Neuronal Enhancers are Hotspots For DNA Single-Strand Break Repair

Wei Wu[1*], Sarah E. Hill[2*], William J. Nathan[1,3*], Jacob Paiano[1*], Kenta Shinoda[1], Jennifer Colon-Mercado[2], Elsa Callen[1], Raffaella de Pace[5], Dongpeng Wang[1], Han-Yu Shih[4], Steve Coon[5], Maia Parsadanian[2], Hana Hanzlikova[6,7], Peter J. McHugh[3], Andres Canela[8], Keith W. Caldecott[6,7‡], Michael E. Ward[2‡], & André Nussenzweig[1‡]

[1] Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA

[2] National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD USA

[3] Department of Oncology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom

[4] National Eye Institute, NIH, Bethesda, MD USA

[5] Eunice Kennedy Shriver National Institute of Child Health and Human Development

[6] Department of Genome Dynamics, Institute of Molecular Genetics of the Czech Academy of Sciences, 142 20 Prague, 4, Czech Republic

[7] Genome Damage and Stability Centre, University of Sussex, Falmer Brighton, BN1 9RQ, UK

[8] The Hakubi Center for Advanced Research and Radiation Biology Center, Graduate School of Biostudies, Kyoto University, Kyoto, Japan

*Authors contributed equally to this work

‡Correspondence: k.w.caldecott@sussex.ac.uk or wardme@nih.gov or andre_nussenzweig@nih.gov

Genome stability is essential for all cell types. However, defects in DNA repair frequently lead to neurodevelopmental and neurodegenerative diseases, underscoring the particular importance of DNA repair in long-lived post-mitotic neurons. The neuronal genome is subjected to a constant barrage of endogenous DNA damage due to high levels of oxidative metabolism in the central nervous system. Surprisingly, we know little about the identity of the lesion(s) that accumulate in neurons and whether they accrue throughout the genome or at specific loci. Here, we show that neurons, but not other post-mitotic cells, accumulate unexpectedly high numbers of DNA single-strand breaks (SSBs) at specific sites within the genome. These recurrent SSBs are found within enhancers, and trigger DNA repair through recruitment and activation of poly(ADP-ribose) polymerase-1 (PARP1) and XRCC1, the central SSB repair scaffold protein. Notably, deficiencies in PARP1, XRCC1, or DNA polymerase β elevate the localized incorporation of nucleotides, suggesting that the ongoing DNA synthesis at neuronal enhancers involves both short-patch and long-patch SSB repair processes. These data reveal unexpected levels of localized and continuous DNA single-strand breakage in neurons, suggesting an explanation for the neurodegenerative phenotypes that occur in patients with defective SSB repair.

Long-lived neurons are thought to be particularly prone to a progressive accumulation of endogenous DNA lesions. Recent studies have shown that single human neurons harbor large numbers of somatic muations, and that the burden of mutations in neurons substantially increases during aging[1,2]. Neuronal somatic mutations are especially abundant in patients with familial neurodegenerative diseases associated with dysfunctional DNA repair genes[1,3], as well as neurodevelopmental diseases such as autism and spectrum disorder (ASD)[4]. Despite a wealth of evidence supporting the importance of neuronal DNA damage in neurological diseases, the precise identity and sources of DNA damage that shape the mutational landscape of the neuronal genome remain unclear.

Several different types of DNA damage have been reported to accumulate in neurons including base modifications, single- and double-strand breaks and interstrand crosslinks[3,5]. While defects in all major DNA repair pathways have been implicated in neurological diseases, single-strand break (SSB) repair appears to be particularly important in neurons, because defects in this pathway leads almost exclusively to neuronal dysfunction and degeneration[3,5]. Importantly, DNA excision and DNA strand break repair pathways are associated with unscheduled DNA synthesis, an obligatory and characteristic step of DNA repair known as gap filling. During gap filling, excised or missing nucleotides are replaced, usually using the undamaged strand as a template[6]. Importantly, if gap filling involves the incorporation of sufficient number of nucleotides, DNA repair synthesis can be a robust measure of the extent and sites of endogenous DNA damage[7].

Regions of the neuronal genome are associated with ongoing DNA synthesis

Using alkaline comet assays, others previously observed endogenous DNA damage in cultured rat neurons[8]. We developed a new method to measure sites of DNA repair synthesis and map their genomic locations by sequencing (synthesis-associated with repair sequencing; SAR-seq). We labeled post-mitotic iPSC-derived glutamatergic neurons (i3Neurons[9,10]) with a thymidine analog EdU for 18 hours, biotinylated the labeled DNA by click chemistry, sonicated it to 150-200 bp, and then isolated the labeled DNA for high-throughput sequencing (**Fig. 1a**). Surprisingly, we identified >55,000 peaks of DNA synthesis at recurrent genomic locations in neurons, which were highly reproducible between different experiments (**Fig. 1b**, **Extended Data Fig. 1a, b**). Peaks were not caused by DNA synthesis during S phase[11,12], because the neurons are post-mitotic and the SAR-seq peaks were unaffected by inhibitors of the replicative DNA polymerases DNA polymerase α (PolA), Pol δ, and Pol ϵ (**Extended Data Fig. 1c, d).** As expected, neuronal DNA synthesis was largely prevented by hydroxyurea, which reduces the availability of deoxyribonucleotides (**Extended Data Fig. 1c, d).**

The tracts of DNA synthesis ranged from about 200-2,000 bp in width (averaging 1,006 bp) (**Fig. 1c**), most likely reflecting multiple clustered sites of DNA repair. The most prominent SAR peaks in the post-mitotic neurons were detectable by pulse labeling with EdU for just 1 hour, and EdU incorporation approached saturation after labeling for 18 hours (**Extended Data Fig. 2a**). On average, we estimate that any given tract was labeled to 89% capacity within ~10 hours of EdU labeling (**Extended Data Fig. 2b**).

Importantly, when iPSCs were differentiated into skeletal muscle cells rather than neurons, we did not detect incorporation of EdU (**Extended Data Fig. 2c**). Similarly, we failed to detect EdU incorporation in G0-arrested pre-B cells, although we could detect EdU

incorporation in the pre-B cells at site-specific DNA double strand breaks (DSBs) (**Extended Data Fig. 2d**). To rule out the possibility that the SAR-seq peaks were an artifact of iPSC differentiation, we also labelled *bona fide* rat neurons with EdU. Similar to i[3]Neurons, we detected robust peaks of EdU incorporation at specific sites in the rat neurons (**Fig. 1d**). Thus, the high frequency of recurrent DNA synthesis appears to be a specific feature of post-mitotic neurons.

The recurrent sites of neuronal DNA synthesis are located in cis-acting enhancers

Neuronal SAR-seq peaks were highly enriched in intragenic regions (**Extended Data Fig. 2e**) and associated with expressed genes (**Extended Data Fig. 2f**), suggesting the involvement of transcription. However, the signal intensity did not correlate with transcript levels as measured by RNA-seq (**Extended Data Fig. 2f**). Moreover, the sites of EdU incorporation were not associated with strand-specificity, because the EdU was incorporated uniformly in both transcribed and non-transcribed strands, resulting in symmetrical strand-specific SAR-seq profiles (**Extended Data Fig. 2g**).

We next searched for specific DNA motifs using the strongest 5,000 SAR-seq peaks. More than 50% of the sites harbored a motif similar to the ONECUT family of transcription factors, and the ONECUT1 motif was centered at SAR-seq peaks (**Extended Data Fig. 3a**). Since ONECUT1 has been implicated in promoting genomic accessibility in neurons[13], we compared SAR-seq peaks with accessible regions using ATAC-seq. We found that 57% of the SAR-seq regions coincided with ATAC-seq peaks (**Extended Data Fig. 3b**) and the width of the SAR-seq peaks correlated with the width of the ATAC-seq peaks, suggesting that open

chromatin structure influences the extent of DNA synthesis (**Extended Data Fig. 3c and Fig. 1c**).

Although the neuronal sites of recurrent DNA synthesis are located predominantly in open chromatin, SAR-seq peaks were not enriched at promoters like ATAC-seq peaks (**Fig. 2a,d and Extended Data Fig. 3d**). Indeed, promoters occupied by RNA polymerase II (POL II), as determined by ChIP-seq, exhibited only modest levels of DNA synthesis (**Extended Data Fig. 3e**). In contrast to promoters, however, we detected a strong positive correlation between the location of DNA synthesis and neuronal enhancers, as measured by ChIP-seq for H3K4me1, H3K27ac and MLL4, the major mammalian histone H3K4 mono-methyltransferase[14] (**Fig. 2a-c**). Indeed, SAR-seq peaks were centered on the nucleosome-free region occupied by MLL4 (**Fig. 2a, b**). Altogether, 87% of SAR-seq peaks overlapped with H3K4me1 peaks and were significantly enriched at enhancers compared to random regions (**Fig. 2c, d**), indicating enhancers are hotspots of neuronal DNA synthesis. Consistent with these results in i[3]Neurons, SAR peaks in primary rat neurons also overlapped and correlated with enhancers (**Extended Data Fig 4a**).

MLL4 has been reported to colocalize with lineage determining transcription factors[14]. To determine whether the enhancers with SAR-seq peaks are specific to those that are active in differentiated neurons, we compared sites of H3K4me1 in neurons with those in the iPSCs from which they were derived. Only 2% of the SAR-seq peaks overlapped with iPSCs-specific H3K4me1peaks, while most of the SAR-seq peaks overlapped with either neuron-specific or H3K4me1 sites that were shared between neurons and iPSCs (**Extended Data Fig. 4b**). Thus, based on H3K4me1 localization, SAR is associated with enhancers that are active in differentiated neurons.

We then performed Gene Ontology (GO) analysis to determine biological processes associated with the genes containing SAR-seq peaks. Many of the top enriched GO terms related to neuronal migration, development, axon formation, and synapse assembly, reflecting nervous system function (**Extended Data Fig. 4c**). Super-enhancers (SE) are a large collection of enhancers that drive transcription of genes involved in cell identity. We found approximately 1,300 SE in i³Neurons as determined by H3K27ac ChIP-seq, 90% of which exhibited SAR-seq peaks, whereas less than 25% of conventional enhancers were enriched in SAR (**Fig. 2e, Extended Data Fig. 4d**). Collectively, these data identify super-enhancers and genes associated with neuronal function as hotpsots of recurrent DNA synthesis.

PARP activation at neuronal sites of DNA synthesis

Next, we addressed the source of the DNA synthesis in post-mitotic neurons. Given the close association of unrepaired DNA breaks with neurodegeneration, we wondered if the sites of EdU incorporation might reflect sites of DNA break repair. To test this, we measured the activity of poly(ADP-ribose) polymerase (PARP) enzymes at the sites of DNA synthesis. PARP1 and PARP2 are activated in response to various types of DNA breakage, including SSBs, DSBs, and single-strand gaps[15,16]. PARP activity signals the presence of these lesions by modifying proteins with poly(ADP-ribose) (PAR), after which PAR is rapidly removed from proteins by poly(ADP-ribose) glycohydrolase (PARG)[15,16]. When we monitored ADP-ribosylation in individual neurons using an anti-pan-ADP ribose-binding reagent,[17] we observed robust focal staining in neuronal nuclei (**Fig. 3a**). To determine whether the poly(ADP-ribose) was located at sites of neuronal DNA synthesis we employed PAR ChIP-seq, which we demonstrated can detect ADP-ribosylation at site-specific DSBs in pre-B cells (**Extended Data Fig. 5a**). Strikingly, when we

used this method in i[3]Neurons, we found that sites of PAR synthesis co-localized with sites of DNA synthesis (**Fig. 3b, c).** These findings strongly suggest that the recurrent sites of DNA synthesis at neuronal enhancers are associated with sites of DNA break repair.

Sites of neuronal DNA synthesis are sites of DNA single-strand break repair

Neuronal activity has been reported to result in DSBs[18,19]. Moreover, topoisomerase 2 (TOP2) is associated with neuronal activity. TOP2-induced DSBs can promote the expression of early response genes[18], associating these DNA breaks with regions of transcriptional activity. However, while we found that treatment of i[3]Neurons with etoposide to trigger TOP2-induced DSBs resulted in DNA synthesis within gene bodies, most of the sites of etoposide-induced DNA synthesis were distinct from those detected in untreated neurons (**Extended Data Fig**. **5b, c**). In addition, we did not detect DSBs in unchallenged i[3]Neurons as measured by either γ-H2AX/53BP1 immunostaining (**Extended Data Fig. 5d**) or by END-seq (**Extended Data Fig. 5e**), which maps DSBs genome-wide at base-pair resolution[20]. Thus, the sites of DNA synthesis in neuronal enhancers are independent of DSBs.

Importantly, PARP1 and/or PARP2 are also activated at SSBs, which subsequently recruits the XRCC1 protein complexes necessary for the repair of DNA single-strand gaps and breaks[21,22]. We therefore examined the genomic localization of XRCC1 by ChIP-seq. Strikingly and similar to PAR, XRCC1 also co-localized with SAR-peaks, and the intensity of XRCC1 binding was correlated with the intensity of EdU incorporation in both human i[3]Neurons and primary rat neurons (**Fig. 3b, c; Extended Data Fig. 5f**). Collectively, these data indicate that the sites of DNA synthesis at neuronal enhancers are sites of PARP activation and XRCC1-associated SSB repair.

DNA synthesis at neuronal enhancers involves both short-patch and long-patch SSB repair.

PARP1 and XRCC1 promote the repair of a wide spectrum of SSBs[23]. We therefore examined the impact of inhibiting and/or depleting these proteins on neuronal DNA synthesis. Intriguingly, we observed a reproducible increase in SAR when neurons were treated with three independent inhibitors of PARP during the EdU incubation, or if PARP1 was depleted using CRISPR interference (CRISPRi)[24] (**Fig. 4a, b, Extended Data Fig. 6a**). Moreover, depletion of XRCC1 similarly led to a prominent increase in neuronal SAR (**Fig. 4c-e; Extended Data Fig. 6b**). These data suggest that if PARP1- and XRCC1-dependent repair is impeded, the amount of EdU incorporation that is associated with SSB repair is increased. The simplest explanation for these data is that in the absence of XRCC1-dependent short-patch repair, long-patch repair is triggered instead (**Fig. 5c**). During short-patch SSB repair a single nucleotide is replaced at the site of the break[23,25], typically by DNA polymerase β (POLβ) which interacts directly with XRCC1[5,21] (**Fig. 5c**). In contrast, during long-patch repair, alternative DNA polymerases can achieve gap filling, at the expense of increased patch sizes for DNA synthesis. Indeed, we found that depletion of POLβ by two independent CRISPRi sgRNAs resulted in a dramatic increase in DNA synthesis at neuronal enhancers (**Fig. 5a, b; Extended Data Fig. 6c, d**). Thus, the increased DNA repair synthesis in PARP1/XRCC1/POLβ-depleted neurons is consistent with increased activity of long-patch SSB repair.

Conclusions

Our study reveals that human post-mitotic neurons are subject to a profound and unexpected degree of localized and recurrent DNA synthesis, which is associated with ongoing sites of SSB

repair at neuronal enhancers. The scaffold protein XRCC1 is of particular importance during SSB repair because it interacts directly with many of the components of short-patch repair including POLβ and LIG3[21]. Interestingly, patients with hereditary mutations in XRCC1 exhibit progressive neurodegeneration, suggesting that intact SP-SSB repair is necessary for long-term neuronal viability[26,27]. One possible explanation for this is that the elevated PARP1 activity that results from loss of XRCC1-dependent SSB repair triggers neuronal dydsfunction and/or ultimately cell death[21]. In addition, our current data raise the possibility that an increased dependency on long-patch DNA repair synthesis in XRCC1-defective cells increases mutational burden in long-lived neurons (**Fig. 5c**). Age-dependent accrual of mutations at neuronal enhancers through hyperactive LP-SSB repair could subsequently lead to aberrant transcription, eventually resulting in neurodegeneration. This notion is consistent with the recent finding that somatic mutations measured in single human neurons are enriched at highly-accessible genomic regions including brain-specific enhancers and genes regulating neuronal function[1,2,4]. Although the mechanism(s) for such mutations were not defined, three distinct somatic mutation signatures were identified: a clock-like signature of aging, a brain region-specific signature, and a mutation class most closely associated with oxidative DNA damage[1,2].

It is unexpected that sites of endogenous SSB repair would be enriched at neuronal enhancers rather than being randomly distributed. Enhancers could be especially vulnerable to single-strand damage due to their increased mobility upon transcriptional activation[28], or because negatively-charged products of oxidative metabolism might preferentially interact with phase-separated enhancer condensates[29]. Intriguingly, active DNA demethylation of cytosine at CpG sites occurs preferentially at enhancers[30], is 10-fold more active in post-mitotic neurons than peripheral cell types [31], and generates SSBs that are intermediates of XRCC1-associated

BER[30,32]. Indeed, we have found that CpG dinucleotides are highly enriched at SAR sites

(**Extended Data Fig. 7**). While the short patch-BER pathway that replaces methylated cytosines

would not generate EdU incorporation, DNA synthesis surrounding this site by long patch-BER

might be a frequent outcome.

In summary, we describe a new method that enables genome-wide mapping of sites of

DNA repair in post-mitotic cells. Our findings identify enhancers as hotspots of endogenous

DNA single-strand break repair in human post-mitotic neurons, and may explain why this

important DNA repair process is crucial for maintaining the functionality and viability during

neuronal aging.

Note: During the preparation of this manuscript, we became aware of the closely related work of

D. Ried *et al.*, which demonstrated recurrent DNA repair sites in embryonic stem cell-derived

neurons. doi: https://doi.org/10.1101/2020.03.25.008490

Figure Legends

Fig. 1: Discrete loci in the neuronal genome are associated with ongoing DNA synthesis

a) Schematic of SAR-seq (DNA synthesis associated with repair sequencing) methodology. A

population of neurons grown in culture (*1*) are incubated with EdU to label sites of DNA repair

synthesis (*2*). Genomic EdU is tagged with biotin via click chemistry (*3*), sheared by sonication

and captured with streptavidin beads (*4*). Enriched DNA sequences are PCR-amplified and

subjected to next-generation sequencing (*5*).

b) Genome browser screenshot displaying SAR-seq profiles as normalized read density (reads per million, RPM) for human iPSC-derived neurons (i³Neurons). Three independent biological replicates are shown as well as input DNA sequenced in parallel.

c) Histogram of individual SAR-seq peak widths, with an average width of 1,006 bp.

d) Genome browser screenshot of SAR-seq performed on rat primary neurons. The culture was simultaneously treated with 5μM aphidicolin to block DNA replication in S phase glial cells.

Fig. 2: SAR-seq peaks occur within nucleosome-free regions of enhancers

a) Genome browser screenshots displaying SAR-seq (red), ATAC-seq (green), and H3K4me1 (light blue), H3K27ac (steel blue), and MLL4 ChIP-seq (indigo) in i³Neurons. Zoomed-in screenshot of the highlighted region is shown below. ATAC-seq and MLL4 peaks align directly with the SAR-seq peak, while H3K4me1 and H3K27ac peaks flank the SAR-seq site. Region highlighted in black in top panel shows the absence of SAR-seq signal at the promoter (see also quantitation in Fig. **2d** and Extended Data Figs. **2e** and **3d** below).

b) Heatmaps of SAR-seq, H3K4me1, H3K27ac, MLL4 ChIP-seq and ATAC-seq signal ±1kb around SAR-seq peak summits in i³Neurons, ordered by SAR-seq intensity. SAR-seq peaks are located within the nucleosome-free region of H3K4me1 and H3K27ac ChIP-seq peaks, and colocalize with MLL4 ChIP-seq and ATAC-seq regions.

c) Venn diagram showing the overlap between H3K4me1 and SAR-seq peaks in i³Neurons. n= 1,000 random datasets were generated to test the significance of overlap (one-sided Fisher's Exact test (p<2.2e-16)).

d) Graph showing the fold enrichment of SAR-seq and ATAC-seq peaks located at enhancers (black) and promoters (grey) compared to 1000 sets of randomly shuffled regions of the same

sizes and chromosome distributions, respectively. (Fisher's Exact test ***p<2.2e-16; NS, p=0.07)).

e) Graph showing the fraction of super-enhancers (left) overlapping with SAR-seq peaks compared to conventional enhancers (right). 1,385 super-enhancers were defined by H3K27ac ChIP-seq intensity in i[3]Neurons (see **Extended Data Fig. 4d**).

Fig. 3: PARP activation and XRCC1 recruitment at neuronal sites of DNA synthesis

a) Representative images of anti-PAR immunofluorescence staining in i[3]Neurons. As a positive control for DNA damage and increased PARylation, cells were treated with 50uM etoposide (ETO) for 18 hr. Primary anti-PAR antibody without secondary antibody was used as a negative control.

b) Genome browser screenshots displaying SAR-seq (red), PAR (light blue) and XRCC1 ChIP-seq (steel blue) signal in i[3]Neurons. Cells were incubated with PARGi 20 minutes prior to fixation for PAR ChIP-seq.

c) Heatmaps of XRCC1 and PAR ChIP-seq signal ±1kb surrounding SAR-seq peak summits in i[3]Neurons, ordered by SAR-seq intensity.

Fig. 4: PARP or XRRC1 deficiency increases SAR-seq peak intensity

a) Genome browser screenshots displaying SAR-seq profiles in i[3]Neurons treated with PARP inhibitors olaparib, talazoparib, veliparib, or CRISPRi-mediated knockdown with a control non-targeting sgRNA (sgControl) or an sgRNA targeting PARP1 (sgPARP1) in duplicates.

b) Heatmaps of SAR-seq intensities ±1kb surrounding SAR-seq peak summits for i[3]Neurons treated with PARP inhibitors and in neurons containing CRISPRi-mediated PARP1 knockdown or non-targeting sgRNA control. Aggregate plot of SAR-seq intensity is shown in the top panel.

c) Genome browser screenshots of SAR-seq profiles in i[3]Neurons expressing CRISPRi non-targeting sgRNAs or targeting XRCC1 (sgXRCC1), in duplicate.

d) Heatmaps of SAR-seq intensities ±1kb surrounding SAR-seq peak summits for i[3]Neurons expressing non-targeting sgRNA or targeting XRCC1. Aggregate plot of SAR-seq intensity is shown in the top panel.

e) Venn diagrams showing the overlap of SAR-seq peaks between control (red) and XRCC1 (purple). n = 1,000 random datasets were generated to test significance of overlap (one-sided Fisher's Exact test (p<2.2e-16)).

Fig. 5: Deficiency in POLβ elevates the localized incorporation of EdU

a) Genome browser screenshots displaying SAR-seq profiles in i[3]Neurons expressing CRISPRi-mediated knockdown of POLβ (sgPOLB, two guides are shown) or a control non-targeting sgRNA (sgControl), in duplicates.

b) Heatmaps of SAR-seq intensities ±1kb surrounding SAR-seq peak summits for i[3]Neurons expressing CRISPRi-mediated knockdown of POLβ or a control non-targeting sgRNA. Aggregate plot of SAR-seq intensity is shown in the top panel.

c) Model suggesting that PARP1, XRCC1, and POLβ influence the choice between SP- and LP-SSB repair. Loss of these proteins results in patches of increased DNA synthesis arising from LP-SSB repair, which may increase mutation frequency.

Extended Data Fig. 1: Mapping regions of unscheduled DNA synthesis in neurons

a) Venn diagram showing the overlap of SAR-seq peaks in i[3]Neurons in three independent biological replicates.

b) Correlations of SAR-seq intensities between three replicates in i³Neurons. Pearson correlation coefficient is indicated.

c) Genome browser screenshot showing SAR-seq in i³Neurons treated with hydroxyurea (HU), aphidicolin (APH) or polymerase alpha inhibitor (POLAi). NT, not treated.

d) Scatterplots showing SAR-seq intensity (SAR-seq reads per kilobase per million mapped reads, RPKM) for HU- (left), and APH (middle)- and POLAi- (right) treated compared to non-treated samples.

Extended Data Fig. 2: Genomic features of SAR-seq peaks

a) Genome browser screenshot showing SAR-seq in i³Neurons harvested after 1 hour, 2 hours, 4 hours, 8 hours, or 18 hours of EdU incubation in otherwise non-treated cells.

b) Graph showing the fraction of EdU labeled in i³Neurons (relative to maximum labeling at 18 hours) as a function of time. Red points and error bars represent the relative levels of EdU labeled from experimental data. Black line represents the theoretical model after fitting, with k being the rate of EdU labeling.

c) Genome browser screenshot showing SAR-seq in iMuscle cells for two independent biological replicates. Cells were pretreated with aphidcolin.

d) Genome browser screenshot displaying SAR-seq peak at a representative AsiS1 restriction enzyme sites (tick mark). AsiS1 expression was induced for 18 hrs (+Dox) vs non-treated (-Dox) in G0-arrested, Abelson virus-transformed murine pre-B cells as described [20].

e) Pie-chart showing distribution of i³Neuron SAR-seq peaks at promoters, exons, introns and distal intergenic regions. Promoters are defined as 1kb upstream of transcription start sites and distal intergenic represents promoter-excluded intergenic regions.

f) Intensity of SAR-seq peaks within intragenic regions compared with transcript levels measured by RNA-seq. 71% of SAR-seq peaks are at expressed genes (FPKM≥0.1, red dashed line).

g) Genome browser screenshot comparing SAR-seq vs. strand-specific SAR-seq that discriminates which strand is labeled with EdU. Both strands show similar labeling, in triplicate.

Extended Data Fig. 3: Motif of SAR-seq peaks and comparison with ATAC-seq peaks

a) Motif analysis for the sequences within ±500bp surrounding the summit of top 5,000 SAR-seq peaks in i[3]Neurons. Upper panel, the best motif discovered by the MEME suite analysis. 2,834 out of 5,000 sites shared this motif. Middle panel, the motif of the transcription factor ONECUT1 was identified as the best matched motif to SAR-seq motif by the TOMTOM motif comparison tool. Bottom panel, position distribution of the best motif (shown in the upper panel) at ±500bp window of the SAR-seq peak summit. The best motif is centered on the SAR-seq peak summit.

b) Venn diagram showing overlap between peaks detected in ATAC-seq (green) and SAR-seq (red) in i[3]Neurons. n = 1,000 random datasets were generated to test significance of overlap (one-sided Fisher's Exact test (p<2.2e-16)).

c) Scatterplot comparing widths of ATAC-seq peaks and SAR-seq peaks in i[3]Neurons. Pearson correlation coefficient is indicated.

d) Distribution of SAR-seq and ATAC-seq peaks with respect to different genomic features compared to genome-wide distribution in the hg19 human reference genome. Promoters are defined as 1kb upstream of transcription start sites and distal intergenic represents promoter-excluded intergenic regions.

e) Heatmaps of RNA Polymerase II ChIP-seq signal and SAR-seq signal within ±1kb surrounding the transcription start site (TSS) in i³Neurons, ordered by RNA Polymerase II ChIP-seq intensity.

Extended Data Fig. 4: SAR enrichment at primary neuron enhancers and i³Neuron super-enhancers

a) Heatmaps of SAR-seq and H3K4me1 ChIP-seq signal at ±1kb around the SAR-seq peak summit in embryonic day 18 primary rat cortical neurons ordered by SAR-seq intensity.

b) Pie-chart showing distribution of i³Neuron SAR-seq peaks in iPSC-specific, i³Neuron-specific and iPSC- and i³Neuron- shared enhancers.

c) Top biological processes enriched for the genes containing top 2000 SAR-seq peaks determined by Gene Ontology (GO) analysis.

d) H3K27ac signal at enhancers in i³Neurons ranked by H3K27ac ChIP-seq intensity. Red dashed line indicates the inflection point of H3K27ac signal used to determine super enhancers (cutoff: 1000). 1,385 enhancers were defined as super enhancers.

Extended Data Fig. 5: Mapping regions of DNA damage and repair in neurons

a) PAR ChIP-seq reads at AsiSI restriction enzyme cut site in Abelson virus-transformed murine pre-B cells. Cells were arrested in G0, and AsiSI double-strand breaks were induced for 18 hr prior to ChIP fixation. PARylation is not observed in non-treated (-AsiSI) cells and is robustly increased after 20 min PARGi treatment prior to fixation (AsiSI + PARGi).

b) Genome browser examples of SAR-seq in either control (NT) or 18 hr 50uM etoposide (ETO) treated neurons, in duplicate.

c) Heatmaps for SAR-seq in either control (NT) or in 18 hrs 50uM etoposide- (ETO) treated i³Neurons at -2kb to +5kb of the transcription start sites (TSS) ordered by ETO SAR-seq signal.

d) Immunofluorescence staining of DSB markers γ-H2AX and 53BP1 in non-treated or 1 hour etoposide-treated i³Neurons

e) SAR-seq vs END-seq in non-treated i³Neurons. END-seq, a method to specifically detect double-strand breaks, shows no enriched signal over background at SAR-seq peaks. END-seq reads are separated by plus and minus DNA strands (black and grey tracks, respectively).

f) Heatmaps of SAR-seq and XRCC1 ChIP-seq in cultured primary rat neurons, centered on SAR-seq peak summits and ordered by SAR-seq intensity.

Extended Fig. 6: Confirmation of PARP1, XRCC1, and POLb knockdowns in i³Neurons

(**a-c**) Quantitative RT-PCR analysis of PARP1 (**a**), XRCC1 (**b**), and POLβ (**c**) transcript levels in i³Neurons after CRISPRi knockdown, cultured in parallel with samples used for SAR-seq.

d) Western blot of POLβ protein after CRISPRi knockdown.

Extended Fig. 7: CG dinucleotide content associated with SAR-seq peaks

Aggregate plots showing the distribution of CG dinucleotide (black) at ±1kb around SAR-seq summit overlaid with SAR-seq signal (red).

Acknowledgments

Author Contributions

W.W., S.E.H., W.J.N., J.P., K.W.C., M.E.W., and A.N. designed the study. S.E.H., W.J.N., J.P., performed most of the experiments with assistance from K.S., J.C-M., E.C., R.P., D.W., H.-Y., S.C., M.P. on certain experiments. A.C. developed SAR-seq in the A.N. lab, W.W. designed bioinformatics pipelines and performed data analysis, and designed the figures. H.H., P.J.M and A.C provided insights. K.W.C., M.E.W., and A.N. analyzed and interpreted data and wrote the paper. W.W., S.E.H., W.J.N., J.P., helped edit the paper. M.E.W. and A.N. supervised the study.

References

1   Lodato, M. A. *et al.* Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555-559, doi:10.1126/science.aao4426 (2018).

2   Lodato, M. A. *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94-98, doi:10.1126/science.aab1785 (2015).

3   McKinnon, P. J. Genome integrity and disease prevention in the nervous system. *Genes Dev* **31**, 1180-1194, doi:10.1101/gad.301325.117 (2017).

4   Rodin, R. E. *et al.* The Landscape of Mutational Mosaicism in Autistic and Normal Human Cerebral Cortex. *bioRxiv*, https://doi.org/10.1101/2020.1102.1111.944413 (2020).

5   Caldecott, K. W. Single-strand break repair and genetic disease. *Nat Rev Genet* **9**, 619-631, doi:10.1038/nrg2380 (2008).

6   Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, 644-656, doi:10.1016/j.cell.2017.01.002 (2017).

7   Setlow, R. B. Repair. Different basic mechanisms in DNA repair. *Arch Toxicol Suppl* **3**, 217-228, doi:10.1007/978-3-642-67389-4_16 (1980).

8   Swain, U. & Subba Rao, K. Study of DNA damage via the comet assay and base excision repair activities in rat brain neurons and astrocytes during aging. *Mech Ageing Dev* **132**, 374-381, doi:10.1016/j.mad.2011.04.012 (2011).

9   Fernandopulle, M. S. *et al.* Transcription Factor-Mediated Differentiation of Human iPSCs into Neurons. *Curr Protoc Cell Biol* **79**, e51, doi:10.1002/cpcb.51 (2018).

10   Wang, C. *et al.* Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. *Stem Cell Reports* **9**, 1221-1233, doi:10.1016/j.stemcr.2017.08.019 (2017).

11   Macheret, M. & Halazonetis, T. D. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **555**, 112-116, doi:10.1038/nature25507 (2018).

12   Tubbs, A. *et al.* Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse. *Cell* **174**, 1127-1142 e1119, doi:10.1016/j.cell.2018.07.011 (2018).

13   van der Raadt, J., van Gestel, S. H. C., Nadif Kasri, N. & Albers, C. A. ONECUT transcription factors induce neuronal characteristics and remodel chromatin accessibility. *Nucleic Acids Res* **47**, 5587-5602, doi:10.1093/nar/gkz273 (2019).

14   Froimchuk, E., Jang, Y. & Ge, K. Histone H3 lysine 4 methyltransferase KMT2D. *Gene* **627**, 337-342, doi:10.1016/j.gene.2017.06.056 (2017).

15   Gupte, R., Liu, Z. & Kraus, W. L. PARPs and ADP-ribosylation: recent advances linking molecular functions to biological outcomes. *Genes Dev* **31**, 101-126, doi:10.1101/gad.291518.116 (2017).

16   Hanzlikova, H. & Caldecott, K. W. Perspectives on PARPs in S Phase. *Trends Genet* **35**, 412-422, doi:10.1016/j.tig.2019.03.008 (2019).

17   Gibson, B. A., Conrad, L. B., Huang, D. & Kraus, W. L. Generation and Characterization of Recombinant Antibody-like ADP-Ribose Binding Proteins. *Biochemistry* **56**, 6305-6316, doi:10.1021/acs.biochem.7b00670 (2017).

18   Madabhushi, R. *et al.* Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* **161**, 1592-1605, doi:10.1016/j.cell.2015.05.032 (2015).

19   Suberbielle, E. *et al.* Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-beta. *Nat Neurosci* **16**, 613-621, doi:10.1038/nn.3356 (2013).

20   Canela, A. *et al.* DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Mol Cell* **63**, 898-911, doi:10.1016/j.molcel.2016.06.034 (2016).

21   Caldecott, K. W. XRCC1 protein; Form and function. *DNA Repair (Amst)* **81**, 102664, doi:10.1016/j.dnarep.2019.102664 (2019).

22   Hanzlikova, H., Gittens, W., Krejcikova, K., Zeng, Z. & Caldecott, K. W. Overlapping roles for PARP1 and PARP2 in the recruitment of endogenous XRCC1 and PNKP into oxidized chromatin. *Nucleic Acids Res* **45**, 2546-2557, doi:10.1093/nar/gkw1246 (2017).

23   Caldecott, K. W. DNA single-strand break repair. *Exp Cell Res* **329**, 2-8, doi:10.1016/j.yexcr.2014.08.027 (2014).

24   Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104**, 239-255 e212, doi:10.1016/j.neuron.2019.07.014 (2019).

25   Beard, W. A., Horton, J. K., Prasad, R. & Wilson, S. H. Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism. *Annu Rev Biochem* **88**, 137-162, doi:10.1146/annurev-biochem-013118-111315 (2019).

26   Hoch, N. C. *et al.* XRCC1 mutation is associated with PARP1 hyperactivation and cerebellar ataxia. *Nature* **541**, 87-91, doi:10.1038/nature20790 (2017).

27   O'Connor, E. *et al.* Mutations in XRCC1 cause cerebellar ataxia and peripheral neuropathy. *J Neurol Neurosurg Psychiatry* **89**, 1230-1232, doi:10.1136/jnnp-2017-317581 (2018).

28   Gu, B. *et al.* Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050-1055, doi:10.1126/science.aao3136 (2018).

29   Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816, doi:10.1016/j.cell.2018.10.042 (2018).

30   Lio, C. J. *et al.* TET methylcytosine oxidases: new insights from a decade of research. *J Biosci* **45** (2020).

31   Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930, doi:10.1126/science.1169786 (2009).

32   Steinacher, R. *et al.* SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation. *EMBO J* **38**, doi:10.15252/embj.201899242 (2019).

33   Watanabe, S. *et al.* MyoD gene suppression by Oct4 is required for reprogramming in myoblasts to produce induced pluripotent stem cells. *Stem Cells* **29**, 505-516, doi:10.1002/stem.598 (2011).

34   Akiyama, T. *et al.* Efficient differentiation of human pluripotent stem cells into skeletal muscle cells by combining RNA-based MYOD1-expression and POU5F1-silencing. *Sci Rep* **8**, 1189, doi:10.1038/s41598-017-19114-y (2018).

35   Selvaraj, S. *et al.* Screening identifies small molecules that enhance the maturation of human pluripotent stem cell-derived myotubes. *Elife* **8**, doi:10.7554/eLife.47970 (2019).

36   Pawlowski, M. *et al.* Inducible and Deterministic Forward Programming of Human Pluripotent Stem Cells into Neurons, Skeletal Myocytes, and Oligodendrocytes. *Stem Cell Reports* **8**, 803-812, doi:10.1016/j.stemcr.2017.02.016 (2017).

37   Guo, X. *et al.* In vitro Differentiation of Functional Human Skeletal Myotubes in a Defined System. *Biomater Sci* **2**, 131-138, doi:10.1039/C3BM60166H (2014).

38   Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661, doi:10.1016/j.cell.2014.09.029 (2014).

39   Farias, G. G., Britt, D. J. & Bonifacino, J. S. Imaging the Polarized Sorting of Proteins from the Golgi Complex in Live Neurons. *Methods Mol Biol* **1496**, 13-30, doi:10.1007/978-1-4939-6463-5_2 (2016).

40   Kirwan, P., Jura, M. & Merkle, F. T. Generation and Characterization of Functional Human Hypothalamic Neurons. *Curr Protoc Neurosci* **81**, 3 33 31-33 33 24, doi:10.1002/cpns.40 (2017).

41   Wong, N., John, S., Nussenzweig, A. & Canela, A. END-seq: An Unbiased, High-Resolution, and Genome-Wide Approach to Map DNA Double-Strand Breaks and Resection in Human Cells. *Methods Mol Biol* **2153**, 9-31, doi:10.1007/978-1-0716-0644-5_2 (2021).

42   Bredemeyer, A. L. *et al.* DNA double-strand breaks activate a multi-functional genetic program in developing lymphocytes. *Nature* **456**, 819-823, doi:10.1038/nature07392 (2008).

43   Santos, M. A. *et al.* DNA-damage-induced differentiation of leukaemic cells as an anti-cancer barrier. *Nature* **514**, 107-111, doi:10.1038/nature13483 (2014).

44   Canela, A. *et al.* Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. *Mol Cell* **75**, 252-266 e258, doi:10.1016/j.molcel.2019.04.030 (2019).

45   Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

46   Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

47   Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

48   Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

49   Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

50   Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

51   Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

52   Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958, doi:10.1093/bioinformatics/btp340 (2009).

53   Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354, doi:10.1038/s41598-019-45839-z (2019).

54   Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774, doi:10.1101/gr.135350.111 (2012).

55   Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

56   Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

57   Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696-1697, doi:10.1093/bioinformatics/btr189 (2011).

58   Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207, doi:10.1093/bioinformatics/btq351 (2010).

59   Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1101/gr.229102 (2002).

60   Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

Figure 1

Figure 2

Figure 3

a

	Control	Etoposide	No primary

PAR

25μM

5μM

b

ACBD6

SAR-seq — 3

PAR ChIP-seq — 0.4

XRCC1 ChIP-seq — 1.5

100 kb

c

SAR-seq   XRCC1 ChIP-seq   PAR ChIP-seq

high

low

SAR-seq intensity

-1 Summit +1   -1 Summit +1   -1 Summit +1 (kb)
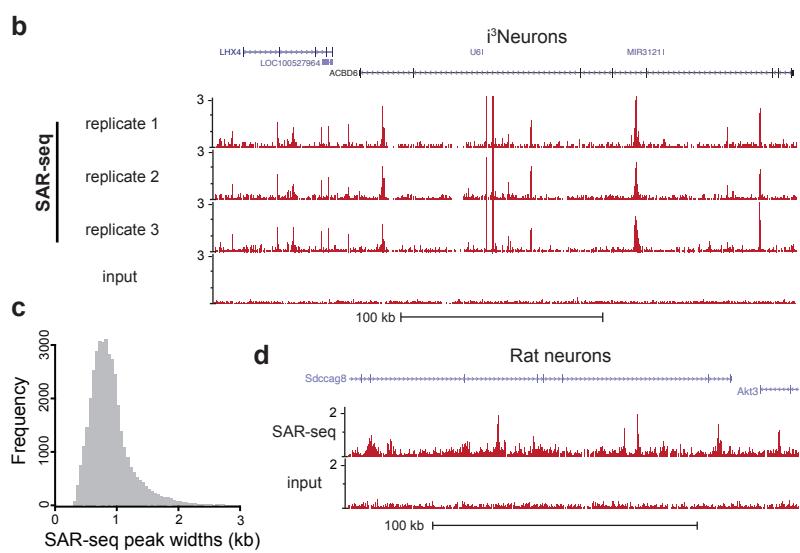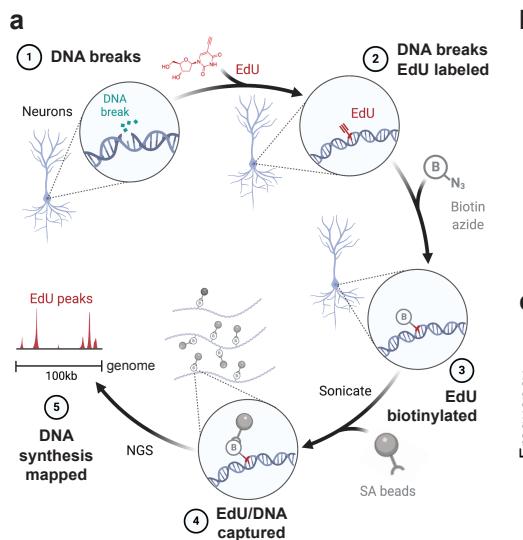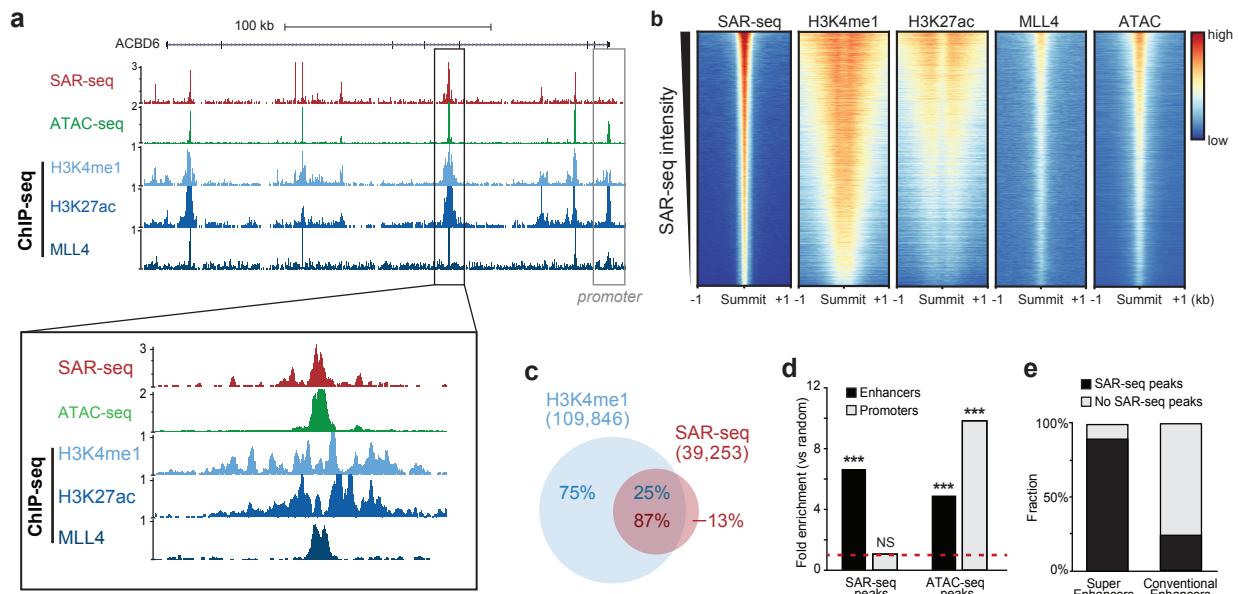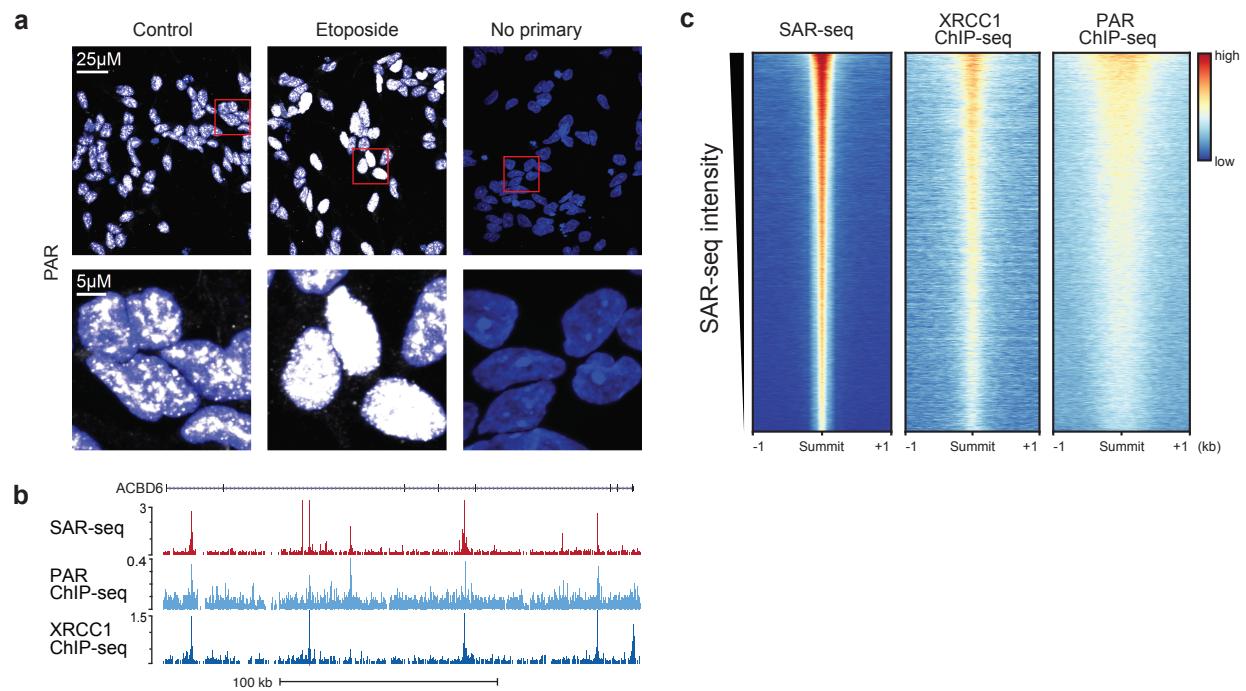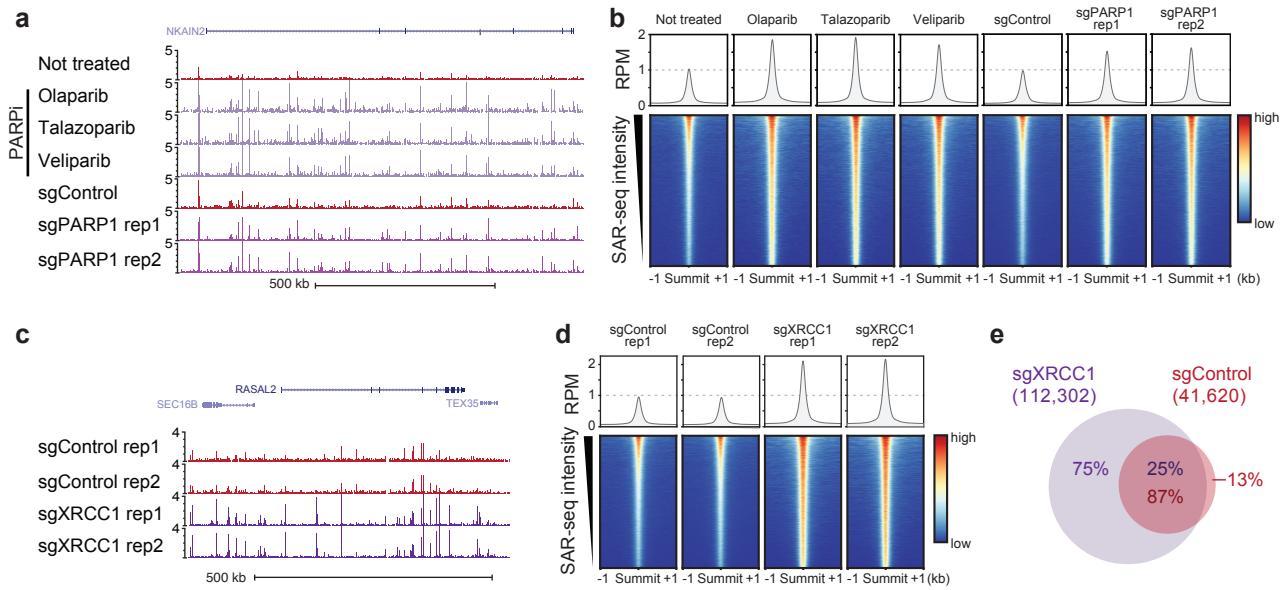
Figure 4

Figure 5

a

b

c

Extended Data Figure 2

a

LHX4
ACBD6

1 hr
2 hrs
4 hrs
8 hrs
18 hrs

100 kb

b

EdU labeld fraction
(relative to 18h EdU Incorporation)

EdU labeling (hours)

$$1 - e^{-k*h}$$
$$k = 0.19\ h^{-1}$$

c

LHX4
ACBD6

iMuscle rep1
iMuscle rep2

100 kb

d

Actr3
AsisI site

AsisI +Dox
AsisI -Dox

2 kb

e

Promoter
Exon
Intron
Distal intergenic

4%
10%
30%
56%

f

SAR-seq intensity
(RPKM)

Gene Expression
(FPKM)

g

LHX4
ACBD6

Regular SAR-seq

Strandness SAR-seq
rep1
rep2
rep3

100 kb

Extended Data Figure 3

a

SAR-seq best motif

E-value: 2.0e-338
Sites: 2,834/5,000

ONECUT1 motif

Motif probability

Position of best sites in sequences

b

ATAC-seq (65,082)   SAR-seq (39,253)

67%   33%   43%
57%

c

R=0.43

Widths of SAR-seq peaks (bp)
Widths of ATAC-seq peaks (bp)

d

Promoter
Exon
Intron
Distal intergenic

e

Pol II ChIP-seq   SAR-seq

POL II ChIPseq intensity

high
low

-1 TSS +1 -1 TSS +1 (kb)

Extended Data Figure 4

a

SAR-seq (rat)   H3K4me1

SAR-seq intensity

-1 Summit +1   -1 Summit +1 (kb)

high

low

b

Both iNeurons and iPSC
iNeuron only
iPSC only
None

11%

26%

2%

61%

c

-log10(FDR)

1 2 3 4 5 6 7 8

axon guidance
intracellular signal transduction
nervous system development
positive regulation of GTPase activity
glutamate receptor signaling pathway
positive regulation of synapse assembly
neuron migration
synapse assembly

d

H3K27ac signal at Enhancers

10000
8000
6000
4000
2000

n=1,385

0 2 4 6 8 (x10^4)

Enhancers ordered by H3K27ac signal

a

PAR ChIP-seq

-AsisI · 50

+AsisI · 50

+AsisI · 50
+ PARGi

| AsisI site

b

LRRC8D ▸—————— ZNF326 ▸▪◂◂

NT rep1 · 1.5
NT rep2 · 1.5
ETO rep1 · 1.5
ETO rep2 · 1.5

200 kb ├——————┤

c

d

	DAPI	53BP1	γH2AX	Merged
Not Treated				
+Etoposide				

e

LHX4 ▸—◂◂   ◂▪◂

ACBD6 ▸——————————————————————————————◂

SAR-seq · 3 ⌐
· 1 ⌐

END-seq · 1 ⌐
· -1 ⌐

100 kb ├——————┤

f

Extended Data Figure 6

Extended Data Figure 7