From the Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

DNA Fragility in the Context of Neural Stem Cell Fate: A Multi-Method Integrative Exploration of Genome Dynamics

Roberto Ballarino

Karolinska Institutet

Stockholm 2022

The cover picture is a representation of the different topics intersecting one another forming the core to my curiosity about brain development and health. Every problem can be scaled to different levels of resolution and with each step taken, the complexity increases. There is no development without flaws, no change without movement and no progress without boldness. This image was wonderfully crafted by Gabriela Stumberger after sharing my initial sketches.

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB.

ISBN 978-91-8016-773-4

DNA Fragility in the Context of Neural Stem Cell Fate: A Multi-Method Integrative Exploration of Genome Dynamics

THESIS FOR DOCTORAL DEGREE (Ph.D.)

To be publicly defended at Eva & Georg Klein hall at Biomedicum,
Karolinska Institutet science park, Solnavägen 9, 171 65 Solna, Stockholms län, Sweden

On November 8th, 2022 at 9:30 AM

By:

Roberto Ballarino

Principal Supervisor:
Assistant Professor Nicola Crosetto
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Genome Biology

Co-supervisor(s):
Professor Marie Carlén
Karolinska Institutet
Department of Neuroscience

Professor Anna Falk
Lund University
Department of Experimental Medical Science
Division of Neural Stem Cells

Associate Professor Magda Bienko
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Dr. Federico Agostini
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Genome Biology

Opponent:
Associate Professor Pelin Sahlén
KTH Royal Institute of Technology
Department of Gene technology

Examination Board:
Professor Neus Visa
Stockholm University
Department of Molecular Biosciences

Associate Professor Qi Dai
Stockholm University
Department of Molecular Biosciences

Associate Professor Anita Göndor
Karolinska Institutet
Department of Oncology-Pathology

Dedicated to family, friends, colleagues and my wonderful wife Serena
for their love and support throughout this journey

"Research is like a voyage of discovery into unknown lands, seeking not for new territory but for new knowledge. It should appeal to those with a good sense of adventure."

- Frederick Sanger

ABSTRACT

Recent advances in mapping the complex genetic architecture underlying various debilitating brain disorders have enabled identification of several genetic risk variants. However, these risk variants only explain part of the heritability and vulnerability to these disorders in early development. Moreover, de novo somatic mutations have been detected in subsets of brain cells, which might account for a significant portion of the missing heritability. However, it remains unclear where these mutations come from and at what developmental stage they might occur.

Genome fragility is subject to the functional activity and spatial chromatin organization characteristic of a distinct cell identity. Under physiological conditions, cells regulate their chromatin structure and organization to express necessary genes. DNA topoisomerases are a key player in all of these processes and in replication. Through generation of transient breaks in the DNA, topoisomerases are able to resolve topological problems and thereby activation of particular sections of the genome. Beyond topoisomerases, the genome is subject to perpetual challenges with DNA double-strand breaks (DSBs) being among the most deleterious. Each cell is estimated to suffer numerous transient DSBs per day, most of which are repaired. Incorrectly repaired DSBs however, pose a major threat to genome stability through formation of mutations or potential genomic rearrangements. Although the exact relationship of DNA damage to differentiation is still unclear, a recent investigation into neural specification demonstrated that loss of DNA repair sensors leads to centrosome amplification, thereby resulting in defective mitosis and chromosomal instability. Ensuing excessive stem cell proliferation and replication stress also happen to be a hallmark of neurodevelopmental disorders (NDDs). Despite the emerging evidence linking endogenous DSBs to NDDs, there has been a lack of genome-wide maps of DSBs spontaneously arising at different stages of human neurogenesis.

This thesis brings together (I) a correlative genomics study describing endogenous DSBs genome-wide during neural differentiation in a cell-type specific manner, and (II) a mechanistic study into the regulatory role of Topoisomerase 1 (TOP1) in transcription and proliferation.

In paper I, we mapped the genomic DSB landscape of cells at various stages of neural differentiation and correlated our maps with genomic and epigenomic features. In so doing, we provide clues on how DSB formation and their incorrect repair might contribute to the pathogenesis of NDDs. The current view is that transcription-associated DSBs seem to be the main driver of de novo mutations. Indeed, we found that DSBs preferentially form around the transcription start site (TSS) of transcriptionally active genes, as well as at chromatin loop anchors in proximity of highly transcribed genes. This follows from the accumulation of DNA torsional stress and topoisomerase activity in these regions. Interestingly, hotspots of endogenous DSBs were detected around the TSS of highly transcribed genes involved in general cellular processes and along the gene body of long, neural-specific genes whose human orthologues had been previously implicated in NDDs. Through our integrative multi-method approach we corroborate previous findings regarding DSB-fragile loci at TSSs and loop anchors, and find a unique distribution pattern for this fragility in post-mitotic neurons.

We show a cell type-specific preference for DSB accumulation in specific NDD genes and begin to describe the relation of DSB fragility and chromatin conformation.

In paper II, we investigated the role of Topoisomerase I (TOP1) in relation to transcription in the context of replication stress across mitosis and as subject of interruption of interphase chromatin conformation. In particular, we investigated different stages of the cell cycle for transcription patterns and transcriptional spiking by RNA polymerase II (RNAPII) in human colon carcinoma cells. TOP1 relieves torsional stress in actively transcribed DNA and facilitates the expression of long genes, many of which are important for neural functions. However, TOP1 also plays a direct role in transcriptional control through interaction with RNAPII Carboxy-Terminal Domain (CTD). We investigated control cells and a knock-in (KI) clone lacking TOP1 exon4, the phosphor-CTD-binding site for RNAPII. We found that in early mitosis TOP1 clears RNAPII during transcriptional elongation. When the TOP1 CTD-binding domain is disrupted, we detected replication stress and delay in mitotic exit. In this case, chromatin becomes topologically stressed, increasing the need for TOP2A cleavage resulting in DSBs. However, we did not detect substantial changes in DSB markers gamma-H2AX and 53BP1 when comparing WT and KI cells across different stages of the cell cycle. Therefore, we conclude that the observed delay in mitotic exit is most likely due to the deregulation of gene expression, rather than to the activation of DNA repair pathways. Acute depletion of TOP1 through the auxin-degron system resulted in absence of RNAPII spiking at the TSS. Efficient removal of RNAPII from chromosomes by TOP1 in early mitosis is both a prerequisite for the timely spike of RNAPII at TSSs in mid mitosis and might affect cellular memory. Indeed, we found that when mitotic transcription is poorly regulated, individual proliferating cells have a greater variance in transcriptional levels and thus could lead to loss of cell identity.

Concluding from these findings, we demonstrate that endogenous DSBs are distributed differentially in a cell type-specific manner. Through our integrative multi-method approach, we corroborate previous findings regarding DSB-fragile loci and discovered a unique distribution pattern for DSBs in post-mitotic neurons. We show a preference for specific NDDs genes and begin to describe the relation of DSB fragility and chromatin conformation in a developmental context. We assessed the role of TOP1 in a model for replication stress and found that outside of its canonical torsional stress function, the direct interaction with RNAPII across the cell cycle is crucial in maintaining transcriptional memory and could feed into loss of cell identity.

While not exhaustive, the findings described in these papers begin to elucidate a complex mystery of human NDDs and provide valuable datasets for further investigation of genome fragility. Taken together, these findings contribute to a better understanding of how neural genome dynamics affect high transcriptional or replicative burden during neurodevelopment.

POPULAR SCIENCE SUMMARY OF THE THESIS

Catching mutations at the right time: It might all be in your head

A broad range of neurodevelopmental disorders of previously unexplained cause might be the result of damage to the DNA occurring during normal brain development. Historically, brain disorders have always been screened for and categorized based on the inherited genetic material, since disorders like schizophrenia and autism spectrum disorder typically run in families. By studying families, the predictive value of specific genetic variations has been estimated and these gene lists are continuously further refined. Nonetheless, over two-thirds of brain disorders remain a mystery. With our study we aim to shed light on the origin of these poorly understood disorders.

Parts of the brain might be differentially targeted by genetic changes, with a subset of brain cells accumulating these changes during early development. A more commonly studied example of many cells containing different sets of DNA can be seen in the study of cancer, where having multiple genetically different cells within a tumor is referred to as tumor heterogeneity: As the tumor cells divide, sub-populations within the tumor collect errors, making some of them less susceptible to treatment. There are many ways of accumulating genetic changes. In cancer, replication stress is one of the major drivers towards genetic changes resulting in accelerated growth and evasion of the immune system. While brain cells are not actively dividing later in life, prenatal stem cells are. These stem cells progress through a rather streamlined process of maturation toward their final form, while ultimately every neuron in the brain ends up being unique and indispensable for brain function. The genetic changes that accumulate in early brain stem cells might have large consequences for the role and function of future neurons. Faulty cell identity assignment or death of these cells can result in loss of specific brain functions.

This thesis presents my contribution to identifying which parts of the DNA are particularly prone to accumulate breaks that precede disease-relevant genetic changes. In addition, it shows how DNA activity or conformation in 3D space of the cells' nucleus might affect the location of DNA break enriched sites. It lays out how temporary DNA breaks could result in lasting genetic changes that have a predictive value to brain disorders. We assess if there is a critical window for vulnerability to breaks during early development. Finally, we investigate if loss of the regulation of gene activity could cause loss of cell identity through gene activity programs or DNA breaks and could be clinically significant.

In the first study "An atlas of endogenous DNA double-strand breaks…" (Paper I), we set out to describe the genetic landscape of early brain cell development. Through implementation of state-of-the-art methods, we investigated DNA fragility, activity and 3D organization. This is one of the first major studies describing genome fragility and the process of DNA breakage in absence of perturbations of neural cell development. The fact that we studied development without including any perturbation is an important detail, because the quantitative nature of DNA damage can be affected through changes in environment or suppression of repair. We took a snapshot at three specific timepoints in the streamlined neural developmental timeframe outlined earlier. Each timepoint was chosen to represent

specialization milestones reached by the cells. Rapidly replicating stem cells (1), primed progenitor cells (2) and terminally specialized neurons (3). Experimentally, we tagged and identified loose DNA ends for each developmental timepoint and associated them with the locally corresponding DNA activity and spatial conformation of the DNA inside the nucleus. By generating an atlas of DNA breaks across the genome for each of the three developmental milestones, we describe a general and genome-wide tendency of DNA breaks to occur at highly active transcription sites and their regulating promotor region. We found that neurons are unique as a consequence of their 3D DNA conformation and significantly stand out from the preceding proliferating cell types in terms of their DNA break distributions. Taken together, our datasets describe many interrelated processes, but do not reveal any direct mechanistic causation.

In the second study "Topoisomerase 1 activity during mitotic transcription…" (Paper II), we focused on the process of DNA activity regulation by DNA-nicking enzyme Topoisomerase 1 (TOP1), which makes temporary breaks in the DNA in the context of cell division and replication stress. We discovered that TOP1 regulates DNA activity directly by binding the key enzyme in RNA production called RNA Polymerase II (RNAPII). When disrupting the interaction between TOP1 and RNAPII, we found cell division was delayed and noticed effects of replication stress. By eliminating TOP1 in healthy cells, we noticed an RNAPII misplacement similar to the mutant cell line. We conclude that absence of TOP1 directly causes destabilization of gene activity programs, loss of cellular memory and thus loss of cell identity.

Concluding from these studies, we show that DNA breaks occur naturally and as a consequence of a particular cell state or identity. We see that the 3D organization and DNA activity of a particular cell allows us to predict fragile DNA break sites. DNA breaks accumulate around the gene activation sites and their promotor areas. We discovered a new regulatory role of TOP1 in these same areas and in maintaining cellular memory across replication. However, we did not find a global increase in DNA damage in absence of TOP1. Taken together, these findings contribute to a better understanding of what happens inside the nucleus of a cell, how DNA is regulated and structured and finally, how perturbation of these processes during development could result in debilitating brain disorders.

Populärvetenskaplig sammanfattning (SWE)

Att hitta mutationer i tid: Allt sitter i ditt huvud

Ett brett spektrum av tidigare oförklarliga neuropsykiatriska störningar kan härröra från skador på DNA som uppstår under normal hjärnutveckling. När det gäller hjärnsjukdomar så screenar och kategoriserar kliniker numera dem utifrån det ärftliga genetiska materialet, eftersom sjukdomar som schizofreni och autismspektrumstörning tenderar att ärvas i släkter. Genom att studera släkter har det prediktiva värdet av specifika genetiska variationer för vissa hjärnsjukdomar uppskattats och dessa listor på gener visar det prediktiva värdet av varje gen eller genetisk koordinat för en hjärnsjukdom. Ändå är ursprunget till mer än två tredjedelar av psykiatriska och utvecklingsstörningar ett mysterium. Vi vill med våra studier försöka hitta möjliga förklaringar till varför de icke-ärftliga sjukdomarna uppstår.

Delar av hjärnan kan utsättas för genetiska förändringar på olika sätt, där en undergrupp av hjärnceller ackumulerar mutationer under tidig utveckling. Ett allmänt studerat exempel på många celler som innehåller olika sammansättningar av DNA är cancerforskning. Det kan finnas genetiskt olika celler i en tumör. När tumörcellerna delar sig, ackumulerar vissa subpopulationer i tumören gradvis fel, vilket gör dem mindre mottagliga för behandling. De ackumulerade genetiska mutationerna har uppstått till följd av olika orsaker. I tumörer är replikationsstress vid celldelning en av de viktigaste orsakerna till genetiska förändringar som leder till accelererad tillväxt och förändrad cellroll. I likhet med tumörer delar sig prenatala stamceller ofta och snabbt, därför är de känsliga för replikationsstress och mutationer. Dessa stamceller går igenom en strömlinjeformad utveckling till sin slutliga form och roll, de delar sig då inte längre. De ackumulerade mutationer som uppstått i det slutliga skede då de prenatala stamcellerna utvecklats till färdiga neuron kommer att ha stor betydelse för neuroners roll och funktion. Felaktig rolltilldelning eller död av dessa celler leder till förlust av specifika hjärnfunktioner.

Denna avhandling presenterar mitt bidrag till att kartlägga vilka delar av DNA:t som är särskilt mottagliga för ackumulering av skador som föregår mutationer. Avhandlingen visar hur processer som genaktivitet eller rumslig organisering av DNA i cellkärnan påverkar känsligheten för skador. På vilka sätt kan de tillfälliga skadorna i DNA:t leda till permanenta genetiska mutationer som har ett prediktivt värde för hjärnsjukdomar? Finns det ett specifikt kritiskt utvecklingsstadium där skadekänsligheten förändras avsevärt? Och slutligen, om störningar av regleringen av genaktivitet kan orsaka förlust av cellroll, med konsekvenser för hjärnans utveckling.

I den första studien "An atlas of endogenous DNA double-strand breaks…" (Paper I), ville vi beskriva det genetiska landskapet för tidig hjärncellsutveckling. Genom att implementera moderna metoder undersökte vi DNA-bräcklighet, DNA-aktivitet och rumslig organisering av DNA. Det här är en av de första större studierna som beskriver genomets bräcklighet och processen för DNA-skador hos neuroner som utvecklats i frånvaro av störningar. Det är värt att notera att vi studerade cellutveckling i en kontrollerad miljö och med fullt fungerade DNA-reparationssystem, eftersom frånvaron av dessa faktorer kan påverka både antalet och fördelningen av DNA-avbrott. Vi skapade en ögonblicksbild vid tre tidpunkter i det strömlinjeformade neurala utvecklingsförloppet som nämndes tidigare. Varje tidpunkt valdes

för de specifika milstolpar som uppnåtts i specialiseringen av cellrollen. Snabbt delande och självförnyande stamceller (1), specialiserande progenitorceller (2) och färdiga neuroner (3). Experimentellt märkte och identifierade vi lösa DNA-ändar för varje utvecklingstid och associerade dem med DNA-aktiviteten och den rumsliga organiseringen av DNA:t i cellkärnan. Genom att sammanställa en atlas av DNA-avbrott över hela DNA:t för varje milstolpe i utvecklingen, finner vi en allmän tendens att skador på DNA inträffar på platser med hög genaktivitet och i promotorregionen som reglerar genaktiviteten. Vi fann att neuroner är unika på grund av deras DNA-organisering och skiljer sig avsevärt i brotthastighet från stamceller och progenitorceller, som fortfarande delar sig. Sammantaget beskriver våra data många inbördes relaterade processer, men visar ingen direkt mekanistisk kausalitet mellan dessa processer.

I den andra studien "Topoisomerase 1 activity during mitotic transcription…" (Paper II), fokuserade vi på regleringen av DNA-aktivitet genom DNA-klyvningsenzymet Topoisomerase 1 (TOP1). TOP1 orsakar tillfälliga skador i DNA:t, något som förstärks ytterligare i samband med celldelning och replikationsstress. Vi fann att TOP1 direkt reglerar DNA-aktivitet genom att binda till nyckelspelaren i RNA-produktion, RNA Polymerase II (RNAPII). Genom att störa interaktionen mellan TOP1 och RNAPII noterade vi ökad replikationsstress, vilket i sin tur saktade ner celldelning. Om man istället eliminerar TOP1 i friska celler, observerar vi att RNAPII blir felplacerad, vilket vi även observerade i den muterade cellinjen. Vi drar slutsatsen att frånvaron av TOP1 direkt orsakar destabilisering av förväntad genaktivitet, vilket leder till förlust av cellulärt minne, och därmed kan orsaka förlust av cellroll.

Avslutningsvis visar vi i dessa studier att skador i DNA sker naturligt och som en konsekvens av ett visst celltillstånd eller cellroll. Vi ser att den rumsliga organiseringen och DNA-aktiviteten hos en viss cell tillåter oss att förutsäga sannolika platser för DNA-skador. DNA-skador ackumuleras runt platser med hög genaktivitet och även i promotorregioner som reglerar denna genaktivitet. Vi upptäckte en ny roll för TOP1: att reglera genaktivitet i tidigare nämnda högaktiva regioner. Följaktligen upprätthålls cellulärt minne under celldelning av TOP1. Vi hittade dock ingen global ökning av DNA-skador i frånvaro av TOP1. Sammantaget bidrar dessa fynd till en bättre förståelse av vad som händer i cellkärnan, hur DNA regleras och struktureras, och slutligen hur störningar av dessa processer under normal hjärncellsutveckling kan leda till hjärnsjukdomar.

POPULAIR-WETENSCHAPPELIJKE SAMENVATTING (NL)

Speuren naar mutaties tijdens de ontwikkeling: misschien zit het wel tussen de oren

Een breed scala aan neuropsychiatrische aandoeningen met een voorheen onverklaarbare oorzaak kan het gevolg zijn van schade aan het DNA die zich voordoet tijdens de normale hersenontwikkeling. In geval van hersenaandoeningen wordt in de kliniek tegenwoordig gescreend en gecategoriseerd op basis van het erfelijke genetische materiaal, aangezien aandoeningen zoals schizofrenie en autisme spectrum stoornis meestal in families overerven. Door families te bestuderen is de voorspellende waarde van specifieke genetische variaties voor bepaalde hersenaandoeningen ingeschat en worden deze lijsten vervolgens steeds verder verfijnd om de voorspellende waarde van elk gen of genetische locatie voor een hersenaandoening te bepalen. Desalniettemin blijft de oorsprong van meer dan twee derde van de psychiatrische en ontwikkelingsstoornissen een mysterie. Met ons onderzoek zoeken we mogelijke verklaringen voor het ontstaan van deze aandoeningen.

Delen van de hersenen kunnen op verschillende manieren onderhevig zijn aan genetische veranderingen, waarbij een subset van hersencellen deze mutaties tijdens de vroege ontwikkeling accumuleert. Een vaker bestudeerd voorbeeld van veel cellen die verschillende samenstellingen van DNA bevatten is het onderzoek naar kanker. Het aanwezig zijn van meerdere genetisch verschillende cellen in een tumor wordt tumorheterogeniteit genoemd: terwijl de tumorcellen zich delen verzamelen sommige subpopulaties binnen de tumor gaandeweg fouten, waardoor deze minder vatbaar worden voor behandeling. De geaccumuleerde genetische mutaties kunnen diverse oorzaken hebben. Bij kanker is replicatiestress tijdens de celdeling één van de belangrijkste oorzaken van genetische veranderingen, die leiden tot bijvoorbeeld versnelde groei en o.a. ontwijking van het immuunsysteem. Terwijl hersencellen zich later in het leven niet langer delen, staan prenatale stamcellen er juist bekend om dat ze veel en snel moeten delen. Deze stamcellen doorlopen een gestroomlijnde ontwikkeling naar hun uiteindelijke vorm en rol, terwijl uiteindelijk elk neuron in de hersenen uniek en onmisbaar wordt voor gezonde hersenfunctie. De mutaties die geaccumuleerd worden in vroege hersenstamcellen hebben potentieel grote gevolgen voor de rol en functie van de toekomstige hersencellen. Een verkeerde rol-toewijzing of de dood van deze cellen leiden tot verlies van specifieke hersenfuncties.

Dit proefschrift presenteert mijn bijdrage aan het in kaart brengen van welke delen van het DNA bijzonder vatbaar zijn voor accumulatie van breuken in het DNA die voorafgaan aan mutaties. Het laat zien hoe processen als genactiviteit of DNA-conformatie in de 3D ruimte van de celkern de locatie van kwetsbare delen van het DNA beïnvloeden. Op welke manieren de tijdelijke breuken in het DNA zouden kunnen leiden tot blijvende genetische mutaties die een voorspellende waarde hebben voor hersenstoornissen. Of er een specifieke kritieke ontwikkelingsfase is waarin de vatbaarheid voor breuken significant verandert. En tenslotte, of verstoring van de regulatie van genactiviteit verlies van cel rol zou kunnen veroorzaken, met gevolgen voor de hersenontwikkeling.

In de eerste studie "An atlas of endogenous DNA double-strand breaks…" (Paper I), wilden we het genetische landschap van vroege hersencel ontwikkeling beschrijven. Door

implementatie van moderne methodologieën hebben we DNA-fragiliteit, DNA-activiteit en 3D-organisatie van het DNA onderzocht. Dit is een van de eerste grote studies die de fragiliteit van het genoom en het proces van DNA-breuk beschrijven in afwezigheid van verstoringen van de ontwikkeling van neurale cellen. Belangrijk is dat we de celontwikkeling hebben bestudeerd zonder enige interventie, omdat de kwantitatieve aard van DNA-breuken kan worden beïnvloed door veranderingen in de omgeving of door onderdrukking van DNA-reparatie systemen. We hebben een momentopname gemaakt op drie tijdstippen in het eerder aangehaalde gestroomlijnde neurale ontwikkelingsverloop. Elk tijdstip is gekozen voor de specifiek behaalde mijlpalen in de specialisatie van de cel rol. Snel delende en self-vernieuwende stamcellen (1), zich specialiserende progenitor cellen (2) en uitontwikkelde neuronen (3). Experimenteel hebben we losse DNA-uiteinden gelabeld en geïdentificeerd voor elk ontwikkeling tijdstip en deze vervolgens geassocieerd met de DNA-activiteit en ruimtelijke conformatie van het DNA in de kern. Door het samenstellen van een atlas van DNA-breuken over de totaliteit van het DNA voor elke ontwikkelingsmijlpaal, vinden we een algemene neiging van DNA-breuken om op te treden op plaatsen in het DNA met hoge genactiviteit, en in het promotorgebied dat die genactiviteit reguleert. We ontdekten dat neuronen uniek zijn als gevolg van hun 3D-DNA-conformatie, en zich aanzienlijk onderscheiden in mate van optredende breuken van de voorgaande nog-delende cellen. Al met al beschrijven onze data veel onderling gerelateerde processen, maar laten ze geen directe mechanistische causaliteit tussen die processen zien.

In de tweede studie “Topoisomerase 1 activity during mitotic transcription…” (Paper II), hebben wij ons gericht op de regulatie van DNA-activiteit regulatie door DNA-knip-enzym Topoisomerase 1 (TOP1). TOP1 maakt tijdelijke breuken in het DNA in de context van celdeling en daarbij optredende replicatie stress. We ontdekten dat TOP1 de DNA-activiteit direct reguleert door te binden aan de hoofdrolspeler in de RNA-productie, RNA Polymerase II (RNAPII). Door de interactie tussen TOP1 en RNAPII te verstoren ontdekten we dat de celdeling daardoor vertraagd werd, en merkten we effecten van replicatiestress op. Door TOP1 in gezonde cellen te elimineren, namen we waar dat RNAPII misplaatst werd. Een vergelijkbaar fenomeen namen we waar in de mutant cellijn. We concluderen dat afwezigheid van TOP1 direct destabilisatie van geprogrammeerde genactiviteit veroorzaakt, leidend tot verlies van cellulair geheugen, en dus verlies van cel rol kan veroorzaken.

Uit deze onderzoeken concluderen we dat DNA-breuken van nature voorkomen, als gevolg van een bepaalde cel toestand of cel rol. We zien dat de 3D-conformatie en DNA-activiteit van een bepaalde cel ons in staat stellen fragiele DNA-breuk plaatsen te voorspellen. DNA-breuken hopen op rondom de plaatsen met hoge genactiviteit en regulatie promotorgebieden. Wij ontdekten een nieuwe rol van TOP1 bij het reguleren van genactiviteit in diezelfde gebieden en het in stand houden van cellulair geheugen gedurende de celdeling. Wij vonden echter geen globale toename van DNA-schade in afwezigheid van TOP1. Alles bij elkaar, dragen deze bevindingen bij aan een beter begrip van wat er in de celkern gebeurt, hoe DNA wordt gereguleerd en gestructureerd en tot slot hoe verstoringen van die processen tijdens de reguliere ontwikkeling van hersencellen tot ernstige hersenaandoeningen zouden kunnen leiden.

RIASSUNTO DELLA TESI A FINI DIVULGATIVI (ITA)

Catturare le mutazioni al momento giusto: potrebbe essere tutto nella tua testa

Un'ampia gamma di disturbi psichiatrici la cui eziologia è sconosciuta potrebbe essere il risultato di danni al DNA che sopraggiungono durante il normale sviluppo cerebrale. Storicamente, le patologie psichiatriche sono state identificate precocemente e classificate sulla base del materiale genetico ereditato, in considerazione del fatto che spesso disordini come la schizofrenia o l'autismo sono familiari. Attraverso lo studio dei suddetti gruppi familiari, è stato possibile identificare il valore predittivo di specifiche variazioni genetiche e tali liste di possibili mutazioni sono in continuo sviluppo. Ciò nonostante, oltre i due terzi dei disturbi psichiatrici rimane ancora ad oggi un mistero: il nostro studio si prefigge l'obiettivo di contribuire a far luce sui poco conosciuti meccanismi patogenetici che stanno alla base di tali disordini cerebrali.

Parti differenti del cervello potrebbero essere il bersaglio di specifiche variazioni genetiche, che porterebbero alla formazione di diverse sottopopolazioni cellulari che continuamente accumulano mutazioni durante le fasi precoci dello sviluppo cerebrale. Per meglio intenderci, un tipico esempio può essere rappresentato dai tessuti tumorali: un insieme di cellule geneticamente differenti (eterogenee) che, dividendosi, danno luogo ad ulteriori sottopopolazioni cellulari che, a loro volta, accumulano errori, alcuni dei quali risultanti in una minore vulnerabilità al trattamento. Nelle cellule tumorali, lo stress replicativo è uno dei maggiori induttori dei cambiamenti genetici, i quali conferiscono alle cellule la capacità di moltiplicarsi in maniera incontrollata e/o di eludere il sistema immunitario. A differenza dei neuroni, che non si dividono attivamente durante le fasi tardive dello sviluppo, le cellule staminali neuronali vanno incontro ad un preciso e definito processo di maturazione verso la loro forma finale. Ciascuna di esse, porterà alla formazione di un neurone unico e indispensabile per il funzionamento cerebrale. L'assegnazione di una falsa identità o la morte di una di queste cellule potrebbe risultare nella perdita di una specifica funzione cerebrale.

Il mio lavoro di tesi vuole rappresentare il contributo nell'identificazione di quelle parti del DNA che sarebbero particolarmente predisposte ad accumulare al loro interno delle rotture potenzialmente responsabili di cambiamenti genetici rilevanti per lo sviluppo di una malattia. Inoltre, procedendo con il nostro studio, vogliamo dimostrare come l'attività del DNA e/o la conformazione tridimensionale del materiale genetico nel nucleo potrebbe influenzare la localizzazione di siti ricchi di rotture del DNA stesso. É stato, in aggiunta, osservato come persino rotture temporanee del DNA potrebbero determinare cambiamenti genetici a lungo termine e dal possibile valore predittivo per quanto riguarda i disturbi psichiatrici. Ancora, abbiamo voluto stimare se esiste una finestra temporale critica durante le fasi precoci di sviluppo per il verificarsi di queste rotture del DNA. In conclusione, abbiamo studiato se un disturbo della regolazione dell'attività dei geni possa causare la perdita dell'identità cellulare attraverso programmi di attivazione genetica o rotture del DNA ed essere, al contempo, clinicamente rilevante.

Nel primo studio "An atlas of endogenous DNA double-strand breaks…" (Primo Articolo), ci siamo proposti di descrivere il panorama genetico delle fasi precoci dello sviluppo

cerebrale. Attraverso l'implementazione di tecniche d'avanguardia, abbiamo potuto analizzare la fragilità del DNA, la sua attività ed organizzazione nello spazio tridimensionale. Il nostro rappresenta uno dei principali studi che descrivono la fragilità del genoma e il processo di rottura del DNA in assenza di perturbazioni nelle cellule neurali in sviluppo. Abbiamo specificatamente scelto di condurre la nostra indagine in assenza di perturbazioni, in quanto la natura quantitativa del danno al DNA potrebbe essere influenzata dai cambiamenti ambientali o dalla soppressione della riparazione. Ci siamo focalizzati su tre momenti critici durante il minuzioso processo di sviluppo neuronale delineato in precedenza. Ciascuno di essi rappresenta una pietra miliare, un traguardo raggiunto dalle cellule in via di specializzazione. Cellule staminali ad alto tasso di replicazione (1), cellule progenitrici multipotenti (2), cellule neuronali definitive. Nei nostri esperimenti, abbiamo identificato e marcato le estremità libere del DNA in ognuno di questi momenti critici di sviluppo cellulare, associandole con la corrispondente attività e conformazione tridimensionale del DNA all'interno del nucleo. Ciò ha reso possibile la realizzazione di una mappa delle suddette rotture del DNA nel contesto del genoma nei vari momenti, nonché l'osservazione che la maggior parte delle rotture del DNA si verifica in prossimità dei siti di trascrizione altamente attivi e del promotore. Abbiamo scoperto che l'unicità dei neuroni è una diretta conseguenza della loro conformazione 3D nello spazio e che si differenziano notevolmente dai loro progenitori in termini di distribuzione dei siti di rottura del DNA. Nel complesso, il nostro dataset ci ha permesso di descrivere diverse interrelazioni, ma non rivela nessun meccanismo causale diretto.

Nel secondo studio "Topoisomerase 1 activity during mitotic transcription…" (Secondo Articolo), abbiamo posto l'attenzione sul processo di regolazione del DNA da parte dell'enzima Topoisomerasi 1 (TOP1), il quale causa rotture temporanee del DNA durante la divisione cellulare e lo stress replicativo. Abbiamo scoperto che TOP1 regola l'attività del DNA legando direttamente l'enzima chiave nella produzione dell RNA: RNA Polymerase II (RNAPII). In particolare, disturbando l'interazione tra TOP1 e RNAPII, abbiamo constatato che, a causa dello stress replicativo, la divisione cellulare veniva rallentata. Attraverso la soppressione di TOP1 nelle cellule sane, abbiamo potuto osservare un mal posizionamento di RNAPII nel contesto del DNA simile a quello delle linee cellulari mutate. Abbiamo così potuto accertare che l'assenza di TOP1 causa direttamente una destabilizzazione dei programmi di attività genetica, perdita della memoria nonché perdita dell'identità cellulare.

Dai nostri studi possiamo concludere che: le rotture del DNA sopraggiungono naturalmente e in conseguenza di un particolare stato cellulare o identità; i siti fragili di rottura del DNA si differenziano nelle varie fasi di sviluppo cerebrale, sulla base dell'organizzazione tridimensionale e l'attività del DNA; le rotture del DNA si verificano con maggiore tendenza in prossimità dei siti di attivazione genetica e dei loro promotori; TOP1 svolge un ruolo regolatore in queste stesse regioni e nel garantire una memoria cellulare durante la replicazione. Tuttavia, non abbiamo riscontrato un aumento globale in termini di danno del DNA in assenza di TOP1. Nel complesso, queste scoperte ci permettono di comprendere meglio cosa accade all'interno del nucleo di una cellula, come il DNA è regolato e strutturato ed, infine, come una perturbazione di questi processi durante lo sviluppo cellulare può risultare nella comparsa di un disturbo cerebrale.

LIST OF SCIENTIFIC PUBLICATIONS

THESIS CONSTITUENT PAPERS

I. **Ballarino R.,** Bouwman B.A.M., Agostini F., Harbers L., Diekmann C., Wernersson E., Bienko M., Crosetto N. (2022). **An atlas of endogenous DNA double-strand breaks arising during human neural cell fate determination.** *Scientific Data* 9:400, 1-19.

II. Wiegard A., Kuzin V., Cameron D.P., Grosser J., Ceribelli M., Mehmood R., **Ballarino R.,** Valant F., Grochowski R., Karabogdan I., Crosetto N., Bizard A.H., Kouzine F., Natsume T., Baranello L. (2021). **Topoisomerase 1 activity during mitotic transcription favors the transition from mitosis to G1.** *Molecular Cell* 81, 5007–5024.

OTHER PUBLICATIONS

Mastropasqua F., Oksanen M., Soldini C., Alatar S., **Ballarino R.**, Molinari M., Agostini F., Poulet A., Watts M., Rabkina I., Becker M., Li1 D., Anderlid B., Isaksson J., Remnelius K.L., Moslem M., Jacob Y., Falk A., Crosetto N., Bienko M., Santini E., Borgkvist A., Bölte S., Tammimies K. (2022). **Deficiency of Heterogeneous Nuclear Ribonucleoprotein U Leads to Delayed Neurogenesis.** *BioRxiv*, 1–35.

Ballarino R., Bouwman B.A.M., Crosetto N. (2021). **Genome-Wide CRISPR Off-Target DNA Break Detection by the BLISS Method.** *Methods in Molecular Biology* 2162, 261–281.

CONTENTS

LIST OF ABBREVIATIONS

2n	Diploid (2 sets of chromosomes)
3C-Seq	Chromatin conformation capture
3D	Three-dimensional space
53BP1	p53-binding protein 1
ADHD	Attention deficit hyperactivity disorder
alt-EJ	Alternative end joining
APH	DNA polymerase inhibitor aphidicolin
ASD	Autism spectrum disorder
ATAC-Seq	Assay for transposase-accessible chromatin using sequencing
BFB	Breakage-fusion-bridge
BIR	Break-induced repair
BLESS	Direct in situ breaks labeling, enrichment on streptavidin and next-generation sequencing
BLISS	Breaks Labeling In Situ and Sequencing
BPD	Bipolar disorder
BRCA2	DNA repair protein breast cancer 2
c-NHEJ	Canonical non-homologous end-joining
CAG	Cytosine-adenine-guanine tri-nucleotide
Cas12a	CRISPR associated protein 12a, previously known as Cpf1
CAS9	CRISPR associated protein 9
CFS	Common fragile site
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
CNA	Copy number alteration
CNCC-seq	Coverage-normalized cross correlation analysis
CNV	Copy number variation
COSMIC	Catalogue of somatic mutations in cancer
CpG island	Genomic regions with high frequency CpG sites
CpG site	Cytosine followed by a guanine dinucleotide along the 5' → 3' direction
CRISPR	Clustered regularly interspaced short palindromic repeats
CT	Chromosome territory
CTCF	CCCTC-binding factor
CUTseq	Restriction enzyme-based method for reduced representation genome sequencing
D-Loop	Displacement loop: two strands dsDNA separated, yet held apart by a third strand

DamID	DNA adenine methyltransferase identification
DDR	DNA damage response
DNA	Deoxyribonucleic acid
DSB	DNA double-strand break
ENCODE	Encyclopedia of DNA elements
ERG	Neural early response genes
ESC	Totipotent embryonic stem cell
ETO	Etoposide
FFPE	Formalin-fixed paraffin embedded
FISH	Fluorescence in situ hybridization
FoSTeS	DNA fork stalling and template switching
G4	G-quadruplexes
gammaH2AX	Histone variant H2AX with phosphorylation on residue Ser-139
GC-content	Guanine-cytosine content in stretch of DNA
gDNA	Genomic DNA
GO	Gene ontology
GPSeq	Genomic loci positioning by sequencing
GUIDE-Seq	Genome-wide, unbiased identification of DSBs enabled by sequencing
GWAS	Genome-wide association studies
HCT116	Human colon cancer cell line
HD	Huntington's disease
HI-C method	Extension of 3C-Seq to map chromatin contacts genome-wide
HR	Homologous recombination
HTGTS	High-throughput genome-wide translocation sequencing
HU	Ribonucleoside diphosphate reductase inhibitor hydroxyurea
ID	Intellectual disability
IDLV	Integrase-defective lentiviral vector
IF	Immunofluorescence
Indel	Genomic insertion/deletion
iPSC	Induced pluripotent stem cell
IVT	In vitro transcription
kb	One kilobase is equal to 1000 bases
KI	Gene knock-in
KU70/80	DNA repair heterodimer of Ku70 (XRCC6) and Ku80 (XRCC5)
LAM-HTGTS	Linear amplification mediated high-throughput genomic translocation sequencing

LOH	Loss of heterozygosity
LRGs	Neural late response genes
Mb	One megabase is equal to 1 million bases
MMEJ	Microhomology-mediated end joining
mTOR	Mammalian target of rapamycin - FK506-binding protein
NAHR	Non-allelic HR
NDD	Neurodevelopmental disorder
NES	Neuroepithelial stem (cell line)
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
PKcs	DNA-dependent protein kinases
PSC	Pluripotent stem cell
PSYCHENCODE	Psychiatry encyclopedia of DNA elements
QTL	Quantitative trait locus
R-loop	A nucleic acid structure containing a DNA:RNA hybrid and a strand of DNA
RAD51	DNA repair protein RAD51 homolog 1
RAG	Recombination activating enzymes
RDC	Recurrent DSB cluster identified through HTGTS
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RNAPII	RNA polymerase II
ROS	Reactive oxygen species
RPA	Replication protein A
sBLISS	in-suspension breaks labeling in situ and sequencing
scCUTseq	Single-cell CUTseq method
SCZ	Schizophrenia
SFARI	Simons foundation autism research initiative
sgRNA	Single guide RNA that enables specificity in every CRISPR experiment
SLAM-seq	Time-resolved measurement of newly synthesized and existing RNA
SMARCAL1	SWI/SNF related, actin dependent regulator of chromatin, subfamily a, like 1
SNP	Single-nucleotide polymorphism is a germline substitution of a single nucleotide
SPLiT-seq	Split-pool ligation-based transcriptome sequencing
SPRITE	Split-pool recognition of interactions by tag extension
SSA	Single-strand annealing
SSB	DNA single-strand break

SSM	DNA polymerase slipped strand mispairing
TAD	Topologically associated DNA domains
TdT	Terminal deoxynucleotidyl transferase
TOP1	Type I DNA topoisomerase
TOP2	Type II DNA topoisomerase
TOPcc	Topoisomerases cleavage complexes
TP53	Phosphoprotein p53
TSS	Transcription start site
UMI	Unique molecular identifier
VDJ	variable, joining, and diversity gene segments
WGS	Whole-genome sequencing
WT	Wild type
XRCC	DNA repair protein X-ray repair cross complementing

1 INTRODUCTION

1.1 NEURODEVELOPMENTAL DISORDERS AND GENOME FRAGILITY

1.1.1 The nervous system and disorders affecting brain function

The adult human brain is comprised of roughly 86 billion neurons that work together to maintain homeostasis at many levels of resolution. The brain can be seen as a complex multi-cellular tissue in which genetic, functional, and cellular architecture need to be regulated. Within the brain, different neural (neurons) and supportive cell types (astrocytes, oligodendrocyte, microglia), each with their specific role and transcriptome, work together to shape the brain's function during development and maintain it for the rest of our lives.

When the balance of the healthy brain is perturbed, there may be large consequences for our wellbeing. Physical and structural injury to the brain often has clear causes and huge consequences, including the loss of various specific cognitive and physical functions[1]. In contrast, other types of injury and disorders affecting the brain, such as neurodevelopmental and neuropsychiatric disorders, have generally remained more enigmatic[2]. For one, this lack of insight may relate to the difficulty of classifying such disorders by brain region, genetic pathway involved, clinical presentation, or even the cell type responsible for their functional defect[3]. A group of several disorders believed to emerge in early human brain development fits this description and we will henceforth refer to this group as neurodevelopmental disorders (NDDs), even though some of these were classified as neuropsychiatric disorders before their early onset or developmental aspect was properly appreciated[4].

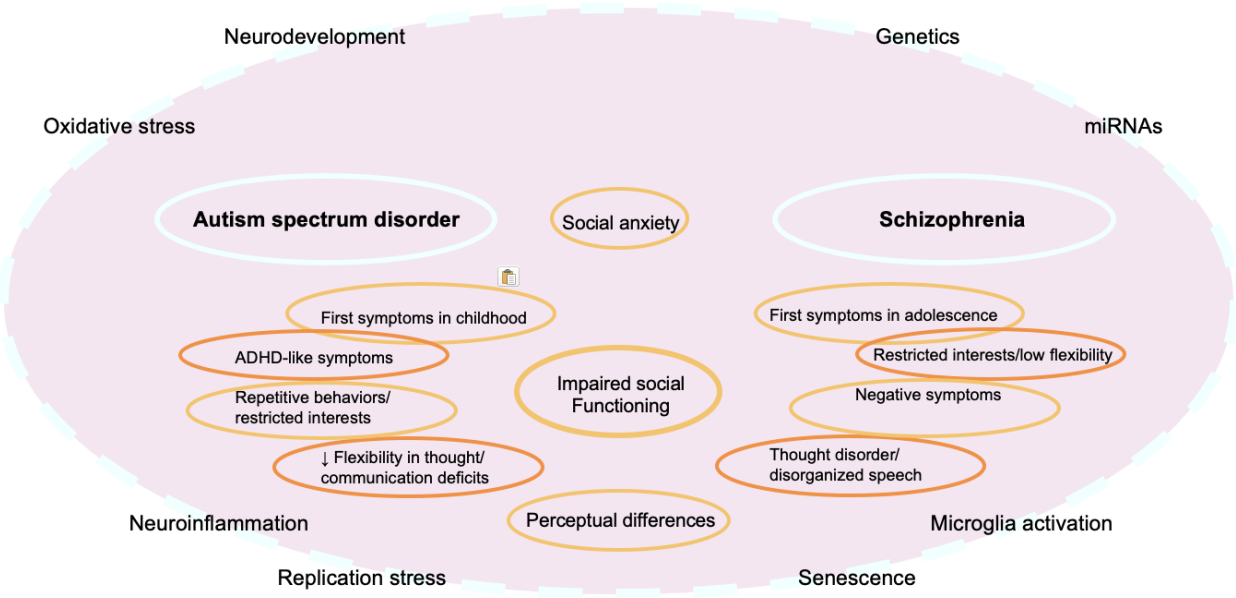

Figure 1. A hypothetical integrative model of ASD and SCZ. The center circles represent intersections and similarities between ASD and SCZ in terms of clinical symptom areas. Cognitive functions as the ones included here often present in a spectrum and are associated with broad neural circuits. Placement of the symptoms around the shared impaired social functioning in the center is meant to represent the relationship between classically differently presenting phenotypes. The outer circle represents underlying biological processes which have been used to explain pathogenesis and initiation of disturbance of neurochemical homeostasis. This figure is inspired by the figure put forward by Prata et al., 2017 explaining overlap in biomarkers across ASD and SCZ[5].

NDDs are multifaceted conditions characterized by early-onset impairments or deficits of variable severity in cognition, communication, behavior, and/or motor skills, resulting from abnormal brain development. NDDs include among others intellectual disabilities (ID), cerebral palsy, attention-deficit/hyperactivity disorder (ADHD), schizophrenia (SCZ), and autism spectrum disorder (ASD). SCZ and ASD both represent a spectrum of disorders, with SCZ referring to severe psychotic disorders that are characterized by a disconnection from reality, including delusions and hallucinations[5] (**Figure 1**). While SCZ generally presents between 15 and 25 years (in men) and 25 and 35 years (in women), recent evidence has suggested that the associated changes in the brain already emerge much earlier in life. In ASD, a disorder spectrum characterized by variations in communication, learning, behavior, and social interaction as compared to neurotypical individuals, there is a very wide range of symptoms and levels of disability in functioning, ranging from intelligently gifted children and adults able to fully perform all facets of life to others needing extensive lifelong support.

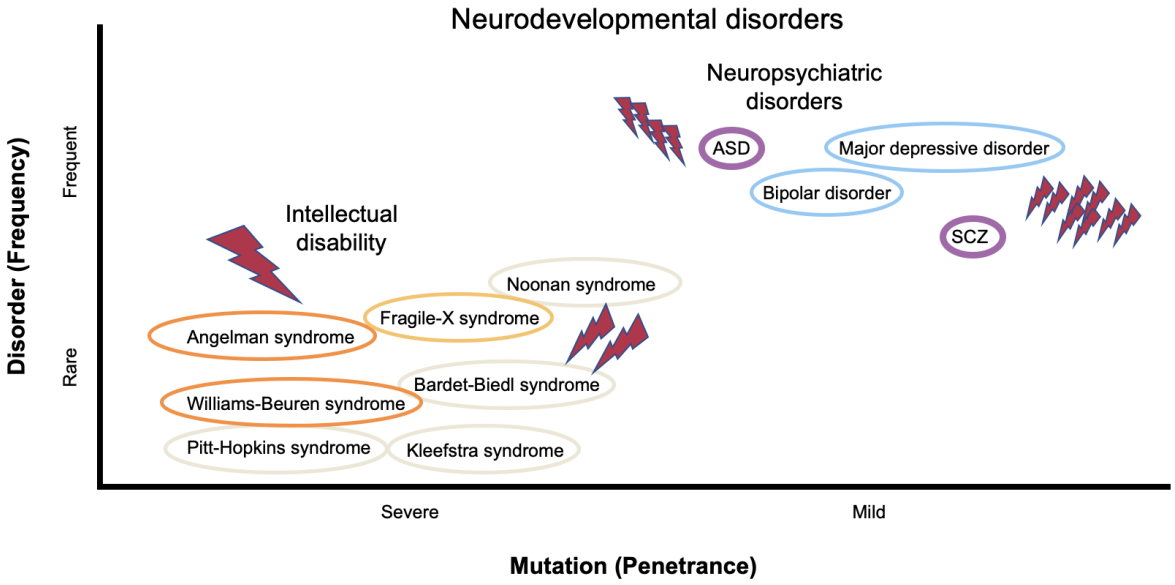

Figure 2. Genetic spectrum and overlap of various neurodevelopmental disorders. On the left, several monogenic ID disorders with rare mutations of severe impact. On the right, multi-gene complex neuropsychiatric disorders which have proven difficult to categorize. The contrast lies between neurodevelopmental disorders which are either multifactorial or complex genetic disorders that often arise from common variants with a weaker effect on gene function. Red thunderbolts represent a quantitative effect of one or more stochastic mutations of the genome or environmental triggers or stimuli. This figure is inspired by the figure put forward by van der Voet et al., 2014 illustrating the concept of genetic penetrance and in neurodevelopmental disorders[6].

NDDs are classified as complex traits, meaning that they do not follow simple Mendelian inheritance and that their inheritance cannot be attributed to a single mutated gene but rather to a group of risk variants in various genes in combination with environmental factors (**Figure 2**). The underlying insult or molecular cause giving rise to perturbation of homeostasis in these NDDs remains unclear, but recent genetic advances are pointing to converging causes and a shared etiology, as is the case for ASD and SCZ. Understanding the role of the genetic architectures of different brain disorders is thus challenging, as they often strongly overlap both in symptoms and associated genetic risk variants (**Figure 3**)[2]. To date,

the (Simons Foundation Autism Research Initiative) SFARI gene database for ASD research is the most sophisticated resource available, with a list of 1,036 genes of significant impact. However, compared to current knowledge, the reproducible yield of candidate gene-association studies has been questioned[7]. In recent years efforts to identify genomic variants with regulatory functions in large scale projects such as the Psychiatry Encyclopedia of DNA Elements (PsychENCODE) project has indicated that NDD genetic risk factors converge at least partially on the same underlying pathogenic biological processes[8]. The biological processes driven by gene expression phenotypes is what is referred to as an functional quantitative trait loci's (eQTL). In other words, eQTLs explain a fraction of the genetic variance of functional or pathological process in relation to genetic changes at particular genomic coordinates, thereby attributing a "weight" to each part of the genome and the role in specific pathological processes. These converging pathological processes and disease etiologies fit into the hypothesis described above, but the discovery of eQTLs often requires further complementary functional approaches to hold water[9]. In order to truly get insights into early origins of homeostasis disruption, it is important to study multiple cellular processes including calcium homeostasis, proteostasis, energy regulation and genome stability[10,11]. In the sections below, I discuss aspects of our current understanding of the etiology of NDDs, with a focus on structural genomic variation.

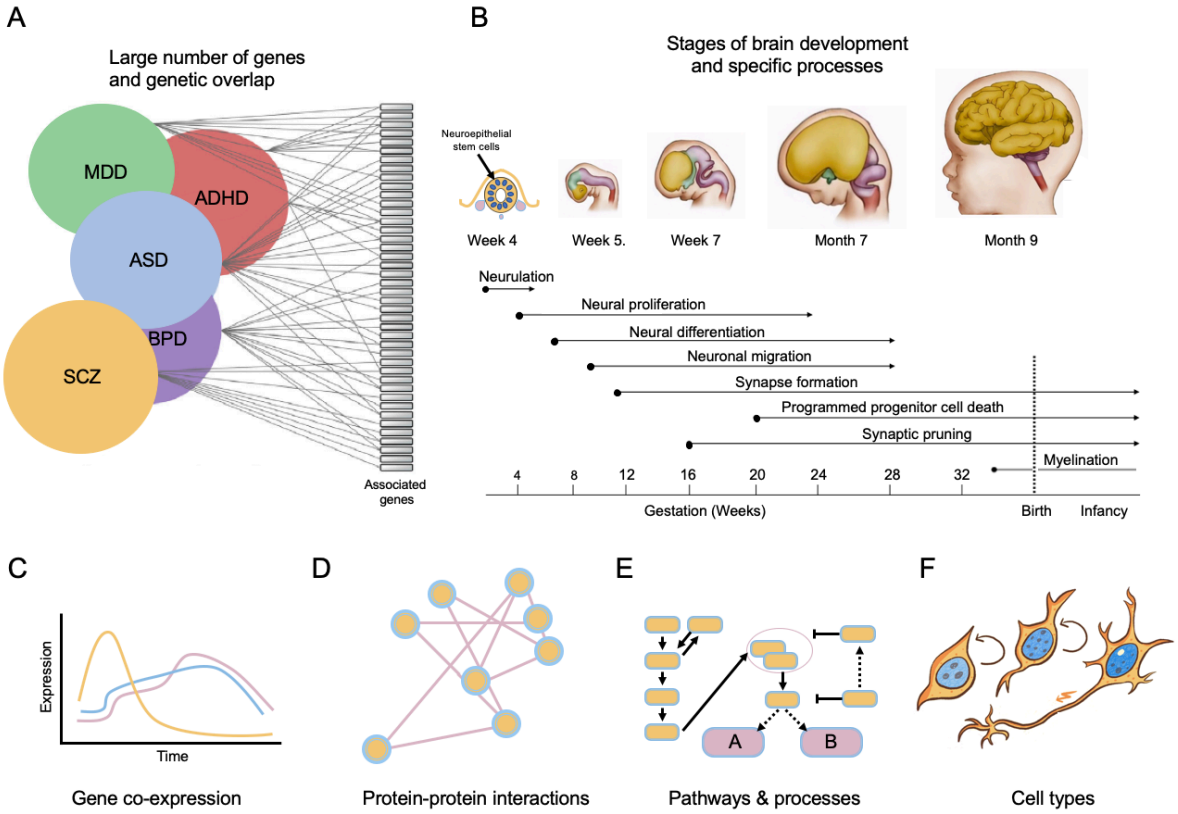

Figure 3. Overlap between different neurodevelopmental and neuropsychiatric disorders. (**A**) Fenn diagram depicting genetic overlap of distinct disorders and shared gene causality. (**B**) Scheme representing milestones in development of the central nervous system from early embryology to birth and the associated cellular processes at different stages of gestation. To better understand how multiple processes are able to give rise to disease, multiple levels of regulation are depicted below. (**C**) Graph depicts how gene dosage and co-expression throughout developmental time may differ in cells with different (epi)genetic background. (**D**) Network of protein-protein interactions depicts how

important is can be to understand how specific enzymes interact and co-regulate each other. (**E**) Scheme illustrating multi-level regulation of signaling cascades and processes all of which are important to establish and maintain a correct balance of transcription factors and cell identity. (**F**) Sketch of the three distinct developmental stages in neural specification. Rapidly replicating self-renewing neuroepithelial stem cells (1), primed neural progenitor cells initiating protrusion migration akin to radial glia (2) and post-mitotic neurons exhibiting early stages of neural activity, and electrochemical transmission (3). Each cell type is highly sensitive to changes of (c), (d) and (e). This figure is inspired by the figure put forward by van der Shohat et al., 2021, illustrating the genetic overlap, developmental timeframe and mechanisms underlying neural pathophysiology[12].

Analyzing structural genomic variation and distinguishing germline and somatic events

The recent increased application of sequencing in clinical settings has made a start towards elucidating the genomic architecture of NDDs and has concomitantly led to a better understanding of NDD pathophysiology, as will be discussed in the next sections. First, I will define some of the terms used to discuss genomic variation.

Among the various forms of genomic variants, a prominent and often impactful type is represented by copy number variants (CNV), which are defined as a change in the normal diploid (2n) copy number of a part of the genome sequence, typically ranging from a few kilobases (kb) up to several megabases (Mb). CNVs are distinguished from aneuploidy, which is instead defined as the presence of one or more gains or losses of entire chromosomes or chromosome arms[13]. The prevalence of disease-related CNVs is estimated to be 10 times higher than the prevalence of disease-related single-nucleotide polymorphisms (SNPs), which represent the other major form of genomic variation[14]. Traditionally, the term CNV has been used to describe both inherited and *de novo* germline events, whereas copy number alterations (CNAs) and single nucleotide variation (SNV) are used to describe somatic events that form in non-germline cells and thus escape hereditary transmission. However, for simplicity and to avoid confusion throughout this thesis, I will distinguish between germline and somatic CNVs where needed, refraining from using the term CNAs.

To understand the outcome of sequencing experiments and acquire an understanding of the accumulation of structural genomic variation over time, it is important to distinguish germline and somatic CNVs (**Figure 4**.). *De novo* germline variants detected in a child, but not in their parents, might be relatively rare and carry increased disease risk, whereas common variants widely present across a population tend to have smaller effect sizes. By definition, germline CNVs are present in all the cells of the organism and can therefore be detected by sequencing genomic DNA (gDNA) extracted from peripheral blood cells. In contrast, somatic CNVs are generally confined to one or a few tissue or cell types and can therefore only be detected in the genome of those cells, depending on when they arise during organismal life[15]. In line, their effects may also be confined to a particular tissue and thus be exempted from hereditary transmission. For example, CNVs that do not form in the germline but early on during embryogenesis will have a wide tissue distribution, whereas somatic CNVs that emerge in a particular stem cell niche will have a much more restricted distribution. Cancer-associated CNVs are a clear example of CNVs that emerge in adult life and that are restricted to a specific group of cells (tumor cells)[16,17]. Due to this tissue/cell-restricted nature, somatic CNVs often escape detection in traditional genome-wide

association studies (GWAS), as their preferred source material is whole blood[18]. These studies, aimed at detecting genomic variation in the population using whole genome sequencing (WGS), will thus generally identify germline CNVs or SNPs, while missing particular tissue- or cell-type specific somatic CNVs that have the capacity to be of large effect, and which have proven important for assessing genetic heterogeneity and evolution in normal tissues and cancers. Similarly, clinical deep sequencing of a patient's genome to acquire a map of their genomic make-up and their genomic variation will also fail to properly identify more tissue-specific variants when the patient's blood is used, as is classically the case.

1.1.2 CNVs in neurological disease and NDDs

Recently developed single-cell sequencing methods[19] have allowed assessment of somatic genomic variation across individual neural cells[20–22], including SNVs associated with neurological diseases such as epilepsy and brain malformations[23]. Their findings have broadly suggested that structural genomic variation (including CNVs) is more frequently associated with early arising NDDs and neuropsychiatric disease, whereas a broader group of mutations can be related to large imbalances in brain functions such as those observed in epilepsy, micro/macrocephaly, and cancer.

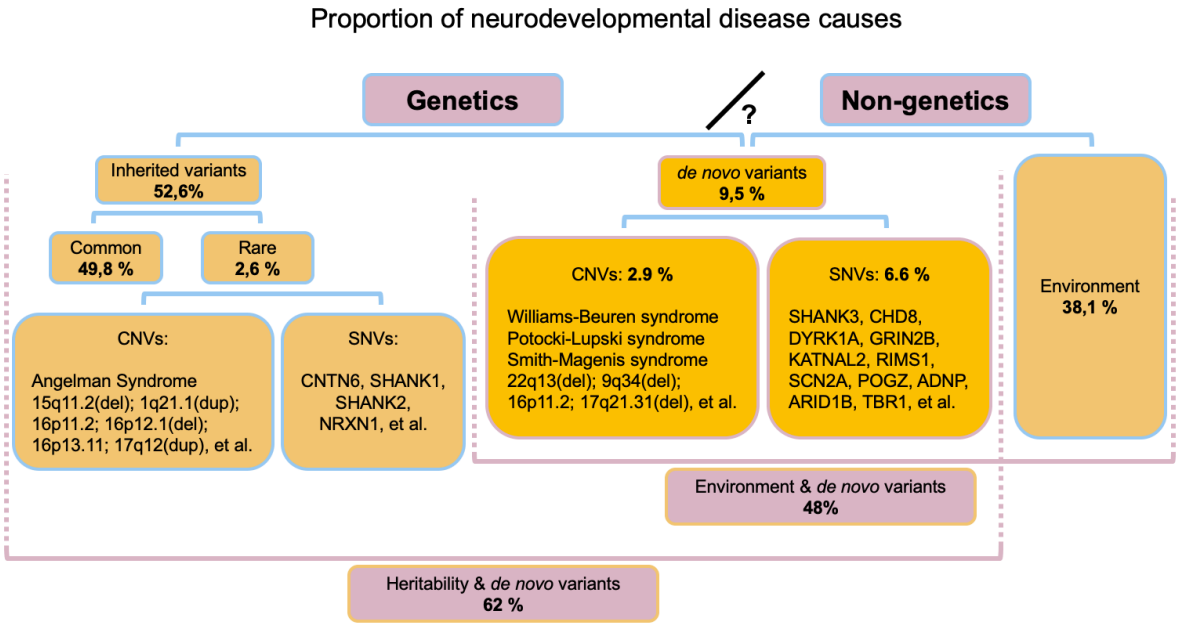

Figure 4. Proportion of neurodevelopmental disease causes. The estimation of relative contribution of genetics and environment to ASD is approximately half based on familial and twin studies. Inherited common variants are observed in the general population, rare variants only contribute a small part. *De novo* mutations are genetic causes, but since they do not contribute to heritability, they are considered environmental causes of ASD that act on the DNA molecule. This figure is based on the figure put forward by Huguet et al., 2016, illustrating proportion of genetic vs non-genetic causes[24].

While a lot of the existing genomic variation has no direct phenotypic consequences, both germline and somatic CNVs can cause or predispose to a variety of diseases[25]. Although CNVs represent a minority of all causative alleles, they can be used to assess disease risk in certain complex disease traits for which the underlying mechanism is more ambiguous[25]. Indeed, various CNVs have been associated with disease and predispose in particular to

NDDs and syndromic forms of autism[26], as well as a broader spectrum of human diseases, in particular brain disorders[27] (**Figure 4**).

CNVs and other types of variation are thought to play a large role in conveying risk for NDDs. *Vice versa*, many genomic risk regions identified to impart risk of NDDs have been found to overlap regions affected by CNVs. A common form of CNV implicated in more than 30 different neurological disorders including Huntington's disease (HD) is a phenomenon referred to as short nucleotide repeats instability, such as the instability of cytosine-adenine-guanine (CAG) trinucleotide repeats[28]. In HD, the inherited repeat length of CAG trinucleotide repeats can be prognostic for disease onset, as these two factors show a strong inverse correlation[29]. Huntington disease is autosomic dominant, so the CAG repeats are present in all cells. Interestingly, the pathogenicity phenotype is limited to the brain. However, as repeat length is also inversely correlated with patient age, it is plausible that the levels of germline and somatic instability of CAG repeats are different within sub-populations of cells. This conclusion can be extended, as the appearance of somatic repeat length gains goes hand in hand with progressive pathogenesis in a cell type-selective manner[30]. Like in HD, there are many CAG trinucleotide repeats that arise in exons of certain genes and induce highly selective neurodegeneration in specific regions or cell types of the brain[31]. Finally, CAG repeats have been shown to modulate DNA repair pathways and could predispose to increased mutagenesis. As such, expanding repeats could modify the overall stability of the genome through both cis and trans-acting mechanisms[31].

More typical large recurrent CNVs, including amplifications and deletions, have been shown to predispose to NDDs and syndromic forms of ASD[26,32]. Etiologically relevant CNVs are found in 2-3% of all SCZ cases, in 10% of ASD cases, and in over 25% of all tested cases of ID[33]. A high prevalence of ASD symptoms is frequently associated with monogenetic syndromes characterized by highly penetrant CNVs[34]. In these cases, the CNV pathogenicity has been attributed to the copy number change of one or more dosage-sensitive genes or genomic regions[35–37]. Such gene dosage alteration has emerged as a widespread phenomenon in neuropsychiatric disease, where it largely manifests in the form of CNVs[38]. Most illustrative of this is the Williams-Beuren syndrome, where a duplication of the 7q11.23 locus spanning several genes gives rise to neurological and behavioral problems, whereas a deletion of the same locus results in increased risk of epilepsy, ID, and neurobehavioral abnormalities[39]. A similar yet distinct phenomenon occurs at the 15q.11-q13 locus, where either deletion or duplication results in several neurobehavioral syndromes associated with ID and epilepsy[40]. **Figure 4** illustrates the relevant contribution of genetic and environmental factors to ASD and illustrates the difficulty of estimating the proportion cause of genetics and environment in NDDs.

While highly penetrant congenital CNVs play a role in disease etiology, the mechanisms by which the resulting complex NDDs arise remain elusive. In addition to congenital CNVs, somatic CNVs in neural cells, although much less frequent, might also have high penetrance and affect parts of the brain differently[41,42]. Despite considerable investigation on NDD-associated pathogenic CNVs, significant gaps in the clinical characterization of NDDs and other brain-associated disorders remain for various reasons. Firstly, as mentioned above, the composition of implicated overall risk variants in the population points towards many

overlapping pathways underlying disease and shared disease etiology, particularly in the case for poorly understood psychiatric disorders with developmental intellectual disability. Secondly, pleiotropy, *i.e.*, non-specificity of NDD-associated pathogenic CNVs, has remained unexplained. Thirdly, rare CNVs are often highly penetrant, whereas other CNVs may confer risk for multiple NDDs.

1.1.3 Somatic variation in the brain of neurotypical individuals

Intriguingly, in addition to identifying CNVs associated with NDDs, single-cell sequencing of cells obtained from different brain regions has also led to the surprising discovery that CNVs are commonly encountered in the brain of individuals without any apparent mental disorders, also referred to as neurotypical individuals[20–22,43]. Usually, somatic neural CNVs are restricted to one type of neuron and/or brain region, a phenomenon broadly known as genetic mosaicism. However, other CNVs are shared by multiple distinct neuron types in brains of neurotypical as well as diseased individuals, implying that these formed post-zygotically, most likely in a committed neural stem cell or progenitor giving rise to a specific cell type in a defined brain region[21].

Based on these single-cell sequencing efforts, it has been estimated that 10–40% of human cortical neurons contains at least Mb-scale *de novo* CNVs. These subchromosomal CNVs were shown to have a two-fold higher tendency towards deletion than amplification[20]. As these deletions were found in both endogenous human frontal cortex neurons and stem-cell derived neurons in culture, certain neural subtypes may be especially prone to large-scale genome alterations[20]. Furthermore, these CNVs were found 10 times more frequently at the somatic level compared to the organismal level, suggesting that these Mb-scale copy number changes may be better tolerated when they occur sporadically in the tissue[43]. In other words, due to the absence of these CNVs in healthy subjects, we can infer an evolutionary selection; the brain is unlikely to cope with these specific mutations if they are brain-wide, but the fact that we do detect single sporadically located cells carrying these mutations, particularly in enriched fashion in patients, indicates that these sites do play a role in pathophysiology. Lastly, sub-Mb somatic CNVs were preferentially detected around telomeres, but were not found to be enriched at known fragile sites or germline CNVs[20], indicating a different cause or means of maintaining genome stability. I elaborate on further on such fragile sites in section 1.2.6 and Figure 9.

1.1.4 Potential origin and pathological function of somatic variation

CNVs can perturb gene expression and consequentially tissue homeostasis in different ways[10,44]. Although the impact of CNVs on the genome structure is not necessarily harmful, the loss, gain, or regulatory disruption of genes affected by these CNVs, resulting in an altered dosage of their RNA and protein products, is often associated with disease, including NDDs[45]. Indeed, CNVs may act directly by amplifying or deleting a gene or functional genomic unit, or more indirectly through positional effects that dysregulate genes in other chromosomal regions *in cis*, for example through chromatin looping[46]. Furthermore, CNVs may also predispose the genome to additional deleterious genetic changes[47]. While this is still a rather new field, work diving into this highly complex subject is currently hard underway under the umbrella of the Brain Mosaicism Network[48]. In sum, the mechanistic

consequences of CNVs in general and of a given specific CNV remain difficult to predict and need to be individually assessed.

Similarly, whether somatic mutations in the brain represent a benign event or whether they predispose to specific neurological disorders is a topic of active research in the field[48,49]. It has for example been hypothesized that CNVs in neurotypical brains might be the result of a mechanism evolved to increase the number of different neural cell types and as such phenotypic variation[47]. These variations found in many cell types, each with different vulnerabilities based on inherent qualities, can result in differential progression in development or cell type specification. In support of this, one study has shown that differential methylation and expression resulting from perinatal stimuli experienced during normal brain activity and development are likely among the mechanisms that contribute to the changes in copy number[50]. However, in humans, an increased burden of brain region-specific CNVs has been described in aged individuals compared to younger subjects, suggesting that new CNVs form throughout lifetime and not only in early life[22]. Several efforts are still underway to discover and assess the impact of these CNVs both in health and disease.

Despite being prevalent in apparently healthy individuals, mosaic brain CNVs might actually work 'behind the scenes' and predispose to disease by altering the balance of neural circuits[10,51,52]. Several studies have shown that a fraction of post-mitotic cells carry genomic changes and display altered phenotypes, which in turn can alter the microenvironment and have large effects in the long run. The impact of these post-mitotic genomic changes on some diverse neural subtypes may therefore be greater than expected based on genes alone: distinct types of neurons show differential gene expression and modes of regulating their electrophysiological activity and may thereby tie into different disease etiology hypotheses for SCZ, ASD and other NDDs (Figure 3). Alterations in gene dosage, which are often implicated in these disorders, might have pronounced effects on neuronal function, for example by i) inducing *trans* differentiation of neurons to other cell types (phenotypic switching); ii) altering the connectivity of neuronal circuits; or iii) inducing neuronal senescence resulting in dead nodes in neural circuits[53]. Somatic mutations that arise during development or during neural network maturation could progressively alter the behavior of individual neurons and thus potentially affect the neural networks they support through secondary degeneration. As such, accumulation of mutations over time could explain why the first signs of disease in certain disease types appear years after the first genetic alteration[52]. Alternatively, CNVs might disrupt the neuronal microenvironment that normally inhibits cell proliferation in the adult brain, and as such increase the risk of cancer[54–56]. However, how much CNV brain mosaicism contributes to functional diversity versus increasing the risk of disease remains largely unknown.

1.1.5 The molecular origin of CNVs

In line with the complexity of understanding the exact pathogenic contribution of CNVs to the onset of brain disorders, the causes and mechanisms that underlie CNV formation are similarly poorly understood[17]. CNVs can be classified into two groups: (i) recurrent and (ii) non-recurrent. The former is characterized by recurring breakpoints in specific regions. Regions with tend to subsequently accumulate large segmental duplications based on WGS

studies, and includes 20-40% of polymorphic CNVs and CNVs associated with distinct clinical phenotypes, such as those identified on chromosomes 16p11.2 and 17q11.2 in individuals with severe ID, ASD, and SCZ[57]. In contrast, non-recurrent CNVs have unique breakpoints that are not associated with large homologous regions, although microhomology between the breakpoints is often found [58]. The majority of polymorphic CNVs and a large fraction of disease-associated CNVs fall into the latter group[57].

Only a few studies have experimentally investigated the formation of CNVs in mammalian cells. In one study, acute exposure of cultured human fibroblasts to the DNA polymerase inhibitor aphidicolin (APH) or the ribonucleoside diphosphate reductase inhibitor hydroxyurea (HU)—conditions that cause replication stress—was shown to cause the formation of recurrent *de novo* CNVs [59]. Etoposide, a TOP2 inhibitor which freezes the TOPcc was shown to also significantly induce large-scale indels and many unique CNVs[60]. A possible explanation for frequent recurrent CNVs was hypothesized to relate to a correlation between transcriptional activity and genome fragility, as roughly a third of the detected recurrent CNVs corresponded to actively transcribed long genes[61]. However, HU and APH do not induce CNVs in germline cells based on adult male mice whom were administered HU[62], suggesting that genomic instability as a consequence of replication stress may only arise later in differentiation, or as a consequence of *in vitro* vulnerability[63]. The breakpoints of large CNVs tend to be associated with late DNA replication timing, low surrounding gene expression and proximity to large genes. By contrast, smaller amplifications show contrasting associations and may represent a new class of replication stress-induced CNVs similar to non-recurrent CNVs [63]. The latter can form independently of genomic repeat regions and are more difficult to identify. Indeed, most human CNVs are labeled as non-recurrent when they form outside of the context of long stretches of homology and thus likely arise through a different mechanism than recurrent CNVs. These more unique events are often classified as disease-associated CNVs when they occur in humans [64,65]. The observation of an increased incidence of non-recurrent CNV formation in stressed cells, in the absence of long stretches of homology, may indicate separate mechanisms at play in the origins of the structural change.

An important type of DNA damage that has been linked to CNV formation are DNA double-strand breaks (DSBs). Indeed, agents that induce replication stress and, in turn, CNVs *in vitro* and *in vivo* also cause the accumulation of DSBs. In the next section, I will discuss how DSBs form in cells, thereby predominantly focusing on endogenously occurring DSBs that emerge in the absence of exogenous causes, and how they can predispose to the formation of CNVs.

1.2 DNA DOUBLE-STRAND BREAKS ORIGINS AND REPAIR

1.2.1 Introduction to DSBs

Constant maintenance of the genome sequence and its structural integrity is crucial for faithful transmission of genetic material to daughter cells during cell division. However, genomic aberrations are inevitable as the DNA molecule is subject to a multitude of different types of damage on a daily basis, such as DNA mismatches, oxidative and hydrolytic cleavage, DNA-protein cross-links, and DNA breaks including single-strand breaks (SSBs) and DSBs. DSBs are among the most deleterious forms of genetic insults as they have the

potential to permanently disrupt genomic integrity, for example by giving rise to structural genomic rearrangements and translocations, if not repaired faithfully. Although DSBs are classically associated with exposure to exogenous factors including UV light, X-ray, and ionizing radiation, or chemotherapeutics such as etoposide, the majority of DSBs is believed to form endogenously, during essential nuclear processes that I will discuss in more detail further below. While DSBs can be highly deleterious and are often associated with cancer, senescence or cell death, they are a frequent occurrence with expected DSB rates between 10 to 50 DSBs per cell per cell cycle[66]. In the next section, I will explore the different mechanisms that might be relevant for understanding this genome-wide fragility and repair in the context of development. First, I will address how DSB repair may have adverse outcome and give rise to structural genomic changes, after which I will discuss the three main sources of endogenous DSB formation.

1.2.2 DSB repair and adverse outcomes

The choice of DSB repair pathway depends on the origin of the DSB, associated proteins at the DSB site, cell type, and cell cycle phase[67–69], and it affects the likelihood of a DSB giving rise to genomic rearrangements and CNVs, or not[67]. Homologous recombination (HR) is a replicative type of DNA repair that reliably repairs DSBs prior to mitosis, during and shortly after DNA replication, by blunt end ligation in the presence of the sister chromatid[70]. HR is thus limited to cycling cells and is not available in post-mitotic cells. In contrast, canonical non-homologous end-joining (c-NHEJ) is a non-replicative repair pathway that can operate throughout the cell cycle[71]. Yet, c-NHEJ has a higher potential for deletions of a few bases in comparison to HR[72], even if recent studies indicated that NHEJ could be guided by pre-mRNA to achieve lower error rates[73]. DSBs may also be repaired through the alternative repair pathways single-strand annealing (SSA) and microhomology-mediated end joining (MMEJ), which is also known as alternative end joining (alt-EJ)[74]. The latter two are known as mutagenic repair pathways and more frequently result in large deletions or insertions and deletions, respectively[67]. **Figure 5** shows a schematic overview of HR and c-NHEJ repair of DSBs.

Although HR represents the most accurate repair pathway for DSBs, its mechanisms can lead to the generation of CNVs, even though this is normally avoided[75–77]. In case of a two-ended DSB, crossing over between the two homologous chromosomes may introduce loss of heterozygosity (LOH), while non-allelic HR (NAHR) may lead to genomic duplications or deletions when repair occurs in a repetitive area and the recombination repair event presents an unequal crossing over[75,78]. In the case of a single-ended DSB, emerging for example during repair of a collapsed or dysfunctional replication fork, a generally faithful and untraceable type of HR repair called break-induced repair (BIR) may take place, which has mainly been studied in yeast although BIR-like processes have also been described in mammals (**Figure 6**)[79,80]. However, LOH may again occur if the broken end finds a homologue rather than a sister molecule. When this homologous sequence resides elsewhere in the genome, BIR occurs in a nonallelic fashion and can give rise to translocations and rearrangements including deletions and duplications[75,79–81].

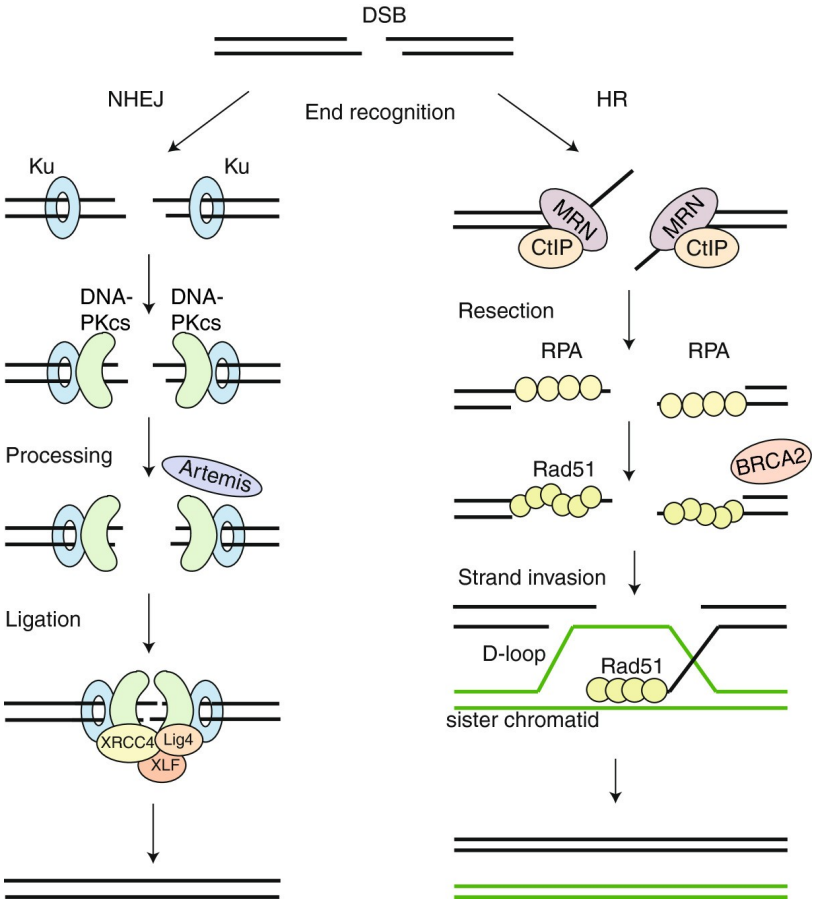

Figure 5. In classical NHEJ (c-NHEJ), the protein heterodimer Ku70/80 recognizes the DSB ends and recruits DNA-dependent protein kinases (PKcs). If needed, incompatible ends are trimmed with a nuclease, such as Artemis, after which the XRCC4-DNA Ligase IV-XLF ligation complex ligates the two DSB ends to seal the break. In HR, the DSB ends are first resected to generate a stretch of single-stranded (ss)DNA. Note that this makes the DSB incompatible with NHEJ repair. RPA then coats the ssDNA, after which RAD51 recruited by BRCA2 takes over and mediates strand invasion on the homologous DNA template. The DSB is repaired by extending the D-loop and then capturing and ligating the second end. This figure is taken from the 2012 review by Brandsma & Gent, illustrating the difference between HR and NHEJ[82].

As already mentioned at the beginning of section 1.2.2, c-NHEJ in contrast to HR, does not depend on a homologous sequence yet it is still relatively faithful. In MMEJ, however, short-range homologous sequences (5–25bp) anneal at the DSB ends as part of the repair process, which can lead to chromosomal rearrangements between regions that only show very limited homology when more than 1 DSB is repaired simultaneously[75,80]. In addition to the above, a few other mechanisms can lead to the formation of CNVs and structural genomic rearrangements due to DSB repair, including breakage-fusion-bridge (BFB) cycles (**Figure 7**, left panel) or fork stalling and template switching (FoSTeS) as a result of replication fork stalling and exposure of the single-stranded template of the lagging strand **(Figure 7**, right panels) without the presence of a DSB[75].

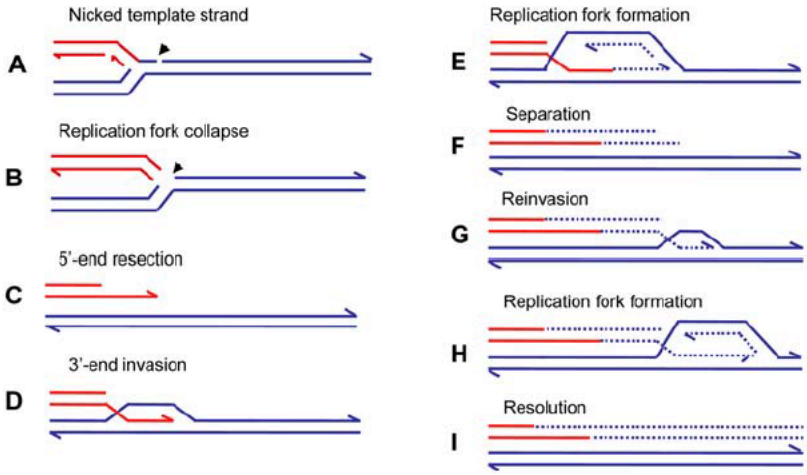

Figure 6. Break-induced repair (BIR) can repair a collapsed replication fork. (A) The replication fork collapses when it comes across a nick in the template strand, causing a single double-stranded break (DSB) where the fork breaks of and collapses (B). BIR resects the 5' strand, creating a 3' overhang (C) that then invades the sister molecule (D), thereby becoming a new replication fork exerting both leading and lagging strand replication (E). A Holliday junction is created at the site of the D-loop, which upon migration or helicase activity will lead to separation of the extended DSB end from its template (F), upon which the DSB end is processed further so the 3' end can again invade the sister and form a replication fork (G, H) that becomes fully processive and can continue replicating the entire chromosome end (H, I). Lines represent DNA strands, polarity is indicated by half arrows, DNA synthesis in dashed lines. This figure taken from Hastings et al., 2009, depicts the step-by-step processing of BIR[83].

When structural variations of large DNA segments accumulate, a condition referred to as genomic instability may arise. Genome instability refers to an increased rate of mutagenesis and copy number changes that is i) characteristic of almost all cancer types; ii) thought to play a critical role in cancer initiation and progression; and iii) is believed to drive intra-tumor heterogeneity. Although proliferation rates in the brain are nowhere near those in cancers, the nervous system has the added difficulty of cells lasting a lifetime with hardly any possibility to replace cells that do not function properly within the network. Moreover, while stem cells in the developing brain require rapid proliferation to give rise to different cortical structures, they do express HR proteins, while fully differentiated post-mitotic neurons that no longer are able to divide deploy the more error-prone c-NHEJ and MMEJ pathways. As elevated levels of DSBs contribute to genomic instability, properly repair of DSB lesions is indispensable for maintaining the stability and integrity of the neuronal genome.

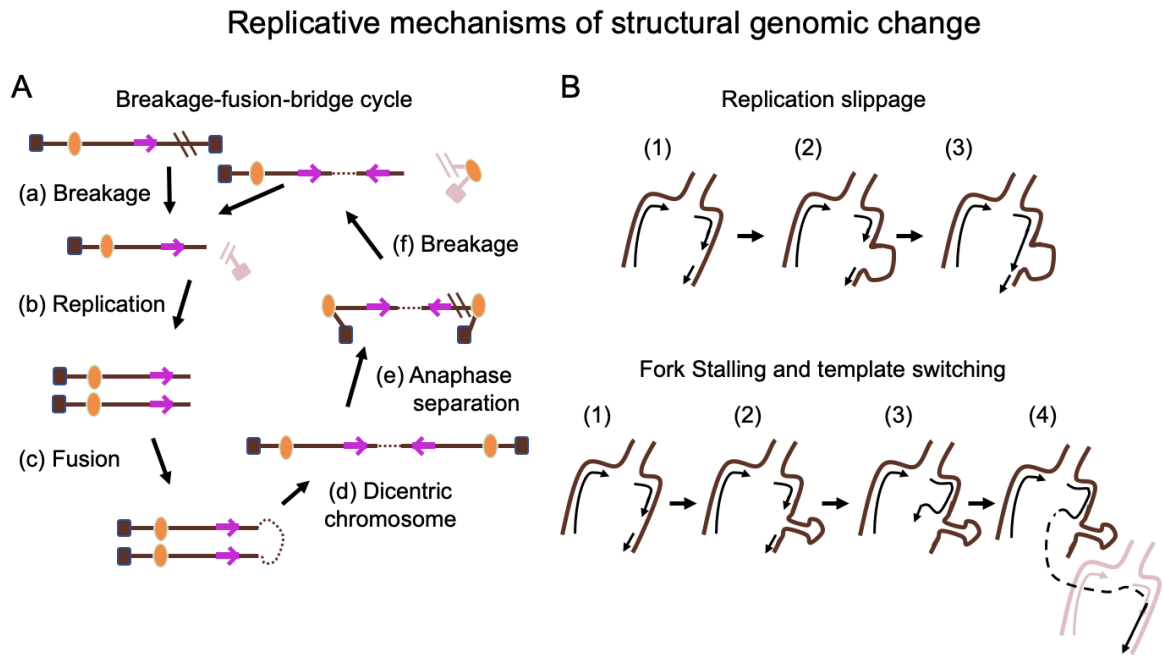

Figure 7. Replicative mechanisms of structural genomic change. (**A**) Left panel shows the breakage-fusion-bridge cycle (BFB). A DSB in an unreplicated chromosome causes loss of a telomere (a), leading to sister chromatids without that telomere after replication (b). Fusion of these two sister chromatid ends (c) is proposed to create a dicentric chromosome (d). In anaphase, the two centromeres of the dicentric chromosome are separated in the telophase nucleus (e), leading to the formation of a bridge between the telophase nuclei. (f) Breakage of the bridge occurs randomly, thereby leading to the formation of an inverted large duplication and a chromosome with an unprotected DSB ending. In the next round of replication, the same cycle is likely followed again, thereby repeating the process until a telomere end is acquired from another source. Centromeres are shown as orange balls, telomeres as brown blocks, genomic sequence as brown lines and mangenta arrows that indicate orientation, breakage sites as double brown lines, lost fragments in beige. (**B**) Right panel, top. Replication slippage exposes a region of the lagging strand as a single strand across timepoints 1, 2 and 3. Right panel, bottom. Fork stalling and template switching (FoSTeS) may occur when the exposed single-strand template of the lagging strand (see top panel), acquires secondary structures that can halt the replication fork, causing the 3' ends of the primer to depart their template (timepoint 3 and 4) and encounter another exposed single-stranded template sequence sharing microhomology. As this other template sequence belongs to another replication fork, duplications, translocations, deletions, or inversions may befall based on the relative genomic position of the encountered replication fork. The mechanisms described in this figure were inspired by both Malkova & Ira 2013 and Hastings et al., 2009 respectively[75,79].

1.2.3 Endogenous sources of DSB formation: DNA replication

Across cellular lifetime there are three principal endogenous causes of DSB formation: (i) proliferation and DNA replication (in cycling cells); (ii) transcription; and (iii) forces acting on the chromatin context (**Figure 9**). The first major driving force of genomic fragility is cell proliferation. As cells progress through the cell cycle, they need to replicate all their genetic material before cell division (mitosis) can be initiated. Mitosis is known to have several checkpoints for genomic integrity and when cells are rapidly dividing, such as stem or cancer cells compartments, they show an increased mutational burden[71,84]. As DNA is replicated, the unwinding of the DNA duplex generates a supercoiling ahead of the replication fork, which requires removal by topoisomerases. Topoisomerases act through the formation of a

short-lived DNA break that is normally re-sealed immediately, but which can also result in a persisting DSB, as will be discussed below in more detail[85]. Moreover, during replication the single-stranded DNA (ssDNA) on the leading strand is more vulnerable to hydrolysis than its double-stranded counterpart and can therefore break, resulting in the formation of a SSB that then results in a one-sided DSB after replication[86].

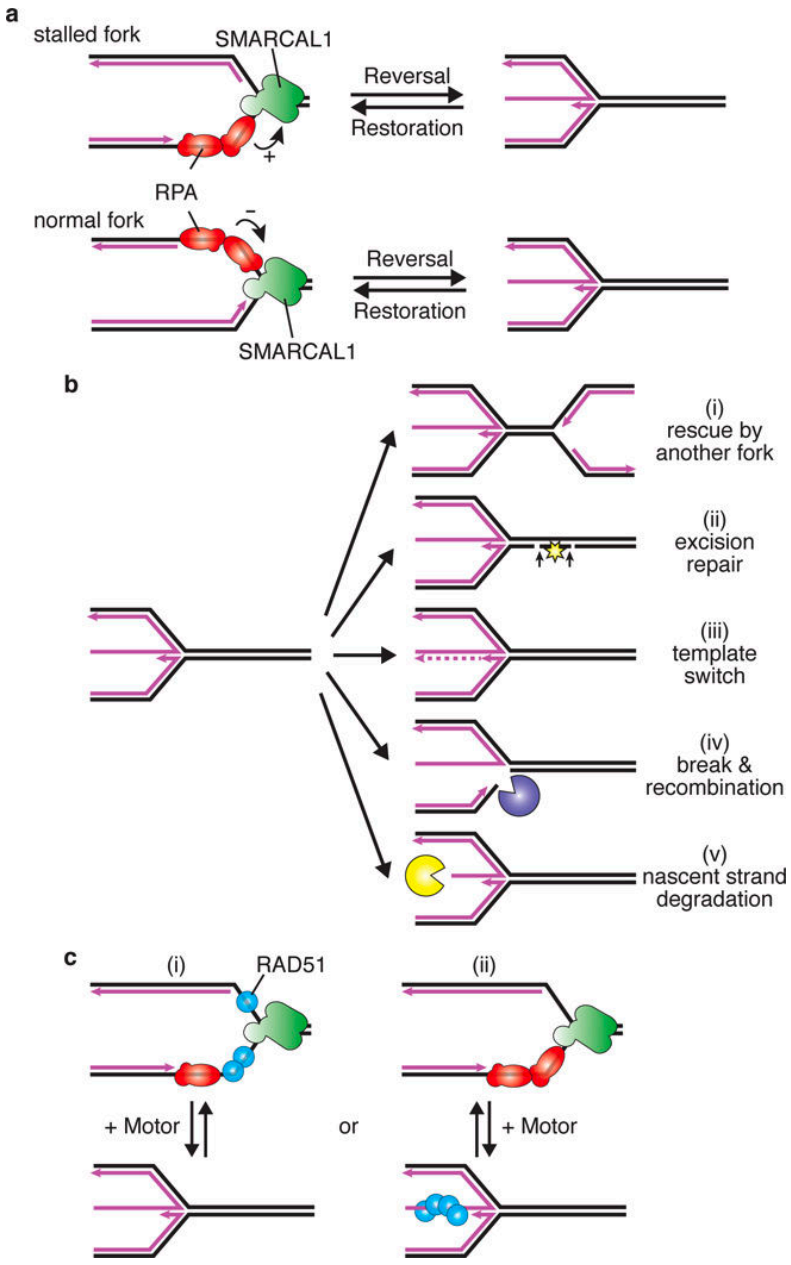

Figure 8. Replication fork reversal. (**a**) Replication fork reversal of stalled forks through SMARCAL1 stimulation of RPA bound to the leading strand template (top). In contrast, SMARCAL1 is bound to the lagging strand in a normal, non-stalled fork. Black lines show template DNA, pink lines show nascent DNA. (**b**) Reversed fork structures represent intermediate structures in the mechanism of fork stabilization and restart, but remain somewhat sensitive to nuclease processing. Reversed forks can be processed further, leading to the outcomes illustrated in (i) to (v). (**c**) Two models for how RAD51 may be involved in promoting fork reversal. This figure is taken from Bhat et al., 2018, illustrating the mechanism of replication fork reversal[87].

One of the major contributors to both DSBs and CNVs is replication stress, which occurs endogenously but can also be enhanced by pharmacological treatments that cause pausing, collapse, or breakage of replication forks. Replication forks refer to the Y-shaped structures indicating genomic sites where DNA replication takes place by moving the replication fork and its associated replisome complex, containing DNA helicase, polymerases and more, along the DNA template. In general, most replication stress (examples described below) only leads to a temporary pausing or slowing down of the replisome, or may not affect the replication fork at all when the stress resides on the lagging strand and the lagging strand polymerases that generate the Okazaki fragments can bypass the lesion. In contrast to pausing, when replication stress such as obstructive DNA damage is present on the leading strand there is a higher chance of longer-term replication fork stalling or arresting and, in most cases, uncoupling of the replicative polymerase ad helicase activities, which is characteristic of replication stress and requires resolving by the repair machinery, involving processes such as fork reversal (described in more detail below, **Figure 8**) and restart, but which also potentially result in DSB formation[86]. However, in many cases replication can still be completed because another fork that initiated replication from an adjacent origin of replication can take over when the damaged fork and obstructing stressor have been removed. Alternatively, in rare situations DNA synthesis needs to be completed from the stalled fork that is stabilized and restarted by actors of the replication checkpoint[88]. When stalled replication forks fail to be stabilized they will collapse, in a process called fork collapse that may entail several processes such as dissociation or disassembly of the replisome proteins and DSB formation, although the latter only occurs in a subsequent round of replication in general[86,89]. Fork collapse and breakage also occur when two replication forks experience head-to-tail collisions[86].

Replication fork progression may for example be obstructed by complex DNA structures such as G-quadruplexes (G4), which often form at telomeric regions and in guanine-rich regions where hydrogen bonds form highly stable tertiary DNA structures that block replication forks, resulting in a DSB[90]. While G4 are the best understood example, other pre-existing and complex DNA lesions, such as base alterations and inter-strand crosslinks caused by ionizing radiation or reactive oxygen species, can all cause the polymerase and helicase activity to be stalled, culminating in replication fork collapse and ultimately in formation of DSBs[91,92–94]. Another replication-associated mechanism that can give rise to DSBs in the form of repeat expansion is DNA polymerase slipped strand mispairing (SSM)[28]. SSM can occur when a repetitive genomic sequence is replicated and mis-paired, resulting in displacement of the DNA strand with possibility of incurring a DSB. In sum, both complex DNA lesions and SSM may interfere with replication and result in DSBs. The origin of these DSBs is most likely replication fork stalling followed by fork reversal, a protective mechanism during which the replication forks reverse their direction in order to gain protection against degradation by nucleases involved in DNA damage repair, giving rise to so-called Holliday junction-like structures[87,95] (**Figure 8**).

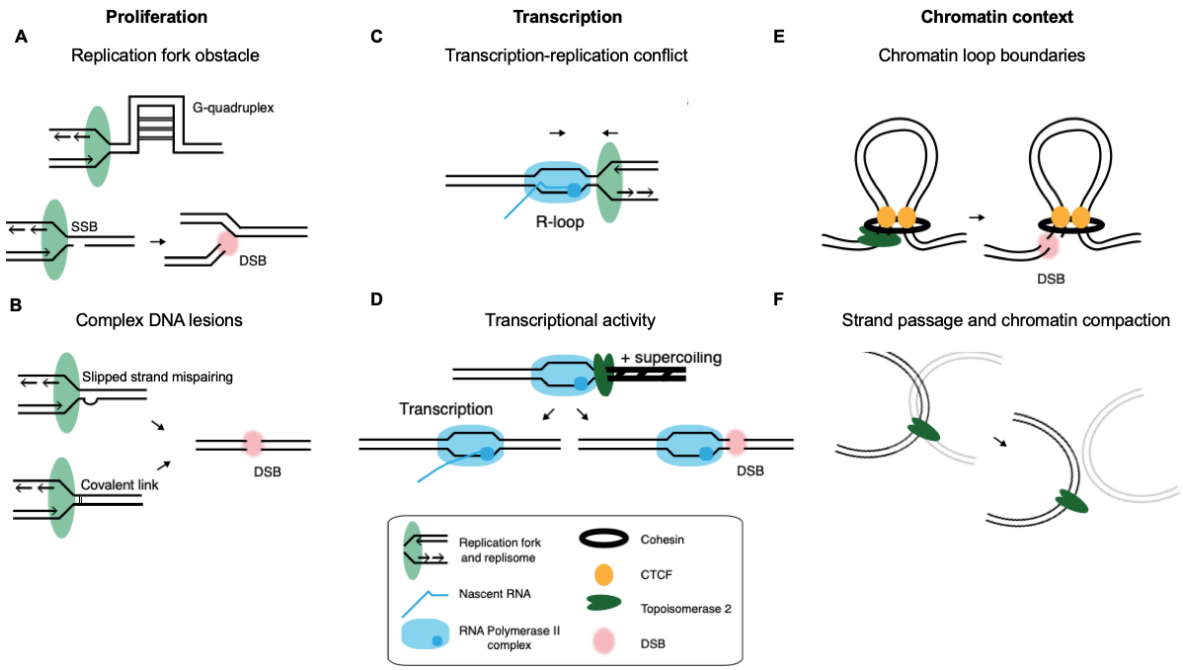

Figure 9. Scheme of mechanisms responsible for DSB formation. Endogenous sources of DSBs can be classified into three distinct main types: DNA replication, transcription and chromatin folding. **(A)** During DNA replication, obstacles such as G-quadruplexes can lead to replication fork stalling. Consequential processing of the structure by the repair machinery that is either recruited to or travels along the replisome may then lead to a single-strand break (SSB) on the leading strand. If not resolved, this lesion gets converted into a DSB during replication. **(B)** Complex DNA lesions like covalent inter-strand crosslinks or slipped strand mispairing may cause similar obstacles during replication, resulting in DSB formation. **(C)** Transcription-replication conflicts can occur when transcription complexes and a replication fork encounter each other. This encounter is particularly detrimental when it occurs head-on and when the transcription complex forms R-loops that concomitantly stabilize the association of the RNA polymerase with the DNA. **(D)** Transcriptional activity itself is associated with topoisomerase 2 (TOP2) endonuclease activity, particularly at gene promoters. TOP2-mediated DSBs are a physiological phenomenon, but when failed to be repaired properly, they result in persistent DSBs. Notably, TOP2 is also active upstream of the replication fork to release torsional stress (not shown). **(E)** Genomic regions that experience torsional stress at 3D genome loop boundaries require TOP2 action to be resolved. These fragile sites are enriched in DSBs as a consequence of chromatin looping and local activity. **(F)** During differentiation, chromatin undergoes global compaction in association with regulating accessibility and gene activity. This happens genome-wide at many sites, but an example of a large-scale reorganization of accessible chromatin is the massive rearrangement of heterochromatin in rod photoreceptor cells, which concentrate all heterochromatin in the nuclear center. As genomic loci are rearranged, strand-passage is mediated by TOP2 action. After strand passage, the transported segment is released from the clamp and the broken ends of the gate segment are re-ligated.

In addition to DSBs emerging during the replication process itself, proliferation exacerbates the mutational burden by collisions between the transcription and replication machineries (Fig. 9C). A common occurrence during transcription is the formation of a stable three-stranded RNA:DNA hybrid (R-loop), which can form when the newly generated RNA hybridizes to its complementary DNA strand and as such displaces the other strand into a looped configuration. R-loops formed during transcription might be associated with transcription termination[96]. However, R-loops may cause a similar obstacle to the replication

fork as is the case for G4. This is underlined by a recent study that showed that R-loop removal is required to maintain genome integrity following production of DSB in active genes. R-loops associated with DSBs sites could be resolved by the RNA:DNA helicase Senataxin by recruiting DNA-repair factors[97]. This is supported by the observation that some DSB-causing agents increased DSB burden in Senataxin-depleted cells, whereas radiation did not. Finally, in absence of Senataxin, spatially clustered R-loops and their associated DSBs gave rise to more translocations[97]. This is a relevant cause of genomic fragility, since DSB-associated R-loops are particularly likely to form in longer genes, which in some cases require longer than one cell cycle to transcribe and are therefore more likely to lead to machinery collisions[98].

1.2.4 Endogenous sources of DSB formation: transcription

The second major challenge to genomic integrity is transcription itself, independent of replication. Transcription is associated with supercoiling at promoters and torsional stress on the DNA strand[99,100]. This force can be alleviated by TOP1 and TOP2 topoisomerases, involving a highly reversible transient SSB or DSB, respectively[99,101]. These transient breaks can be subject to faulty repair or lead to lasting damage. To release torsional stress, topoisomerases form cleavage complexes (TOPcc), catalytic intermediates that are normally rapidly reversible and that have no strict sequence preference and no dependency on supercoiling. Hence, it is assumed that topoisomerase activity is strictly regulated to prevent its action on supercoiled regions that are required for transcription and replication[99,101–103].
Various anticancer drugs act as topoisomerase blockers that target the TOPcc to stabilize (‘trap’) TOPcc to the DNA strand. This prevent further processing or immediate repair of the cleavage site and remains present for a longer period of time. Moreover, TOP1 occasionally fails to relegate the DNA strands, which also generates a trapped TOP1ccs[104]. Trapped TOPccs are pathological and can induce Pol II arrest and stall transcription complexes. Moreover, collision of a replication fork with a trapped TOP1cc can either result in an irreversible TOP1cc and a DSB or a reversed replication fork[101]. The processing and removal of TOP1ccs occurs via various repair routes that may introduce DNA damage, in particular when the bases around the trapped complex are either abasic sites or already carry damage. For example, when a TOP1cc forms on one strand while the other carries a nick, a DSB forms that may be (erroneously) resealed by TOP1 creating a mutation[101]. Indeed, aberrant TOP1 DNA lesions have been found to be pathogenic and related to genome instability syndromes and neurodegeneration[105].

TOP2 isoforms TOP2A and TOP2B, in contrast to TOP1, form a transient DSB and enzymatically re-ligate it as part of their physiological cycle of action. Failure to re-ligate is called abortive catalysis of TOP2, which can occur due to topoisomerase poisons or by spontaneous abortion for other reasons. This leads directly to DSB persistence, requiring removal of the TOP2ccs and DSB repair[104]. Trapped TOP2ccs induced by the TOP2 poison topoisomerase etoposide that has frequently been used in cancer therapy are strongly linked to the emergence of oncogenic translocations that underlie secondary leukemias[106]. In addition to drug-induced stabilized pathogenic TOP2ccs, spontaneous abortive catalysis of TOP2 has in recent years been shown to be much more frequent than previously anticipated[104]. Cells can process and remove TOP2ccs by degradation mediated by the proteasome, followed by repair of the remaining DSB. Intriguingly, a recent study

demonstrated that such proteasome-mediated DSB repair following degradation of the TOP2ccs is highly error prone and that proteasome inhibition led to suppression of the DNA damage response and protection against etoposide-induced genome instability. Instead, the TOP2 enzyme was found to uncouple itself in an error-free manner[107].

In the nervous system, neuronal activity triggers the rapid transcription of early response genes (ERGs) and late response genes (LRGs), and the transcripts generated during neuronal activity-dependent gene transcription mediate lasting changes to neuronal morphology and synapse organization[108,109]. Importantly, DSBs have been shown to rapidly form and accumulate across multiple brain regions following behavioral stimulation of mice[110]. Moreover, depolarization of neural cells was found to induce DSBs, particularly in the promoter regions of neural ERGs[111], which are genes involved in neuroplasticity and regulation of neural signaling. These targeted DSBs have been proposed to be part of a physiological, controlled gene activation process, acting as a switch that turns on transcription through TOP2 action, as their expression is not induced in the absence of TOP2[111]. As such, TOP2 activity not only serves to release torsional stress, but seems required for successful transcription. In line with this, TOP2 was recently found to facilitate enhancer-promoter interactions and RNA polymerase loading during transcription[112]. As genes differ in their composition and expression patterns[113–115], it is conceivable that stochastic errors in the repair of transcription-regulating DSBs might result in the formation of CNVs[107,116]. Moreover, disruptions in the activity-dependent transcription programs mentioned above are thought to underlie the development of various neurological disorders, including NDDs, major depressive disorder, and addiction[117–119].

1.2.5 Endogenous sources of DSB formation: 3D chromatin folding

The third mechanism affecting DNA fragility is the folding of chromatin in the nucleus, which is referred to as 3D genome organization[120]. 3D genome organization is predominantly studied in interphase, during which chromosomes fold into more or less dynamic structural genomic units or domains referred to as topologically associated domains (TADs)[121], mediated by chromatin loop extrusion and architectural proteins such as CCCTC-binding factor (CTCF) and cohesin[122]. I will discuss in more detail how these processes contribute to shaping the 3D genome of different cell types in section 1.3.

TADs and chromatin loops at diverse length scales serve to spatially organize various genomic transactions in the nucleus. Loop anchors, the sequences that form the base of a loop or TAD and that often harbor CTCF binding sites, provide contacts between sites far apart along the linear genome, thereby for example allowing regulatory regions such as enhancers and gene promoters to come into spatial proximity[123,124]. Recently, these loop anchor regions were found to be enriched in DSBs and represent hotspots of structural variation[120,125,126]. DSBs around chromatin loop anchors are thought to occur as a consequence of TOP2 activity at CTCF sites in the presence of cohesin, serving to dissipate the torsional stress generated during transcription and the topological constraints that arise as a result of loop extrusion dynamics. The observed enrichment of DSBs at loop anchors indicates that these regions may determine local fragility, in particular when highly expressed genes reside in their vicinity[120]. Another role of 3D genome dynamics is regulating DSB repair. As discussed below in more detail, depending on a cell's function, various loci relocate in 3D space over the course of

cell differentiation[127,128]. An example of such a developmental reorganization takes place across neurodifferentiation, where specific enhancer-promoter loops and TADs are altered, indicating reorganization of the nucleus[119,129,130]. In addition to cell differentiation, 3D genome organization also changes during the different phases of the cell cycle, with the most pronounced changes occurring during mitosis, as revealed by high-throughput chromosome conformation capture (Hi-C)[131], which I will discuss later in more detail.

Considering that different phases of the cell cycle are known to have different active DNA repair pathways, different topoisomerase isoform preferences, and different 3D genome topologies, it is reasonable to expect interdependence between these dynamic processes. Interestingly, recent findings on chromosome dynamics revealed that the 3D topology of damaged chromatin is locally stabilized to facilitate DSB repair and protect the DNA ends from aberrant processing. Depletion of proteins involved in this stabilization was shown to disrupt this arrangement, decompact the DSB-flanking regions, mislocalization of DNA repair proteins, and excessive resection of the DSB ends[132]. On the other hand, other recent work studying the phenomenon of increased DSB repair during sleep in zebrafish suggested increased levels of chromosome dynamics in neurons, but not in two other cell types. However, the authors also showed that DSBs accumulate during wakefulness, while chromosome dynamics are low, thereby potentially confirming the reduced dynamics observed in relation to DSB stabilization. Sleep then appears to increase the dynamics in order to repair DSBs encountered during the day, while few new DSBs are formed during sleep, thereby suggesting that genomic maintenance is one of the important restorative functions of sleep[133]. One of the important drivers of this chromatin motility is a different TOP2-dependent mechanism called strand passage[134,135], in which chromatin structures are facilitated to pass one another by TOP2 cleavage of one DNA strand, a mechanism that allows for resolution and re-formation of facultative chromatin[134,135,136]. As such, strand passage represents yet another process in which transient DNA breaks are necessary for cell specification[137,138].

Taken together, DNA replication, transcription, and chromatin dynamics, as well as the interplay between these processes, all contribute to endogenous formation of DSBs and potential loss of genomic integrity, with topoisomerases and DSB repair pathways as key players. To find fragile sites that are relevant for disease it is therefore important to leverage methods that can map which parts of the genome break more frequently and the associated mechanisms. In the next sections, I will discuss methods for mapping of DSBs genome-wide and how such methods can reveal disease-related DSB hotspots in the genome.

1.2.6 DSB form non-randomly and preferentially in fragile regions

Although DSBs can occur throughout the genome, as already discussed above certain genomic regions such as the promoters of active genes are more prone to undergo breakage[139]. Another type of regions in the genome that frequently break under specific conditions are so-called common fragile sites (CFSs). CFSs are defined as the cytobands in metaphase chromosome preparations in which breaks or gaps can be seen when DNA synthesis is partially blocked[140]. CFSs represent hotspots for chromosomal rearrangements and genomic alterations frequently found in cancers, in particular recurrent deletions[140]. Several genomic features of CFSs have been proposed to contribute to their fragility,

including i) the fact that they tend to contain AT-rich sequences that are more difficult to replicate due to their tendency to form secondary structures; ii) their replication in late S phase; iii) a shortage of origins of replication inside them, and iv) the tendency for very large genes to overlap with CFSs (>80% of CFSs overlaps with genes larger than 300 kb)[140,141]. A particular type of CFSs are early replication FSs or ERFS, which represent regions that replicate early, harbor actively transcribed gene clusters in open chromatin, and are marked by a high GC-content, a high density of replication origins, and repetitive elements[142]. Endogenous DSBs spontaneously form in ERFS during replication, and conditions that cause replication stress further increase break formation at these sites[142].

Another type of fragile genomic regions is recurrent DSB clusters (RDCs), which are hotspots of endogenous DSBs detected around the transcriptional start site (TSS) of highly transcribed genes or along the gene body of long, neural-specific genes in mouse neuronal stem/progenitor cells[143]. Human orthologues of these genes have been previously implicated in NDDs, including SCZ and ASD[37,143,144]. Interestingly, many RDCs were found to be conserved in normal human neural stem cells[145] , and in neural stem cells derived from patients with a particular form of ASD characterized by increased susceptibility to replication stress[146]. At present, it is not clear how these RDCs affect neural function or when they arise during human neurogenesis. In order to improve our understanding of the impact of recurrent endogenous DSBs during human neurogenesis, genome-wide maps of endogenous DSBs that arise are required.

1.2.7 Methods for identifying DSBs in the genome

To understand the fragility of the genome, in the past years several imaging and sequencing-based methods have been developed to map the frequency and location of DSBs[139,147].

Indirect methods for profiling of DSBs

One of the canonical ways of detecting DSBs through imaging strategies is monitoring the accumulation of DNA-damage response proteins at break sites, such as TP53 binding protein 1 (53BP1) [148]. 53BP1 is a key player in DSB repair that promotes NHEJ repair by rapidly accumulation on the chromatin surrounding the detected DSB and antagonizes DSB overhang resection[149,150]. 53BP1 has been reported to form large focal clusters which form to facilitate DSB repair[151,152]. Another DSB marker is the histone variant H2AX phosphorylated on serine 139 (gammaH2AX) which spans damaged regions [84,153]. GammaH2AX decorates the sequence surrounding a DSB for several kilobases and can be detected as bright fluorescent foci under the microscope using immunofluorescence. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) using gammaH2AX specific antibodies has been used to map the genomic locations of DSB in yeast[154] and mammalian cells[155]. Although ChIP-seq allows identifying DSBs genome-wide, its main disadvantages are that the method is indirect (it does not detect the DSB itself, but relies on recruited proteins as markers that may not be DSB-specific) and cannot identify DSBs at single-nucleotide resolution[147].

A second group of indirect methods for DSB profiling detects DSBs by relying on integration events of ectopic pieces of DNA into the DSB site or on capture of the DSB ends via generated translocations or chromosomal rearrangements. Examples of these methods

include translocation-capture sequencing (TC-Seq), GUIDE-seq, integrase-defective lentiviral vector (IDLV)-mediated DNA break capture, and linear amplification-mediated high-throughput, genome-wide, translocation sequencing (LAM-HTGTS), as reviewed recently in[139,147]. The latter, LAM-HTGTS[156,157] detects 'prey' DSB ends genome-wide through their translocation to a 'bait' DSB end generated via CRISPR/Cas9 at a fixed genomic location. Bait-prey combinations are amplified from isolated gDNA and then ligated to sequencing adapters enabling paired-end sequencing. LAM-HTGTS has for example been harnessed to identify RDCs in primary mouse neural stem/progenitor cells[145]. Although these indirect approaches do enable identification of actual DSB ends at near nucleotide-resolution, depending on the level of end resection that occurs during NHEJ, they do all rely on an active DSB repair pathway and live cells, making the methods less applicable to certain types of cancer cells and less flexible to a variety of sample types, respectively[139,147].

Direct methods for profiling of DSBs

The number of direct methods is extensive and includes methods with lower and higher resolution. Lower resolution including methods such as Break-seq and DSB-seq label DSB directly using biotinylated nucleotides incorporated using the terminal deoxynucleotidyl transferase (TdT) DNA polymerase, followed by gDNA fragmentation and immunoprecipitation of fragments with biotin-labeled DSB ends, followed by sequencing[147]. A few of the direct labeling methods label DSBs in extracted genomic DNA, rather than directly in situ in the (fixed or non-fixed) nuclear chromatin. Although convenient, this approach increases the chance of identifying false-positives and non-endogenous DSBs introduced during sample handling.

Methods for nucleotide-resolution DSB mapping include BLESS[158] and its successors BLISS[159] and sBLISS[160], END-seq[161], DSBCapture[162], several adaptations of these approaches. The first method for direct, genome-wide, nucleotide-resolution *in situ* mapping of DSBs was breaks labeling, enrichment on streptavidin and next-generation sequencing (BLESS)[158]. In BLESS, cells are cross-linked with formaldehyde, and lysed to extract intact nuclei, after which (endogenous and/or induced) DSB ends are blunted, 5′-phosphorylated and *in situ* labeled by a short hairpin-like biotinylated adapter. The adapter-bound DSB ends are then captured on streptavidin beads and ligated to another hairpin-like distal adapter, after which polymerase chain reaction (PCR) amplification using primers binding to proximal and distal adapters is used prior to library preparation for high-throughput sequencing.

The most recent advancement in the BLESS family of genome-wide DSB mapping methodologies is sBLISS (in-suspension breaks labeling in situ and sequencing)[160]. Many of these methods have also been applied to chart off-target DSB events of CRISPR/Cas-based genome editing approaches[163].

Genome-wide mapping of DSBs by the BLISS method

When investigating fragile sites in relation to genome stability it is important to know where exactly damage is incurred. Breaks Labeling In Situ, Enrichment on Streptavidin and Sequencing (BLESS) [158] and later Breaks Labeling In Situ and Sequencing (BLISS) [159] detect DSBs in their native chromatin context by ligating DSB ends to specialized adapters in cross-linked nuclei. To map endogenous DSBs, it is important to use a direct labeling technique

capable of detecting even transient DSBs. BLISS, allows generation of a snapshot of all DSBs including very transient ones, as well as intermediates of DSB repair or replication fork remodeling. This is different from non-direct labeling techniques like where detection is based on repair and thus differences might occur between DSBs over a period of several hours. Alternative approaches using direct detection of DSBs include iBLISS, qDSB-seq which build on the previous techniques. Break-seq, DSBcapture, END-seq offer alternative workflows but are all direct detection methods similar to BLISS. The advantage of BLISS over the others is that it detects DSBs in their native chromatin context by ligating DSB ends to specialized adapters in cross-linked nuclei without use of agarose plugs[159].

DSB mapping in the nervous system

Previous applications of DSB mapping in neural progenitors was mainly done using high-throughput genome-wide translocation sequencing (HTGTS), which is an indirect DSB labeling technique dependent on DNA repair pathways [164]. In particular, they used a DNA-repair deficient cell line which is poised to maintain loose DNA break-ends for a longer period of time due to XRCC4 and P53 mutations. Work of this group demonstrated that a small group of long neural genes accumulate DSBs as a consequence of transcription, which in turn result in the formation of translocations (Wei et al., 2016). Neural progenitors undergo a period of rapid expansion correlated with a short cell cycle and positivity for gammaH2AX, a hallmark of double stranded DNA damage. At the end of this progenitor expansion, approximately 50% of cells undergo apoptosis[165]. 27 recurrent DSB clusters (RDCs) were found in neural stem/progenitor cells, the vast majority of which overlap with long, transcribed, and late-replicating genes[143]. The genes affected by these break clusters could be divided in three classes related to cell adhesion, neurogenesis and synapse plasticity[144]. The majority of these RDCs are conserved between mouse and human, supporting a functional mechanism for this subset of genes[146]. A large proportion of recurrent DSB clusters occurs after commitment to neural lineage[145]. While dominant homologous recombination in ESCs might protect RDC DSBs from occurring, human NES cells do have an active HR-repair, whereas progenitors are likely to be more dependent on C-NHEJ pathways. This was supported by very recent findings studying ESC to NPC transition in mouse[145].

In this large body of work using repair-based HTGTS several features of genomic instability in the neural system has been uncovered[37,144–146,156,164]. The question remains whether endogenous DSBs mapped by BLISS in an unperturbed system behave similarly to those detected using HTGTS and if they may give complementary insights on the effect of genomic instability. While BLISS is a powerful method to detect DSBs at any particular time, it is limited in that BLISS data represents a snapshot of the sample at time of fixation and thus is particularly suited to show where DSBs arise. Whether the identified DSBs are repaired or result in structural changes in the genome.

Novel approaches in the field

In addition to identifying genomic coordinated, new methods are being developed to further investigate the characteristics of genome fragility. coverage-normalized cross correlation sequencing (CNCC-seq) shows promise to add more details about the loose-end overhangs and specifics of the mechanism[166]. TOP2 inhibition by etoposide for instance increases break densities at promoters and TSSs, but reveals a skewed profile with increased genome-wide

3′-overhang end structures, and displays the progression of 5′ to 3′ resection[166]. This analysis approach elucidating the DSB end structure and allows for patterns to be identified within noisy and sometimes sparse data. Moreover, several new technologies are beginning to investigate also DNA SSBs, but their application has been technically limited due to challenges regarding their resolution and empirical reliability of capturing transient events. SSB-Seq[147], SSiNGLe[167], GLOE-Seq[168] are showing promising results by, in slightly different ways directly tagging 3′-OH termini of DNA breaks. A more complex approach is Nick-seq which utilizes both nick-translation and TdT-mediated tailing[169]. A huge step forward in this are new SSB mapping methods using 5-bromo-20-deoxyuridine BrdU incorporated at the break sites, allowing a direct readout of interruptions to the DNA[170]. This protocol has been used in several slightly different ways[171–173]. However, both direct and indirect means of measuring SSBs have proven difficult to implement reliably due to high background signal, large input and fixation issues. As the technology develops further and gets applied more regularly to mammalian systems, this will open up a whole new field which has yet remained out of reach.

1.3 3D GENOME IN THE NERVOUS SYSTEM

1.3.1 Introduction to 3D genome organization

The eukaryotic nucleus is a highly organized organelle within which gene expression is highly regulated[174] and cell function and identity at any time are a consequence of the gene expression program. A pivotal factor in the correct activation and maintenance of cell type-specific gene expression programs is the spatial orchestration of the genome, which is distributed across 23 pairs of chromosomes, packaged and compacted into chromatin, and smartly and dynamically folded in order to fit inside the nuclear space, while maintaining access to the required genes and regulatory regions at the right time in cellular life. In brief, this organization is referred to as the 3D genome, and in the past 1-2 decades there has been an enormous increase in our knowledge of 3D genome folding, in large part thanks to a variety of sequencing- and imaging-based techniques[175], of which I will discuss a few below in section 1.3.4.

In interphase, each chromosome takes up a certain space in the nucleus that is referred to as its chromosome territory (CT) (**Figure 10**). Although the relative positioning of a given chromosome's CT may differ between cells and during consecutive cell cycles, certain gene-rich chromosomes with lots of transcriptional activity have a strong preference to reside near the nuclear center, while others that are more gene-dense tend to localize in the nuclear periphery adjacent to the nuclear lamina. Within each CT, chromatin organizes into domains of self-associating chromatin, which is generally thought to be driven by homotypic interactions between chromatin decorated with similar histone modifications and comparable expression state. Based on results from many high-throughput chromatin conformation capture (Hi-C) experiments carried out over the years by different labs, it has become evident that chromatin folds into dynamic structural units within which genomic regions frequently associate and form chromatin contacts. These units, called topologically associating domains (TADs), differ in size from 100 kb to 5 mb. TADs cluster together with other TADs, often with a similar chromatin state, to form so-called A/B compartments that respectively refer to

more active and less condensed chromatin, and more silenced and compacted chromatin (**figure 10**). Although TADs were initially identified in bulk Hi-C assays, recent single-cell Hi-C (scHi-C) assays as well as imaging approaches have revealed the presence of TAD-like behavior at the single cell level too, albeit with more variability in TAD border regions between cells.

TADs take shape through chromatin looping between two boundary/base regions, which are frequently associated with architectural protein complexes including CTCF and/or cohesion (**Figure 10**). Below the level TADs, sub-TADs have been defined, and both are considered fundamental physical units of genome organization in individual cells as they allow for intra-TAD contacts, while minimizing inter-TAD contacts and thereby for example enabling insulation of regulatory action. Mostly within TADs, but also across TAD borders, additional and potentially more dynamic chromatin loops form, for example between gene promoters and regulatory elements such as enhancers. Importantly, although such regulatory (enhancer-promoter) contacts generally occur over shorter distances, they may loop over several Mbs and can even occur between different chromosomes.

1.3.2 Spatial genome rearrangements during development

Pluripotent stem cells (PSCs) such as embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) are able to self-renew and to give rise to all the embryonic germ layers. Pluripotent cells are characterized by a highly plastic chromatin landscape and a relatively small group of core regulators that maintain the pluripotent state, while simultaneously enabling controlled exit of this state during early differentiation. Transcriptionally, pluripotency is unique as well, as PSCs have a characteristic widespread low-level transcriptional activity of exons as well as introns and non-genic regions, which altogether cause RNA levels to be doubled compared to differentiated cells[176]. While PSCs are transcriptionally promiscuous, expression of key genes underlying lineage fate determination is suppressed in order to maintain pluripotency. To sustain and help regulate this particular state, pluripotent chromatin is relatively devoid of heterochromatin and shows highly dynamic association of architectural proteins and high histone turnover as well as instability of histone H1 binding[177]. To embark upon differentiation, the core of the pluripotency transcription circuitry collapses, lineage-specific genes are activated, chromatin structure compresses and widespread transcription ceases. Upon cell fate establishment, gene expression programs are generally stably maintained and transferred across daughter cells. Differentiation is further accompanied by a severe transformation of the regulatory landscape, involving among others rapid enhancer switching, allowing developmental fine-tuning of gene expression even when similar gene activity is maintained.

As the 3D genome can help to instruct, establish, and maintain expression regulation, differentiation is accompanied by a spatial reorganization across all scales of 3D genome organization. One feature that seems specific to PSC genomes is that the B compartment appears less strictly organized, or at least harboring fewer specific long-range contacts that are shared among cells in a population, and with chromocenters being more dispersed and randomly positioned than in differentiated cells. Hence, differentiation is accompanied by i) A/B compartment rearrangements early on, even before transcription programs change[178], ii) changes in radial positioning for various key pluripotency factors, iii) loss of the 3D

chromatin interactions between the key pluripotency loci, iv) increase in the specific organization of the B compartment, and v) relocation of the inactivated X chromosome towards the periphery in female cells that underwent X chromosome inactivation.

1.3.3 The 3D genome during neurodifferentiation

During development of the brain, various types of neuronal cells emerge from progressively differentiating stem cells during a process called neurodifferentiation. This is accompanied by a rearrangement of the 3D genome and the upregulation of neuron-specific genes[179]. Most studies of 3D genome dynamics during development of the brain have focused on TADs and chromatin loops[122] and not on so-called higher-order features of 3D genome organization (e.g. beyond the level of individual TADs). Although the global landscape of TADs remains relatively unaltered, brain development is accompanied by a rewiring of intra- and inter-TAD contacts[129,180]. Furthermore, the average size of a TAD has been reported to increase when human neural progenitor cells differentiate to neurons[181]. Both in human and mouse, neurodifferentiation was found to be marked by a widespread pruning of short-range contacts (<100–200 kb), and a concomitant increase of the number of longer-range contacts leading to the emerging of enhancer-promoter loops bound by neuron-specific transcription factors[129,181,182]. In agreement with this, adult neurons have been observed to harbor stable long-range contacts and larger TAD structures[180,183,184]. At the larger-scale of 3D genome organization, neurodifferentiation has been reported to be accompanied by i) A/B compartment changes [129,185], ii) repositioning of lamina-associated domains (LADs)[186], and iii) neuron-specific radial repositioning of specific chromosomes[187], which is all in line with previous reports of developmental 3D genome changes[129].

These neurodifferentiation-associated changes to the 3D genome affect gene loci that have previously been related to the risk for NDDs including ASD and SCZ[181,188]. Moreover as discussed above in sections 1.1.1 and 1.1.2, both 3D genome dynamics and NDDs have been linked to DSBs[37,120,126,143,146]. However, at present there is no clear understanding of how 3D genome dynamics are coordinated during neurodifferentiation, nor how these dynamics relate to transcriptional changes and DNA damage.

1.3.4 Methods in 3D genome mapping using imaging

Light microscopy is a key technology in modern cell biology and in combination with immunofluorescence and in-situ hybridization allows investigation of the nucleus. Typically, DNA is dyed using DAPI or HOECHST dyes to contrast the nucleus from the cell body which distinguishes heterochromatin from euchromatin. In addition to standard dyes, many novel less-GC-biased dyes have been developed and are being used. FISH allows investigation of individual loci by hybridizing sequence-targeted fluorescently labeled oligos which when used in high throughput can identify multiple chromosome domains[189]. Super resolution microscopy has become trendy and now allows sub-diffraction multicolor imaging of the nuclear periphery and observed holes in DAPI signal shown in traditional wide-field imaging[190]. Nuclear pore and nucleolus details are important functional structures which play both a regulatory role for the rest of the spatial DNA organization while also being subject to cellular processes This now allows us to investigate how the hypothetical 3D genome organization features hold up and is there are any biases in individual cells or cell types[175,191].

1.3.5 Methods in 3D genome using sequencing

Another powerful method to investigate 3D genome organization is through many novel sequencing strategies. The most prominent one is the family of chromosome conformation capture (3C) methods which measure proximity of DNA elements within the confines of the nucleus. All 3C methods follow a general principle of fixation, enzymatic digestion of chromatin and re-ligation which allows identification of genomic loci which tend to cluster together. Hi-C in turn, quantifies interactions between all possible pairs of fragments simultaneously allowing reproduction of genome-wide interaction maps. Important novel methods in 3D genome are Hi-C, GPseq, DamID, SPRITE among others[192].

Hi-C is a proximity ligation approach that captures the organizational structure of chromatin in three dimensions, allowing genomic sequences that are remote in linear distance to be closer in 3D space[121,193,194]. The Hi-C data-generated high-resolution, genome-wide map of interacting genetic loci can then be used for a variety of genomic applications, such as identifying promoter-enhancer interactions for gene regulation studies, detecting structural rearrangements, and scaffolding contigs for genome assemblies to define chromosomes from scratch. Using Hi-C in 3D genomics offers information on both the sequencing and structure of the genome. Proximity ligation, in which nuclear DNA is crosslinked, is digested with restriction enzymes, biotin-labeled, and proximity-ligated before being sequenced using normal NGS methods. This allows you to obtain crucial 3D genome organization as well as long-term genetic information in a single test.

GPSeq uses a basic and elegant concept: nuclear diffusion from the nucleus perimeter to its inside[195]. By digesting fixed cells with specific restriction enzymes at several time points, rings across the radial orientation of the nucleus are tagged. These open restriction sites can be labeled with so-called YFISH labels for visual inspection of enzyme diffusion or sequencing tags to generate libraries for each of the concentric digestion periods. GPSeq combines the sequencing of genomic loci with varied digestion times into a "GPSeq score," a credible assessment of locus centrality. GPSeq in combination with HI-C allows for 3D genome reconstruction approach to show how a centrality restriction improves recovered architectures. 3D genome architectures built by a GPSeq-informed algorithm are proving useful now that spatial information of nuclear processes becomes increasingly important.

DamID offers an alternative approach by mapping DNA- and chromatin-binding protein locations[191,196]. DamID discovers binding sites by fusing DNA-binding protein with DNA methyltransferase. DamID fuses Dam to a protein or chromatin component. The target protein binds to Dam's in vivo binding site, methylating nearby sites. Methyl PCR then detects the binding sites. This approach permits mapping proteins for which no antibody exists. Chromatin Immuno-Precipitation (ChIP) measures protein binding at specific genomic loci. Unlike ChIP, DamID doesn't require a protein-specific antibody, DamID assays where the protein has been, whereas ChIP assays where it is now.

SPRITE enables genome-wide detection of multiple simultaneously occurring higher-order DNA interactions within the nucleus and provides a global picture of inter-chromosomal spatial arrangement around nuclear bodies[197].

2 DOCTORAL THESIS

2.1 RESEARCH AIMS

The aim of my research was to investigate endogenous fragile sites (DNA double-strand breaks, DSBs) genome-wide and across neural development in multiple ways. First, to identify DSB hotspots and relate them to multiple parallel processes of genome dynamics. Second, to describe the fragility of the genome as a consequence of cell identity. Finally, to contribute to mechanistic insights to how the regulation of transcription takes place in the context of replication stress.

Paper I ("An atlas of endogenous DNA double-strand breaks…"):

- To implement novel genomics methods to a well-characterized neural cell line to better understand how the genome adjusts its activity and spatial organisation during differentiation.
- To find whether specific genomic loci are more vulnerable to endogenous DSBs than others and whether developmental stage has an impact on the generated DSB maps.

Paper II ("Topoisomerase 1 activity during mitotic transcription…"):

- To investigate whether the dramatic changes in chromatin structure across mitosis affect the transcriptional regulatory role of TOP1 and its genome-wide chromatin occupancy.
- To find whether disruption of the RNAPII interaction domain or ectopic degradation of TOP1 promotes DNA damage and/or transcriptional defects.

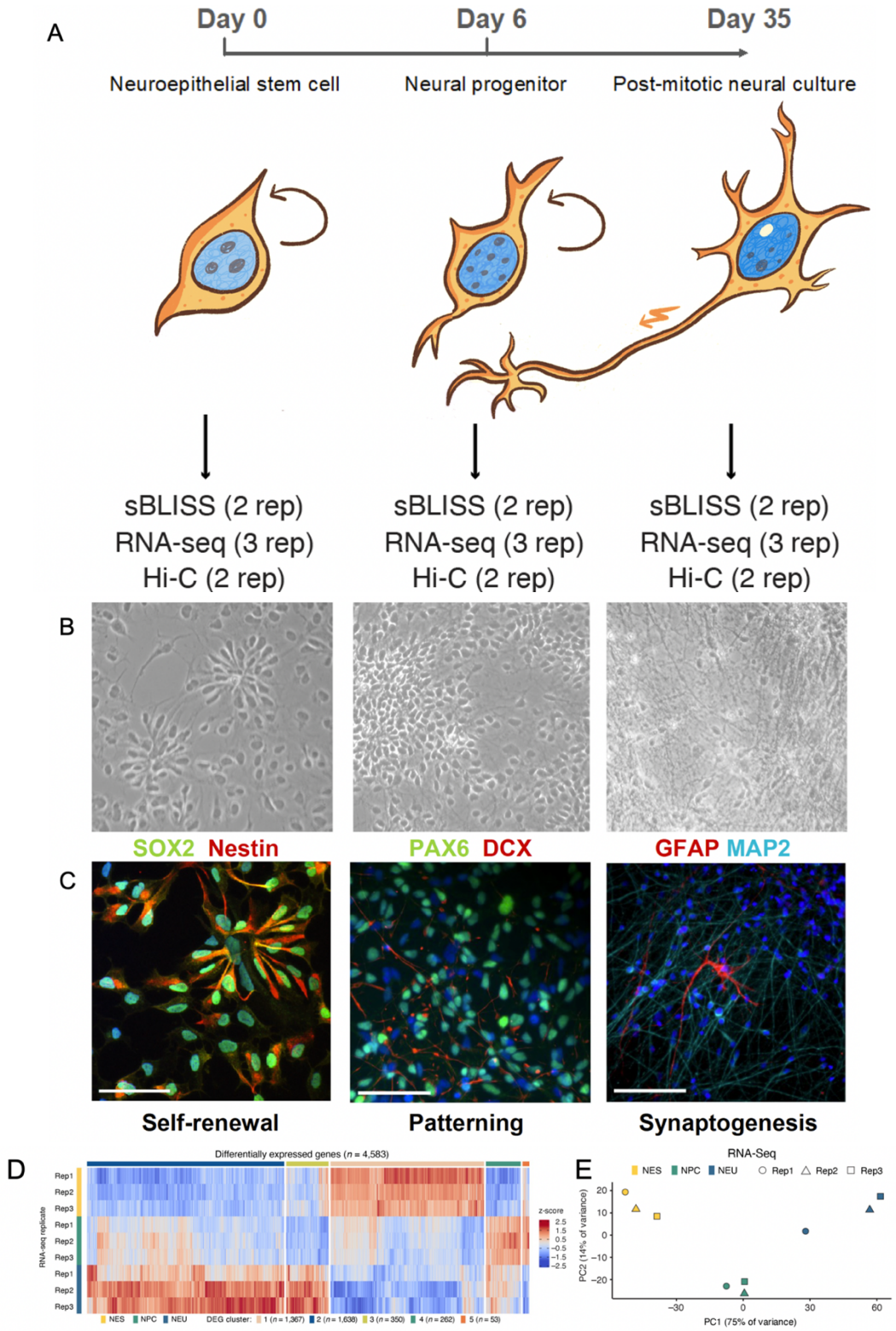

Figure 10. Validation of the model system of human neurogenesis used in this study. (A) Scheme representing NES differentiation to NPC and NEU cells. Cells were harvested at three timepoints and processed for sBLISS, RNA-Seq, and Hi-C. (B) Phase-contrast imaging of live cultures before harvest. (C) Maximum z-projections of wide- field epifluorescence microscopy z-stacks showing the expression of different markers of neural lineage (D) Hierarchical clustering of differentially expressed genes (DEG) between NES, NPC, and NEU cells. (E) Principal component analysis of the RNA-Seq datasets.

2.2 SUMMARY OF RESEARCH PAPERS

2.2.1 Paper I: An atlas of endogenous DNA double-strand breaks arising during human neural cell fate determination

Background

To assess how DSB localize depending on we mapped the genomic DSB landscape of cells at various stages of neural differentiation and correlated our maps with genomic and epigenomic features. In so doing, we provide clues on how DSB formation and their incorrect repair might contribute to the pathogenesis of NDDs. The current view is that transcription-associated DSBs seem to be the main driver of de novo mutations. Indeed, we found that DSBs preferentially form around the transcription start site (TSS) of transcriptionally active genes, as well as at chromatin loop anchors in proximity of highly transcribed genes. This follows from the accumulation of DNA torsional stress and topoisomerase activity in these regions. Interestingly, hotspots of endogenous DSBs were detected around the TSS of highly transcribed genes involved in general cellular processes and along the gene body of long, neural-specific genes whose human orthologues had been previously implicated in NDDs.

Motivation and methods

When investigating the basis of pathogenesis is crucially important to have a well-controlled model system. Here we work with a 3-step differentiation of long term self-renewing neuroepithelial stem cells (NES) derived from a female donor (AF-22 obtained from the KI iPS Cell core facility). This stage of cell type specification is ideal to study due to its' highly controlled environment, lack of environmental stimuli and a constant media for the cells to naturally differentiate. The model represents a developmentally immature neural stem cell state with the ability to progress towards a terminal cell fate and which have been thoroughly characterized[165,198,199]. Interestingly, as neural cells differentiate, they go through sequential transcriptional waves as different developmental processes are initiated[200]. To capture different cell types in this gradual differentiation process, I chose to use: undifferentiated NES cells (day 0) for their self-renewing property and rapid cell cycle progression, thus representing early neural tube development; differentiation media-primed neural progenitors (day 5), which have significantly reduced their proliferation, migrate and produce projections, thus representing cortical radial migration; and 5 week old post-mitotic neural cultures (day 35), which are electrophysiologically active and regulate their synaptic contacts to stabilize neural circuits, similar to what happens in the developing cortex (Figure 3).

Main findings

We set out to assess genome fragility in the form of DSBs in the context of a naturally adjusting dynamic nucleus. Moreover, we performed whole-genome sequencing of both NES cells and NEU cultures to confirm absence of genomic differences or abnormalities. The chosen developmental timepoints were picked to represent important distinct stages of neural specification and determination of cell fate. We first validated the differentiation of the individual timepoints by daily visual inspection, immunofluorescence labeling of molecular processes and performed total-RNA-seq to assess differences in expression of both coding and not coding gene expression (**Figure 10**). The chosen timepoints for the showed high correlation between replicates and were significantly distinct across differentiation timepoints.

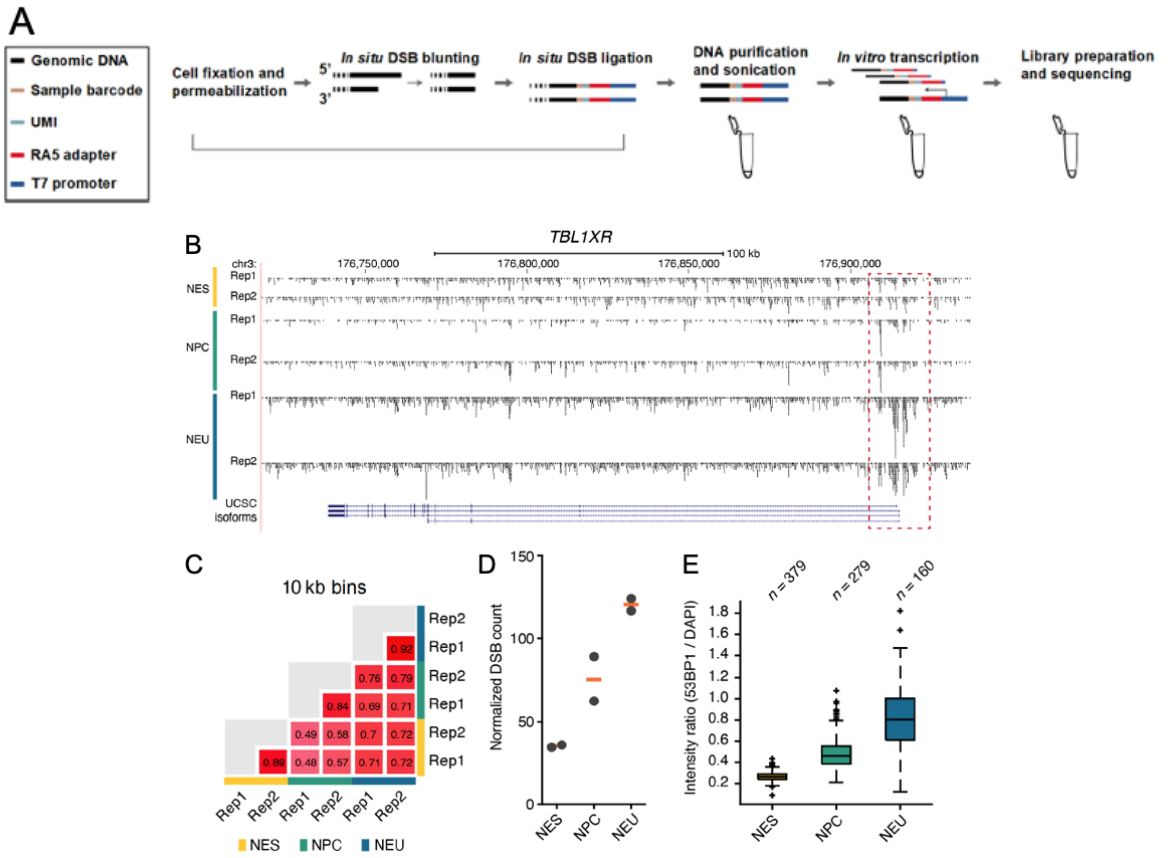

Figure 11. Overview and validation of sBLISS. (A) sBLISS workflow and schematic representation of the adapters used to tag individual DSB ends and to amplify the gDNA sequence downstream by in vitro transcription. UMI, unique molecular identifier. T7 phage RNA polymerase. RA3/5 adapters. (B) Visualization of mapped DSBs along one of the top-fragile genes shown in and using the squish option in the UCSC genome browser. The dashed red rectangles indicate the enrichment of DSBs around the TSS of the two genes (C) Normalized counts of DSB ends detected by sBLISS in each of the six sBLISS datasets described here. Each grey dot represents one replicate experiment. Orange bars, mean value. (E) Normalized 53BP1 nuclear intensity. For each segmented nucleus, we normalized the intensity in the fluorescence channel of the 53BP1 antibody to the intensity of the DNA staining channel. Black dots, outliers.

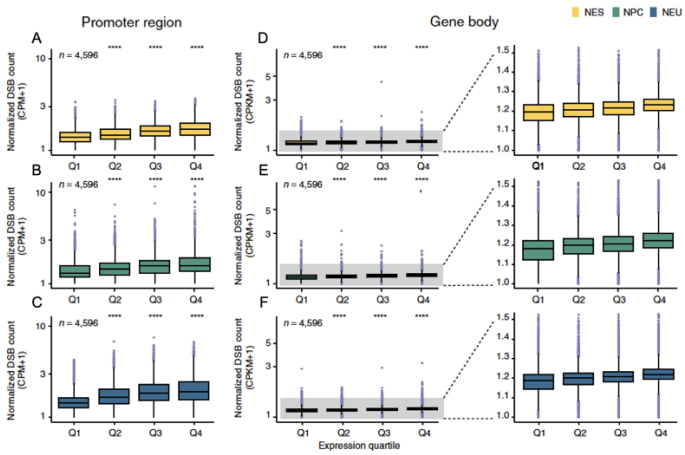

Figure 12. Endogenous DSBs are enriched in the promoter region and along the gene body of highly expressed protein-coding genes. (A-C) Distributions of normalized DSB counts in a 3 kb window around the TSS of human protein-coding genes classified in four different quartiles (Q) based on their expression levels determined by RNA-Seq. (D–F) Same as in (A–C), but for DSBs along the gene body from the first TSS to the last transcription end site of each gene. The part of the boxplots highlighted in grey is magnified on the right.

We first generated two sBLISS biological replicate datasets Our efforts to assess DSB distribution genome-wide sBLISS yielded highly correlated DSB distributions for each developmental cell stage between replicates at different resolutions. We found that sBLISS reproducibly detects endogenous DSBs and that we observe differences within the same cell line, purely subject to the differentiation process (**Figure 11**). In other words, cell type is a determining characteristic in the DSB-landscape as a consequence of developmental changes.

To investigate the activity-induced DSB hypothesis, we correlate sBLISS data with total-RNA-seq derived from the same timepoint. We examined the DSB distribution in the promoter and in the gene body of highly expressed protein-coding genes, which our and other groups have previously shown to be hotspots of DSB accumulation in different cell types, using sBLISS or other genome-wide DSB detection methods. Here we found a correlation of breakage with expression in the same cell types (**Figure 12**). Interestingly, DSBs are enriched at gene promoters. During neural cell maturation, CpG island and their methylation plays an important role in driving maturation processes. Assessing and the promoter's CpG content indicated that CpG-rich promoters are more enriched in DSBs (**Figure 13**). NEU showed an increase in CpG-DSB correlation beyond what we would expect based on expression alone.

We know that during neural cell fate determination the nucleus is reorganized[129,181]. We generated Hi-C data and correlated them with sBLISS revealing that DSBs were enriched in active A compartments, at the boundary between consecutive topologically associating domain (TAD), and around chromatin loop anchors, in line with previous reports linking 3D genome dynamics and genome fragility. Through our integrative multi-method approach we investigate individual cross-chromosome interaction changes find a unique DSB distribution pattern for this fragility in post-mitotic neurons (**Figure 14**).

Finally, we assessed the prevalence of DSBs at genes previously associated with increased risk for SCZ and ASD, revealing that the promoter region and the gene body of these genes are hotspots of spontaneous DSB accumulation and are significantly more fragile compared to the same regions in all other human protein-coding genes, especially in post-mitotic differentiated NEU cells (**Figure 15**).

Conclusions

Through our integrative multi-method approach we corroborate previous findings regarding DSB-fragile loci at TSSs and in relation to high levels of expression. We identify specific genomic sites which are fragile in a cell-type specific manner. We find high levels of similarity between the cell types, but with distinct details at specific genomic sites. Finally, we find a unique distribution pattern for DSB in post-mitotic neurons which might be related to chromatin compaction associated to differentiation. To better understand the relation of DSB fragility and chromatin conformation. Additional orthogonal methods assessing 3D genome conformation are needed. We show a cell type-specific preference for DSB accumulation in specific NDD genes. Interestingly, we find a subset of genes which have increased fragility at the earlier differentiation time points, indicating that these genes might be particularly prone to replication-stress associated DSBs.

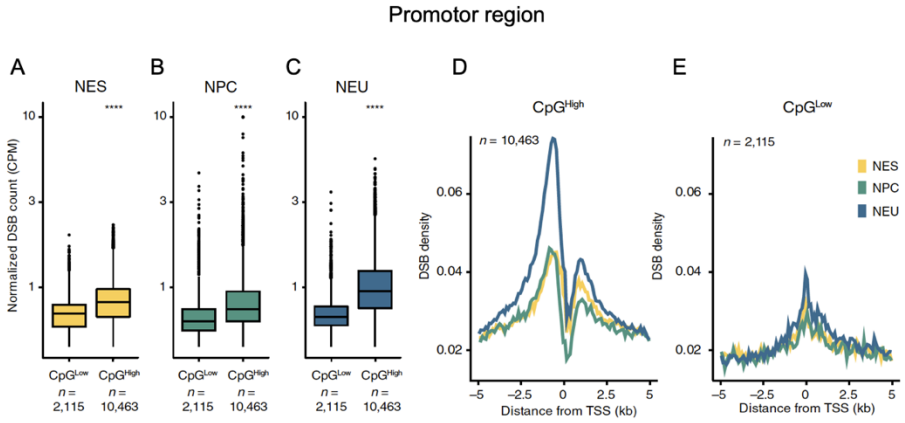

Figure 13. CpG-rich promoters are highly fragile. (A–C) Distributions of normalized DSB counts in a 3 kb window around the TSS of human protein-coding genes, for genes with high (CpGHigh) or low (CpGLow) levels of CpG dinucleotides in their promoter region. (D,E) Metaprofiles of the DSB density around the TSS of human protein-coding genes classified as CpGHigh (D) or CpGLow (E) based on the frequency of CpG dinucleotides in their promoter region. n, number of genes.

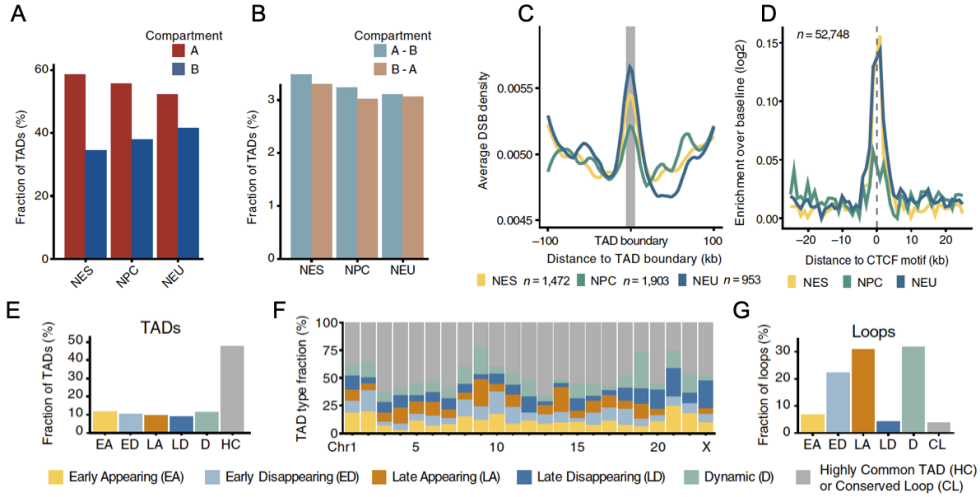

Figure 14. Endogenous DSBs are enriched at dynamic 3D genome sites. Fraction of TADs spanning genomic regions belonging to the same (A) or to a different (B) compartment type. (C) Metaprofile of DSB density around TAD boundaries. (D) Metaprofiles of DSB enrichment around CTCF factor binding motifs. (E) Fraction of TADs belonging to one of six categories: (1) Early Appearing; (2) Early Disappearing; (3) Late Appearing; (4) Late Disappearing ; (5) Dynamic; and (6) Highly Common, based on whether and when TADs disappear or appear during the differentiation of NES cells to NEU. (F) Same as in (E) but separately for each chromosome. (G) Same as in (E), but for chromatin loops. Note that the last category (grey) is now referred to as Conserved Loop (CL).

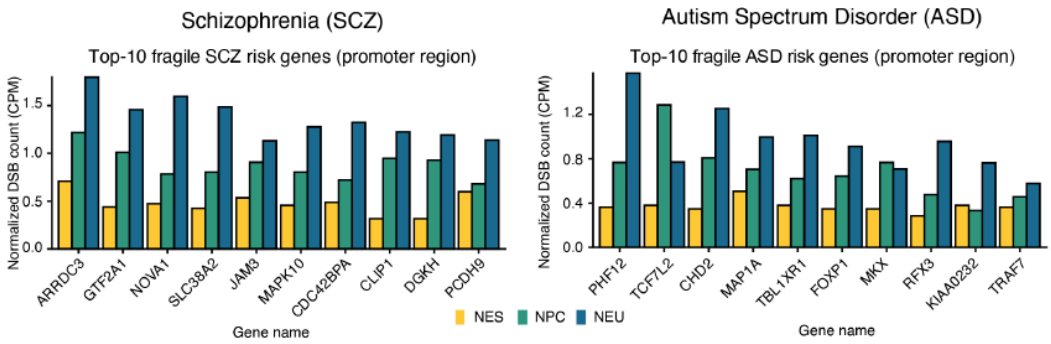

Figure 15. Top-fragile genes are associated with increased risk NDDs. Normalized DSB counts in the promoter region for the ten most fragile genes associated with SCZ and ASD risk in NES, NPC, and NEU cells. CPKM, DSB count per kilobase per million reads calculated as number of DSBs divided by number of reads times one million divided by gene width.

2.2.2 Paper II: Topoisomerase 1 activity during mitotic transcription favors the transition from mitosis to G1.

Background

Our knowledge of transcriptional regulatory mechanisms comes mainly from study of transcription factors and chromatin regulation, while the mechanical and topological properties of the DNA during transcription have been less investigated. Topoisomerases are nuclear enzymes that play essential roles in DNA replication, transcription, chromosome segregation, and recombination. All cells have two major forms of topoisomerases: type I, which makes single-stranded cuts in DNA, and type II enzymes, which cut and pass double-stranded DNA. TOP1 is known to be regulated through RNAPII pause release. However, as cells enter mitosis, chromatin needs to be adjusted for the segregation of chromosomes to occur correctly. This compaction in addition to high levels of expression by RNAPII can halt the process of transcriptional elongation and require TOP1 intervention to proceed. The role of TOP1 in removal of torsional stress and resume elongation is well characterized in interphase, but less so in increased compaction of chromatin during mitosis. The increased opposing supercoiling at this stage of the cell cycle could impair its progression and clearance with important consequences for cell integrity.

Motivation and methods

We investigated the role of TOP1 in relation to nascent transcription through SLAM-seq[201] in the context of replication stress across mitosis and as subject of interruption of interphase chromatin conformation. By monitoring chromatin occupancy of TOP1 and RNAPII genome-wide through ChIP-seq and along mitosis timeframe. In particular, we investigated early, mid and late mitosis by labeling the DNA with propidium iodide (PI) and anti-phospho-histone h3 (Ser10) antibody. We then used nocodazole-synchronized human colon carcinoma HCT116WT cells and their derived clone HCT116KI cells[102] which contain a mutation that genetically disrupts TOP1 exon4, responsible for interaction with the RNAPII Carboxy-Terminal (CTD)-interaction domain. Through time-lapse microscopy, immunofluorescence and several inhibiting drug treatments contemporary processes are measured and quantified. By then investigating transcription patterns and transcriptional spiking by RNAPII in HCT116 across time, we strived to elucidate the role of TOP1 specifically in mitosis.

Main findings

We first set out distinct phases of mitosis, establishing the details of our model system (**Figure 16**). We began by mapping RNAPII, H3K4me3 (as a measure of promoter accessibility) and TOP1 in actively elongating genes across early, mid and late mitosis. We demonstrate TOP1-RNAPII association both through sequencing and imaging, again for early, mid and late mitosis. We found that in early mitosis TOP1 clears RNAPII during transcriptional elongation, but that the localization was disrupted in HCT116KI. By labeling nascent transcripts using SLAM-seq, we noticed a higher variance in global transcriptome and transcriptional noise in HCT116KI (**Figure 19**). Disruption of peri-mitotic transcriptional factors and accumulation of important factors during a time-sensitive process as cell cycle start predicted to perturb progression to the next cell cycle phase. All in all, we can conclude that TOP1 acts as a transcription factor and promotes RNAPII transcription and clearance from chromosomes in prometaphase (**Figure 17**).

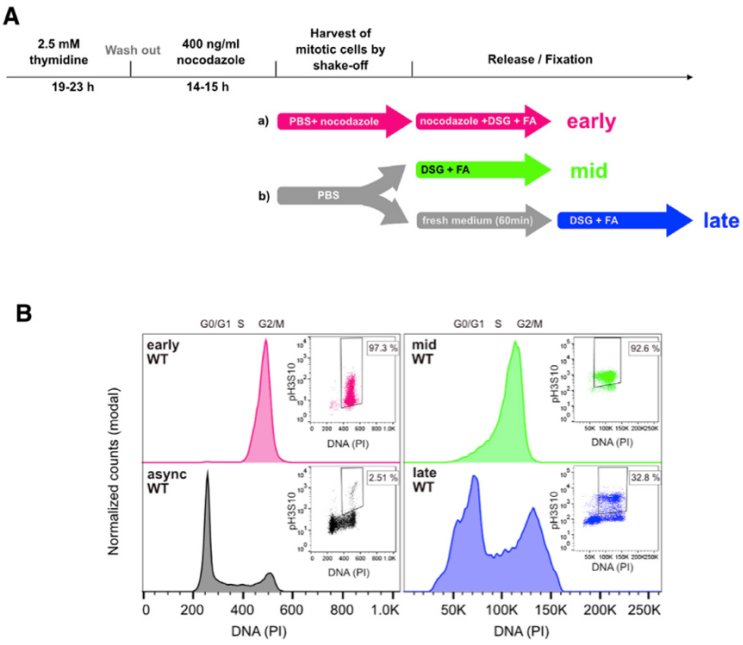

Figure 16. Workflow of HCT116 cell synchronization and categorization of stages of mitosis

(A) Schematic of thymidine and nocodazole treatments used to enrich HCT116 cells in prometaphase. Cells were either kept in prometaphase by collecting and fixing them in the presence of nocodazole (early mitosis) or collected in the absence of nocodazole and directly fixed (mid mitosis) or released for 1 h before fixation (late mitosis). (B) Flow cytometry analysis of asynchronous and mitotic cells after propidium iodide (PI) and anti-phospho-histone h3 (Ser10) antibody staining.

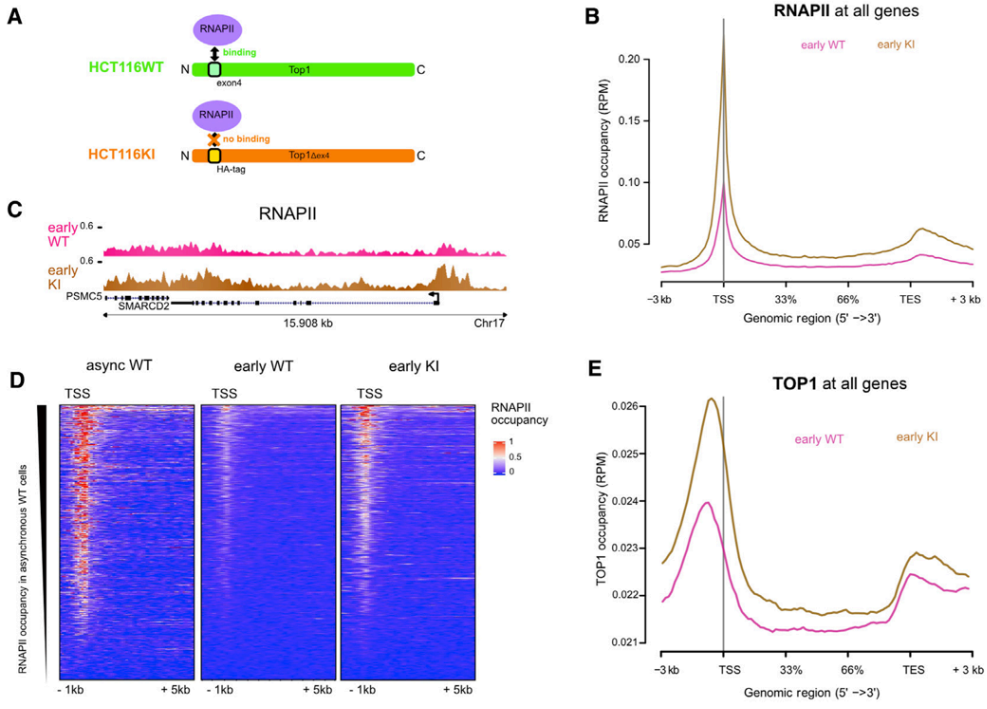

Figure 17. RNAPII-TOP1 interaction is necessary to ensure proper RNAPII clearance from chromosomes in early mitosis

(A) HCT116KI cells express a TOP1 mutant where exon4 is replaced by 3xHA tags. TOP1 mutant cannot interact with RNAPII. (B) Average RNAPII occupancy (RPM ± SEM) at all genes. Gray line indicates the TSS. (C) UCSC Genome Browser tracks of RNAPII at SMARCD2 locus. (D) Heatmaps of RNAPII density at protein coding genes ranked from highest to lowest RNAPII level in asynchronous HCT116WT cells. (E) Average TOP1 occupancy at all genes (RPM ± SEM).

Loss of the TOP1-RNAPII interaction causes supercoil buildup and segregation defects. By using time-lapse imaging, we confirmed a slowing cell cycle in HCT116KI (**Figure 18**). Specifically, we noticed delay in G1 progression and sought to exclude the possibility that the observed delay in cell cycle progression could be a consequence of DNA repair pathways triggered by segregation defects. Here, we expected a higher demand for TOP2 catenation activity to compensate for supercoiling accumulation in early mitosis resulting in ultrafine anaphase bridges which are DNA segments that fail to segregate and are tethered between the sister chromatids and result in mitosis stalling. We measured DNA damage markers gammaH2AX and 53BP1 using immunofluorescence labeling and did not detect substantial changes in gammaH2AX and 53BP1 foci when comparing HCT116WT and HCT116KI cells (**Figure 18**). Therefore, we concluded that the delay in mitotic exit is most likely due to the deregulation of transcription rather than to the activation of DNA repair.

We then set out to characterize the transcriptional changes as a consequence of TOP1-RNAPII disruption through drug treatments. As expected, this analysis revealed that HCT116KI were more resistant to TOP1 inhibitors given the inability of RNAPII to efficiently stimulate TOP1[102]. Notably, HCT116KI cells were more sensitive to inhibitors of pathways controlling G1 growth phase, in particular to drugs targeting members of the mTOR pathway[202]. This increased sensitivity was restricted to drugs regulating the G1 growth program, as inhibitors of S phase were equally effective in HCT116WT and HCT116KI cells.

We computed TOP1 and RNAPII profiles at the top 30% expressed genes on the basis of RNA-seq in asynchronous cells. Whereas the enrichment of TOP1 was low in early mitosis, TOP1 peaked around TSSs in mid mitosis and then declined in late mitosis, following the same profile as RNAPII. In late mitosis, TOP1 was abundant along gene bodies where RNAPII was elongating and accumulated at TSSs where RNAPII was paused, further highlighting that TOP1 binding mirrors RNAPII distribution. TOP1 localization on mitotic chromosomes was also confirmed by immunofluorescence microscopy. Thus, TOP1 is an integral part of the transcription complex and is associated with mitotic chromosomes

Finally, we set out to characterize the consequences of disrupting the TOP1-RNAPII interaction in HCT116WT cells through the auxin-degron system, removing TOP1 at the onset of mitosis. We found that similarly to the HCT116KI, transition to G1 was delayed. Disrupting the TOP1-RNAPII binding affects cell growth and sensitizes cells to mTOR drugs, confirming our hypotheses (**Figure 18**). Taken together, we present a novel regulatory role for TOP1.

Conclusions

TOP1 relieves torsional stress in actively transcribed DNA and facilitates the expression of long genes. In the context of replication, we found that during mitotic exit TOP1 importantly assists RNAPII promoter loading to restart transcription. By disrupting TOP1-RNAPII interaction, the cell cycle is delayed and mitotic transcription is affected. When mitotic transcription is poorly regulated, individual proliferating cells have a greater variance in transcriptional levels and RNAPII bursting is necessary for the hierarchical program of transcription re-activation and cell memory thus could lead to loss of cell identity.

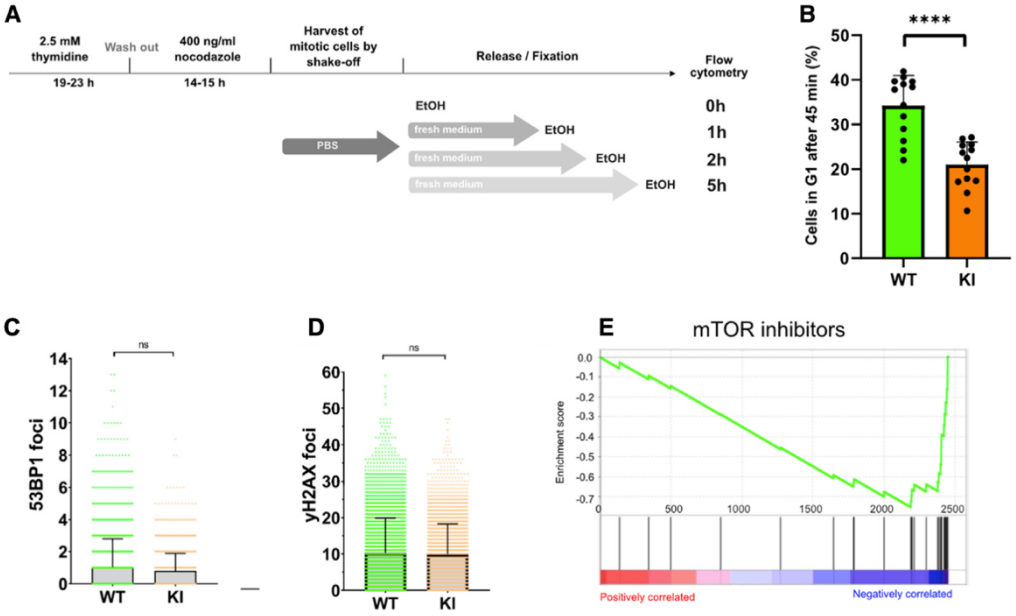

Figure 18. Interfering with RNAPII-TOP1 interaction leads to cell cycle delays (A) Cells were synchronized with nocodazole, released, and fixed with EtOH. (B) Cells in G1 phase 45 min after release comparing HCT116WT and HCT116KI). Significance was determined using paired t test. (C and D) Immunofluorescence microscopy of asynchronous HCT116WT and HCT116KI cells stained with DAPI and anti-53BP1 (C) or anti-gammaH2AX (D). Significance was calculated using Mann-Whitney test. (E) Drug set enrichment analysis of differential sensitivity between HCT116WT and HCT116KI cells toward drugs targeting mTOR.

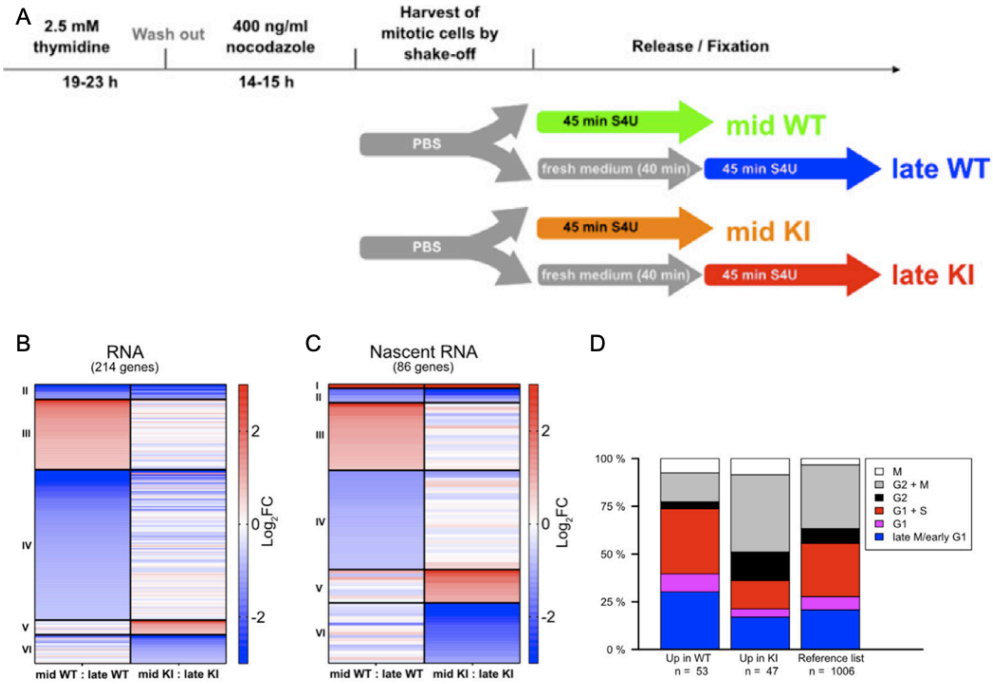

Figure 19. TOP1-RNAPII interaction is required for the coordinated transcriptional reactivation during mitosis (A) Schematic of labeling cells for SLAM-seq after release from nocodazole arrest. (B and C) Differential expression analysis of total (B) and nascent RNA(C), comparing expression levels in mid and late mitosis within theHCT116WT (left) and HCT116KI (right). Heatmaps display the log2 fold change (FC) and are horizontally clustered in groups of genes that are differentially upregulated in (I) both mid WT and mid KI, (II) both late WT and late KI, (III) mid WT (intermediate WT genes), (IV) late WT (late WT genes), (V) mid KI (intermediate KI genes), or (VI) late KI (late KI genes). (D) Categorization of RNAs differentially enriched in HCT116WT and HCT116KI into cell cycle phases.

3 DISCUSSIONS AND CONCLUSIONS

3.1 DISCUSSION OF FINDINGS

This thesis brings together (I) a correlative genomics study leveraging novel genome-wide methods on a complex 2D cell culture model of neurotypical cell type specification under unperturbed conditions describing endogenous DNA double-strand breaks (DSBs) genome-wide, and (II) a mechanistic study into the regulatory role of Topoisomerase 1 (TOP1) in transcription and proliferation. We report an sBLISS dataset to chart the first-ever atlas of endogenous DSBs forming as a result of endogenous processes at three sequential stages of human neural stem cell differentiation and reproducibly capture changes in the burden of endogenous DSBs. To investigate this, we provided high-quality datasets of total-RNA RNA-Seq, and Hi-C as proxy for the morphological and transcriptional changes that occur during neural cell lineage specification (Paper I). To investigate the relationship between transcription, topoisomerases, and cell cycle, we investigated the DNA-damage response in a replication-stress model system and contributed to the elucidation of the mechanical TOP1-RNAPII interaction across G1 progression (Paper II).

3.1.1 Analysis and evaluation of Paper I

The driving force behind the DSB origin remains unknown and requires in-depth mechanistical studies which go beyond the scope of constituent paper I. A summary of possible processes driving DSB formation is described in the introduction (Section 1.2). As described in the introduction (Section 1.2.7), we are not the first to provide genome-DSB maps in neural stem cells. Specifically, in mouse and human neural stem cells, endogenous DSBs have been detected in the context of genes associated with risk for neurodevelopmental disorders such as SCZ and ASD. Despite the importance of DSBs in the context of these diseases, an atlas of endogenous DSBs forming during human neurogenesis had not been charted before. Paper I is a step towards this goal and leverages on recently published sBLISS[160] method by charting endogenous DSBs at high resolution (10 kb) by applying sBLISS to a 2D cell culture model of human neurogenesis, specification and cell fate determination. We were surprised to find DSBs enriched at CpG islands nearby TSSs as typically the methylation taking place ought to be protective and a functional regulator of cell type specificity[142,203–206]. Indeed, CpG islands in proximity to several important NDD genes are known to be methylated as part of neural specification or transition from totipotent cells to the neural lineage[204,207–209]. Genes high in CpG islands enriched over others seems counterintuitive, but is a promising lead for further investigation[204,206,210,211]. In addition to this finding, we present several examples of total-RNA-seq and Hi-C's association to sBLISS tracks and show global trends in a cell-type context. While the findings we describe are not particularly novel, the quality of the data, model system and multi method approach across well-controlled cell fate acquisition process is unique and offers a valuable resource for further investigations into the role of genome fragility during neural cell fate determination.

The closest insights into the role of genome fragility during neural cell fate determination is the work by the Alt lab. Researchers from this group apply an ingenious alternative to sBLISS, HTGTS to assess genome fragility hotspots (See section 1.2.7). DSBs are identified by means of their ability to translocate to a fixed "bait", HTGTS has been used in both

mouse[143,144,146] and human neural stem cells[145] to identify RDC fragile sites. All of these publications are vital in shaping the context of genome fragility. To facilitate HTGTS DSB detection these studies leverage P53 and XRCC4 deficient stem cell lines. This is slightly problematic due to recent reports that P53 might play important roles in neural cell fate acquisition[165]. Interestingly, they found that over 2/3rd of the replication-stress susceptible DSB hotspot genes identified in neural precursor cells were orthologs of primary mouse neural stem/progenitor cell RDC-genes. Due to the difference in signal and coverage between the indirect HTGTS and direct sBLISS method we were not able to adequately compare our datasets. Moreover, in all the papers covered above the authors tend to choose totipotent iPSC/ESCs and pluripotent progenitors as timepoints. These chose timepoints are less than ideal because the differentiation step between these two cell roles is rather large and requires strong induction to differentiate to a neural lineage. Finally, the neural cells described in these papers never truly specialize into post-mitotic neural fate. We did find that the RDC-genes found in neural stem/progenitor cells typically ranked high on our top DSB-enriched gene lists, but also found many genes which are not described in these papers in each of our three neural time points. Whether this is an effect of the DSB-mapping strategy or genetic background remains unclear.

Indeed, much is still left to be uncovered. Regardless, increasing evidence suggests that endogenous DSBs associated with DNA transcription and 3D genome dynamics play an important role in physiological and pathological processes in the nervous system. While the controversial paper by Madabhushi et al., 2015 did kickstart an avalanche of interest in the topic of transcription activity-induced DSBs and their role in regulating neural ERG activation. Their work was extremely recently expanded with a thorough mechanistic investigation of activity-induced DSBs, regulated by TOP2 during a fear conditioning paradigm[212]. Once again work by these researchers is opening many doors for questions. In particular the spatial location of DSBs at the periphery of the nucleus, which we are not able to assess with our sequencing approach is likely to have a lot of merit and deserve in depth investigation. However, due to the choice of mapping and quantifying DSBs through gammaH2AX alone, in absence of an orthogonal direct-mapping method is a missed opportunity. The major limitation of this study is the fact that gammaH2AX was used as DSB measure and that the use of semi-quantitative methods is used to make very strong claims. Taken together it does take away some credibility of the proposed mechanism.

A rather novel report investigating the role of R-loops in DSB formation demonstrated that these structures do not ultimately drive replication stress-induced, recurrent DSB cluster formation[213]. This gives more credibility that another regulatory structure or protein might be important specifically for neural cell fate determination. Finally, a striking paper recently linked TOP1-induced replication stress results and its role in p53 driven stem cell fate decisions during human pluripotent stem cell-based neurogenesis[214]. Moreover, TOP1 was also found to induce DSBs and regeneration in the nervous system[215]. More evidence is necessary to assess the true mechanistic nature of de novo mutations and NDDs.

3.1.2 Analysis and evaluation of Paper 2

Transcription supercoils DNA to levels that can impede further progression of RNAPII unless it is removed by TOP1. Using ChIP-seq on mitotic cells, we found that TOP1 is required for RNAPII translocation along genes. While both TOP1 and TOP2 are mainly known for their canonical function of "relaxing" the DNA (Section 1.2.3), Baranello et al., 2016 showed that RNAPII drives TOP1's canonical activity to mediate transcriptional elongation[102]. However, Paper II actually presents a novel mechanistic function of TOP1 in promoting RNAPII transcription and clearance during prometaphase. We assess this mechanistic interaction in HCT116 colorectal cells both by genetically disrupting the CTD-interaction domain, or by using an auxin-degron system. We argue that this regulatory function of TOP1 in mitosis has important consequences for RNAPII chromatin occupancy, RNAPII's restart after mitosis and in the larger picture, maintaining cellular memory across subsequent generations. As such, we can deduce that TOP1 is crucial also for cell identity and function.

To better understand the dramatic changes in DNA structure and disruption of chromatin interactions and transcription during mitosis, it is important to distinguish different stages of the cell cycle and how they interrelate to transcription. Palozola, Lerner & Zaret 2019, laid out the importance of properly re-establishing transcription gene regulatory networks for diverse cells[216]. Indeed, we observe in paper II that interfering with the TOP1-RNAPII interaction has consequences for regulation of stochastic gene expression and introduces transcriptional noise which is important for specifying cell fates[217,218].

We propose that the activity of TOP1 during mitotic transcription is more important than during interphase to remove supercoiling that would otherwise oppose RNAPII elongation and clearance before the re-initiation of transcription in mid mitosis. Indeed, in the degron experiments we observe an accumulation of supercoiling at stressed regions of the genome (i.e., promoters or where transcription and replication collide). If sufficiently intense, supercoiling can also provoke alternative DNA structures[219]. The expected compensatory action of TOP2 could be insufficient to relieve the supercoiling. As a consequence, elevated supercoiling might increase TOP2 catenation activity, forcing the enzyme into aberrant cleavage complexes[220] and triggering segregation defects[221,222]. Although we did not investigate this directly, segregation defects at sites of cell-type specific transcriptional bursts has a high likelihood of causing larger indels or even CNAs (See section 1.1.1). Studying the specific disruptive effects observed in different cell types would be an intriguing avenue of bridging fundamental biological questions about molecular function and pathogenesis. Taken together, we contributed to the elucidation of the mechanical TOP1-RNAPII interaction across G1 progression in a model system ideally suited for replication-stress.

3.1.3 Caveats to the combined hypotheses

While these two studies are not directly linked, a connection between the description of neural genome dynamics and the mechanics driving transcriptional regulation can be argued when we consider the DNA nicks or DSBs resulting from topoisomerase activity. While we did not find global increases in DNA damage resulting from TOP1 depletion in Paper II, the immunofluorescence-based approach and analysis cannot exclude that physiologically relevant DSBs or CNVs accumulated over time. Indeed, another important caveat is that the studies are not part of a bigger collaborative effort and, as such, use different cellular models

and contexts to infer biological processes. Cross-model system experiments will always yield higher variability and as we observed in Paper I, cell identity can affect how the genome is structured and activated. We can infer that TOP1 occupancy would differ based on the cell-types' transcriptome. Likewise, it is important to consider that the balance of TOP1 and TOP2 is strongly regulated[220] and that both TOP1 and TOP2 are actively studied in the context of NDDs.

3.2 CONCLUSIONS AND FUTURE PERSPECTIVES

In conclusion, in this PhD study my colleagues and I contributed a non-exhaustive view of the DSB landscape and genome fragility in the context of endogenous neural cell specification and uncovered a small part of transcriptional regulation as it relates to replication stress. While these processes are related, the fragility landscape in neural cell specification generated in Paper I does not translate to the HCT116 colorectal cell line studied in Paper II, and thus the performed experiments would need to be repeated in the same model system in order to draw more generalized conclusions from the two studies performed during my PhD.

Building on what we have got

As summarized in the introduction chapter, most of our knowledge about NDDs comes down to changes in gene dosage and CNVs. It follows that DSBs are a key player in the fragility giving rise to *de novo* mutations and that replication stress and transcriptional regulation are mechanistically related. DNA damage has long been associated with the aging brain, but our findings here show that early neurodevelopmental time points may very well be critical windows during which very specific genomic loci and their associated functions may be affected. To better understand the details of genome fragility in development, there is still a lot that could be improved. Examples of how to further increase the adoption of these datasets rely on improvements in experimental setup, usage of technology or by setting up future projects digging at some outstanding questions.

Experimental setup:

The majority of state-of-the art mechanistical studies investigating the role of gene regulation, DNA repair or chromatin conformation have been conducted in non-brain cell lines or tissues. To really understand how these processes are regulated, using the appropriate model is a great feat. Regardless, the experimental setup can always be improved upon. While the work presented here would benefit greatly from expanded application to additional developmental time points to improve the temporal definition of critical periods of cell fate determination; using FACS to sort specific cell types based on surface markers to increase the purity of cell identities; or, including perturbation treatments such as etoposide, aphidicolin or camptothecins to perform mechanistic studies within the model system of interest, it is not always feasible to expand labor-intensive steps or high-cost experiments. That said, some of the major details that could be of added value is to also investigate multiple NES cell lines derived from different donors to reduce the experimental variability and correcting for possible confounding effects of genomic background. This would allow the DSB atlas to be more generally applicable than it currently is and significantly reduce confounding effects. Catching and correcting these confounding effects would allow for a

deeper biological insight and greater reliability of the findings and mechanistic follow-up studies.

Usage of Technology:
One of the exciting findings that could be elaborated on is the CpG-rich genes described earlier. While there is a large number of DSB-enriched CpG-rich TSS regions, there are some unexpected consequences that follow. Typically, methylated genes should not be expressed and therefore this finding goes against the activity-induced DSB hypothesis. While, counter-intuitive, this phenomenon deserves further investigation. With some adjustments the BLISS adapter and UMI could be leveraged by combining the sBLISS protocol with a long-read nanopore sequencing approach to gain insights of methylation status of individual CpGs.

One major issue with sBLISS is that it presents a snapshot of currently present loose ends *in situ*. By performing sBLISS in bulk, it becomes possible to catch a sufficient number of endogenous DSBs and map them on the genome, yielding a sufficiently complex library. However, the density of hits or the presence of a loose DSB end does not tell us anything about the temporal aspect of DSBs. Due to the presence of transcription machinery, DBSs localized at promoters could be shortly present whereas DSBs inside the gene-body or inter-gene regions could be present for a longer timeframe, skewing the DSB enrichment analysis. If different DSBs are repaired at different rates, for example due to proximate gene activity or DNA repair factories, those sites that are repaired quickly (i.e. activity-induced DSBs) will be underrepresented in our DSB atlas. On the other hand, DSBs which are not considered a threat may be present in the cell for longer periods of time and thus be overrepresented. As such, survivorship bias might be another feature that deserved to be tackled. - In World War II, damaged aircrafts were inspected for bullet holes and future aircrafts reinforced at those most-hit places. Abraham Wald at Columbia University proved that this was the wrong conclusion. He pointed out that those parts of the aircrafts which lacked bullet holes needed reinforcing, since aircrafts that were hit there never returned to base. – Likewise, DSBs which are so disastrous that cells cannot cope with them will hardly ever be detected. In cancer, individual cells with diminished function might not give cause for alarm. But in the brain, where individual cells last a lifetime and play integral roles in neural circuits, those few cells that might be exposed to deleterious DSBs or contain de novo structural variation might just deserve more attention.

Multi-omics & single cell technologies:
When the current projects were envisioned, the biotechnological progress could hardly have been accounted for. Genomic methods a being mixed and matched on a daily basis and the amount of data grows as our questions increase in complexity. Combinations of for example 3C and bisulphite sequencing are revealing complex pathways and disease mechanisms[130]. Multiple layers of proteomics, transcriptomics, epigenetics and genomics information connect genotype to phenotype and will provide researchers with novel scientific insights that cannot be found from single omic methods alone. All of the data presented in the two constituent papers, with exception of the immunofluorescence-based quantification of DNA damage in Paper II, relies on bulk data. Single cell applications are becoming increasingly common and with them, a whole world of new analyses is evolving. Single cell applications, allow us to sift through the variability I described earlier, by clustering cells together based

on their characteristics. While single cell sBLISS might be challenging to set up due to limits of sequencing depth, I see a future of more sophisticated mapping of biological complexity. Temporal encoding of single cells and lineage tracing experiments, through ERDU or something similar and SPLiT-seq labeling to track temporary breaks and lasting changes within cell populations. This hypothetical investigation is a glimpse of my take on what the future may hold.

Future project ideas

- Investigating multiple cell lines and including perturbations would have a great added value to stratfiying our observations. To further interpret our findings regarding CpG-coupled fragility, it might be possible to develop sBLISS further to include oxford nanopore sequecning or something similar, to recognise the methylation of CpGs. Incorporating accessability assesment like ATAC-seq would also further establish the DSB atlas validity by allowing accessability to be used as a benchmark for DSB maps.

- Investigating TOP1-TOP2 balance, specifically in the neural specification model. TOP1 and TOP2 are likely to play a synergistic role and need to be studied within the neural system to elucidate their role in development and disease. While it has proven difficult to investigate TOP2 by means of ChIP, technological advancements might provide better antibodies to pursue this intruiging question. Moreover, post-mitotic neurons are known to rely on different DNA repair mechanisms and have a different balance of topoisomerases to deal with the absence of cell cycle progression. DNA repair systems are regulated differently as a cell progresses through mitosis, differentiation and finally acquires post-mitotic cell fate.

- Investigating the location of DSBs and topoisomerases in 3D space. While Hi-C may offer an idea of chromatin compaction, conformation and inter/trans chromosomal contacts, it is not able to provide visual information about which part of the nucleus might be affected. Additional ortogonal methods like DamID and/or GPSeq would be able to provide a spatial vector for genomic locations. This is particularly interesting to assess in light of important processes taking place at the nuclear periphery alluded to in paper II.

- While we allude to the importance of DSB fragile sites and demonstrate endogenous occurance in cultured cells, investigating the lasting changes in the genome in line with the brain mosaisism network effort would be highly illuminating for understanding NDDs. As we get deeper insights in the connection between DSB hotspots, the transcriptome and chromatin organization, it will become increasingly relevant to perform single cell sequencing efforts to assess the real-world impact of DNA conformation and fragility. Performing targeted sequencing of promoters will be able to assess promotor mutagensis in time, while reduced representation sequencing might inform us if CNAs occur within those sites identified by sBLISS, HTGTS and others.

3.3 FINAL REMARKS

Biotechnology seems to have taken the example of microchips and is progressing and innovating at an exponential rate comparable to what we observed for computers Moore's law. Since NGS became available in the early 2000's and the first draft of the human genome being completed at 2003 with a major expansion with T2T consortium earlier this year filling in large parts of what was still uncharted. An unprecedented field of inquiry has opened up.

Once the genome became accessible, the amount of epigenetic variability and other layers of regulation by transcription factor networks have become apparent. The next generation is single-cell genomics which is allowing us to categorize this variability further and pick out biologically relevant processes within single cells.

Future studies will greatly benefit from the integration of multiple molecular-omic layers, ideally at the single-cell level, and together with advancing technologies such as long-read sequencing and live-cell imaging will likely be the key to decode the functional importance of genome organization for gene regulation in development and disease.

This is a field which is only now coming of age and is unlikely to directly affect health policy. The search for understanding however, will continue to drive us scientists to new levels.

4 ACKNOWLEGEMENTS

The past 5 years of working on my PhD project have been a rollercoaster ride with beautiful moments as well as challenging times. I am happy to have had more than my fair share of supportive, competent and inspiring people standing with me. However, I would particularly like to thank my great friend and wife Serena with whom I faced the hardships of wrapping up what I consider to be my major individual contribution to science to date. She has been an absolute pillar of honesty, support and patience during the closing of the curtains of my work in the Crosetto group, which continued during everything else that has happened in our lives no matter how pressing or important. It seems only fitting to thank you before my colleagues.

To begin, I would like express my deepest gratitude to my main supervisor **Nicola Crosetto** for inspiring me in early 2017, inviting me to present my previous work in lab and allowing me the possibility of joining his group as a PhD student. You have built the foundation of the BiCro-group and have grown out roots and a sturdy oak trunk. Your insights into leveraging new technologies in innovative and groundbreaking ways is something I deeply admire, but what is even more impressive is your ability to sense and put to rest preoccupations which I would inevitably accumulate with prolonged struggles in the lab. I am very grateful to have worked with you and I hope my contributions have provided you new perspectives and will leave a lasting impression not unlike like your mentorship has done for me.

Many thanks are owed also to my co-supervisors. **Federico Agostini**, our discussions both practical and philosophical about the datasets and ways of analyzing them were very enjoyable and you have taught me a lot about the ins and outs of science. **Anna Falk**, you have always been ready to answer my questions and through the iPS cell core provided practical cell culture support. **Magda Bienko**, your work-ethic is unparalleled and you are a great example as a scientist and a group leader. I will strive to take as much after you as I possibly can. **Marie Carlén**, thank you so much for meeting with me the very beginning of my visit for the job-interview and for your caring attitude throughout these 5 years. I wish we would have been able to discuss the mechanistic and neuroscientific implications of my research more at length, but am very grateful for the time you continued to make for me.

I also extend a great appreciation to my opponent **Dr. Pelin Sahlén**, the exam board comprised of **Prof. Neus Visa**, **Dr. Qi Dai** and **Dr. Anita Göndor** and the chairperson **Dr. Bennie Lemmens** in advance, for finding time to read my thesis and being part of my public defense. Of course, I am also thankful of my half time committee members: **Prof. Gonçalo Castello-Branco**, **Dr. Anita Göndor** and **Dr. Bennie Lemmens**, for the time and precious advice they offered me in back May 2020. While Gonçalo was kind to offer suggestions on our RNA-Seq experiments and an inspiration with his developing brain symposium series, I was saddened to not have him on my exam board. I do however want to extend a heartfelt thank you for his support.

I am grateful also to **Elias Arnér** for following my educational trajectory and **Victoria Balabanova** for sorting out all of my administrative needs and arranging the defense.
Within the Bienko-Crosetto lab

Being part of the BiCro lab at Karolinska Institutet has taught me a lot and I do not think I could have asked for a more stimulating environment to work towards my PhD. It was a privilege to see the first generation of PhD students grow up and take wing. There are many who I appreciate endlessly. **Michele Simonetti** and **Reza Mirzazadeh** were the first to take my hand and kickstart my productivity. Reza's sober attitude and mental fortitude set the tone as we started with cryo-sectioning tissue blocks and set up increasingly large BLISS experiments. Michele's organizational skills and search for perfection is something I have greatly missed since his graduation in late 2021. **Britta Bouwman** has been a great mentor and teacher, both for writing and at the bench. **Gabriele Girelli,** while our projects together were cut short, I deeply respect your way of establishing systems and being friendly and helpful to a fault, regardless of how heavily charged with responsibility you were. Thank you for showing me an example to aspire to and for maintaining your welcoming attitude always. **Ana Mota,** you deserve a major thanks for your straightforwardness and much needed moral support in the lab. And then there's **Luuk Harbers**, thank you for your ability to relativize and remind me of my Dutch origins in a sea of Italians and Polish druktemakers!

To produce scientific data can hardly ever be done alone and people who deserve a special mention are the following for great discussions, collaborations and help in completing our team efforts: **Ning Zhang, Dora Grassini, Constantin Diekmann, Erik Wernersson, Tomesz Kallas, Merula Stout, Katarina Gradin,** and **Wenjing Kang** for our work on the BLISS project, the 3D genome project and the single cell CUTseq project. At the bench and particularly at the microscope, I learned a great deal from **Eleni Gelali** and **Quim Custodio** who had a level of talent and experience most researchers would dream of.

Learning is often coupled to teaching and I have had the honor to supervise, collaborate and exchange lessons with many visiting students in the lab. **Francesco Rossi**, we started together on the same day in the BiCro lab and I could not have asked for a better welcome to Sweden! **Jesko Wagner,** we worked together on the early BLISS datasets and you taught me a lot about the bioinformatic approach. Perhaps our teamwork was among one of my favorite experiences of the early PhD. **Spark Yaddav**, you reminded me that science should be fun at a time where I had lost sight of that. Our pilot experiments on live imaging the 3D genome late into the night gave me back some childish energy which was much needed at the time where results where thin on the ground. **Radosław Grochowski, Gustaw Eriksson, Gesa Voigt**, you were all indispensable for performing several heavy experiments which unfortunately were not fully utilized further. Each of you taught me a lot, in addition to the pleasures of working with you all. **Gabriela Stumberger** and **Yudish Varma**, thank you for your smiles and heartwarming conversations. Recognizing the start of one's journey becomes so much more beautiful once you are coming to the end of your own.

The BiCro-lab has too many amazing people to thank, both in past and present, however I want to list all of you who come to mind for creating a great environment, sharing interesting conversations and for expanding scientific horizons on a daily basis: **Emma Inns, Su Wang, Xiaoze Li-Wang, Carla Rossell, Masahiro Matsumoto, Cynthia Perez Estrada, Quentin Verron, Szymon Berezicki, Xinge Li, Marcin Nicoś, Silvano Garnerone, Maud Schweitzer, Katalin Ferenc, Xiaolu Zhang, Sólrún Kolbeinsdóttir, Gertjan Laenen, Eleonora Oliani, Linxuan Zhao, Eléna Coulier,** and the bright future of the new BiCro

PhD all-star team: **Kaja Harton, Jinxin Chen** and **Wing Hin Yip.** I expect great things and wish you all the best**!**

Collaborators, colleagues and friends
Together we stand stronger and I definitely am able to stand taller thanks to the research collaborations, coffee meetings and friends motivating me to push my boundaries.

Annika Wiegard, Donald Cameron, Marta Trullols, Laura Baranello and the rest of the people involved with the topoisomerase projects. I am super grateful for having been involved and engaged with your team, digging into the mechanics of what happens to TOP1 during mitosis. It was refreshing to see the creativity, curiosity and can-do attitude in your group. Annika, your great optimism and drive have been a beacon of light. Laura, you have mentored me through much of the PhD and you deserve all the more appreciation for always having your door open for me. Thank you for the many conversations about research and life.

Francesca Mastropasqua and **Kristiina Tammimies**, thank you for helping me apply the biotechnology to a biologically-relevant disease model. Working together with you has been a pleasure and I am sure the beautiful manuscript will soon be accepted in its proper place.

Laura Curti, Ramiro Vazquez, Ottavio Croci and **Stefano Campaner,** thank you for providing us with cancer samples and for your rigorous discussions on tackling the sBLISS mapping and analysis. Contributing to your efforts has been a pleasure and a privilege.

Giuseppe Testa, Alejandro López Tobón, thank you for your discussions on life and neuroscience. Thank you for your playful attitude and the efforts on investigating gene-dosage in spheroids. It was a very welcome diversion from the standard 2D cultures and I had a lot of fun assembling, testing the smFISH targets in the beautiful organoid structures.

I deeply appreciate **Ana Marin Navarro, Robin Pronk, Carlos Villaescusa, Malin Kele, Moshen Moslem** and co-supervisor **Anna Falk** for providing the cell lines and all the suggestions in setting up and validating the neurodifferentiation protocols, including moral support for when the differentiations would inevitably stop working for a few weeks.

Alessandro Bonetti, and co-supervisors **Gonçalo Castelo-Branco & Federico Agostini**, thank you for involving me in the FANTOM6-RADICL-seq project. Working with you has allowed me to reflect and test experimentally several DNA:RNA interactions and has given me a deeper appreciation for the depth of complexity behind the dynamic 3D genome.

A massive rocking thanks to **Ola Hermanson** for his support around the half-time seminar and readiness to discuss the scientific merit of my work. Together with **Fredrik Lanner** you made my exchange to Toronto positively unforgettable. Thank you for allowing me to return the hospitality for our Canadian guests some years later with the help of the ebullient **Laura Baque Vidal**. From the University of Toronto: **Stephen Matthews, Stephen Lye & Brian Cox**, thank you for setting up a very inspiring curriculum and for discussing alternate ways of investigating genome structure and fragility and for helping me perform analyses and pushing me as a bioinformatician and a scientist.

I have much appreciation for **Daniella Tropea** and **Karl Thomas** from Trinity College Dublin, for allowing me to present and discuss my science in interdisciplinary context. My stay in Ireland taught me about innovation and development of ideas and helped me to later organize team building activities back in the BiCro-lab. The summer school you organized provided me with lifelong friendships and priceless memories.

Massive thanks to **Feng Zhang** and **Winston Yan** from the broad institute / MIT, for allowing me access to the CRISPR-off target experiments used in the methods in molecular biology book-chapter. I was deeply familiar with your papers before enrolling at KI and it was a huge surprise to meet and discuss together with great minds such as yourselves.

As I approached the end of my PhD, nearing the end of the COVID-19 pandemic, I had the pleasure of presenting my work internationally at the Deciphering Stem Cell Fate by Single Cell, Multiomic & Inference Approaches conference 2022 in Turin. Thank you to **Valentina Proserpio, Salvatore Oliviero, Davide Cittaro, Alexander van Oudenaarden** and all others who provided valuable input during the revision of Paper I, and for the countless intriguing discussions during the conference. Each and every one of you left a deep impression on me and I learned a massive amount from all of you that I will take with me in my future endeavors!

There are many great scientists with whom I have talked, discussed. All of them have helped me grow into the person I am today. For that, I am very grateful to all of you in Biomedicum/Bioclinicum, SciLifeLab and in the old KI campus: **Lars-Gunnar Larsson, Simon Elsasser, Hans Blom, Illaria Testa, Stefania Giacomello, Vicent Pelechano, Adnane Achour, Marc Friedländer, Oscar-Fernandez-Capetillo, Paul Hudson, Olli Kallioniemi, Mathias Uhlén**. Thank you all for providing the beautiful backdrop to science.

Other colleagues who helped me get started or supported me during my time in SciLifeLabs: **Carolyn Marks, Dimitris Kanellis, Jaime Espinoza Ruiz, Sharath Tippur Narayana Iyengar, Melania Zauri, Mirela Balan, Valeria Vanda, Rozina Caridha, Ann-Sofie Nilsson, Anna-maria Katsori, Dörte Schlesinger, Jing Lyu, Banushree Kumar, Flor Pineiro, Eva Brinkman, Alisa Alekseenko, Kyle Kimler, Lea van Husen, Markus Janasch, Marcel Tarbier, Johannes Heimgärtner, Anna-Sophia,** Thank you.

Past inspirations without whom I would not have continued down the path of science: **Adri Thomas** and **Wim Dictus**, you thought me early on about the cell, about transcription and the genome; about studying, performing experiments and about life. Thank you so much for all the selfless effort you have put into what you both do.

Thank you **Joost Verhaagen**, **Ruben Eggers**, **Fred de Winter** & **Adriaan Klop** for serving my first taste of real academic science. Studying peripheral neural regeneration at the Neuroscience Institute Netherlands has given me a unique perspective on how optogenetics can be used for real-world applications. But also, to always keep in mind to apply everything.

Those who taught me about iPS Cells and neurodifferentiation and the complexity that is psychiatric aetiology and the biology of mental illness helped shape me in incredible ways. **Shashini Munshi, Jeffrey Stedenhouder, Edwin Mientjes, Bas Lendemeijer, Geeske van Woerden, Mark van der Kroeg, Denise Slump, Femke de Vrij, Stefano Rivetti** and **Steven Kushner**, thank you for making my time in the molecular neuropsychiatry department amazing and lifechanging. The Angelman Syndrome expertise center ENCORE allowed me to learn from the best and showed me the extends of scientific excellence.

A massive thanks to **Richard**, **Agneta** and **Julia Cohen** for warmly welcoming me and helping me settle in when I first arrived in Stockholm. Moving to a new country is always tough, but just like cells during embryonic development, we cannot change if we do not move. Thank you for helping me take this step and helping me assimilate with my Swedish.

I cannot end the acknowledgements and leave my dearest friends and family unmentioned. In SciLifeLab there was one person who was able to support me even when I was unsupportable. Thank you, **Yerma Pareja Sanchez,** for the practical advice, courageous pushes forward and hugs when I most needed them. **Franziska Bonath**, your golden heart, welcoming nature and stern scientific support were truly invaluable. **Francesco Righetti**, you made the weekends more exciting with the occasional hike, board game or climbing invitation, the few times that I would actually not be stuck in the lab. You are awesome!

Some other friends for good measure, **Machteld van Rooijen, Valentina Lorenzi, Isaac Klop, Amir Homavar, Wayne Plaxton, Julien van Kralingen, Viola Ong, Adam Kamínski, Ryan Spencer, Markella Moraki,** thank you all for being by my side always. Helping me to spar on ideas, proofread my drafts or making me smile, always. Thank you to all of my friends, also those who I might have forgot to mention.

Family

To my father **Antonino Ballarino**, thank you for being my hero. You are the example who taught me to follow my passion no matter where it will take me. The more we change, the more we stay the same and in the 5 years past I feel like I become more and more like you.

To my mother **Minke Bonthuis**, thank you for stimulating my curiosity and critical attitude always. I feel blessed for the love and trust you have instilled in me. It is in large part in your honor that I made it here pursuing growth and righteousness.

To my sister **Marcella Ballarino**, thank you for being a source of loving encouragement and at times, my personal psychologist and HR-manager. I was happy and proud to see you graduate and am sure you will return the favor.

Peter Eisinga, thank you for our discussions through the years and for your feedback on the Dutch popular science summary which in my opinion has made it way more pleasurable to read. Thank you, **Anja Kellner,** for your care and kindness. Thank you to all the invisible supporters who are not mentioned but appreciated regardless. …and thank you to the rest of my family for all their love.

And finally, again, to my wife **Serena**, without whom I would have not been able to complete this journey.

Thank you.

5 BIBLIOGRAPHY

1. D'Gama, A. M. & Walsh, C. A. Somatic mosaicism and neurodevelopmental disease. *Nat. Neurosci.* **21**, 1504–1514 (2018).
2. Sullivan, P. F. & Geschwind, D. H. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162–183 (2019).
3. Akbarian, S. *et al.* The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).
4. Owen, M. J. & O'Donovan, M. C. Schizophrenia and the neurodevelopmental continuum:evidence from genomics. *World Psychiatry* **16**, 227–235 (2017).
5. Prata, J., Santos, S. G., Almeida, M. I., Coelho, R. & Barbosa, M. A. Bridging Autism Spectrum Disorders and Schizophrenia through inflammation and biomarkers - pre-clinical and clinical investigations. *J. Neuroinflammation* **14**, 1–33 (2017).
6. van der Voet, M., Nijhof, B., Oortveld, M. A. W. & Schenck, A. Drosophila models of early onset cognitive disorders and their clinical applications. *Neurosci. Biobehav. Rev.* **46**, 326–342 (2014).
7. Farrell, M. S. *et al.* Evaluating historical candidate genes for schizophrenia. *Mol. Psychiatry* **20**, 555–562 (2015).
8. Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509–516 (2022).
9. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery. *bioRxiv* 2022.05.07.491045 (2022).
10. Frere, S. & Slutsky, I. Alzheimer's Disease: From Firing Instability to Homeostasis Network Collapse. *Neuron* **97**, 32–58 (2018).
11. Styr, B. & Slutsky, I. Imbalance between firing homeostasis and synaptic plasticity drives early-phase Alzheimer's disease. *Nat. Neurosci.* **21**, 463–473 (2018).
12. Shohat, S., Amelan, A. & Shifman, S. Convergence and Divergence in the Genetics of Psychiatric Disorders From Pathways to Developmental Stages. *Biol. Psychiatry* **89**, 32–40 (2021).
13. Ben-David, U. & Amon, A. Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.* **21**, 44–62 (2020).
14. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97 (2006).
15. Ben-Yosef, D. *et al.* Genomic Analysis of hESC Pedigrees Identifies De Novo Mutations and Enables Determination of the Timing and Origin of Mutational Events. *Cell Rep.* **4**, 1288–1302 (2013).
16. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
17. Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).
18. Trost, B. *et al.* Impact of DNA source on genetic variant detection from human whole-genome sequencing data. *J. Med. Genet.* **56**, 809–817 (2019).
19. Eberwine, J., Sul, J. Y., Bartfai, T. & Kim, J. The promise of single-cell sequencing. *Nat. Methods* **11**, 25–27 (2014).
20. McConnell, M. J. *et al.* Mosaic Copy Number Variation in Human Neurons. *Science* **342**, 632–637 (2013).
21. Cai, X. *et al.* Single-Cell, Genome-wide Sequencing Identifies Clonal Somatic Copy-Number Variation in the Human Brain. *Cell Rep.* **8**, 1280–1289 (2014).
22. Villela, D. *et al.* Increased DNA Copy Number Variation Mosaicism in Elderly

Human Brain. *Neural Plast.* **2018**, 1–9 (2018).
23. Lodato, M. A. *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).
24. Huguet, G., Benabou, M. & Bourgeron, T. The Genetics of Autism Spectrum Disorders. in 101–129 (2016).
25. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
26. Rice, A. M. & McLysaght, A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nat. Commun.* **8**, 1–11 (2017).
27. Zoghbi, H. Y. & Warren, S. T. Neurogenetics: Advancing the "Next-Generation" of brain research. *Neuron* **68**, 165–173 (2010).
28. Kim, J. C. & Mirkin, S. M. The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* **23**, 280–288 (2013).
29. Swami, M. *et al.* Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* **18**, 3039–3047 (2009).
30. Shelbourne, P. F. *et al.* Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum. Mol. Genet.* **16**, 1133–1142 (2007).
31. Massey, T. H. & Jones, L. The central role of DNA damage and repair in CAG repeat diseases. *DMM Dis. Model. Mech.* **11**, (2018).
32. Coe, B. P. *et al.* Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat. Genet.* **51**, 106–116 (2019).
33. Sullivan, P. F. & Owen, M. J. Increasing the Clinical Psychiatric Knowledge Base About Pathogenic Copy Number Variation. *Am. J. Psychiatry* **177**, 204–209 (2020).
34. Redin, C. *et al.* The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat. Genet.* **49**, 36–45 (2017).
35. Sanchis-Juan, A. *et al.* Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Med.* **10**, 1–10 (2018).
36. Adamo, A. *et al.* 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat. Genet.* **47**, 132–141 (2015).
37. Schwer, B. *et al.* Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc. Natl. Acad. Sci.* **113**, 2258–2263 (2016).
38. Pollack, J. R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23**, 41–46 (1999).
39. Ramocki, M. B. *et al.* Recurrent distal 7q11.23 deletion including HIP1 and YWHAG identified in patients with intellectual disabilities, epilepsy, and neurobehavioral problems. *Am. J. Hum. Genet.* **87**, 857–865 (2010).
40. Kalsner, L. & Chamberlain, S. J. Prader-Willi, Angelman, and 15q11-q13 Duplication Syndromes. *Pediatr. Clin. North Am.* **62**, 587–606 (2015).
41. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
42. Halvorsen, M. *et al.* Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. *Nat. Commun.* **11**, (2020).
43. Knouse, K. A., Wu, J. & Amon, A. Assessment of megabase-scale somatic copy number variation using single-cell sequencing. *Genome Res.* **26**, 376–384 (2016).
44. Ramocki, M. B. & Zoghbi, H. Y. Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature* **455**, 912–918 (2008).

45. Zarrei, M. *et al.* A large data resource of genomic copy number variation across neurodevelopmental disorders. *npj Genomic Med.* **4**, (2019).
46. Torres, F., Barbosa, M. & Maciel, P. Recurrent copy number variations as risk factors for neurodevelopmental disorders: Critical overview and analysis of clinical implications. *J. Med. Genet.* **53**, 73–90 (2015).
47. Cook, E. H. & Scherer, S. W. Copy-number variations associated with neuropsychiatric conditions. *Nature* **455**, 919–923 (2008).
48. McConnell, M. J. *et al.* Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science* **356**, (2017).
49. Lodato, M. A. *et al.* Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).
50. Bedrosian, T. A., Quayle, C., Novaresi, N. & Gage, F. H. Early life experience drives structural variation of neural genomes in mice. *Science* **359**, 1395–1399 (2018).
51. Lodato, M. A. & Walsh, C. A. Genome aging: somatic mutation in the brain links age-related decline with disease and nominates pathogenic mechanisms. *Hum. Mol. Genet.* **28**, R197–R206 (2019).
52. Paquola, A. C. M., Erwin, J. A. & Gage, F. H. Insights into the role of somatic mosaicism in the brain. *Curr. Opin. Syst. Biol.* **1**, 90–94 (2017).
53. Verheijen, B. M., Vermulst, M. & van Leeuwen, F. W. Somatic mutations in neurons during aging and neurodegeneration. *Acta Neuropathol.* **135**, 811–826 (2018).
54. Charles, N. A., Holland, E. C., Gilbertson, R., Glass, R. & Kettenmann, H. The brain tumor microenvironment. *Glia* **59**, 1169–1180 (2011).
55. Cuddapah, V. A., Robel, S., Watkins, S. & Sontheimer, H. A neurocentric perspective on glioma invasion. *Nat. Rev. Neurosci.* **15**, 455–465 (2014).
56. Lutz, H., Nguyen, T. A., Joswig, J., Rau, K. & Laube, B. NMDA receptor signaling mediates cfos expression via Top2β-induced DSBS in glioblastoma cells. *Cancers (Basel).* **11**, 306 (2019).
57. Arlt, M. F., Ozdemir, A. C., Birkeland, S. R., Wilson, T. E. & Glover, T. W. Hydroxyurea induces de novo copy number variants in human cells. *Proc. Natl. Acad. Sci.* **108**, 17360–17365 (2011).
58. Lee, J. A., Carvalho, C. M. B. & Lupski, J. R. A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders. *Cell* **131**, 1235–1247 (2007).
59. Arlt, M. F. *et al.* Replication Stress Induces Genome-wide Copy Number Changes in Human Cells that Resemble Polymorphic and Pathogenic Variants. *Am. J. Hum. Genet.* **84**, 339–350 (2009).
60. Zachariadis, V., Cheng, H., Andrews, N. & Enge, M. A Highly Scalable Method for Joint Whole-Genome Sequencing and Gene-Expression Profiling of Single Cells. *Mol. Cell* **80**, 541-553.e5 (2020).
61. 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).
62. Arlt, M. F. *et al.* Effects of hydroxyurea on CNV induction in the mouse germline. *Environ. Mol. Mutagen.* **59**, 698–714 (2018).
63. Shaikh, N. *et al.* Replication Stress Generates Multiple Distinct Classes of Copy Number Alterations. *bioRxiv Mol. Biol.* (2019).
64. Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
65. Conrad, D. F. *et al.* Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nat. Genet.* **42**, 385–391 (2010).

66. Vilenchik, M. M. & Knudson, A. G. Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12871–12876 (2003).
67. Ceccaldi, R., Rondinelli, B., D'Andrea, A. D. & D'Andrea, A. D. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol.* **26**, 52–64 (2016).
68. Nouspikel, T. DNA repair in differentiated cells: Some new answers to old questions. *Neuroscience* **145**, 1213–1221 (2007).
69. Nouspikel, T. & Hanawalt, P. C. Terminally Differentiated Human Neurons Repair Transcribed Genes but Display Attenuated Global DNA Repair and Modulation of Repair Gene Expression. *Mol. Cell. Biol.* **20**, 1562–1570 (2000).
70. Li, X. & Heyer, W.-D. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res.* **18**, 99–113 (2008).
71. Nouspikel, T. Genetic instability in human embryonic stem cells: Prospects and caveats. *Futur. Oncol.* **9**, 867–877 (2013).
72. Lieber, M. R. NHEJ and its backup pathways in chromosomal translocations. *Nat. Struct. Mol. Biol.* **17**, 393–395 (2010).
73. Chakraborty, A. *et al.* Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes. *Nat. Commun.* **7**, 1–12 (2016).
74. Sfeir, A. & Symington, L. S. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem. Sci.* **40**, 701–714 (2015).
75. Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nat. Rev. Genet.* **10**, 551–564 (2009).
76. Fitzgerald, D. M., Hastings, P. J. & Rosenberg, S. M. Stress-Induced Mutagenesis: Implications in Cancer and Drug Resistance. *Annu. Rev. Cancer Biol.* **1**, 119–140 (2017).
77. Rohrback, S. *et al.* Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing. *Proc. Natl. Acad. Sci.* **115**, 10804–10809 (2018).
78. Carruthers, R. D. *et al.* Replication stress drives constitutive activation of the DNA damage response and radioresistance in glioblastoma stem-like cells. *Cancer Res.* **78**, 5060–5071 (2018).
79. Malkova, A. & Ira, G. Break-induced replication: functions and molecular mechanism. *Curr. Opin. Genet. Dev.* **23**, 271–9 (2013).
80. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **20**, 698–714 (2019).
81. Zhang, B. N., Bueno Venegas, A., Hickson, I. D. & Chu, W. K. DNA replication stress and its impact on chromosome segregation and tumorigenesis. *Semin. Cancer Biol.* **55**, 61–69 (2019).
82. Brandsma, I. & Gent, D. C. Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integr.* **3**, 9 (2012).
83. Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* **5**, (2009).
84. Bonner, W. M. *et al.* γH2AX and cancer. *Nat. Rev. Cancer* **8**, 957–967 (2008).
85. Branzei, D. & Foiani, M. Maintaining genome stability at the replication fork. *Nat. Rev. Mol. Cell Biol.* **11**, 208–219 (2010).
86. Alexander, J. L. & Orr-Weaver, T. L. Replication fork instability and the consequences of fork collisions from rereplication. *Genes Dev.* **30**, 2241–2252

(2016).
87. Bhat, K. P. & Cortez, D. RPA and RAD51: Fork reversal, fork protection, and genome stability. *Nat. Struct. Mol. Biol.* **25**, 446–453 (2018).
88. Cortez, D. Preventing replication fork collapse to maintain genome integrity. *DNA Repair (Amst).* **32**, 149–157 (2016).
89. Labib, K. & Hodgson, B. Replication fork barriers: Pausing for a break or stalling for time? *EMBO Rep.* **8**, 346–353 (2007).
90. Lemmens, B., Van Schendel, R. & Tijsterman, M. Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat. Commun.* **6**, 1–8 (2015).
91. 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. *J. Radiat. Res.* **59**, 261–271 (2018).
92. Dianov, G. L., O'Neill, P. & Goodhead, D. T. Securing genome stability by orchestrating DNA repair: Removal of radiation-induced clustered lesions in DNA. *BioEssays* **23**, 745–749 (2001).
93. Shibata, A. & Jeggo, P. A historical reflection on our understanding of radiation-induced DNA double strand break repair in somatic mammalian cells; interfacing the past with the present. *Int. J. Radiat. Biol.* **95**, 945–956 (2019).
94. Santini, E. *et al.* Mitochondrial superoxide contributes to hippocampal synaptic dysfunction and memory deficits in Angelman syndrome model mice. *J. Neurosci.* **35**, 16213–16220 (2015).
95. Qiu, S., Jiang, G., Cao, L. & Huang, J. Replication Fork Reversal and Protection. *Front. Cell Dev. Biol.* **9**, 1–8 (2021).
96. Belotserkovskii, B. P. *et al.* Transcription blockage by homopurine DNA sequences: Role of sequence composition and single-strand breaks. *Nucleic Acids Res.* **41**, 1817–1828 (2013).
97. Cohen, S. *et al.* Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nat. Commun.* **9**, (2018).
98. Helmrich, A., Ballarino, M. & Tora, L. Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. *Mol. Cell* **44**, 966–977 (2011).
99. Kouzine, F. *et al.* Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* **20**, 396–403 (2013).
100. Baranello, L., Levens, D., Gupta, A. & Kouzine, F. The importance of being supercoiled: How DNA mechanics regulate dynamic processes. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**, 632–638 (2012).
101. Pommier, Y., Sun, Y., Huang, S. Y. N. & Nitiss, J. L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* **17**, 703–721 (2016).
102. Baranello, L. *et al.* RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **165**, 357–371 (2016).
103. Wiegard, A. *et al.* Topoisomerase 1 activity during mitotic transcription favors the transition from mitosis to G1. *Mol. Cell* 1–18 (2021).
104. Morimoto, S. *et al.* Type II DNA topoisomerases cause spontaneous double-strand breaks in genomic DNA. *Genes (Basel).* **10**, 1–18 (2019).
105. Katyal, S. *et al.* Aberrant topoisomerase-1 DNA lesions are pathogenic in neurodegenerative genome instability syndromes. *Nat. Neurosci.* **17**, 813–821 (2014).
106. Pendleton, M., Lindsey, R. H., Felix, C. A., Grimwade, D. & Osheroff, N. Topoisomerase II and leukemia. *Ann. N. Y. Acad. Sci.* **1310**, 98–110 (2014).

107. Sciascia, N. *et al.* Suppressing proteasome mediated processing of Topoisomerase II DNA-protein complexes preserves genome integrity. *Elife* **9**, (2020).
108. West, A. E. & Greenberg, M. E. Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb. Perspect. Biol.* **3**, 1–21 (2011).
109. Yap, E. L. & Greenberg, M. E. Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. *Neuron* **100**, 330–348 (2018).
110. Suberbielle, E. *et al.* Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-β. *Nat. Neurosci.* **16**, 613–621 (2013).
111. Madabhushi, R. *et al.* Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* **161**, 1592–1605 (2015).
112. Göndör, A. & Ohlsson, R. Enhancer functions in three dimensions: Beyond the flat world perspective [version 1; referees: 3 approved]. *F1000Research* **7**, 1–12 (2018).
113. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226 (2008).
114. Bahar Halpern, K. *et al.* Bursty gene expression in the intact mammalian liver. *Mol. Cell* **58**, 147–156 (2015).
115. Larsson, A. J. M. *et al.* Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 (2019).
116. Dellino, G. I. *et al.* Release of paused RNA polymerase II at specific loci favors DNA double-strand-break formation and promotes cancer translocations. *Nat. Genet.* **51**, 1011–1023 (2019).
117. Ebert, D. H. & Greenberg, M. E. Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* **493**, 327–337 (2013).
118. Nestler, E. J., Peña, C. J., Kundakovic, M., Mitchell, A. & Akbarian, S. Epigenetic Basis of Mental Illness. *Neurosci.* **22**, 447–463 (2016).
119. Ballarino, R. *et al.* An atlas of endogenous DNA double-strand breaks arising during human neural cell fate determination. *Sci. Data* **9**, 400 (2022).
120. Gothe, H. J. *et al.* Spatial Chromosome Folding and Active Transcription Drive DNA Fragility and Formation of Oncogenic MLL Translocations. *Mol. Cell* **75**, 267-283.e12 (2019).
121. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
122. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
123. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
124. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
125. Uusküla-Reimand, L. *et al.* Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol.* **17**, 1–22 (2016).
126. Canela, A. *et al.* Genome Organization Drives Chromosome Fragility. *Cell* **170**, 507-521.e18 (2017).
127. Sati, S. *et al.* 4D Genome Rewiring during Oncogene-Induced and Replicative Senescence. *Mol. Cell* 1–17 (2020).
128. Krijger, P. H. L. *et al.* Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming. *Cell Stem Cell* **18**, 597–610 (2016).
129. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).
130. Lee, D.-S. *et al.* Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019).
131. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell

resolution. *Nature* **547**, 61–67 (2017).
132. Ochs, F. *et al.* Stabilization of chromatin topology safeguards genome integrity. *Nature* **574**, 571–574 (2019).
133. Zada, D., Bronshtein, I., Lerer-Goldshtein, T., Garini, Y. & Appelbaum, L. Sleep increases chromosome dynamics to enable reduction of accumulating DNA damage in single neurons. *Nat. Commun.* **10**, 895 (2019).
134. Xie, P. Dynamics of strand passage catalyzed by topoisomerase II. *Eur. Biophys. J.* **39**, 1251–1259 (2010).
135. Chen, S. F. *et al.* Structural insights into the gating of DNA passage by the topoisomerase II DNA-gate. *Nat. Commun.* **9**, 1–13 (2018).
136. Miller, E. L. *et al.* TOP2 synergizes with BAF chromatin remodeling for both resolution and formation of facultative heterochromatin. *Nat. Struct. Mol. Biol.* **24**, 344–352 (2017).
137. Mata-Garrido, J., Casafont, I., Tapia, O., Berciano, M. T. & Lafarga, M. Neuronal accumulation of unrepaired DNA in a novel specific chromatin domain: structural, molecular and transcriptional characterization. *Acta Neuropathol. Commun.* **4**, 41 (2016).
138. See, K. *et al.* Lineage-specific reorganization of nuclear peripheral heterochromatin and H3K9me2 domains. *Development* **146**, dev174078 (2019).
139. 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**, 632 (2018).
140. Glover, T. W., Wilson, T. E. & Arlt, M. F. Fragile sites in cancer: More than meets the eye. *Nat. Rev. Cancer* **17**, 489–501 (2017).
141. Li, S. & Wu, X. Common fragile sites: protection and repair. *Cell Biosci.* **10**, 29 (2020).
142. Barlow, J. H. *et al.* Identification of early replicating fragile sites that contribute to genome instability. *Cell* **152**, 620–632 (2013).
143. Wei, P. C. *et al.* Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* **164**, 644–655 (2016).
144. Wei, P.-C. *et al.* Three classes of recurrent DNA break clusters in brain progenitors identified by 3D proximity-based break joining assay. *Proc. Natl. Acad. Sci.* 201719907 (2018).
145. Tena, A. *et al.* Induction of recurrent break cluster genes in neural progenitor cells differentiated from embryonic stem cells in culture. *Proc. Natl. Acad. Sci.* **117**, 10541–10546 (2020).
146. 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.e6 (2020).
147. Rybin, M. J. *et al.* Emerging Technologies for Genome-Wide Profiling of DNA Breakage. *Front. Genet.* **11**, 610386 (2021).
148. Fernandez-Capetillo, O. *et al.* DNA damage-induced G2-M checkpoint activation by histone H2AX and 53BP1. *Nat. Cell Biol.* **4**, 993–997 (2002).
149. Schultz, L. B., Chehab, N. H., Malikzay, A. & Halazonetis, T. D. p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *J. Cell Biol.* **151**, 1381–1390 (2000).
150. Lei, T., Du, S., Peng, Z. & Chen, L. Multifaceted regulation and functions of 53BP1 in NHEJ-mediated DSB repair (Review). *Int. J. Mol. Med.* **50**, 1–19 (2022).
151. Harrigan, J. A. *et al.* Replication stress induces 53BP1-containing OPT domains in G1 cells. *J. Cell Biol.* **193**, 97–108 (2011).
152. Lukas, C. *et al.* 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nat. Cell Biol.* **13**,

243–253 (2011).
153. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).
154. Szilard, R. K. *et al.* Systematic identification of fragile sites via genome-wide location analysis of γ-H2AX. *Nat. Struct. Mol. Biol.* **17**, 299–305 (2010).
155. Iacovoni, J. S. *et al.* High-resolution profiling of γh2AX around DNA double strand breaks in the mammalian genome. *EMBO J.* **29**, 1446–1457 (2010).
156. Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* **33**, 179–188 (2015).
157. Hu, J. *et al.* Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* **11**, 853–71 (2016).
158. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–5 (2013).
159. Yan, W. X. *et al.* BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).
160. Bouwman, B. A. M. *et al.* Genome-wide detection of DNA double-strand breaks by in-suspension BLISS. *Nat. Protoc.* (2020).
161. Canela, A. *et al.* DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Mol. Cell* **63**, 898–911 (2016).
162. Lensing, S. V. *et al.* DSBCapture: In situ capture and sequencing of DNA breaks. *Nat. Methods* **13**, 855–857 (2016).
163. Ballarino, R., Bouwman, B. A. M. B. A. M. & Crosetto, N. Genome-Wide CRISPR Off-Target DNA Break Detection by the BLISS Method. in *Methods in Molecular Biology* **2162**, 261–281 (2021).
164. Chiarle, R. *et al.* Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107–119 (2011).
165. Marin Navarro, A. *et al.* P53 Controls Genomic Stability and Temporal Differentiation of Human Neural Stem Cells and Affects Neural Organization in Human Brain Organoids. *Cell Death Dis.* **11**, 52 (2020).
166. Szlachta, K., Raimer, H. M., Comeau, L. D. & Wang, Y.-H. CNCC: an analysis tool to determine genome-wide DNA break end structure at single-nucleotide resolution. *BMC Genomics* **21**, 25 (2020).
167. Cao, H. *et al.* Novel approach reveals genomic landscapes of single-strand DNA breaks with nucleotide resolution in human cells. *Nat. Commun.* **10**, 1–14 (2019).
168. Sriramachandran, A. M. *et al.* Genome-wide Nucleotide-Resolution Mapping of DNA Replication Patterns, Single-Strand Breaks, and Lesions by GLOE-Seq. *Mol. Cell* **78**, 975-985.e7 (2020).
169. Cao, B. *et al.* Nick-seq for single-nucleotide resolution genomic maps of DNA modifications and damage. *Nucleic Acids Res.* **48**, 6715–6725 (2020).
170. Kilgas, S., Kiltie, A. E. & Ramadan, K. Immunofluorescence microscopy-based detection of ssDNA foci by BrdU in mammalian cells. *STAR Protoc.* **2**, 100978 (2021).
171. Broderick, R. *et al.* EXD2 promotes homologous recombination by facilitating DNA end resection. *Nature Cell Biology* **18**, 271–280 (2016).
172. He, Y. J. *et al.* DYNLL1 binds to MRE11 to limit DNA end resection in BRCA1-deficient cells. *Nature* **563**, 522–526 (2018).
173. Halder, S. *et al.* SPRTN protease and checkpoint kinase 1 cross-activation loop safeguards DNA replication. *Nature Communications* **10**, (2019).
174. Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, 0826–0842 (2005).

175. Bouwman, B. A. M., Crosetto, N. & Bienko, M. The era of 3D and spatial genomics. *Trends Genet.* **38**, 1062–1075 (2022).
176. Efroni, S. *et al.* Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell* **2**, 437–447 (2008).
177. Bouwman, B. A. M. & De Laat, W. Architectural hallmarks of the pluripotent genome. *FEBS Lett.* **589**, 2905–2913 (2015).
178. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
179. Harabula, I. & Pombo, A. The dynamics of chromatin architecture in brain development and function. *Curr. Opin. Genet. Dev.* **67**, 84–93 (2021).
180. Fraser, J. *et al.* Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* **11**, 852 (2015).
181. Rajarajan, P. *et al.* Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science* **362**, eaat4311 (2018).
182. Noack, F. *et al.* Multimodal profiling of the transcriptional regulatory landscape of the developing mouse cortex identifies Neurog2 as a key epigenome remodeler. *Nature Neuroscience* **25**, 154–167 (2022).
183. Winick-Ng, W. *et al.* Cell-type specialization is encoded by specific chromatin topologies. *Nature* **599**, 684–691 (2021).
184. Lu, L. *et al.* Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. *Mol. Cell* **79**, 521-534.e15 (2020).
185. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
186. Peric-Hupkes, D. *et al.* Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Mol. Cell* **38**, 603–613 (2010).
187. Tan, L. *et al.* Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* **184**, 741-758.e17 (2021).
188. Beagan, J. A. *et al.* Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. *Nat. Neurosci.* **23**, 707–717 (2020).
189. Gelali, E. *et al.* iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. *Nat. Commun.* **10**, 1–15 (2019).
190. Schermelleh, L. *et al.* Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320**, 1332–1336 (2008).
191. Jerkovic´, I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nat. Rev. Mol. Cell Biol.* **22**, 511–528 (2021).
192. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).
193. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).
194. Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.* **51**, 1252–1262 (2019).
195. Girelli, G. *et al.* GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat. Biotechnol.* (2020).
196. Rang, F. J. *et al.* Single-cell profiling of transcriptome and histone modifications with EpiDamID. *Mol. Cell* **82**, 1956-1970.e14 (2022).
197. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757.e24 (2018).
198. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).

199. Falk, A. *et al.* Capture of neuroepithelial-like stem cells from pluripotent stem cells provides a versatile system for in vitro production of human neurons. *PLoS One* **7**, 1–13 (2012).
200. Telley, L. *et al.* Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science* **351**, 1443–1446 (2016).
201. Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **14**, 1198–1204 (2017).
202. Sabatini, D. M. Twenty-five years of mTOR: Uncovering the link from nutrients to growth. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11818–11825 (2017).
203. Yamashita, R., Suzuki, Y., Sugano, S. & Nakai, K. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. *Gene* **350**, 129–136 (2005).
204. Bertucci, E. M. & Parrott, B. B. Is CpG Density the Link between Epigenetic Aging and Lifespan? *Trends Genet.* **36**, 725–727 (2020).
205. Dileep, V. & Tsai, L. H. Neuronal enhancers get a break. *Neuron* **109**, 1766–1768 (2021).
206. Langerman, J., Lopez, D., Pellegrini, M. & Smale, S. T. Species-Specific Relationships between DNA and Chromatin Properties of CpG Islands in Embryonic Stem Cells and Differentiated Cells. *Stem Cell Reports* **16**, 899–912 (2021).
207. Sutcliffe, J. S. *et al.* Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. *Nat. Genet.* **8**, 52–58 (1994).
208. Kosaki, K., McGinniss, M. J., Veraksa, A. N., McGinnis, W. J. & Jones, K. L. Prader-Willi and Angelman syndromes: Diagnosis with a bisulfite-treated methylation-specific PCR method. *Am. J. Med. Genet.* **73**, 308–313 (1997).
209. de Goede, O. M., Lavoie, P. M. & Robinson, W. P. Cord blood hematopoietic cells from preterm infants display altered DNA methylation patterns. *Clin. Epigenetics* **9**, 1–13 (2017).
210. Lister, R. *et al.* Epigenetic reconfigulation in Mammlian Brain Development. *Science* **341**, 1–21 (2013).
211. Wu, W. *et al.* Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* **593**, 440–444 (2021).
212. Delint-Ramirez, I. *et al.* Calcineurin dephosphorylates topoisomerase IIβ and regulates the formation of neuronal-activity-induced DNA breaks. *Mol. Cell* 1–16 (2022).
213. Thongthip, S., Carlson, A., Crossley, M. P. & Schwer, B. Relationships between genome-wide R-loop distribution and classes of recurrent DNA breaks in neural stem/progenitor cells. *Sci. Rep.* **12**, 1–12 (2022).
214. Michel, N. *et al.* Transcription-associated DNA DSBs activate p53 during hiPSC-based neurogenesis. *Sci. Rep.* **12**, 12156 (2022).
215. Cheng, Y.-C. *et al.* Topoisomerase I inhibition and peripheral nerve injury induce DNA breaks and ATF3-associated axon regeneration in sensory neurons. *Cell Rep.* **36**, 109666 (2021).
216. Palozola, K. C., Lerner, J. & Zaret, K. S. A changing paradigm of transcriptional memory propagation through mitosis. *Nat. Rev. Mol. Cell Biol.* **20**, 55–64 (2019).
217. Urban, E. A. & Johnston, R. J. Buffering and Amplifying Transcriptional Noise During Cell Fate Specification. *Front. Genet.* **9**, 1–14 (2018).
218. Voortman, L. & Johnston, R. J. Transcriptional repression in stochastic gene expression, patterning, and cell fate specification. *Dev. Biol.* **481**, 129–138 (2022).
219. Kouzine, F. *et al.* Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst.* **4**, 344-356.e7 (2017).

220. Bossaert, M. *et al.* Transcription-associated topoisomerase 2α (Top2a) activity is a major effector of cytotoxicity induced by g-quadruplex ligands. *Elife* **10**, 1–30 (2021).
221. Bizard, A. H. & Hickson, I. D. Anaphase: a fortune-teller of genomic instability. *Curr. Opin. Cell Biol.* **52**, 112–119 (2018).
222. Gamble, D., Shaltz, S. & Jinks-Robertson, S. Recombinational Repair of Nuclease-Generated Mitotic Double-Strand Breaks with Different End Structures in Yeast. *G3 Genes|Genomes|Genetics* **10**, 3821–3829 (2020).