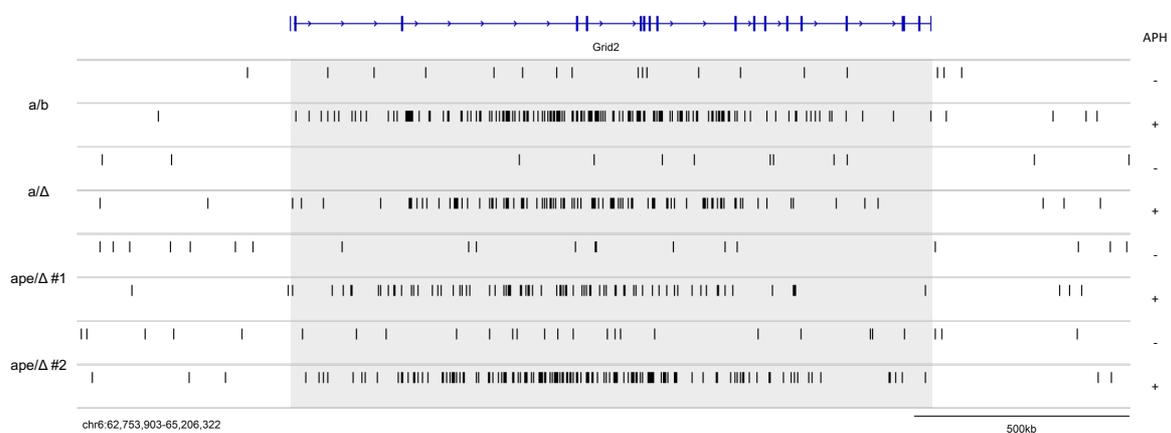


Figure 16: **Break distribution around the locus of *Grid2*.** The gene body of *Grid2* is illustrated at the top in blue. On the left the breaks are assigned to the cell line labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of *Grid2*'s gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 10,000 junctions from three replicates for each sample group. $n=3$.

The quantitative analysis at the *Grid2* locus is comparable to the *Ccser1* data shown above. The break density is lower in DMSO-treated samples than in APH-treated samples. For all cell lines except the *ape/Δ* 1 clone, the difference is also statistically significant. The lack of statistical significance can stem from the previously mentioned variation seen between the different values in junctions per thousand per Megabase. Especially, one value in the APH-treated samples of *ape/Δ* 1 is very low, putting that data point together with the DMSO-treated samples. Together with the lower sample number of 3 in comparison to the other

cell lines, this causes this sample group to have no statistically significant difference, even though a clear shift is visible in the means between the treatments. After APH treatment, the parental cells show an increase in mean break density from 11.5 to 36.6 Junctions per thousand per Megabase ($p = 0.0001$). For the founder cell lines the observable change is from a new of 10.5 to 27.6 Junctions per thousand per Megabase ($p = 0.0022$). The mean break density increases from $ape/\Delta 1$ from 5.4 to 21.0 Junctions per thousand per Megabase is not statistically significant due to the previously mentioned low break density in one of the repeats. For $ape/\Delta 2$ aphidicolin treatment led to a mean break density increase from 9.5 to 42.6 Junctions per thousand per Megabase ($p = 0.0193$).

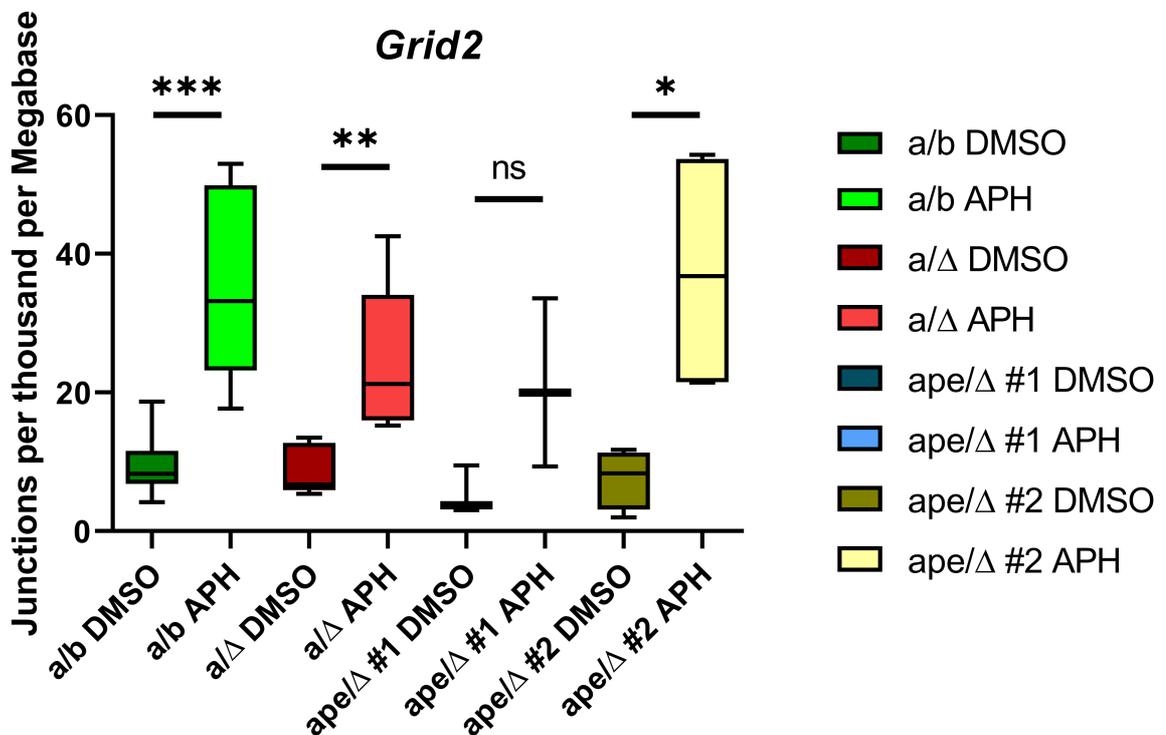


Figure 17: **Break density analysis at Grid2 across the experimental cell lines.** On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in “Junctions per thousand per Megabase” is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (a/b), the founder cells (a/Δ), the first p/e clone (ape/Δ #1) and the second p/e clone (ape/Δ #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. *** represents $p = 0.0001$, ** indicates $p = 0.0022$ and * represents $p = 0.0193$. ns = not significant. $n \geq 3$.

4.1.1.9 Replication timing at *Ctnna2* in *Ctnna2* promoter and enhancer deleted cell lines.

Since transcription has been altered and the break formation pattern has changed at the *Ctnna2* locus, I also investigated a change in replication timing. E/L-Repli-seq was chosen to compare the replication timing between the different clones at untreated conditions. For this experiment, cells were treated with BrdU for two hours, harvested, fixed with ethanol, and afterwards sorted into two fractions (early- and late-replicating cells between G1 and G2/M) based on their DNA content. The genomic DNA has been extracted, sonicated, and BrdU-incorporating DNA was further processed into libraries. Two replicates of each cell line have been prepared, sequenced, and analyzed. In Figure 18, on the x-axis the coordinates on chr6 are indicated (chr6:74,255,709-80,599,457). In general, all samples have a similar replication timing around the *Ctnna2* locus when plotting them on an early (top) to late (bottom) y-axis, all of them clearly remaining in the late replicating half of the figure. However, when comparing the two *ape/Δ* cell lines with their control cell line (*a/Δ*) more closely, the deviations in the replication timing at the *Ctnna2* locus are statistically significant for both *ape/Δ* 1 ($p=0.012$) and *ape/Δ* 2 ($p=0.023$) (statistical analysis performed by Li-Chin Wang).

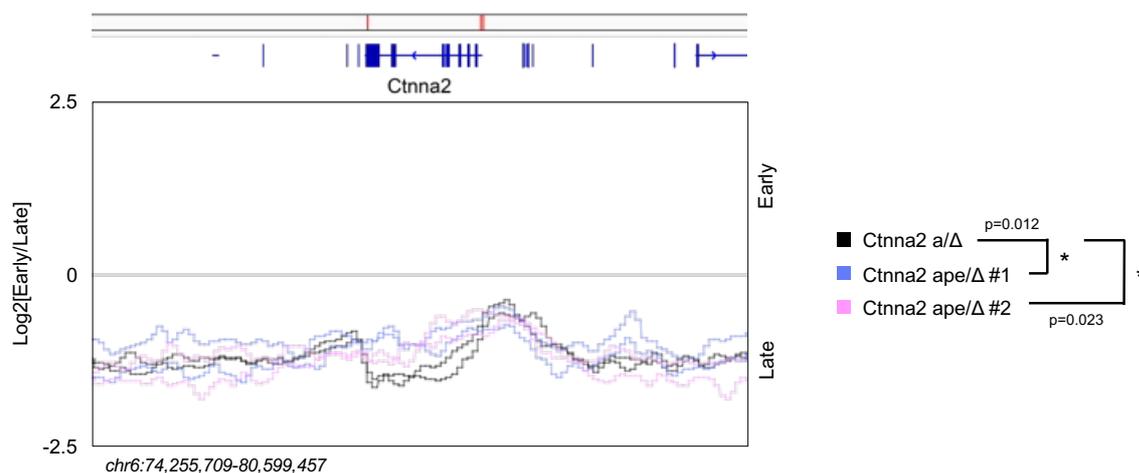


Figure 18: **E/L-Repli-seq of the experimental cell lines at the *Ctnna2* locus.** At the top of the figure, gene bodies based on the reference genomes are illustrated with *Ctnna2*'s locus being in the middle. Above in red the locations of the sgRNAs used to generate the different cell lines are indicated. The two replicates of the samples show variation amongst their counterpart but follow the same trend. *Ctnna2*-*a/Δ* founder cells are represented in black. The cell lines with promoter proximal deletions, *Ctnna2*-*ape/Δ* #1 and *Ctnna2*-*ape/Δ* #2, are shown in blue and in pink, respectively. * represents $p < 0.05$. $n=2$.

4.1.2 Neurexin 1 (*Nrxn1*)

4.1.2.1 Validation of the *Nrxn1* founder cell line.

The cell line carrying a whole locus allele deletion of *Nrxn1* has been generated by Dr. Pei-Chi Wei using CRISPR/Cas9. The parental embryonic stem cell line used for the *Nrxn1* experiments is NXP047. To generate the founder cell line with a one-allelic deletion, NXP047 has been nucleofected with two sgRNAs targeting areas downstream and upstream of *Nrxn1*. The gene locus has been checked and two targeting sgRNAs have been designed. In Supplementary Figure 3, the area around *Nrxn1* is visible in the UCSC Genome Browser displaying the mm10 mouse reference genome. To assess regulatory elements, also the ENCODE Regulation Tracks for histone modifications were activated.

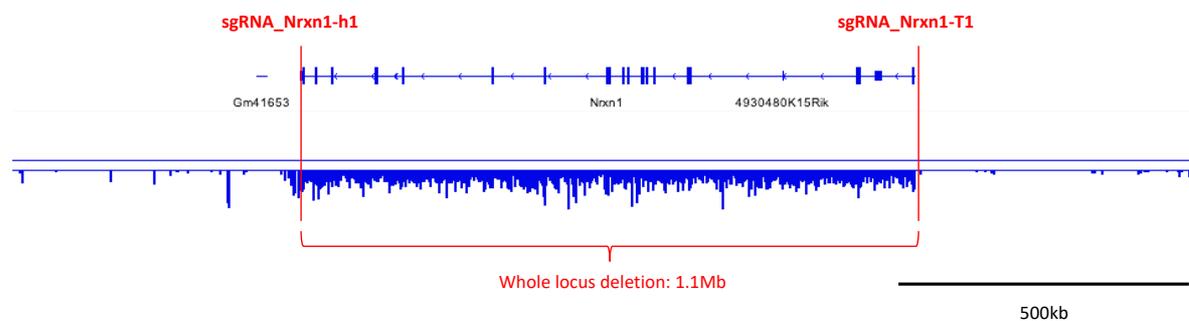


Figure 19: **Whole locus deletion of *Nrxn1***. Gene locus of *Nrxn1* (blue - gene body at the top) ± 0.5 Mb (chr17:89,531,644-91,594,802) with corresponding transcription signal (blue - at the bottom) from nascent RNA in the parental cell line NXP047. Red vertical lines indicate the location of the two sgRNAs used to target *Nrxn1* and delete the whole locus of one allele. *sgRNA_Nrxn1_h1* is located at chr17:90,032,402-90,032,421 and *sgRNA_Nrxn1_T1* at chr17:91,097,116-91,097,135. The whole locus deletion spans across 1.1 Mb. Scale bar at the bottom right indicates 500 kb.

The resulting 1.1 Mb deletion of one allele has been validated by a PCR screen of 96 colonies. One clone has been then selected as a founder cell line to produce the experimental cell lines, which are going to be referred to as the promoter and enhancer deleted clones. This work has been done by Dr. Pei-Chi Wei.

Subsequently, when I started my work using this cell line, I again validated one preselected, PCR validated clone and additionally analyzed it with Sanger sequencing.

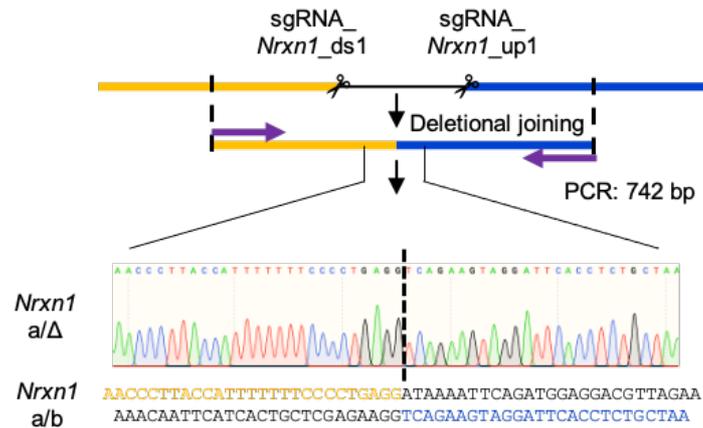


Figure 20: **Workflow for PCR validation and Sanger sequencing of positive clones after CRISPR/Cas9 cutting.** The *Nrxn1* locus is here represented here in black. sgRNAs are guiding CRISPR/Cas9 to cut downstream and upstream of the gene body (scissors). After deletional joining the yellow and blue part of the genome are merged. The deletion can be validated by using the PCR primers indicated as purple arrows, which produce a PCR product with a length of 742 bp. Positive clones have additionally been analyzed using Sanger sequencing. The result of the selected founder cell line (*a/Δ*) is shown at the bottom as sequencing peaks and maintained sequence (yellow and blue) compared to the deleted original sequence (black) from the parental cell line (*a/b*).

4.1.2.2 Validation of the *Nrxn1* promoter and enhancer deleted cell lines.

Promoter-proximal regulatory elements have been deleted in the founder cell line that has a whole locus deletion on one allele of the *Nrxn1* gene. As stated before, the gene area has been evaluated in the UCSC Genome Browser on mm10 and special care has been taken to remove regulatory elements, indicated on the ENCODE Regulation Tracks of histone modifications (see Supplementary Figure 4). Two sgRNAs have been designed to target the promoter of *Nrxn1* up- and downstream and yield a deletion of the promoter and promoter-proximal enhancer regions.

Another genome editing step was undertaken to generate the experimental cell lines. *Nrxn1 a/Δ* cells have been nucleofected with the two sgRNAs that are targeting the regulatory elements close to the promoter and thereby, this area on the remaining second allele was deleted. This led to a 11 kb deletion on the second allele and was validated by a colony PCR screen. This work has also been done by Dr. Pei-Chi Wei.

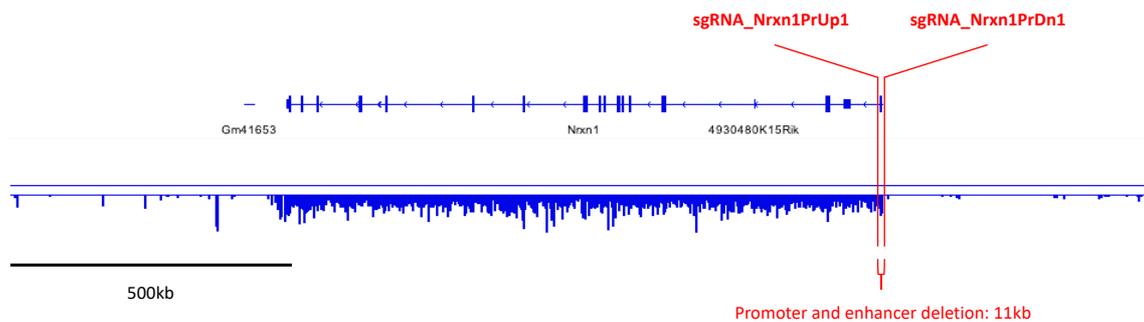


Figure 21: **Promoter-proximal deletion of *Nrxn1*'s promoter.** Gene locus of *Nrxn1* (blue gene body at the top) ± 0.5 Mb (chr17:89,531,644-91,594,802) with corresponding transcription signal (blue at the bottom) from nascent RNA in the parental cell line NXP047. Red vertical lines indicate the location of the two sgRNAs used to target the promoter area of *Nrxn1* and delete the regulatory elements. *sgRNA_Nrxn1PrUp1* is located at chr17:91,083,495-91,083,514 and *sgRNA_Nrxn1PrDn1* at chr17:91,094,628-91,094,647. The promoter and enhancer deletion spans across 11 kb. Scale bar at the bottom left indicates 500 kb.

Subsequently, when I started to work with the positive cell lines, I again validated preselected PCR validated clones and additionally analyzed them with Sanger sequencing as well. Two unique clones with the shortest end resections after Cas9 cutting were selected as the promoter and enhancer deleted clones (*ape/Δ*) of the experimental cell lines.

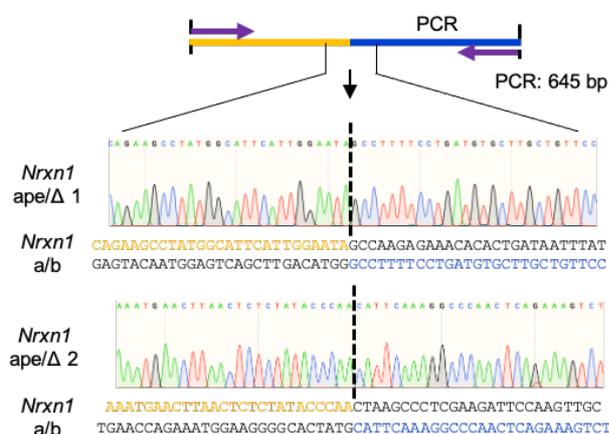


Figure 22: **Workflow for PCR validation and Sanger sequencing of positive clones after the second CRISPR/Cas9 cutting.** After deletional joining the yellow and blue part of the genome are merged and the proximal promoter area of *Nrxn1* is removed. The deletion was validated by using the PCR primers indicated as purple arrows, which produce a PCR product with a length of 645 bp. Positive clones have additionally been analyzed using Sanger sequencing. The result of the selected promoter and enhancer deleted clones (*ape/Δ 1* and *ape/Δ 2*) are shown at the bottom as sequencing peaks and maintained sequence (yellow and blue) compared to the deleted original sequence (black) from the parental cell line (a/b).

4.1.2.3 Expected alleles are present in the *Nrxn1* deleted cell lines.

Furthermore, also the allele numbers have also been confirmed by TaqMan probes. The probes were designed to target the middle of the *Nrxn1* gene (Probe A - blue) to assess the whole locus of *Nrxn1* in the experimental cell lines, while Probe B (orange) has been designed in the area that was deleted during the generation of the promoter and enhancer deleted clones. The parental cell line (a/b) has two alleles for both regions (Probe A and Probe B). Both probes could still target the founder cell line (a/ Δ) with the one allelic deletion of the whole locus, however, the signal was only half in comparison to the parental cell lines, confirming that one allele has been successfully removed from this cell line. It was also able to validate the two promoter and enhancer deleted clones (ape/ Δ 1 and ape/ Δ 2) with the probes. While the area around the middle of the *Nrxn1* gene can still be detected and shows an allelic density of one (corresponding with the founder cell line), the promoter proximal area is no longer detectable (n.d. = not detectable) by a TaqMan probe, validating the deletion of this area on the remaining second allele.

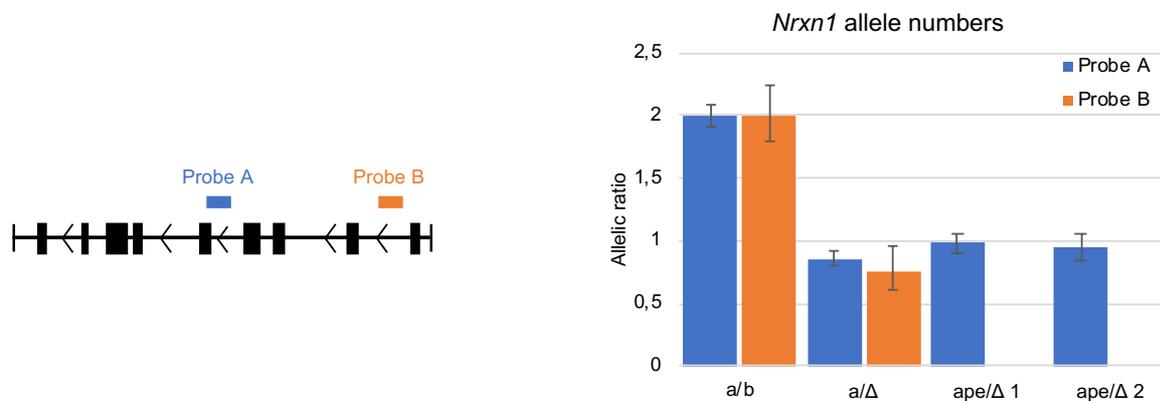


Figure 23: **TaqMan-qPCR validates the deletion of the first allele in the a/ Δ cell line and the promoter proximal area on the second allele in the ape/ Δ clones.** Left: Schematic representation of the *Nrxn1* gene body (black) with the transcriptional direction from right to left (Transcription start site on the right and transcription termination on the left). Probe A (blue) is represented in the middle of the gene body and detects the presence of the locus. Probe B (orange) is shown around the transcription start site and detects the area that is deleted during the generation of the promoter and enhancer deleted cell lines. Right: Allelic ratio of *Nrxn1* (on the y-axis) in the experimental cell lines (on the x-axis) based on the data generated with the TaqMan probes. Alleles of the whole locus (blue - Probe A) as well as the promoter proximal area (orange - Probe B) are detected. Error bars correspond to the standard deviation of the values. n.d. = not detectable. n=2.

4.1.2.4 Transcription is shortened in the *Nrxn1* promoter and enhancer deleted cell lines.

The validated cell lines have subsequently been tested to examine their transcriptional level at the *Nrxn1* locus across the different cell lines. Since *Nrxn1* is not transcribed in ESCs as well, the cell lines had to be induced to NPCs before the experiment. To check transcription initiation events via nascent RNA, cells were harvested, and nuclei were isolated for Global Run On-sequencing (GRO-seq). At the *Nrxn1* locus, the parental cell line (a/b) shows active transcription across the whole gene body in both replicates, while the founder cell line with the one allelic deletion (a/ Δ) already shows a decrease in transcriptional activity. In the promoter and enhancer deleted clones (ape/ Δ 1 and ape/ Δ 2) transcriptional activity is eradicated at the original *Nrxn1* transcription start site as anticipated after deletion of the promoter proximal area; however, a shorter isoform of approximately 400 kb in length is still being transcribed due to another promoter being active around the middle of the gene body.

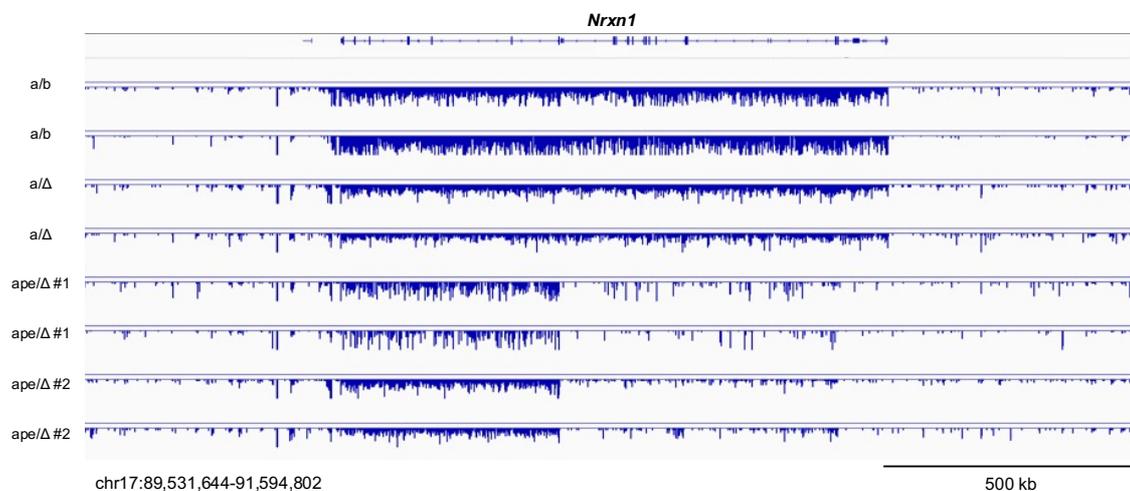


Figure 24: **A short isoform of *Nrxn1* is transcribed after deleting the promoter proximal area near the transcription start site.** The locus of *Nrxn1* ± 0.5 Mb (chr17:89,531,644-91,594,802) with *Nrxn1*'s gene body shown on the top. GRO-seq signal corresponding to active transcription is shown in dark blue on the minus strand for the different samples. From top to bottom: Two replicates of the parental cell line (a/b) shown at the top. Underneath, two replicates from the founder cell lines (a/ Δ) are shown. The lower samples are two replicates for each promoter and enhancer deleted cell line with ape/ Δ 1 and ape/ Δ 2 for the last two rows. $n=2$. Scale bar at the bottom right indicates 500 kb.

4.1.2.5 Analysis shows lower break formation at Nrnx1 after aphidicolin treatment in Nrnx1 promoter and enhancer deleted cell lines.

After observing the change in transcriptional activity in Nrnx1, the DNA double-strand break dynamics were investigated as well. Parental, founder, as well as promoter and enhancer deleted cell lines were induced to NPCs and nucleofected with Cas9 together with a sgRNA targeting chromosome 17, on which Nrnx1 is situated. Aphidicolin was used to induce additional breaks that will then translocate to the CRISPR-induced cut and generate junctions, which can be captured via sequencing.

In the supplement, in Supplementary Table 3 and Supplementary Table 4, all libraries prepared and used for analysis of this chapter are presented and compared to each other in terms of reads and resulting junctions.

Firstly, the locus of *Nrxn1* has been further investigated. The libraries were aligned to the mm10 reference genome after sequencing, demultiplexed, and run through the HTGTS pipeline to yield the sequences which joined the induced CRISPR/Cas9 cut. This allows me to extract the exact position of the joining event from translocation. In qualitative plots, this break positions are represented as individual black vertical bars across the genome. In Figure 25 below, the break distribution around *Nrxn1* is shown. The gene body is in the middle of the plot and is flanked by 500 kb on the left and right side. The area inside *Nrxn1*'s gene body is highlighted with a gray background for the break distribution. Starting with the parental cell line which has two alleles (a/b), it is possible to see a strong increase in break frequency with aphidicolin treatment (APH +) in comparison to the DMSO control (APH -). Already in the founder cell line with one allelic deletion (a/ Δ), a reduction is observable in terms of break frequency, however, the recurrent DNA break cluster is still visible in the aphidicolin treated cells. In the two experimental cell lines with the deletion of the promoter proximal area (ape/ Δ 1 and ape/ Δ 2) which consequently express a shorter isoform of the *Nrxn1* transcript, the aphidicolin treatment no longer induces the same break cluster formation (75 to 90% reduction in the break junction number at *Nrxn1* compared to the parental cells under aphidicolin) as in the cells with one or two transcribable alleles.

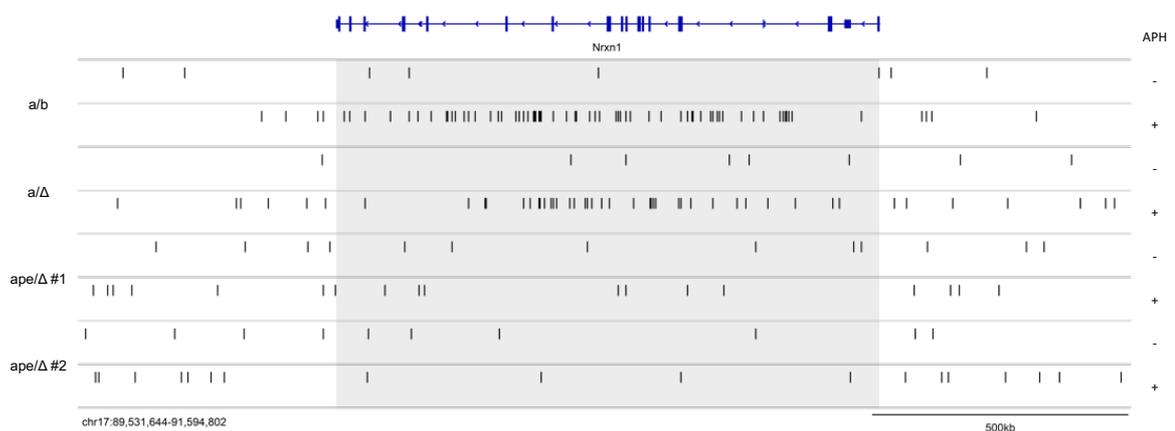


Figure 25: Break distribution around the locus of *Nrxn1*. The gene body of *Nrxn1* is illustrated at the top in blue. On the left the breaks are assigned to the cell line labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of *Nrxn1*'s gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 30,000 junctions from three replicates for each sample group. $n=3$.

The visualized data from the experimental cell lines under the different conditions in Figure 25 have also been quantified. To be able to compare the samples with each other, the unit “Junctions per thousand per Megabase” was applied. It was calculated by counting the number of junctions inside the gene body of *Nrxn1*, on chromosome 17 as well as in the area around the bait break (± 10 kb). Bait breaks were subtracted from the total number of breaks on chromosome 17, giving an adjusted value for aphidicolin induced breaks on chromosome 17. The number of breaks inside *Nrxn1* was then divided by the adjusted value for chromosome 17 breaks and divided by 1,000. This new value indicates the “Junctions per thousand”. The length of *Nrxn1* was then considered in the calculation by dividing the “Junctions per thousand” by the length of *Nrxn1* converted into megabases. The resulting “Junctions per thousand per Megabase” was then used to compare the break density between the different conditions.

Just as for the qualitative view of the *Nrxn1* locus, the parental and founder cell lines with two and one transcribable allele, respectively, show a significantly increased level of break density after aphidicolin treatment in comparison to the DMSO vehicle control samples. The parental cell line shows an increase of their mean break density from 3.5 to 15.7 Junctions per thousand per Megabase ($p \leq 0.01$), while the founder cell line already has an increase from their mean break density of only 1.3 to 5.3 Junctions per thousand per Megabase which also remains statistically significant ($p \leq 0.01$). After the deletion of promoter and enhancer elements on the remaining second allele, which leads to the expression of a shorter transcript, the break density also remains low upon aphidicolin treatment and no statistically significant difference is observable between the two treatments in both *ape*/ Δ 1 and *ape*/ Δ 2.

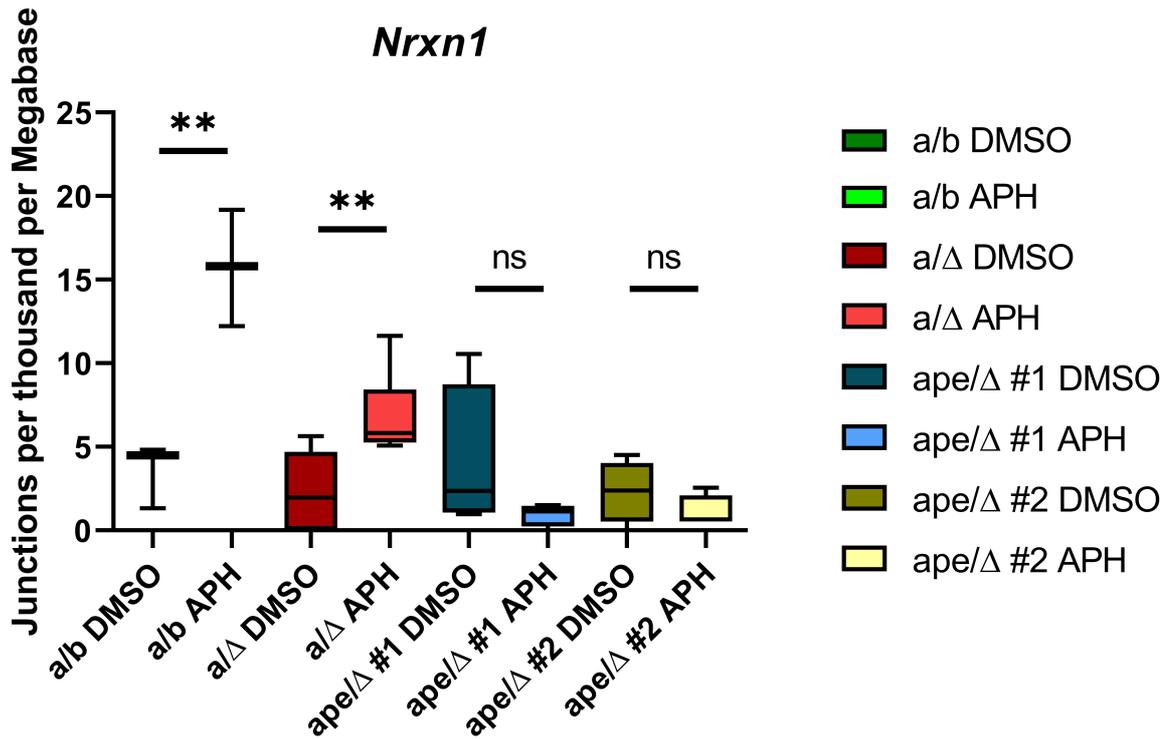


Figure 26: **Break density analysis at Nrnx1 across the experimental cell lines.** On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in “Junctions per thousand per Megabase” is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (a/b), the founder cells (a/Δ), the first p/e clone (ape/Δ #1) and the second p/e clone (ape/Δ #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. ** represents $p \leq 0.01$. ns = not significant. $n \geq 3$.

4.1.2.6 Analysis shows no change in break formation at *Ptprm* after aphidicolin treatment in *Nrxn1* promoter and enhancer deleted cell lines.

Other RDC-genes on chromosome 17 were also investigated as a control. For this purpose, the RDC-genes *Ptprm* and *Pde10a* were evaluated more closely. Firstly, *Ptprm* was displayed and assessed qualitatively. All experimental cell lines formed a visible cluster of DNA double-strand breaks at the *Ptprm* locus only after aphidicolin treatment when matching aphidicolin-treated samples with their non-treated counterparts. In comparison to the behavior at the *Nrxn1* locus (seen in Figure 25), no change in break distribution is observable when comparing the different experimental cell lines treated with aphidicolin. All cell lines show an increase in the number of break junctions detected of around 86% when treated with APH.

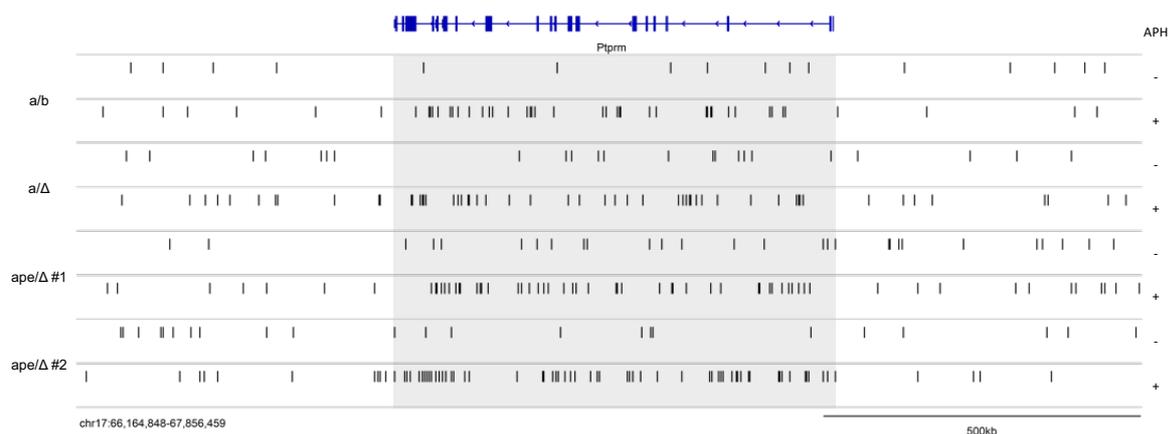


Figure 27: **Break distribution around the locus of *Ptprm*.** The gene body of *Ptprm* is illustrated at the top in blue. On the left the breaks are assigned to the cell line labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of *Ptprm*'s gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 30,000 junctions from three replicates for each sample group. $n=3$.

Like before, also the break density at the *Ptprm* locus has been analyzed and quantified to compare the different conditions. While the increase of break density in “Junctions per thousand per Megabase” is visible in all cell lines across the gene body of *Ptprm*, regardless of the number of *Nrxn1* alleles and transcription, the results for *Ptprm* are more variable and therefore, only one matched DMSO- and APH-treated cell line has statistical significance.

A possible reason could be the lower number of breaks at this control locus, since a smaller number of APH-induced breaks at *Ptprm* could lead to more variation and less prominent distinctions from the DMSO-treated control. Nonetheless, the trend that *Ptprm*, which has not been altered in any of the cell lines used in the experiments, shows cluster formation under aphidicolin treatment but not when treated with DMSO is visible, although not significant statistically in the quantitative break density analysis.

The concrete numbers for the increases in Junctions per thousand per Megabase are the following: The mean break density for the parental cell line changes from 9.7 Junctions per thousand per Megabase in DMSO-treated conditions to 17.1 Junctions per thousand per Megabase in APH-treated conditions. For the founder cells the mean changes are from 5.6 Junctions per thousand per Megabase to 12.4 Junctions per thousand per Megabase. The *ape/Δ 1* shows mean break density changes from 8.4 to 11.9 Junctions per thousand per Megabase in response to aphidicolin treatment. The only significant change at the *Ptprm* locus was observed for *ape/Δ 2*, namely the increase in mean break density from 5.2 to 17.5 Junctions per thousand per Megabase after APH ($p = 0.0019$).

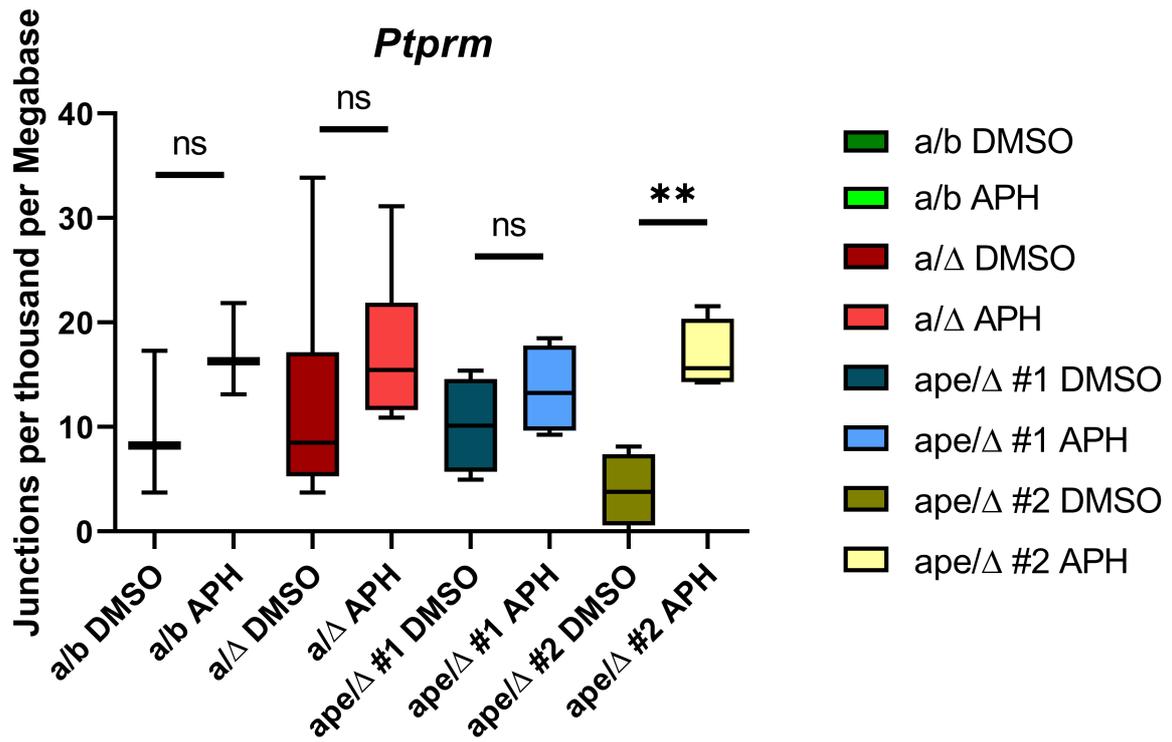


Figure 28: **Break density analysis at *Ptpm* across the experimental cell lines.** On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in “Junctions per thousand per Megabase” is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (a/b), the founder cells (a/Δ), the first p/e clone (ape/Δ #1) and the second p/e clone (ape/Δ #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. ** indicates $p = 0.0019$. ns = not significant. $n \geq 3$.

4.1.2.7 Analysis shows no change in break formation at *Pde10a* after aphidicolin treatment in *Nrxn1* promoter and enhancer deleted cell lines.

Another control locus for *Nrxn1* on chromosome 17 is the RDC-gene *Pde10a*. Similar as for *Ptprm*, more breaks were observed in *Pde10a* upon aphidicolin treatment; however, the difference between DMSO- and APH-treated samples is not very prominent at this locus. The same pattern is evident for all cell lines, including the experimental cell lines with the proximal promoter deletion. The deletion of *Nrxn1*'s first promoter does not seem to affect the break formation at *Pde10a*; nevertheless, the total number of junctions identified at this gene is very low.

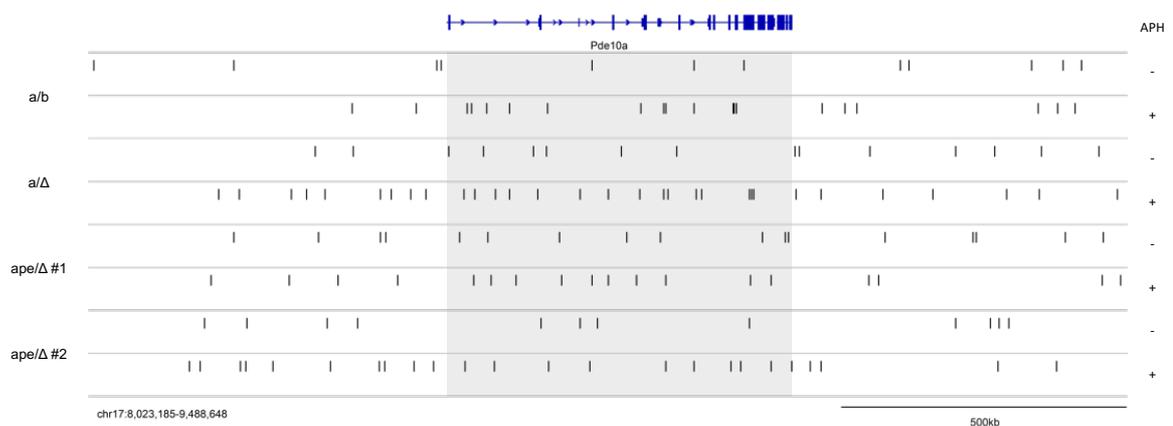


Figure 29: **Break distribution around the locus of *Pde10a*.** The gene body of *Pde10a* is illustrated at the top in blue. On the left the breaks are assigned to the cell line labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of *Pde10a*'s gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 30,000 junctions from three replicates for each sample group. $n=3$.

Quantitative analysis of the *Pde10a* locus confirms that the break density at *Pde10a* is low. When comparing the DMSO- and the APH-treated samples, no significant difference can be detected between the samples since the number of APH-induced breaks is small at this locus. Similar to the previously discussed control loci (*Ccser1*, *Grid2* and *Ptpm*), the low numbers of junctions lead to a high variation in the “Junction per thousand per Megabase” values at *Pde10a* as well. None of the break density responses to aphidicolin are significant at this control locus.

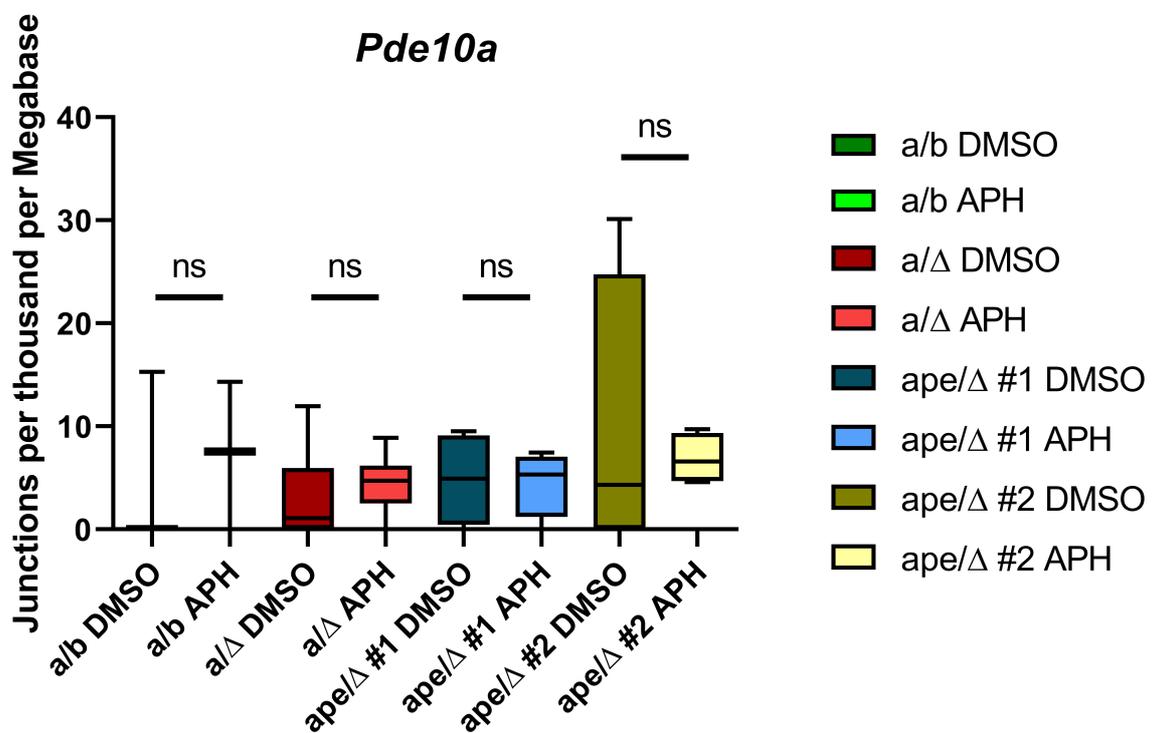


Figure 30: **Break density analysis at *Pde10a* across the experimental cell lines.** On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in “Junctions per thousand per Megabase” is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (a/b), the founder cells (a/Δ), the first p/e clone (ape/Δ #1) and the second p/e clone (ape/Δ #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. ns = not significant. n ≥ 3.

4.1.2.8 Replication timing at *Nrxn1* in *Nrxn1* promoter and enhancer deleted cell lines.

Due to the change in the transcription program and the alteration of the break formation pattern at *Nrxn1*, I decided to investigate a potential change in the replication timing. For this experiment, I chose E/L-Repli-seq to compare the replication timing between the different clones at untreated conditions. Two replicates of each cell line have been prepared, sequenced, and analyzed. In Figure 31, on the x-axis the coordinates on chr17 are represented (chr17:88,543,123-92,567,360). In general, all samples have a similar replication timing around the *Nrxn1* locus when plotting them on an early (top) to late (bottom) y-axis. When comparing the two *ape/Δ* cell lines harboring a promoter and enhancer deletion with their control cell line (*a/Δ*), the deviations are not statically significant ($p=0.11$ for *ape/Δ* #1; $p=0.85$ for *ape/Δ* #2), meaning that while transcription and break formation did change, the replication timing did not change between the cell lines (statistical analysis performed by Li-Chin Wang).

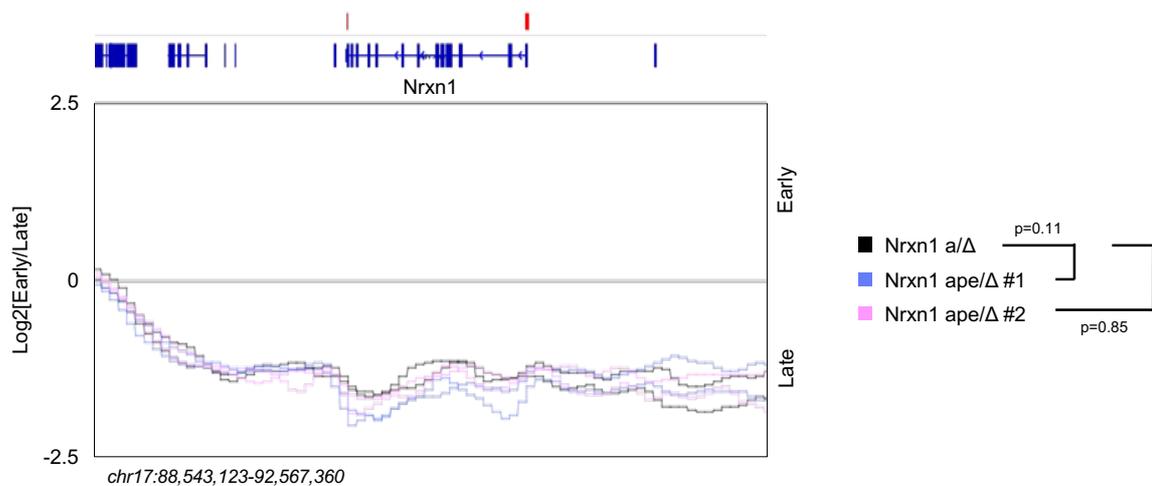


Figure 31: **E/L-Repli-seq of the experimental cell lines at the *Nrxn1* locus.** On the x-axis, the coordinates around the *Nrxn1* are shown, while on the y-axis on the right neutral (bottom) to late (top) replication timing is indicated. At the top of the figure, gene bodies based on the reference genomes are illustrated with *Nrxn1*'s locus being in the middle. Additionally, above in red the locations of the sgRNAs used for cell line generation are indicated. The two replicates of the samples show variation amongst their counterpart but follow the same trend. *Nrxn1-a/Δ* founder cells are represented in black. The cell lines with promoter proximal deletions, *Nrxn1-ape/Δ* #1 and *Nrxn1-ape/Δ* #2, are shown in blue and in pink, respectively. $n=2$.

4.2 RDCs are single-ended DSBs.

Since RDCs are induced through replication stress, I hypothesized that the observed double-strand breaks are single-ended DNA double-strand breaks resulting from daughter strands breaking off a collapsed replisome. Instead of the induced two-ended double-strand breaks, that were generated by CRISPR/Cas9 cutting and distribute within an area of approximately 50 kb around the target sequence, the chromosome locations of the aphidicolin-induced breaks distribute more broadly across the whole gene body of the RDC-genes (up to 2 Mb).

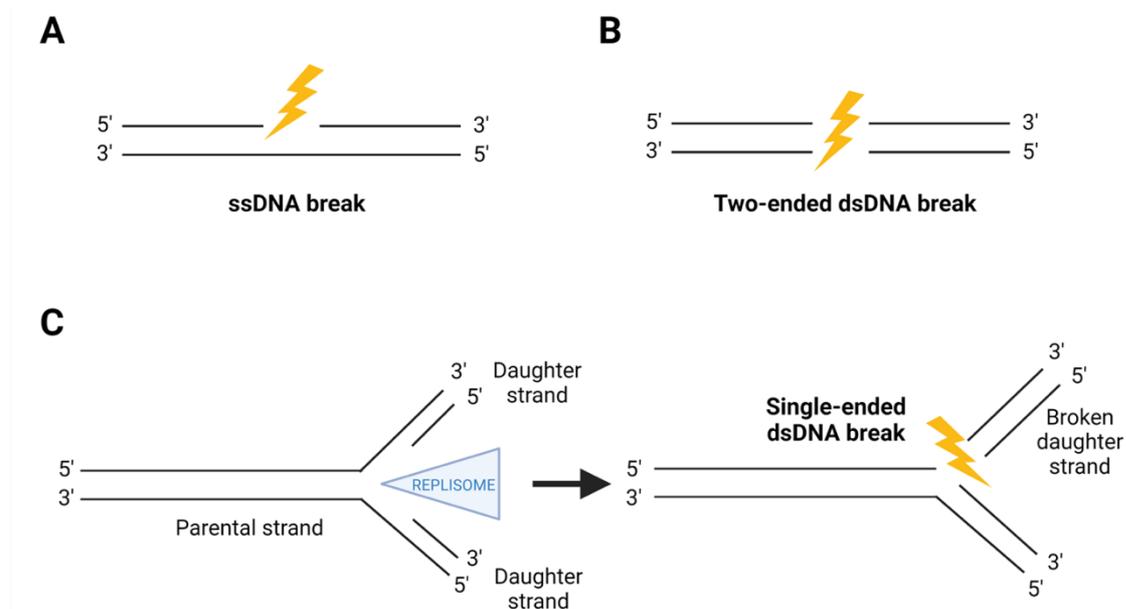


Figure 32: **Scheme of different DNA breaks.** A: Single-stranded DNA break, in which only the upper of the two DNA strands shows a gap. B: Two-ended DNA double-strand break, which is the “typical” double-strand break. C: Single-ended DNA double-strand break, which is formed after replisome collapse and leads to a broken off daughter strand.

To investigate the break peak pattern of CRISPR-induced versus aphidicolin-induced breaks, HTGTS data from the parental cell line NXP010, which has been shown before, has been separated according to the two DNA strands. In this way, the directionality of the break and its subsequent junction formation can be detected. Each distinct break position has been extended from one precise base position to 50,000 bases, which resulted in long horizontal signals instead of single vertical bars. By doing this as well as piling up the data, it is possible to observe

distinct peaks forming, and these peaks would present differently, based on the type of DNA break.

Firstly, the locus of the CRISPR/Cas9-induced bait break was studied to understand how double-ended DNA double-strand breaks would present in this representation. Therefore, I used sample files in which the bait break has not been removed from the data. The bait break can be distinguished from the pattern observed at RDC-genes by its additional presence in the vehicle control-treated libraries (DMSO). Additionally, the break pattern does not change when comparing DMSO- with APH-treated samples. The break pattern at the bait break disseminates around an area of 50 kb with the target sequence located in the middle of the peak area. It presents itself as a rectangle, showing a steady signal with a sudden drop on either side.

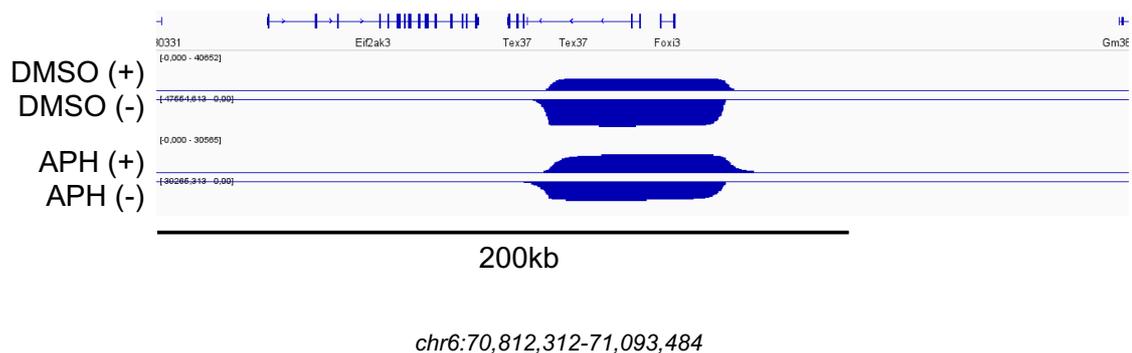


Figure 33: **Junction pattern from double-ended DSB at the induced CRISPR/Cas9 cut site.** Bait break site in extended HTGTS data. Base positions of single junctions have been extended to 10kb and split to their alignment to the + and - strand, accordingly. On the top, gene annotations are visible with the gene bodies of *Eif2ak3*, *Tex37* and *Foxi3* being visible. Underneath, DMSO and APH samples are illustrated, and the signal data of the DNA is represented on the two strands as separate data. Scale bar indicates 200 kb. The chromosome coordinates are specified (*chr6:70,812,312-71,093,484*).

Secondly, an off-target location of the CRISPR-induced cut has been examined. The pattern closely resembles the rectangular bait break peak. The peaks' presence in DMSO- as well as APH-treated libraries indicates that this type of break is not induced by replication stress, but by Cas9 cutting. Additionally, the peak continues to be symmetrical on the + and - strands and shows a very similar size to the bait break at approximately 50 kb.

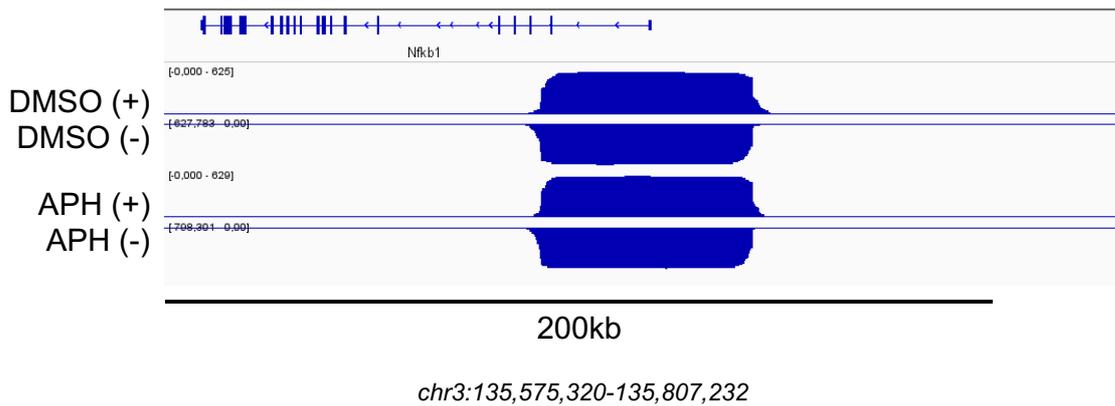


Figure 34: **Junction pattern from double-ended DSB at an CRISPR off-target location.** Off-target site in extended HTGTS data. Base positions of single junctions have been extended to 10kb and split to their alignment to the + and - strand, accordingly. On the top, gene annotations are visible with the gene body of *Nfkb1* being visible. Underneath, DMSO and APH samples are illustrated, and the signal data of the DNA is represented on the two strands as separate data. Scale bar indicates 200 kb. The chromosome coordinates are specified (*chr3:135,575,320-135,807,232*).

At the RDC-gene *Ctnna2*, the junction pattern is completely different when compared to the CRISPR-induced breaks seen before in Figure 33 and Figure 34. RDCs mostly present as double peaks at the RDC-genes. The lack of peaks in the DMSO-treated samples indicates that the breaks are induced only after applying replication stress to the cells. Additionally, when comparing the coordinates on the + and - strands, the peaks are not symmetrical but instead shifted apart.

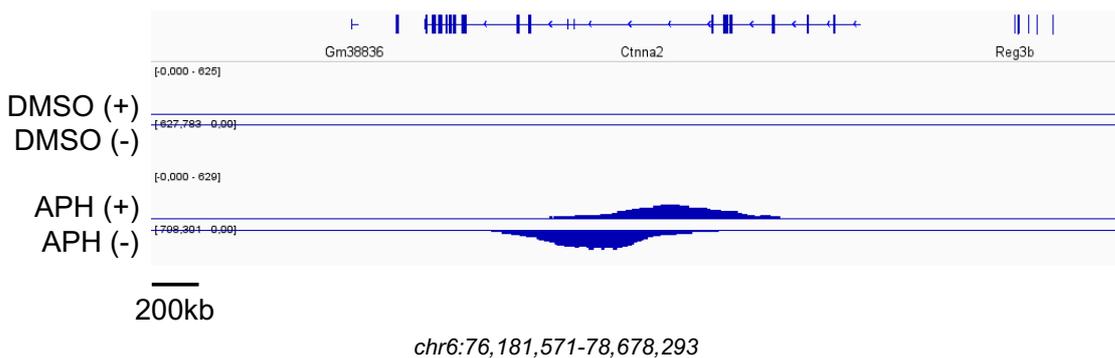


Figure 35: **RDC junction pattern differs from double-ended DSB.** RDC-gene *Ctnna2* in extended HTGTS data. Base positions of single junctions have been extended to 10 kb and split to their alignment to the + and - strand, accordingly. On the top, gene annotations are visible with the gene bodies of *Gm38836*, *Ctnna2* and *Reg3b* being visible. Underneath, DMSO and APH samples are illustrated, and the signal data of the DNA is represented on the two strands as separate data. Scale bar indicates 200 kb. The chromosome coordinates are specified (*chr6:76,181,571-78,678,293*).

As seen in Figure 35, the peaks appear to be moving apart from each other. Therefore, the next set of experiments aimed to assess if RDCs are aphidicolin dosage-dependent. Here, I hypothesized that the shifting indicates the fork stalling and fork collapse before the formation of the junction. Consequently, more aphidicolin would lead to more breaks and higher peaks in the data.

When applying low to high concentrations (0.2 to 0.6 μM) of aphidicolin, a significant DNA break cluster shift was observable at *Ctnna2* that likely corresponds to the stalling of the replication fork. Furthermore, I observed a total increase in the number of DNA breaks. Additionally, when taking transcription and replication direction into consideration, it is evident that the gene body of *Ctnna2* is transcribed into the direction of the centromere (shown here as from right to left). In Figure 36, the left illustration shows the right moving replication fork colored in blue, while the left moving fork is colored in pink. Therefore, the pink replication fork and transcription are traveling into the same direction. The replication moving towards the right likely leads to more conflict, which might be the reason why the blue junction density peak is higher under the same treatment conditions.

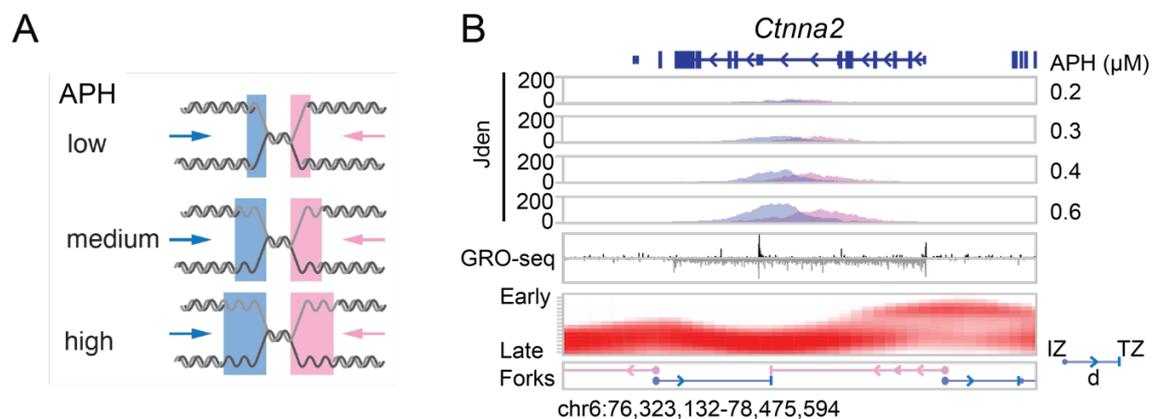


Figure 36: Recurrent DNA break clusters corresponds to replication fork stalling. Left: Illustration to indicate fork stalling under different aphidicolin (Aph) concentrations. Low levels of aphidicolin appear to have less replication fork stalling than high levels of aphidicolin treatment. The blue arrow indicates replication from left to the right and corresponds to the stalling at the blue area, while the pink is corresponding to the replication fork, which is traveling from the right to the left and has its own pink area of stalling. Right: The gene body of *Ctnna2* is illustrated in green. Underneath the data for different aphidicolin (APH) concentrations (in μM) is indicated, increasing from the top to the bottom tracks. The data tracks represent junction density (Jden) of the breaks. Blue peaks match the right-moving replication fork while the pink peaks match the left-moving fork.

4.3 Head-on collision leads to more DNA breaks.

Based on these data, I consequently hypothesized that not only the encounter of transcription and replication influences the recurrent DNA break cluster formation, but also the directionality of the encounter. From the first data, it seemed that conflicts that lead to collision of transcription and replication facing each other scar the genome at a greater extent, which is also in accordance with the literature. These conflicts are termed “head-on”, while on the other hand, collisions of transcription machinery and replication fork, that are moving into the same direction are termed “codirectional”.

Another example is the RDC-gene *Tenm3*, which is transcribed into the direction of the centromere (right to left) and shown below in Figure 37. The blue peaks representing junctions are higher than the pink peaks. The pink replication fork is traveling into the same direction as the transcription, while the blue replication fork is traveling against the transcription and will lead to a head-on encounter.

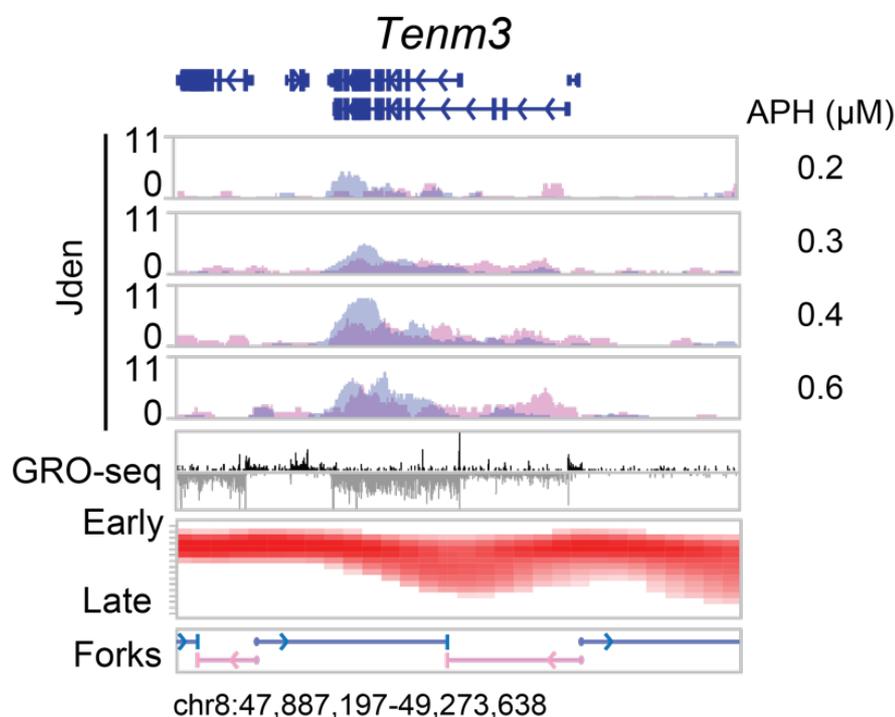


Figure 37: **Head-on collision leads to more DNA breaks at the *Tenm3* locus.** The gene body of *Tenm3* is illustrated in green. Underneath the data for different aphidicolin (APH) concentrations (in μM) is indicated, increasing from the top to the bottom tracks. The data tracks represent junction density of the breaks. Blue peaks match the right-moving replication fork, while the pink peaks match the left-moving fork. Underneath the junction data, signal of nascent RNA is plotted from GRO-seq indicating transcription on the negative strand

(gray). The red curve corresponds to published High-resolution Repli-Seq and indicates which part of the shown coordinates are early (top) or late (bottom) replicating. This data was used to extract the fork directionality shown as blue and pink arrows underneath. The chromosome coordinates are indicated as chr8:47,887,297-49,273,638.

This behavior was also investigated on a whole genome level. For this purpose, a genome-wide quantification of all RDC-genes was necessary. The junctions found in centromeric or telomeric sequences were counted and proportions have been calculated to assess if a significant difference can be detected. In the control (DMSO) condition, the proportion of formed junctions after codirectional and head-on collisions is similar. However, starting from the 0.2 to the 0.6 μM aphidicolin concentration, the proportion of formed junctions decreases for gene areas that encounter codirectional collisions, while the proportion of head-on collisions increases. In conclusion, unbiased genome-wide analyses revealed that head-on collisions generate 40% more DNA breaks than co-directional transcription and replication conflicts.

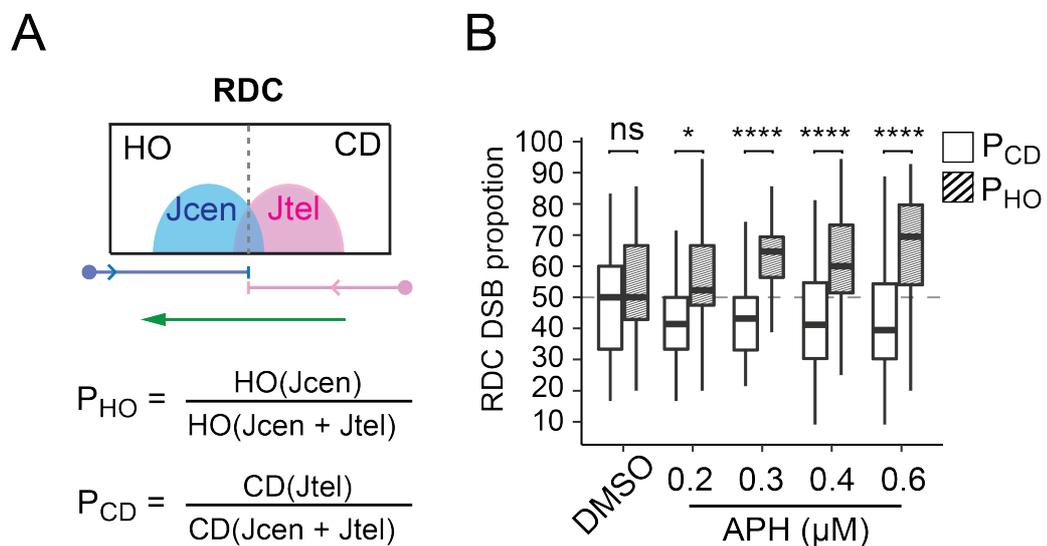


Figure 38: **Genome-wide increase of DNA breaks due to head-on collisions.** A: Illustration depicting the transcription directionality in the context of the replication fork. In the case of a telomere to centromere transcription, a right-moving (blue) replication fork will collide with it in a head-on manner (“HO” white) while a left-moving (pink) replication fork will collide co-directionally (“CD” dashed). In the case of a centromere to telomere transcription direction, head-on and co-direction collision are exchanged. B: Proportions of centromeric to telomeric junctions per RDC (y-axis) were calculated and displayed according to their aphidicolin concentration (x-axis).

5 Discussion

The motivation behind this work was to investigate whether and how transcription is involved in the formation of recurrent DNA break clusters. With the core hypothesis being that transcription/replication collisions are the driver of the breaks; transcription was modulated in a replication stress scenario. To this extent, the promoter-proximal regulatory elements were deleted using CRISPR/Cas9 to generate cell lines in which the expression of either the highly transcribed *Ctnna2* or the long isoform of the lowly transcribed *Nrxn1* has been ablated. The DNA break density was then compared between the cell lines under normal conditions and after inhibition of active transcription in response to a low dose aphidicolin treatment, which induces replication fork stalling. I found that the break density at the clusters was significantly decreased once transcription had been abolished or shortened. Additionally, no changes in replication timing were observed as a direct response to altered transcription and its subsequent changes in the break clusters. I was also able to identify the breaks as single-ended double-strand breaks, again linking them to the hypothesis of transcription/replication collision driving development of RDCs. When also taking into consideration the directionalities of transcription as well as replication during the TRC, I found a 40% increase in double-strand breaks in head-on versus co-directional collisions.

In this chapter, I will first discuss the results and concepts used for the conclusions from this study. Moreover, I will also go into the caveats of the study and explain its limitations. Lastly, I will compare the findings with published literature in the field before giving examples of future perspectives following this project.

5.1 Transcription activity is required for *Ctnna2* and *Nrxn1* RDC formation.

Recurrent DNA break clusters have been identified in various cell types such as ESCs as well as NSPCs. Additionally, RDC formation has been investigated with regards of its locus specificity and its cell-type dependence^{70-72,107,117}. However, since it is difficult to compare different cell types that have different transcription, replication and DNA repair programs, the goal was to establish cell lines that are distinguished only in their levels of transcription. By using genome-edited murine

ESC-derived NPCs with an ablation of either *Cttna2* or the long isoform of *Nrxn1* and then performing LAM-HTGTS after 96 hours of aphidicolin treatment, I found that the cluster of recurrent DNA breaks at the gene-of-interest is no longer being formed. In contrast, the parental cells that continue to transcribe the full gene-of-interest also form the previously identified break cluster. The two targeted long genes, *Cttna2* and *Nrxn1*, share characteristics with the common fragile sites that have been primarily observed in large transcription units, as described by Wilson *et al.*¹¹⁸. In this study, the authors used murine ESCs as well as human fibroblasts, which were treated with either hydroxyurea (HU) or aphidicolin to induce replication stress. Wilson *et al.* found that primarily very large transcription units are affected, which are replicated late and have CNV hotspots in their transcribed and flanking regions¹¹⁸.

5.2 LAM-HTGTS and RDC detection.

The technique applied to detect the recurrent DNA break clusters is the same that has been used to identify RDCs in NSPCs in the first place^{71,72}: LAM-HTGTS exploits the formation of translocations that can be sequenced as a readout of junctions, which consist of repaired DSBs. They are merged from induced “bait” breaks due to Cas9 cutting at a pre-defined location and aphidicolin-induced “prey” DSBs from stalled replication forks at unknown locations throughout the genome^{119,120}. Nonetheless, physical proximity (*cis*-translocation, on the same chromosome, or beneficial 3D genome conformation) remains necessary to allow detection of the translocations¹²¹.

This might explain that a difference in break density was detected when comparing *Cttna2* and *Nrxn1* to their control RDCs on the same chromosomes. Using the chromosome 6 bait break as a viewpoint, *Cttna2* is located towards the telomere of the chromosome (on the right), while the two control regions *Ccser1* and *Grid2* are in a different orientation to the bait cut, towards the centromere of the chromosome (on the left). This could cause that when translocating, the bait break and *Cttna2* translocations would preferentially form a functional chromosome with a long-range deletion but with a centromere and a telomere, while for *Ccser1* and

Grid2, the translocation could also lead to the loss of the centromere, which destabilizes the structure and could cause the loss of chromosome 6 altogether.

Another issue that needs to be taken into consideration is the distance of the control RDCs. While *Grid2* (chr6:63-64 Mb) has the same distance (approximately 6-7 Mb) from the CRISPR-induced bait break (chr6:70 Mb) as *Ctnna2* (chr6:76-77Mb), this does not translate to the control loci on chromosome 17: the bait break on chromosome 17 (chr17:41 Mb) is further away (approximately 50 Mb) from *Nrxn1* (chr17:90-91 Mb), but still allows detection of RDCs at that location. Additionally, the control RDC *Ptpm* (chr17:66-67 Mb) is situated between the bait break and *Nrxn1*, however, it still shows high variability in junction numbers, regardless of its similar orientation to *Nrxn1* and closer proximity to the bait.

This hypothesis can also explain the high variability in resulting junction values for the control loci related to the gene-of-interest and translates directly to the “Junctions per thousand per Megabase” unit that I used. This unit is required for comparing different RDC-genes with different gene lengths with each other; however, it might still come with potential issues such as the gene orientation dilemma touched upon before. Due to the genotype of the cells used in this study, it might be less likely, but not impossible that some breaks are repaired in a way that is no longer detectable with my approach. Another limitation of the assay is the alignment to the reference genome; repetitive regions cannot be aligned appropriately, and information could be lost in these areas. Nonetheless, the method allows studying of break formation at a nucleotide resolution, which makes it possible to extract maximum information from the experimental data generated as seen in the investigation of RDCs being single-ended DSBs.

5.3 Replication timing changes.

E/L-Repli-seq to assess possible replication timing changes due to the deletion of the promoter-proximal regulatory elements has been performed in untreated cells (without APH). While the experimental cells for *Nrxn1* do not show a significant change in their replication timing, the *Ctnna2* cell lines do show a deviation. However, it is known that some genomic regions change their timing only when actively undergoing replication stress, which is the case for APH-treated cells¹²².

Therefore, it would have been more informative to compare the vehicle-treated experimental cell lines based on their response to aphidicolin-treatment.

In a preliminary experiment, I attempted this comparison combined with a high-resolution Repli-seq approach¹²³ that utilizes 16 instead of 2 fractions (data not shown). Nevertheless, this challenging protocol, which includes ethanol fixation with subsequent fluorescence-activated cell sorting (FACS) based on DNA content, could not have been optimized for neural progenitor cells in time to include in this study. The main problem was that the differentiation protocol which generates ESC-derived NPCs did not support cell growth for cells outside the neural progenitor cell pool. Thus, all cells that did not undergo successful induction eventually died. As a consequence, a small proportion of these dead cells remained in the cell culture dish until they were harvested after BrdU incorporation and caused a clumpy and sticky cell suspension after ethanol fixation. Due to the high number of cells (80,000) required for each fraction to be able to generate Repli-seq libraries, a concentrated cell suspension has been sorted, which consequently made the FACS procedure more elaborate. While this issue was already troublesome for DMSO-treated cells during the optimization, the APH-treated cells posed an even more complex challenge in handling since their 96-hour APH treatment led to massive cell death. A shorter period of 72 hours of aphidicolin treatment was also tested, although, it did not help to increase the success of the sorting procedure.

5.4 Head-on collision in this study.

The replication fork directionality information used in this project and shown in figures related to head-on versus co-directional transcription/replication collision is based on a novel convolutional neural network model developed by Sergej Andrejev. Information from Okazaki fragment sequencing (OK-seq) of ESCs¹²⁴ has been extracted to predict initiation zones in NPCs together with information from high-resolution Repli-seq of NPCs¹²³ to predict termination zones of the replication, respectively. The transcription directionality information from my own GRO-seq results was validated with public information from the RefSeq transcript annotations¹²⁵ and is in line with these data. Additionally, the Repli-seq data shown

in the same context are results from NPC Repli-seq by Zhao *et al.*¹²³, which was used for predicting the replication fork directions.

While the data used to create the network is of high quality, one major limitation of these predictions is that different cell types and/or genetic backgrounds have been utilized in Petryk *et al.*¹²⁴ and Zhao *et al.*¹²³. Nevertheless, the sophisticated protocols of both, OK-seq¹²⁶ as well as the previously discussed high-resolution Repli-seq, would have taken more time to optimize before allowing me to generate my own data with the correct genetic background in the experimental cell lines.

5.5 CRISPR off-target during genome editing.

One pitfall of CRISPR/Cas9 genome editing is the possibility of off-target cuts during the generation of the experimental cell lines. While the used sgRNAs have been closely evaluated for sequence homology throughout the mouse reference genome (mm10) before usage, it is still possible that unwanted cuts are happening *in vitro* due to the differences in the genetic background of the mouse strain from which the parental cells have been harvested. The cell lines used in this project were checked and validated for the expected on-site edit, in order to ensure that they have the correct anticipated deletions as well as exhibit only short resections, which might also affect the cells' behavior.

I was able to observe that one of the *Ctnna2* promoter and enhancer deleted clones (namely: *Ctnna2* ape/ Δ 1) was growing slower (data not shown) than the parental cells, founder cells or other experimental cell line clones. This is particularly interesting since interference at a tumor suppressor gene, which *Ctnna2* is reported to be¹²⁷, should lead to a growth advantage instead. The second experimental cell line clone, *Ctnna2* ape/ Δ 2, did not display an altered growth behavior in comparison with its parental and founder cells, and this was also the case for the *Nrxn1*-edited cell lines. Furthermore, none of the used cell lines carry insertions at the target gene and all of the detected resections, which are a result of deletions, are in the range of ≤ 50 bp. Altogether, structural variants (SVs) have not been observed at the target site in the cells after genome editing.

To compensate for this clonal variety in behavior, two experimental clones were generated and used for all the experiments from the start to rule out that the observed results are only due to statistical variation instead of actual biological differences. One possibility to assess off-target sites of the applied sgRNAs would have been to perform LAM-HTGTS using the parental cell lines and the sgRNAs designed for genome editing to introduce bait breaks. In this way, it would have been possible to see where and if the genome was also edited at other locations in addition to the projected on-target edit¹¹⁹. However, since I was able to generate as well as validate, the anticipated experimental cell lines by means of PCR, Sanger sequencing and TaqMan-qPCR in the first round of experiments without impediment, this step has been omitted.

5.6 Differentiation into neural progenitor cells.

The usage of ESC-induced NPCs instead of isolating a pool of fresh neural stem and progenitor cells (NSPCs) from frontal brains of postnatal mice has been already validated and published by Tena *et al.* as a suitable model for studying RDC formation¹¹⁷. The induction protocol used for the aforementioned study as well as for this study is based on Ying *et al.*¹²⁸. The advantages of using a differentiation approach are that after the isolation of the murine embryonic stem cells, there is no further need to isolate cells for any of the experiments. This is especially beneficial due to the genome editing steps required to generate the experimental cell lines since the generation would consume additional time if implemented into the cell lines by crossing of mice strains instead of direct genome editing using CRISPR/Cas9.

A 2D cell culture system as used in my project instead of the highly complex brain structures *in vivo* leads to an artificial cell pool, which might contain artifacts due to the absence of the cells' canonical environment such as stroma cells and cerebrospinal fluid, the lack of morphogen signaling, the lower complexity of the cell type pool, as well as cellular stress. Furthermore, the protocol published by Ying *et al.* refers to the resulting cell population solely as neuroectodermal precursors, which indicates that no specific class of cells is generated, but instead these cells could be even less determined towards the neural progenitor cell fate.

However, this approach also allowed minimization of batch effect by not introducing any inter-mouse variation; all experimental cell lines that are compared within their experimental group are derived from the same parental cells. Additionally, using ESC-induced NPCs eliminated the limitation in cell number for experiments. Scaling up the differentiation by using a higher embryonic stem cell number at the start of the protocol lead to a higher output of NPCs available for use in the experiments. Moreover, the possibility to use hiPSCs derived from patients with neuropsychiatric disorders for the same type of experiments allows assessment of their individual DSB hotspots *in vitro* as shown in Wang *et al.*, to gain deeper understanding into disease mechanisms⁷⁰.

5.7 Deficiency of p53 and XRCC4.

In an *in vivo* study, it was shown that the deficiency of the non-homologous end joining (NHEJ) double strand break (DSB) repair mechanism by inactivation of XRCC4 in p53-deficient mice leads to the development of early-onset medulloblastoma. XRCC4 deficiency without the additional deactivation of TP53 causes embryonic lethality due high levels of genomic instability⁸⁹. From these data, I concluded that my cells would need to be both XRCC4- as well as p53-deficient in order to be able to enrich for translocations that are later detected as RDCs. Cell lines with a wild-type genetic background could in theory also be used, however, without enrichment and this artificial push towards the alternative end-joining repair pathway, the recurrent DNA breaks would be less frequent. Therefore, a deeper sequencing depth would be necessary to be able to observe the same results, translating to a higher cost of the experiments for the same information.

5.8 Cell cycle dependence of replication.

On the other hand, this genetic background causes one major disadvantage in the study; the cells are not synchronizable due to their p53 deficiency. Normally, with p53 active, it would be possible to arrest cells between G1 and S phase or between G2 and mitosis¹²⁹. This would have allowed to uncouple transcription and replication to observe at which point during the cell cycle, the RDCs are formed. Similar studies as well as therapy regimens in the clinic are widely using CDK4/6

inhibitors in order to suppress the G1 to S cell cycle transition^{130,131}. Furthermore, what remains to be investigated is how the cells would respond to the arrest in terms of DNA repair. A prolonged stay at the G1/S checkpoint could increase the involvement of homologous recombination in the repair of the prey double-strand breaks¹³², which consequently could decrease the number of translocations that are detectable using LAM-HTGTS.

5.9 Mechanisms of Common Fragile Site Formation versus Recurrent DNA Break Cluster Formation.

The origins of CFSs are still to be discovered. Some CFSs have AT-rich sequences, while no such common sequence signature has been found for RDCs¹³³. The mechanism of CFS formation proposed by Wilson *et al.*¹¹⁸ is based on transcription-dependent double-fork failure. The main features of this current working hypothesis are based on large gene units with active transcription. The hypothesis assumes two replication forks that are approaching each other while transcription is delaying origin firing, resulting in unreplicated DNA. Additionally, CFSs and CNVs hotspots seem to colocalize in the same cell-type specific loci. The consequence of the double-fork failure also explains the formation of CNVs. The observed deletions could be justified by skipping of the unreplicated DNA area, while amplification of a genomic area could be explained by the restart of replication at a location downstream of the fork collapse. Altogether, large actively-expressed genes are predicted to be CFSs and CNV hotspots in replicating cells¹¹⁸.

This is also in line not only with my study of the RDC-genes *Ctnna2* and *Nrxn1*, but also recent transcription studies such as work by Park *et al.*¹³⁴ on the CFS FRA3B, which is situated on the *FHIT* gene. In this study, the common fragile site FRA3B on chromosome 3 was targeted via CRISPR/Cas9 to knock down transcription of the *FHIT* gene in the mouse-human hybrid cell line GM11713, which harbors one copy of the human chromosome 3. In response, the fragility at the same locus has been reduced while other non-targeted control CFS remained fragile¹³⁴. This confirms that transcription seems to be a necessary factor to be able to observe fragility at CFSs as well as RDCs; nonetheless, they also came to the conclusion that transcription does not exclusively induce fragility.

5.10 The effect of transcription on the genome.

The link between copy number variations (CNVs) and transcription has been the subject of multiple research projects. For example, a study by Hull *et al.*¹³⁵ observed the copper-resistance gene *CUP1* in budding yeast. They were able to demonstrate that formation of CNVs was stimulated by increased levels of environmental copper acting on an inducible promoter. They concluded that promoter activity together with H3K56ac levels are key elements in the *CUP1* CNV formation in budding yeast¹³⁵. H3K56ac plays an essential role in transcription initiation and supports elongation and/or termination, however, it was also observed that H3K56ac represses transcription directly after replication fork passage to ensure accurate nucleosome assembly¹³⁶.

In the previously discussed study by Wilson *et al.*¹¹⁸, CFSs and CNVs often colocalized, presenting as two different outcomes of the same genome instability events. They found that CNVs are often found in areas with large active transcription units that are cell-type-specific and often late replicating. Transcription of large genes continues into the S-phase and therefore, prevents late replication origin firing. As a result, unreplicated DNA forms breaks that are presenting as CFSs and/or deletions and amplifications that lead to CNVs in the region around the large transcription unit¹¹⁸.

It remains to be clarified whether transcription interferes with replication timing, and if such an effect would be of direct or indirect nature. This was touched upon in the previous paragraph when discussing that transcription in S-phase prevents late origin firing¹¹⁸. This is due to eviction of pre-replication complexes by the movement of the RNA polymerase on its template^{137,138}. Changes in the replication timing indeed are observable in response to 3D genome organization as well as transcription. Therefore, one can conclude that transcription and replication timing are affecting each other; however, no direct relationship could be observed, which suggests that they have an indirect association. Rivera-Mulia *et al.*¹³⁹ used genome-wide replication timing (RT) profiles of 15 human cell types as well as intermediate stages from differentiated human embryonic stem cells to construct regulatory networks for replication timing. These RT networks were then compared to matched cell-type transcriptional regulatory networks, and it was noticeable that

they overlapped with each other. Following up, a composite network was created, and these networks were used to further investigate regulators of the replication timing. Finally, they found, that transcription factors indeed act on the replication timing of genes in *trans* in a cell-type-specific manner.

Nonetheless, this computational study leaves some questions unanswered. Specifically, whether the transcription factors can regulate the replication timing independently of their role in transcription and whether all transcription factors are able to alter the replication timing remains the topic of future studies. Another possibility could be the binding of multiple transcription factors, together with activation of superenhancers, might be required to remodel the replication dynamics¹³⁹. In a study by Sarni *et al.*¹⁴⁰, expression levels of genes were investigated in response to APH treatment. Aphidicolin did not show any effect on transcriptional activity; however, expressed large fragile genes with an average replication timing of early S were delayed by APH. This observation was only true for expressed large genes, while silent genes were not affected by aphidicolin treatment. Additionally, the delay of expressed large genes was further increased when the genes were situated at a TAD boundary, making them more vulnerable in replication stress conditions. Therefore, high transcription, replication timing delay and location in inter-TAD regions are properties of fragile genes¹⁴⁰.

A study by Liu *et al.*¹⁴¹ proposed a “transcription bulldozing” model in which RNA polymerase II redistributes MCM complexes to non-transcribed regions of the genome to prevent early DNA replication at actively transcribed genes. The authors were also able to demonstrate increased levels of genome instability upon induction of transcription/replication-collision through the restart of transcription in regions with dysregulated DNA replication initiation¹⁴¹.

Another topic connected to the effects of transcription on the genome is the role of torsional stress. Transcription can have various influences on the DNA since it is necessary to unwind complementary DNA strands. For example, at the upstream promoter region of all transcribed genes, negative DNA supercoiling can in turn be observed, while nucleosomes also are moved up- and downstream by the RNA polymerase which can release or redistribute torsional stress in the form of

negative supercoils on the chromatin. Additionally, torsional stress might be a prerequisite to reorganize the chromatin in a way where enhancer and promoter move into close proximity of each other¹⁴². Under normal conditions, topoisomerases I and II are responsible to alleviate the buildup of torsional stress by the introduction of DNA breaks, which, if deregulated, could be a direct source of DNA damage. Teves *et al.*¹⁴³ showed that in cells with inhibited topoisomerases the torsional strain accumulates, which leads to the stalling of RNA polymerase II directly behind the TSSs as well as to destabilization of the nucleosomes¹⁴³. If torsional stress due to transcription is not alleviated, it can impact not only transcription itself, but also induce the formation of DNA secondary structures that could in turn lead to genome instability¹⁴⁴.

Andrade-Lima *et al.*¹⁴⁵ investigated how DNA repair and recovery behaves in short versus long genes. In general, long genes have a higher probability to accumulate DNA damage and experience inhibition of transcription elongation than shorter genes, with the probability being proportional to the gene length. Transcription elongation is inhibited in this case because the RNA polymerase II needs to be either removed or backed up in order to allow access to the repair factors¹⁴⁵. In a study by Mabb *et al.*¹⁴⁶, the transcription of long genes was specifically examined: treatment with topotecan, a topoisomerase 1 inhibitor, was found to decrease the expression levels of long neural genes (>100 kb). Similar to the treatment, the conditional deletion of *Top1* in cortical neuron cultures also lead to a reduced gene expression as well as reduced levels of synaptic adhesion proteins, such as neurexin 1 transcribed from the RDC-gene *Nrxn1*. Top1 normally forms an intermediate with DNA, which is known as the Top1 cleavage complex (Top1cc). In their study, they found a total of 500 long genes with an average length of 318 kb to be downregulated following Top1 disruption. Therefore, the regulation of gene expression in neurons seems to be Top1cc-dependent¹⁴⁶.

In the context of genome instability, transcription-associated mutagenesis is not to be underestimated, with up to 80% of the human genome being transcriptionally active. By opening the chromatin for transcription, transient regions of ssDNA are created which are not only chemically more reactive, but also more susceptible to

damage than dsDNA¹⁴⁷. All in all, comparative genome analyses were able to link active transcription to mutagenesis^{52,148}.

5.11 Epigenetic changes and how they translate to replication timing.

Different cell types have individual characteristics, such as the differences in the epigenetic landscape across the genome between different cell types. Transcription can also be heavily influenced by changes of the epigenetic landscape. If the chromatin is in a heterochromatin conformation or the gene's promoter is not in proximity of its enhancers, the gene might not be transcriptionally active. Therefore, an important factor that has not been taken into consideration for this study were the epigenetic differences between embryonic stem cells and neural progenitor cells.

Two major studies in the field connect the epigenetic state of the cells to replication timing changes. The first study by Sarni *et al.*¹⁴⁰ discovered that delayed replication timing and transcription are not sufficient for the expression of common fragile sites. Moreover, delayed replication timing, active transcription, and the location of the gene at a TAD boundary are essential parts of the CFS signature. Therefore, inter-TAD location is a third prerequisite that needs to be fulfilled in order for a gene to be a CFS¹⁴⁰. This hypothesis could be easily tested for RDCs by performing Hi-C, allowing detection of the situation of RDC-genes in the 3D genome and revealing whether they are also mainly located at TAD boundaries. Nonetheless, as mentioned above, 3D conformation and TAD boundaries might fluctuate largely from cell type to cell type, particularly during differentiation, which would make differentiation timepoints necessary in my experimental setting in order to draw any valid conclusions.

The second study by Klein *et al.*¹⁴⁹ proposed the knockout of Replication Timing Regulatory Factor 1 (RIF1) which led to a heterogeneous association of replication initiation factors with both active and repressive chromatin. This study showed that RIF1 controls replication timing by reducing variation, and that these alterations lead to global changes of chromatin compartments and epigenetic state of the cells. The effect on the epigenome requires replication and occurs already during

the first S phase after degradation of RIF1, but is further exacerbated as the cells cycle without replication timing control¹⁴⁹.

Additionally, a study by Courtot *et al.*¹⁵⁰ investigated the effect of low replicative stress on replication timing and found that advances are actually inherited by the next generation of cells based on enhanced chromatin accessibility. Modifications in replication timing, replication origin firing as well as gene expression were all induced long-term by altered chromatin organization in response to a low-dose (0.2 μ M) aphidicolin treatment in experiments using the colon cancer cell line RKO¹⁵⁰.

Chromosomal rearrangements in cancer were studied in the context of replication timing and the epigenome in a study performed by Du *et al.*¹⁵¹. They demonstrated that areas of the cancer cell genome which undergo long-range epigenetic deregulation also exhibit modifications in the replication timing. Specifically, the late-replicating areas of prostate and breast cancer cell lines showed reduced levels of DNA methylation. The loci which were observed to switch their replication timing showed an increased probability of chromosomal rearrangements. When comparing the replication timing profiles, the different cancer cell types clustered together in a principal component analysis, making them more similar to each other than to non-cancer cells and thus indicating that these changes in replication timing might be a common feature of tumorigenesis¹⁵¹. Additionally, a study investigating the mutational signatures of great apes also identified an increased level of mutations in regions of the late-replicating genome¹⁵².

5.12 Future perspectives.

As stated before, it would be valuable to study the behavior of the investigated cell lines also in terms of their epigenetic makeup. The focus of such follow-up experiments should not lay only on cells at the ESC or NPC states, but also in between during their differentiation, to see at which point the epigenetic landscape might change in such a way that transcription is turned on and the replication profile is altered, likely due to conformational changes in the chromatin. The observed effect might also be only a passenger of another event due to modifications in the 3D genome conformation.

Additionally to the closer investigation of the role of epigenetics in RDC formation, other questions also remain to be answered. First of all, it would be of great interest to unravel how RDCs are formed in the first place. Since RDCs are associated with stalled and collapsed replication forks, the formation mechanism is thought to be involved in the resolvment of the DNA structure by two essential proteins, MUS81 and RECQ5. MUS81 is a endonuclease which resolves intermediates during DNA repair after inter-strand cross-linking, replication fork collapse, or DSBs¹²⁵. On the other hand, RECQ5 is a helicase that displaces RAD51 from ssDNA to facilitate the activity of MUS81¹⁵³. When acting together in the context of CFS locations or mitotic DNA synthesis (MiDAS)¹⁵⁴, fork cleavage and accurate chromosome segregation lead to breaks and gaps in the chromosomes. However, when RECQ5 is inactive, the RAD51 filament is not removed which subsequently leads to cleavage inhibition and consequently causes gross chromosome instability in the form of micronuclei¹⁵⁵. Different approaches are currently being used to inhibit and knock down MUS81 in order to observe if RDCs are still able to form in response to aphidicolin treatment. These experiments should clarify the overlap between RDC and MiDAS to test the hypothesis whether RDCs are a precursor of MiDAS, which then could act as a repair pathway for these fragile locations in the genome.

Another question remaining open is whether the formation of these RDCs could be used to induce synthetic lethality in cells prior to them developing a disease phenotype. Here, I performed preliminary experiments (data not shown) in which cells were treated with aphidicolin and additionally with inhibitors of DNA damage repair pathways (*i.e.* Wee1 inhibitor, ATM inhibitor and ATR inhibitor). While

treatment with aphidicolin and a Wee1 inhibitor, as expected, leads to a replication catastrophe due to an increase in DSBs, not only at RDCs, the cells responded differently to the ATM and ATR inhibition. In general, ATR responds to breaks originating from stalled replication forks whereas ATM responds to DSBs. ATR inhibition together with APH treatment has been observed to be more lethal to my cells than ATM + APH treatment, which is in line with my previous results showing that RDC are formed after replication fork collapse. Likely, the ATM inhibition did not have any effect on the cells since ATM is situated upstream of p53 in the signaling pathway and the cells used in my experiments are p53-deficient.

Looking at the artificial model that has been applied in this work, one should also consider how RDCs could be generated *in vivo*. Likely, the cells will not experience replication stress through aphidicolin treatment; however, lack of nucleotides after rapid proliferation, oncogene activation, tumor suppressor gene inactivation and/or reactive oxygen species could all lead to replication fork stalling and subsequently, trigger tumorigenesis or cell death. Consequently, another project in the lab is studying if the formation of RDCs can be triggered by rapid proliferation (based on: Dai *et al.*, Nature Protocols, 2020.¹⁵⁶), as well as the effects of replication stress in the developing mouse brain. This model allows the investigation of the behavior of cells in their natural niche, while the *in vitro* studies available so far were conducted in isolated neural stem and progenitor cells as well as ESC-induced neural progenitor cells.

Nonetheless, it was shown that some of these fragile RDC-genes are also fragile in cancer patients¹⁵⁷ and in people with neurodevelopmental or neuropsychiatric disorders, as they are showing copy number variations such as deletions, inversions, translocations, or amplifications⁷². These genome instabilities could be possible results of RDC formation; however, it remains unclear if these rearrangements are driver or passenger events. The current working hypothesis describes that these observed CNVs are consequently inducing altered gene expression^{158,159} of alternative transcripts¹⁶⁰ and thus possibly lead to translation of alternative proteins. Such genomic alterations are expected to generate heterogeneity in the neuronal cells, which could cause brain mosaicism¹⁶¹. The adult human brain is made up of clones that can be traced back to 50-100 founder

progenitors, supporting that this genetic mechanism is indeed important for typical brain function¹⁶². While most somatic variants are benign, deletions in sensitive regions such as tumor suppressor genes and proteins involved in neural function, for the latter especially during neurodevelopment, are able to cause brain pathology (reviewed in: Bizzotto *et al.*, Nature Reviews. Neuroscience, 2022.¹⁶¹). For the two RDC-genes studied, *Ctnna2* and *Nrxn1*, both generate intragenic deletions in the human ortholog in patients. These kinds of rearrangements cause various phenotypes such as mild mental retardation and speech delay in a reported case of a *CTNNA2* intragenic deletion at intron 7¹⁶³ or in the case of deletions in *NRXN1* with 3,540 individuals analyzed: association with autism spectrum disorders, mental retardation and language delays¹⁶⁴.

6 Conclusion

Taken together, this study was able to assess whether transcription is a licensing factor for the formation of recurrent DNA break clusters. By deleting single promoters of RDC-genes, I was able to compare RDC formation in the same cell type. The break density at the clusters was significantly decreased once transcription had been abolished or shortened. Additionally, I found no changes in replication timing as a direct response to altered transcription and its subsequent changes in the break clusters. In the second part, the hypothesis of transcription/replication collision driving development of RDCs was investigated. The detected DNA double-strand breaks stem from collapsed replication forks; increasing levels of fork slowing led to an increase in breaks while also exhibiting a shift in the junction peak pattern. Fork stalling in response to aphidicolin treatment was investigated and allowed me to identify recurrent DNA break clusters as single-ended DSBs. The link between break formation and collision directionality was also clarified by taking transcription and replication directionalities into consideration. This made it possible to find a 40% increase in double-strand breaks in head-on versus co-directional collisions.

The obtained data is in line with the presented hypothesis and settles the ground for further investigation in the context of genome instability in neuropsychiatric disorders as well as cancer. This research has the potential to not only unlock new treatment regimens aimed at addressing genomic instability in cancer, but can also serve as a foundational framework for further basic research. This additional exploration could shed light on the formation of CNVs in RDC-genes before the development of cancer or neurological diseases.

7 Materials and Methods

7.1 Materials

7.1.1 Cell culture

7.1.1.1 Cell growth media

Mouse Embryonic Fibroblast medium (=MEF medium) [DMEM, high glucose; 10 % heat inactivated FBS, 20mM HEPES, 1 X NEAA, 1 X Pen/Strep, 0.1mM β -ME]

450 mL	DMEM, high glucose (4.5 g/L); containing 2 mM L-glutamine and sodium pyruvate (Gibco 11965092)
50 mL	Heat inactivated fetal bovine serum (Gibco 10500064)
10 mL	1M HEPES (Gibco 15630056)
6 mL	100 X Non-Essential Amino Acids Solution (Gibco 11140035)
6 mL	100 X Penicillin-Streptomycin-Glutamine (Gibco 10378016)
4 μ L	β -Mercaptoethanol (stock: 14.2 M) (Sigma-Aldrich M6250-250ML)

The different solutions were all combined in a clean 500 mL glass bottle, filtered through a 0.22 μ M filter to sterilize the cell growth medium. Medium was stored at 4°C and remained good for up to 4 weeks.

Embryonic Stem Cell medium (=ESC medium) [DMEM, high glucose; 15 % ES-grade FBS, 20mM HEPES, 1 X NEAA, 1 X Pen/Strep, 0.1mM β -ME, supplied with 2×10^4 U LIF/mL]

425 mL	KnockOut DMEM, high glucose (4.5 g/L); containing sodium pyruvate (Gibco 10829018)
75 mL	ES grade, Heat inactivated fetal bovine serum (Gibco 16141079)
10 mL	1M HEPES (Gibco 15630056)
6 mL	100 X Non-Essential Amino Acids Solution (Gibco 11140035)
6 mL	100 X Penicillin-Streptomycin-Glutamine (Gibco 10378016)
4 μ L	β -Mercaptoethanol (stock: 14.2 M) (Sigma-Aldrich M6250-250ML)
25 μ L	LIF (10^7 U/mL) (Gibco A35934)

The different solutions were all combined in a clean 500 mL glass bottle, filtered through a 0.22 µM filter to sterilize the cell growth medium. Medium was stored at 4°C and remained good for up to 2 weeks.

N2B27 medium [DMEM/F12: NeuroBasal = 1:1; 1% B27 supplement without retinyl acetate; 0.5% modified N2 supplement, 1X GlutaMax]

50 mL	NeuroBasal medium (Gibco 21103049)
50 mL	DMEM/F12 medium (Gibco 11320033)
1 mL	B27 minus vitamin A (Gibco 12587010)
0.5 mL	Modified N2 (home-made)
100 µL	GlutaMax (Gibco 35050061)

The different solutions were all combined in a clean 500 mL glass bottle, filtered through a 0.22 µM filter to sterilize the cell growth medium. Medium was stored at 4°C and remained good for up to 2 weeks.

NBBG medium [NeuroBasal A; 2% B27 supplement w/o retinyl acetate; GlutaMax, 1X Gentamicin, 1X EGF/FGF/PDGF]

100 mL	Neurobasal A (Gibco 10888022)
2 mL	B27 minus vitamin A (Gibco 12587010)
250 µL	GlutaMax (Gibco 35050061)
100 µL	Gentamicin, 10mg/mL (Gibco 15710064)
30 µL	Growth Factor mixture (1X EGF/FGF/PDGF) (home-made)

The different solutions were all combined in a clean 500 mL glass bottle, filtered through a 0.22 µM filter to sterilize the cell growth medium. Medium was stored at 4°C and remained good for 1 week.

7.1.1.2 Other reagents

2X Freezing medium [ES-grade FBS with 20% DMSO]

8mL ES grade, Heat inactivated fetal bovine serum (Gibco 16141079)

2mL Dimethyl sulfoxide (DMSO) (ACROS organics 414885000)

Both solutions were combined in a 15 mL tube and transferred to a syringe with a 0.22 μ M filter. To sterilize the 2X freezing medium the liquid was pushed through the filter into a fresh sterile 15 mL tube and stored at 4°C for up to 2 weeks.

1X Dulbecco's Phosphate Buffered Saline Solution (DPBS)

50 mL of 10X DPBS without calcium and magnesium (Gibco 14200075) were diluted in 450 mL of deionized and filtered water. 1X DPBS was autoclaved before use and stored at room temperature indefinitely.

0.2 % gelatin in 1X DPBS

1 g of gelatin from porcine skin (Sigma-Aldrich G1890-500G) has been weighed in and 500 mL of 1X DPBS were added. The gelatin solution was autoclaved before use and stored at room temperature indefinitely.

Poly-L-Ornithine (PLO) solution

100 μ L of PLO (Sigma-Aldrich P4957-50ML) were diluted (1:100) in 9.9 mL of 1X DPBS. The solution was then ready to use.

Laminin solution

100 μ L of laminin (Sigma-Aldrich L2020-1MG) were diluted (1:100) in 9.9 mL of DMEM/F12 (Gibco 11320033). The solution was then ready to use.

100 mg/mL human apo-transferrin solution

1 mL of sterilized water were added to 100 mg of human apo-transferrin (R&D Systems 3188-AT-100MG). After a 10 minutes incubation at room temperature, the solution was resuspended by pipetting up and down with a P1000 pipet. The solution was then ready to use.

1 mg/mL progesterone solution

To prepare a stock solution of 10 mg/mL, 10mg of progesterone (Sigma-Aldrich P8783-1G) were solved in 1 mL of 100% ethanol and mixed by inverting the tube. 100 µL of the 10 mg/mL stock solution were further diluted with 900 µL of 100% ethanol to prepare the working solution of 1mg/mL progesterone. The solution was then ready to use.

1 M (=160 mg/mL) putrescine dihydrochloride solution

160 mg of putrescine dihydrochloride (Sigma-Aldrich P5780-5G) were weighted and solved in 1 mL of sterilized water. The solution was mixed by inverting the tube and was then ready to use.

3 mM sodium selenite solution

5.18 mg of sodium selenite (Sigma-Aldrich S5261-10G) were solved in 1 mL of sterilized water and mixed by inverting the tube to prepare a 30 mM stock solution. To achieve a 3 mM working solution 100 µL of the 30 mL stock were further diluted with 900 µL of sterilized water. The solution was then ready to use.

Modified N2 supplement

2.5 mL	human insulin solution (Sigma-Aldrich I9278-5ML)
1 mL	100 mg/mL human apo-transferrin
6 µL	1 mg/mL progesterone
100 µL	1.6 mg/mL putrescine dihydrochloride
10 µL	3 mM sodium selenite
667 µL	7.5% Bovine Albumin Fraction V solution (Sigma-Aldrich A8412-100ML)
5.717 mL	DMEM/F12 (Gibco 11320033)

All the above solutions were combined to make a 10 mL modified N2 supplement solution. 500 µL aliquots were prepared and stored at -20°C for up to 6 months.

10% (w/v) Bovine Serum Albumin

1 g of BSA (Sigma-Aldrich A2153-100G) were solved in 10 mL of deionized water and 1 mL aliquots were stored in cryovials at -80°C.

0.1% BSA/100mM acetic acid

288 μ L of glacial acetic acid (=17.4 M) (Fisher Chemical A/0400/PB15) and 500 μ L of 10% BSA were added to 50 mL of deionized water. The solution was sterilized through a 0.22 μ M sterile filter and 1 mL aliquots were stored in cryovials at -80°C.

0.1% BSA/PBS

100 μ L of 10% BSA were diluted in 10 mL of 1X DPBS, the solution was sterilized through a 0.22 μ M sterile filter and 1 mL aliquots were stored in cryovials at -80°C.

Growth Factor mixture

250 μ L of 1X DPBS were added to 25 μ g of recombinant human EGF (Gibco PHG0314) to prepare a 100 μ g/mL EGF stock solution. 250 μ L of 0.1% BSA/100 mM acetic acid were added to 25 μ g of recombinant mouse PDGF-BB (Gibco PMG0044) to prepare a 100 μ g/mL PDGF-BB stock solution. 250 μ L of sterile water were added to 25 μ g of recombinant mouse FGFb (Gibco PMG0034) to prepare a 100 μ g/mL FGFb stock solution. To prepare the 100X growth factor working solution, 10 μ L of each of the previously prepared stocks (EGF, PDGF-BB and FGFb) were combined and 970 μ L of 0.1% BSA/PBS were added to a total volume of 1000 μ L. Aliquots of 30 μ L each were prepared and stored at -20°C for up to 6 months.

2X Trypsin [0.5% Trypsin without phenol red]

10 mL of trypsin (Gibco 15090046) were diluted (1:5) in 40 mL of 1X DPBS. The solution was stored at 4°C and remained good for up to 2 months.

Accutase

Accutase (Invitrogen 00-4555-56) was thawed, aliquoted into tubes with 10 mL each and then stored at -20°C indefinitely. Solution is ready to use after thawing and remained good at 4°C for up to 2 months.

Dimethyl sulfoxide (DMSO) treatment

10 μ L of DMSO (ACROS organics 414885000) were diluted in 5.9 mL of NBBG medium and transferred to a syringe with a 0.22 μ M filter. To sterilize the DMSO

treatment, the liquid was filtered into a fresh sterile 15 mL tube and used immediately.

Aphidicolin (APH) treatment

10 μ L of APH (Sigma-Aldrich A4487) were diluted in 5.9 mL of NBBG medium and transferred to a syringe with a 0.22 μ M filter. To sterilize the APH treatment, the liquid was filtered into a fresh sterile 15 mL tube wrapped with an aluminum foil as a protection from light and used immediately.

PK cell lysis buffer [10 mM Tris-HCl (pH 8), 200 mM NaCl, 1 mM EDTA, 0.5% SDS]

0.5 mL	1 M Tris-HCl (pH=8.0) (Sigma-Aldrich T1503-5KG)
2 mL	5 M Sodium chloride (Fisher Chemicals S/3160/65)
0.1 mL	0.5 M EDTA (pH=8.0) (Sigma-Aldrich EDS-500G)
2.5 mL	10% SDS (Sigma-Aldrich 75746-1KG)

The solutions above were combined and filled up to a total volume of 50 mL with deionized water. The lysis buffer was stored at room temperature indefinitely. Prior to usage, Proteinase K (20 mg/mL, 1:50) was added to a target concentration of 0.4 mg/mL in the final buffer before use, and the buffer then had to be used immediately.

SB1 buffer

5.47 g	Sucrose (Sigma-Aldrich S7903-1KG)
150 μ L	1 M CaCl ₂ (Sigma-Aldrich C8106-1KG)
100 μ L	1 M MgAc ₂ (Sigma-Aldrich M5661-250G)
10 μ L	0.5 M EDTA (Sigma-Aldrich EDS-500G)
500 μ L	1 M Tris-HCl (pH 8) (Sigma-Aldrich T1503-5KG)
2.5 mL	10% IGEPAL CA-630 (Santa Cruz sc-281108)
50 μ L	1 M DTT (Thermo Scientific R0861)
2 tablets	1X EDTA-free protease inhibitors (Thermo Scientific A32965)

The sucrose and the solutions above were combined and filled up with deionized water to a final volume of 50 mL. Solution can be used immediately, and leftover can be stored at -20°C for one additional run after thawing.

SB2 buffer

34.23 g	Sucrose (Sigma-Aldrich S7903-1KG)
250 μ L	1 M MgCl ₂ (Sigma-Aldrich M8266-1KG)
10 μ L	0.5 M EDTA (Sigma-Aldrich EDS-500G)
500 μ L	1 M Tris-HCl (pH 8) (Sigma-Aldrich T1503-5KG)
50 μ L	1 M DTT (Thermo Scientific R0861)

The sucrose and the solutions above were combined and filled up with deionized water to a final volume of 50mL. The solution can be used immediately, and leftover solution can be stored at -20°C for one additional run after thawing.

Glycerol storage buffer

2.5 mL	1 M Tris-HCl (pH 8.3) (Sigma-Aldrich T1503-5KG)
20 mL	Glycerol (Roth 3783.1)
250 μ L	1 M MgCl ₂ (Sigma-Aldrich M8266-1KG)
10 μ L	0.5 M EDTA (Sigma-Aldrich EDS-500G)
27.24 mL	Deionized water

The solutions above were combined and the glycerol storage buffer was stored at 4°C for up to 1 year.

BrdU stock solution

BrdU (Roche 10280879001) was diluted to 10 mg/mL in deionized water and warmed up to 37°C in a water bath to ensure complete dissolving. 1 mL aliquots can be stored at -20°C and used for up to 2 years if protected from light.

PBS/1% (vol/vol) FBS

0.5 mL of heat inactivated fetal bovine serum (Gibco 10500064) was diluted in 49.5 mL of 1X DPBS. The solution was stored at 4°C and used for up to 2 months.

7.1.2 Molecular biology lab work

7.1.2.1 Cloning

2X LB broth, low salt

5 g	Sodium chloride (Fisher Chemicals S/3160/65)
10 g	Tryptone (Gerbu 16540500)
5 g	Yeast Extract (Gerbu 11330500)

All components were weighted in and dissolved in 500 mL of water. The solution was autoclaved prior to use and stored at 4°C afterwards for up to 2 months.

LB agar plates with Ampicillin

7.5 g	Agar (Roth 400405000)
500 mL	2X LB broth, low salt
500 µL	1,000X Ampicillin stock (50mg/mL) (Roth K0294)

The agar powder was mixed with an un-autoclaved 2X LB broth solution. Afterwards the solution was autoclaved and cooled down to approximately 56°C. The ampicillin was added and mixed into the liquid LB agar prior to pouring it into 10cm plates (around 15mL per plate). Once the agar plates solidified, they could be packed in a plastic bag and stored at 4°C upside down for up to 2 months.

sgRNAs

Nrxn1 cell lines:

sgRNA names	Sequences (5' to 3')
Nrxn1-h1	TGAGGATAAAATTCAGATGG
Nrxn1-T1	CAATTCATCACTGCTCGAGA
Nrxn1PrUp1	AAGCAACTTGGAATCTTCGA
Nrxn1PrDn1	CAATGGAGTCAGCTTGACAT

Ctnna2 cell lines:

sgRNA names	Sequences (5' to 3')
Ctnna2_ds1	ACTTAGATGTGATGTAGTGG
Ctnna2_up1	ATTCACTACAGAGATATCCA
Ctnna2_pe_ds1	TGAGGAGGCAGAAGGTACAG
Ctnna2_pe_up1	CAAGGAAAGCTTGGGAACTG

Bait nucleofection for LAM-HTGTS:

Bait name	sgRNA bait sequence (5' to 3')
Chr6_70Mb	GAGGGTTGGAAATGCTCGTA
Chr17_41Mb	TCCAGGGGTTCCCTAGGTGTA

7.1.2.2 PCR screenings

Tail digestion buffer stock [50 mM KCl, 10 mM Tris-HCL (pH 9), 0.1% Triton X-100]

2.5 mL	1 M Potassium chloride (Roth 6781.1)
0.5 mL	1 M Tris-HCl (pH 9.0) (Sigma-Aldrich T1503-5KG)
0.5 mL	10% Triton X-100 (Sigma-Aldrich X100-500ML)

The solutions were combined and filled up to 50 mL with sterilized deionized water. Buffer can be stored at 4°C for up to 6 months and is ready to use after adding proteinase K (20 mg/mL) in a dilution of 1:50 for a final concentration of 0.4 mg/mL.

10X TBE buffer

108 g	Tris (Sigma-Aldrich T1503-5KG)
55 g	Boric acid (Roth 6943.2)
9.3 g	0.5 M EDTA (pH 8.0) (Sigma-Aldrich EDS-500G)

All chemicals were weighted in together, dissolved in 900 mL of deionized water and the volume was adjusted to 1 L. The solution can be stored at room temperature for up to 6 months.

Reagents

- *Taq* polymerase (Invitrogen 10342053)
- 10X PCR buffer, without magnesium (Invitrogen 10342053)
- 50 mM Magnesium chloride (Invitrogen 10342053)
- 10 mM dNTP (Thermo Scientific R0181; combined and diluted 1:10)
- Agarose (Roth 3810.3)
- GeneRuler 1 kb DNA Ladder (Thermo Scientific SM0311)

Primers

Names	Sequences (5' to 3')
Nrxn1 H1F	TGCTAGCCCCTTTTGTGGT
Nrxn1_wld_R	CCTGCTTGCTTAGCTGTTCC
Nrxn1PrUp1_F	CCTGCTAAATGGCAGACCTC
Nrxn1PrDn1_R	CTGAAGCTTGCAAGGTGAAATG
Ctnna2_wld_F	AGGAAGCAATGGGCACCTTA
Ctnna2_wld_R	GAGAGGGATGGAGTGGTCAG
Ctnna2_wld_seq_F	TCATGCAAGTTCAATGTTCCA
Ctnna2_wld_seq_R	TGCCTGATTGTGCCTTAAATC
Ctnna2_pe_F	GGCATTACCGAGAAGCAGAG
Ctnna2_pe_R	TGAGGAGGCAGAAGGTACAG

7.1.2.3 Sanger sequencing

- QIAquick Gel Extraction Kit (Qiagen 28706)
- LightRun Tube Barcodes (Eurofins Genomics)

7.1.2.4 TaqMan-qPCR

Reagents

- TaqMan Universal PCR Master Mix (Applied Biosystems 4304437)

TaqMan probes

Names	Sequences (5' to 3')
Nrxn1_wld_probe	FAM-ACAAGTGGGAAGGCAAGGAGGAAC-QSY
Nrxn1_P_E_probe	ABY-TTTATGGCACTGCATTGTGGCAGC-QSY

Ctnna2_wld_probe	FAM-AGGGCATGTGCTGATGTCATGGT-QSY
Ctnna2_P_E_probe	VIC-TCCCATAACTCTTACAGAGAGACAGGG-QSY
TmoIMR0105	VIC-CCAATGGTCCGGCACTGCTCAA-QSY
Csmd1_wld_probe	ABY-ACTGGGTGTGAAGTTTGGAGCAGA-QSY

Primers

Names	Sequences (5' to 3')
Nrxn1_wld_F	CCACTTGGGAGAGAGAAGAAAG
Nrxn1_wld_R	TCCTCTACTTGCCTACCTACTT
Nrxn1_pe_F	GACACAGCTCACTAGCCAATTA
Nrxn1_pe_R	GATTCAGTTCACCAGAGAG
Ctnna2_wld_F	GACTGAGAACAGGCATAGAAGG
Ctnna2_wld_R	CAATGCTGTCTGATCCCAAATG
Ctnna2_pe_F	GCTACACAAGTATGAACACTTCC
Ctnna2_pe_R	GAGACAGGATTAACCATTGTGATAC
oIMR1544	CACGTGGGCTCCAGCATT
oIMR3580	TCACCAGTCATTTCTGCCTTTG
Csmd1_wld_F	GACCAGCTCCTCTGTTTGTT
Csmd1_wld_R	AGAACCACTCTCAGCTGTTTAC

7.1.2.5 Genomic DNA isolation

- Ethanol absolute ≥99.8% (VWR 20821330)
- Phenol/Chloroform/Isoamyl alcohol, 25:24:1, pH 7.5-8.0 (Roth A156.1)

7.1.2.6 LAM-HTGTS

2X Bind & Wash (B&W) Buffer [2M NaCl, 10mM Tris-HCl (pH 7.4), 1mM EDTA]

20 mL	5 M Sodium chloride
0.5 mL	1 M Tris-HCl (pH 7.4)
1 mL	0.5 M EDTA

The solutions were combined and filled up to 50 mL with sterilized deionized water. The buffer can be stored at room temperature for up to 1 year.

Biotinylated Primers:

Primer name	Sequences (5' to 3')
Bio-chr6	/5Biosg/CTTGA ACTCATCCTGGGATGCTAAG
Bio-chr17	/5Biosg/CATGGTTTATTCCATCTGCCTAT

RED (=Barcoding) Primers:

Primer name	Primer Sequence (5' to 3') (MID-Seq in bold)
I5+Chr6_70Mb	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M1V11	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T A TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M2V11	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T AT TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M2V12	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T TG TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M3V11	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T AGT TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M3V12	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T TCA TTC GGC TTC TGT TCT CTG AAA GAG C
I5_Chr6_70Mb _M1PW1	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T T TTC GGC TTC TGT TCT CTG AAA GAG C
I5_Chr6_70Mb _M2PW1	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T TA TTC GGC TTC TGT TCT CTG AAA GAG C

I5_Chr6_70Mb _M3PW1	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T GCG TTC GGC TTC TGT TCT CTG AAA GAG C
I5_Chr6_70Mb _M4PW1	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T TAGT TTC GGC TTC TGT TCT CTG AAA GAG C
I5_Chr6_70Mb _M4PW2	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T CATG TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M5LC1	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T CGATT TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M5LC2	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T CTGAC TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M5LC3	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T CATGC TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M5LC4	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T GTCAA TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M5LC5	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T ATCGT TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M5LC6	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T ACTGC TTC GGC TTC TGT TCT CTG AAA GAG C
I5+Chr6_70Mb _M5LC7	AC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T TCGAA TTC GGC TTC TGT TCT CTG AAA GAG C

Other primers used:

Primer name	Sequences (5' to 3')
P7-I7	CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT C

P5-I5	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T
I7-AP2	CTC GGC ATT CCT GCT GAA CCG CTC TTC CGA TCT ACT ATA GGG CAC GCG TGG T
ssA-upper-TN	GCG ACT ATA GGG CAC GCG TGG TNN NNN /3AmMO/
ssA-lower-A	/5Phos/ A CCA CGC GTG CCC TAT AGT CGC /3AmMO/
ssA-lower-A1	/5Phos/ CCA CGC GTG CCC TAT AGT CGC /3AmMO/

7.1.2.7 qPCR primers for the *Ctnna2* locus

Name	Sequences (5' to 3')	Exons
Ctnna2_17_1_1_F	GAGGAGGAGGCGAGAACTC	EX1-2
Ctnna2_17_1_1_R	CCTGGCACTGGAGCTATGAG	EX1-2
Ctnna2_13_1_2_F	ATGTAGCAGCAAGACGGCAG	EX6-7
Ctnna2_13_1_2_R	TGTACAGCATGGTGGCATT	EX6-7
Ctnna2_12_1_2_F	TGCAGCCCTGAATGAGTTTG	EX7-8
Ctnna2_12_1_2_R	GTCTCTCCTCCAGGGATGGT	EX7-8
Ctnna2_10_1_1_F	TGCGATTGACAAGATGACCA	EX9-10
Ctnna2_10_1_1_R	AGGGGACATTGGTTTCCAAG	EX9-10
Ctnna2_6_1_2_F	CAGAGAAAGTGCTGGAAGCC	EX13-14
Ctnna2_6_1_2_R	TCAACTTGTTTCAGCAAAGCG	EX13-14

7.1.2.8 GRO-seq

- SUPERase RNase Inhibitor (20 U/μL) (Invitrogen AM2696)
- TRIzol (Invitrogen 15596026)
- GlycoBlue (Invitrogen AM9516)
- anti-BrdU antibodies (Santa Cruz SC-32323 AC)
- RNA 5' Pyrophosphohydrolase (=RppH) (NEB M0356S)
- 10X ThermoPol Reaction Buffer (NEB B9004S)
- T4 Polynucleotide Kinase (=T4 PNK) (NEB M0201S)
- T4 RNA ligase 1 (NEB M0204S)
- Reverse transcriptase (SSIII) (Invitrogen 18080051)

- Phusion High-Fidelity DNA polymerase (Thermo Scientific F530L)
- Micro Bio-Spin P-30 Gel Columns, Tris Buffer (RNase-free) (Bio-Rad 7326250)
- QIAquick Gel Extraction Kit (Qiagen 28706)
- AMPure XP beads (Beckman Coulter A63881)

Primer and adaptor sequences:

from TruSeq Small RNA Library Prep Kit -Set A (Illumina RS2000012)

Primer name	Sequence (5' to 3')
5' RNA adapter (CG361)	GUUCAGAGUUCUACAGUCCGACGAUC
3' RNA adapter (CG362)	/Phospho/UGGAAUUCUCGGGUGCCAAGG
RT primer (CG363)	GCCTTGGCACCCGAGAATTCCA
RNA PCR primer (RP1); CG364	AATGATACGGCGACCACCGAGATCTACACGTTTCAGA GTTCTACAGTCCGA
RNA PCR primer, Index 1 (RPI1); CG365	CAAGCAGAAGACGGCATAACGAGAT <u>CGTGAT</u> GTGAC TGGAGTTCCTTGGCACCCGAGAATTCCA
RNA PCR primer, Index 2 (RPI2); CG366	CAAGCAGAAGACGGCATAACGAGAT <u>ACATCGG</u> TGAC TGGAGTTCCTTGGCACCCGAGAATTCCA
RNA PCR primer, Index 3 (RPI3); CG367	CAAGCAGAAGACGGCATAACGAGAT <u>GCCTAAG</u> TGAC TGGAGTTCCTTGGCACCCGAGAATTCCA
RNA PCR primer, Index 4 (RPI4); CG368	CAAGCAGAAGACGGCATAACGAGAT <u>TGGTCAG</u> TGAC TGGAGTTCCTTGGCACCCGAGAATTCCA
RNA PCR primer, Index 6 (RPI-6, reverse complementary)	CAAGCAGAAGACGGCATAACGAGAT <u>ATTGGC</u> GTGAC TGGAGTTCCTTGGCACCCGAGAATTCCA
RNA PCR primer, Index 12 (RPI-12, reverse complementary)	CAAGCAGAAGACGGCATAACGAGAT <u>TACAAGG</u> TGAC TGGAGTTCCTTGGCACCCGAGAATTCCA

7.1.2.9 Repli-seq

- Zymo Quick-DNA Microprep Kit (Zymo D3021)
- Zymo DNA Clean & Concentrator-5 (Zymo D4014)
- NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7645L)
- anti-BrdU antibody (Santa Cruz SC-32323)
- Protein G Magnetic Beads (Thermo Scientific 88847)
- NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) (NEB E7600S)
- AMPure XP beads (Beckman Coulter A63881)

7.1.2.10 Quality Control of sequencing libraries

- Qubit 4 Fluorometer (Invitrogen Q33226)
- Qubit dsDNA HS Reagent (Invitrogen Q32854)
- Qubit dsDNA HS Buffer (Invitrogen Q32854)
- Qubit Assay tubes (Invitrogen Q32856)

- D1000 Ladder (Agilent 50675586)
- D1000 Sample Buffer (Agilent 50675602)
- D1000 ScreenTape (Agilent 50675582)

- Blue S'Green qPCR Kit (Biozym 331416S)

Primers

Name	Sequence (5' to 3')
Kapa Q1 (P5)	AATGATACGGCGACCACCGAGATCT
Kapa Q2 (P7)	CAAGCAGAAGACGGCATACTGAGAT

7.2 Methods

7.2.1 Cell culture

All experiments were conducted using murine embryonic stem cell (ESC) lines that are *p53*- and *XRCC4*-deficient. The two parental cell lines for the experiments were NXP010 and NXP047. Additionally, irradiated mouse embryonic fibroblasts (MEF) were used as feeder cells to maintain pluripotency.

7.2.1.1 Pre-gelatinized plates

Cell culture plates were incubated for 5 minutes at 37°C with 5% CO₂ and 92% humidity with a 0.2% gelatin in 1X DPBS solution to prepare the surface. 2 mL were used to cover the surface of each well in a 6-well plate, while for 6 cm dishes 5 mL and for 10 cm dishes 10 mL were used.

7.2.1.2 MEF feeder cell culture

A cryotube of irradiated mouse embryonic fibroblasts (MEF) was thawed for 2 minutes at 37°C in the water bath. After thawing, the cell suspension was immediately diluted with 9 mL of MEF medium prior to centrifugation (300 x g for 5 minutes). The supernatant was discarded, and the pellet was resuspended in 12 mL of fresh MEF medium. The cell suspension was then transferred to a pre-gelatinized 6-well plate and 2 mL of cell suspension were distributed to each well. The plate was placed in the incubator at 37°C with 5% CO₂ and 92% relative humidity. The irradiated MEF were used in the next five days as a feeder cell layer to ensure the quality of the embryonic stem cell culture.

7.2.1.3 ESC culture

The embryonic stem cells were thawed mostly similarly to the MEF above. However, after thawing, the suspension was diluted with 9 mL of ESC medium instead and centrifuged for 5 minutes at 300 x g. After centrifugation the supernatant was discarded, and the cell pellet was resuspended in 3 mL of fresh ESC medium. The MEF medium of the prepared feeder cell well was then removed and 3 mL of ESC suspension were seeded on top of the MEF layer. The plate was then placed back into the incubator at 37°C with 5% CO₂ and 92% relative humidity. After 24 hours, the ESC medium was changed once and after 36 to 48 hours, when

ESCs reached approximately 50% confluency, the cells were normally split in a ratio between 1:3 to 1:8.

7.2.1.4 ESC nucleofection

For the generation of the experimental cell lines, ESCs first were depleted of MEFs. To this extent, ESCs were detached with 2X trypsin and reseeded onto a pre-gelatinized plate. In the meantime, the two solutions of the Mouse Embryonic Stem Cell Nucleofector Kit (Lonza VPH-1001) were mixed. After 30 minutes in the incubator, the cell suspension was aspirated and counted, a total of 2 million ESCs were subsequently used for each nucleofection reaction. The cells were centrifuged at only 90 x g for 5 minutes at room temperature. After the centrifugation, the supernatant was removed, first by vacuum aspiration followed by careful aspiration with a P200 pipet. The cell pellet was resuspended in 100µL of the premixed nucleofection solution and transferred to a tube containing 2µg of each pX330 plasmid used per reaction as well as 0.5µg of pMAX-GFP plasmid as a spike-in control. The nucleofection reaction was then transferred to the nucleofector cuvette and put into the Nucleofector 2b device (Lonza). The “ES cells, mouse (A-023)” program was selected and after the quick transfection, the cells were immediately transferred to 10mL of fresh ESC medium to dilute the nucleofection solution. The cell concentration was first diluted to 200,000 cells/mL and then further diluted to 20,000 cells/mL. 500 µL (=1,000 cells) or 700 µL (=1,400 cells) were seeded to prepared 6 cm MEF plates. The plates were placed back into the incubator at 37°C with 5% CO₂ and 92% relative humidity. After the nucleofection, the ESCs grew slower, therefore cells were normally picked 3 to 5 days post-nucleofection.

7.2.1.5 Colony picking

Four 24-well MEF plates were prepared for ESC colony picking to allow the assessment of 96 colonies. Prior to the colony picking, the MEF medium in the 24-well plates was changed with 800µL of fresh ESC medium per well. In the meantime, also a 96-well plate was prepared containing 45 µL of 1X DPBS per well. For picking colonies, a P20 pipet was adjusted to a volume of 15 µL. With the help of a microscope, a colony with round and shiny appearance was selected and

then aspirated using the P20 pipet. 24 colonies were picked at a time and then trypsinized for 5 minutes at 37°C in 60 µL of 4X trypsin. 120 µL of ESC medium were then added and the cell suspension was mixed prior to the transfer to the 24-well plates. The plates were placed back into the incubator at 37°C with 5% CO₂ and 92% relative humidity. After the colony picking, the colonies were growing out after 3 to 4 days. On the 4th day, the colonies needed to be split.

7.2.1.6 Colony splitting

To split the colonies, the medium in the 24-well plates was discarded by aspiration. After a wash with 1X DPBS, the cells were trypsinized by adding 120 µL of 2X trypsin and incubating for 8 minutes at 37°C. The reaction was stopped by adding 300 µL of fresh prewarmed ESC medium and the suspension was resuspended. The plates were then put back for 30 minutes into the incubator to allow MEF depletion. During the MEF depletion, the medium of fresh 24-well MEF plates was changed from MEF medium to ESC medium; furthermore, a 96-well flat bottom plate was pregelatinized. After 30 minutes, the cell suspension was mixed again and 300µL were transferred to the 96-well plate, while 80 µL were transferred to the fresh 24-well plates. The plates were placed back into the incubator at 37°C with 5% CO₂ and 92% relative humidity. The 96-well plate was lysed for genotyping on the next day, while the 24-well plates were cultured for up to 3 days prior to splitting.

7.2.1.7 PLO and laminin-coated plates

Two days before the induction to NPCs, plates were coated with PLO and laminin. PLO was diluted as stated above at 1:100 in 1X DPBS and 1.5 mL of the solution were added to each well of a 6-well plate. The plate was sealed with parafilm and stored in the fridge at 4°C. After an overnight incubation at 4°C, the solution was aspirated, and each well was washed twice with 1X DPBS. In the meantime, a 1:100 dilution of laminin in DMEM/F12 was prepared (see above) and 1.5 mL were added to each well. The plate was again sealed with parafilm and incubated overnight at 4°C. On the next day, the plates are ready for NPC induction and could be stored for another two weeks if kept at 4°C. Prior to cell seeding the coated plates were prewarmed to at least 30 minutes. The same coating was also applied

for the maintenance of cells as NPCs, however, the volumes used were increased to 4 mL per plate since 6 cm plates were used instead of 6-well plates.

7.2.1.8 ESC-NPC induction

After two splits, ESCs were ready for induction. Therefore, ESCs were MEF-depleted on a pre-gelatinized 6 cm plate before cell counting. For each three wells of NPC induction, 0.3×10^6 ES cells were aliquoted and centrifuged at 1,200 rpm for 5 minutes. The supernatant was aspirated, and the cell pellet was resuspended in 6 mL of N2B27 medium. 2mL of cell suspension were seeded per well of PLO/laminin-coated plates. The plates were placed back into the incubator at 37°C with 5% CO₂ and 92% relative humidity. The N2B27 medium was exchanged every day for seven days. From day one to day three, 2 mL were used and starting from the fourth day, the volume was increased to 4 mL.

7.2.1.9 NPC maintenance

After 7 days of NPC induction, cells were split and transferred onto bigger plates. Therefore, the cells were detached with Accutase prior to centrifugation at 300 x g for 5 minutes. The supernatant was discarded, and cells were resuspended in 16 mL of NBBG medium. Each three wells of induced cells were then split to four PLO/laminin-coated 6cm plates. The plates were placed back into the incubator with 37°C with 5% CO₂ and 92% relative humidity. The NBBG medium was changed every second day or if the medium turned yellow. After four days of NPC maintenance, the cells were ready to be used in experiments.

7.2.1.10 NPC nucleofection for LAM-HTGTS

NPCs were detached with Accutase. In the meantime, the two solutions of the Mouse Neural Stem Cell Nucleofector Kit (Lonza VPG-1004) were mixed. The cell suspension was aspirated, cells were counted, and a total of 5 million NPCs were used for each nucleofection reaction. The cells were centrifuged at only 80 x g for 5 minutes at room temperature. After the centrifugation, the supernatant was first aspirated by vacuum, followed by careful aspiration with a P200 pipet. The cell pellet was resuspended in 100 µL of the premixed nucleofection solution and transferred to a tube containing 5 µg of the pX330 plasmid containing the bait-

sgRNA as well as 0.5 µg of pMAX-GFP plasmid as a spike-in control. The nucleofection reaction was then transferred to the nucleofector cuvette and put into the Nucleofector 2b device (Lonza). The “NSC, mouse (A-033)” program was selected and after the short transfection, the cells were immediately transferred to prewarmed PLO/laminin-coated plates containing 5 mL of fresh NBBG medium in order to dilute the nucleofection solution. The plates were placed back into the incubator at 37°C with 5% CO₂ and 92% relative humidity. After all reactions were nucleofected, the DMSO- and APH-treatments were prepared as stated before and 500 µL were added to cells accordingly (APH: 0.5 µM for 72 hours, 0.25 µM for 24 hours). The treatment was refreshed after 48 hours and diluted by half at 72 hours. 96 hours post-nucleofection, the cells were collected for the isolation of genomic DNA.

7.2.1.11 Cell harvest

NPCs were harvested 96 hours after nucleofection. For harvesting, the medium was aspirated, and cells were treated with Accutase. After a 10-minute incubation at 37°C, Neurobasal A medium (Gibco 10888022) was added to the cells and resuspended. The cells were collected in 15mL tubes and centrifuged at 200 x g for 5 minutes at room temperature. After aspirating the supernatant, the cell pellet was resuspended in 1 mL of 1X DPBS and transferred to a 1.5mL tube. The cell suspension was centrifuged again at 200 x g for 5 minutes. In the meantime, the PK cell lysis buffer was prepared. After centrifugation, the 1X DPBS was discarded, the tubes were flicked for mixing, and 300 µL of cell lysis buffer were added directly to the pellet. After a brief resuspension, the tubes were securely closed and incubated overnight at 56°C for PK digestion prior to gDNA isolation for LAM-HTGTS.

7.2.1.12 Nuclei isolation for Global Run On (GRO)-seq

Cells were harvested as stated above using Accutase but centrifuged at 4°C and kept on ice after collection. The cell pellet was gently resuspended in 20 mL of ice-cold 1X DPBS prior to another centrifugation step. After aspiration, the cell pellet was loosened by gentle vortexing and resuspended in 4 mL of ice-cold SB1 buffer. The suspension was then transferred to a Kontes Dounce Tissue Grinder with a

type B pestle and the cell walls were broken through homogenization by 14 up-and-down strokes. Subsequently, 4 mL of SB2 buffer were added and the solutions were mixed by pipetting and inversion. In the meantime, ultracentrifuge tubes (Beckman 344059) were filled with 4 mL of SB2 buffer. The nuclei suspension was then slowly pipetted on top of the SB2 layer and ultracentrifuge buckets were balanced with SB1 buffer. To separate the nuclei from the other cell components, the suspension was centrifuged for 45 minutes at 30,000xg (=15,600 rpm, SW 41Ti rotor; maximum acceleration and slow braking) at 4°C. The supernatant was removed by aspiration and the nuclei were resuspended in 100 µL ice-cold glycerol storage buffer. Nuclei were counted and aliquoted in cryovials with a total of 5-10 million nuclei per vial for each GRO-seq run. The nuclei can be stored for up to 2 weeks at -80°C or up to 2 years in liquid nitrogen.

7.2.1.13 Cell fixation for Repli-seq

NPCs were seeded onto PLO/laminin-coated 6 cm plates and cultured for 72 hours. 2 hours before cell harvest, BrdU was added to the medium to a final concentration of 100 µM and cells were further incubated at 37°C. Cells were then gently rinsed for 5 times with 5 mL of ice-cold 1X DPBS and detached using Accutase. After a 10-minute incubation, the cells were transferred to a 15 mL tube and centrifuged at 200 x g for 5 minutes at room temperature. After the supernatant was aspirated, the cell pellet was resuspended in 2.5 mL of ice-cold 1X DPBS with 1% of FBS. At this step, cells could be counted to check the final cell number. Afterwards, detached cells were fixed by adding 7.5 mL of ice-cold ethanol to the cell suspension. The tubes were then sealed with Parafilm, mixed and inverted before storing the fixed cells at -20°C protected from light. Fixed cells can be stored for more than a year and then thawed for subsequent sorting.

7.2.1.14 RNA harvest

ESCs and NPCs grown in a monolayer were lysed directly in the 6 cm culture dish with 1 mL of TRIzol (Invitrogen 15596026) by pipetting up and down and transferring the cell suspension to a reaction tube. This cell suspension was either frozen at -80°C or immediately processed to extract the RNA (see 7.2.2.7).

7.2.2 Molecular biology lab work

7.2.2.1 CRISPR/Cas9 cloning of the pX330 plasmid

sgRNAs have been designed after selecting the area of interest with the sgRNA Designer: CRISPRko by the Broad Institute. Here the mouse GRCm38 reference genome and the SpyoCas9 enzyme have been selected. The target sequence was entered and the best sgRNA candidate was selected from the generated sgRNAs by validating the sequence for uniqueness via BLAT Search (provided by UCSC). The reverse complement was also generated to retrieve the paired sequence which can be cloned into the vector.

For cloning, it was necessary to add overhangs to the sgRNA oligos to enable generation of the correct plasmid. I added, "CACCG" at the 5' end of the upper sequence, while the lower sequence had "AAAC" added at its 5' end and a "C" at its 3' end. The paired oligos were resuspended in sterilized deionized water to a concentration of 100 μ M. 1 μ L of each oligo was combined with 1 μ L of 10 mM ATP, 1 μ L of 10X T4 PNK Buffer (NEB M0201S), 5.5 μ L of sterilized deionized water as well as 0.5 μ L of T4 PNK (NEB M0201S) to a total volume of 10 μ L. The phosphorylation and annealing were performed in a thermocycler at 30 minutes at 37°C and 5 minutes at 95°C before cooling down to 25°C at a ramp rate of 5°C/minute. After annealing the oligos were further diluted at 1:250 for the ligation with the plasmid. If not used immediately, the annealed oligos were stored at -20°C.

Additionally, the plasmid backbone had to be digested; therefore, 1 μ g of the pX330 plasmid (Addgene 42230) was combined with 1 μ L of Bpil restriction enzyme (Thermo Scientific ER1011), 1 μ L of FastAP (Thermo Scientific EF0651), 2 μ L of 10X FastDigest Buffer (Thermo Scientific B64) and 15 μ L of sterilized deionized water. The mixture was incubated at 37°C overnight to ensure complete digestion. On the next day, the plasmid mixture was purified on a 0.7% agarose gel and the plasmid was extracted using the QIAquick Gel Extraction Kit (Qiagen 28706) according to the manufacturer's manual.

For the ligation, 20-50 ng of the Bpil-digested pX330 plasmid were combined with 1 μ L of the annealed oligos (diluted 1:250), 5 μ L of 2X rapid ligation buffer (Promega C6711) and 3 μ L of sterilized deionized water. 1 μ L of T4 DNA ligase (Promega M1801) was added last and the suspension was mixed. After 10 minutes at room temperature, half of the ligation mix was used for direct transformation while the other half was transferred to 4°C overnight.

DH5 α competent bacteria (Invitrogen 18265017) were transformed using a heat shock protocol: A vial of competent bacteria was thawed on ice for 10 minutes. Afterwards, 50 μ L of competent bacteria were carefully added into a tube with 5 μ L of the ligation mixture and incubated for 5 minutes on ice. Then, the tube was placed in a water bath at 42°C for 35 seconds and immediately put back on ice for another 5 minutes. 500 μ L of 2X LB broth were added to the bacteria mixture and incubated at 37°C for 20 minutes. After a centrifugation at 5,000 rpm for 10 minutes, the supernatant was discarded, and the bacteria pellet was resuspended in 100 μ L of fresh 2X LB broth. The bacteria were then plated out on an agar plate containing ampicillin and incubated at 37°C overnight. On the next day, the agar plate was removed from the incubator, sealed with parafilm and stored upside down at 4°C.

To prepare a liquid bacteria culture, 4 mL of 2X LB broth were aliquoted into a glass reaction tube with a loose cap to allow air flow. The agar plate with bacteria was retrieved from 4°C and a colony was punctured with a 200 μ L pipet tip. This pipet tip was then transferred to the glass reaction tube and incubated overnight at 37°C.

On the next day, the plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen 27106) according to the manufacturer's manual.

7.2.2.2 Quick DNA isolation for PCR screenings

To harvest DNA for the colony PCR screen, culture medium has been discarded from the 96-well plate by aspiration. The cells were washed twice with 200 μ L of 1X DPBS. The lysis buffer has been prepared by mixing 16.5 mL of tail digestion

buffer with 330 μL of proteinase K stock solution (20 mg/mL). 150 μL of lysis buffer were added to each well before incubating the plate at room temperature for 30 minutes. After checking for successful cell lysis under the microscope, the cell lysate was slowly resuspended and transferred to a sealable 96-well PCR plate. The lysate was incubated at 56°C for 12 hours, 95°C for 10 minutes and then the temperature was held at 10°C. Samples are then ready for PCR and can be stored at 4°C for a week or indefinitely at -20°C.

7.2.2.3 PCR screening

The isolated DNA from 7.2.2.2 was used for screening the colonies for positive clones after the CRISPR/Cas9 genome editing.

PCR reaction (1X)

2.5 μL	10X PCR buffer, without magnesium (Invitrogen 10342053)
0.4 μL	10 mM dNTP (Thermo Scientific R0181)
0.75 μL	50 mM Magnesium chloride (Invitrogen 10342053)
0.5 μL	10 μM primer mix (Forward plus reverse)
0.2 μL	Taq polymerase (Invitrogen 10342053)
2.0 μL	DNA/sample
19.65 μL	Sterilized deionized water

PCR program

1. 95°C 3 minutes
2. 95°C 30 seconds
3. 58/60/62°C 30 seconds
4. 72°C 45 seconds
5. Go to step 2, repeat 34 times.
6. 72°C 4 minutes
7. 4°C Hold
8. End

The PCR products were run on an 1.2% TBE agarose gel at 160 V for 45 minutes. GeneRuler 1 kb DNA Ladder (Thermo Scientific SM0311) was added as a reference. Bands were visualized under an ultraviolet (UV) light transilluminator.

7.2.2.4 Sanger sequencing

For Sanger sequencing, the positive bands were cut out of the agarose gel after PCR screening. These gel pieces were weighted, and the PCR products were extracted with the QIAquick Gel Extraction Kit (Qiagen 28706) according to the manufacturer's manual. In the next step, the DNA concentration was assessed with a Nanodrop One Microvolume UV-Vis Spectrophotometer (Thermo Scientific). To prepare the sequencing reaction, the DNA samples were diluted to a concentration of 5 ng/ μ L according to the requirements by Eurofins Genomics. Two reactions were premixed per sample (using 5 μ L of DNA), one with the forward PCR primer used and one with the reverse PCR primer used (each primer: 5 μ L at a concentration of 5 pmol/ μ L). The result sequences were evaluated with SnapGene Viewer 6.0.4.

7.2.2.5 Allele density assessment via TaqMan-qPCR

Frozen DNA samples from 7.2.2.2, primers and TaqMan probes were thawed on ice. After the samples thawed, the DNA concentration was quantified and diluted to 10 ng/ μ L. The TaqMan Universal PCR Master Mix (Applied Biosystems 4304437) was used to prepare the following PCR reaction:

TaqMan-qPCR reaction (1X)

1 μ L	10 μ M Internal Control Primer (Forward)
1 μ L	10 μ M Internal Control Primer (Reverse)
0.5 μ L	10 μ M Internal Control Probe
1 μ L	10 μ M Whole Locus Deletion Primer (Forward + Reverse)
0.5 μ L	10 μ M Whole Locus Deletion Probe
1 μ L	10 μ M Promoter/Enhancer Del. Primer (Forward + Reverse)
0.5 μ L	10 μ M Promoter/Enhancer Del. Probe
10 μ L	TaqMan Universal PCR Master Mix (2X)
2.5 μ L	Sterilized deionized water

2 μ L Genomic DNA sample (10 ng/ μ L)

The 20 μ L reactions for each sample were run as triplicates in the QuantStudio 3 Real-Time PCR System. A comparative CT ($\Delta\Delta$ CT) experiment was set up with the parental cell and the internal control set as a reference. In the analyzed results, the detection of two alleles leads to a “Relative Quantification” value of 1. Therefore, to assess the number of present loci in the experimental cell lines, the values were multiplied by 2 to represent the allele number.

PCR program:

1. 50°C 2 minutes
2. 95°C 10 minutes
3. 95°C 15 seconds
4. 60°C 1 minute
5. Data collection
6. Go back to step 3, repeat 39 times.
7. End

7.2.2.6 Genomic DNA isolation

Harvested cells were lysed overnight at 56°C in cell lysis buffer with proteinase K. After taking the samples out of the oven, 700 μ L of ice-cold 100% ethanol were added to the lysate to precipitate the DNA. The DNA was pelleted down by centrifugation at 21,000 x g for 30 seconds. After two washes with 70% ethanol, the pelleted DNA was air dried and resuspended in 100 μ L of sterilized deionized water. The samples were incubated overnight at 4°C to let the DNA resolve. On the next day, the volume was increased with sterilized deionized water to 400 μ L prior to a phenol-chloroform-isoamyl alcohol purification with the same volume. The top phase was transferred to a new tube and washed with 400 μ L of chloroform. In the next step, the new top phase was transferred to a new tube and combined with 40 μ L of 3 M sodium acetate as well as 900 μ L of ice-cold 100% ethanol. The DNA was precipitated overnight at -20°C. Subsequently, the pellet was washed twice with 70% ethanol and air dried. 100 μ L of sterilized deionized water were added to resuspend the DNA and after a 30-minute incubation at room temperature the DNA

concentration was quantified using a Nanodrop One Microvolume UV-Vis Spectrophotometer.

7.2.2.7 RNA extraction

For further processing, the samples from 7.2.1.14 were incubated at room temperature for 5 minutes. 0.2 mL of chloroform was added, the tubes were vortexed for 15 seconds and afterward incubated for additional 2 to 3 minutes at room temperature. The mixture was separated by centrifugation at 12,000 x g for 5 minutes at 4°C and the upper aqueous phase was transferred to another tube. The upper phase was precipitated on ice for 10 minutes with 0.5 mL of isopropanol. The precipitated RNA was centrifugated at 4°C with 12,000 x g for 10 minutes and washed twice with 75% ethanol. Each time the centrifugation was performed as stated above at 4°C with 12,000 x g for 5 minutes. The excess of ethanol was removed before letting the pellet air dry. The dried pellet was resuspended in 50 µL of RNase-free water before the RNA concentration was quantified using a NanoDrop microvolume spectrophotometer.

7.2.2.8 LAM-HTGTS

For each sample 20 µg of genomic DNA from 7.2.2.6 were aliquoted into 0.65 mL microtubes and sonicated in a Bioruptor Pico sonication device (Diagenode) for 13 seconds at 2°C. Afterwards, the sonication efficiency was evaluated on a 1% TBE agarose gel by comparing it to GeneRuler 1 kb DNA Ladder as well as 1 µg of unsonicated DNA per sample as a control. The desired size range to proceed with the library preparation was 500 to 2,000 bp. Linear amplification was prepared as follows:

Linear amplification-mediated PCR (1X)

40 µL	5X Phusion HF PCR Buffer (Thermo Scientific F530L)
8 µL	2.5 mM dNTPs
4 µL	Biotinylated primer (1 µM)
2 µL	Phusion polymerase (Thermo Scientific F530L)
-	Sonicated DNA (20 µg)

All reagents were combined, and the final PCR volume was 200 μ L for 20 μ g of genomic DNA input. The PCR was divided to four PCR tubes to allow efficient temperature regulation.

LAM-PCR program:

1. 98°C 2 minutes
2. 98°C 30 seconds
3. 58°C 30 seconds
4. 72°C 90 seconds
5. Go back to step 2, repeat 79 times.
6. 72°C 6 minutes
7. 10°C Hold
8. End

Dynabeads MyOne Streptavidin C1 (Invitrogen 65001) were washed with 1X B&W buffer to prepare them for biotin pull-down. In the meantime, the LAM-PCR reaction of each sample was pooled in a 1.5 mL tube, the volume was increased to 400 μ L with sterilized deionized water and 100 μ L of 5 M sodium chloride as well as 5 μ L of 0.5 M EDTA were added. 40 μ L of Dynabeads were added to each sample and mixed on a rotating wheel for 3 hours. The supernatant was discarded upon collection of the Dynabeads on a magnetic stand and three washes with 1X B&W buffer as well as one wash with sterilized deionized water was performed. Dynabeads were resuspended in 45 μ L of sterilized deionized water before proceeding with the on-beads ligation. Therefore, 10 μ L of 10X T4 ligation buffer, 5 μ L of 50 μ M TN/A adapter, 5 μ L of 20 mM HCC, 5 μ L of T4 ligase and 30 μ L of 50% PEG8000 were added. The suspension was mixed by pipetting and then split in two PCR tubes with a volume of 50 μ L each.

On-beads ligation program:

1. 25°C 1 hour
2. 22°C 2 hours
3. 16°C Hold
4. End

The ligation mixture was pooled again and the Dynabeads were washed three times with 1X B&W buffer followed by one wash using sterilized deionized water before resuspending the beads in 200 μ L of sterilized deionized water.

Nested PCR to barcode the libraries (1X)

80 μ L	5X Phusion HF PCR buffer (Thermo Scientific F530L)
32 μ L	2.5 mM dNTPs
16 μ L	10 μ M I5-Red Barcoding primer
16 μ L	10 μ M AP2-I7
4 μ L	Phusion polymerase (Thermo Scientific F530L)
200 μ L	DNA on Dynabeads
52 μ L	Sterilized deionized water

A master mix was prepared, which contained all the reagents mentioned above except the I5-Red Barcoding primers and the DNA on Dynabeads. 400 μ L of PCR mix were aliquoted into eight PCR tubes. The 98°C step in the PCR was lowered to 95°C in order to not disrupt the Biotin-streptavidin interaction of the Dynabeads with the bound DNA.

Nested PCR program:

1. 95°C 5 minutes
2. 95°C 1 minute
3. 60°C 30 seconds
4. 72°C 1 minute
5. Go back to step 2, repeat 14 times.
6. 72°C 10 minutes
7. 10°C Hold
8. End

The Nested PCR was pooled, and the DNA was eluted from the Dynabeads using the QIAquick Gel Extraction Kit (Qiagen 28706) following the manufacturer's manual. The DNA was eluted in 60 μ L of sterilized deionized water and the DNA concentration was quantified using a Nanodrop One Microvolume UV-Vis Spectrophotometer.

To amplify the barcoded libraries and introduce the adapters necessary for sequencing, another PCR step (“Tagged PCR”) was performed.

Tagged PCR (1X)

40 μ L	5X Phusion HF PCR buffer
16 μ L	2.5 mM dNTPs
8 μ L	10 μ M P5-I5
8 μ L	10 μ M P7-I7
2 μ L	Phusion polymerase
60 μ L	DNA library
66 μ L	Sterilized deionized water

200 μ L of PCR mix were aliquoted into four PCR tubes before starting the reaction.

Tagged PCR program:

1. 98°C 3 minutes
2. 98°C 30 seconds
3. 60°C 30 seconds
4. 72°C 1 minutes
5. Go back to step 2, repeat 14 times.
6. 72°C 6 minutes
7. 10°C Hold
8. End

If the DNA concentration was higher than 10 ng/ μ L after the Nested PCR, the amplification cycles of the Tagged PCR could be reduced to 14 cycles instead of 15.

The PCR reaction were pooled, and each sample was loaded onto four lanes of a 1.5% agarose gel. As a reference, a GeneRuler 100bp DNA Ladder was loaded so that the fragments between 500 to 1,000 bp could be extracted. The DNA was extracted from the agarose gel piece using the QIAquick Gel Extraction Kit (Qiagen 28706) following the manufacturer’s manual implementing a minor change: the gel

fragment was melted at 30°C instead of 50°C. The libraries were finally eluted in 60 µL of sterilized deionized water.

Two-sided beads purification was performed using AMPure XP beads (Beckman Coulter A63881). To select for fragments with a size around 400 to 500 bp, first, a reaction with 0.56X beads was performed to remove the larger fragments from the DNA solution. After a 5-minute incubation, the beads were separated on a magnetic stand and after the solution had cleared, the supernatant was transferred to a new tube and incubated in a second step with 0.24X beads according to the original volume. Since now the desired fragment size binds to the beads, in this second step, the supernatant was discarded, the beads were washed with 80% ethanol twice, air dried and resuspended in 20 µL of sterilized deionized water. The fragments can be stored at -20°C or directly proceeded to the quality control (see 7.2.2.11) before sequencing.

7.2.2.9 *GRO-seq*

Frozen nuclei from 7.2.1.12 were thawed on ice for 30 minutes prior to the global nuclear run on. After 5 minutes at 30°C for the nuclear run on, nuclei were immediately lysed in TRIzol, RNA was extracted and purified with acid phenol/chloroform. The cleaned-up RNA was precipitated overnight in ethanol with GlycoBlue (Invitrogen AM9516) at -80°C. On the next day, the RNA was hydrolyzed with sodium hydroxide before the first BrU pulldown with agarose-coupled anti-BrdU antibodies (Santa Cruz SC-32323 AC). The RNA was purified with acid phenol/chloroform after elution and precipitated overnight in ethanol with GlycoBlue at -80°C. On the third day, end repair and decapping for the RNA were performed. After another RNA purification step, the 5'-adapters were ligated overnight. A second round of BrU pulldown was performed and RNA was purified prior to the 3'-adapter ligation overnight. The fifth day starts with the final BrU pulldown and is followed by another acid phenol/chloroform purification. The RNA was reverse transcribed to cDNA and libraries were indexed. The 200 to 500bp fragments of the indexed cDNA libraries were isolated with the QIAquick Gel Extraction Kit. A test PCR was performed to assess the amplification cycles (between 11 and 17 cycles) required for construction of the final libraries depending

on their concentration and overamplification signature on a TBE-PAGE. The final PCR was performed in a similar protocol to the indexing PCR and test PCR and PCR products were further cleaned up by size-selection.

To this extent, one-sided beads purification with AMPure XP beads (Beckman Coulter A63881) was performed using a 1.2X volume of beads to select a fragment size of around 200 bp. The desired fragments were bound to the beads, washed with 80% ethanol twice, air dried and resuspended in 20 μ L of sterilized deionized water and could be stored at -20°C or directly advanced to quality control (see 7.2.2.11) before sequencing.

7.2.2.10 *Repli-seq*

Fixed cells from 7.2.1.13 were thawed for 30 minutes at room temperature before proceeding with propidium iodide staining to be able to sort the cells according to their DNA content into early and late S phase. A total of 100,000 cells were sorted for each fraction. Subsequently, sorted cells were lysed with an SDS-PK buffer for 2 hours before isolation of DNA using the Quick-DNA Microprep Kit (Zymo D3021). Genomic DNA was sonicated using a Covaris ultrasonicator to obtain fragments with a length of around 200bp. DNA libraries were constructed with the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7645L) before the BrdU pulldown that was facilitated by the anti-BrdU antibody (Santa Cruz SC-32323) and the Protein G Magnetic Beads (Thermo Scientific 88847). In the next step, the libraries were indexed with the NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) (NEB E7600S) and amplified.

For the post-PCR clean-up, one-sided beads purification with AMPure XP beads (Beckman Coulter A63881) was performed using a 0.9X volume of beads to select a fragment size of around 300 bp. The desired fragments were bound to the beads, washed with 80% ethanol three times, air dried and resuspended in 30 μ L of 10 mM Tris-HCl and could be stored at -20°C or directly proceeded to the quality control (see 7.2.2.11) before sequencing.

7.2.2.11 *Quality control of sequencing libraries*

Qubit

To quantify the DNA concentration of the generated sequencing libraries, the Qubit 4 Fluorometer (Invitrogen Q33226) was used. The Qubit dsDNA HS Reagent (Invitrogen Q32854) was diluted 1:200 in Qubit dsDNA HS Buffer (Invitrogen Q32854). Using Qubit Assay tubes (Invitrogen Q32856), the standard solutions were diluted in a ratio of 1:20, while samples were diluted 1:100 in the Qubit working solution. After dilution, all samples were incubated at room temperature for 2 minutes before reading. The DNA concentration was given in ng/ μ L for each sample.

TapeStation

1 μ L of DNA libraries as well as D1000 Ladder (Agilent 50675586) were mixed with 3 μ L of D1000 Sample Buffer (Agilent 50675602). The solutions were vortexed using a MS3 vortexer (IKA 0003617000) and adapter (IKA 0003428000) at 2,000 rpm for 1 minute. The samples were then loaded into the TapeStation 2200 (Agilent) instrument, which loaded the DNA onto the D1000 ScreenTape (Agilent 50675582). After the miniature electrophoresis was done, the fragment sizes of the DNA libraries could be read out in bp.

qPCR

DNA libraries with a similar fragment length and known concentration have been used as a standard to assess the tagging efficiency of sequencing libraries. Therefore, the standard DNA has been diluted with 0.1% Tween 20 to 20 pM, 10 pM, 5 pM, 2.5 pM and 1.25 pM. The DNA concentration from Qubit and the fragment length determined by TapeStation measurement was used to calculate the molarity of the samples using the formula below. The samples were diluted to 5 pM and compared to the 5 pM standard readout to calculate the tagging efficiency and adjust the concentration for submission to sequencing if necessary.

$$\frac{(\text{concentration in } \frac{\text{ng}}{\mu\text{L}})}{(660 \frac{\text{g}}{\text{mol}} \times \text{average library size in bp})} \times 10^6 = \text{concentration in nM}$$

7.2.3 Analyses of the sequencing data

7.2.3.1 LAM-HTGTS

Libraries were sequenced on MiSeq V3 (Illumina) as 300 bp paired-end reads. An in-house pipeline was used for preprocessing to trim the Illumina adapters and demultiplex pooled libraries. Reads were aligned to the mouse mm10/GRCm38 genome using Bowtie2¹⁶⁵ (Version 2.5.1). Unique bait-prey junctions are kept and classified as filtered junctions while duplicates were removed. The in-detail process of the analysis is explained in Hu *et al.*¹²⁰ and the original associated documentation can be found at the GitHub repository https://robinmeyers.github.io/transloc_pipeline/index.html (Alt Lab, 2016). A package which includes the used pipeline inside a docker container can be found at the GitHub repository <https://github.com/brainbreaks/HTGTS> (Wei Lab, 2021).

7.2.3.2 GRO-seq

Libraries were sequenced on NextSeq 550 as 75 bp single-ended high-output reads. An in-house pipeline is used for preprocessing to trim the Illumina adapters and demultiplex pooled libraries. Reads were aligned to the mouse mm10/GRCm38 genome using Bowtie2 (Version 2.5.1). Unique sequences are kept and while duplicates were removed. Transcripts were identified and gene expression was estimated based on the workflow previously published by Meng *et al.*¹⁶⁶. A package which includes the used pipeline inside a docker container can be found at the GitHub repository <https://github.com/brainbreaks/GROseq> (Wei Lab, 2021).

7.2.3.3 Repli-seq

Libraries were sequenced on NovaSeq 6K as 100 bp single-ended reads on a SP flow cell. Reads were aligned to the mouse mm10/GRCm38 genome using Bowtie2 (Version 2.5.1). The protocol by Marchal *et al.*¹⁶⁷ was followed and analyzed as described. The statistical significance testing was performed using the SwitchRT package, which can be found at the GitHub repository <https://github.com/ClaireMarchal/SwitchRT> (March 2023).

8 References

- 1 Tsegay, P. S., Lai, Y. & Liu, Y. Replication Stress and Consequential Instability of the Genome and Epigenome. *Molecules* **24** (2019). <https://doi.org:10.3390/molecules24213870>
- 2 Petermann, E., Orta, M. L., Issaeva, N., Schultz, N. & Helleday, T. Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol Cell* **37**, 492-502 (2010). <https://doi.org:10.1016/j.molcel.2010.01.021>
- 3 Schwartz, E. K. & Heyer, W. D. Processing of joint molecule intermediates by structure-selective endonucleases during homologous recombination in eukaryotes. *Chromosoma* **120**, 109-127 (2011). <https://doi.org:10.1007/s00412-010-0304-7>
- 4 Kotsantis, P., Petermann, E. & Boulton, S. J. Mechanisms of Oncogene-Induced Replication Stress: Jigsaw Falling into Place. *Cancer Discov* **8**, 537-555 (2018). <https://doi.org:10.1158/2159-8290.Cd-17-1461>
- 5 Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability — an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology* **11**, 220-228 (2010). <https://doi.org:10.1038/nrm2858>
- 6 Hinds, P. W. *et al.* Regulation of retinoblastoma protein functions by ectopic expression of human cyclins. *Cell* **70**, 993-1006 (1992). [https://doi.org:10.1016/0092-8674\(92\)90249-c](https://doi.org:10.1016/0092-8674(92)90249-c)
- 7 Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864-870 (2005). <https://doi.org:10.1038/nature03482>
- 8 Murga, M. *et al.* Exploiting oncogene-induced replicative stress for the selective killing of Myc-driven tumors. *Nat Struct Mol Biol* **18**, 1331-1335 (2011). <https://doi.org:10.1038/nsmb.2189>
- 9 Curti, L. & Campaner, S. MYC-Induced Replicative Stress: A Double-Edged Sword for Cancer Development and Treatment. *Int J Mol Sci* **22** (2021). <https://doi.org:10.3390/ijms22126168>
- 10 Bester, A. C. *et al.* Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell* **145**, 435-446 (2011). <https://doi.org:10.1016/j.cell.2011.03.044>
- 11 Vasjari, L., Bresan, S., Biskup, C., Pai, G. & Rubio, I. Ras signals principally via Erk in G1 but cooperates with PI3K/Akt for Cyclin D induction and S-phase entry. *Cell Cycle* **18**, 204-225 (2019). <https://doi.org:10.1080/15384101.2018.1560205>

- 12 Kotsantis, P. *et al.* Increased global transcription activity as a mechanism of replication stress in cancer. *Nat Commun* **7**, 13087 (2016).
<https://doi.org:10.1038/ncomms13087>
- 13 Weyemi, U. *et al.* ROS-generating NADPH oxidase NOX4 is a critical mediator in oncogenic H-Ras-induced DNA damage and subsequent senescence. *Oncogene* **31**, 1117-1129 (2012).
<https://doi.org:10.1038/onc.2011.327>
- 14 Bartkova, J. *et al.* Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* **444**, 633-637 (2006). <https://doi.org:10.1038/nature05268>
- 15 Beck, H. *et al.* Cyclin-dependent kinase suppression by WEE1 kinase protects the genome through control of replication initiation and nucleotide consumption. *Mol Cell Biol* **32**, 4226-4236 (2012).
<https://doi.org:10.1128/mcb.00412-12>
- 16 Sørensen, C. S. & Syljuåsen, R. G. Safeguarding genome integrity: the checkpoint kinases ATR, CHK1 and WEE1 restrain CDK activity during normal DNA replication. *Nucleic Acids Res* **40**, 477-486 (2012).
<https://doi.org:10.1093/nar/gkr697>
- 17 Diehl, F. F. *et al.* Nucleotide imbalance decouples cell growth from cell proliferation. *Nature Cell Biology* **24**, 1252-1264 (2022).
<https://doi.org:10.1038/s41556-022-00965-1>
- 18 Chabes, A. *et al.* Survival of DNA Damage in Yeast Directly Depends on Increased dNTP Levels Allowed by Relaxed Feedback Inhibition of Ribonucleotide Reductase. *Cell* **112**, 391-401 (2003).
[https://doi.org:10.1016/S0092-8674\(03\)00075-8](https://doi.org:10.1016/S0092-8674(03)00075-8)
- 19 Blow, J. J. & Dutta, A. Preventing re-replication of chromosomal DNA. *Nat Rev Mol Cell Biol* **6**, 476-486 (2005). <https://doi.org:10.1038/nrm1663>
- 20 Siddiqui, K., On, K. F. & Diffley, J. F. Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol* **5** (2013).
<https://doi.org:10.1101/cshperspect.a012930>
- 21 Rampazzo, C. *et al.* Regulation by degradation, a cellular defense against deoxyribonucleotide pool imbalances. *Mutat Res* **703**, 2-10 (2010).
<https://doi.org:10.1016/j.mrgentox.2010.06.002>
- 22 Giannattasio, M. & Branzei, D. S-phase checkpoint regulations that preserve replication and chromosome integrity upon dNTP depletion. *Cell Mol Life Sci* **74**, 2361-2380 (2017). <https://doi.org:10.1007/s00018-017-2474-4>
- 23 Zhao, X., Chabes, A., Domkin, V., Thelander, L. & Rothstein, R. The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage. *Embo j* **20**, 3544-3553 (2001). <https://doi.org:10.1093/emboj/20.13.3544>

- 24 Aird, K. M. & Zhang, R. Nucleotide metabolism, oncogene-induced senescence and cancer. *Cancer Lett* **356**, 204-210 (2015).
<https://doi.org:10.1016/j.canlet.2014.01.017>
- 25 Azhar, A., Begum, N. A. & Husain, A. Nucleotide Pool Imbalance and Antibody Gene Diversification. *Vaccines (Basel)* **9** (2021).
<https://doi.org:10.3390/vaccines9101050>
- 26 Niedernhofer, L. J., Lalai, A. S. & Hoeijmakers, J. H. Fanconi anemia (cross)linked to DNA repair. *Cell* **123**, 1191-1198 (2005).
<https://doi.org:10.1016/j.cell.2005.12.009>
- 27 Noll, D. M., Mason, T. M. & Miller, P. S. Formation and Repair of Interstrand Cross-Links in DNA. *Chemical Reviews* **106**, 277-301 (2006).
<https://doi.org:10.1021/cr040478b>
- 28 Wang, G. & Vasquez, K. M. Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst)* **19**, 143-151 (2014). <https://doi.org:10.1016/j.dnarep.2014.03.017>
- 29 Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Research* **43**, 8627-8637 (2015).
<https://doi.org:10.1093/nar/gkv862>
- 30 Ravichandran, S., Subramani, V. K. & Kim, K. K. Z-DNA in the genome: from structure to disease. *Biophys Rev* **11**, 383-387 (2019).
<https://doi.org:10.1007/s12551-019-00534-1>
- 31 Jain, A., Wang, G. & Vasquez, K. M. DNA triple helices: Biological consequences and therapeutic potential. *Biochimie* **90**, 1117-1130 (2008).
<https://doi.org:https://doi.org/10.1016/j.biochi.2008.02.011>
- 32 Bikard, D., Loot, C., Baharoglu, Z. & Mazel, D. Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiol Mol Biol Rev* **74**, 570-588 (2010). <https://doi.org:10.1128/mnbr.00026-10>
- 33 Gaillard, H., Herrera-Moyano, E. & Aguilera, A. Transcription-Associated Genome Instability. *Chemical Reviews* **113**, 8638-8661 (2013).
<https://doi.org:10.1021/cr400017y>
- 34 Hamperl, S. & Cimprich, K. A. Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell* **167**, 1455-1467 (2016).
<https://doi.org:https://doi.org/10.1016/j.cell.2016.09.053>
- 35 Aguilera, A. & García-Muse, T. R Loops: From Transcription Byproducts to Threats to Genome Stability. *Molecular Cell* **46**, 115-124 (2012).
<https://doi.org:10.1016/j.molcel.2012.04.009>
- 36 Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nature Reviews Genetics* **16**, 583-597 (2015).
<https://doi.org:10.1038/nrg3961>

- 37 Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**, 814-825 (2012). <https://doi.org:10.1016/j.molcel.2012.01.017>
- 38 Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365-378 (2005). <https://doi.org:10.1016/j.cell.2005.06.008>
- 39 Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **170**, 774-786.e719 (2017). <https://doi.org:10.1016/j.cell.2017.07.043>
- 40 Allison, D. F. & Wang, G. G. R-loops: formation, function, and relevance to cell stress. *Cell Stress* **3**, 38-46 (2019). <https://doi.org:10.15698/cst2019.02.175>
- 41 Cohen, S. *et al.* Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nature Communications* **9**, 533 (2018). <https://doi.org:10.1038/s41467-018-02894-w>
- 42 Cerritelli, S. M. & Crouch, R. J. Ribonuclease H: the enzymes in eukaryotes. *Febs j* **276**, 1494-1505 (2009). <https://doi.org:10.1111/j.1742-4658.2009.06908.x>
- 43 Hegazy, Y. A., Fernando, C. M. & Tran, E. J. The balancing act of R-loop biology: The good, the bad, and the ugly. *J Biol Chem* **295**, 905-913 (2020). <https://doi.org:10.1074/jbc.REV119.011353>
- 44 Sollier, J. *et al.* Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Molecular Cell* **56**, 777-785 (2014). <https://doi.org:10.1016/j.molcel.2014.10.020>
- 45 Sollier, J. & Cimprich, K. A. Breaking bad: R-loops and genome integrity. *Trends in Cell Biology* **25**, 514-522 (2015). <https://doi.org:https://doi.org/10.1016/j.tcb.2015.05.003>
- 46 Percharde, M., Bulut-Karslioglu, A. & Ramalho-Santos, M. Hypertranscription in Development, Stem Cells, and Regeneration. *Developmental Cell* **40**, 9-21 (2017). <https://doi.org:https://doi.org/10.1016/j.devcel.2016.11.010>
- 47 Bowry, A., Kelly, R. D. W. & Petermann, E. Hypertranscription and replication stress in cancer. *Trends Cancer* **7**, 863-877 (2021). <https://doi.org:10.1016/j.trecan.2021.04.006>
- 48 Vilette, D., Ehrlich, S. D. & Michel, B. Transcription-induced deletions in plasmid vectors: M13 DNA replication as a source of instability. *Mol Gen Genet* **252**, 398-403 (1996). <https://doi.org:10.1007/BF02173004>

- 49 Srivatsan, A., Tehrani, A., MacAlpine, D. M. & Wang, J. D. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet* **6**, e1000810 (2010).
<https://doi.org/10.1371/journal.pgen.1000810>
- 50 Kornberg, A. & Baker, T. A. *DNA replication*. Vol. 3 (Wh Freeman New York, 1992).
- 51 Wang, J. D., Berkmen, M. B. & Grossman, A. D. Genome-wide coorientation of replication and transcription reduces adverse effects on replication in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* **104**, 5608-5613 (2007).
- 52 Prado, F. & Aguilera, A. Impairment of replication fork progression mediates RNA polIII transcription-associated recombination. *The EMBO journal* **24**, 1267-1276 (2005).
- 53 Guy, L. & Roten, C.-A. H. Genometric analyses of the organization of circular chromosomes: a universal pressure determines the direction of ribosomal RNA genes transcription relative to chromosome replication. *Gene* **340**, 45-52 (2004).
- 54 Huvet, M. *et al.* Human gene organization driven by the coordination of replication and transcription. *Genome research* **17**, 1278-1285 (2007).
- 55 Trautinger, B. W., Jaktaji, R. P., Rusakova, E. & Lloyd, R. G. RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol Cell* **19**, 247-258 (2005).
<https://doi.org/10.1016/j.molcel.2005.06.004>
- 56 Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* **58**, 235-263 (2017).
<https://doi.org/10.1002/em.22087>
- 57 Kirkwood, T. B. L. Understanding the Odd Science of Aging. *Cell* **120**, 437-447 (2005). <https://doi.org/10.1016/j.cell.2005.01.027>
- 58 Lindahl, T. & Barnes, D. E. Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* **65**, 127-133 (2000).
<https://doi.org/10.1101/sqb.2000.65.127>
- 59 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351 (2001). <https://doi.org/10.1126/science.1058040>
- 60 Cheung-Ong, K., Giaever, G. & Nislow, C. DNA-Damaging Agents in Cancer Chemotherapy: Serendipity and Chemical Biology. *Chemistry & Biology* **20**, 648-659 (2013).
<https://doi.org/https://doi.org/10.1016/j.chembiol.2013.04.007>
- 61 Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nature Genetics* **27**, 247-254 (2001).
<https://doi.org/10.1038/85798>

- 62 Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, 644-656 (2017). <https://doi.org:10.1016/j.cell.2017.01.002>
- 63 Jackson, S. P. Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**, 687-696 (2002). <https://doi.org:10.1093/carcin/23.5.687>
- 64 Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol* **26**, 52-64 (2016). <https://doi.org:10.1016/j.tcb.2015.07.009>
- 65 Li, S. & Wu, X. Common fragile sites: protection and repair. *Cell & Bioscience* **10**, 29 (2020). <https://doi.org:10.1186/s13578-020-00392-5>
- 66 Ji, F. *et al.* Genome-wide high-resolution mapping of mitotic DNA synthesis sites and common fragile sites by direct sequencing. *Cell Research* **30**, 1009-1023 (2020). <https://doi.org:10.1038/s41422-020-0357-y>
- 67 Smith, D. I., Zhu, Y., McAvoy, S. & Kuhn, R. Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett* **232**, 48-57 (2006). <https://doi.org:10.1016/j.canlet.2005.06.049>
- 68 Martínez-Cerdeño, V. & Noctor, S. C. Neural Progenitor Cell Terminology. *Frontiers in Neuroanatomy* **12** (2018). <https://doi.org:10.3389/fnana.2018.00104>
- 69 Matarredona, E. R. & Pastor, A. M. Neural Stem Cells of the Subventricular Zone as the Origin of Human Glioblastoma Stem Cells. Therapeutic Implications. *Frontiers in Oncology* **9** (2019). <https://doi.org:10.3389/fonc.2019.00779>
- 70 Wang, M. *et al.* Increased Neural Progenitor Proliferation in a hiPSC Model of Autism Induces Replication Stress-Associated Genome Instability. *Cell Stem Cell* **26**, 221-233.e226 (2020). <https://doi.org:https://doi.org/10.1016/j.stem.2019.12.013>
- 71 Wei, P.-C. *et al.* Three classes of recurrent DNA break clusters in brain progenitors identified by 3D proximity-based break joining assay. *Proceedings of the National Academy of Sciences* **115**, 1919-1924 (2018). <https://doi.org:doi:10.1073/pnas.1719907115>
- 72 Wei, P. C. *et al.* Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* **164**, 644-655 (2016). <https://doi.org:10.1016/j.cell.2015.12.039>
- 73 Krokan, H., Wist, E. & Krokan, R. H. Aphidicolin inhibits DNA synthesis by DNA polymerase alpha and isolated nuclei by a similar mechanism. *Nucleic Acids Res* **9**, 4709-4719 (1981). <https://doi.org:10.1093/nar/9.18.4709>

- 74 Baranovskiy, A. G. *et al.* Structural basis for inhibition of DNA replication by aphidicolin. *Nucleic Acids Res* **42**, 14013-14021 (2014).
<https://doi.org:10.1093/nar/gku1209>
- 75 Sessa, C. *et al.* Phase I and clinical pharmacological evaluation of aphidicolin glycinate. *J Natl Cancer Inst* **83**, 1160-1164 (1991).
<https://doi.org:10.1093/jnci/83.16.1160>
- 76 Edelson, R. E., Gorycki, P. D. & MacDonald, T. L. The mechanism of aphidicolin bioinactivation by rat liver in vitro systems. *Xenobiotica* **20**, 273-287 (1990). <https://doi.org:10.3109/00498259009046847>
- 77 Branzei, D. & Foiani, M. Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol* **9**, 297-308 (2008).
<https://doi.org:10.1038/nrm2351>
- 78 Aziz Sancar, Laura A. Lindsey-Boltz, Keziban Ünsal-Kaçmaz & Linn, S. Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. *Annual Review of Biochemistry* **73**, 39-85 (2004).
<https://doi.org:10.1146/annurev.biochem.73.011303.073723>
- 79 Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *The EMBO Journal* **17**, 5497-5508 (1998).
<https://doi.org:https://doi.org/10.1093/emboj/17.18.5497>
- 80 Delacôte, F. & Lopez, B. S. Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: the trans-S double-strand break repair model. *Cell Cycle* **7**, 33-38 (2008). <https://doi.org:10.4161/cc.7.1.5149>
- 81 Haber, J. E. Deciphering the DNA Damage Response. *Cell* **162**, 1183-1185 (2015). <https://doi.org:10.1016/j.cell.2015.08.034>
- 82 Simsek, D. & Jasin, M. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4–ligase IV during chromosomal translocation formation. *Nature Structural & Molecular Biology* **17**, 410-416 (2010). <https://doi.org:10.1038/nsmb.1773>
- 83 Hefferin, M. L. & Tomkinson, A. E. Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA Repair* **4**, 639-648 (2005). <https://doi.org:https://doi.org/10.1016/j.dnarep.2004.12.005>
- 84 Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology* **18**, 495-506 (2017).
<https://doi.org:10.1038/nrm.2017.48>
- 85 Weterings, E. & Chen, D. J. The endless tale of non-homologous end-joining. *Cell Research* **18**, 114-124 (2008).
<https://doi.org:10.1038/cr.2008.3>

- 86 Oksenyich, V. *et al.* Functional redundancy between the XLF and DNA-PKcs DNA repair factors in V(D)J recombination and nonhomologous DNA end joining. *Proc Natl Acad Sci U S A* **110**, 2234-2239 (2013). <https://doi.org:10.1073/pnas.1222573110>
- 87 Li, Z. *et al.* The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83**, 1079-1089 (1995). [https://doi.org:https://doi.org/10.1016/0092-8674\(95\)90135-3](https://doi.org:https://doi.org/10.1016/0092-8674(95)90135-3)
- 88 Gao, Y. *et al.* Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature* **404**, 897-900 (2000). <https://doi.org:10.1038/35009138>
- 89 Yan, C. T. *et al.* XRCC4 suppresses medulloblastomas with recurrent translocations in p53-deficient mice. *Proc Natl Acad Sci U S A* **103**, 7378-7383 (2006). <https://doi.org:10.1073/pnas.0601938103>
- 90 Iliakis, G., Murmann, T. & Soni, A. Alternative end-joining repair pathways are the ultimate backup for abrogated classical non-homologous end-joining and homologous recombination repair: Implications for the formation of chromosome translocations. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **793**, 166-175 (2015). <https://doi.org:https://doi.org/10.1016/j.mrgentox.2015.07.001>
- 91 Sallmyr, A. & Tomkinson, A. E. Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *Journal of Biological Chemistry* **293**, 10536-10546 (2018). <https://doi.org:https://doi.org/10.1074/jbc.TM117.000375>
- 92 Costantino, L. *et al.* Break-Induced Replication Repair of Damaged Forks Induces Genomic Duplications in Human Cells. *Science* **343**, 88-91 (2014). <https://doi.org:doi:10.1126/science.1243211>
- 93 Iliakis, G. *et al.* Mechanisms of DNA double strand break repair and chromosome aberration formation. *Cytogenetic and Genome Research* **104**, 14-20 (2004). <https://doi.org:10.1159/000077461>
- 94 Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* **7**, 2902-2906 (2008). <https://doi.org:10.4161/cc.7.18.6679>
- 95 Filippo, J. S., Sung, P. & Klein, H. Mechanism of Eukaryotic Homologous Recombination. *Annual Review of Biochemistry* **77**, 229-257 (2008). <https://doi.org:10.1146/annurev.biochem.77.061306.125255>
- 96 Liu, S. & Kong, D. End resection: a key step in homologous recombination and DNA double-strand break repair. *Genome Instability & Disease* **2**, 39-50 (2021). <https://doi.org:10.1007/s42764-020-00028-5>

- 97 Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. Homologous recombination and its regulation. *Nucleic Acids Research* **40**, 5795-5818 (2012). <https://doi.org:10.1093/nar/gks270>
- 98 Levine, A. J., Momand, J. & Finlay, C. A. The p53 tumour suppressor gene. *Nature* **351**, 453-456 (1991). <https://doi.org:10.1038/351453a0>
- 99 Oren, M. Decision making by p53: life, death and cancer. *Cell Death & Differentiation* **10**, 431-442 (2003). <https://doi.org:10.1038/sj.cdd.4401183>
- 100 Toufektchan, E. & Toledo, F. The Guardian of the Genome Revisited: p53 Downregulates Genes Required for Telomere Maintenance, DNA Repair, and Centromere Structure. *Cancers (Basel)* **10** (2018). <https://doi.org:10.3390/cancers10050135>
- 101 Menon, V. & Povirk, L. Involvement of p53 in the repair of DNA double strand breaks: multifaceted Roles of p53 in homologous recombination repair (HRR) and non-homologous end joining (NHEJ). *Subcell Biochem* **85**, 321-336 (2014). https://doi.org:10.1007/978-94-017-9211-0_17
- 102 Yao, Y. & Dai, W. Genomic Instability and Cancer. *J Carcinog Mutagen* **5** (2014). <https://doi.org:10.4172/2157-2518.1000165>
- 103 Choi, E.-H., Yoon, S., Koh, Y. E., Seo, Y.-J. & Kim, K. P. Maintenance of genome integrity and active homologous recombination in embryonic stem cells. *Experimental & Molecular Medicine* **52**, 1220-1229 (2020). <https://doi.org:10.1038/s12276-020-0481-2>
- 104 Nik-Zainal, S. From genome integrity to cancer. *Genome Medicine* **11**, 4 (2019). <https://doi.org:10.1186/s13073-019-0617-y>
- 105 Furgason, J. M. & Bahassi, E. M. Targeting DNA repair mechanisms in cancer. *Pharmacology & Therapeutics* **137**, 298-308 (2013). <https://doi.org:https://doi.org/10.1016/j.pharmthera.2012.10.009>
- 106 McKinnon, P. J. Genome integrity and disease prevention in the nervous system. *Genes Dev* **31**, 1180-1194 (2017). <https://doi.org:10.1101/gad.301325.117>
- 107 Schwer, B. *et al.* Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc Natl Acad Sci U S A* **113**, 2258-2263 (2016). <https://doi.org:10.1073/pnas.1525564113>
- 108 Pomerantz, R. T. & O'Donnell, M. What happens when replication and transcription complexes collide? *Cell Cycle* **9**, 2537-2543 (2010). <https://doi.org:10.4161/cc.9.13.12122>
- 109 Bouwman, B. A. M. & Crosetto, N. Endogenous DNA Double-Strand Breaks during DNA Transactions: Emerging Insights and Methods for Genome-Wide Profiling. *Genes (Basel)* **9** (2018). <https://doi.org:10.3390/genes9120632>

- 110 So, A., Le Guen, T., Lopez, B. S. & Guirouilh-Barbat, J. Genomic rearrangements induced by unscheduled DNA double strand breaks in somatic mammalian cells. *The FEBS Journal* **284**, 2324-2344 (2017). [https://doi.org:https://doi.org/10.1111/febs.14053](https://doi.org/https://doi.org/10.1111/febs.14053)
- 111 Bednarski, J. J. & Sleckman, B. P. Lymphocyte development: integration of DNA damage response signaling. *Adv Immunol* **116**, 175-204 (2012). <https://doi.org:10.1016/b978-0-12-394300-2.00006-5>
- 112 Rousseau, L. *et al.* In vivo importance of homologous recombination DNA repair for mouse neural stem and progenitor cells. *PLoS One* **7**, e37194 (2012). <https://doi.org:10.1371/journal.pone.0037194>
- 113 Kashiwagi, H., Shiraishi, K., Sakaguchi, K., Nakahama, T. & Kodama, S. Repair kinetics of DNA double-strand breaks and incidence of apoptosis in mouse neural stem/progenitor cells and their differentiated neurons exposed to ionizing radiation. *Journal of Radiation Research* **59**, 261-271 (2018). <https://doi.org:10.1093/jrr/rrx089>
- 114 Brambati, A., Colosio, A., Zardoni, L., Galanti, L. & Liberi, G. Replication and transcription on a collision course: eukaryotic regulation mechanisms and implications for DNA stability. *Front Genet* **6**, 166 (2015). <https://doi.org:10.3389/fgene.2015.00166>
- 115 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002). <https://doi.org:10.1101/gr.229102>
- 116 Zeisel, A., Yizhaky, A., Bossel Ben-Moshe, N. & Domany, E. An accessible database for mouse and human whole transcriptome qPCR primers. *Bioinformatics* **29**, 1355-1356 (2013). <https://doi.org:10.1093/bioinformatics/btt145>
- 117 Tena, A. *et al.* Induction of recurrent break cluster genes in neural progenitor cells differentiated from embryonic stem cells in culture. *Proceedings of the National Academy of Sciences* **117**, 10541-10546 (2020). <https://doi.org:doi:10.1073/pnas.1922299117>
- 118 Wilson, T. E. *et al.* Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome Res* **25**, 189-200 (2015). <https://doi.org:10.1101/gr.177121.114>
- 119 Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnology* **33**, 179-186 (2015). <https://doi.org:10.1038/nbt.3101>
- 120 Hu, J. *et al.* Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nature Protocols* **11**, 853-871 (2016). <https://doi.org:10.1038/nprot.2016.043>
- 121 Zhang, J. A., Mortazavi, A., Williams, B. A., Wold, B. J. & Rothenberg, E. V. Dynamic transformations of genome-wide epigenetic marking and

- transcriptional control establish T cell identity. *Cell* **149**, 467-482 (2012).
<https://doi.org:10.1016/j.cell.2012.01.056>
- 122 Hellman, A. *et al.* Replication Delay along FRA7H, a Common Fragile Site on Human Chromosome 7, Leads to Chromosomal Instability. *Molecular and Cellular Biology* **20**, 4420-4427 (2000).
<https://doi.org:10.1128/MCB.20.12.4420-4427.2000>
- 123 Zhao, P. A., Sasaki, T. & Gilbert, D. M. High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. *Genome Biology* **21**, 76 (2020).
<https://doi.org:10.1186/s13059-020-01983-8>
- 124 Petryk, N. *et al.* Replication landscape of the human genome. *Nature Communications* **7**, 10208 (2016). <https://doi.org:10.1038/ncomms10208>
- 125 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745 (2016). <https://doi.org:10.1093/nar/gkv1189>
- 126 Kit Leng Lui, S. *et al.* Monitoring genome-wide replication fork directionality by Okazaki fragment sequencing in mammalian cells. *Nature Protocols* **16**, 1193-1218 (2021). <https://doi.org:10.1038/s41596-020-00454-5>
- 127 Qiao, J. *et al.* Identification of β -N-catenin as a novel tumor suppressor in neuroblastoma. *Oncotarget* **10** (2019).
- 128 Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature Biotechnology* **21**, 183-186 (2003).
<https://doi.org:10.1038/nbt780>
- 129 Bargonetti, J. & Manfredi, J. J. Multiple roles of the tumor suppressor p53. *Curr Opin Oncol* **14**, 86-91 (2002). <https://doi.org:10.1097/00001622-200201000-00015>
- 130 O'Leary, B., Finn, R. S. & Turner, N. C. Treating cancer with selective CDK4/6 inhibitors. *Nat Rev Clin Oncol* **13**, 417-430 (2016).
<https://doi.org:10.1038/nrclinonc.2016.26>
- 131 Trotter, E. W. & Hagan, I. M. Release from cell cycle arrest with Cdk4/6 inhibitors generates highly synchronized cell cycle progression in human cell culture. *Open Biol* **10**, 200200 (2020).
<https://doi.org:10.1098/rsob.200200>
- 132 Orthwein, A. *et al.* A mechanism for the suppression of homologous recombination in G1 cells. *Nature* **528**, 422-426 (2015).
<https://doi.org:10.1038/nature16142>
- 133 Ji, F. *et al.* New Era of Mapping and Understanding Common Fragile Sites: An Updated Review on Origin of Chromosome Fragility. *Frontiers in Genetics* **13** (2022). <https://doi.org:10.3389/fgene.2022.906957>

- 134 Park, So H. *et al.* Locus-specific transcription silencing at the FHIT gene suppresses replication stress-induced copy number variant formation and associated replication delay. *Nucleic Acids Research* **49**, 7507-7524 (2021). <https://doi.org:10.1093/nar/gkab559>
- 135 Hull, R. M., Cruz, C., Jack, C. V. & Houseley, J. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLOS Biology* **15**, e2001333 (2017). <https://doi.org:10.1371/journal.pbio.2001333>
- 136 Topal, S., Vasseur, P., Radman-Livaja, M. & Peterson, C. L. Distinct transcriptional roles for Histone H3-K56 acetylation during the cell cycle in Yeast. *Nature Communications* **10**, 4372 (2019). <https://doi.org:10.1038/s41467-019-12400-5>
- 137 Snyder, M., Sapolsky, R. J. & Davis, R. W. Transcription interferes with elements important for chromosome maintenance in *Saccharomyces cerevisiae*. *Mol Cell Biol* **8**, 2184-2194 (1988). <https://doi.org:10.1128/mcb.8.5.2184-2194.1988>
- 138 Lööke, M. *et al.* Relicensing of transcriptionally inactivated replication origins in budding yeast. *J Biol Chem* **285**, 40004-40011 (2010). <https://doi.org:10.1074/jbc.M110.148924>
- 139 Rivera-Mulia, J. C. *et al.* Replication timing networks reveal a link between transcription regulatory circuits and replication timing control. *Genome Res* **29**, 1415-1428 (2019). <https://doi.org:10.1101/gr.247049.118>
- 140 Sarni, D. *et al.* 3D genome organization contributes to genome instability at fragile sites. *Nature Communications* **11**, 3613 (2020). <https://doi.org:10.1038/s41467-020-17448-2>
- 141 Liu, Y. *et al.* Transcription shapes DNA replication initiation to preserve genome integrity. *Genome Biology* **22**, 176 (2021). <https://doi.org:10.1186/s13059-021-02390-3>
- 142 Kouzine, F., Levens, D. & Baranello, L. DNA topology and transcription. *Nucleus* **5**, 195-202 (2014). <https://doi.org:10.4161/nucl.28909>
- 143 Teves, S. S. & Henikoff, S. Transcription-generated torsional stress destabilizes nucleosomes. *Nature Structural & Molecular Biology* **21**, 88-94 (2014). <https://doi.org:10.1038/nsmb.2723>
- 144 Cristini, A., Géraud, M. & Sordet, O. in *International Review of Cell and Molecular Biology* Vol. 364 (eds Urbain Weyemi & Lorenzo Galluzzi) 195-240 (Academic Press, 2021).
- 145 Andrade-Lima, L. C., Veloso, A., Paulsen, M. T., Menck, C. F. M. & Ljungman, M. DNA repair and recovery of RNA synthesis following exposure to ultraviolet light are delayed in long genes. *Nucleic Acids Research* **43**, 2744-2756 (2015). <https://doi.org:10.1093/nar/gkv148>

- 146 Mabb, A. M. *et al.* Topoisomerase 1 Regulates Gene Expression in Neurons through Cleavage Complex-Dependent and -Independent Mechanisms. *PLoS One* **11**, e0156439 (2016). <https://doi.org:10.1371/journal.pone.0156439>
- 147 Jinks-Robertson, S. & Bhagwat, A. S. Transcription-Associated Mutagenesis. *Annual Review of Genetics* **48**, 341-359 (2014). <https://doi.org:10.1146/annurev-genet-120213-092015>
- 148 Green, P. *et al.* Transcription-associated mutational asymmetry in mammalian evolution. *Nature Genetics* **33**, 514-517 (2003). <https://doi.org:10.1038/ng1103>
- 149 Klein, K. N. *et al.* Replication timing maintains the global epigenetic state in human cells. *Science* **372**, 371-378 (2021). <https://doi.org:doi:10.1126/science.aba5545>
- 150 Courtot, L. *et al.* Low Replicative Stress Triggers Cell-Type Specific Inheritable Advanced Replication Timing. *International Journal of Molecular Sciences* **22**, 4959 (2021).
- 151 Du, Q. *et al.* Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nature Communications* **10**, 416 (2019). <https://doi.org:10.1038/s41467-019-08302-1>
- 152 Goldberg, M. E. & Harris, K. Mutational Signatures of Replication Timing and Epigenetic Modification Persist through the Global Divergence of Mutation Spectra across the Great Ape Phylogeny. *Genome Biology and Evolution* **14** (2021). <https://doi.org:10.1093/gbe/evab104>
- 153 Andrs, M., Hasanova, Z., Oravetzova, A., Dobrovolna, J. & Janscak, P. RECQ5: A Mysterious Helicase at the Interface of DNA Replication and Transcription. *Genes* **11**, 232 (2020).
- 154 Macheret, M. *et al.* High-resolution mapping of mitotic DNA synthesis regions and common fragile sites in the human genome through direct sequencing. *Cell Research* **30**, 997-1008 (2020). <https://doi.org:10.1038/s41422-020-0358-x>
- 155 Di Marco, S. *et al.* RECQ5 Helicase Cooperates with MUS81 Endonuclease in Processing Stalled Replication Forks at Common Fragile Sites during Mitosis. *Mol Cell* **66**, 658-671.e658 (2017). <https://doi.org:10.1016/j.molcel.2017.05.006>
- 156 Dai, H.-Q. *et al.* Direct analysis of brain phenotypes via neural blastocyst complementation. *Nature Protocols* **15**, 3154-3181 (2020). <https://doi.org:10.1038/s41596-020-0364-y>
- 157 Glover, T. W., Wilson, T. E. & Arlt, M. F. Fragile sites in cancer: more than meets the eye. *Nature Reviews Cancer* **17**, 489-501 (2017). <https://doi.org:10.1038/nrc.2017.52>

- 158 Henrichsen, C. N. *et al.* Segmental copy number variation shapes tissue transcriptomes. *Nature Genetics* **41**, 424-429 (2009). <https://doi.org:10.1038/ng.345>
- 159 Ishibashi, M. *et al.* Copy number variants in patients with intellectual disability affect the regulation of ARX transcription factor gene. *Hum Genet* **134**, 1163-1182 (2015). <https://doi.org:10.1007/s00439-015-1594-x>
- 160 Holt, R. *et al.* CNVs leading to fusion transcripts in individuals with autism spectrum disorder. *Eur J Hum Genet* **20**, 1141-1147 (2012). <https://doi.org:10.1038/ejhg.2012.73>
- 161 Bizzotto, S. & Walsh, C. A. Genetic mosaicism in the human brain: from lineage tracing to neuropsychiatric disorders. *Nat Rev Neurosci* **23**, 275-286 (2022). <https://doi.org:10.1038/s41583-022-00572-x>
- 162 Bizzotto, S. *et al.* Landmarks of human embryonic development inscribed in somatic mutations. *Science* **371**, 1249-1253 (2021). <https://doi.org:doi:10.1126/science.abe1544>
- 163 Paganelli, V., Giordano, M., Meazza, C., Schena, L. & Bozzola, M. An intragenic deletion within CTNNA2 intron 7 in a boy with short stature and speech delay: A case report. *SAGE Open Med Case Rep* **5**, 2050313x17693967 (2017). <https://doi.org:10.1177/2050313x17693967>
- 164 Ching, M. S. *et al.* Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am J Med Genet B Neuropsychiatr Genet* **153b**, 937-947 (2010). <https://doi.org:10.1002/ajmg.b.31063>
- 165 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359 (2012). <https://doi.org:10.1038/nmeth.1923>
- 166 Meng, F.-L. *et al.* Convergent Transcription at Intragenic Super-Enhancers Targets AID-Initiated Genomic Instability. *Cell* **159**, 1538-1548 (2014). <https://doi.org:https://doi.org/10.1016/j.cell.2014.11.014>
- 167 Marchal, C. *et al.* Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nature Protocols* **13**, 819-839 (2018). <https://doi.org:10.1038/nprot.2017.148>

9 Appendix

9.1 Abbreviations

Abbreviation	Definition
2D	Two dimensional
3D	Three dimensional
a-EJ	Alternative end joining
ai	Artificial intelligence
APH	Aphidicolin
ATM	Ataxia-telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3 related
<i>B. subtilis</i>	Bacillus subtilis
B&W	Bind and wash
BLAT	BLAST-like alignment tool
bp	Base pairs
BRCA2	Breast Cancer gene 2
BrdU	Bromodeoxyuridine
BrU	Bromouridine
BSA	Bovine serum albumin
Ccser1	Coiled-coil serine rich protein 1
CD	Co-directional collision
CDK	Cyclin-dependent kinase
cDNA	Complementary DNA
CFS	Common Fragile Site
ChIP	Chromatin Immunoprecipitation
chr	Chromosome
CNV	Copy number variation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Ctnna2	Catenin Alpha 2
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit

DNA2	DNA Replication Helicase/Nuclease 2
dNTPs	Deoxynucleotides
DPBS	Dulbecco's Phosphate Buffered Saline
Dr.	Doctor
ds	Downstream
DSB	Double-strand break
DSBR	Double-strand break repair
dsDNA	double-stranded DNA
E/L-Repli-seq	Early/Late-Replication-timing-sequencing
EDTA	Ethylendiaminetetraacetic acid
EGF	Epidermal growth factor
ENCODE	Encyclopedia of DNA Elements
ESC	Embryonic stem cell
ESC-NPC	Embryonic Stem Cell-induced Neural Progenitor Cell
EX	Exon
EXO1	Exonuclease 1
FACS	Fluorescence-Activated Cell Sorting
FGFb	Basic fibroblast growth factor
FHIT	Fragile Histidine Triad Diadenosine Triphosphatase
g	Gram
G1	Growth 1 phase
G2 phase	Growth 2 phase
gDNA	Genomic DNA
GRCm38	Genome Reference Consortium Mouse Build 38
Grid2	Glutamate Ionotropic Receptor Delta Type Subunit 2
GRO-seq	Global Run-On sequencing
H3K56ac	Acetylation of the 56th lysine residue of the histone H3
HF	High Fidelity
hiPSC	Human induced pluripotent stem cells
HO	Head-on collision
HPRT	Hypoxanthine-guanine phosphoribosyltransferase
HR	Homologous recombination

HS	High Sensitivity
HTGTS	High throughput, genome-wide translocation sequencing
HU	Hydroxyurea
ICL	interstrand cross-links
Jden	Junction density
kb	Kilobases
ko	Knockout
LAM-HTGTS	Linear amplification-mediated high-throughput genome-wide translocation sequencing
LB	Lysogeny broth
LC	Lorenzo Corazzi
LIF	Leukemia inhibitory factor
M phase	Mitosis
M.Sc.	Master of Science
Mb	Megabases
MEF	Mouse embryonic fibroblasts
MiDAS	Mitotic DNA Synthesis
mm10	Mus musculus 10 (=GRCm38)
MRN	MRE11-RAD50-NBS1
mRNA	Messenger RNA
n.d.	Not detectable
ng	Nanogram
NHEJ	Non-homologous end joining
nM	Nanomolarity
NPC	Neural Progenitor Cell
Nrxn1	Neurexin 1
NSC	Neural Stem Cell
NSPC	Neural Stem and Progenitor Cell
OK-seq	Okazaki fragments-sequencing
p-value	Calculated probability
PAGE	Polyacrylamide gel electrophoresis
PARP1	Poly(ADP-ribose) polymerase 1

PCR	Polymerase Chain Reaction
Pde10a	Phosphodiesterase 10A
PDGF-BB	Two B subunits of the platelet-derived growth factor
PE	Paired-end
PK	Proteinase K
PLO	Poly-L-Ornithine
pM	Picomolarity
Pol θ	DNA polymerase theta
Prof.	Professor
Ptprm	Protein Tyrosine Phosphatase Receptor Type M
PW	Pei-Chi Wei
qPCR	Quantitative Polymerase Chain Reaction
RAG1	Recombination Activating 1
RAG2	Recombination Activating 2
RAS	Rat sarcoma
RDC	Recurrent DNA break cluster
RefSeq	Reference Sequences
Repli-seq	Replication-timing-sequencing
RIF1	Replication Timing Regulatory Factor 1
RNA	Ribonucleic acid
RNAP	RNA polymerase
ROS	Reactive oxygen species
RPA	Replication protein A
RT	Replication timing
RT-qPCR	Reverse Transcription-quantitative Polymerase Chain Reaction
S phase	Synthesis phase
SDS	Sodium dodecyl sulfate
SDSA	Synthesis-dependent strand annealing
SEM	Standard error of the mean
sgRNA	Single guide RNA
SNV	Single nucleotide variant

SR	Single-read
ssDNA	Single-stranded DNA
SV	Structural variant
TAD	Topologically associating domain
TBE	Tris/Borate/EDTA
Tenm3	Teneurin Transmembrane Protein 3
TFO	Triplex-forming oligonucleotide
Top1	DNA topoisomerase 1
TP53	Tumor Protein 53
TSS	Transcription start site
UCSC	University of California Santa Cruz
up	Upstream
UV	Ultraviolet
VI	Vivien Ionasz
Vis	Visible
XRCC4	X-Ray Repair Cross Complementing 4
$\Delta\Delta$ CT	Comparative CT
μ L	Microliter
μ M	Micromolarity

9.2 List of Figures

Figure 1: **Transcription-replication collision and the fundamental hypothesis.**

Top: Schematic representation of a head-on collision between RNA polymerase II (RNAP - orange) and the replication fork (Pol ϵ - green, Pol δ - light blue and MCM - dark blue). The red lightning bolt represents replication stress and a potentially resulting DNA double-strand break. Bottom: Illustrative gene body shown in dark blue, with active transcription (RNA - green) on the left and abolished transcription on the right. This change in transcription is expected to also translate to the level of DNA breaks. For the transcribed gene in the lower left panel (DNA breaks), black bars are visible, these indicate double-strand breaks at the corresponding location of the gene body above. For the non-transcribed gene, the expectation is to have less DNA double-strand breaks or even eradicate the formation of a cluster altogether.20

Figure 2: **Experimental approach for Aim 1.** [A] Wild-type expression is depicted

by a green transcript along the total gene length. [B] Promoter and enhancer deletion using CRISPR/Cas9 leads to the abolishment of transcription. Gene bodies are depicted in blue, deleted areas are shown in gray, transcripts are shown in green, and the red crossed out circle illustrates transcriptional inactivation upon deletion of the proximal regulatory elements on the b allele.21

Figure 3: **Experimental system for the deactivation of transcription.** Embryonic

stem cell lines with p53- and Xrcc4-deficiency have been used as starting material for genome editing. These parental cell lines (yellow) are derived from mice and therefore have two alleles (a/b) of the RDC-genes that have been targeted. In the first step, one allele of the RDC-gene of interest has been deleted completely by CRISPR/Cas9. This intermediary cell line is going to be referred to as the founder cell line (a/ Δ). The experimental promoter and enhancer deleted clones (ape/ Δ) have been established by another round of CRISPR/Cas9 genome editing by long-range deletion of the promoter-proximal regulatory elements of the gene-of-interest on the remaining second allele.23

Figure 4: **Workflow of the experiments.** After induction of the embryonic stem

cells to neural progenitor cells, experiments were able to be conducted. For the Ctnna2 as well as the Nrnx1 cell lines nuclei have been harvested for GRO-seq (Global Run On-sequencing) and cells were nucleofected with CRISPR/Cas9 to induce the bait break that is necessary for HTGTS (High-Throughput Genome-wide Translocation Sequencing), the method applied to assess break dynamics by sequencing translocations. In a test run, also RNA has been isolated from the Ctnna2 cell lines to assess if RT-qPCR and GRO-seq are both able to confirm the deactivation of transcription.24

Figure 5: **Whole locus deletion of Ctnna2.** Gene locus of Ctnna2 (blue - gene

body at the top) \pm 0.5Mb (chr6:76,379,637-78,481,703) with corresponding transcription signal (blue - at the bottom) from nascent RNA in the parental cell line NXP010. Red vertical lines indicate the location of the two sgRNAs used to target Ctnna2 and delete the whole locus of one allele. sgRNA_Ctnna2_ds1 is located at chr6:76,879,005-76,879,024 and sgRNA_Ctnna2_up1 at chr6:77,981,693-77,981,712. The whole locus deletion spans across 1.1 Mb. The scale bar at the bottom right indicates 500 kb.25

Figure 6: Workflow for PCR validation and Sanger sequencing of positive clones after CRISPR/Cas9 cutting. The *Ctnna2* locus is here represented in black. sgRNAs are cutting downstream and upstream of the gene body (scissors). After deletional joining the yellow and blue part of the genome are merged. The deletion can be validated by using the PCR primers indicated as purple arrows, which produce a PCR product with a length of 443 bp. Positive clones have been additionally analyzed by Sanger sequencing. The result of the selected founder cell line (a/ Δ) is shown at the bottom as sequencing peaks and maintained sequence (yellow and blue) compared to the deleted original sequence (black) from the parental cell line (a/b)..... 26

Figure 7: Promoter-proximal deletion of *Ctnna2*'s promoter. Gene locus of *Ctnna2* (blue - gene body at the top) \pm 0.5 Mb (chr6:76,379,637-78,481,703) with corresponding transcription signal (blue - at the bottom) from nascent RNA in the parental cell line NXP010. Red vertical lines indicate the location of the two sgRNAs used to target the promoter area of *Ctnna2* and delete the regulatory elements. sgRNA_ *Ctnna2*_pe_ds1 is located at chr6:77,962,891-77,962,910 and sgRNA_ *Ctnna2*_pe_up1 at chr6:77,984,629-77,984,648. The promoter and enhancer deletion spans across 22 kb. Scale bar at the bottom left indicates 500 kb..... 27

Figure 8: Workflow for PCR validation and Sanger sequencing of positive clones after the second CRISPR/Cas9 cutting. After deletional joining the yellow and blue part of the genome were merged and the proximal promoter area of *Ctnna2* was removed. The deletion was validated by using the PCR primers indicated as purple arrows, which produces a PCR product with a length of 510 bp. Positive clones have been additionally analyzed using Sanger sequencing. The result of the selected promoter and enhancer deleted clones (ape/ Δ 1 and ape/ Δ 2) are shown at the bottom as sequencing peaks and maintained sequence (yellow and blue) compared to the deleted original sequence (black) from the parental cell line (a/b)..... 27

Figure 9: TaqMan-qPCR validates the deletion of the first allele in the a/ Δ cell line and the promoter proximal area on the second allele in the ape/ Δ clones. Left: Schematic representation of the *Ctnna2* gene body (black) with the transcriptional direction from right to left (Transcription start site on the right and transcription termination on the left). Probe A (blue) is represented in the middle of the gene body and detects the presence of the locus. Probe B (orange) is shown around the transcription start site and detects the area that is deleted during the generation of the promoter and enhancer deleted cell lines. Right: Allelic ratio of *Ctnna2* (on the y-axis) in the experimental cell lines (on the x-axis) based on the data generated with the TaqMan probes. Alleles of the whole locus (blue - Probe A) as well as the promoter proximal area (orange - Probe B) are detected. Error bars correspond to the standard deviation of the values. n.d. = not detectable. n=2. 28

Figure 10: Transcription is abolished upon deletion of the promoter proximal area. The locus of *Ctnna2* \pm 0.5 Mb (chr6:76,379,637-78,481,703) with *Ctnna2*'s gene body is shown on the top. GRO-seq signal corresponding to active transcription is shown in dark blue on the minus strand for the different samples. From top to bottom: Two replicates of the parental cell line (a/b) shown at the top. Underneath, two replicates from the founder cell lines (a/ Δ) are shown. The lower

samples are two replicates for each promoter and enhancer deleted cell line with *ape/Δ 1* and *ape/Δ 2* for the last two rows. Scale bar at the bottom right indicates 500 kb. *n*=2.29

Figure 11: Reduced expression of *Ctnna2* in the experimental promoter and enhancer deleted clones. Cell lines and primer locations are illustrated on the x-axis and the expression fold change of *Ctnna2* is on the y-axis. Each bar is labeled with its value and standard deviation shown as error bars. The two clones with the deletion at the promoter proximal area show a significant reduction in terms of gene expression of *Ctnna2*.31

Figure 12: Break distribution around the locus of *Ctnna2*. The gene body of *Ctnna2* is displayed at the top in blue. On the left the breaks are assigned to the cell line labels. On the right, the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of *Ctnna2*'s gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 10,000 junctions from three replicates for each sample group. *n*=3.33

Figure 13: Break density analysis at *Ctnna2* across the experimental cell lines. On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in “Junctions per thousand per Megabase” is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (*a/b*), the founder cells (*a/Δ*), the first p/e clone (*ape/Δ #1*) and the second p/e clone (*ape/Δ #2*) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. **** represents *p* < 0.0001. ns = not significant. *n* ≥ 3.35

Figure 14: Break distribution around the locus of *Ccser1*. The gene body of *Ccser1* is illustrated at the top in blue. On the left the breaks are assigned to the cell line labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of *Ccser1*'s gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 10,000 junctions from three replicates for each sample group. *n*=3.36

Figure 15: Break density analysis at *Ccser1* across the experimental cell lines. On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in “Junctions per thousand per Megabase” is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (*a/b*), the founder cells (*a/Δ*), the first p/e clone (*ape/Δ #1*) and the second p/e clone (*ape/Δ #2*) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. **** represents *p* < 0.0001, ** indicates *p* = 0.0072 and * represents *p* = 0.0153. ns = not significant. *n* ≥ 3.37

Figure 16: Break distribution around the locus of *Grid2*. The gene body of *Grid2* is illustrated at the top in blue. On the left the breaks are assigned to the cell line

labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of Grid2's gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 10,000 junctions from three replicates for each sample group. n=3..... 38

Figure 17: Break density analysis at Grid2 across the experimental cell lines. On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in "Junctions per thousand per Megabase" is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (a/b), the founder cells (a/Δ), the first p/e clone (ape/Δ #1) and the second p/e clone (ape/Δ #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. *** represents p = 0.0001, ** indicates p = 0.0022 and * represents p = 0.0193. ns = not significant. n ≥ 3. 39

Figure 18: E/L-Repli-seq of the experimental cell lines at the Ctnna2 locus. At the top of the figure, gene bodies based on the reference genomes are illustrated with Ctnna2's locus being in the middle. Above in red the locations of the sgRNAs used to generate the different cell lines are indicated. The two replicates of the samples show variation amongst their counterpart but follow the same trend. Ctnna2-a/Δ founder cells are represented in black. The cell lines with promoter proximal deletions, Ctnna2-ape/Δ #1 and Ctnna2-ape/Δ #2, are shown in blue and in pink, respectively. * represents p < 0.05. n=2. 40

Figure 19: Whole locus deletion of Nrnx1. Gene locus of Nrnx1 (blue - gene body at the top) ± 0.5 Mb (chr17:89,531,644-91,594,802) with corresponding transcription signal (blue - at the bottom) from nascent RNA in the parental cell line NXP047. Red vertical lines indicate the location of the two sgRNAs used to target Nrnx1 and delete the whole locus of one allele. sgRNA_Nrxn1_h1 is located at chr17:90,032,402-90,032,421 and sgRNA_Nrxn1_T1 at chr17:91,097,116-91,097,135. The whole locus deletion spans across 1.1 Mb. Scale bar at the bottom right indicates 500 kb. 41

Figure 20: Workflow for PCR validation and Sanger sequencing of positive clones after CRISPR/Cas9 cutting. The Nrnx1 locus is here represented here in black. sgRNAs are guiding CRISPR/Cas9 to cut downstream and upstream of the gene body (scissors). After deletional joining the yellow and blue part of the genome are merged. The deletion can be validated by using the PCR primers indicated as purple arrows, which produce a PCR product with a length of 742 bp. Positive clones have additionally been analyzed using Sanger sequencing. The result of the selected founder cell line (a/Δ) is shown at the bottom as sequencing peaks and maintained sequence (yellow and blue) compared to the deleted original sequence (black) from the parental cell line (a/b). 42

Figure 21: Promoter-proximal deletion of Nrnx1's promoter. Gene locus of Nrnx1 (blue gene body at the top) ± 0.5 Mb (chr17:89,531,644-91,594,802) with corresponding transcription signal (blue at the bottom) from nascent RNA in the parental cell line NXP047. Red vertical lines indicate the location of the two sgRNAs used to target the promoter area of Nrnx1 and delete the regulatory elements. sgRNA_Nrxn1PrUp1 is located at chr17:91,083,495-91,083,514 and

sgRNA_Nrxn1PrDn1 at chr17:91,094,628-91,094,647. The promoter and enhancer deletion spans across 11 kb. Scale bar at the bottom left indicates 500 kb.43

Figure 22: Workflow for PCR validation and Sanger sequencing of positive clones after the second CRISPR/Cas9 cutting. After deletional joining the yellow and blue part of the genome are merged and the proximal promoter area of Nrxn1 is removed. The deletion was validated by using the PCR primers indicated as purple arrows, which produce a PCR product with a length of 645 bp. Positive clones have additionally been analyzed using Sanger sequencing. The result of the selected promoter and enhancer deleted clones (ape/ Δ 1 and ape/ Δ 2) are shown at the bottom as sequencing peaks and maintained sequence (yellow and blue) compared to the deleted original sequence (black) from the parental cell line (a/b).43

Figure 23: TaqMan-qPCR validates the deletion of the first allele in the a/ Δ cell line and the promoter proximal area on the second allele in the ape/ Δ clones. Left: Schematic representation of the Nrxn1 gene body (black) with the transcriptional direction from right to left (Transcription start site on the right and transcription termination on the left). Probe A (blue) is represented in the middle of the gene body and detects the presence of the locus. Probe B (orange) is shown around the transcription start site and detects the area that is deleted during the generation of the promoter and enhancer deleted cell lines. Right: Allelic ratio of Nrxn1 (on the y-axis) in the experimental cell lines (on the x-axis) based on the data generated with the TaqMan probes. Alleles of the whole locus (blue - Probe A) as well as the promoter proximal area (orange - Probe B) are detected. Error bars correspond to the standard deviation of the values. n.d. = not detectable. n=2.44

Figure 24: A short isoform of Nrxn1 is transcribed after deleting the promoter proximal area near the transcription start site. The locus of Nrxn1 \pm 0.5 Mb (chr17:89,531,644-91,594,802) with Nrxn1's gene body shown on the top. GRO-seq signal corresponding to active transcription is shown in dark blue on the minus strand for the different samples. From top to bottom: Two replicates of the parental cell line (a/b) shown at the top. Underneath, two replicates from the founder cell lines (a/ Δ) are shown. The lower samples are two replicates for each promoter and enhancer deleted cell line with ape/ Δ 1 and ape/ Δ 2 for the last two rows. n=2. Scale bar at the bottom right indicates 500 kb.45

Figure 25: Break distribution around the locus of Nrxn1. The gene body of Nrxn1 is illustrated at the top in blue. On the left the breaks are assigned to the cell line labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of Nrxn1's gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 30,000 junctions from three replicates for each sample group. n=3.47

Figure 26: Break density analysis at Nrxn1 across the experimental cell lines. On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in "Junctions per thousand per Megabase" is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (a/b), the founder cells (a/ Δ),

the first p/e clone (ape/ Δ #1) and the second p/e clone (ape/ Δ #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. ** represents $p \leq 0.01$. ns = not significant. $n \geq 3$ 49

Figure 27: Break distribution around the locus of Ptpm. The gene body of Ptpm is illustrated at the top in blue. On the left the breaks are assigned to the cell line labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of Ptpm's gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 30,000 junctions from three replicates for each sample group. $n=3$ 50

Figure 28: Break density analysis at Ptpm across the experimental cell lines. On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in "Junctions per thousand per Megabase" is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (a/b), the founder cells (a/ Δ), the first p/e clone (ape/ Δ #1) and the second p/e clone (ape/ Δ #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. ** indicates $p = 0.0019$. ns = not significant. $n \geq 3$ 52

Figure 29: Break distribution around the locus of Pde10a. The gene body of Pde10a is illustrated at the top in blue. On the left the breaks are assigned to the cell line labels. On the right the presence or absence of aphidicolin (APH) is indicated. Black vertical bars indicate a position of a unique double-strand break. The area inside of Pde10a's gene body is highlighted in gray. On the bottom left, the coordinates of the shown area are indicated. On the bottom right a scale bar represents 500 kb in size. The data shown has been normalized to a total of 30,000 junctions from three replicates for each sample group. $n=3$ 53

Figure 30: Break density analysis at Pde10a across the experimental cell lines. On the x-axis, the different cell lines with and without aphidicolin are shown, while on the y-axis, the break density in "Junctions per thousand per Megabase" is displayed. DMSO refers to the samples without aphidicolin, while APH indicates samples treated with aphidicolin. The parental cell line (a/b), the founder cells (a/ Δ), the first p/e clone (ape/ Δ #1) and the second p/e clone (ape/ Δ #2) are compared. Box plots are representing the minimum and maximum value with the whiskers, while the box is drawn between the 25% and 75% quantile. The mean is marked inside the box by a horizontal line. ns = not significant. $n \geq 3$ 54

Figure 31: E/L-Repli-seq of the experimental cell lines at the Nrnx1 locus. On the x-axis, the coordinates around the Nrnx1 are shown, while on the y-axis on the right neutral (bottom) to late (top) replication timing is indicated. At the top of the figure, gene bodies based on the reference genomes are illustrated with Nrnx1's locus being in the middle. Additionally, above in red the locations of the sgRNAs used for cell line generation are indicated. The two replicates of the samples show variation amongst their counterpart but follow the same trend. Nrnx1-a/ Δ founder cells are represented in black. The cell lines with promoter proximal deletions,

Nrxn1-ape/ Δ #1 and Nrxn1-ape/ Δ #2, are shown in blue and in pink, respectively. n=2.55

Figure 32: **Scheme of different DNA breaks.** A: Single-stranded DNA break, in which only the upper of the two DNA strands shows a gap. B: Two-ended DNA double-strand break, which is the “typical” double-strand break. C: Single-ended DNA double-strand break, which is formed after replisome collapse and leads to a broken off daughter strand.56

Figure 33: **Junction pattern from double-ended DSB at the induced CRISPR/Cas9 cut site.** Bait break site in extended HTGTS data. Base positions of single junctions have been extended to 10kb and split to their alignment to the + and - strand, accordingly. On the top, gene annotations are visible with the gene bodies of Eif2ak3, Tex37 and Foxi3 being visible. Underneath, DMSO and APH samples are illustrated, and the signal data of the DNA is represented on the two strands as separate data. Scale bar indicates 200 kb. The chromosome coordinates are specified (chr6:70,812,312-71,093,484).57

Figure 34: **Junction pattern from double-ended DSB at an CRISPR off-target location.** Off-target site in extended HTGTS data. Base positions of single junctions have been extended to 10kb and split to their alignment to the + and - strand, accordingly. On the top, gene annotations are visible with the gene body of Nfkb1 being visible. Underneath, DMSO and APH samples are illustrated, and the signal data of the DNA is represented on the two strands as separate data. Scale bar indicates 200 kb. The chromosome coordinates are specified (chr3:135,575,320-135,807,232).58

Figure 35: **RDC junction pattern differs from double-ended DSB.** RDC-gene Ctnna2 in extended HTGTS data. Base positions of single junctions have been extended to 10 kb and split to their alignment to the + and - strand, accordingly. On the top, gene annotations are visible with the gene bodies of Gm38836, Ctnna2 and Reg3b being visible. Underneath, DMSO and APH samples are illustrated, and the signal data of the DNA is represented on the two strands as separate data. Scale bar indicates 200 kb. The chromosome coordinates are specified (chr6:76,181,571-78,678,293).58

Figure 36: **Recurrent DNA break clusters corresponds to replication fork stalling.** Left: Illustration to indicate fork stalling under different aphidicolin (Aph) concentrations. Low levels of aphidicolin appear to have less replication fork stalling than high levels of aphidicolin treatment. The blue arrow indicates replication from left to the right and corresponds to the stalling at the blue area, while the pink is corresponding to the replication fork, which is traveling from the right to the left and has its own pink area of stalling. Right: The gene body of Ctnna2 is illustrated in green. Underneath the data for different aphidicolin (APH) concentrations (in μ M) is indicated, increasing from the top to the bottom tracks. The data tracks represent junction density (Jden) of the breaks. Blue peaks match the right-moving replication fork while the pink peaks match the left-moving fork.59

Figure 37: **Head-on collision leads to more DNA breaks at the Tenm3 locus.** The gene body of Tenm3 is illustrated in green. Underneath the data for different aphidicolin (APH) concentrations (in μ M) is indicated, increasing from the top to the bottom tracks. The data tracks represent junction density of the breaks. Blue

peaks match the right-moving replication fork, while the pink peaks match the left-moving fork. Underneath the junction data, signal of nascent RNA is plotted from GRO-seq indicating transcription on the negative strand (gray). The red curve corresponds to published High-resolution Repli-Seq and indicates which part of the shown coordinates are early (top) or late (bottom) replicating. This data was used to extract the fork directionality shown as blue and pink arrows underneath. The chromosome coordinates are indicated as chr8:47,887,297-49,273,638. 60

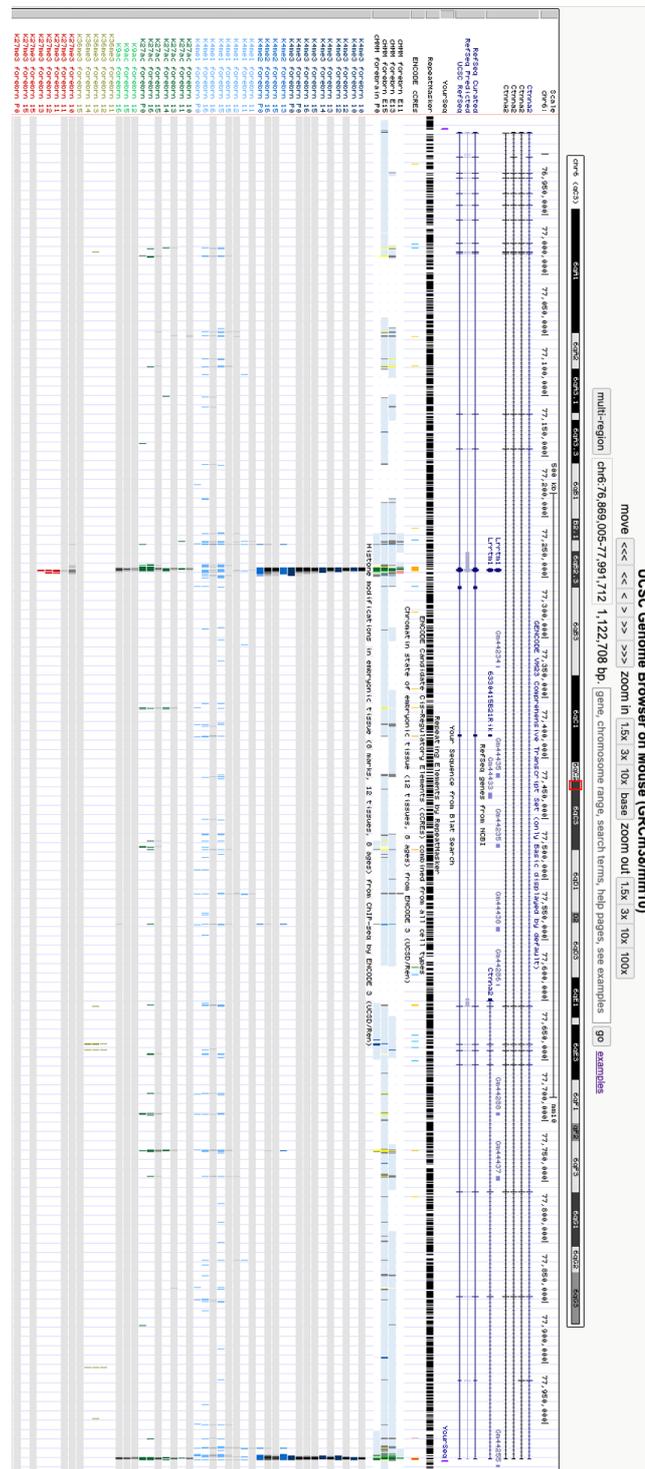
Figure 38: Genome-wide increase of DNA breaks due to head-on collisions.

A: Illustration depicting the transcription directionality in the context of the replication fork. In the case of a telomere to centromere transcription, a right-moving (blue) replication fork will collide with it in a head-on manner (“HO” white) while a left-moving (pink) replication fork will collide co-directionally (“CD” dashed). In the case of a centromere to telomere transcription direction, head-on and co-direction collision are exchanged. B: Proportions of centromeric to telomeric junctions per RDC (y-axis) were calculated and displayed according to their aphidicolin concentration (x-axis). 61

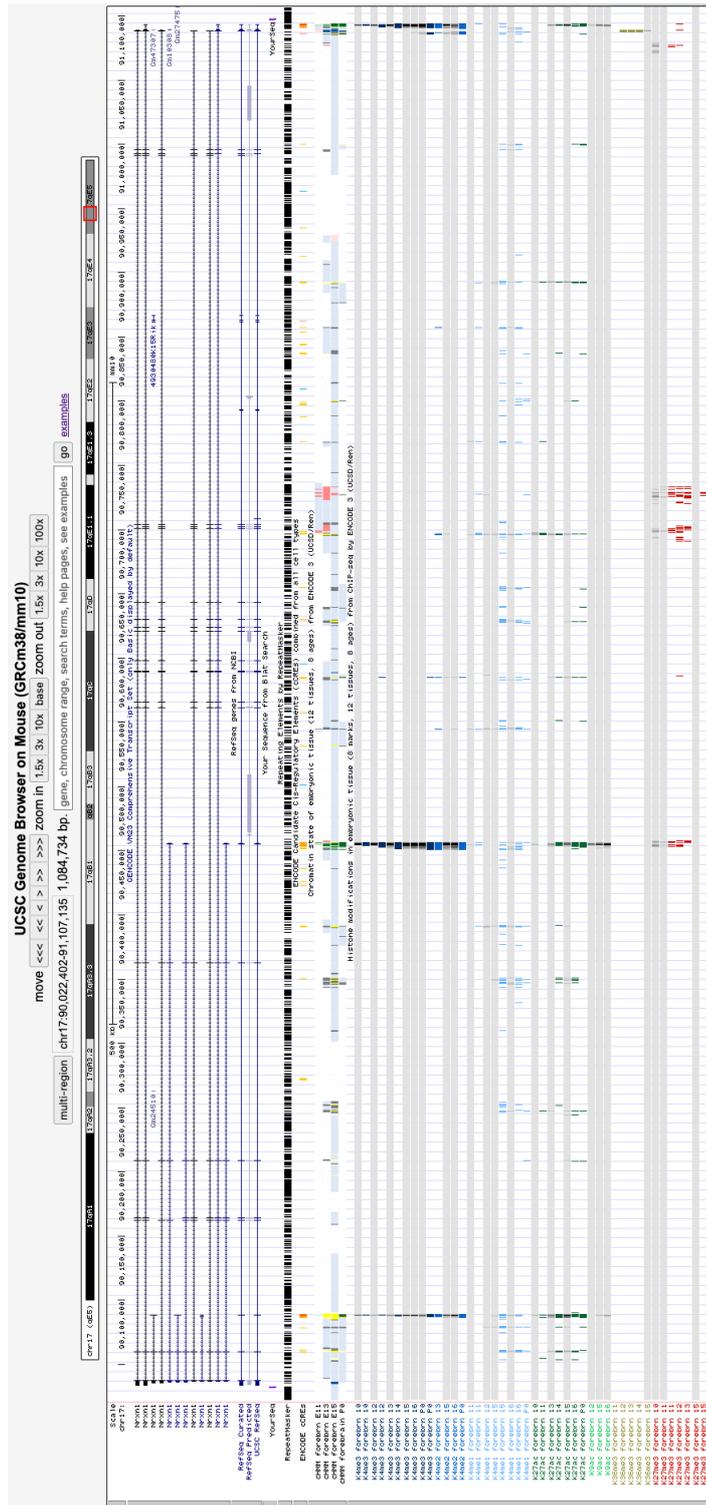
9.3 List of Tables

Table 1: **Primers used to target Ctnna2's gene body for RT-qPCR.** Name refers to the official qPCR primer names from the UCSC Genome Browser track "Mouse (mm10) Whole Transcriptome qPCR Primers" based on Zeisel A et al., *Bioinformatics* (2013).¹¹⁶. Sequences are specified for all forward and reverse primer pairs. Additionally, the primers are all designed to be intron spanning to ensure the detection of mRNA; therefore, all primers detect areas spanning from one exon to another. The exact targeted exons of Ctnna2 are indicated here for each primer pair.31

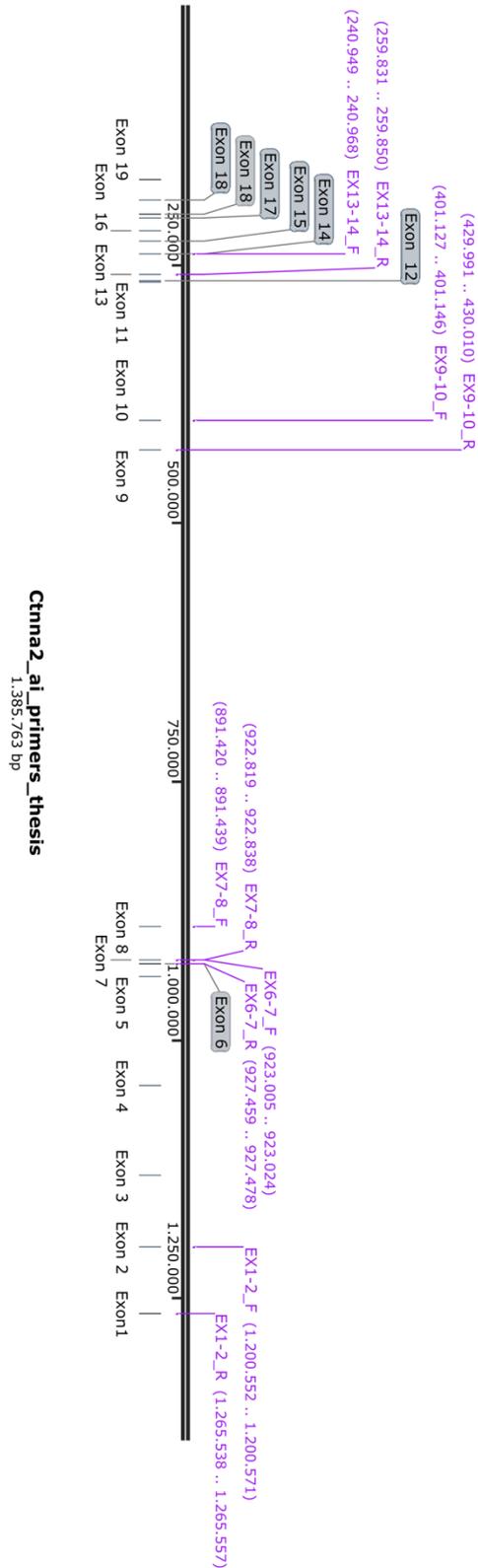
9.4 Supplementary Figures



Supplementary Figure 1: **UCSC Genome Browser view of the *Ctnna2* locus and the sgRNA positions.** *Ctnna2* \pm 10kb after sgRNA positions (chr6:76,869,005-77,991,712). From the top: gene body based on UCSC and RefSeq annotations of mm10. Your Seq: Position of the two sgRNAs targeting the *Ctnna2* locus to delete one whole allele of *Ctnna2*. Bottom: Peaks of ChIP-seq data from ENCODE Regulation Tracks showing histone modifications around the regulatory elements of the gene body. Screenshot taken from <http://genome.ucsc.edu>.



Supplementary Figure 4: **UCSC Genome Browser** view of the *Nrxn1* promoter and the sgRNA positions. *Nrxn1* promoter area ± 10 kb after sgRNA positions (chr17:91,073,491-91,104,666). From the top: gene based on UCSC and RefSeq annotations of mm10. Your Seq: Position of the two sgRNAs targeting the *Nrxn1* promoter to delete promoter-proximal regulatory elements of *Nrxn1*. Bottom: Peaks of ChIP-seq data from ENCODE Regulation Tracks showing histone modifications around the regulatory elements of the gene body. Screenshot taken from <http://genome.ucsc.edu>.



Supplementary Figure 5: **Primers and location at the Cttna2 locus.** The black horizontal line corresponds to the gene body of Cttna2 (transcribed from right to left) with a total length of 1.385.763 bp as indicated at the bottom. Across the gene, a total of 19 exons are indicated with vertical lines and labeled accordingly. Additionally, the used primers are indicated here in purple with their name referring to the exons they target.

9.5 Supplementary Tables

Number	Library number	Seq run	Cell line	Treatment	Experiment	Bait chromosome	Junction reads	Total reads	Junction reads/total reads	Result junction	Result junction/total reads
1	LC0021	B400_005	Parental	DMSO	1	chr6	181645	1094559	17%	48683	4%
2	LC0020	B400_005	Parental	APH	1	chr6	145080	822151	18%	57298	7%
3	LC0024	B400_005	Parental	DMSO	2	chr6	46143	588402	8%	17446	3%
4	VI021	B400_006	Parental	APH	2	chr6	102038	1497687	7%	46736	3%
5	VI020	B400_006	Parental	DMSO	3	chr6	98540	1578555	6%	29340	2%
6	VI028	B400_006	Parental	APH	3	chr6	133350	1537928	9%	60282	4%
7	VI035	B400_007	Parental	DMSO	4	chr6	106315	1980674	5%	32188	2%
8	VI036	B400_009	Parental	APH	4	chr6	36451	947589	4%	19362	2%
9	PW055	B400_007	Parental	DMSO	5	chr6	28750	443222	6%	12524	3%
10	PW056	B400_008	Parental	APH	5	chr6	87794	1292959	7%	39348	3%
11	VI047	B400_011	Parental	DMSO	6	chr6	63095	1027273	6%	22183	2%
12	VI048	B400_011	Parental	APH	6	chr6	97754	1355515	7%	43712	3%
13	VI053	B400_012	Parental	DMSO	7	chr6	112518	1746268	6%	30766	2%
14	VI054	B400_012	Parental	APH	7	chr6	79805	1692231	5%	33228	2%
15	VI059	B400_015	Parental	DMSO	8	chr6	31264	335653	9%	12016	4%
16	VI060	B400_015	Parental	APH	8	chr6	49298	421990	12%	24426	6%
17	LC0027	B400_005	Founder	DMSO	1	chr6	28140	564396	5%	11906	2%
18	VI029	B400_006	Founder	APH	1	chr6	121616	1669280	7%	52817	3%
19	VI022	B400_006	Founder	DMSO	2	chr6	102187	1695635	6%	31345	2%
20	VI023	B400_006	Founder	APH	2	chr6	115063	2081002	6%	37067	2%
21	VI037	B400_007	Founder	DMSO	3	chr6	125528	1729468	7%	39195	2%
22	VI038	B400_007	Founder	APH	3	chr6	70559	1185004	6%	32403	3%

Supplementary Table 1: **Ctnna2** HTGTS libraries – part 1. Libraries generated for the investigation regarding break formation at the *Ctnna2* locus. Libraries starting with LC have been prepared by Lorenzo Corazzi, PW libraries by Dr. Pei-Chi Wei and VI libraries by myself, Vivien Ionasz. Further columns represent the sequencing run, cell line, treatment, and further experimental information.

Number	Library number	Seq run	Cell line	Treatment	Experiment	Bait chromosome	Junction reads	Total reads	Junction reads/total reads	Result junction	Result junction/total reads
23	PW057	B400_007	Founder	DMSO	4	chr6	36891	467648	8%	16082	3%
24	PW058	B400_007	Founder	APH	4	chr6	32044	433590	7%	17346	4%
25	VI043	B400_010	Founder	DMSO	5	chr6	127106	1015119	13%	43643	4%
26	VI044	B400_010	Founder	APH	5	chr6	111527	952843	12%	45562	5%
27	VI049	B400_011	Founder	DMSO	6	chr6	58640	747338	8%	21775	3%
28	VI050	B400_011	Founder	APH	6	chr6	103081	942051	11%	47924	5%
29	VI055	B400_012	Founder	DMSO	7	chr6	110184	1427447	8%	32100	2%
30	VI056	B400_012	Founder	APH	7	chr6	68986	1031250	7%	29304	3%
31	LC0025	B400_005	ape clone #1	DMSO	1	chr6	31790	468517	7%	12485	3%
32	VI034	B400_006	ape clone #1	APH	1	chr6	308619	1640973	19%	115121	7%
33	VI057	B400_012	ape clone #1	DMSO	2	chr6	169849	1592608	11%	47472	3%
34	VI058	B400_012	ape clone #1	APH	2	chr6	188718	1513709	12%	70522	5%
35	VI065	B400_015	ape clone #1	DMSO	3	chr6	61405	409061	15%	22616	6%
36	VI066	B400_015	ape clone #1	APH	3	chr6	61845	393729	16%	26301	7%
37	LC0026	B400_005	ape clone #2	DMSO	1	chr6	14726	460075	3%	6965	2%
38	VI033	B400_006	ape clone #2	APH	1	chr6	54293	1336650	4%	24666	2%
39	VI045	B400_010	ape clone #2	DMSO	2	chr6	102078	1129285	9%	35444	3%
40	VI046	B400_010	ape clone #2	APH	2	chr6	58478	738845	8%	28346	4%
41	VI051	B400_011	ape clone #2	DMSO	3	chr6	104484	1429375	7%	31343	2%
42	VI052	B400_011	ape clone #2	APH	3	chr6	69648	1045764	7%	34217	3%
43	VI063	B400_015	ape clone #2	DMSO	4	chr6	41945	444803	9%	14925	3%
44	VI064	B400_015	ape clone #2	APH	4	chr6	54693	549137	10%	24428	4%

Supplementary Table 2: **Ctnna2 HTGTS libraries – part 2**. Libraries generated for the investigation regarding break formation at the *Ctnna2* locus. Libraries starting with LC have been prepared by Lorenzo Corazzi, PW libraries by Dr. Pei-Chi Wei and VI libraries by myself, Vivien Ionasz. Further columns represent the sequencing run, cell line, treatment, and further experimental information.

Number	Library number	Seq run	Cell line	Treatment	Experiment	Bait chromosome	Junction reads	Total reads	Junction reads/total reads	Result junction	Result junction/total reads
1	PW045	B400_005	Parental	DMSO	1	chr17	45745	814653	6%	16857	2%
2	PW046	B400_005	Parental	APH	1	chr17	32872	1071983	3%	14615	1%
3	VI085	B400_022	Parental	DMSO	2	chr17	36955	1364768	3%	11301	1%
4	VI086	B400_022	Parental	APH	2	chr17	50683	1932909	3%	18267	1%
5	VI087	B400_022	Parental	DMSO	3	chr17	39531	1686926	2%	10454	1%
6	VI088	B400_022	Parental	APH	3	chr17	25885	1636396	2%	8323	1%
7	PW010	B400_001	Founder	DMSO	1	chr17	29276	1111400	3%	10305	1%
8	PW009	B400_001	Founder	APH	1	chr17	34370	1200375	3%	14122	1%
9	PW047	B400_005	Founder	DMSO	2	chr17	34879	864497	4%	11771	1%
10	PW048	B400_005	Founder	APH	2	chr17	39107	824152	5%	17853	2%
11	PW079	B400_010	Founder	DMSO	3	chr17	28795	963651	3%	10186	1%
12	PW080	B400_010	Founder	APH	3	chr17	25775	885547	3%	12511	1%
13	VI067	B400_020	Founder	DMSO	4	chr17	120006	1554079	8%	29898	2%
14	VI072	B400_020	Founder	APH	4	chr17	181459	1358145	13%	38910	3%
15	VI068	B400_020	Founder	DMSO	5	chr17	106965	1517701	7%	29154	2%
16	VI073	B400_020	Founder	APH	5	chr17	90181	850236	11%	30659	4%
17	VI077	B400_021	Founder	DMSO	6	chr17	14900	427039	3%	5201	1%
18	VI078	B400_021	Founder	APH	6	chr17	27397	487199	6%	15388	3%

Supplementary Table 3: **Nrxn1 HTGTS libraries – part 1**. Libraries generated for the investigation regarding break formation at the *Nrxn1* locus. Libraries starting with PW libraries have been prepared by Dr. Pei-Chi Wei and VI libraries by myself, Vivien Ionasz. Further columns represent the sequencing run, cell line, treatment, and further experimental information.

Number	Library number	Seq run	Cell line	Treatment	Experiment	Bait chromosome	Junction reads	Total reads	Junction reads/total reads	Result Junction	Result Junction/total reads
19	PW086	B400_012	ape clone #1	DMSO	1	chr17	32452	1177578	3%	11510	1%
20	PW087	B400_012	ape clone #1	APH	1	chr17	15426	1550784	1%	3702	0%
21	VI069	B400_020	ape clone #1	DMSO	2	chr17	112221	1567196	7%	29568	2%
22	VI074	B400_020	ape clone #1	APH	2	chr17	177816	991352	18%	36211	4%
23	VI070	B400_020	ape clone #1	DMSO	3	chr17	139232	1895266	7%	29593	2%
24	VI075	B400_020	ape clone #1	APH	3	chr17	262747	1379624	19%	40266	3%
25	VI079	B400_021	ape clone #1	DMSO	4	chr17	22531	538510	4%	7832	1%
26	VI080	B400_021	ape clone #1	APH	4	chr17	21076	520591	4%	11456	2%
27	PW014	B400_001	ape clone #2	DMSO	1	chr17	29137	1148735	3%	10281	1%
28	PW013	B400_001	ape clone #2	APH	1	chr17	32926	1194902	3%	12928	1%
29	VI071	B400_020	ape clone #2	DMSO	2	chr17	147104	1961135	8%	28315	1%
30	VI076	B400_020	ape clone #2	APH	2	chr17	288324	1659110	17%	36743	2%
31	VI081	B400_021	ape clone #2	DMSO	3	chr17	20348	477198	4%	6759	1%
32	VI082	B400_021	ape clone #2	APH	3	chr17	31422	436410	7%	17052	4%
33	VI083	B400_021	ape clone #2	DMSO	4	chr17	37105	719706	5%	11382	2%
34	VI084	B400_021	ape clone #2	APH	4	chr17	43850	597290	7%	19585	3%

Supplementary Table 4: **Nrxn1 HTGTS libraries – part 2.** Libraries generated for the investigation regarding break formation at the Nrxn1 locus. Libraries starting with PW libraries have been prepared by Dr. Pei-Chi Wei and VI libraries by myself, Vivien Ionasz. Further columns represent the sequencing run, cell line, treatment, and further experimental information.

10 Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Pei-Chi Wei, for giving me the possibility to work on this fascinating project and letting me do my PhD in her lab. I learned countless things in the last three and a half years in the lab. Starting with DNA breaks, stem cell culture, neuroscience, bioinformatics, scientific collaboration, as well as myself. I really appreciate the guidance that I received during the course of my research.

Additionally, I want to thank my whole group, B400, for providing me with a nice and collaborative research atmosphere on the daily, I will always cherish our great time. Marco, who always made sure that the lab is up and running by ordering all the reagents and organizing our space. Lorenzo, who helped me out with cell culture and supported the core of this project from a different angle. Giulia, that always provided an open ear and helping hands whenever it was necessary, I am glad to have found a friend in you. Thank you for all the tasty Italian food over the years, that basically fueled this research. And also, Boyu for all of our in-detail discussions about molecular mechanisms and letting me teach you about lab work while you teach me about Chinese cuisine.

My dear friend Yassin, who listened to all of my ideation processes, presentations and complaints over the years and read all my reports as well as this thesis. Thank you for your valuable input and uplifting feedback. The only person that I was able to meet for almost two months due to the lockdown and never failed to show up for me ever since. Our coffee breaks were crucial for me managing to finish this PhD.

My partner Jonas, which I found solely because of this degree. We first interacted about me being equally fascinated and envious that you already had a finished Master's thesis while I was still writing mine (almost exactly four years ago). I am glad that we both went to that ice hockey game in January 2020 in Mannheim. You helped me stay sane on a daily basis for already more than 3 years. I would not have wanted (or managed) to do my PhD simultaneously with any other partner. It was a real challenge, but I am happy that we got to go on this journey together. I am very much looking forward to the future that we will build together.

I am also incredibly thankful to my parents and bonus parents, which made it possible for me to be in the position to pursue a doctoral degree and move to Heidelberg. You always supported me and my ideas, even when they meant that I will be moving to Iceland for 5 months without ever having been there. I am happy to finally finish my formal education after 22 years and cannot express how much I appreciate the hard work you put in for me every single day. I hope that one day I will be able to be as strong and courageous as you are. I love you, thank you for always trying your best!

My best friend Petra, which always was by my side, no matter the 580 km distance between us. I am grateful that we managed to grow up together and made it through different stages of our life while never losing our connection. You are truly my north star bestie and I know I can always count on you, no matter how much time has passed. I hope I am and will be able to give you the same amount of support for your studies as you did for me. To many more years of friendship and growth.

All my other friends and family, who I cannot all mention by name here, that have always supported me, believed in me, and helped me study and come to an end with this chapter of my life. Your endorsement to just try and do the next step almost made me feel like the sky is the limit and made my educational path a straight line. You all are an amazing support network, and I could not have done it without you.

Maia, îți mulțumesc foarte mult că m-ai crescut. Încă îmi povestești despre momentul în care încă mă puteai ține în mâini, pentru că eram doar un bebeluș. M-ai văzut cu adevărat crescând sub ochii tăi și m-ai acompaniat de la grădiniță până la universitate. Am petrecut atât de mult timp împreună și mă bucur foarte mult că am făcut-o. Îmi pare incredibil de rău că a trebuit să te părăsesc când am venit la Heidelberg pentru a-mi continua doctoratul și că nu ne-am văzut prea mult în ultimii ani. Sper că am reușit să te fac să fii mândră și că mai avem încă ceva timp împreună.

