UNIVERSITY OF CALIFORNIA SAN DIEGO

Replication stress associated genome instability in the pathogenesis of autism

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Biology

by

Meiyan Wang

Committee in charge:

Professor Fred H. Gage, Chair
Professor Cornelis Murre, Co-Chair
Professor Christopher K. Glass
Professor Albert La Spada
Professor Jonathan Sebat

2019

The Dissertation of Meiyan Wang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California San Diego

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Fred H. Gage for his support as the chair of my committee.

I would also like to acknowledge Professors Cornelis Murre, Christopher Glass, Albert La Spada, and Jonathan Sebat for their support as the members of my committee.

Introduction, in part, is a reprint of the material as it appears in Protein Cell 2019. Wang, Meiyan; Zhang, Lei; Gage, Fred, Springer Nature Switzerland AG. The dissertation author was the primary investigator and author of this paper.

Chapter 1, in full, has been submitted for publication of the material as it may appear in Cell Stem Cell, Wang, Meiyan; Wei, Pei-Chi; Lim, Christina; Gallina, Iryna; Marshall, Sara; Marchetto, Carol; Alt, Frederick; Gage, Fred, Elsevier Press, 2019. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, has been submitted for publication of the material as it may appear in Cell Stem Cell, Wang, Meiyan; Wei, Pei-Chi; Lim, Christina; Gallina, Iryna; Marshall, Sara; Marchetto, Carol; Alt, Frederick; Gage, Fred, Elsevier Press, 2019. The dissertation author was the primary investigator and author of this paper.

Discussion, in part, has been submitted for publication of the material as it may appear in Cell Stem Cell, Wang, Meiyan; Wei, Pei-Chi; Lim, Christina; Gallina, Iryna; Marshall, Sara; Marchetto, Carol; Alt, Frederick; Gage, Fred, Elsevier Press, 2019. The dissertation author was the primary investigator and author of this paper.

VITA

2012 Bachelor of Science, Tsinghua University, China

2019 Doctor of Philosophy, University of California San Diego

FIELDS OF STUDY

Major Field: Biology

Studies in Neuroscience
Professor Fred H. Gage

ABSTRACT OF THE DISSERTATION

Replication stress associated genome instability in the pathogenesis of autism

by

Meiyan Wang

Doctor of Philosophy in Biology

University of California San Diego, 2019

Professor Fred H. Gage, Chair

Professor Cornelis Murre, Co-Chair

The association between macrocephaly and autism spectrum disorder (ASD) suggests that the mechanisms underlying excessive neural growth could contribute to the pathogenesis of the disorder. Neural progenitor cells (NPCs) derived from induced pluripotent stem cells of ASD individuals with early developmental brain enlargement are inherently more proliferative than control NPCs. Here, we demonstrate that NPCs derived from ASD patients with macrocephaly display an altered DNA replication program and increased DNA damage. When compared to the control NPCs, high throughput genome-wide translocation sequencing demonstrates that ASD-derived NPCs harbored elevated DNA double-strand breaks in replication stress-susceptible genes, some of which are associated with the pathogenesis of ASD. Our results identify a mechanism linking hyperproliferation of NPCs with the pathogenesis of ASD by disrupting long neural genes involved in cell-cell adhesion and migration.

Introduction

Replication stress and genome instability

Replication stress is a complex phenomenon, which often causes genome instability and has severe impacts on cell survival and human diseases. In the eukaryotes, DNA replication originates at individual firing origins that form bidirectional replication forks (Zeman and Cimprich, 2014). Replication firing origins are tightly regulated and divided into early- and late-replicating origins. Replication stress arises from many different sources, including a wide range of physical obstacles. While there is no unifying definition of replication stress, some features have been described during this process, such as the slowing or stalling of the replication fork progression (Zeman and Cimprich, 2014).

Formation of aberrant replication fork containing single-stranded DNA (ssDNA) activates the replication stress response. The exposure of ssDNA, bound by replication protein A (RPA), recruits a number of replication stress-response proteins, including the ataxia-telangiectasia mutated (ATM)- and Rad3-related (ATR) (Zou and Elledge, 2003, Marechal and Zou, 2013). ATR activation is crucial for cell survival and faithfully duplication of the genome under stress (Zeman and Cimprich, 2014).

Persistent replication stress induces genome instability. For example, fork collapse often involves double-strand breaks (DSBs) formation at the stalled fork. The ATR kinase is required to stabilize stalled forks and prevent DNA breaks at these sites. Certain regions of the genome are more prone to replication stress-induced DNA damage.

Even at mild levels of stress, these “common fragile sites” (CFS) are sensitive to breakage (Debatisse et al., 2012).

Another constant challenge to DNA replication is the transcription activity. Since transcription and replication occur at the same time during S-phase and both operate on DNA, it is inevitable that the two processes will interfere with each other (Zeman and Cimprich, 2014, Techer et al., 2017). Collision of the replication and transcription complexes are known to induce DSBs. For example, in the mammalian cells, a set of highly transcribed genomic regions replicated early in the S-phase, are especially prone to DSB formation (Barlow et al., 2013). Moreover, transcription of very large genes often correlates with CFS instability (Wei et al., 2016, Wilson et al., 2015, Helmrich et al., 2011). Although the mechanisms by which transcription destabilizes certain genomic regions remain obscure, these results emphasize the impact of transcription on fork progression and genome instability.

LINE-1 retrotransposon

Retrotransposons are genetic elements that move in the host genome utilizing a “copy-and-paste” mechanism. 17% of human genome is composed of the Long Interspersed Nuclear Element-1 (LINE-1) retrotransposon (Lander et al., 2001). LINE-1 is the only autonomous retrotransposon in the human genome that is able to “jump” in the host genome through a process called retrotransposition. The full length, active LINE-1 consists of a 5’ translated region (5’UTR) (Swergold, 1990, Speek, 2001), two open reading frames (ORFs), ORF1 and ORF2, respectively, and a 3’ UTR with a weak polyadenylation signal (Doucet et al., 2015, Dombroski et al., 1991, Mathias et al., 1991). ORF0, a primate specific ORF in the antisense orientation, was also described in the

5'UTR of LINE-1 with a potential role to form fusion proteins with proximal exons (Denli et al., 2015). The 6kb bicistronic LINE-1 mRNA is poly-adenylated and exported into the cytoplasm. The newly translated ORF1 and ORF2 proteins show strong cis-preference and bind the LINE-1 mRNA that encodes them (Wei et al., 2001).

ORF1p is an RNA binding protein with chaperone activity and ORF2p has endonuclease (EN) and reverse transcriptase (RT) activity (Mathias et al., 1991, Khazina et al., 2011, Feng et al., 1996). Ribonucleoprotein particles (RNPs), composed of ORF1p, ORF2p, and LINE-1 mRNA, are then imported back into the nucleus (Khazina et al., 2011). In the nucleus, ORF2p EN domain nicks the DNA at the A/T rich consensus target sites (TTTT/AA) (Feng et al., 1996). A complementary DNA strand is synthesized using the LINE-1 mRNA by the ORF2p RT domain through a process called target primed reverse transcription (TPRT) (Luan et al., 1993). A second strand of cDNA is then synthetized through an unknown process and joined to the genomic DNA.

LINE-1 remains active in both human and mouse genomes and contributes to genome mosaicism. Until recently, retrotransposition was believed to occur primarily in the germ cells and cancer tissues in mammals. However, in the past decade, several lines of evidence have demonstrated that LINE-1 retrotransposition occurs in the brain (Coufal et al., 2009, Muotri et al., 2005), at a rate of 0.2-16.3 events per neuron in the healthy human brains (Erwin et al., 2016, Upton et al., 2015, Evrony et al., 2012).

Disease modeling using human induced pluripotent stem cells

The recent advances in technologies that enable adult human somatic cells to be reprogrammed into induced pluripotent stem cells (iPSCs) (Takahashi et al., 2007, Yu et al., 2007) hold enormous potential in the field of regenerative medicine (Shi et al., 2017).

Despite the promise of iPSCs as an autologous cell source for cell transplantation therapy, concerns have been raised regarding the clinical applications of iPSCs (Zhao et al., 2015, Zhao et al., 2011, Kang et al., 2016, Ji et al., 2012, Gore et al., 2011). While much work remains to be done to improve the safety and reliability of the technology, iPSCs have provided a unique opportunity to study developmental neuropsychiatric diseases. In the past, this type of research has been challenging due to the inaccessibility of relevant live tissues and cell types. The lack of reliable models has significantly hindered the progress in understanding the pathogenesis of the disease and development of novel treatments.

Animal models provide valuable insights into the mechanisms of specific functions of genes. However, modeling of human neuropsychiatric disorders in animals is extremely challenging given the complexity of the disorders, the subjective nature of many symptoms, and the lack of biomarkers and reliable tests (Nestler and Hyman, 2010). The differences in the brain structures of human and mouse have also made it difficult to study behaviors related to higher-function brain areas such as the dorsolateral prefrontal cortex, one of the most recently derived parts of the human brain that is often implicated in neuropsychiatric disorders (Tekin and Cummings, 2002). In addition, it is difficult to assess some complex phenotypes in mouse, such as psychosis.

Postmortem human brains serve as valuable sources for examining pathological changes in patients. However, these tissues only represent the disease endpoint, whereas the alterations that lead to the development of the diseases often occur early in development. Postmortem brain studies are thus limited in their ability to reveal dynamic neuronal changes that are often important in the disease mechanisms.

iPSC technology has opened the field to new discoveries in neuropsychiatric diseases. iPSCs can be directly generated from somatic cells by the introduction of reprogramming factors (Takahashi et al., 2007). Moreover, iPSCs are capable of differentiating into adult cell types through directed differentiation with a combination of small molecules, growth factors, and morphogens, allowing the derivation of unlimited disease-relevant cells carrying the variations that caused or facilitated the development of the disease. These cells include human neurons, a previously inaccessible cell type. Most neuropsychiatric diseases, including ASD, schizophrenia, bipolar disorder, and major depression, have a strong but complex genetic component. Multiple low penetrance genetic variants contribute to the etiology of those disorders (Schizophrenia Working Group of the Psychiatric Genomics, 2014, Bipolar et al., 2018, Grove et al., 2019, Wray et al., 2018). Because human iPSCs capture the genetic diversity of the patient, they are particularly useful for modeling how complex genetic variants lead to the pathogenesis of the disease.

Macrocephaly and autism

ASD, a class of clinically heterogenous neurodevelopmental disorders, is a highly heritable condition defined by deficits in social interaction and communication, as well as restrictive, repetitive behaviors. About 1 in 5 autistic children has macrocephaly, a clinical condition refers to an abnormal enlargement of head size, including the scalp, the cranial bones, and brain size. Increased neuronal number has been reported in the dorsolateral prefrontal cortex of the autism group compared to control group (Courchesne et al., 2011). Although exact cause of this peculiar phenotype in autism is as yet unknown, genes that are associated with this abnormal pattern of brain growth in autism

have been identified. For example, mutations in *CHD8* (O'Roak et al., 2012), chromodomain helicase DNA binding protein 8, a regulator of Wnt/β-catenin signaling.

iPSC models of macrocephalic ASD have been established (Marchetto et al., 2017, Mariani et al., 2015). Interestingly, abnormal cell proliferation was found in the NPCs derived from the autistic iPSCs. Genetic analysis has identified mutations in genes involved in the Wnt/β-catenin signaling in these autistic iPSC lines (Marchetto et al., 2017). In accordance, dysregulation of Wnt/β-catenin signaling was also described in the NPCs from iPSC lines.

Abnormal cell proliferation induces replication stress in cells. Here, we characterize genomic regions sensitive to DNA damage in human NPCs upon replication stress by high-throughput genome-wide translocation sequencing, and investigate the role of replication stress associated genome instability in the pathogenesis of ASD using patients-derived iPSCs.

Acknowledgements

Introduction, in part, is a reprint of the material as it appears in Protein Cell 2019. Wang, Meiyan; Zhang, Lei; Gage, Fred, Springer Nature Switzerland AG. The dissertation author was the primary investigator and author of this paper.

Chapter 1 Replication stress induces DNA double-strand breaks in a group of transcribed long genes in human neural progenitor cells

Introductions

Brain development requires that the neural precursor cells undergo millions of cell division during embryonic development and early years of life to give rise to most of the 80 billion neurons in the human brain (Lui et al., 2011). DNA damage, especially DNA DSBs, can generate *de novo* somatic mutations during development. Somatic mutations contribute to genomic diversity and cause many human genetic disorders and cancers (Greenman et al., 2007, Poduri et al., 2013, McConnell et al., 2017). Because neurons are among the longest-lived cells in the body, accumulation of somatic mutations during brain development could influence neuronal development and function (McConnell et al., 2017).

Efficient DNA repair is imperative for neural development. For example, mice deficient in certain components of the classical non-homologous end-joining (C-NHEJ) [*e.g.*, DNA ligase IV (*Lig4*) and X-Ray Repair Complementing Defective Repair In Chinese Hamster Cells 4 (*Xrcc4*)] pathway exhibit late embryonic lethality due to extensive neuronal apoptosis (Gao et al., 1998, Barnes et al., 1998, Frank et al., 2000). Moreover, neuronal death and embryonic lethality in C-NHEJ-deficient mice are rescued by p53 deficiency (Frank et al., 2000, Gao et al., 2000), indicating a role of DNA damage-induced apoptosis in generating these phenotypes.

Even in the presence of functional DNA repair pathways, DSBs can still induce *de novo* somatic mutations. Errors in DNA replication and repair can lead to the formation of copy number variations (CNVs). Somatic CNVs can then reshape neural

development, function and diversity. While abundant CNVs in neurons have been detected both *in vitro* and *in vivo* (McConnell et al., 2013, Cai et al., 2014), the rate of brain somatic mutations and their underlying mechanisms remain unclear.

Recent studies have identified numerous recurrent DSB clusters (RDCs) in mouse neural precursors that are located in genes, the majority of which occur in long neural-specific genes associated with neuropsychiatric diseases and cancers, suggesting potential impacts of DNA damage on neural development and function (Wei et al., 2016, Wei et al., 2018). A subset of RDC-containing genes are found within CNVs, of which some correspond to genomic regions harboring known CFSs (Wei et al., 2016, Wei et al., 2018, Glover and Wilson, 2016).

In this chapter, we applied high-throughput genome-wide translocation sequencing to investigate genomic regions susceptible to replication stress. We mapped DSB sites at nucleotide resolution in NPCs. We showed that replication stress induced a plethora of DSBs in the longest genes of the genome in NPCs. Intriguingly, replication stress attenuated the expression of many of these susceptible genes involved in adherens junctions, apical polarity, cell migration, and disrupted relative functions in NPCs.

High-throughput mapping of DSBs in human NPCs

To investigate genomic regions susceptible to DSBs upon replication stress in human NPCs, we took advantage of recent advances in the derivation of NPCs (Yu et al., 2014) from human iPSCs (hiPSCs) and human embryonic stem cells (hESCs). We treated NPCs with low doses of aphidicolin (APH) (Wei et al., 2016, Wei et al., 2018), a reversible inhibitor of eukaryotic DNA replication (Glover et al., 1984), and performed high-throughput genome-wide translocation sequencing (HTGTS) assay (Frock et al.,

2015, Hu et al., 2016, Wei et al., 2016, Wei et al., 2018) that maps DSBs genome-wide at nucleotide resolution based on their ability to translocate to bait DSBs (Figure 1.1A). We employed a Cas9:single-guide RNA (sgRNA) approach to generate a HTGTS bait DSB at 1p36.22 in NPCs derived from hiPSCs. A biotinylated primer that specifically recognizes the telomeric broken end of the bait DSB was used to amplify endogenous prey DSBs that joined the bait DSBs. We detected 32649 unique prey DSBs located more than 10kb away from the bait DSB in NPCs treated with vehicle control (DMSO). Approximately 75% of those DSBs were located on chromosomes other than chromosome (Chr) 1 (Figure 1.1B). After removing DSB hotspots within low complexity/repeat regions identified by the RepeatMasker track of the UCSC genome browser (http://genome.ucsc.edu/) (Table 1.1), six prey DSB hotspots were identified (Figure 1.1B, example shown in 1.1C). SgRNA off-target (OT) analysis revealed that all of the detected DSB hotspots were Chr1-sgRNA OTs (Table 1.1). By comparing them to the HTGTS data of NPCs treated with APH, we found several new prey DSB hotspots that were unique to APH-treated NPCs (Figure 1.1D, example shown in 1.1E). Therefore, replication stress can induce DSB hotspots in NPCs derived from hESCs/hiPSCs.

Human NPCs harbor replication stress-susceptible genes

To profile genome-wide replication stress-induced DSB hotspots (Wei et al., 2016, Wei et al., 2018) in NPCs, we carried out HTGTS analysis of two hiPSC-derived NPC lines and one hESC-derived NPC line with bait DSBs on either Chr1 or Chr11 (Figure 1.2A and Table 1.1) and performed DSB hotspot calling with a modified RDC-identification pipeline (Wei et al., 2016, Wei et al., 2018). We performed at least three independent HTGTS experiments on DMSO- or APH-treated cells for each of the baits

and NPC lines. We employed the spatial clustering approach for the identification of chromatin immunoprecipitation-enriched regions (SICER) algorithm (Zang et al., 2009) and focused on common regions in multiple lines revealed by at least one bait; we identified 37 DSB hotspots (Table 1.2). The size of DSB hotspots ranged from 180kb to 1.7Mb, with a median length of 570kb (Figure 1.2B). To facilitate the comparison of HTGTS libraries generated by different baits and cell lines, we calculated DSB densities defined as number of DSBs per Mb per 10,000 total DSBs. As expected from proximity effects (Alt et al., 2013), we found higher DSB densities in replication stress-induced DSB hotspots located on the same chromosome as the bait DSBs. Therefore, we analyzed intra-chromosomal DSB hotspots (hotspots located on the same chromosome as the bait DSB) and inter-chromosomal DSB hotspots (hotspots located on chromosomes other than the bait chromosome) separately. We found significantly higher DSB density in 36 of the 37 DSB hotspots in APH-treated NPCs when compared to DMSO-treated NPCs (Figure 1.2C and 1.2D). Notably, 36 of the 37 DSB hotspots overlapped protein-coding genes (Figure 1.3A, 1.3B, and Table 1.3). In summary, our results indicate that replication stress induced DSB hotspots in 36 susceptible genes in human iPSCs/ESCs-derived NPCs.

Replication stress in NPCs induces DSBs in long genes

To assess the DSB enrichment in genes, we calculated the DSB density in the 36 susceptible genes that overlapped with DSB hotspots using the gene coordinates from the hg19 reference genome. Thirty-one of the 36 susceptible genes harbored significantly greater DSB density in APH-treated NPCs than DMSO-treated NPCs (Figure 1.4A and 1.4B; ANOVA corrected for two-stage linear step-up procedure of Benjamini, Krieger

and Yekutieli, FDR < 0.05). Consistent with previously published studies of mouse cells (Wei et al., 2016, Wei et al., 2018), we found that replication stress-susceptible genes in human NPCs were also enriched in long genes (Figure 1.4C). To determine whether the extent of replication stress-induced DSBs in genes was directly correlated with gene length, we quantified the number of DSBs in genes longer than 100kb and interrogated the enrichment of DSBs upon replication stress (fold-change of DSB densities of genes in APH-treated versus DMSO-treated cells) versus gene length. We found length-dependent replication stress-induced DSB enrichment in NPCs, with the longest genes in the genome displaying the highest level of DSBs after induction of replication stress (Figure 1.4D and 1.4E). The length dependence of gene fragility upon replication stress was reproducibly detected in all NPC lines (Figure 1.4D and 1.4E). We verified that DSB densities of long genes (genes longer than 800kb located on chromosomes other than Chr1) in APH-treated NPCs were significantly greater than those in DMSO-treated NPCs (Figure 1.4F). Markedly, 30 of the 36 susceptible genes were longer than 800kb. Furthermore, in APH-treated NPCs, long genes harbored significantly greater DSB densities than medium genes (genes between 400kb to 800kb) (Figure 1.4G). Taken together, our data suggest that replication stress induced DSBs in long genes.

DNA damage correlates with transcription in long genes

The occurrence of DSB hotspots in genes suggests that transcription could potentially influence DSB sites. To investigate this possibility, we performed Global Run-On sequencing (GRO-seq) (Core et al., 2008) (Figure 1.5A). Thirty-four of the 36 susceptible genes were actively transcribed in NPCs (Figure 1.6A). In addition, enriched DSBs were observed at the actively transcribed region of the gene, suggesting the

involvement of transcription in generating DSBs (Figure 1.5B, 1.6B and 1.6C). To further explore this finding, we divided the long genes into three groups based on the gene expression level from GRO-seq: high, medium, and low. We then compared the DSB densities of the genes in the high expression group to the genes in the low expression group. We found that, upon replication stress, genes in the high expression group had more DSBs than genes in the low expression group (Figure 1.6D). We also observed that the expression level of long genes was positively correlated with DSB density of genes (Figure 1.6E). Analysis of the replication program showed that APH-treated cells were accumulated in mid to late S-phase (Figure 1.5C-E). To investigate whether replication stress induced conflicts between replication fork and transcription machineries, we monitored their interaction by proximity ligation assay (PLA) (Hamperl et al., 2017). Antibodies against RNA polymerase II (RNAPII) and Proliferating Cell Nuclear Antigen (PCNA) were used to detect transcription machineries and replication forks, respectively (Figure 1.5F). This analysis revealed increased PLA foci in APH-treated NPCs compared to DMSO-treated NPCs (Figure 1.6F and 1.6G). Collectively, our data demonstrate that replication stress induced conflicts of replication fork and transcription machineries, leading to increased DNA damage in transcribed long genes.

Aberrant adherens junctions, apical polarity, and cell migration of NPCs upon replication stress

DSB repair interferes with replication and transcription. The induction of a single DSB at a human RNAPII-transcribed gene can lead to inhibition of transcription elongation and re-initiation (Pankotai et al., 2012). Moreover, increased conflicts between replication machineries and replication forks may also attenuate transcription of the

genes. The 36 susceptible genes play important roles in neural function, including cell-cell adhesion and cell migration (Figure 1.7A). We first carried out qPCR analysis of Hues6-derived NPCs and found that many of the susceptible genes implicated in cell-cell adhesion were significantly downregulated after replication stress (Figure 1.8A). Neural rosettes from Hues6 NPCs showed strong expression of the neural cell adhesion molecule N-cadherin at the center of the luminal surface of each rosette (Figure 1.7B), representing typical formation of adherens junctions of NPCs. Interestingly, while the structure of the rosettes was largely maintained, the expression of N-cadherin was disrupted after replication stress (Figure 1.7C), as indicated by a scattered expression of N-cadherin (Figure 1.7B). Because adherens junctions are crucial for maintaining cell polarity, we asked whether the apical-basal polarity of NPCs was affected. Neural rosettes showed robust expression of the apical polarity marker (Figure 1.7D), atypical PKCλ, representing the typical formation of apical-basal polarity of NPCs. Upon replication stress, disruption of the PKCλ structure was observed in NPCs, as indicated by the disrupted structure or absence of PKCλ in the center of neural rosettes (Figure 1.7D and 1.7E). We observed no apparent cell death when the neural cultures were treated with 0.25 μM APH for 2 days and a slight increase in cell death at 0.5 μM APH for 2 days (Figure 1.8C). Aberrant cell-cell adhesion and apical polarity of the neural rosettes were observed at the low dose (Figure 1.7C and 1.7E), indicating that these phenotypes were not induced by cell death.

N-cadherin is also involved in cell migration. A number of susceptible genes implicated in neuronal migration were significantly downregulated upon replication stress (Figure 1.8B). To investigate whether replication stress affected cell migration, we

generated neurospheres from Hues6-derived NPCs and performed a neurosphere migration assay in the presence or absence of APH (Figure 1.8D). Cell body distances from the neurosphere were measured after 48 hours of treatment. We found migration defects of neurospheres treated with 0.25 μM APH (Figure 1.7F). Taken together, these results suggest that replication stress attenuates the expression of many of the susceptible genes and induces aberrant adherens junctions, apical polarity, and cellular migration of NPCs.

Methods

Cell culture

Human ESC- and iPSC-derived NPCs were cultured as described before (Yu et al., 2014, Marchetto et al., 2017). Protocols describing the use of iPSCs and hESCs were previously approved by the University of California San Diego and Salk Institute Institutional Review Board and the Embryonic Stem Cell Research Oversight Committee. All the assays were carried out using passage 6 to 8 NPCs. The cells were plated at 100-150 k/cm^2 and cultured for two or three days for the analysis.

DSB induction

Bait DSB induction was achieved with a Cas9:sgRNA approach (Frock et al., 2015). Briefly, NPCs were culture until confluent and dissociated with Accutase. 5 million cells were nucleofected with 5 μg of Cas9:sgRNA expression vector using the Nucleofector reagent for Rat Neural Stem Cell (Lonza, VPG-1005) as per manufacturer's instruction. Cells were plated at 200-300 k/cm^2 post nucleofection and cultured for 4 days before harvesting. Replication stress was induced by the addition of 0.5 μM aphidicolin (APH, Sigma) for 3 days and then 0.25 μM APH for 1 day.

Cas9:sgRNA plasmid construction

Cas9:sgRNA expression vectors were constructed as described (Cong et al., 2013). Annealed oligonucleotides (see Table 1.1 for details) were ligated into BbsI digested pX330-U6-Chimeric_BB-CBh-hSpCas9 vector (Addgene plasmid #42230).

HTGTS and related bioinformatic analyses

LAM-HTGTS was performed and analyzed as previously described (Hu et al., 2016, Wei et al., 2016, Wei et al., 2018). FASTQ output files were de-multiplexed, and unique reads aligned to genome build hg19 by Bowtie2 (Langmead and Salzberg, 2012) were processed through the HTGTS pipeline (Hu et al., 2016). Reads with less than 50bp bait sequence were excluded and only unique HTGTS junctions were kept. Primers used are described in Table 1.1.

HTGTS junction enrichment analysis

A modified RDC identification pipeline (Wei et al., 2016) was used to identify DSB hotspot candidates. The analyses were performed by SICER (Zang et al., 2009) of concatenated control (DMSO) or treated (APH) HTGTS libraries (excluding junctions within 10kb of the bait break-site) using the following parameters (Wei et al., 2016): *SICER.sh Species- hg19; redundancy threshold- 5; window size- 30000; fragment size- 1; effective genome fraction- 0.8; gap size (bp)- 90000; FDR- 0.1.* Only clusters with more than five junctions (more than 10 junctions if on the same chromosome as the bait DSB) from APH-treated libraries were considered.

To identify DSB hotspots in NPCs, we separated the candidates into intra-chromosomal and inter-chromosomal DSB hotspots. For intra-chromosomal DSB hotspots, clusters had to be independently identified in three NPC lines by the bait

located on the same chromosome. For inter-chromosomal DSB hotspots, clusters identified in three NPC lines or identified in two NPC lines by both baits were kept. The remaining inter-chromosomal DSB hotspots identified in two NPC lines were subjected to statistical test (two tailed t-test) and only the significant ones were reported.

Identification of recurrent translocation to Cas9:sgRNA off-target sites

Translocations between Cas9:sgRNA bait and off-target DSBs were identified as described (Frock et al., 2015) by MACS2 (Zhang et al., 2008) with the following parameters: *-g hs --keep-dup all --nomodel --extsize 500 -q 0.001 --llocal 10000000*. Hotspots ≥100 kb from the bait DSB break-site with an FDR-adjusted P-value threshold of 1×10^{-9} were considered translocations between Cas9:sgRNA bait and off-target DSBs if they shared >30% sequence with the bait site in multiple libraries.

Global run-on sequencing

GRO-seq libraries were prepared as previously described (Meng et al., 2014) from 5 to 10 million NPC nuclei. Two biological replicates of hiPSC-1 derived NPCs were performed. GRO-seq data were aligned to human genome build hg19 by Bowtie2 and non-redundant, uniquely mapped sequence reads were retained. Gene expression levels were analyzed by HOMER (analyzeRepeats.pl rna hg19 -count genes -strand - -rpkm) (Heinz et al., 2010).

Immunofluorescence

Cells cultured on slides were fixed in 4% paraformaldehyde (10 min, room temperature), permeabilized with 0.1% Triton X-100, blocked in 5% horse serum, and incubated with primary antibodies overnight at 4 °C. After wash, cells were incubated with secondary antibodies for 1 h at room temperature, washed, incubated with DAPI for

10 min, and mounted using ProLong™ Gold Antifade Mountant (Life Technologies). Image acquisition was performed using Zeiss LSM 710 or LSM 880 Laser Scanning Confocal Microscope or Zeiss LSM 880 Laser Scanning Confocal Microscope (Carl Zeiss).

Proximity ligation assay

Proximity ligation assay was performed using Duolink® In Situ Orange Kit (MilliporeSigma) according to manufacturer's instruction with minor modifications. Briefly, cells were fixed 4% paraformaldehyde (10 min, room temperature), permeabilized with 0.1% Triton X-100, blocked in 2% horse serum and 2% BSA, and incubated with primary antibodies overnight at 4 °C (1:2000 rabbit PCNA (abcam, ab18197) alone; 1:1000 mouse RNAPII (CTD4H8, santa cruz biotechnology, sc-47701); or 1:2000 rabbit PCNA and 1:1000 mouse RNAPII). Cells were then processed according to manufacturer's instructions. Slides were imaged using Zeiss LSM 710 Laser Scanning Confocal Microscope (Carl Zeiss).

Neural rosette formation assay

Neural rosettes from Hues6 hESCs were generated based on a previously published protocol (Yu et al., 2014). Briefly, EBs were formed by mechanical dissociation of Hues6 colonies using collagenase IV and cultured in low-adherent plates. For EB differentiation, floating EBs were treated with DKK1 (0.5 μg/ml), SB431542 (10 μM), Noggin (0.5 μg/ml), and cyclopamine (1 μM) in DMEM/F12 (Gibco) plus N2 and B27 supplements for 20 days. To obtain neural rosettes, EBs were plated on poly-ornithine and laminin-coated plates in DMEM/F12 (Gibco) plus N2 and B27 supplements and laminin (Invitrogen, 1μg/ml). Rosettes were manually collected and dissociated with

Accutase after 1 week and plated onto poly-ornithine and laminin-coated plates. Rosettes were passaged again at high density with gentle dissociation with Accutase. Only individual non-overlapped neural rosettes with typical morphology were included for quantification. The quantifications were performed with at least three independent cultures.

Neurosphere migration assay

The neurosphere migration assay was performed according to previously published protocols (Brennand et al., 2015, Marchetto et al., 2019). Briefly, NPCs were dissociated with Accutase and then cultured for 3 days in low-adherent plates to generate neurospheres. Neurospheres were then manually picked and plated in Matrigel matrix (0.5 mg Matrigel was used to coat one 24-well plate for at least 1 hour before plating). The next morning, neurospheres derived from Hues6 NPCs were treated with NPC media with DMSO or 0.25 μM APH and fixed 48 hours later to assess NPC migration. To compare cell migration of ASD-derived NPCs and control NPCs, neurospheres were manually picked and plated in Matrigel-coated plates and maintained for 60 hours in NPC media. Cell migration distance from each neurosphere was measured using Image J software (NIH).

RNA extraction and quantitative PCR

Total cellular RNA was extracted from 2-5 x10^6 cells using the RNA-BEE (TEL-TEST, INC) and RNA Clean & Concentrator Kit (Zymo research), according to the manufacturer's instructions, and reverse transcribed using SuperScriptIII First-Strand Synthesis System (Invitrogen). qPCR was done using SYBR green (Life Technologies).

Acknowledgements

Chapter 1, in full, has been submitted for publication of the material as it may appear in Cell Stem Cell, Wang, Meiyan; Wei, Pei-Chi; Lim, Christina; Gallina, Iryna; Marshall, Sara; Marchetto, Carol; Alt, Frederick; Gage, Fred, Elsevier Press, 2019. The dissertation author was the primary investigator and author of this paper.

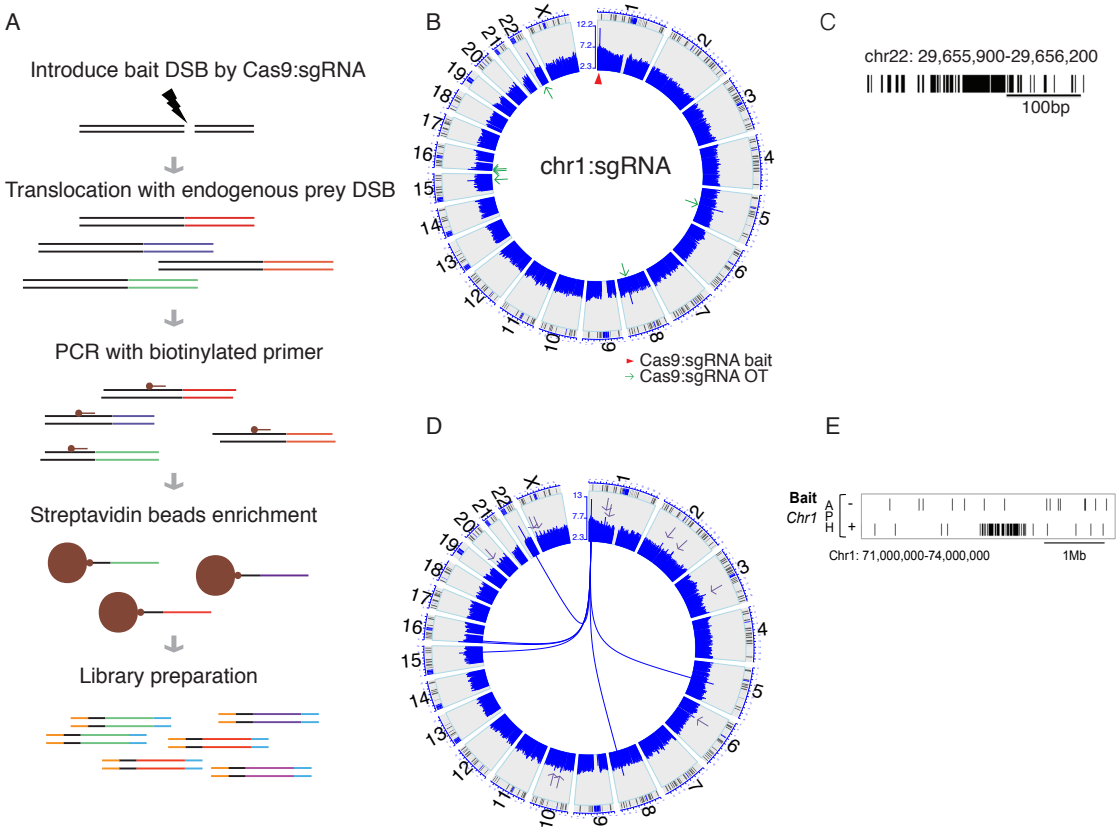


Figure 1.1:High-throughput mapping of DSBs in NPCs.

(A) The workflow of HTGTS.
(B) The genome-wide HTGTS junctions of *Chr1*-sgRNA-mediated bait DSBs of DMSO-treated hiPSC-derived NPCs binned into 2Mb regions are plotted in Circos plot. Bins containing more than 4 unique junctions are kept. Bar height indicates the number of unique DSB junctions per bin on a log scale. DSB junctions within the detected low complexity/repeat regions are not shown. Red triangle indicates the genomic location of *Chr1*-sgRNA-mediated bait DSB. Green arrows indicate the location of six Cas9:sgRNA OTs.
(C) Junctions captured by *Chr1*-sgRNA-mediated bait DSB of DMSO-treated NPCs within chr22: 29,655,900-29,656,200 where Cas9:sgRNA OT was detected.
(D) Circos plot of the human genome shows the genome-wide HTGTS junctions of *Chr1*-sgRNA-mediated bait DSBs in APH-treated NPCs binned into 2Mb. DSB junctions within the detected low complexity/repeat regions are not shown. Bins containing more than 4 unique junctions are plotted. Bar height shows the number of unique DSB junctions per bin on a log scale. Blue lines connect the bait and Cas9:sgRNA OT sites. Purple arrows indicate several bins with increased junctions in APH-treated cells compared to DMSO-treated cells.
(E) Junctions captured by *Chr1*-sgRNA-mediated bait DSB of DMSO-treated (-) or APH-treated (+) NPCs within chr1:71,000,000-74,000,000.

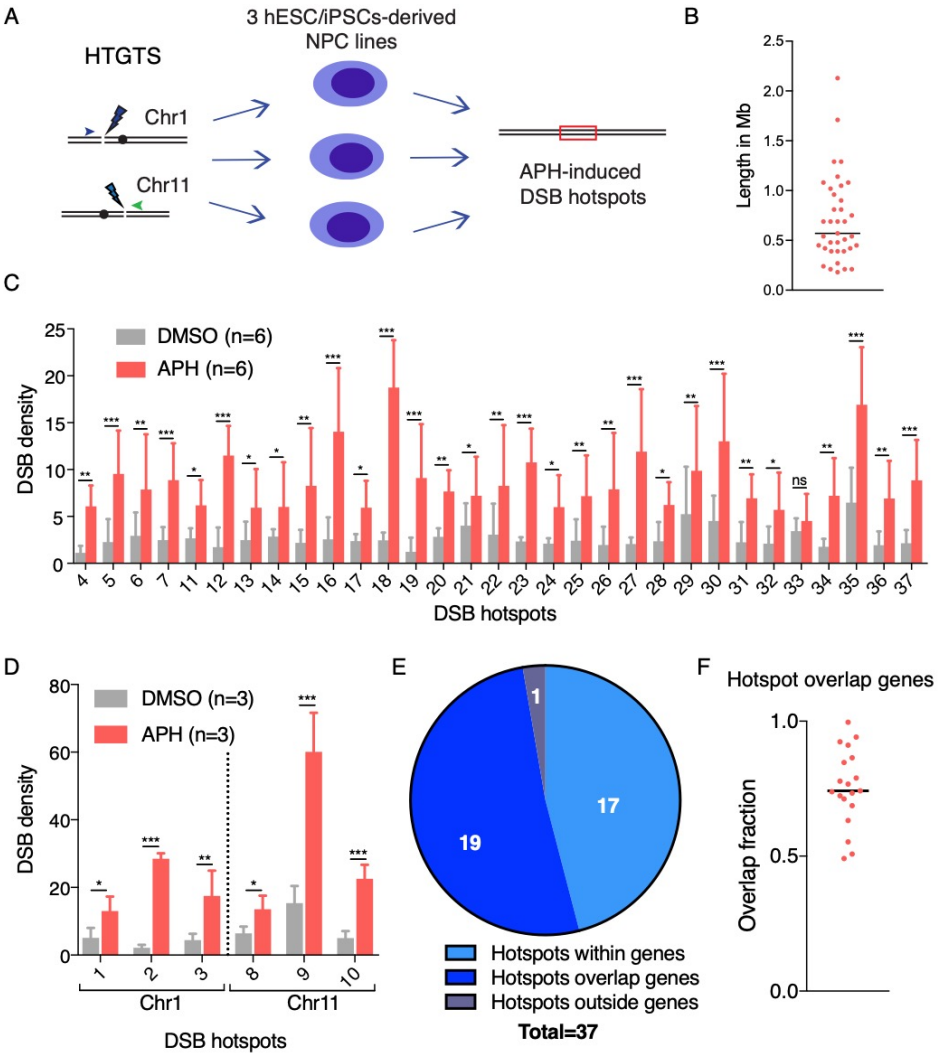


Figure 1.2: NPCs harbor replication stress-induced DSB hotspots in genes.

(A) Illustration shows the identification of replication stress-induced DSB hotspots in NPCs.
(B) Scatter plot of the size of DSB hotspots in Mb. Line represents median length (570kb).
(C) Bar plot shows the DSB densities captured by Chr1 bait and Chr11 bait within the indicated inter-chromosomal DSB hotspots. Mean ± SD. ANOVA corrected for two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli, *** FDR < 0.001, ** FDR < 0.01, * FDR < 0.05, ns, not significant.
(D) Bar plot shows the DSB densities captured by Chr1 bait (left) or Chr11 bait (right) within the indicated intra-chromosomal DSB hotspots located on Chr1 or Chr11. Mean ± SD. ANOVA corrected for two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli, *** FDR < 0.001, ** FDR < 0.01, * FDR < 0.05.
(E) Pie chart shows the location of the DSB hotspots.
(F) Scatter plot of the overlap fraction of DSB hotspot interval in genes. 19 DSB hotspots overlapping genes are plotted.

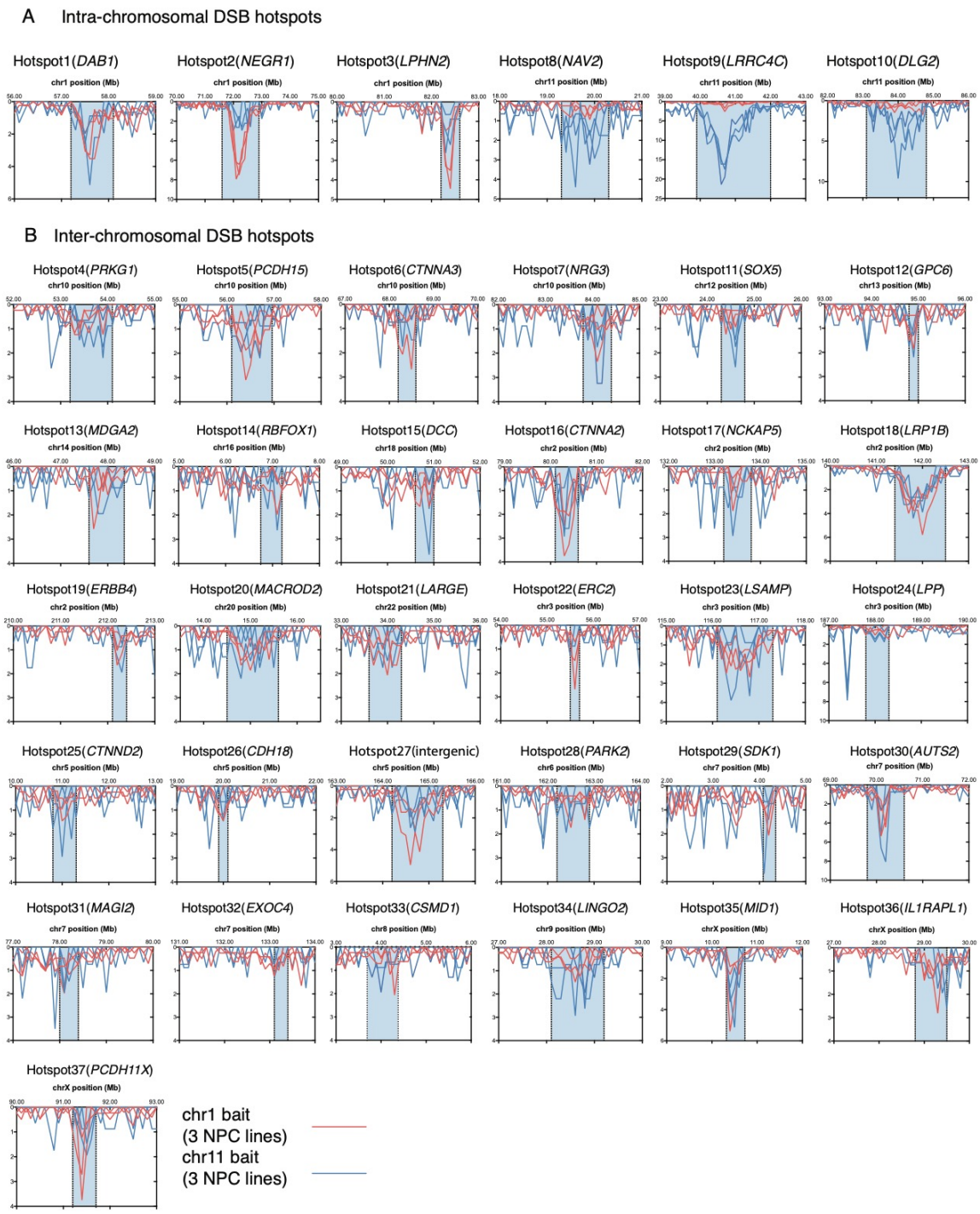


Figure 1.3: Replication stress-induced DSB hotspots in NPCs.

(A) Prey junctions within the indicated regions per 10000 total junctions of APH-treated NPCs captured by chr1 bait (red) and chr11 bait (blue), bin size 100kb. Genomic region corresponding to the indicated DSB hotspot is highlighted in blue.
(B) Prey junctions within the indicated regions per 10000 total junctions of APH-treated NPCs captured by chr1 bait (red) and chr11 bait (blue), bin size 100kb. Genomic region corresponding to the indicated DSB hotspot is highlighted in blue.

Figure 1.4: NPCs harbor replication stress-induced DSBs in long genes.

(A) Bar plot shows the DSB densities captured by Chr1 bait and Chr11 bait within the indicated genes located on chromosomes other than Chr1 and Chr11. Mean ± SD. ANOVA corrected for two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli, *** FDR < 0.001, ** FDR < 0.01, * FDR < 0.05.
(B) Bar plot shows the DSB densities captured by Chr1 bait (NEGR1) or Chr11 bait (LRRC4C and DLG2) within the indicated genes located on Chr1 or Chr11. Mean ± SD. ANOVA corrected for two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli, *** FDR < 0.001, * FDR < 0.05.
(C) Gene length comparison of all protein coding genes that are expressed (left) and the 36 replication stress-susceptible genes (right). Box-and-whiskers plot shows the gene length in kb on a log scale. Min to max is plotted. Mann-Whitney test, two-tailed, **** $p < 0.0001$.
Genome-wide changes in DSB density assessed by HTGTS captured by Chr1 bait (D) or Chr11 bait (E) of APH-treated versus DMSO-treated NPCs. Lines represent mean fold-change (APH/DMSO) of DSB density in genes on a log scale for genes binned according to gene length (100 gene bins, 25 gene step); the ribbon is the SEM of each bin.
(F) Box-and-whiskers plot shows the DSB densities of genes longer than 800kb located on chromosomes other than Chr1 in DMSO-treated or APH-treated NPCs. Min to max is plotted. Wilcoxon matched-pairs signed rank test, ****$p < 0.0001$.
(G) Box-and-whiskers plot shows the DSB densities of genes longer than 800kb or genes between 400kb and 800kb located on chromosomes other than Chr1 in APH-treated NPCs. Min to max is plotted. Mann-Whitney test, two-tailed, ****$p < 0.0001$, **$p < 0.01$.

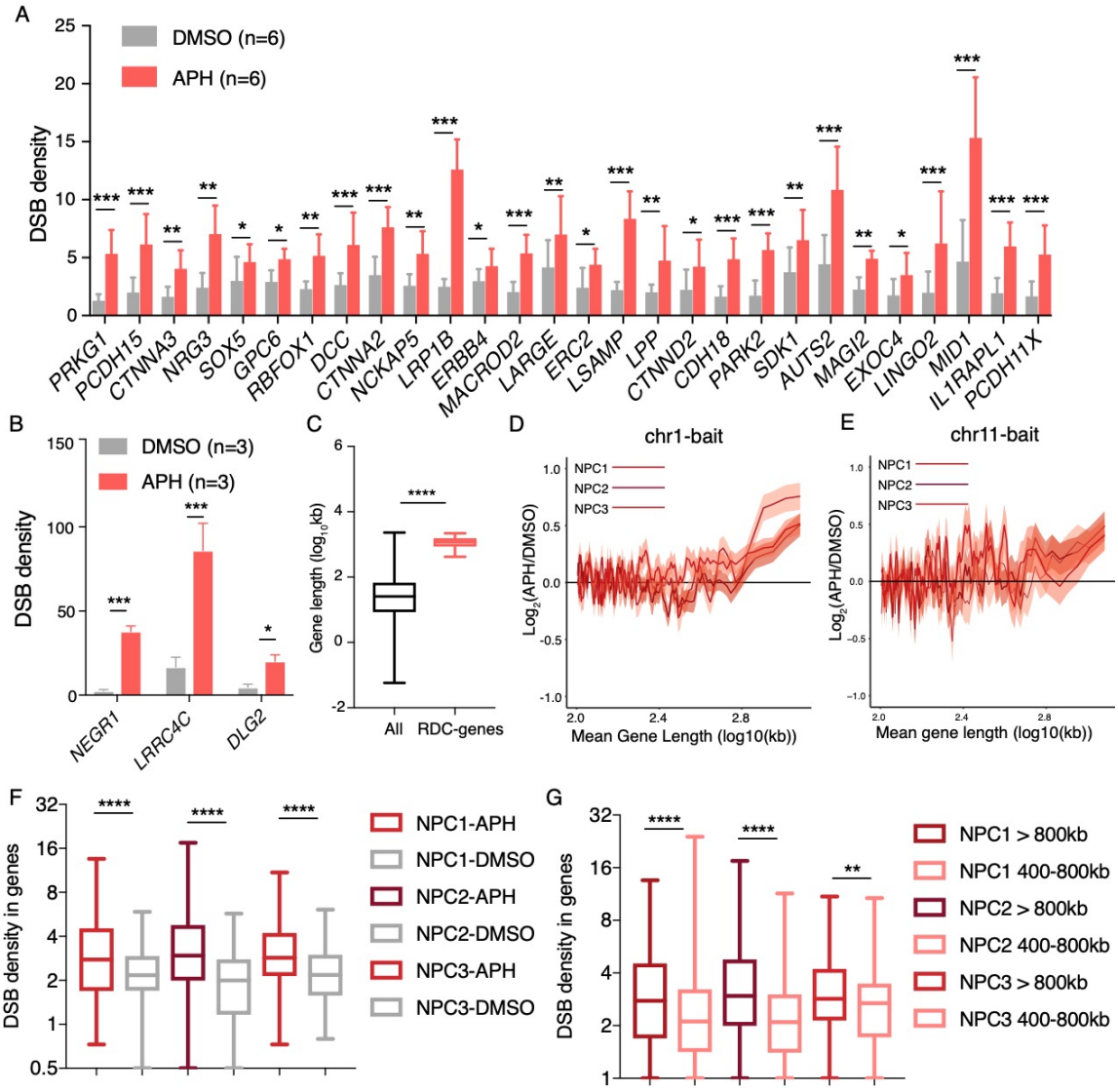

A
DMSO (n=6)
APH (n=6)
DSB density
PRKG1
PCDH15
CTNNA3
NRG3
SOX5
GPC6
RBFOX1
DCC
CTNNA2
NCKAP5
LRP1B
ERBB4
MACROD2
LARGE
ERC2
LSAMP
LPP
CTNND2
CDH18
PARK2
SDK1
AUTS2
MAGI2
EXOC4
LINGO2
MID1
IL1RAPL1
PCDH11X
B
DMSO (n=3)
APH (n=3)
DSB density
NEGR1
LRRC4C
DLG2
C
Gene length (log10 kb)
All
RDC-genes
D
chr1-bait
NPC1
NPC2
NPC3
Log2(APH/DMSO)
Mean Gene Length (log10(kb))
E
chr11-bait
NPC1
NPC2
NPC3
Log2(APH/DMSO)
Mean gene length (log10(kb))
F
DSB density in genes
NPC1-APH
NPC1-DMSO
NPC2-APH
NPC2-DMSO
NPC3-APH
NPC3-DMSO
G
DSB density in genes
NPC1 > 800kb
NPC1 400-800kb
NPC2 > 800kb
NPC2 400-800kb
NPC3 > 800kb
NPC3 400-800kb

Figure 1.5: Replication stress-induced DSB hotspots correlate with transcription.

(A) Smooth scatter plot shows the Spearman correlation of two GRO-seq replicates of hiPSC1-derived NPCs.
(B) DSB cluster between Chr1 bait DSB and prey DSBs within LPHN2 locus. Prey DSB junctions within the indicated region captured by Chr1 bait and Chr11 bait were shown (middle). RefGene (top) and GRO-seq (bottom) are shown (ordinate indicates normalized GRO-seq counts; reads are shown in plus [red] and minus [gray] orientations). Genomic region corresponding to the actively transcribed region of the gene detected by GRO-seq is highlighted in yellow.
Cells treated with DMSO (C) or 0.5 μM APH (D) for 24 h were incubated with 20 μM BrdU for 2 h. S-phase was divided to early, mid, and late according to DNA content.
(E) Bar plots show the fraction of cells in early, mid, and late S-phase in DMSO- or APH- treated NPCs. n = 3 technical replicates. Mean ± SD. Student's t test, two-tailed, ns, not significant, *** $p < 0.001$, **** $p < 0.0001$.
(F) Representative images of RNAPII alone, PCNA alone, and PCNA-RNAPII PLA foci per nucleus. RNAPII alone and PCNA alone are single antibody controls. Scale bar: 10 μm. Red, PLA foci; blue, DAPI.

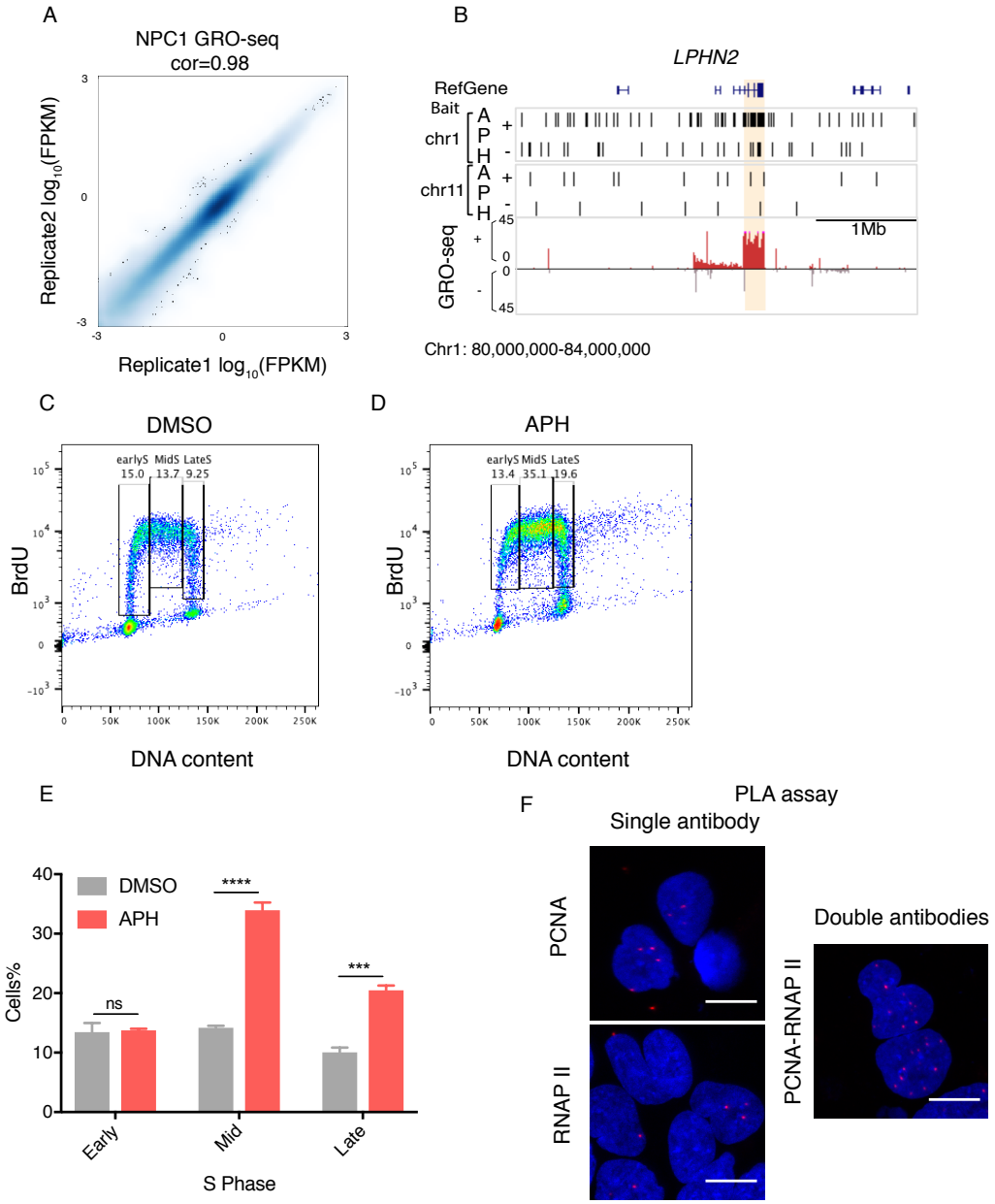
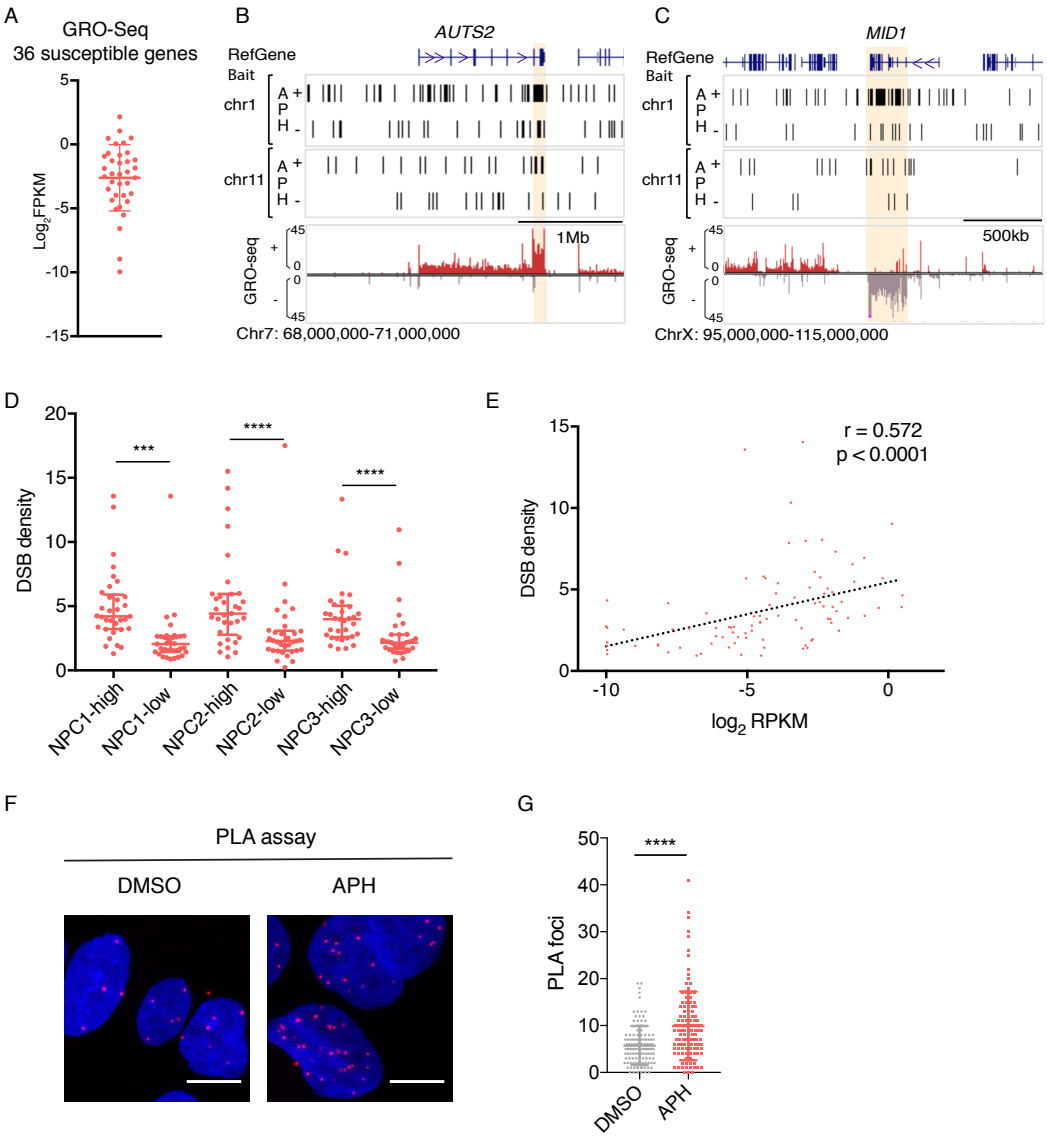

A
NPC1 GRO-seq
cor=0.98
Replicate2 log10(FPKM)
Replicate1 log10(FPKM)
B
LPHN2
RefGene
Bait
chr1
chr11
GRO-seq
1Mb
Chr1: 80,000,000-84,000,000
C
DMSO
earlyS MidS LateS
15.0 13.7 9.25
BrdU
DNA content
D
APH
earlyS MidS LateS
13.4 35.1 19.6
BrdU
DNA content
E
DMSO
APH
Cells%
ns
****
***
Early
Mid
Late
S Phase
F
PLA assay
Single antibody
PCNA
RNAP II
Double antibodies
PCNA-RNAP II

Figure 1.6: Replication stress induces replication-transcription conflicts.

(A) Scatter plot shows the average RPKM of two GRO-seq replicates of hiPSC1-derived NPCs of the 36 replication stress-susceptible genes on a log scale. Line represents mean and SD.
DSB cluster between Chr1 bait DSB and prey DSBs within *AUTS2* locus (B) and *MID1* locus (C). Prey DSB junctions within the indicated region captured by Chr1 bait and Chr11 bait are shown (middle). RefGene (top) and GRO-seq (bottom) are shown (ordinate indicates normalized GRO-seq counts; reads are shown in plus [red] and minus [gray] orientations). Genomic region corresponding to the actively transcribed region of the gene detected by GRO-seq is highlighted in yellow.
(D) Long genes (>800kb) were divided into three groups based on their expression level from GRO-seq: high (1-33%), medium (34-66%), and low (67-100%). Scatter plot shows the DSB densities of the high expression group and low expression group. Median ± interquartile range. Mann-Whitney test, two-tailed, *** $p < 0.001$, **** $p < 0.0001$.
(E) XY-plot shows the correlation of DSB density and expression (log_2RPKM) detected by GRO-seq. Genes longer than 800kb located on chromosomes other than Chr1 were plotted. Nonparametric Spearman correlation, correlation coefficient $r = 0.572$, two tailed, $P < 0.0001$.
(F) Images of NPCs treated with DMSO (left) or 0.25 μM APH (right) for 24 h. DAPI (blue), PCNA-RNAPII PLA (red), scale bar, 10 μm.
(G) Scatter plot shows the number of PLA foci of NPCs treated with DMSO (left) or 0.25 μM APH (right) for 24 h; $n > 100$ nuclei. Mean ± SD. Mann-Whitney test, two-tailed, **** $p < 0.0001$.


A
GRO-Seq
36 susceptible genes
Log2FPKM
5
0
-5
-10
-15
B
AUTS2
RefGene
Bait
chr1
chr11
A
P
H
GRO-seq
1Mb
Chr7: 68,000,000-71,000,000
C
MID1
RefGene
Bait
chr1
chr11
GRO-seq
500kb
ChrX: 95,000,000-115,000,000
D
DSB density
***
****
****
NPC1-high
NPC1-low
NPC2-high
NPC2-low
NPC3-high
NPC3-low
E
DSB density
r = 0.572
p < 0.0001
log2 RPKM
F
PLA assay
DMSO
APH
G
PLA foci
****
DMSO
APH

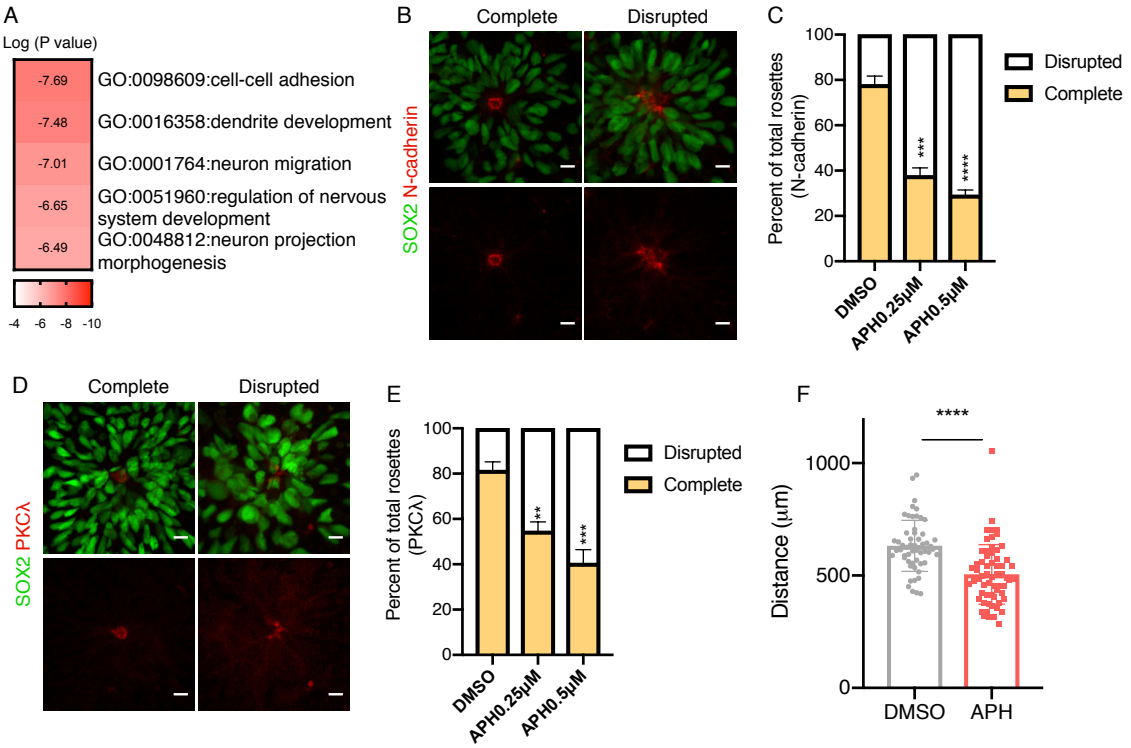


Figure 1.7: Aberrant adherens junction, apical polarity, and NPC migration of NPCs upon replication stress.

(A) Heatmap shows the P value on a log scale of selected top GO terms.
(B) Defects in adherens junctions of Hues6-derived NPCs treated with APH for 48 h. Sample confocal images of immunostaining of N-cadherin for neural rosettes are shown. Scale bar, 10 μm.
(C) Quantification of neural rosettes with complete or scattered N-cadherin in Hues6-derived NPCs treated with DMSO, 0.25 μM APH, or 0.5 μM APH for 48 h. n = 3 cultures. Mean ± SD. Student's t test, two-tailed, **** $p < 0.0001$, *** $p < 0.001$.
(D) Defects in atypical PKCλ of Hues6-derived NPCs treated with APH for 48 h. Sample confocal images of immunostaining of PKCλ for neural rosettes are shown. Scale bar, 10 μm.
(E) Quantification of neural rosettes with complete or disrupted PKCλ in Hues6-derived NPCs treated with DMSO, 0.25 μM APH, or 0.5 μM APH for 48 h. n = 3 cultures. Mean ± SD. Student's t test, two-tailed, *** $p < 0.001$, ** $p < 0.01$.
(F) Quantification of cell migration from neurospheres treated with DMSO or 0.25 μM APH for 48 h. Each point represents one neurosphere. Student's t test, two-tailed, **** $p < 0.0001$.

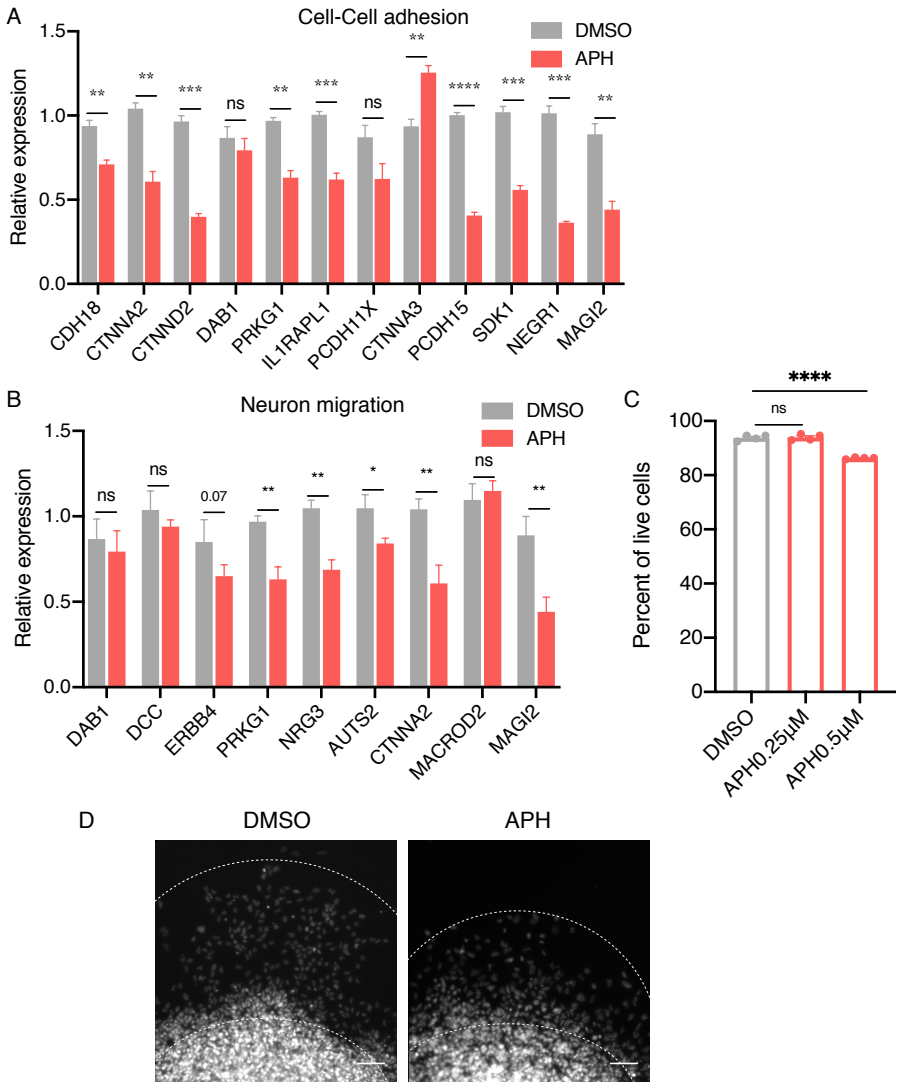


Figure 1.8: Aberrant gene expression and NPC migration of NPCs upon replication stress.

(A) qPCR analysis of expression of susceptible genes implicated in cell-cell adhesion (GO: 0098609) of Hues6-derived NPCs treated with DMSO or 0.5 μM APH for 48 h. n=3. Mean ± SD. Student's t test, two-tailed, *** $p < 0.001$, ** $p < 0.01$, ns, not significant.
(B) qPCR analysis of expression of susceptible genes implicated in neuron migration (GO: 0001764) of Hues6-derived NPCs treated with DMSO or 0.5 μM APH for 48 h. n=3. Mean ± SD. Student's t test, two-tailed, ** $p < 0.01$, * $p < 0.05$, ns, not significant.
(C) Percentage of live cells of Hues6-derived NPCs treated with DMSO, 0.25 μM APH, or 0.5 μM APH for 48 h. n=4. Mean ± SD. Student's t test, two-tailed, **** $p < 0.0001$, ns, not significant.
(D) Representative images of neurospheres treated with DMSO (left) or 0.25 μM APH (right) for 48 h. Scale bar: 100 μm.

Table 1.1: Genomic coordinates of the DSB hotspots in low complexity/repeat regions, Chr1-sgRNA OT sites, and Chr11-sgRNA OT sites identified by HTGTS.

Low complexity/ repeat region		
Chromosome	Start	End
chr2	89820000	89879999
chr4	49080000	49169999
chr4	49620000	49679999
Chr10	37128000	37157999
Chr10	38760000	39179999
Chr10	42330000	42629999

Chr1-sgRNA		
On-target	Off-target (identified by HTGTS)	Off-target locus
GTCCGGGGCCTCT CTAGTCC AGG	ATACGGGGTCTCTCTAGTCA AGG	5 + 92908024 92908043
	GCCCTGGGCCTCTCTGGTCC TGG	8 - 103669535 103669554
	ACCCGGGGCCTCCCTATTCC AGG	16 + 9058754 9058773
	GTCCTGGGCCTCCCTAGCCC AGG	22 - 29656046 29656065
	GCCCAGGGCCCCTCCAGTCC AGG	15 - 72448424 72448443
	CTCCGGGCCCTCCCTAGCCC CGG	16 + 850622 850641

Chr11-sgRNA		
On-target	Off-target (identified by HTGTS)	Off-target locus
GACTTGTTTTCAT TGTTCTC AGG	CATTTGTTTTCATTGTTCTC TGG	4 + 120560949 120560968
	TCCTTGTTTTCATTGTTCTC TGG	14 + 89055053 89055072

Table 1.2: HTGTS junction enrichment within identified regions.

Associated gene	P-value (if detected)					
	NPC1-Chr1bait	NPC2-Chr1bait	NPC3-Chr1bait	NPC1-Chr11bait	NPC2-Chr11bait	NPC3-Chr11bait
DAB1	1.94E-13	3.55E-13	3.50E-09	9.75E-10	8.40E-07	9.03E-05
NEGR1	1.07E-109	4.14E-131	6.02E-157	2.45E-15	6.67E-12	9.03E-05
LPHN2	9.79E-16	6.32E-24	1.45E-02		5.47E-09	7.99E-08
PRKG1	5.83E-08	1.21E-08			7.74E-04	3.84E-15
PCDH15	1.27E-27	3.27E-34		3.42E-13	2.18E-05	
CTNNA3		4.95E-22	2.41E-06		2.35E-06	
NRG3	9.14E-20	4.89E-23		1.32E-05		
NAV2				5.89E-03	5.59E-10	7.39E-04
LRRC4C	2.68E-22		5.66E-09	3.05E-43	2.62E-49	1.65E-43
DLG2		3.92E-10	5.71E-11	1.47E-13	8.76E-13	1.51E-34
SOX5		4.43E-04	3.92E-03	1.69E-06	2.35E-06	5.94E-02
GPC6		3.29E-06	5.51E-03			
MDGA2	4.13E-10	4.43E-04	2.41E-06	9.04E-06		
RBFOX1		8.39E-08	1.71E-08			
DCC	2.77E-12	1.68E-07			7.97E-12	
CTNNA2	1.69E-39	5.05E-22	7.23E-23	1.12E-03	7.74E-04	
NCKAP5	1.52E-08	8.77E-13		5.91E-03	5.47E-09	
LRP1B	1.01E-70	1.78E-92	1.13E-30	1.32E-24	5.58E-11	8.13E-27
ERBB4	2.40E-03	1.84E-05		4.48E-10		
MACROD2	1.95E-08	2.74E-17	5.91E-10		2.35E-06	
LARGE	1.90E-04	2.31E-15				
ERC2		1.25E-03	6.32E-04			
LSAMP	3.05E-45	3.02E-48	1.46E-14	1.13E-15	4.57E-07	
LPP	3.84E-04	4.43E-04			3.12E-06	5.40E-04
CTNND2	4.10E-04	1.92E-07			7.97E-12	
CDH18	2.77E-12	1.21E-08				
*Intergenic**	2.08E-18	5.71E-83	1.94E-21	7.70E-04	8.19E-08	1.28E-10
PARK2			5.71E-09	1.79E-16		
SDK1	3.84E-04	2.08E-10	8.32E-12		2.20E-10	
AUTS2	5.88E-07	2.14E-39	5.78E-20	3.41E-04	4.93E-06	7.99E-08
MAGI2	6.85E-07	1.76E-18	6.09E-18	1.56E-02		
EXOC4	8.88E-05	1.94E-11	5.66E-09			
CSMD1	4.10E-04	7.22E-09		1.32E-05	3.84E-03	
LINGO2		3.13E-10	3.37E-21	1.67E-08	1.21E-07	
MID1	8.04E-14	1.94E-11	1.57E-02	1.14E-04	2.21E-16	
IL1RAPL1	9.28E-22	1.71E-21	8.32E-12	5.97E-06		5.40E-04
PCDH11X	9.73E-13	1.88E-03	5.78E-20	3.41E-04		

Table 1.3: Genomic coordinates of the DSB hotspots and associated genes.

DSB hotspots	chr	Start	End	Associated gene	Gene coordinates (hg19)
1	Chr1	57210000	58169999	*DAB1*	Chr1:57460451-59012406
2	Chr1	71670000	72959999	*NEGR1*	Chr1:71861623-72748417
3	Chr1	82260000	82649999	*LPHN2*	Chr1:81771845-82458120
4	Chr10	53280000	54179999	*PRKG1*	Chr10:52750945-54058110
5	Chr10	56160000	56969999	*PCDH15*	Chr10:55562531-57387702
6	Chr10	68220000	68609999	*CTNNA3*	Chr10:67672276-69455927
7	Chr10	83880000	84449999	*NRG3*	Chr10:83635070-84746935
8	Chr11	19320000	20369999	*NAV2*	Chr11:19372271-20143144
9	Chr11	39900000	42029999	*LRRC4C*	Chr11:40135753-41481323
10	Chr11	83160000	84869999	*DLG2*	Chr11:83166055-85338966
11	Chr12	24390000	24839999	*SOX5*	Chr12:23682438-24715524
12	Chr13	94830000	95039999	*GPC6*	Chr13:93879095-95059655
13	Chr14	47610000	48359999	*MDGA2*	Chr14:47308826-48144157
14	Chr16	6750001	7229999	*RBFOX1*	Chr16:6069095-7763340
15	Chr18	50640000	51029999	*DCC*	Chr18:49866542-51057784
16	chr2	80100000	80609999	*CTNNA2*	chr2:79412357-80875905
17	chr2	133290000	133829999	*NCKAP5*	chr2:133429374-134326034
18	chr2	141420000	142559999	*LRP1B*	chr2:140988992-142889270
19	chr2	212190000	212429999	*ERBB4*	chr2:212240446-213403565
20	chr20	14580000	15599999	*MACROD2*	chr20:13976015-16033842
21	chr22	33690000	34379999	*LARGE*	chr22:33558212-34318829
22	chr3	55530000	55739999	*ERC2*	chr3:55542336-56502391
23	chr3	116100000	117389999	*LSAMP*	chr3:115521235-117716095
24	chr3	187830000	188369999	*LPP*	chr3:187871072-188608460
25	chr5	10860000	11339999	*CTNND2*	chr5:10971952-11904155
26	chr5	19950000	20129999	*CDH18*	chr5:19473060-20575982
27	chr5	164280000	165359999	*Intergenic**	
28	chr6	162210000	162899999	*PARK2*	chr6:161768452-163148803
29	chr7	4080000	4349999	*SDK1*	chr7:3341080-4308632
30	chr7	69810001	70619999	*AUTS2*	chr7:69063905-70258054
31	chr7	78030000	78449999	*MAGI2*	chr7:77646393-79082890
32	chr7	133110000	133319999	*EXOC4*	chr7:132937829-133751342
33	chr8	3690000	4379999	*CSMD1*	chr8:2792875-4852494
34	chr9	28140000	29219999	*LINGO2*	chr9:27948076-28670283
35	chrX	10320000	10739999	*MID1*	chrX:10413350-10851773
36	chrX	28860000	29549999	*IL1RAPL1*	chrX:28605516-29974840
37	chrX	91260000	91709999	*PCDH11X*	chrX:91034260-91878229

Chapter 2 NPCs derived from autistic patients with macrocephaly have increased replication stress and chronic DNA damage

Introduction

Somatic mutations are an important cause of brain overgrowth and neuronal migration disorders (Lim et al., 2015, Lee et al., 2012, Poduri et al., 2012). Moreover, focal patches of abnormal laminar cytoarchitecture and cortical disorganization of neurons in the prefrontal and temporal cortical tissue of children with autism have been reported (Stoner et al., 2014), suggesting a potential contribution of somatic mutations to ASD. Despite growing evidence that NPCs derived from ASD patients with macrocephaly undergo rapid cell cycle progression (Marchetto et al., 2017, Mariani et al., 2015), little is known about how perturbed cellular proliferation affects genome stability in NPCs and its contribution to neurodevelopmental disorders such as ASD. In this chapter, we demonstrate that NPCs derived from iPSCs reprogrammed from fibroblasts of ASD patients with macrocephaly exhibited accelerated S-phase progression, increased replication stress, and chronic DNA damage compared to NPCs derived from control subjects.

ASD-derived NPCs have perturbed S-phase progression

A previous study demonstrated that NPCs derived from iPSCs reprogrammed from fibroblasts of ASD subjects with macroscopic early brain overgrowth (Table 2.1) displayed rapid cellular proliferation (Marchetto et al., 2017). Exome sequencing revealed damaging mutations in genes in the canonical Wnt pathway, cell cycle regulation, mitotic checkpoints, and DNA repair in ASD subjects, including genes central

to maintaining genome stability (e.g., *ATM*, *BRCA1*, *CDK7*, and *ERCC4*) and several components of the anaphase-promoting complex/cyclosome (e.g., *ANAPC1* and *CDC27*) (Marchetto et al., 2017). Mutations that attenuate mitotic checkpoints and DNA repair promote transmission of errors that occur during DNA replication to daughter cells. Quantification of Ki-67$^+$ cells indicated an increased percentage of proliferating cells in ASD-derived NPCs (Figure 2.1A and 2.1B). To explore whether ASD-derived NPCs had an altered replication program, asynchronously growing NPCs were pulse-labeled with 5-bromo-2'-deoxyuridine (BrdU) for 30 min and then chased for 3 hours in fresh media before fluorescence-activated cell sorting (FACS) analysis. ASD-derived NPCs exhibited a greater fraction of BrdU$^+$ cells that reached 4N DNA content 3 hours following the labeling (Figure 2.1C, 2.2A and 2.2B). To validate this finding, the asynchronously growing NPCs were pulse-labeled with CldU for 30 min, chased for 3 hours, and then pulse-labeled with IdU for 30 min (Figure 2.1D and 2.2C). The ASD-derived NPCs progressed through S-phase much faster than the control NPCs (Figure 2.1E and 2.1F).

To determine whether accelerated S-phase progression altered the replication program, we examined the spatiotemporal pattern of replication factories. Analysis of the replication program showed similar proportions of cells in early, mid, and late S-phase in ASD-derived NPCs and control NPCs (Figure 2.2D and 2.2E). We analyzed IdU-labeled replication foci and found that ASD-derived NPCs exhibited increased intensity of individual foci compared to control NPCs (Figure 2.1G), indicating a higher number of replication forks within each focus. Collectively, these data reveal that ASD-derived NPCs displayed accelerated S-phase progression and altered nuclear organization of replication factories.

ASD-derived NPCs exhibit replication stress, chronic activation of the ATR-CHK1 pathway, and elevated DNA damage

Rapid S phase progression may cause perturbation of DNA replication. To directly test this possibility, we pulse-labeled NPCs with BrdU for 30 min to identify replicating regions and then carried out a DNA combing assay to examine the length of BrdU tracks (Figure 2.3A). The analysis revealed a greater reduction in replicative DNA fiber length in ASD-derived NPCs compared to control NPCs (Figure 2.3A and 2.3B). To identify stalled or collapsed forks, we performed a DNA combing assay to monitor the progression of replicating forks, during which replicating DNA was pulse-labeled with 5-Iodo-2'-deoxyuridine (IdU) first and then with 5-Chloro-2'-deoxyuridine (CldU) (Figure 2.3C). We then analyzed fork symmetry between the first and second pulse in IdU/CldU dual-labeled DNA fibers. Analysis revealed asymmetric fork progression in ASD-derived NPCs (Figure 2.3D). To estimate origin firing, we measured the distance between origins of replication (Figure 2.4A). A reduction in origin-to-origin distance was observed in ASD-derived NPCs (Figure 2.3E), indicating increased fork density. Collectively, these data point to DNA replication stress in ASD-derived NPCs.

Replication stress induces fork stalling and promotes genome instability (Cimprich and Cortez, 2008). Mild treatment of NPCs with APH dramatically induced DNA damage detected by the phosphorylation of histone H2AX on serine 139 (γH2AX) in control NPCs (Figure 2.4C and 2.4D). We then asked whether ASD-derived NPCs also showed elevated DNA damage. We observed a significant increase in the percentage of cells with three or more γH2AX foci in ASD-derived cells compared to control NPCs (Figure 2.3F and 2.3G) without an apparent increase in cell death (Figure 2.4B). We next

characterized the key molecular events induced by DNA damage, i.e., the activation of the ATM and ATR pathways, which are primarily triggered by DSBs and replication stress, respectively. ASD-derived NPCs showed increased CHK1 phosphorylation at Ser345 and RPA32 phosphorylation at Ser 33 (Figure 2.4E and 2.4F), a hallmark for activation of the ATR pathway. ATR and CHK1 kinases are key for the response to replication stress and are essential for cell viability (Cimprich and Cortez, 2008). In contrast, the ATM pathway was not activated in ASD-derived cells, as indicated by the lack of ATM phosphorylation at Ser1981 and RPA32 phosphorylation at Ser 4/8 (Figure 2.4E and 2.4F). These observations are consistent with the finding that replication stress induces head-on transcription-replication conflicts that lead to fork stalling and selectively activate the ATR-CHK1 pathway but not the ATM pathway (Hamperl et al., 2017). We hypothesized that perturbation of DNA replication would induce increased sensitivity to external replication stress in ASD-derived NPCs. Indeed, ASD-derived NPCs exhibited elevated DNA damage upon mild treatment with APH (Figure 2.4G). These findings suggest that ASD-derived NPCs had increased replication stress that led to chronic activation of the ATR-CHK1 pathway and elevated DNA damage.

ASD-derived NPCs have more DNA damage in long genes and exhibit aberrant apical polarity and migration

Remarkably, 20 susceptible genes were found in the SFARI Gene dataset (https://www.sfari.org/resource/sfari-gene/) that consisted of genes implicated in ASD (Figure 2.5A and Table 2.2) and 19 susceptible genes were located within CNV modules of SFARI Gene (Table 2.3). We next sought to ascertain, using PLA, whether ASD-derived NPCs had increased conflicts of replication fork and transcription machineries.

We found increased PLA foci in ASD-derived NPCs compared to control NPCs (Figure 2.5B and 2.6A). The observation that replication stress induced increased conflicts of replication fork and transcription machineries (Figure 1.6G) corroborates the observation that ASD-derived NPCs with altered replication program had increased PCNA-RNAPII PLA foci (Figure 2.5B), leading to a hypothesis that ASD-derived NPCs harbor more DNA damage in replication stress-susceptible genes. To investigate whether the altered S-phase progression and replication program in ASD-derived NPCs induced DNA damage in long genes, we interrogated DSB sites by HTGTS. We observed a small, yet significant increase in DSB densities in the long genes (genes longer than 800kb+ located on chromosomes other than Chr1) in ASD-derived NPCs (Figure 2.5C). This difference was more significant in replication stress-susceptible genes (33 susceptible genes located on chromosomes other than Chr1) (Figure 2.5D and Table 1.3).

Increased replication stress in neural culture led to gene expression changes and defects in neural function. To determine whether replication stress and increased DNA damage in replication stress-susceptible genes attenuated gene expression in ASD-derived NPCs, we carried out qPCR analysis of genes involved in cell-cell adhesion and neuron migration. Consistently, ASD-derived NPCs showed decreased expression of several genes in these pathways (Figure 2.6B and 2.6C), including *AUTS2*, which has been shown to regulate neuronal migration (Hori et al., 2014). To determine whether ASD-derived cells exhibited defects in apical polarity and adherens junctions, we generated neural rosettes using a monolayer differentiation protocol (Shi et al., 2012). A greater percentage of disrupted neural rosettes, as evidenced by both N-cadherin and PKCλ expression, were observed in ASD-derived cultures (Figure 2.5E and 2.5F).

Moreover, neurosphere migration assay also revealed slightly decreased cell migration in ASD-derived NPCs (Figure 2.5G). Taken together, our findings suggest that replication stress in ASD-derived NPCs induces elevated DSBs in long ASD genes and leads to expression and function defects, providing a novel mechanistic link between an abnormal replication program and defects related to ASD risk.

Methods

FACS analysis

Cell cycle analysis. For BrdU pulse-chase experiment, cells were pulse-labeled with 20 μM BrdU for 30 min, collected immediately or chased for 3 h in fresh media. For cell cycle analysis, cells were pulse-labeled with 20 μM BrdU for 2 h and harvested. Cells were fixed in 70% ice cold ethanol for at least 30 min, permeabilized with 0.1% Triton X-100, and treated with 2N HCl for 30 min prior to BrdU antibody (BioLegend) labeling. Cells were washed with PBS and treated with 20 μg/ml RnaseA (Life Technologies) and DNA was stained with 20 μg/ml propidium iodide (Invitrogen). Cells were analyzed on LSR II (Becton Dickinson) and acquired data were analyzed using FlowJo software.

Cell viability assay. Cells were collected, washed in cold PBS, and stained with propidium iodide (Invitrogen). Cells were analyzed on LSR II (Becton Dickinson) and acquired data were analyzed using FlowJo software. Live cells were defined as cells negative for propidium iodide staining.

DNA combing assay

To measure fork speed, cells were pulse-labeled with 20 μM BrdU (Sigma-Aldrich) for 30 min and collected. To measure fork symmetry and estimate fork density,

cells were sequentially labeled with 25 μM IdU (Sigma-Aldrich) for 20 min and then with 100 μM CldU for 20 min. DNA fiber spreads were prepared as previously described (Marechal et al., 2014). Briefly, 2 μl of cell suspension was spotted onto a cleaned glass slide and lysed with 7 μl of lysis buffer (50mM EDTA, 0.5% SDS, and 200mM Tris-HCl pH 7.5). Slides were tilted to allow DNA to spread slowly down the slide, followed by air-drying and fixation in methanol/acetic acid (3:1) for 10 min. The DNA spread was then denatured in 2.5 M HCl for 80 min and then blocked with 5% BSA in PBS for 30 min. Mouse anti-BrdU antibody (BD Biosciences, 347580) was used to detect IdU, and rat anti-BrdU antibody (Accurate, OBT0030) was used to detect BrdU or CldU. SsDNA antibody (Enzo life Sciences, F7-26) was used to detect single-stranded DNA. Antibodies were diluted in blocking solution and applied to the slides followed by incubation in a humidified chamber for overnight at 4°C. After three washes with PBS, secondary antibodies were applied for 1 h at room temperature. The slides were washed and then mounted with ProLong™ Gold Antifade Mountant (Life Technologies). Images of well-spread DNA fibers were acquired using Zeiss LSM 710 or LSM 880 Laser Scanning Confocal Microscope (Carl Zeiss) and measured using the ImageJ software (NIH).

Immunofluorescence

Cells cultured on slides were fixed in 4% paraformaldehyde (10 min, room temperature), permeabilized with 0.1% Triton X-100. For the CldU/ IdU experiment, cells were treated with 2N HCl for 30 min prior to blocking. Cells were incubated with primary antibodies overnight at 4 °C. After wash, cells were incubated with secondary antibodies for 1 h at room temperature, washed, incubated with DAPI for 10 min, and mounted using ProLong™ Gold Antifade Mountant (Life Technologies). Image

acquisition was performed using Zeiss LSM 710 or LSM 880 Laser Scanning Confocal Microscope or Zeiss LSM 880 Laser Scanning Confocal Microscope (Carl Zeiss). IdU foci intensity was analyzed using unprocessed 0.31μm stacks with the FociPicker3D algorithm (Du et al., 2011).

Acknowledgements

Chapter 2, in full, has been submitted for publication of the material as it may appear in Cell Stem Cell, Wang, Meiyan; Wei, Pei-Chi; Lim, Christina; Gallina, Iryna; Marshall, Sara; Marchetto, Carol; Alt, Frederick; Gage, Fred, Elsevier Press, 2019. The dissertation author was the primary investigator and author of this paper.

Figure 2.1: ASD-derived NPCs display accelerated S-phase progression and increased DNA damage.

(A) Images of NPCs derived from control (left) or ASD (right) subjects. Ki-67 (white), DAPI (blue), NESTIN (red). Arrows indicate Ki-67 negative cells. Scale bar, 50 μm.
(B) Bar plot shows the quantification of percentage of Ki-67+ cells. Each point represents one cell line. Average of six randomly selected 20x images per line. Mean ± SD. Student's t test, two-tailed, ** $p < 0.01$.
(C) NPCs were pulse-labeled with 20 μM BrdU for 30 min, collected immediately or chased in fresh media for 3 h before collecting. Percentage of BrdU+ cells reaching 4N DNA content was quantified. Mean ± SD. Student's t test, two-tailed, * $p < 0.05$.
(D) NPCs were pulse-labeled with 20 μM CIdU (green) for 30 min, chased in fresh media for 3 h, and pulse-labeled with 20 μM IdU (red) for 30 min. S phase progression was determined by the characteristic replication foci detected by IdU or CldU staining. Representative images of S phase progression: top, early; middle, mid; bottom, late; scale bar, 10 μm. Transition of replication patterns was classified as early-early (cells that remained in early S-phase during the experiment), early-mid (cells that progressed from early to mid S-phase during the experiment), and early-late/exit (cells that progressed through S-phase).
(E) Percentage of early-early cells. $n > 100$ nuclei per line. Mean ± SD. Student's t test, two-tailed, * $p < 0.05$.
(F) Percentage of early-late/exit cells. $n > 100$ nuclei per line. Mean ± SD. Student's t test, two-tailed, ** $p < 0.01$.
(G) Foci intensity of IdU-labeled cells. $n > 100$ nuclei per line. Average of each cell line was used for statistical test. Student's t test, two-tailed, * $p < 0.05$.

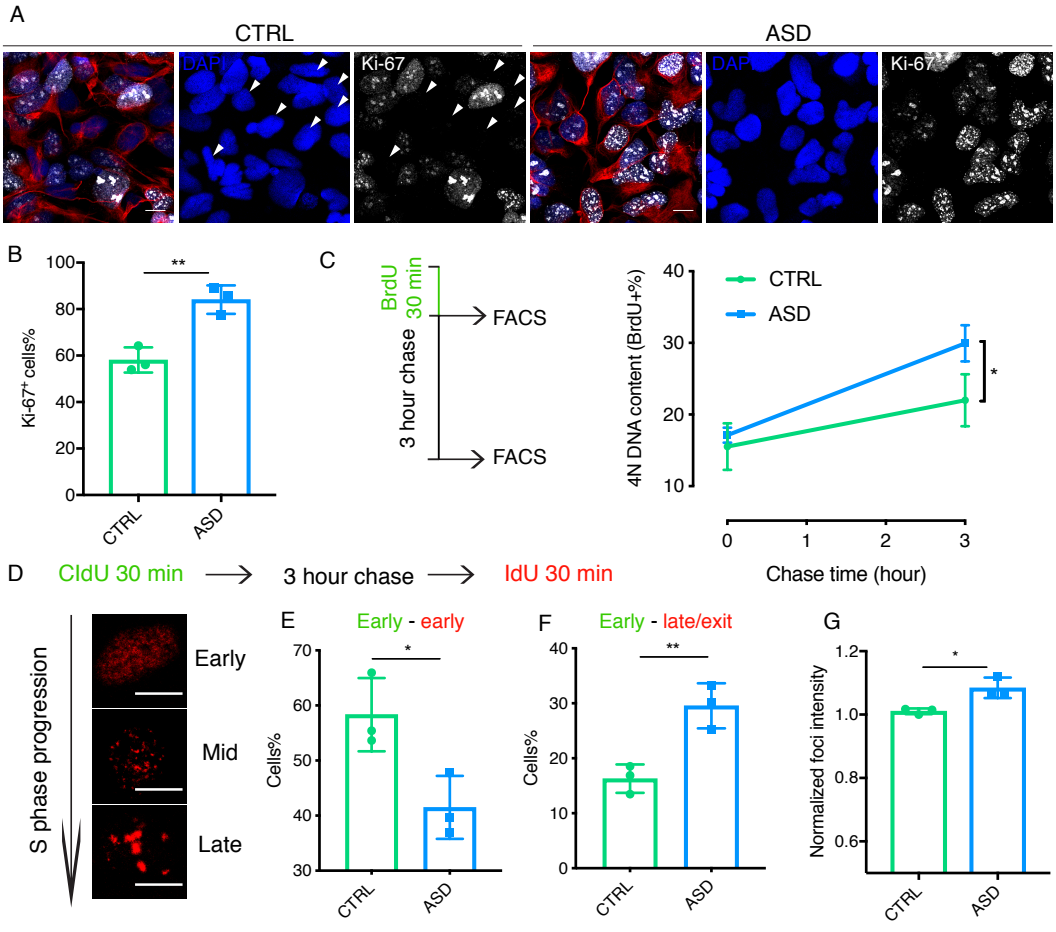

A
CTRL
ASD
DAPI
Ki-67
B
Ki-67+ cells%
CTRL
ASD
C
BrdU 30 min
3 hour chase
FACS
FACS
4N DNA content (BrdU+%)
CTRL
ASD
Chase time (hour)
D
CldU 30 min
3 hour chase
IdU 30 min
S phase progression
Early
Mid
Late
E
Early - early
Cells%
F
Early - late/exit
Cells%
G
Normalized foci intensity

Figure 2.2: ASD-derived NPCs display rapid cell cycle progression.

(A) CTRL-derived NPCs (top) and ASD-derived NPCs (bottom) were pulse-labeled with 20 μM BrdU for 30 min, and collected immediately (left) or chased in fresh media for 3 h (right). $BrdU^+$ cells and late S-phase cells are indicated. Fraction of replicating cells in late S-phase is calculated as (late S-phase cells/BrdU+ cells)%. X-axis, DNA content, y-axis, BrdU.
(B) Bar plot shows the rate of late S-phase accumulation calculated as differences (3h – 0h) of fraction of proliferating cells in late S-phase. Each point represents one line. Mean ± SD. Student's t test, two-tailed, ** $p < 0.01$.
(C) Cells were pulse-labeled with 20 μM CldU (green) for 30 min, chased in fresh media for 3 h, and pulse-labeled with 20 μM IdU (red) for 30 min. Transition of replication patterns was classified as early-early (cells that remained in early S-phase during the experiment), early-mid (cells that progressed from early to mid S-phase during the experiment), and early-late/exit (cells that progressed through S-phase). Representative images of early-early (top) and early-late (bottom). DAPI (blue), CldU (green), IdU (red). Scale bar, 10 μm.
(D) Percentage of cells in early, mid, or late S-phase. Each point represents one line. Mean ± SD. Student's t test, two-tailed, ns, not significant.
(E) Representative cell cycle FACS plot. Control NPCs (left) and ASD NPCs (right) were pulse-labeled with 20 μM BrdU for 2 h and collected. Cells in G1, early S, mid S, late S, and G2/M are indicated. X-axis, DNA content, y-axis, BrdU.

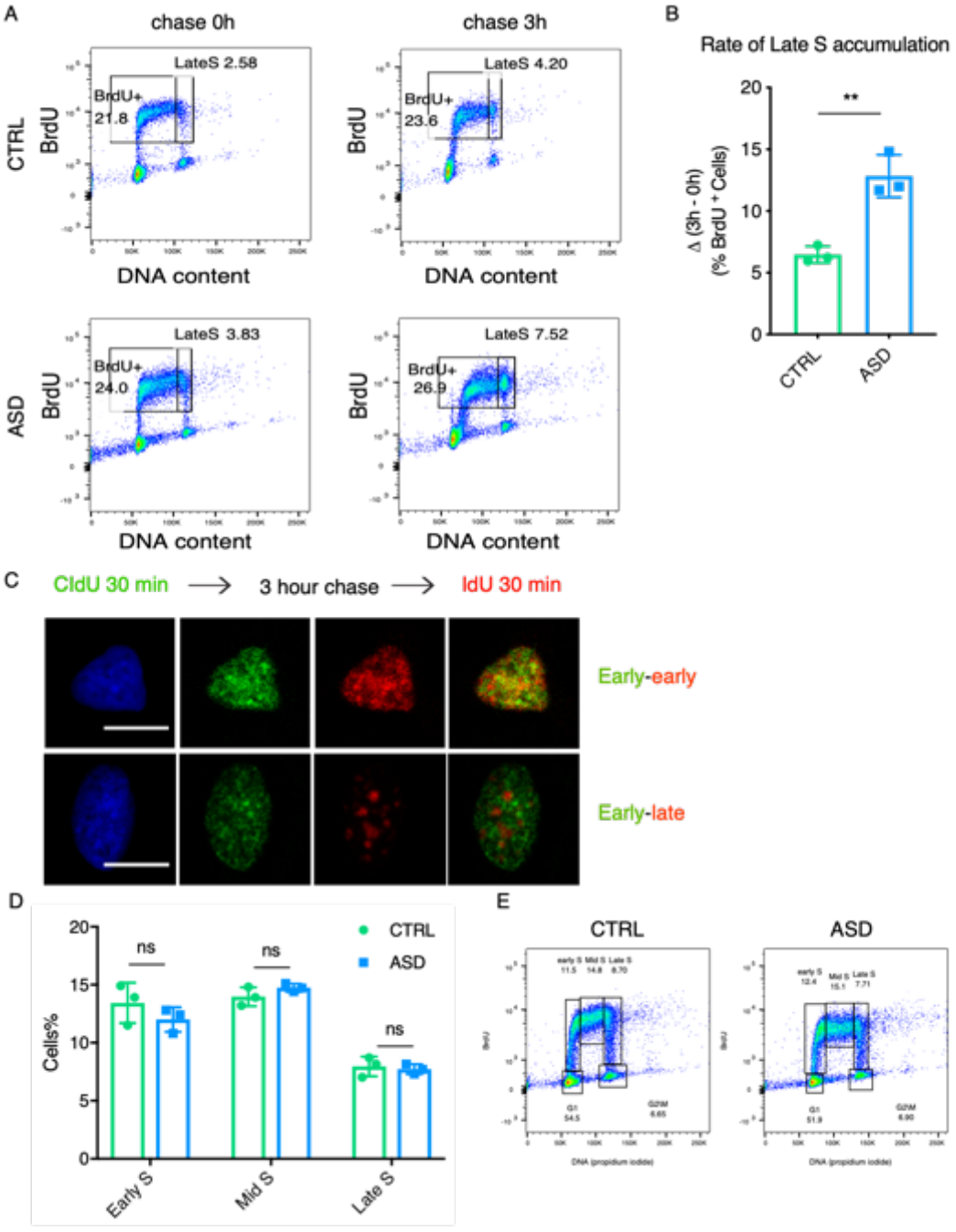

A
chase 0h
chase 3h
CTRL
ASD
BrdU
DNA content
LateS 2.58
BrdU+
21.8
LateS 4.20
23.6
LateS 3.83
24.0
LateS 7.52
26.9
B
Rate of Late S accumulation
**
Δ (3h - 0h)
(% BrdU + Cells)
CTRL
ASD
C
CldU 30 min
3 hour chase
IdU 30 min
Early-early
Early-late
D
Cells%
CTRL
ASD
ns
Early S
Mid S
Late S
E
CTRL
ASD

Figure 2.3: ASD-derived NPCs display increased replication stress and chronic DNA damage.

(A) Representative images of replicating DNA labeled with BrdU by DNA combing assay. BrdU (green). Top, control NPC. Bottom, ASD NPC. Scale bar, 10 μm.
(B) Scatter plot shows the DNA fiber length. Line represents average DNA fiber length. Each bar represents one cell line. Average of each cell line was used for statistical test. Student's t test, two-tailed, ** $p < 0.01$.
(C) Representative images of IdU/CldU dual-labeled DNA fibers. IdU (red), CldU (green). Top, control NPC. Bottom, ASD NPC. Scale bar, 10 μm.
(D) Scatter plot shows the fork symmetry calculated by IdU/CldU ratio. Line represents average IdU/CldU ratio. Each bar represents one cell line. Average of each cell line was used for statistical test. Student's t test, two-tailed, * $p < 0.05$.
(E) Scatter plot shows the ori-ori distance. Line represents average ori-ori distance. Each bar represents one cell line. Average of each cell line was used for statistical test. Student's t test, two-tailed, * $p < 0.05$.
(F) Representative images of γH2AX staining. Top, control NPC. Bottom, ASD NPC. Blue, DAPI. Red, γH2AX. Arrows indicate cells with three or more γH2AX foci. Scale bar, 10 μm.
(G) Bar plot shows the percentage of cells with three or more γH2AX foci. Each dot represents one cell line. Mean ± SD. Average of each cell line was used for statistical test. Student's t test, two-tailed, ** $p < 0.01$.

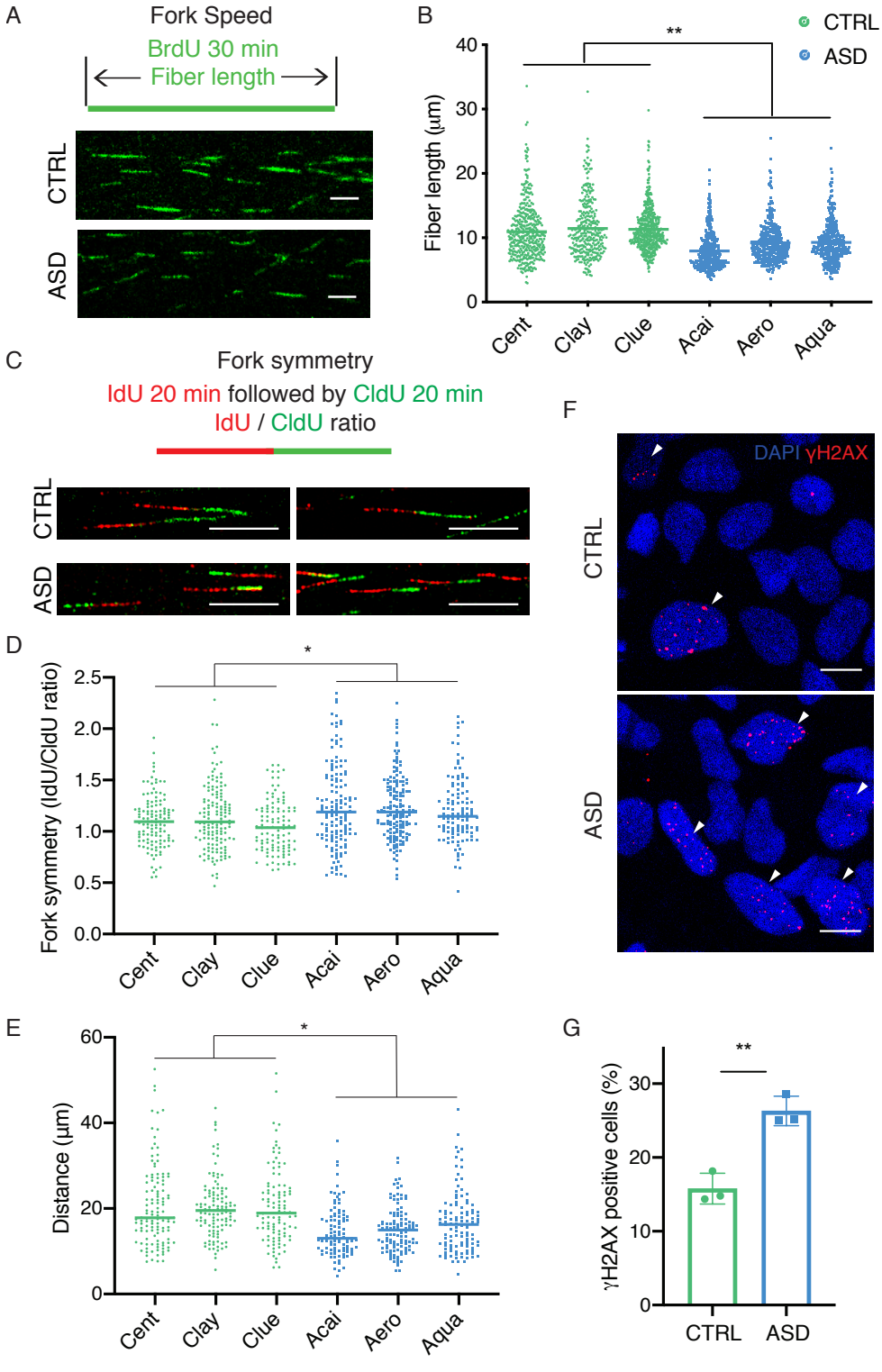

A
Fork Speed
BrdU 30 min
Fiber length
CTRL
ASD
B
CTRL
ASD
**
Fiber length (µm)
40
30
20
10
0
Cent
Clay
Clue
Acai
Aero
Aqua
C
Fork symmetry
IdU 20 min followed by CldU 20 min
IdU / CldU ratio
CTRL
ASD
D
*
Fork symmetry (IdU/CldU ratio)
2.5
2.0
1.5
1.0
0.5
0.0
Cent
Clay
Clue
Acai
Aero
Aqua
E
*
Distance (µm)
60
40
20
0
Cent
Clay
Clue
Acai
Aero
Aqua
F
DAPI γH2AX
CTRL
ASD
G
**
γH2AX positive cells (%)
30
20
10
0
CTRL
ASD

Figure 2.4: ASD-derived NPCs exhibit replication stress, chronic activation of the ATR-CHK1 pathway, and elevated DNA damage.

(A) Representative images of replicating DNA labeled with IdU for 20 min by DNA combing assay. Green, IdU. White, single-stranded DNA (ssDNA). Top, control NPC. Bottom, ASD NPC. Scale bar, 10 μm.
(B) Bar plots shows the percentage of live cells from each group. Each point represents one line. Student's t test, two-tailed, not significant.
(C) Representative images of γH2AX staining in control NPCs treated with 0.25μM APH for 24 h (left) or 50μM etoposide for 6 h (right). Blue, DAPI. Red, γH2AX. Scale bar, 10 μm.
(D) Quantification of percent of cells with three or more γH2AX foci in control, APH-treated, or etoposide-treated NPCs.
(E) Representative images of PCHK1-Ser345 (top), pATM-Ser1981 (middle), and pRPA32-Ser4+Ser8 (bottom). Left, control NPCs, middle, ASD NPCs, right, etoposide-treated NPCs. Blue, DAPI. Scale bar, 10 μm.
(F) Quantification of percent of cells with three or more foci in control- or ASD-derived NPCs. Each point represents one line. Mean ± SD. Student's t test, two-tailed, ** $p < 0.01$.
(G) Quantification number of γH2AX foci in control- or ASD-derived NPCs treated with 0.25μM APH for 24 h. Each dot represents one cell. Average of each line was used for statistical test. Student's t test, two-tailed, * $p < 0.05$.

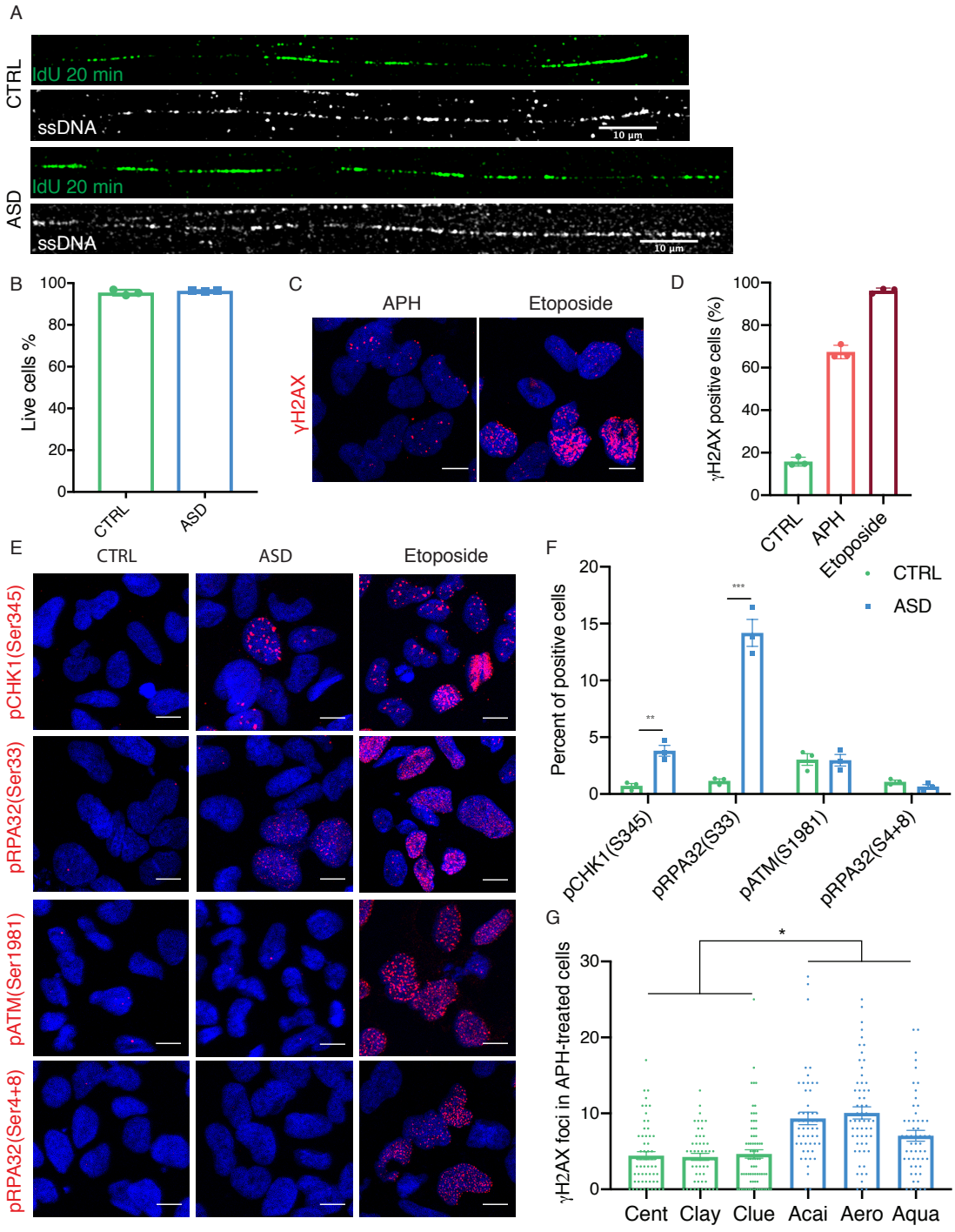
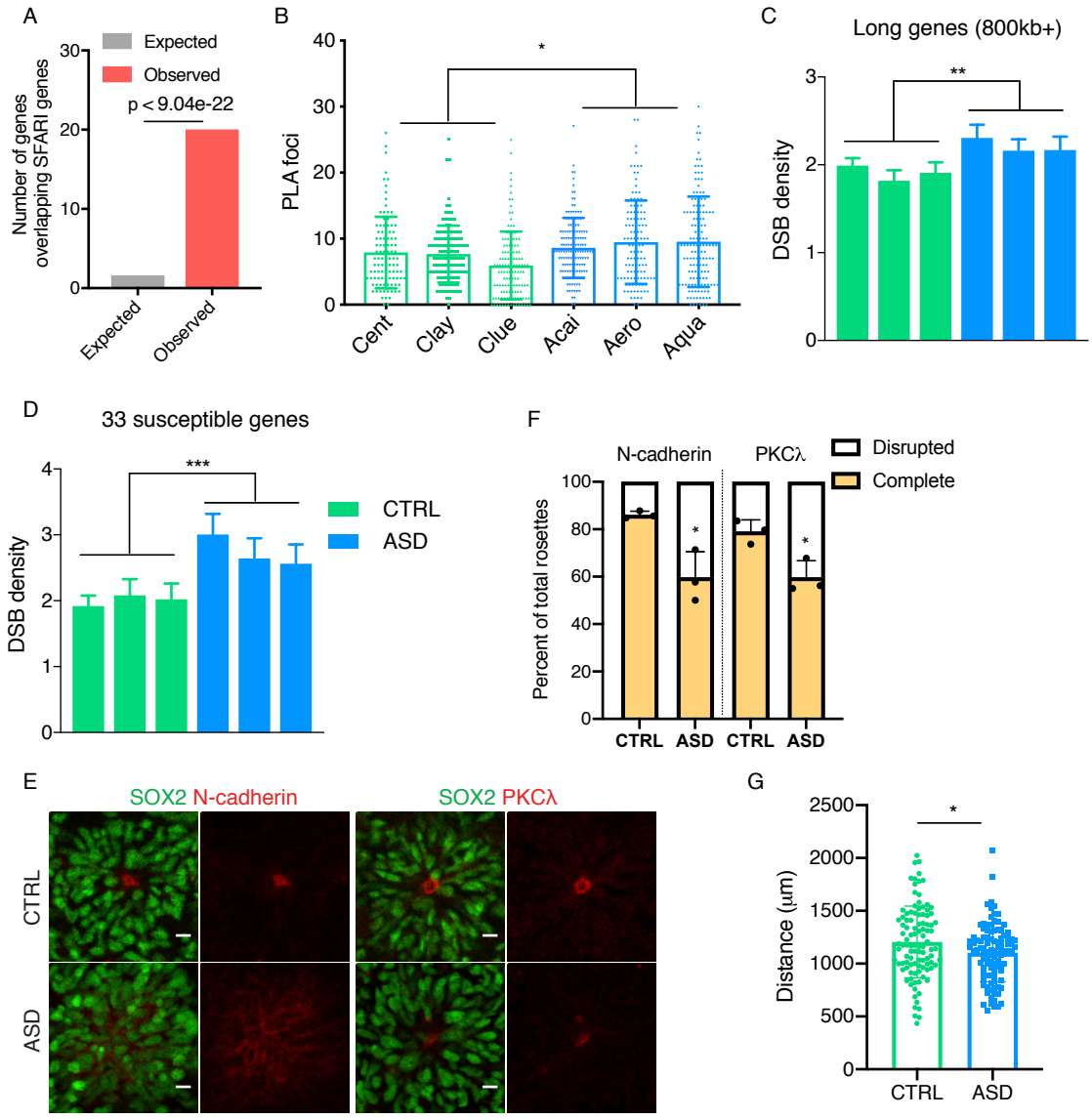

A
CTRL
IdU 20 min
ssDNA
10 µm
ASD
IdU 20 min
ssDNA
10 µm
B
Live cells %
100
80
60
40
20
0
CTRL
ASD
C
APH
Etoposide
γH2AX
D
γH2AX positive cells (%)
100
80
60
40
20
0
CTRL
APH
Etoposide
E
CTRL
ASD
Etoposide
pCHK1(Ser345)
pRPA32(Ser33)
pATM(Ser1981)
pRPA32(Ser4+8)
F
Percent of positive cells
20
15
10
5
0
CTRL
ASD
**
***
pCHK1(S345)
pRPA32(S33)
pATM(S1981)
pRPA32(S4+8)
G
γH2AX foci in APH-treated cells
30
20
10
0
*
Cent
Clay
Clue
Acai
Aero
Aqua

Figure 2.5: ASD-derived NPCs exhibit elevated DNA DSBs in long genes and defects in adherens junctions, apical polarity, and cell migration.

(A) Bar plot shows the expected and observed number of replication stress-susceptible genes overlapping SFARI genes. Hypergeometric test; $p < 9.04e-22$.
(B) Scatter plot shows the PCNA-RNAPII PLA foci. Each bar represents one cell line. Mean ± SD. Average of each line was used for statistical test. Student's t test, two-tailed, * $p < 0.05$.
(C) Bar plot shows the DSB densities assessed by HTGTS captured by Chr1 bait of long genes (> 800kb) located on chromosomes other than Chr1. Each bar represents one line. Mean ± SEM. Wilcoxon matched-pairs signed rank test, two-tailed, ** $p < 0.01$.
(D) Bar plot shows the DSB densities assessed by HTGTS captured by Chr1 bait of 33 NPC susceptible genes located on chromosomes other than Chr1 (Table 1.3). Each bar represents one line. Mean ± SEM. Wilcoxon matched-pairs signed rank test, two-tailed, *** $p < 0.001$.
(E) Defects in adherens junctions (N-cadherin, left) and apical polarity (PKCλ, right) in ASD-derived NPCs. Sample confocal images of immunostaining of N-cadherin and PKCλ for neural rosettes are shown. SOX2(green), N-cadherin (red, left), PKCλ (red, right). Top, control neural rosette. Bottom, ASD neural rosette. Scale bar, 10 μm.
(F) Quantification of neural rosettes with complete or disrupted N-cadherin or PKCλ expression in control or ASD NPCs. Average percentage of 4 cultures per line. Each dot represents one line. Mean ± SD. Student's t test, two-tailed, * $p < 0.05$.
(G) Quantification of cell migration from neurospheres generated from control or ASD NPCs after 60 h. Each point represents one neurosphere. Student's t test, two-tailed, * $p < 0.05$.


A
Expected
Observed
p < 9.04e-22
Number of genes overlapping SFARI genes
Expected
Observed
B
PLA foci
Cent
Clay
Clue
Acai
Aero
Aqua
C
Long genes (800kb+)
DSB density
D
33 susceptible genes
DSB density
CTRL
ASD
E
SOX2 N-cadherin
SOX2 PKCλ
CTRL
ASD
F
N-cadherin
PKCλ
Disrupted
Complete
Percent of total rosettes
CTRL
ASD
CTRL
ASD
G
Distance (μm)
CTRL
ASD

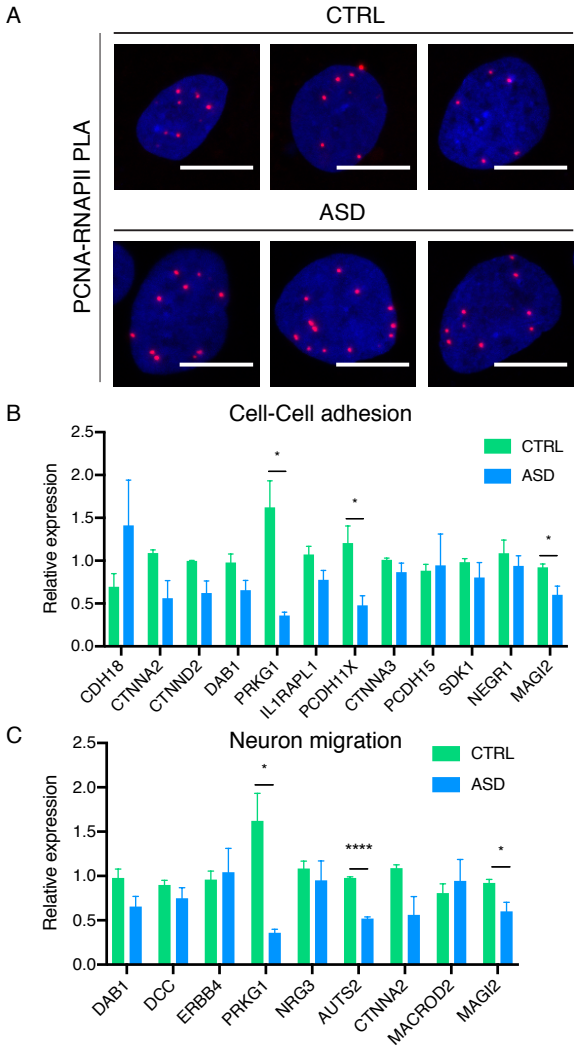


Figure 2.6: ASD-derived NPCs harbor elevated PCNA-RNAPII PLA foci and show expression changes.

(A) Representative images of PCNA-RNAPII PLA in the control NPCs (top) or ASD NPCs (bottom). PLA, red. DAPI, blue . Scale bar, 10 µm.
(B) qPCR analysis of expression of susceptible genes implicated in cell-cell adhesion (GO: 0098609) of Hues6-derived NPCs treated with DMSO or 0.5 µM APH for 48 h. n=3. Mean ± SD. Student's t test, two-tailed, *** $p < 0.001$, ** $p < 0.01$, ns, not significant.
(C) qPCR analysis of expression of susceptible genes implicated in neuron migration (GO: 0001764) of Hues6-derived NPCs treated with DMSO or 0.5 µM APH for 48 h. n=3. Mean ± SD. Student's t test, two-tailed, ** $p < 0.01$, * $p < 0.05$, ns, not significant.

Table 2.1: ASD subjects and matched control subjects-derived NPC lines used in this study (Marchetto et al., 2017).

	iPSC code	Race/ Ethnicity	Scan Age, years	Total Brain Volume
ASD				
1	Acai	Caucasian	4.5	1458
2	Aero	Caucasian	3.2	1338
3	Aqua	Caucasian Hispanic	4.05	1409
CONTROL				
1	Cent	Caucasian	3.58	1389
2	Clay	Caucasian	2.70	1182
3	Clue	Caucasian	3.41	1206

All cases were male. ASD cases met or exceeded cutoffs for a diagnostic classification of ASD on the ADOS/PL-ADOS and DSM-IV.

Table 2.2: Genes identified in human gene module in SFARI database updated on November 21, 2018.

Gene symbol	Gene name	Genetic category
DAB1	Disabled homolog 1 (Drosophila)	Rare Single Gene Mutation, Functional
NEGR1	Neuronal growth regulator 1	Rare Single Gene Mutation, Functional
CTNNA2	catenin alpha 2	Syndromic
PCDH15	Protocadherin-related 15	Rare Single Gene Mutation, Genetic Association
CTNNA3	Catenin (cadherin-associated protein), alpha 3	Rare Single Gene Mutation, Genetic Association
NAV2	Neuron navigator 2	Rare Single Gene Mutation
SOX5	SRY (sex determining region Y)-box 5	Rare Single Gene Mutation
GPC6	Glypican 6	Rare Single Gene Mutation, Genetic Association
MDGA2	MAM domain containing glycosylphosphatidylinositol anchor 2	Rare Single Gene Mutation, Genetic Association
RBFOX1	RNA binding protein, fox-1 homolog (C. elegans) 1	Rare Single Gene Mutation, Genetic Association
NCKAP5	NCK-associated protein 5	Rare Single Gene Mutation
ERBB4	V-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)	Rare Single Gene Mutation
MACROD2	MACRO domain containing 2	Rare Single Gene Mutation, Genetic Association
CTNND2	Catenin (cadherin-associated protein), delta 2	Rare Single Gene Mutation
PARK2	Parkinson disease (autosomal recessive, juvenile) 2, parkin	Rare Single Gene Mutation, Genetic Association
SDK1	Sidekick cell adhesion molecule 1	Rare Single Gene Mutation, Genetic Association
AUTS2	Autism susceptibility candidate 2	Rare Single Gene Mutation, Syndromic, Genetic Association
CSMD1	CUB and Sushi multiple domains 1	Rare Single Gene Mutation, Genetic Association
IL1RAPL1	Interleukin 1 receptor accessory protein-like 1	Rare Single Gene Mutation
PCDH11X	Protocadherin 11 X-linked	Rare Single Gene Mutation

Table 2.3: Genes reported in the CNV module of SFARI Gene.

Del: deletion; Dup: duplication; #: number.

CNV-locus	CNV-type	Del-values	Dup-values	# of reports	# of case population	# of case individuals	Associated gene
1p31.1	Del-Dup	131	43	28	45	175	*NEGR1*
2p12	Del-Dup	161	82	17	28	243	*CTNNA2*
2q21.2	Dup	10	27	12	21	37	*NCKAP5*
2q34	Del-Dup	77	30	23	39	107	*ERBB4*
5p15.2	Del-Dup	218	35	27	43	253	*CTNND2*
6q26	Del-Dup	169	109	35	58	278	*PARK2*
7p22.2	Del-Dup	67	22	16	28	89	*SDK1*
7q11.22	Del-Dup	119	82	34	59	201	*AUTS2*
8p23.2	Del-Dup	175	119	31	50	294	*CSMD1*
10q21.1	Del-Dup	140	104	19	32	244	*PCDH15*
10q21.3	Del-Dup	317	26	35	53	343	*CTNNA3*
11p15.1	Del-Dup	31	15	18	28	47	*NAV2*
12p12.1	Del-Dup	71	25	13	23	96	*SOX5*
13q31.3-q32.1	Del-Dup	2	1	2	4	3	*GPC6*
14q21.3	Del-Dup	276	43	13	25	319	*MDGA2*
16p13.3	Del-Dup	275	188	45	70	465	*RBFOX1*
20p12.1	Del-Dup	240	32	30	46	272	*MACROD2*
Xp21.3-p21.2	Dup	5	2	6	8	7	*IL1RAPL1*
Xq21.31	Del-Dup	53	28	15	23	81	*PCDH11X*

Chapter 3 Replication stress induces LINE-1 retrotransposition in the human NPCs

Introduction

Many fundamental features of brain somatic LINE-1 retrotransposition remain to be determined. It is unclear if the rate of retrotransposition differs in different regions of the brain, considering adult neurogenesis may especially contribute to the rate of somatic retrotransposition in the adult neurogenic niche. Moreover, the mechanisms that regulate the rate and targeting regions remain poorly understood.

LINE-1 EN activity and DNA replication has been suggested to dictate insertion preferences (Flasch et al., 2019, Mita et al., 2018). During neurogenesis, millions of cell divisions of NPCs are required to generate most of the 80 billion neurons in the adult human brain (Lui et al., 2011). Perturbations in this process may influence the rate and insertion preferences of LINE-1 retrotransposition. We have demonstrated in previous chapters that replication stress induces genomic instability. However, it remains to be explored if increased replication stress induces more somatic mutations. LINE-1 retrotransposition is an important cause of somatic mutations. In this chapter, I will examine whether replication stress alters somatic mutation rate, focusing on LINE-1 retrotransposition rate and potential mechanisms.

Replication stress induces LINE-1 retrotransposition

To monitor LINE-1 retrotransposition, an engineered human LRE3 LINE-1 element containing a retrotransposition indicator cassette in the 3'UTR was applied. The cassette encodes an enhanced green fluorescent protein (EGFP) under the control of a

CMV promoter in the inverse orientation relative to the LRE sequence. The EGFP sequence is interrupted by an intron sequence in the same orientation as LRE3. The EGFP expression is only observed upon retrotranposition into the host genome. The Hues6-derived NPCs were transfected with the LRE3–EGFP reporter construct, treated with low dose of APH or vehicle control (DMSO) for 5 days, and L1 retrotransposition was monitored by flow cytometry (EGFP-positive cells) 14 days post transfection (Figure 3.1A and 3.1B). A LINE-1 construct containing a pair of missense mutations in the ORF1p (JM111) was used as a negative control. We reliably detected increased EGFP-positive cells of the NPCs treated with APH compared to the control, while the negative control JM111 remains the same in the APH-treated NPCs compared to the controls (Figure 3.1C). Taken together, these data suggest increased somatic LINE-1 insertion rate in the NPCs upon replication stress.

Replication stress has minimal effects on LINE-1 expression

To understand the mechanisms underlying increased LINE-1 retrotransposition upon replication stress, we first monitored the promoter activity of LINE-1 using a reporter plasmid where the 5'UTR of the LINE-1 element was cloned into a dual-luciferase reporter plasmid. We transfected the cells with the reporter plasmid, treated the cells with APH and DMSO for 24 hours, and collected to measure luciferase expression level. We didn't observe any significant changes of the luciferase expression level of the APH-treated cells compared to the controls (Figure 3.1D). We next evaluated whether the endogenous LINE-1 expression has been altered. We collected the RNA from NPCs treated with APH or DMSO for 24 hours. To reliable detect LINE-1 mRNA expression, we applied two sets of qPCR primers targeting the 5'UTR, ORF1, ORF2, and 3'UTR of

the human endogenous LINE-1 elements. qPCR analysis revealed no overall changes in LINE-1 expression between the APH-treated and control NPCs (Figure 3.1E). Taken together, these data suggest that the increased LINE-1 retrotransposition in NPCs treated with APH was not a result from changes in the expression of LINE-1.

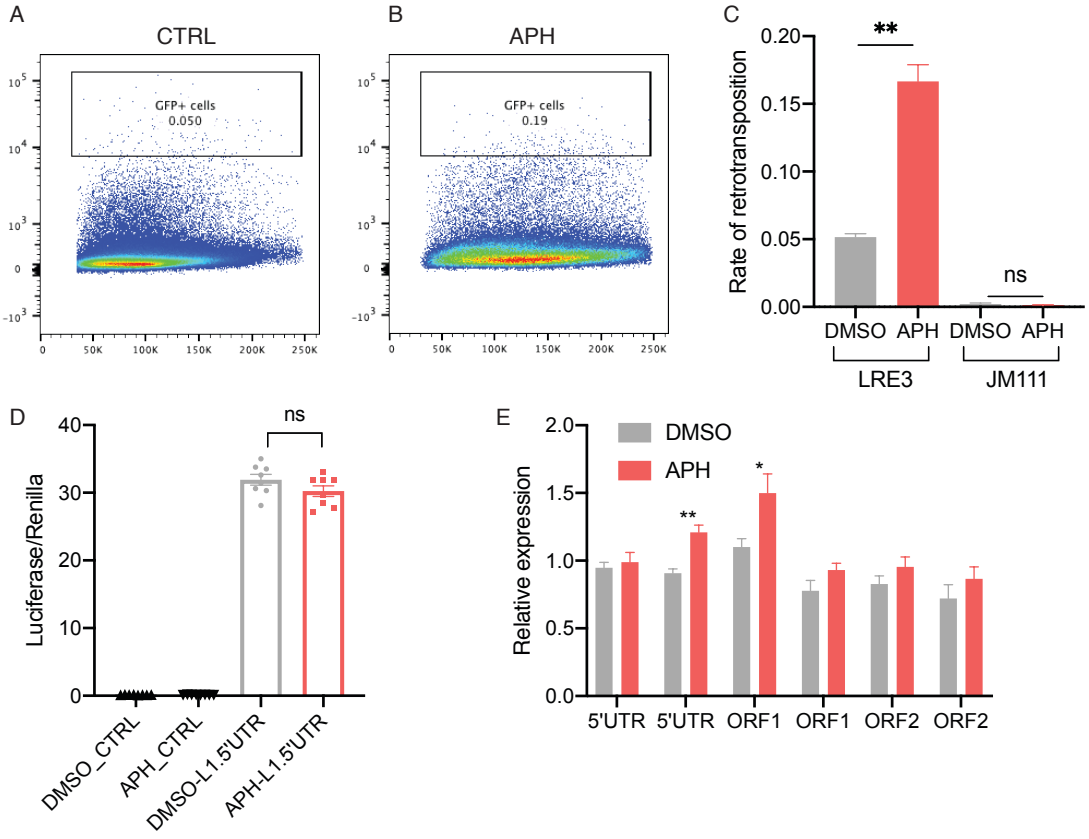


Figure 3.1: Replication stress induces LINE-1 retrotransposition without affecting transcription level.

(A) Representative FACS plot of the retrotransposition assay of the Hues6 NPCs treated with vehicle control.
(B) Representative FACS plot of the retrotransposition assay of the Hues6 NPCs treated with APH.
(C) Bar plot shows the quantification of rate of LINE-1 retrotransposition by the reporter assay. LRE3, retrotransposition competent plasmid. JM111, retrotransposition imcompetent plasmid. Mean ± SEM. Students' t test, two-tailed, ** $p < 0.01$., ns, not significant.
(D) Bar plot shows the promoter activity assessed by luciferase assay. Left, negative control plasmid. Right, plasmid containing LINE-1 5'UTR. Mean ± SEM. Students' t test, two-tailed, ns, not significant.
(E) Bar plot shows the qPCR analysis of LINE-1 expression. Two sets of primers were designed to examine 5'UTR, ORF1, and ORF2. Mean ± SEM. Students' t test, two-tailed, ** $p < 0.01$, * $p < 0.05$.

Discussion

DNA repair by classical, non-homologous end-joining is required for neural development (Barnes et al., 1998, Gao et al., 1998, Frank et al., 2000), suggesting critical roles for DNA repair during embryonic neurogenesis. Studies of the DSB hotspots in human neural precursors are essential for elucidating the details of nervous system development and for understanding the mechanisms underlying brain somatic mosaicism and its contribution to human-specific neurodevelopmental diseases such as autism and schizophrenia (McConnell et al., 2017). We observed that, in human NPCs, replication stress induced DSBs in a number of actively transcribed long genes critical for nervous system development. Notably, 26 susceptible genes identified in human NPCs corresponded to RDC-containing genes in mouse neural precursors (Wei et al., 2016, Wei et al., 2018). Remarkably, 10 susceptible genes were unique to human NPCs. Eight of the human-specific susceptible genes were reliably identified by both baits in our study.

Our work reveals a previously unknown mechanism by which accelerated S-phase progression in ASD-derived NPCs may contribute to DNA damage via increased replication stress. We demonstrated that ASD-derived NPCs displayed accelerated S-phase progression accompanied by an altered replication program and that perturbed S-phase progression and potentially other factors induced replication stress and activated the ATR-CHK1 pathway. Collisions between transcription and replication activate distinct DNA damage responses depending on the conflict orientation. A head-on orientation collision between the transcription machineries and replication fork leads to fork stalling and robust activation of the ATR pathway (Hamperl et al., 2017) .The activation of DNA damage and ATR-CHK1 pathway in the ASD-derived NPCs

corroborates the observation that replication stress induces head-on transcription-replication conflict (Hamperl et al., 2017). Consistently, ASD-derived NPCs harbor more transcription-replication conflicts, reflected by increased PCNA-RNAPII PLA foci. Head-on collision may block transcription, which can in turn lead to diminished gene expression. In fact, in both APH-treated NPC cultures and ASD-derived NPCs, gene expression was attenuated in many of the replication stress-susceptible genes. Our data suggest an intriguing mechanism by which replication stress in the ASD-derived NPCs activates the ATR pathway and chronic DNA damage and induces transcription-replication conflicts, which then leads to attenuated gene expression and functional defects.

Our study reveals that replication stress caused defects in adherens junctions, apical polarity, and migration of NPCs, reminiscent of what was observed in NPCs carrying 15q11.2 CNVs (Yoon et al., 2014). 15q11.2 CNVs are prominent risk factors for various neuropsychiatric disorders, including schizophrenia, ASD, and intellectual disability (Malhotra and Sebat, 2012). Knockdown of CYFIP1, a gene within 15q11.2, caused ectopic localization of radial glial cells in the developing mouse cortex (Yoon et al., 2014), similar to a recent study in which a high incidence of patches of cortical laminar disorganization in autistic brains was identified (Stoner et al., 2014). Aberrant cell migration has been reported in neuropsychiatric disorders (Penagarikano et al., 2011, Wegiel et al., 2010). Recent studies have found migration defects using NPCs derived from patients with neuropsychiatric disorders (Han et al., 2016, Brennand et al., 2015). *CTNNA2*, a gene robustly downregulated upon replication stress, was reported in pachygyria syndrome (Schaffer et al., 2018), where disordered cortical neuronal

migration was observed in the cerebral cortex. *SOX5* contributes to reduction in regional differences in ASD based on expression analysis of the postmortem brain (Parikshak et al., 2016). *RBFOX1*, a major neuronal splicing regulator, is linked to isoform-level dysregulation in ASD and other psychiatric diseases (Gandal et al., 2018).

Our study also suggests that replication stress induces increased LINE-1 retrotransposition, a type of somatic mutations found in the brain, which has been associated with many psychiatric disorders including schizophrenia. While it remains to be explored the mechanism by which LINE-1 retrotransposition is regulated and whether other types of mutations are associated with replication stress, our study suggests an intriguing mechanistic link between replication stress associated genome instability and neurodevelopmental disorders. Future studies on identifying the genomic location and the cause of brain somatic mutations and how they contribute to neurodevelopmental disorders will facilitate the understanding of disease etiology and drug discovery.

Acknowledgements

Discussion, in part, has been submitted for publication of the material as it may appear in Cell Stem Cell, Wang, Meiyan; Wei, Pei-Chi; Lim, Christina; Gallina, Iryna; Marshall, Sara; Marchetto, Carol; Alt, Frederick; Gage, Fred, Elsevier Press, 2019. The dissertation author was the primary investigator and author of this paper.

References

ALT, F. W., ZHANG, Y., MENG, F. L., GUO, C. & SCHWER, B. 2013. Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell,* 152**,** 417-29.

BARLOW, J. H., FARYABI, R. B., CALLEN, E., WONG, N., MALHOWSKI, A., CHEN, H. T., GUTIERREZ-CRUZ, G., SUN, H. W., MCKINNON, P., WRIGHT, G., CASELLAS, R., ROBBIANI, D. F., STAUDT, L., FERNANDEZ-CAPETILLO, O. & NUSSENZWEIG, A. 2013. Identification of early replicating fragile sites that contribute to genome instability. *Cell,* 152**,** 620-32.

BARNES, D. E., STAMP, G., ROSEWELL, I., DENZEL, A. & LINDAHL, T. 1998. Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. *Curr Biol,* 8**,** 1395-8.

BIPOLAR, D., SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS CONSORTIUM. ELECTRONIC ADDRESS, D. R. V. E., BIPOLAR, D. & SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C. 2018. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell,* 173**,** 1705-1715 e16.

BRENNAND, K., SAVAS, J. N., KIM, Y., TRAN, N., SIMONE, A., HASHIMOTO-TORII, K., BEAUMONT, K. G., KIM, H. J., TOPOL, A., LADRAN, I., ABDELRAHIM, M., MATIKAINEN-ANKNEY, B., CHAO, S. H., MRKSICH, M., RAKIC, P., FANG, G., ZHANG, B., YATES, J. R., 3RD & GAGE, F. H. 2015. Phenotypic differences in hiPSC NPCs derived from patients with schizophrenia. *Mol Psychiatry,* 20**,** 361-8.

CAI, X., EVRONY, G. D., LEHMANN, H. S., ELHOSARY, P. C., MEHTA, B. K., PODURI, A. & WALSH, C. A. 2014. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep,* 8**,** 1280-9.

CIMPRICH, K. A. & CORTEZ, D. 2008. ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol,* 9**,** 616-27.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

CORE, L. J., WATERFALL, J. J. & LIS, J. T. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science,* 322**,** 1845-8.

COUFAL, N. G., GARCIA-PEREZ, J. L., PENG, G. E., YEO, G. W., MU, Y., LOVCI, M. T., MORELL, M., O'SHEA, K. S., MORAN, J. V. & GAGE, F. H. 2009. L1 retrotransposition in human neural progenitor cells. *Nature,* 460**,** 1127-31.

COURCHESNE, E., MOUTON, P. R., CALHOUN, M. E., SEMENDEFERI, K., AHRENS-BARBEAU, C., HALLET, M. J., BARNES, C. C. & PIERCE, K. 2011. Neuron number and size in prefrontal cortex of children with autism. *JAMA,* 306**,** 2001-10.

DEBATISSE, M., LE TALLEC, B., LETESSIER, A., DUTRILLAUX, B. & BRISON, O. 2012. Common fragile sites: mechanisms of instability revisited. *Trends Genet,* 28**,** 22-32.

DENLI, A. M., NARVAIZA, I., KERMAN, B. E., PENA, M., BENNER, C., MARCHETTO, M. C., DIEDRICH, J. K., ASLANIAN, A., MA, J., MORESCO, J. J., MOORE, L., HUNTER, T., SAGHATELIAN, A. & GAGE, F. H. 2015. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell,* 163**,** 583-93.

DOMBROSKI, B. A., MATHIAS, S. L., NANTHAKUMAR, E., SCOTT, A. F. & KAZAZIAN, H. H., JR. 1991. Isolation of an active human transposable element. *Science,* 254**,** 1805-8.

DOUCET, A. J., WILUSZ, J. E., MIYOSHI, T., LIU, Y. & MORAN, J. V. 2015. A 3' Poly(A) Tract Is Required for LINE-1 Retrotransposition. *Mol Cell,* 60**,** 728-741.

DU, G., DREXLER, G. A., FRIEDLAND, W., GREUBEL, C., HABLE, V., KRUCKEN, R., KUGLER, A., TONELLI, L., FRIEDL, A. A. & DOLLINGER, G. 2011. Spatial dynamics of DNA damage response protein foci along the ion trajectory of high-LET particles. *Radiat Res,* 176**,** 706-15.

ERWIN, J. A., PAQUOLA, A. C., SINGER, T., GALLINA, I., NOVOTNY, M., QUAYLE, C., BEDROSIAN, T. A., ALVES, F. I., BUTCHER, C. R., HERDY, J. R., SARKAR, A., LASKEN, R. S., MUOTRI, A. R. & GAGE, F. H. 2016. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat Neurosci,* 19**,** 1583-1591.

EVRONY, G. D., CAI, X., LEE, E., HILLS, L. B., ELHOSARY, P. C., LEHMANN, H. S., PARKER, J. J., ATABAY, K. D., GILMORE, E. C., PODURI, A., PARK, P. J. & WALSH, C. A. 2012. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell,* 151**,** 483-96.

FENG, Q., MORAN, J. V., KAZAZIAN, H. H., JR. & BOEKE, J. D. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell,* 87**,** 905-16.

FLASCH, D. A., MACIA, A., SANCHEZ, L., LJUNGMAN, M., HERAS, S. R., GARCIA-PEREZ, J. L., WILSON, T. E. & MORAN, J. V. 2019. Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. *Cell,* 177**,** 837-851 e28.

FRANK, K. M., SHARPLESS, N. E., GAO, Y., SEKIGUCHI, J. M., FERGUSON, D. O., ZHU, C., MANIS, J. P., HORNER, J., DEPINHO, R. A. & ALT, F. W. 2000. DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. *Mol Cell,* 5**,** 993-1002.

FROCK, R. L., HU, J., MEYERS, R. M., HO, Y. J., KII, E. & ALT, F. W. 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol,* 33**,** 179-86.

GANDAL, M. J., ZHANG, P., HADJIMICHAEL, E., WALKER, R. L., CHEN, C., LIU, S., WON, H., VAN BAKEL, H., VARGHESE, M., WANG, Y., SHIEH, A. W., HANEY, J., PARHAMI, S., BELMONT, J., KIM, M., MORAN LOSADA, P., KHAN, Z., MLECZKO, J., XIA, Y., DAI, R., WANG, D., YANG, Y. T., XU, M., FISH, K., HOF, P. R., WARRELL, J., FITZGERALD, D., WHITE, K., JAFFE, A. E., PSYCH, E. C., PETERS, M. A., GERSTEIN, M., LIU, C., IAKOUCHEVA, L. M., PINTO, D. & GESCHWIND, D. H. 2018. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science,* 362.

GAO, Y., FERGUSON, D. O., XIE, W., MANIS, J. P., SEKIGUCHI, J., FRANK, K. M., CHAUDHURI, J., HORNER, J., DEPINHO, R. A. & ALT, F. W. 2000. Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature,* 404**,** 897-900.

GAO, Y., SUN, Y., FRANK, K. M., DIKKES, P., FUJIWARA, Y., SEIDL, K. J., SEKIGUCHI, J. M., RATHBUN, G. A., SWAT, W., WANG, J., BRONSON, R. T., MALYNN, B. A., BRYANS, M., ZHU, C., CHAUDHURI, J., DAVIDSON, L., FERRINI, R., STAMATO, T., ORKIN, S. H., GREENBERG, M. E. & ALT, F. W. 1998. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell,* 95**,** 891-902.

GLOVER, T. W., BERGER, C., COYLE, J. & ECHO, B. 1984. DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum Genet,* 67**,** 136-42.

GLOVER, T. W. & WILSON, T. E. 2016. Molecular biology: Breaks in the brain. *Nature,* 532**,** 46-7.

GORE, A., LI, Z., FUNG, H. L., YOUNG, J. E., AGARWAL, S., ANTOSIEWICZ-BOURGET, J., CANTO, I., GIORGETTI, A., ISRAEL, M. A., KISKINIS, E., LEE, J. H., LOH, Y. H., MANOS, P. D., MONTSERRAT, N., PANOPOULOS, A. D., RUIZ, S., WILBERT, M. L., YU, J., KIRKNESS, E. F., IZPISUA BELMONTE, J. C., ROSSI, D. J., THOMSON, J. A., EGGAN, K., DALEY, G. Q., GOLDSTEIN, L. S. & ZHANG, K. 2011. Somatic coding mutations in human induced pluripotent stem cells. *Nature,* 471**,** 63-7.

GREENMAN, C., STEPHENS, P., SMITH, R., DALGLIESH, G. L., HUNTER, C., BIGNELL, G., DAVIES, H., TEAGUE, J., BUTLER, A., STEVENS, C., EDKINS, S., O'MEARA, S., VASTRIK, I., SCHMIDT, E. E., AVIS, T., BARTHORPE, S., BHAMRA, G., BUCK, G., CHOUDHURY, B., CLEMENTS, J., COLE, J., DICKS, E., FORBES, S., GRAY, K., HALLIDAY, K., HARRISON, R., HILLS, K., HINTON, J., JENKINSON, A., JONES, D., MENZIES, A., MIRONENKO, T., PERRY, J., RAINE, K., RICHARDSON, D., SHEPHERD, R., SMALL, A., TOFTS, C., VARIAN, J., WEBB, T., WEST, S., WIDAA, S., YATES, A., CAHILL, D. P., LOUIS, D. N., GOLDSTRAW, P., NICHOLSON, A. G., BRASSEUR, F., LOOIJENGA, L., WEBER, B. L., CHIEW, Y. E., DEFAZIO, A., GREAVES, M. F., GREEN, A. R., CAMPBELL, P., BIRNEY, E., EASTON, D. F., CHENEVIX-TRENCH, G., TAN, M. H., KHOO, S. K., TEH, B. T., YUEN, S. T., LEUNG, S. Y., WOOSTER, R., FUTREAL, P. A. & STRATTON, M. R. 2007. Patterns of somatic mutation in human cancer genomes. *Nature,* 446**,** 153-8.

GROVE, J., RIPKE, S., ALS, T. D., MATTHEISEN, M., WALTERS, R. K., WON, H., PALLESEN, J., AGERBO, E., ANDREASSEN, O. A., ANNEY, R., AWASHTI, S., BELLIVEAU, R., BETTELLA, F., BUXBAUM, J. D., BYBJERG-GRAUHOLM, J., BAEKVAD-HANSEN, M., CERRATO, F., CHAMBERT, K., CHRISTENSEN, J. H., CHURCHHOUSE, C., DELLENVALL, K., DEMONTIS, D., DE RUBEIS, S., DEVLIN, B., DJUROVIC, S., DUMONT, A. L., GOLDSTEIN, J. I., HANSEN, C. S., HAUBERG, M. E., HOLLEGAARD, M. V., HOPE, S., HOWRIGAN, D. P., HUANG, H., HULTMAN, C. M., KLEI, L., MALLER, J., MARTIN, J., MARTIN, A. R., MORAN, J. L., NYEGAARD, M., NAERLAND, T., PALMER, D. S., PALOTIE, A., PEDERSEN, C. B., PEDERSEN, M. G., DPOTERBA, T., POULSEN, J. B., POURCAIN, B. S., QVIST, P., REHNSTROM, K., REICHENBERG, A., REICHERT, J., ROBINSON, E. B., ROEDER, K., ROUSSOS, P., SAEMUNDSEN, E., SANDIN, S., SATTERSTROM, F. K., DAVEY SMITH, G., STEFANSSON, H., STEINBERG, S., STEVENS, C. R., SULLIVAN, P. F., TURLEY, P., WALTERS, G. B., XU, X., AUTISM SPECTRUM DISORDER WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C., BUPGEN, MAJOR DEPRESSIVE DISORDER WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C., ANDME RESEARCH, T., STEFANSSON, K., GESCHWIND, D. H., NORDENTOFT, M., HOUGAARD, D. M., WERGE, T., MORS, O., MORTENSEN, P. B., NEALE, B. M., DALY, M. J. & BORGLUM, A. D. 2019. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet,* 51**,** 431-444.

HAMPERL, S., BOCEK, M. J., SALDIVAR, J. C., SWIGUT, T. & CIMPRICH, K. A. 2017. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell,* 170**,** 774-786 e19.

HAN, J., KIM, H. J., SCHAFER, S. T., PAQUOLA, A., CLEMENSON, G. D., TODA, T., OH, J., PANKONIN, A. R., LEE, B. S., JOHNSTON, S. T., SARKAR, A., DENLI, A. M. & GAGE, F. H. 2016. Functional Implications of miR-19 in the Migration of Newborn Neurons in the Adult Brain. *Neuron,* 91**,** 79-89.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

HELMRICH, A., BALLARINO, M. & TORA, L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell,* 44**,** 966-77.

HORI, K., NAGAI, T., SHAN, W., SAKAMOTO, A., TAYA, S., HASHIMOTO, R., HAYASHI, T., ABE, M., YAMAZAKI, M., NAKAO, K., NISHIOKA, T., SAKIMURA, K., YAMADA, K., KAIBUCHI, K. & HOSHINO, M. 2014. Cytoskeletal regulation by AUTS2 in neuronal migration and neuritogenesis. *Cell Rep,* 9**,** 2166-79.

HU, J., MEYERS, R. M., DONG, J., PANCHAKSHARI, R. A., ALT, F. W. & FROCK, R. L. 2016. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat Protoc,* 11**,** 853-71.

JI, J., NG, S. H., SHARMA, V., NECULAI, D., HUSSEIN, S., SAM, M., TRINH, Q., CHURCH, G. M., MCPHERSON, J. D., NAGY, A. & BATADA, N. N. 2012. Elevated coding mutation rate during

the reprogramming of human somatic cells into induced pluripotent stem cells. *Stem Cells,* 30**,** 435-40.

KANG, E., WANG, X., TIPPNER-HEDGES, R., MA, H., FOLMES, C. D., GUTIERREZ, N. M., LEE, Y., VAN DYKEN, C., AHMED, R., LI, Y., KOSKI, A., HAYAMA, T., LUO, S., HARDING, C. O., AMATO, P., JENSEN, J., BATTAGLIA, D., LEE, D., WU, D., TERZIC, A., WOLF, D. P., HUANG, T. & MITALIPOV, S. 2016. Age-Related Accumulation of Somatic Mitochondrial DNA Mutations in Adult-Derived Human iPSCs. *Cell Stem Cell,* 18**,** 625-36.

KHAZINA, E., TRUFFAULT, V., BUTTNER, R., SCHMIDT, S., COLES, M. & WEICHENRIEDER, O. 2011. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol,* 18**,** 1006-14.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, Y., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LEE, J. H., HUYNH, M., SILHAVY, J. L., KIM, S., DIXON-SALAZAR, T., HEIBERG, A., SCOTT, E., BAFNA, V., HILL, K. J., COLLAZO, A., FUNARI, V., RUSS, C., GABRIEL, S. B., MATHERN, G. W. & GLEESON, J. G. 2012. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat Genet,* 44**,** 941-5.

LIM, J. S., KIM, W. I., KANG, H. C., KIM, S. H., PARK, A. H., PARK, E. K., CHO, Y. W., KIM, S., KIM, H. M., KIM, J. A., KIM, J., RHEE, H., KANG, S. G., KIM, H. D., KIM, D., KIM, D. S. & LEE, J. H. 2015. Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy. *Nat Med,* 21**,** 395-400.

LUAN, D. D., KORMAN, M. H., JAKUBCZAK, J. L. & EICKBUSH, T. H. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell,* 72**,** 595-605.

LUI, J. H., HANSEN, D. V. & KRIEGSTEIN, A. R. 2011. Development and evolution of the human neocortex. *Cell,* 146**,** 18-36.

MALHOTRA, D. & SEBAT, J. 2012. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell,* 148**,** 1223-41.

MARCHETTO, M. C., BELINSON, H., TIAN, Y., FREITAS, B. C., FU, C., VADODARIA, K., BELTRAO-BRAGA, P., TRUJILLO, C. A., MENDES, A. P. D., PADMANABHAN, K., NUNEZ, Y., OU, J., GHOSH, H., WRIGHT, R., BRENNAND, K., PIERCE, K., EICHENFIELD, L., PRAMPARO, T., EYLER, L., BARNES, C. C., COURCHESNE, E., GESCHWIND, D. H., GAGE, F. H., WYNSHAW-BORIS, A. & MUOTRI, A. R. 2017. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol Psychiatry,* 22**,** 820-835.

MARCHETTO, M. C., HRVOJ-MIHIC, B., KERMAN, B. E., YU, D. X., VADODARIA, K. C., LINKER, S. B., NARVAIZA, I., SANTOS, R., DENLI, A. M., MENDES, A. P., OEFNER, R., COOK, J., MCHENRY, L., GRASMICK, J. M., HEARD, K., FREDLENDER, C., RANDOLPH-MOORE, L., KSHIRSAGAR, R., XENITOPOULOS, R., CHOU, G., HAH, N., MUOTRI, A. R., PADMANABHAN, K., SEMENDEFERI, K. & GAGE, F. H. 2019. Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. *Elife,* 8.

MARECHAL, A., LI, J. M., JI, X. Y., WU, C. S., YAZINSKI, S. A., NGUYEN, H. D., LIU, S., JIMENEZ, A. E., JIN, J. & ZOU, L. 2014. PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry. *Mol Cell,* 53**,** 235-246.

MARECHAL, A. & ZOU, L. 2013. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb Perspect Biol,* 5.

MARIANI, J., COPPOLA, G., ZHANG, P., ABYZOV, A., PROVINI, L., TOMASINI, L., AMENDUNI, M., SZEKELY, A., PALEJEV, D., WILSON, M., GERSTEIN, M., GRIGORENKO, E. L., CHAWARSKA, K., PELPHREY, K. A., HOWE, J. R. & VACCARINO, F. M. 2015. FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell,* 162**,** 375-390.

MATHIAS, S. L., SCOTT, A. F., KAZAZIAN, H. H., JR., BOEKE, J. D. & GABRIEL, A. 1991. Reverse transcriptase encoded by a human transposable element. *Science,* 254**,** 1808-10.

MCCONNELL, M. J., LINDBERG, M. R., BRENNAND, K. J., PIPER, J. C., VOET, T., COWING-ZITRON, C., SHUMILINA, S., LASKEN, R. S., VERMEESCH, J. R., HALL, I. M. & GAGE, F. H. 2013. Mosaic copy number variation in human neurons. *Science,* 342**,** 632-7.

MCCONNELL, M. J., MORAN, J. V., ABYZOV, A., AKBARIAN, S., BAE, T., CORTES-CIRIANO, I., ERWIN, J. A., FASCHING, L., FLASCH, D. A., FREED, D., GANZ, J., JAFFE, A. E., KWAN, K. Y., KWON, M., LODATO, M. A., MILLS, R. E., PAQUOLA, A. C. M., RODIN, R. E., ROSENBLUH, C., SESTAN, N., SHERMAN, M. A., SHIN, J. H., SONG, S., STRAUB, R. E., THORPE, J., WEINBERGER, D. R., URBAN, A. E., ZHOU, B., GAGE, F. H., LEHNER, T., SENTHIL, G., WALSH, C. A., CHESS, A., COURCHESNE, E., GLEESON, J. G., KIDD, J. M., PARK, P. J., PEVSNER, J., VACCARINO, F. M. & BRAIN SOMATIC MOSAICISM, N. 2017. Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science,* 356.

MITA, P., WUDZINSKA, A., SUN, X., ANDRADE, J., NAYAK, S., KAHLER, D. J., BADRI, S., LACAVA, J., UEBERHEIDE, B., YUN, C. Y., FENYO, D. & BOEKE, J. D. 2018. LINE-1 protein localization and functional dynamics during the cell cycle. *Elife,* 7.

MUOTRI, A. R., CHU, V. T., MARCHETTO, M. C., DENG, W., MORAN, J. V. & GAGE, F. H. 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature,* 435**,** 903-10.

NESTLER, E. J. & HYMAN, S. E. 2010. Animal models of neuropsychiatric disorders. *Nat Neurosci,* 13**,** 1161-9.

O'ROAK, B. J., VIVES, L., FU, W., EGERTSON, J. D., STANAWAY, I. B., PHELPS, I. G., CARVILL, G., KUMAR, A., LEE, C., ANKENMAN, K., MUNSON, J., HIATT, J. B., TURNER, E. H., LEVY, R., O'DAY, D. R., KRUMM, N., COE, B. P., MARTIN, B. K., BORENSTEIN, E., NICKERSON, D. A., MEFFORD, H. C., DOHERTY, D., AKEY, J. M., BERNIER, R., EICHLER, E. E. & SHENDURE, J. 2012. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science,* 338**,** 1619-22.

PANKOTAI, T., BONHOMME, C., CHEN, D. & SOUTOGLOU, E. 2012. DNAPKcs-dependent arrest of RNA polymerase II transcription in the presence of DNA breaks. *Nat Struct Mol Biol,* 19**,** 276-82.

PARIKSHAK, N. N., SWARUP, V., BELGARD, T. G., IRIMIA, M., RAMASWAMI, G., GANDAL, M. J., HARTL, C., LEPPA, V., UBIETA, L. T., HUANG, J., LOWE, J. K., BLENCOWE, B. J., HORVATH, S. & GESCHWIND, D. H. 2016. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature,* 540**,** 423-427.

PENAGARIKANO, O., ABRAHAMS, B. S., HERMAN, E. I., WINDEN, K. D., GDALYAHU, A., DONG, H., SONNENBLICK, L. I., GRUVER, R., ALMAJANO, J., BRAGIN, A., GOLSHANI, P., TRACHTENBERG, J. T., PELES, E. & GESCHWIND, D. H. 2011. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell,* 147**,** 235-46.

PODURI, A., EVRONY, G. D., CAI, X., ELHOSARY, P. C., BEROUKHIM, R., LEHTINEN, M. K., HILLS, L. B., HEINZEN, E. L., HILL, A., HILL, R. S., BARRY, B. J., BOURGEOIS, B. F., RIVIELLO, J. J., BARKOVICH, A. J., BLACK, P. M., LIGON, K. L. & WALSH, C. A. 2012. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron,* 74**,** 41-8.

PODURI, A., EVRONY, G. D., CAI, X. & WALSH, C. A. 2013. Somatic mutation, genomic variation, and neurological disease. *Science,* 341**,** 1237758.

SCHAFFER, A. E., BREUSS, M. W., CAGLAYAN, A. O., AL-SANAA, N., AL-ABDULWAHED, H. Y., KAYMAKCALAN, H., YILMAZ, C., ZAKI, M. S., ROSTI, R. O., COPELAND, B., BAEK, S. T., MUSAEV, D., SCOTT, E. C., BEN-OMRAN, T., KARIMINEJAD, A., KAYSERILI, H., MOJAHEDI, F., KARA, M., CAI, N., SILHAVY, J. L., ELSHARIF, S., FENERCIOGLU, E., BARSHOP, B. A., KARA, B., WANG, R., STANLEY, V., JAMES, K. N., NACHNANI, R., KALUR, A., MEGAHED, H., INCECIK, F., DANDA, S., ALANAY, Y., FAQEIH, E., MELIKISHVILI, G., MANSOUR, L., MILLER, I., SUKHUDYAN, B., CHELLY, J., DOBYNS, W. B., BILGUVAR, K., JAMRA, R. A., GUNEL, M. & GLEESON, J. G. 2018. Biallelic loss of human CTNNA2, encoding alphaN-catenin, leads to ARP2/3 complex overactivity and disordered cortical neuronal migration. *Nat Genet,* 50**,** 1093-1101.

SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature,* 511**,** 421-7.

SHI, Y., INOUE, H., WU, J. C. & YAMANAKA, S. 2017. Induced pluripotent stem cell technology: a decade of progress. *Nat Rev Drug Discov,* 16**,** 115-130.

SHI, Y., KIRWAN, P., SMITH, J., ROBINSON, H. P. & LIVESEY, F. J. 2012. Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. *Nat Neurosci,* 15**,** 477-86, S1.

SPEEK, M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol,* 21**,** 1973-85.

STONER, R., CHOW, M. L., BOYLE, M. P., SUNKIN, S. M., MOUTON, P. R., ROY, S., WYNSHAW-BORIS, A., COLAMARINO, S. A., LEIN, E. S. & COURCHESNE, E. 2014. Patches of disorganization in the neocortex of children with autism. *N Engl J Med,* 370**,** 1209-1219.

SWERGOLD, G. D. 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol,* 10**,** 6718-29.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131**,** 861-72.

TECHER, H., KOUNDRIOUKOFF, S., NICOLAS, A. & DEBATISSE, M. 2017. The impact of replication stress on replication dynamics and DNA damage in vertebrate cells. *Nat Rev Genet,* 18**,** 535-550.

TEKIN, S. & CUMMINGS, J. L. 2002. Frontal-subcortical neuronal circuits and clinical neuropsychiatry: an update. *J Psychosom Res,* 53**,** 647-54.

UPTON, K. R., GERHARDT, D. J., JESUADIAN, J. S., RICHARDSON, S. R., SANCHEZ-LUQUE, F. J., BODEA, G. O., EWING, A. D., SALVADOR-PALOMEQUE, C., VAN DER KNAAP, M. S., BRENNAN, P. M., VANDERVER, A. & FAULKNER, G. J. 2015. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell,* 161**,** 228-39.

WEGIEL, J., KUCHNA, I., NOWICKI, K., IMAKI, H., WEGIEL, J., MARCHI, E., MA, S. Y., CHAUHAN, A., CHAUHAN, V., BOBROWICZ, T. W., DE LEON, M., LOUIS, L. A., COHEN, I. L., LONDON, E., BROWN, W. T. & WISNIEWSKI, T. 2010. The neuropathology of autism: defects of neurogenesis and neuronal migration, and dysplastic changes. *Acta Neuropathol,* 119**,** 755-70.

WEI, P. C., CHANG, A. N., KAO, J., DU, Z., MEYERS, R. M., ALT, F. W. & SCHWER, B. 2016. Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell,* 164**,** 644-55.

WEI, P. C., LEE, C. S., DU, Z., SCHWER, B., ZHANG, Y., KAO, J., ZURITA, J. & ALT, F. W. 2018. Three classes of recurrent DNA break clusters in brain progenitors identified by 3D proximity-based break joining assay. *Proc Natl Acad Sci U S A,* 115**,** 1919-1924.

WEI, W., GILBERT, N., OOI, S. L., LAWLER, J. F., OSTERTAG, E. M., KAZAZIAN, H. H., BOEKE, J. D. & MORAN, J. V. 2001. Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol,* 21**,** 1429-39.

WILSON, T. E., ARLT, M. F., PARK, S. H., RAJENDRAN, S., PAULSEN, M., LJUNGMAN, M. & GLOVER, T. W. 2015. Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome Res,* 25**,** 189-200.

WRAY, N. R., RIPKE, S., MATTHEISEN, M., TRZASKOWSKI, M., BYRNE, E. M., ABDELLAOUI, A., ADAMS, M. J., AGERBO, E., AIR, T. M., ANDLAUER, T. M. F., BACANU, S. A., BAEKVAD-HANSEN, M., BEEKMAN, A. F. T., BIGDELI, T. B., BINDER, E. B., BLACKWOOD, D. R. H., BRYOIS, J., BUTTENSCHON, H. N., BYBJERG-GRAUHOLM, J., CAI, N., CASTELAO, E., CHRISTENSEN, J. H., CLARKE, T. K., COLEMAN, J. I. R., COLODRO-CONDE, L., COUVY-DUCHESNE, B., CRADDOCK, N., CRAWFORD, G. E., CROWLEY, C. A., DASHTI, H. S., DAVIES, G., DEARY, I. J., DEGENHARDT, F., DERKS, E. M., DIREK, N., DOLAN, C. V., DUNN, E. C., ELEY, T. C., ERIKSSON, N., ESCOTT-PRICE, V., KIADEH, F. H. F., FINUCANE, H. K., FORSTNER, A. J., FRANK, J., GASPAR, H. A., GILL, M., GIUSTI-RODRIGUEZ, P., GOES, F. S., GORDON, S. D., GROVE, J., HALL, L. S., HANNON, E., HANSEN, C. S., HANSEN, T. F., HERMS, S., HICKIE, I. B., HOFFMANN, P., HOMUTH, G., HORN, C., HOTTENGA, J. J., HOUGAARD, D. M., HU, M., HYDE, C. L., ISING, M., JANSEN, R., JIN, F., JORGENSON, E., KNOWLES, J. A., KOHANE, I. S., KRAFT, J., KRETZSCHMAR, W. W., KROGH, J., KUTALIK, Z., LANE, J. M., LI, Y., LI, Y., LIND, P. A., LIU, X., LU, L., MACINTYRE, D. J., MACKINNON, D. F., MAIER, R. M., MAIER, W., MARCHINI, J., MBAREK, H., MCGRATH, P., MCGUFFIN, P., MEDLAND, S. E., MEHTA, D., MIDDELDORP, C. M., MIHAILOV, E., MILANESCHI, Y., MILANI, L., MILL, J., MONDIMORE, F. M., MONTGOMERY, G. W., MOSTAFAVI, S., MULLINS, N., NAUCK, M., NG, B., et al. 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet,* 50**,** 668-681.

YOON, K. J., NGUYEN, H. N., URSINI, G., ZHANG, F., KIM, N. S., WEN, Z., MAKRI, G., NAUEN, D., SHIN, J. H., PARK, Y., CHUNG, R., PEKLE, E., ZHANG, C., TOWE, M., HUSSAINI, S. M., LEE, Y., RUJESCU, D., ST CLAIR, D., KLEINMAN, J. E., HYDE, T. M., KRAUSS, G., CHRISTIAN, K. M., RAPOPORT, J. L., WEINBERGER, D. R., SONG, H. & MING, G. L. 2014. Modeling a genetic risk for schizophrenia in iPSCs and mice reveals neural stem cell deficits associated with adherens junctions and polarity. *Cell Stem Cell,* 15**,** 79-91.

YU, D. X., DI GIORGIO, F. P., YAO, J., MARCHETTO, M. C., BRENNAND, K., WRIGHT, R., MEI, A., MCHENRY, L., LISUK, D., GRASMICK, J. M., SILBERMAN, P., SILBERMAN, G., JAPPELLI, R. & GAGE, F. H. 2014. Modeling hippocampal neurogenesis using human pluripotent stem cells. *Stem Cell Reports,* 2**,** 295-310.

YU, J., VODYANIK, M. A., SMUGA-OTTO, K., ANTOSIEWICZ-BOURGET, J., FRANE, J. L., TIAN, S., NIE, J., JONSDOTTIR, G. A., RUOTTI, V., STEWART, R., SLUKVIN, II & THOMSON, J. A. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science,* 318**,** 1917-20.

ZANG, C., SCHONES, D. E., ZENG, C., CUI, K., ZHAO, K. & PENG, W. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics,* 25**,** 1952-8.

ZEMAN, M. K. & CIMPRICH, K. A. 2014. Causes and consequences of replication stress. *Nat Cell Biol,* 16**,** 2-9.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol,* 9**,** R137.

ZHAO, T., ZHANG, Z. N., RONG, Z. & XU, Y. 2011. Immunogenicity of induced pluripotent stem cells. *Nature,* 474**,** 212-5.

ZHAO, T., ZHANG, Z. N., WESTENSKOW, P. D., TODOROVA, D., HU, Z., LIN, T., RONG, Z., KIM, J., HE, J., WANG, M., CLEGG, D. O., YANG, Y. G., ZHANG, K., FRIEDLANDER, M. & XU, Y. 2015. Humanized Mice Reveal Differential Immunogenicity of Cells Derived from Autologous Induced Pluripotent Stem Cells. *Cell Stem Cell,* 17**,** 353-9.

ZOU, L. & ELLEDGE, S. J. 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science,* 300**,** 1542-8.