

Neuronal Enhancers are Hotspots For DNA Single-Strand Break Repair

Wei Wu^{1*}, Sarah E. Hill^{2*}, William J. Nathan^{1,3*}, Jacob Paiano¹, Elsa Callen¹, Dongpeng Wang¹, Kenta Shinoda¹, Niek Van Wietmarschen¹, Jennifer M. Colón-Mercado², Dali Zong¹, Raffaella de Pace⁴, Han-Yu Shih⁵, Steve Coon⁴, Maia Parsadonian², Raphael Pavani¹, Hana Hanzlikova^{6,7}, Solji Park^{8,9}, Seol Kyoung Jung^{8,9}, Peter J. McHugh³, Andres Canela¹⁰, Chongyi Chen¹¹, Rafael Casellas^{8,9}, 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 Eunice Kennedy Shriver National Institute of Child Health and Human Development

⁵ National Eye Institute, NIH, Bethesda, MD USA

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 Lymphocyte Nuclear Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases and National Cancer Institute, NIH, Bethesda, MD, USA; ⁹ NIH Regulome Project, NIH, Bethesda, MD, USA

¹⁰ The Hakubi Center for Advanced Research and Radiation Biology Center, Graduate School of Biostudies, Kyoto University, Kyoto, Japan

¹¹ Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, NIH, Bethesda, MD, USA

***Authors contributed equally to this work**

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

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 cellular genome is subjected to a constant barrage of endogenous DNA damage, but surprisingly little is known 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 post-mitotic neurons accumulate unexpectedly high levels of DNA single-strand breakage at specific sites within the genome. Genome-wide mapping reveals that these single-strand breaks (SSBs) are located within super-enhancers at or near to CpG dinucleotides and sites of DNA demethylation, and are repaired by PARP1 and XRCC1-dependent mechanisms. Notably, deficiencies in XRCC1-dependent short-patch repair elevate the extent of DNA repair synthesis at neuronal enhancers, whereas deficiencies in long-patch repair reduce synthesis, suggesting that the high steady-state level of SSB repair in neuronal enhancers is sustained by both short-patch and long-patch processes. These data provide the first evidence of site- and cell type-specific SSB repair, revealing unexpected levels of localized and continuous DNA single-strand breakage in neurons. In addition, these data suggest an explanation for the neurodegenerative phenotypes that occur in patients with defective SSB repair.

Long-lived neurons are particularly prone to defects in DNA repair^{1,2}. Recent studies have shown that single human neurons harbor large numbers of somatic mutations, and that the burden of mutations in neurons increases substantially during aging^{3,4}. Neuronal somatic mutations are especially abundant in patients with familial neurodegenerative diseases associated with dysfunctional DNA repair^{2,3}, as well as neurodevelopmental diseases such as autism spectrum disorder (ASD)⁵. 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 impact pathology in neurons including base modifications, single- and double-strand breaks and interstrand crosslinks^{1,2}. 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 alterations in this pathway lead almost exclusively to neuronal dysfunction and degeneration^{1,2}. 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 template⁶. If gap filling involves the incorporation of a sufficient number of nucleotides, DNA repair synthesis can be a robust measure of the extent and location of endogenous DNA damage⁷.

Regions of the neuronal genome are associated with ongoing DNA synthesis

We developed a 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 (i^3 Neurons^{8,9}) on day 6 post induction of neuronal differentiation with the thymidine analog EdU for 18 hours, biotinylated the labeled DNA by click chemistry, reduced it to 150-200 bp by sonication, and then isolated the biotinylated 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^{10,11} because the neurons are post-mitotic (**Extended Data Fig. 1c**) and the SAR-seq peaks were unaffected by inhibitors of the replicative DNA polymerase DNA polymerase α (PolA) (**Extended Data Fig. 1d-f**). In contrast, as expected, neuronal DNA repair synthesis was largely prevented by hydroxyurea (HU), which reduces the availability of deoxyribonucleotides (**Extended Data Fig. 1d, e**).

Given the absence of replicative DNA synthesis in the neurons, we reasoned that the SAR-seq peaks reflected regions of site-specific DNA repair. The SAR-seq peaks ranged from about 200-2,000 bp in width (averaging 1,006 bp) (**Fig. 1c**), with each peak presumably comprising multiple clustered sites of DNA repair rather than an individual site spanning ~1000 bp (see below). 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 after labeling cells for 18 hours (**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 induced 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 i³Neurons, we detected robust peaks of EdU incorporation at 22,196 specific sites in the rat neurons (**Fig. 1d and Extended Data Fig. 4e**). Thus, the high frequency of recurrent DNA synthesis appears to be a specific feature of post-mitotic neurons.

Recurrent sites of neuronal DNA synthesis are located at enhancers

Neuronal SAR-seq peaks were enriched in intragenic regions (**Extended Data Fig. 2e, f**) and further within expressed genes (**Extended Data Fig. 2f**), suggesting the involvement of transcription. However, SAR signal intensity did not correlate with transcript levels as measured by RNA-seq (**Extended Data Fig. 2fg**). 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 largely symmetrical strand-specific SAR-seq profiles (**Extended Data Fig. 2h**).

We next searched for specific DNA motifs using the strongest 5,000 SAR-seq peaks. More than 25% 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 neurons¹², we compared SAR-seq peaks with accessible regions using ATAC-seq. 54% of all SAR-seq regions coincided with ATAC-seq peaks (**Fig. 2a-c**) and the widths of SAR-seq and ATAC-seq peaks were correlated, suggesting that open chromatin structure influences the extent of DNA synthesis (**Extended Data Fig. 3b**).

Despite their localization in open chromatin, SAR-seq peaks were not enriched at promoters (**Fig. 2a,d and Extended Data Fig. 3c**). Indeed, promoters occupied by RNA polymerase II (POL II), as determined by ChIP-seq, exhibited only modest levels of DNA synthesis (**Extended Data Fig. 3d**). However, we detected a strong positive correlation between the location of DNA synthesis and of neuronal enhancers, as measured by ChIP-seq for H3K4me1 and H3K27ac (**Fig. 2a-c; Extended Data Fig. 4a**). Furthermore, SAR-seq peaks were centered on the nucleosome-free region occupied by MLL4, the major mammalian histone H3K4 mono-methyltransferase¹³ (**Fig. 2a, b**). Altogether, 87% of SAR-seq peaks overlapped with H3K4me1 peaks (**Fig. 2c**). SAR-seq peaks were also significantly enriched at H3K4me1/H3K27ac/MLL4/ATAC-positive regions compared to random sites, or to other epigenetic indicators of open or condensed chromatin such as H3K9me3, H3K27me3, H3K36me3 and CTCF binding sites (**Fig. 2c, d and Extended Data Fig. 4b, c**). Consistent with these results, SAR peaks in primary rat neurons also overlapped and correlated with rat H3K4me1 ChIP-seq (**Extended Data Fig. 4 d-f**).

We performed ultra-deep Hi-C in i³Neurons to examine chromosomal interactions associated with SAR-seq peaks within the context of topological associated domains (TADs) (**Extended Data Fig. 5a**). We observed that enhancers with SAR-seq peaks had greater number of intra-TAD interactions if they colocalized with SAR peaks (**Extended Data Fig. 5a**). Consistently, H3K27ac, a mark of active enhancers, was significantly enriched at enhancers with SAR-seq peaks (**Extended Data Fig. 5a**). The VISTA Enhancer Browser database contains experimentally validated human enhancer elements as assessed in transgenic mice¹⁴. Chromatin looping interactions detected by promoter capture Hi-C (pcHi-C) enables the linkage of distal enhancer elements to their target genes¹⁵. pcHi-C analysis of i³Neurons revealed enrichment of

the ONECUT1 binding motif at distal promoter interacting regions¹⁵. Compared to random sites, we observed an 8-fold enrichment of SAR sites among *in vivo* validated enhancers that overlapped with neuronal H3K4me1 histone marks and a 4-fold enrichment using pcHi-C data sets measured in i³Neurons (**Extended Data Fig 5b**)¹⁵. The latter included enhancers whose activity has been validated in human neuronal cells using CRISPR techniques (**Extended Data Fig 5c, d**). Taken together, these data indicate that distal-acting regulatory enhancers are hotspots of neuronal DNA synthesis.

Patterns of chromatin interactions are cell-type specific¹⁵. To determine whether the enhancers with SAR-seq peaks are specific to neurons, we compared sites of H3K4me1 in i³Neurons with those in the iPSCs from which they were derived. Only 2% of the SAR-seq peaks overlapped with iPSC-specific H3K4me1 peaks, while most of the SAR-seq peaks overlapped with H3K4me1 sites that were either neuron-specific or shared between neurons and iPSCs (**Extended Data Fig. 5e**). 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. 5f**). 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. 5g**). Collectively, these data identify enhancers and genes associated with neuronal function as hotspots of recurrent DNA synthesis.

PARP is active 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 strand breaks with neurodegeneration^{1,2}, 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) polymerases (PARPs) 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^{16,17}. PARP activity signals the presence of these lesions by modifying localized proteins with poly(ADP-ribose) (PAR), after which PAR is rapidly degraded by poly(ADP-ribose) glycohydrolase (PARG)^{16,17}. When we monitored ADP-ribosylation in individual neurons using an anti-ADP ribose-binding reagent,¹⁸ we detected focal sites of nuclear ADP-ribose, which were enhanced along with pan-nuclear staining by treatment with the genotoxin methylmethanesulfate (MMS) (**Fig. 3a**). To determine whether the sites of endogenous ADP-ribosylation were localized to sites of neuronal DNA synthesis we employed ADP-ribose ChIP-seq, which as a positive control we first confirmed could detect ADP-ribosylation occurring at site-specific DSBs (**Extended Data Fig. 6a**). Moreover, the ADP-ribose ChIP-seq signal was greatly enhanced by incubation with PARG inhibitor, confirming that this assay can identify genomic sites of poly(ADP-ribose) (**Extended Data Fig. 6a**). Importantly, when we employed this method in i³Neurons, we found that the endogenous sites of ADP-ribosylation co-localized with sites of DNA synthesis (**Fig. 3c, d**). These findings strongly suggest that the recurrent sites of DNA synthesis at neuronal enhancers are associated with sites of DNA strand break repair.

Neuronal DNA synthesis is not associated with DSBs

Neuronal activity has been reported to result in DSBs generated by topoisomerase 2 (TOP2)^{19,20}. TOP2-induced DSBs can promote the expression of early response genes¹⁹, associating these DNA breaks with regions of transcriptional activity^{21,22}. We found that treatment of i³Neurons with etoposide to trigger TOP2-induced DSBs resulted in DNA synthesis within gene bodies (**Extended Data Fig. 6b, c**). However, most of the sites of etoposide-induced DNA synthesis were distinct from those detected in untreated neurons (**Extended Data Fig. 6b, c**). In addition, we did not detect DSBs in unchallenged i³Neurons as measured by either γ -H2AX/53BP1 immunostaining (**Extended Data Fig. 6d**) or by END-seq (**Extended Data Fig. 6e**), which maps DSBs genome-wide at base-pair resolution²³. Thus, the sites of DNA synthesis in neuronal enhancers are independent of DSBs induced by TOP2 or by other sources.

Single-strand breaks (SSBs) at sites of neuronal DNA synthesis

In addition to DSBs, PARP1 and/or PARP2 are also activated at SSBs and subsequently recruit the XRCC1 protein complex necessary for SSB repair^{24,25}. We therefore examined the genomic localization of XRCC1 by ChIP-seq. Similar to sites of PARP activity, XRCC1 also co-localized with SAR-peaks, and the intensity of XRCC1 binding was correlated with the intensity of EdU incorporation in both human i³Neurons and primary rat neurons (**Fig. 3c, d; Extended Data Fig. 6f, g**). These data indicate that the sites of DNA synthesis at neuronal enhancers colocalize with sites of PARP activation and XRCC1-associated SSB repair.

To directly map SSBs at nucleotide resolution, we treated agarose-embedded i³Neuron plugs with recombinant single strand-specific S1 nuclease *in situ* to convert SSBs into DSBs, which we then detected using END-seq. To test whether this method would successfully convert

DNA nicks into DSBs, we first incubated the i³Neuron agarose plugs with the site-specific nicking endonuclease Nt.BspQI prior to S1 nuclease treatment. Nt.BspQI nicks were readily converted into DSBs by S1 (**Extended Data Fig. 7a**). Moreover, S1-END-seq could detect nicks generated by an inducible CAS9 nickase expressed in G1-arrested MCF10 cells (**Extended Data Fig. 7b,c**). However, S1-END-seq alone failed to detect endogenous SSBs in i³Neurons (**Extended Data Fig. 7a**).

We surmised that failure to detect endogenous SSBs in i³Neurons could be due to their rapid repair. To increase the half-life of SSBs, we incubated cells with a mixture of chain terminating dideoxynucleosides (ddA+ddC+ddT+ddG, denoted ddN). Strikingly, treatment of i³Neurons with ddN led to robust SSB accumulation as detected by S1-END-seq (**Fig. 4a**), and these SSBs colocalized extensively with SAR-seq peaks (**Fig. 4a-c**, and **Extended Data Fig. 7d**). Many of the SAR-seq peaks contained multiple S1-END-seq peaks (**Fig. 4a, inset**), each of which spanned an average of ~25 nucleotides (**Fig. 4d**), and thus were comprised of multiple-clustered single-strand gaps. Interestingly, we also detected a very low-level of DSBs at SAR-seq peaks following chain termination that were independent of S1 nuclease treatment, which most likely reflected closely apposed SSBs on opposite DNA strands (**Fig. 4c**).

Importantly, S1-END-seq enabled the identification of the location of the endogenous SSBs with much higher resolution than did SAR-seq, the resolution of which is limited to the size of the sonicated fragments (150-200 nt). Consequently, we were able to detect a prevalence of C/G nucleotides at the beginning of S1-END-seq reads (**Fig 4e**), and CpG dinucleotides were highly enriched at SAR sites (**Extended Data Fig. 7e**). Collectively, these data provide direct evidence for the site-specific formation of clustered SSBs in neuronal enhancers at or near C/G nucleotides.

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

SSB repair is comprised of short-patch and long-patch sub pathways, in which single or multiple nucleotides are replaced at the site of the SSB, respectively^{26,27}. PARP1 and XRCC1 promote the repair of a wide spectrum of SSBs, primarily by short-patch repair. We therefore examined the impact of inhibiting and/or depleting these proteins on neuronal DNA synthesis. We observed a reproducible increase in EdU incorporation at SAR sites if neurons were co-incubated with any of three independent inhibitors of PARP1, or if PARP1 was depleted using CRISPR interference (CRISPRi)²⁸ (**Fig. 5a and Extended Data Fig. 8a, b**). Depletion of XRCC1 similarly led to a prominent increase in EdU incorporation at sites of SAR (**Fig. 5b and Extended Data Fig. 8c-e**). These data suggest that if PARP1/XRCC1-dependent short-patch SSB repair is impeded, long-patch SSB repair is increased (**Fig. 6d**).

During short-patch SSB repair a single nucleotide is replaced at the site of the break^{26,27,29}. Nucleotide replacement typically requires DNA polymerase β (POL β) which interacts directly with XRCC1²⁴ (**Fig. 6d**). In contrast, during long-patch repair, alternative DNA polymerases such as POL ϵ and POL δ can generate a DNA repair patch the size of 2-20 nucleotides. This occurs through displacement of the lesion-containing strand to generate a 5'-flap that is then excised by the flap endonuclease 1 (FEN1), prior to DNA ligation²⁷. We found that depletion of POL β by two independent CRISPRi sgRNAs resulted in a dramatic increase in DNA synthesis at neuronal enhancers (**Fig. 5c; Extended Data Fig. 8f-h**). In contrast, depletion of FEN1 decreased this DNA synthesis (**Fig. 5d and Extended Data Fig. 9a, b**). In addition, incubation with aphidicolin (Aph), an inhibitor of POL ϵ and POL δ , at concentrations (50 μ M) shown to inhibit cellular repair synthesis^{7,30} greatly reduced SAR in both wild type and POL β -

depleted neurons (**Fig. 5e and Extended Data Fig. 9c**). Collectively, these data indicate that neuronal site-specific SSBs are repaired by both short-patch and long-patch repair, with the latter pathway being the primary source of SAR.

Localized neuronal SSBs are associated with sites of cytosine methylation and hydroxylation.

The detection of DNA repair synthesis specifically at neuronal enhancers could indicate that the latter are sites of increased DNA damage or, alternatively, that they are sites of preferential DNA repair. However, our finding using S1-END-seq that SSBs are enriched at neuronal enhancers strongly supports the first hypothesis. To test this further, we treated i³Neurons with MMS to introduce SSBs stochastically across the genome, with the prediction that if DNA repair is targeted to enhancer sites the introduction of additional SSBs randomly across the genome should not affect the location of the SAR-seq peaks. However, whilst MMS increased, as expected, the overall level of EdU incorporation 2-5 fold it ablated the appearance of SAR-seq peaks, suggesting that it is the elevated level of SSBs at neuronal enhancers that accounts for the elevated level of DNA repair synthesis at these sites (**Extended Data Fig. 10a**).

What is the source of the localized SSBs? Oxidative DNA lesions such as 8-oxoguanine (8-OxoG) are strongly implicated sources of DNA damage in brain, but unlike the sites of DNA repair synthesis detected here such lesions have not been shown to preferentially accumulate in open chromatin or at enhancers³¹. In contrast, enhancers may be especially vulnerable to SSBs induced by TOP1, a topoisomerase enzyme that is implicated in their activation^{32,33}. However, such SSBs are also unlikely to be the source of the localized SSBs identified here, because depletion of the polynucleotide kinase 3'-phosphatase (PNKP) enzyme that is recruited by and

required with XRCC1 for the repair of TOP1-induced SSBs^{34,35}, had only a small impact on the SAR-seq peaks (**Extended Data Fig. 10 b,c**). This was in contrast to the DNA repair synthesis associated with *bona fide* TOP1-induced SSBs that we triggered by treating neurons with the TOP1 poison camptothecin, which instead were located in gene bodies and were increased by PNKP depletion (**Extended Data Fig. 10d**).

Given our finding that SSBs associated with SAR-seq peaks are enriched at C/G nucleotides (**Fig. 4e and Extended Data Fig. 7e**), we favor the possibility that these SSBs are sites of cytosine demethylation. Consistent with this idea, active DNA demethylation of cytosine at CpG sites occurs preferentially at enhancers³⁶, is 10-fold more active in post-mitotic neurons than peripheral cell types³⁷, and generates SSBs that are intermediates of XRCC1-associated, but not PNKP-associated, BER^{36,38}. Active demethylation via ten-eleven translocation (TET) family (TET1, TET2, and TET3) enzymes is initiated through progressive oxidation of 5mC to 5hmC, 5fC, or 5caC, and steady state levels of 5hmC account for approximately 40% of modified cytosines in the brain³⁷. Using highly efficient labeling methods to detect 5hmC and 5fC^{39,40}, we therefore mapped oxidized forms of 5-methylcytosine genome-wide in i³Neurons. We found that the sites of both DNA repair synthesis (measured by SAR-seq) and SSBs (measured by S1-END-seq) overlapped with peaks of 5hmC and 5fC, and that localized intensity of SSBs correlated with that of 5hmC and 5fC (**Fig. 6a-c and Extended Data Fig. 10e**). While future studies will be needed to determine the precise source(s) of SSBs at neuronal enhancers, our data implicate cycles of cytosine methylation and demethylation at neuronal enhancers as a potential source of neuronal site-specific DNA single-strand breakage.

Conclusions

Our study reveals that human post-mitotic neurons are subject to an unexpected level of localized DNA synthesis that 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 SSB repair, including the DNA polymerase ($\text{POL}\beta$) and DNA ligase (LIG3), activities that are central to this repair²⁴. Interestingly, patients with hereditary mutations in XRCC1 exhibit progressive neurodegeneration, suggesting that intact short-patch SSB repair is necessary for long-term neuronal functionality^{41,42}. One possible explanation for this is that elevated and/or aberrant activity of the SSB sensor protein PARP1 that results from inefficient XRCC1-dependent SSB repair triggers neuronal dysfunction and/or death²⁴. In addition, our current data raise the possibility that an increased dependency on long-patch DNA repair synthesis in XRCC1-defective cells increases the mutational burden in long-lived neurons, as a result of increased and localized genome turnover. Age-dependent accrual of mutations at neuronal enhancers as a result of increased DNA repair synthesis could thus lead to aberrant gene expression, resulting in neurological dysfunction over time.

In summary, we describe new methods that enable genome-wide mapping of sites of frequent single-strand breakage and DNA repair synthesis in post-mitotic neurons. Our findings identify enhancers as hotspots of endogenous DNA single-strand breakage and repair in human post-mitotic neurons, perhaps explaining why this DNA repair process is important for neurological functionality during development and normal ageing.

Note: During the preparation of this manuscript, we became aware of the closely related work of D. Reid *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 genomic loci in neurons are associated with ongoing DNA synthesis

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

Neurons grown in culture (1) are incubated with EdU to label sites of DNA repair synthesis (2).

The incorporated genomic EdU is then conjugated to biotin via click chemistry (3), sheared by sonication to fragments 150-200 bp and captured with streptavidin beads (4), and enriched DNA sequences PCR-amplified and subjected to next-generation sequencing.

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. Neurons were labeled with EdU for 18 hours and harvested on day 7 after induction of neuronal differentiation. The exon-intron structure of three genes including ACBD6 is shown above with exons represented by vertical tick marks. Genome coordinates: Chr 1: 180,200,000-180,500,000.

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

d) Genome browser screenshot of SAR-seq performed on rat primary neurons, and input DNA sequenced in parallel. The culture was co-incubated with 5μM aphidicolin to block DNA replication of S phase glial cells.

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

a) Genome browser screenshots of SAR-seq, ATAC-seq, and of H3K4me1, H3K27ac, and MLL4 ChIP-seq in i³Neurons. *Inset*, zoomed-in screenshot of the indicated region highlighting overlapping SAR-seq, ATAC-seq and MLL4 peaks and flanking H3K4me1 and H3K27ac peaks.

The grey box highlighted *top right* shows the absence of SAR-seq signal at a promoter (ACBD6). (See also quantitation in Fig. 2d and Extended Data Figs. 2e, f and 3c below).

Genome coordinates: Chr1: 180,250,000-180,500,000.

b) Heatmaps of SAR-seq, H3K4me1 ChIP-seq, H3K27ac ChIP-seq, MLL4 ChIP-seq and ATAC-seq signal ± 1 kb around SAR-seq peak summits in i³Neurons, ordered by SAR-seq intensity. Note that 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 peaks.

c) Venn diagram illustrating the overlap between H3K4me1 ChIP-seq, ATAC-seq and SAR-seq peaks in i³Neurons. Statistical significance of the overlaps between SAR-seq, H3K4me1 ChIP-seq and ATAC-seq peaks was determined using randomly shuffled datasets (N=1,000) and one-sided Fisher's Exact test (the p value for overlap between H3K4me1 ChIP-seq/SAR-seq peaks is $p < 2.2 \times 10^{-16}$ and for ATAC-seq/SAR-seq peaks is $p < 2.2 \times 10^{-16}$).

d) Bar 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 (one-sided Fisher's Exact test, *** $p < 2.2 \times 10^{-16}$, NS: not significant).

e) Bar graph showing the fraction of super-enhancers (left) and conventional enhancers (right) that overlap with SAR-seq peaks. The super-enhancers (1,385) in the i³Neurons were defined by H3K27ac ChIP-seq intensity (see **Extended Data Fig. 5f**).

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

a) Left panel: Representative images of anti-ADP-ribose immunofluorescence staining (green) and neuronal marker Tubulin Beta 3 (red) in i³Neurons. Cells were counterstained with DAPI (blue). Cells were incubated with 10 μ M PARGi for 20 mins prior to fixation to preserve nascent

poly(ADP-ribose) (PAR). As a positive control for DNA damage and increased ADP ribosylation, cells were treated with 0.1 mg/ml MMS for 15 min. Boxed regions in upper panels are enlarged in lower panels.

b) Quantification of ADP ribosylation levels before and after MMS treatment. Statistical significance was determined using Mann-Whitney test (***, $p < 0.0001$).

c) Genome browser screenshots illustrating the overlap between SAR-seq, ADP-ribose ChIP-seq, and XRCC1 ChIP-seq signals in i³Neurons. Cells were incubated with PARGi for 20 minutes prior to fixation to preserve nascent poly(ADP-ribose). Genome coordinates: Chr1: 180,250,000-180,500,000.

d) Heatmaps of SAR-seq, XRCC1 ChIP-seq and ADP-ribose ChIP-seq signal ± 1 kb surrounding SAR-seq peak summits in i³Neurons, ordered by SAR-seq intensity.

Fig. 4: S1-END-seq following ddN incubation detects single-strand breaks

a) Genome browser screenshots illustrating the overlap in i³Neurons between sites of DNA repair synthesis (SAR-seq) and SSBs (S1-END-seq) detected in the presence of dideoxynucleosides (ddN) to block DNA ligation, or in the absence of dideoxynucleosides (“NT”). Positive- and negative-strand S1-END-seq reads are displayed in black and grey, respectively. *Insets*, zoomed screenshots demonstrating the presence of multiple SSBs within SAR-seq peaks. Genome coordinates: Chr1: 245,450,000-246,050,000.

b) Heatmaps of SAR-seq and S1-END-seq reads from i³Neurons with (ddN) or without (NT) incubation with dideoxynucleosides. Data are plotted ± 1 kb surrounding the SAR-seq peak summits and ordered by SAR-seq intensity.

- c) Aggregate plots of S1-END-seq (red: positive-strand reads, blue: negative-strand reads) and END-seq (black: positive-strand reads, grey: negative-strand reads) ± 500 bp around SAR-seq peak summits following ddN incubation.
- d) Distribution of the size of the gap between positive- and negative-strand S1-END-seq peaks in i³Neurons incubated with ddN. The median gap size is 25 bp (red dashed line). Positive-strand peaks mark the left end and negative-strand peaks mark the right end of a detected DSB.
- e) Composite DNA sequence motif within ± 5 bp surrounding the SSB summits on the positive strand (right panel) and on the negative strand (left panel) of the 10,000 most prominent S1-END-seq peaks.

Fig. 5: Localized SSB repair in neurons is comprised of short-patch and long-patch sub-pathways

- a) Heatmaps of SAR-seq intensities ± 1 kb surrounding SAR-seq peak summits for i³Neurons treated with the indicated PARP inhibitors or following non-targeted control (sgControl) or PARP1-targeted (sgPARP1) CRISPRi. Aggregate plots of SAR-seq intensity are shown in the top panel. NT: non-treated; rep: replicate.
- b) Heatmaps of SAR-seq intensities ± 1 kb surrounding SAR-seq peak summits for i³Neurons following non-targeted control (sgControl) or XRCC1-targeted (sgXRCC1) CRISPRi. Aggregate plots of SAR-seq intensity are shown in the top panel.
- c) Heatmaps of SAR-seq intensities ± 1 kb surrounding SAR-seq peak summits for i³Neurons following non-targeted (sgControl) or POLB-targeted (sgPOLB) CRISPRi. Aggregate plots of SAR-seq intensity are shown in the top panel.

d) Heatmaps of SAR-seq intensities ± 1 kb surrounding SAR-seq peak summits for i^3 Neurons following non-targeted (sgControl) or FEN1-targeted (sgFEN1) CRISPRi. Aggregate plots of SAR-seq intensity are shown in the top panel.

e) Heatmaps of SAR-seq intensities ± 1 kb surrounding SAR-seq peak summits for i^3 Neurons following non-targeted (sgControl) or POLB-targeted (sgPOLB) CRISPRi. Where indicated, neurons were treated with 50 μ M Aphidicolin (Aph) for 24 hours prior to EdU incorporation. Aggregate plots of SAR-seq intensity are shown in the top panel.

Fig. 6: Localised SSB repair in neurons correlates with sites of oxidized 5-methylcytosine

a) Genome browser screenshot illustrating the overlap in POLB-depleted i^3 Neurons between sites of 5hmC (5hmC-SEAL), 5fC (5fc-SEAL), SSBs (ddN S1-END-seq), DNA repair synthesis (SAR-seq), and H3K4me1 (ChIP-seq). Both positive (black) and negative (grey) S1-END-seq reads are shown. Genome coordinates: Chr1: 245,750,000-245,756,000.

b) Heatmaps of the genome-wide signals for DNA repair synthesis (SAR-seq), SSBs (ddN+S1-END-seq), 5fC (5fc-SEAL), and 5hmC (5hmC-SEAL) in POLB-depleted i^3 Neurons, plotted ± 1 kb surrounding the summits of the SAR-seq peaks and ordered by SAR-seq intensity.

c) Scatter plots showing correlations of intensities (RPKM) between SSBs (ddN+S1-END-seq) and 5fC (left) or 5hmC (right), respectively, at ± 1 kb surrounding SAR-seq peak summits for POL β -depleted i^3 Neurons. Spearman correlation coefficients are indicated.

d) Model depicting the balance between short-patch and long-patch SSB repair at neuronal enhancers.

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

- a) Venn diagram showing the overlap of SAR-seq peaks in i³Neurons for three independent biological replicates.
- b) Scatter plots showing correlations of SAR-seq intensities (SAR-seq reads per kilobase per million mapped reads, RPKM) between three replicates in i³Neurons. Pearson correlation coefficient is indicated.
- c) Left panel: Representative images of EdU-biotin staining (green) showing cell proliferation in iPSCs, but not in post-mitotic i³Neurons. i³Neurons were treated with EdU on day 3 or day 6 and fixed on day 7. iPSCs were treated with EdU for 24h and fixed. Cells were counterstained with DAPI (blue). Note that different imaging conditions were used for iPSCs and i³Neurons in the representative images. Right panel: quantification of EdU positive cells.
- d) Genome browser screenshot showing SAR-seq in i³Neurons treated with hydroxyurea (HU) or polymerase alpha inhibitor (POLAi). NT: non-treated.
- e) Scatter plots showing SAR-seq intensities (RPKM) for HU- (left) and POLAi- (right) treated compared to non-treated samples.
- f) Cell cycle profile of pre B cells treated or non-treated with POLAi. Cells were pulsed with EdU for 30 minutes before collecting cells for flow cytometry. Cells were counterstained with DAPI.

Extended Data Fig. 2: Genomic characteristics 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. Genome coordinates: Chr1: 180,200,000-180,500,000.
- b) Graph showing the fraction of EdU labeling in i³Neurons (relative to maximum labeling at 18 hours) as a function of time. Red data points and error bars represent mean and standard

deviation of the relative levels of EdU measured from experimental data. Black line represents the theoretical model after fitting, with k being the rate of EdU labeling.

c) Genome browser screenshot of chromosome 7 showing lack of localized DNA synthesis in two independent biological replicates of SAR-seq performed in iMuscle cells treated with EdU for 18 hours compared to SAR-seq in i³Neurons as well as input DNA.

d) Genome browser screenshot displaying SAR-seq peak at a representative AsiS1 restriction enzyme site (tick mark). AsiS1 expression was induced for 18 hours (+Dox) vs non-treated (-Dox) in G0-arrested, Abelson virus-transformed murine pre-B cells as described²³. Mouse genome coordinates: Chr1: 125,434,000-125,437,000.

e) Distribution of SAR-seq 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. Distal intergenic represents promoter-excluded intergenic regions.

f) Fold enrichment of SAR-seq peaks in intergenic regions, intragenic regions and expressed genes compared to 1000 sets of randomly shuffled regions of the same sizes and chromosome distribution (one-sided Fisher's Exact test, *** $p < 2.2 \times 10^{-16}$, NS: not significant).

g) Left panel: Scatterplot showing correlation of SAR-seq intensity (RPKM) with transcript level of genes containing SAR-seq peaks measured by RNA-seq (Fragments Per Kilobase of transcript per Million mapped reads, FPKM) in i³Neurons. 71% of SAR-seq peaks are at expressed genes (FPKM \geq 0.1; red dashed line: FPKM=0.1). Right panel: correlation of SAR-seq intensity with transcript levels of linked genes determined by pcHiC in i³Neurons (red dashed line: FPKM=0.1). Spearman correlation coefficients are indicated.

h) Genome browser screenshot comparing SAR-seq vs. strand-specific SAR-seq that discriminates which strand is labeled with EdU in i³Neurons. Both strands show labeling, in

triplicate. Strand-specific SAR-seq reads are separated by positive (black) and negative (grey) strands. Genome coordinates: Chr1: 180,200,000-180,500,000.

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

a) Motif analysis for sequence enrichment within ± 500 bp surrounding the summit of the top 5,000 SAR-seq peaks in i³Neurons. Upper panel; the best motif discovered by the MEME suite analysis. 1,384 out of 5,000 sites shared this motif. Middle panel: TOMTOM motif tool used to compare SAR-seq motif shown above with databases of known motifs. The transcription factor ONECUT1 was identified as the most similar motif and its consensus sequence is shown. P value for motif comparison with ONECUT1 motif is indicated. Bottom panel: position distribution of the best motif (upper panel) within ± 500 bp of the SAR-seq peak summit. The best motif is centered on the SAR-seq peak summit.

b) Scatter plot comparing widths of ATAC-seq peaks and SAR-seq peaks in i³Neurons. Pearson correlation coefficient and p value are indicated.

c) 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 1 kb upstream of transcription start sites and distal intergenic represents promoter-excluded intergenic regions.

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

Extended Data Fig. 4: The correlation between SAR-seq and chromatin features

a) Scatter plots showing the correlation between SAR-seq intensity and ATAC-seq, H3K4me1, H3K27ac and MLL4 ChIP-seq intensities (RPKM) ± 1 kb around the SAR-seq peak summits in i³Neurons. Spearman correlation coefficients and p values are indicated.

- b)** Fold enrichment of SAR-seq peaks at ATAC-seq peaks, ChIP-seq peaks of enhancer related marks (H3K4me1, H3K27ac and MLL4), additional chromatin marks at accessible regions (H3K4me3, H3K36me3 and CTCF) and chromatin silencing marks (H3K27me3 and H3K9me3) in i³Neurons. N=1,000 randomly shuffled datasets were generated to test the significance using one-sided Fisher's Exact test, *p<0.01, **p<0.001, ***p <2.2e⁻¹⁶, NS: not significant).
- c)** Heatmaps of SAR-seq and ChIP-seqs of chromatin markers at accessible regions (H3K4me3, H3K36me3 and CTCF) and chromatin silencing markers (H3K27me3 and H3K9me3) ±1kb around the SAR-seq peak summit in i³Neurons, ordered by SAR-seq intensity.
- d)** Heatmaps of SAR-seq and H3K4me1 ChIP-seq signal ±1kb around the SAR-seq peak summit in embryonic day 18 primary rat cortical neurons, ordered by SAR-seq intensity.
- e)** Venn diagram showing the overlap between H3K4me1 and SAR-seq peaks in rat primary neurons. N=1,000 randomly shuffled datasets were generated to test the significance using one-sided Fisher's Exact test: p <2.2e⁻¹⁶.
- f)** Scatter plots showing the correlation between SAR-seq and H3K4me1 ChIP-seq intensities (RPKM) ±1kb around the SAR-seq peak summits in rat primary neurons. Spearman correlation coefficient and p values are indicated.

Extended Data Fig. 5: SAR-seq enrichment at neuronal enhancers

- a)** Box plot showing Hi-C contacts (left) or H3K27Ac levels (right) associated with H3K4me1+SAR-seq+ (red) or H3K4me1+SAR-seq- (grey) elements.
- b)** Fold enrichment of SAR-seq peaks at *in vivo* validated enhancers from VISTA Enhancer Browser databases that overlap with H3K4me1 ChIP-seq peaks (left) or SAR-seq peaks within distal promoter-interacting regions determined by pcHiC (right). N=1,000 randomly shuffled

datasets were generated to test the significance using one-sided Fisher's Exact test, $**p < 0.001$, $***p < 2.2e^{-16}$).

c) Genome browser screenshots showing SAR-seq, H3K4me1 and H3K27ac ChIP-seq, and pcHiC and Hi-C profiles at representative enhancers (highlighted in orange) interacting with CDK5RAP3 promoter and **(c)** the DRD2 promoter. Both enhancers have been validated to promote transcription of their respective genes using CRISPR techniques in i³Neurons¹⁵. The CDK5RAP3 enhancer also overlaps with *in vivo* validated enhancers from VISTA Enhancer Browser database. Genome coordinates: **(b)** Chr17: 46,040,000-46,100,00 and **(c)** Chr11: 113,250,000-113,390,000.

d) Pie-chart showing distribution of i³Neuron SAR-seq peaks in iPSC-specific, i³Neuron-specific and shared iPSC- and i³Neuron-enhancers. Approximately 56 million and 49 million single end reads were sequenced for the H3K4me1 ChIP-seq in iPSC and i³Neurons, respectively, with approximately 100,000 peaks called in both cell types.

e) Top biological processes enriched for the genes containing the 2,000 most intense SAR-seq peaks determined by Gene Ontology (GO) analysis. The x-axis represents the enrichment value as the logarithm of False Discovery Rates (FDR).

f) 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). Accordingly, 1,385 enhancers were defined as super enhancers.

Extended Data Fig. 6: Mapping regions of DNA damage and repair in neurons

a) Anti-ADP-ribose ChIP-seq reads at an AsiSI restriction enzyme cut site (tick mark) in Abelson virus-transformed murine pre-B cells. Cells were arrested in G0, and AsiSI double-strand breaks were induced for 18 hours prior to ChIP. Note that ADP-ribose is enriched at

cleaved AsiSI sites and is increased by 20 min treatment with PARGi prior to fixation (AsiSI + PARGi), which is indicative of the presence of poly(ADP-ribose). Mouse genome coordinates:

Chr 1: 131,825,000-131,830,000.

b) Genome browser examples of SAR-seq profiles in non-treated (NT) or etoposide (ETO)-treated (18 hours, 50 μ M) i³Neurons. Data are from two biological repeats. Genome coordinates:

Chr1: 90,200,000-90,550,000.

c) Heatmaps for SAR-seq in nontreated (NT) or etoposide- (ETO) treated (18 hours 50 μ M) i³Neurons at -2kb to +5kb of the transcription start sites (TSS) ordered by ETO SAR-seq intensity.

d) Immunofluorescence staining of DSB markers γ -H2AX (red) and 53BP1 (green) in non-treated or one-hour ETO treated i³Neurons.

e) Genome browser showing SAR-seq and END-seq profiles in non-treated i³Neurons. Note that END-seq, which detects DSBs specifically²³, does not detect any enriched signal (i.e. above background) at SAR-seq peaks. END-seq reads are separated by positive (black) and negative (grey) strands. Chromosome coordinates: Chr1: 180,200,000-180,500,000.

f) Heatmaps of SAR-seq and XRCC1 ChIP-seq \pm 1kb around SAR-seq peak summits in cultured rat primary neurons, ordered by SAR-seq intensity.

g) Scatter plots showing the correlation between SAR-seq and XRCC1 ChIP-seq intensities (RPKM) \pm 1kb around the SAR-seq peak summits in i³Neuron (left) or rat primary neurons (right), respectively. Spearman correlation coefficients and p values are indicated.

Extended Data Fig. 7: S1-END-seq mapping of SSBs

a) Genome browser screenshot showing profiles of SAR-seq and S1-END-seq. Agarose plugs were incubated with or without (NT) the restriction enzyme Nt.BspQ1 prior to S1 treatment.

Zoomed-in view of Nt.BspQ1 sites (tick mark) display S1-END-seq detection upon Nt.BspQ1 treatment. S1-END-seq reads are separated by positive (black) and negative (grey) strands.

Genome coordinates: Chr1: 181,100,000-181,550,000.

b) FACS profile of G1-arrested MCF10A cells pulsed with EdU. For G1 arrest, MCF10A cells were treated with Palbociclib (1 μ M) for 20 hours to arrest cells in G1. Right panel:

c) Genome browser screenshot showing S1-END-seq profiles at three Cas9 nickase targeting sites (tick marks: sgRNAs 1-3) in the G1-arrested MCF10A cells treated with doxycycline (+Dox) to induce Cas9 expression. S1-END-seq reads are separated by positive (black) and negative (grey) strands. Genome coordinates: Chr19: 2,055,000-2,057,000.

d) Venn diagram showing the overlap between S1-END-seq peaks incubated with ddN and SAR-seq peaks in i³Neurons. N=1,000 randomly shuffled datasets were generated to test the significance using one-sided Fisher's Exact test: $p < 2.2e-16$.

e) Left panel: Aggregate plots showing the distribution of CG dinucleotides (black) at ± 1 kb surrounding SAR-seq peak summits overlaid with SAR-seq signal (red).

f) Aggregate plots showing the distribution of CG dinucleotides (black) at ± 1 kb around SAR-seq peak summits or summits of ATAC-seq peaks (green) that are H3K4me1 positive but do not overlap with SAR-seq.

Extended Data Fig. 8: PARP, XRCC1 or POLB deficiency increases SAR

a) Quantitative RT-PCR analysis showing PARP1 mRNA transcript level in i³Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; sgPARP1: an sgRNA targeting PARP1), cultured in parallel with samples used for SAR-seq. ****: $p < 0.00001$

b) Genome browser screenshots displaying SAR-seq profiles in i³Neurons 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. NT: nontreated. rep: replicate. Genome coordinates: Chr6: 124,100,000-125,200,000.

c) Quantitative RT-PCR analysis showing XRCC1 mRNA transcript level in i³Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; sgXRCC1: an sgRNA targeting XRCC1), cultured in parallel with samples used for SAR-seq. ****: $p < 0.00001$

d) Genome browser screenshots of SAR-seq profiles in i³Neurons expressing CRISPRi non-targeting sgRNAs (sgControl) or targeting XRCC1 (sgXRCC1), in duplicate. Genome coordinates: Chr1: 177,800,000-178,600,000.

e) Venn diagram showing the overlap of SAR-seq peaks between i³Neurons expressing non-targeting sgRNA (sgControl) or targeting XRCC1 (sgXRCC1). N=1,000 random datasets were generated to test significance of overlap (one-sided Fisher's Exact test: $p < 2.2e^{-16}$).

f) Quantitative RT-PCR analysis showing POL β mRNA transcript levels in i³Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; POLB sg1 or sg2: two independent sgRNAs targeting POL β), cultured in parallel with samples used for SAR-seq. ****: $p < 0.00001$.

g) Genome browser screenshots of SAR-seq profiles in i³Neurons expressing CRISPRi non-targeting sgRNAs (sgControl) or targeting POLB (sgPOLB). Genome coordinates: Chr1: 180,200,000-180,500,000.

h) Western blot showing POLB protein levels in i³Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; POLB sg1 or sg2: two independent sgRNAs targeting POLB), cultured in parallel with samples used for SAR-seq.

Extended Data Fig. 9: Aphidicolin or FEN1 deficiency decreases SAR

a) Quantitative RT-PCR analysis showing FEN1 mRNA transcript level in i³Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; FEN1 sg1: an sgRNA targeting FEN1), cultured in parallel with samples used for SAR-seq. ****: p < 0.00001.

b) Genome browser screenshots of SAR-seq profiles of i³Neurons expressing CRISPRi non-targeting sgRNAs (sgControl) or targeting FEN1 (sgFEN1), from two biological replicates. Genome coordinates: Chr1: 174,800,000-175,900,000.

c) Genome browser screenshots of SAR-seq profiles from two biological replicates of i³Neurons expressing CRISPRi non-targeting sgRNAs (sgControl) or targeting POLβ (sgPOLβ). Cells were treated or non-treated (NT) with 50 μM Aph for 24 hours, and then also during incubation with EdU. Genome coordinates: Chr1: 180,200,000-180,500,000.

Extended Data Fig. 10: Localized SSB repair in neurons correlates with sites of oxidized 5-methylcytosine

a) Genome browser of chromosome 7 showing SAR-seq profiles and input DNA sequence from i³Neurons without MMS treatment (NT) or after treatment with 0.1mg/ml MMS for the final 15 min of an 18 hr incubation with EdU. After Streptavidin pull-down and PCR amplification, total DNA was quantified: NT rep1: 0.95 μg; NT rep2: 1.7 μg; MMS rep1: 3.8 μg; MMS rep2: 4.5 μg.

b) Quantitative RT-PCR analysis showing PNKP mRNA transcript level in i³Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; sg PNKP: an sgRNAs targeting PNKP), cultured in parallel with samples used for SAR-seq. ***: p < 0.0001.

c) Heatmaps of SAR-seq intensities ±1kb surrounding SAR-seq peak summits for i³Neurons expressing non-targeting sgRNA (sgControl) or sgRNA targeting PNKP (sgPNKP). Aggregate plots of SAR-seq intensity are shown in the top panel.

d) Heatmap of SAR-seq ± 1 kb surrounding the transcription start site (TSS) in i³Neurons, ordered by SAR-seq intensity. i³Neurons expressing CRISPRi non-targeting sgRNAs (sgControl) or targeting PNKP (sgPNKP) were either non-treated (NT) or treated with 25 μ M CPT during incubation with EdU. Aggregate plots of SAR-seq intensity are shown in the top panel.

e) Scatter plots showing i³Neurons expressing CRISPRi-mediated knockdown of POL β correlations between the intensities (RPKM) of DNA synthesis (SAR-seq) and SSBs (S1-END-seq + ddN) (left), 5fC (middle) or 5hmC (right), at ± 1 kb surrounding the peak summits of SAR-seq. Spearman correlation coefficients and p values are indicated.

Methods

iPSC cell culture

All induced pluripotent stem cell (iPSC) experiments used the WTC11 line, which was derived from a healthy human male participant and obtained from the Coriell cell repository. All policies of the NIH Intramural Research Program for the registration and use of this iPSC line were followed. iPSC culture was performed as described previously⁸. Tissue culture treated dishes were coated with hESC-qualified matrigel (Corning, REF 354277). Matrigel was removed and iPSCs were plated in Essential 8 Medium (E8; Thermo Fisher Scientific, Cat. No. A1517001) and 10 μ M ROCK inhibitor (RI; Y-27632; Selleckchem, Cat. No. S1049). iPSCs were maintained in a 37°C, 5% CO₂ incubator and fed every 1-2 days as needed. Cells were split using either accutase (Life Technologies, Cat. No. A1110501) for enzymatic dissociation into single cells or EDTA (0.5mM; Life Technologies, Cat. No. 15575020) for routine passaging. Media was supplemented with 10 μ M RI to promote survival during passaging. As necessary iPSCs were frozen in 90% ES-qualified fetal bovine serum (FBS) (Sigma Aldrich, Cat. No. ES-009-B)

and 10% DMSO (Mediatech Inc., Cat. No. 25-950-CQC), and then thawed rapidly at 37°C, followed by removal of FBS/DMSO and plating in E8+RI medium.

i³Neuron culture

The human iPSCs used in this study were previously engineered^{8,9} to express mouse Neurogenin-2 (NGN2) under a doxycycline-inducible promoter integrated at the AAVS1 safe harbor in the WTC11 background, +/- CAG-dCas9-BFP-KRAB at the CLYBL promoter²⁸. For neuronal differentiation, 20-25 million iPSCs were plated on day 0 onto a 15 cm plate in N2 media composed of knockout DMEM/F12 media (Life Technologies Corporation, Cat. No. 12660012) with N2 supplement (Life Technologies Corporation, Cat. No. 17502048), 1x GlutaMAX (ThermoFisher Scientific, Cat. No. 35050061), 1x MEM nonessential amino acids (NEAA) (ThermoFisher Scientific, Cat. No. 11140050), 10 µM ROCK inhibitor (Y-27632; Selleckchem, Cat. No. S1049), and 2 µg/mL doxycycline (Clontech, Cat. No. 631311). N2 media was changed once a day for two more days. Day 3 cells were replated onto freshly-prepared poly-L-ornithine- (PLO; 0.1 mg/ml; Sigma, Cat. No. P3655-10MG) coated dishes as follows: Cells were washed with PBS, dissociated with accutase for 10 minutes at 37°C, washed and plated in i³Neuron Culture Media: BrainPhys media (STEMCELL Technologies, Cat. No. 05790) supplemented with 1x B27 Plus Supplement (ThermoFisher Scientific, Cat. No. A3582801), 10 ng/mL BDNF (PeproTech, Cat. No. 450-02), 10 ng/mL NT-3 (PeproTech, Cat. No. 450-03), 1 mg/mL mouse laminin (Sigma, Cat. No. L2020-1MG), and 2 µg/mL doxycycline (Clontech, Cat. No. 631311). For 10 cm plates used in SAR-seq or CHIP-seq, 12-15 million neurons were plated. For 15 cm plates 30-45 million neurons were plated. For ibidi slides used in imaging experiments, 0.2 million neurons per well were plated. Unless otherwise noted, i³Neurons were fed on day 6 during a half media change and harvested on day 7. For i³Neurons

cultured beyond 7 days, half media changes were conducted three times per week. In some experiments pre-differentiated i³Neurons were frozen on day 3 in 90% fetal bovine serum (Sigma Aldrich, Cat. No. ES-009-B) and 10% DMSO (Mediatech Inc., Cat. No. 25-950-CQC), and then thawed rapidly at 37°C, followed by removal of FBS/DMSO and plating in i³Neuron Culture Media. We did not detect any differences for experiments where day 3 neurons were thawed or plated immediately following differentiation.

iMuscle culture

To generate skeletal muscle myoblasts from human iPSCs, we engineered a doxycycline-inducible vector harboring a MyoD1 transcription factor transgene immediately followed by a co-inducible short hairpin RNA targeting Oct4 and selection marker/fluorescent protein cassette (MyoD-O iPSCs), similar to previous methods⁴³. We used the PiggyBacTM system to facilitate delivery and genome integration of the transgene cassette. To increase transposase expression in iPSCs, we subcloned the PiggyBacTM transposase under a long version of the eF1a promoter, and co-transfected this transposase vector with the MyoD1 donor vector into iPSCs using lipofectamine stem (ThermoFisher Scientific, Cat. No. STEM00015). Transfection rates were approximately 55% based on fluorescence of a co-expressed reporter gene, and we were able to obtain 100% MyoD1-O iPSCs after puromycin selection of iPSCs harboring stable integration of the transcription factor cassette.

To induce myogenic differentiation, MyoD-O iPSCs were dissociated using accutase (37°C for 10 minutes) and resuspended in myogenic induction media (MIM) and plated with 3x10⁶ iPSCs per 10 cm dish. The MyoD-O iPSCs differentiation was modified from a recent report⁴⁴. The MIM contained DMEM/F12 HEPES (Gibco, Cat. No. 11-330-032) supplemented to a final concentration of 1 mM sodium pyruvate, 1x MEM nonessential amino acids (NEAA)

(Thermofisher Scientific, Cat. No. 11140050), 0.1 mM 2-mercaptoethanol (Gibco, Cat. No. 21985023), 10 µg/mL insulin (Roche , Cat. No. 11376497001), 2 µg/mL doxycycline and 10 µM RI.

After plating the MyoD-O iPSCs in MIM at day 0, fresh media was supplied daily until 80% confluency or more (day 6). Myoblast morphology was observed beginning at day 2 followed by myotube-like morphology at day 5. On day 6, the glycogen synthase kinase 3 (GSK-3) inhibitor, CHIR 99021, was administered for 48 hours to enhance myoblast differentiation and stimulate fusion (3 µM, Tocris, Cat. No. 4423) as reported by others^{45,46}. At day 9 post-doxycycline, the differentiating myoblasts were harvested.

Lenti-virus production for CRISPRi guide delivery

Lenti-X Human Embryonic Kidney (HEK) cells were plated onto dishes coated with poly-L-ornithine (0.1 mg/ml PLO) at a density of 10-15 million cells (10 cm dish) or 30-45 million cells (15 cm dish) per dish in warm DMEM, high glucose GlutaMAX™ Supplement media (Life Technologies, Cat. No. 10566024) with 10% FBS (Sigma, Cat. No. TMS-013-B) and then cultured overnight to achieve approximately 90% confluency. The next morning cells were transfected using Lipofectamine 3000 (Life Technologies, Cat. No. L3000150). For each transfection, 2.4 mL of room temperature Opti-MEM media (Fisher Scientific, Cat. No. 31985062) and 60 µL of Lipo 3K were combined and incubated at room temperature for 5-20 minutes. Then in another tube, 2.4 mL of room temperature Opti-MEM, 80 µL of P3000, 13.3 µg of psPAX2, 4.5 µg of pMD2G, 1.8 µg of pAdVantage, and 19.5µg of the lenti vector of interest were combined. The contents of the two tubes were combined and incubated at room temperature for 30 minutes. This mixture was then added dropwise to a plate of Lenti-X cells and cultured overnight. The next morning the media was changed with 36 mL of fresh warm

DMEM+Glutamax+10% FBS media supplemented with 36 μ L viral boost reagent (ALSTEM, Cat. No. VB100). Then 2-3 days later the media containing virus was collected, centrifuged to remove cell debris and the supernatant was concentrated using Lenti-X concentrator, following the manufacturer's protocol (Takara Bio, Cat. No. 631231). The viral pellet was gently dissolved in PBS at either 1:10 or 1:100. Concentrated virus was then aliquoted and stored at -80°C for future use.

CRISPRi plasmids

The sgRNAs used in this study were cloned into either the pU6-sgRNA EF1Alpha-puro-T2A-BFP vector (gift from Jonathan Weissman; Addgene #60955)^{28,47} or the pMK1334 CROPSeq vector (gift from Martin Kampmann; Addgene #127965)²⁸. Guides were driven under the mouse U6 promoter. The sgRNA sequences were as follows: non-targeting controls:

GTCCACCCTTATCTAGGCTA or GACCAGGATGGGCACCAACC; PARP1:

GGGTGCGGCGTGTTTCGGTGG; XRCC1: GGTAAGATATGGGGTCCGAG. POLB sg1:

GCGCCGGAGGGAGATCCCCA, POLB sg2: GCCAGCTTGAAGGAGGTACC, PNKP:

GCCAGGGCTTGCCCGTCCGA, FEN1: GACTGGCCCAAGGCTCACAG.

Puromycin selection for CRISPRi knockdown i³Neurons

To perform CRISPRi knockdown 1-4 million iPSCs were transduced with an aliquot of sgRNA-expressing virus immediately following an accutase split prior to cell attachment. Media was changed the next day to remove residual virus. Two days after transduction, iPSCs were split into accutase and plated at low density in E8+RI media containing 10 μ g/mL puromycin. The following morning cells were washed with PBS and given fresh E8 or E8+RI media. Cells were then expanded for 1-2 days before inducing differentiation into neurons. sgRNA knockdown efficiency was tested at the iPSC stage and confirmed in day 7 i³Neurons via QT-PCR.

Immunofluorescence and Microscopy

For imaging, i³Neurons were plated onto 96-well plates (0.05x10⁶ cells per well; Perkin Elmer, Cat. No. 6055302) or on μ -Slide glass bottom IBIDI slides (or 0.2x10⁶ per well; IBIDI). Prior to fixation, if indicated, cells were treated with 0.1 mg/ml MMS for 15 min, 10 μ M PARGi for 20 min, or 50 μ M Etoposide (ETO) for 1h. Cells were then washed with PBS and fixed in 4% paraformaldehyde in PBS for 15 minutes at room temperature. Cells were then washed 3 times in PBS, permeabilized in 0.5% Triton X-100 for 5 minutes at room temperature and blocked in 1% BSA/ 0.2% Triton in PBS for 1 hour at room temperature before incubation with primary antibodies. Primary antibodies and dilutions used were as follows: anti-poly-ADP-ribose (MABE1031, 1:500, Sigma-Aldrich), anti-53BP1 (1:1000, Novus Biologicals), anti-Phospho-Histone H2AX (1:5000, Millipore), and anti-tubulin β 3 (TUBB3, 1:5000, Biolegend, Cat. No. 801201). Immunofluorescence detection was achieved using fluorochrome-conjugated secondary antibodies (Invitrogen). EdU was visualized using Click-iT™ Plus EdU Cell Proliferation Kit for Imaging, Alexa Fluor™ 488 dye (Life Technologies, Cat. No. C10637). Finally, DNA was counterstained with DAPI (Thermo Fisher Scientific). Images were acquired on an inverted Nikon spinning-disk confocal microscope (Nikon Eclipse T1), using a 60x 1.40 NA oil-immersion objective.

Rat primary neuron cell culture

All animal procedures were conducted following the NIH Guide for the Care and Use of Laboratory Animals, under the Animal Study Proposal #19-011 approved by the NICHD Animal Care and Use Committee. Pregnant albino rats were delivered to our facility on day 17 of gestation. They were housed under a 12 h light-dark cycle for 24 h with access to food and water ad libitum. The following day, the animals were sacrificed by carbon dioxide inhalation followed

by decapitation prior to embryo extraction and preparation of neurons. Neurons from embryos of the same litter were pooled for each experiment.

Primary rat cortical neurons were prepared at embryonic day 18 (E18) as previously described⁴⁸. Rats at E18 were harvested and euthanized by decapitation. The brain was collected and meninges were removed, after which cortexes were isolated in sterile Hank's medium (Hanks' Balanced Salt Solution (HBSS), 20 mM HEPES, pH 7.5). Cortexes were then collected and treated with 0.25% trypsin (Gibco), and 100 µg/ml DNase (Roche) for 15 minutes at 37°C. One volume of adhesion medium (Dulbecco's Modified Eagle Medium (DMEM) without phenol red, 4.5 g/L glucose, 25 mM HEPES, 10% heat-inactivated horse serum (Gibco), 100 U/mL penicillin and 100 mg/mL streptomycin) was added to stop trypsin enzymatic action. The tissue was then disrupted mechanically by pipetting it through a 10 ml serological pipet. Cells were then strained through a 70 µm nylon filter (Corning) and centrifuged at 700 g for 10 minutes. The cell pellet was resuspended in 5 ml of adhesion medium and cells counted. 10-25 million cells were plated on 10 cm culture dishes previously coated with poly-L-lysine (Sigma) and 5 µg/mL laminin (Roche). After 2 hours, the neurons were adherent to the plate and the medium was changed to complete neurobasal medium (CNB) (neurobasal medium (Gibco), 1X B27 serum-free (Gibco), 4.5 g/L glucose, and 100 U/mL, penicillin-streptomycin (Gibco)) and supplemented with 5µM aphidicolin (Aph) to eliminate residual dividing cells.

SAR-seq

Neurons and iMuscle cells were incubated with 20 µM EdU for 18 hours, unless otherwise noted. Cells were harvested and fixed as follows. Cells were washed with PBS, incubated with accutase for 5-10 mins, collected with a cell scraper, pelleted at 200 × g for 5 minutes and resuspended in cold 0.1% BSA in PBS. Cold methanol was then added dropwise during slow

vortexing to 80% final concentration. Samples were kept on ice for 20 minutes and then stored at -20°C until processing.

Copper catalyzes azide-alkyne click chemistry. For biotin labeling via Click-iT reaction, cells were first washed 1x in PBS, permeabilized with 0.2% Triton-X100/PBS for 10 minutes on ice, and then washed 1x in PBS. Then the following were added in order: 3 mM copper sulfate (Sigma), 50 μ M biotin azide (ThermoFisher, Cat. No. B10184), and 1X Click-iT additive (ThermoFisher, Cat. No. C10424) for 2 hours shaking at room temperature. Cells were then washed one time in PBS and lysed in 50 mM Tris pH 8.0 with 1% SDS and Proteinase K overnight at 37°C. DNA was extracted with UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v) (Invitrogen) according to manufacturer's instructions, followed by 2.5:1 volume ethanol and 1:10 volume sodium acetate precipitation. DNA pellets were resuspended in TE buffer and sheared to 150-200 bp fragments using Covaris S220 sonicator at 10% duty cycle, 175 peak incident power, 200 cycles per burst, for 240 seconds. DNA was again precipitated by 2.5:1 volume ethanol and 1:10 volume sodium acetate and resuspended in TE buffer. Biotin-EdU fragments were pulled down using MyOne Streptavidin C1 Beads (ThermoFisher, Cat. No. 650-01). Before pulldown, 35 μ L of Dynabeads were washed two times with 1 mL 1X Wash and Binding buffer (1X W&B) (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 M NaCl, 0.1% Tween20) on a DynaMag-2 magnetic separator (Invitrogen, Cat. No. 12321D) and resuspended in 2X W&B (10 mM Tris-HCl pH8.0, 2 mM EDTA, 2 M NaCl, 0.2% Tween20). Equal volume of Dynabeads in 2X W&B were added to DNA in TE and incubated at 24°C shaking in a ThermoMixer C at 800 rpm for 30 minutes. Dynabeads bound to biotin-EdU fragments were washed three times in 1 mL of 1X W&B, two times in 1 mL EB, and one time in 1mL 1X T4 DNA Ligase Buffer (NEB). Dynabeads were resuspended in 50 μ L end-repair reaction mix (1X

T4 DNA Ligase Buffer, 0.4 mM of dNTPs, 2.7 U of T4 DNA polymerase (NEB), 9 U of T4 Polynucleotide Kinase (NEB), and 1 U of Klenow fragment (NEB)) and incubated at 24°C shaking at 800 rpm for 30 minutes. Dynabeads were washed one time in 1mL 1X W&B, two times in 1mL EB, and one time in 1 mL NEBuffer 2 (NEB) and resuspended in 50 µL of A-tailing reaction mix (1X NEB dA-Tailing Buffer and 20 U Klenow fragment exo- (NEB)), followed by incubation at 37°C shaking at 800 rpm for 30 minutes. Dynabeads were then washed again 1x in 1 mL NEBuffer 2 and resuspended in 115 µL of ligation reaction mix (1X Quick Ligase Buffer (NEB), 6000 U Quick Ligase (NEB), 5 nM annealed TruSeq truncated adapter) and incubated at 25°C shaking at 600 rpm for 20 minutes. Ligation reaction was stopped by adding 50 mM EDTA, and Dynabeads were washed three times in 1 mL 1X W&B, three times in 1 mL EB, and finally resuspended in 8 µL EB + 10 µL 2X Kapa HiFi HotStart Ready Mix (Kapa Biosciences). 10 mM primers 5'-CAAGCAGAAGACGGCATAACGA-GATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T-3' and 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T-3' (* indicates a phosphothiorate bond and a NNNNNN TruSeq index sequence) were added with 37 µL PCR reaction mix (20 µL 2X Kapa HiFi HotStart Ready Mix, 17 µL H₂O) for final volume of 60 µL. DNA was amplified using PCR program: 98°C, 45 seconds; 15 cycles [98°C, 15 seconds; 63°C, 30 seconds; 72°C, 30 seconds]; 72°C, 5 minutes. PCR products were separated from DynaBeads and cleaned using 1.8X volume AMPure Beads XP. 150-200 bp bands were isolated on 2% agarose gel and purified using QIAquick Gel Extraction Kit (Qiagen). Prior to sequencing using Illumina NextSeq 550 (75 bp single read), library concentrations were calculated by KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems).

To specifically sequence only EdU-incorporated strands of DNA (Strand-Specific SAR-seq), SAR-seq was followed exactly as above with additional steps prior to PCR amplification. After washing three times in 1 mL 1X W&B and three times in 1 mL EB post-ligation, Dynabeads were washed in 50 μ L 1X SSC buffer and resuspended in 20 μ L of 0.15 M NaOH for 10 minutes at room temperature to denature DNA strands. Beads were placed back on the DynaMag-2 magnetic separator and washed one time with 20 μ L 0.1 M NaOH, one time with 1 mL 1X W&B, two times with 1 mL EB, and resuspended in 8 μ L EB + 10 μ L 2X Kapa HiFi HotStart Ready Mix. Primers and PCR reaction mix were added as above.

In the indicated experiments, i^3 Neurons were treated with the following compounds: Aphidicolin (Aph; 5 μ M), etoposide (50 μ M), olaparib (1 μ M or 10 μ M), velaparib (10 μ M) or tazaparib (1 μ M) were added along with EdU at 18 hours prior to harvest. POLAi (1 μ M) or Aph (50 μ M) was added with EdU 14 hours or 24 hours prior to harvest, respectively. Hydroxyurea (HU; 10 mM) powder was dissolved fresh into water to make a 1 M stock each time it was used. HU was added on day 3 of differentiation and again on day 6 at 18 hours prior to harvest along with EdU for a total of 4 days of treatment. For dideoxynucleoside (ddN) chain termination, 5 μ M of each of ddA, ddT, ddG, and ddC chain-terminating nucleosides (20 μ M total) were added to i^3 Neuron culture for 18 hours prior to cell harvesting for END-seq and S1-END-seq experiments.

END-seq and S1-END-seq

To dissociate i^3 Neurons for use in END-Seq, we performed a papain dissociation protocol modified from a previous protocol⁴⁹. Papain (Worthington Biochemical Cat: LK003178) was dissolved into TrypLE™ Express Enzyme (1X) no phenol red and warmed at 37°C for 10 minutes. Then, day 7 i^3 Neurons on a 15cm plate were washed with PBS and treated with 5mL

papain/TRPLE for 1 minute at 37°C. Papain was removed with gentle pipetting and 5mL of trituration solution was added (30mL i³Neuron Culture Media, 10 µM ROCK inhibitor, and 1 vial of DNase freshly dissolved). Cells were collected and gently pipetted 3-10 times in a conical tube using a wide-bore 10mL pipet, being careful not to over-digest the sample. Cells were then washed with PBS, pelleted and resuspended in PBS containing 0.1% BSA and 0.5mM EDTA, and kept on ice. Cells were processed for END-seq as described⁵⁰. For S1-END-seq, cells were collected and embedded in 1% agarose plugs, lysed and digested with Proteinase K (1 hour at 50°C, followed by 7 hours at 37°C), washed with TE buffer, and then treated with RNase A for 1 hour at 37°C. Plugs were then washed in EB and equilibrated in S1 nuclease buffer (40 mM sodium acetate pH 4.5, 300 mM NaCl, 2 mM ZnSO₄) for 30 minutes. 1.8 U of S1 nuclease was added to 100 µL of S1 nuclease buffer per plug and incubated on ice for 15 minutes to allow for the enzyme to diffuse into the plug. The reaction mix was then placed at 37°C for 20 minutes before addition of EDTA (10 mM final concentration) to terminate the reaction. Finally, plugs were processed through the standard END-seq protocol.

AsiSI induction in pre-B cells

Abelson-transformed murine pre-B cells⁵¹ were retrovirally transduced with tetracycline-inducible ER-AsiSI⁵². Cells were arrested in G1 with 3 µM imatinib for 24 hours, followed by addition of 3 µg/mL doxycycline for 24 hours, and then further addition of 1 µM 4OHT for 18 hours to induce AsiSI nuclear localization, as previously described²³.

Construction of the Dox-Cas9-D10A nickase

Dox-inducible Cas9-D10A was constructed using isothermal assembly⁵³. Briefly, a plasmid encoding Dox-inducible Cas9 nuclease was obtained from Addgene. pCW-Cas9-Blast was a gift from Mohan Babu (Addgene plasmid # 83481; <http://n2t.net/addgene:83481>; RRID:

Addgene_83481). This plasmid was digested with NheI/BamHI and assembled with 2 PCR fragments (Nickase-P1, Nickase-P2) and transformed into competent cells. Gel extracted PCR fragments of Nickase-P1, Nickase-P2 were generated using the Q5 HotStart 2X mastermix with the primers listed below using Addgene-83481 as template.

<i>Fragment</i>	<i>Forward Primer</i>	<i>Reverse Primer</i>
Nickase-P1	GTCAGATCGCCTGGAGAATTG	tgC CAGGCCGATGCTGTACTTCT
Nickase-P2	AGAAGTACAGCATCGGCCTG Gca ATCGGCACCAACTCTGTGGG	TGCCTTGAAAAGGCGCAAC

MCF10A Cas9D10A inducible cell line

To produce the MCF10A Cas9-D10A inducible cell line, we infected MCF10A cells with lentivirus containing Dox-Cas9-D10A and cells were selected with 10 µg/mL blasticidin. Cas9-D10A expression was induced by 3 µg/mL doxycycline and confirmed by Western Blotting. Three guide RNAs (sequences: 5'-TGGGGCGTTTATCCGATGTC-3'; 5'-GCACTAGCCGGCCCGGACGT-3'; 5'-CCAGCCTGGTAGCGCCCCCA-3') were cloned into Lenti-Guide-NLS-GFP vector⁵⁴ and the MCF10A Cas9-D10A inducible cell line containing the three guide RNAs were selected with 2 µg/mL puromycin. For identifying the nicks by S1-END-seq, cells were arrested in G1 for 48 hours with 5 µM Palbociclib with doxycycline added during the last 24 hours to induce Cas9D10A, followed by cell harvesting and S1-END-seq processing.

ChIP-seq and Western blotting

15 million i³Neuron or rat neurons were fixed in 1% formaldehyde at 37°C for 10 minutes. The fixation reaction was quenched with glycine at a final concentration of 125 mM. Cells were spun down and washed twice with chilled PBS, and pellets were then snap frozen on dry ice and finally stored at -80°C until sonication. Sonication, immunoprecipitation, and library preparation steps were done as previously reported²¹. All antibodies were pre-conjugated to 40 µL of

magnetic Protein A beads prior to immunoprecipitations: H3K4me1 (5 µg, Abcam #8895); MLL4 (antibody courtesy of Kai Ge); H3K27ac (5 µg, Abcam #4729); H3K27me3 (5 µg, Millipore, Cat. No. 07-449); H3K9me3 (10 µg, Active Motif, Cat. No. 39765); CTCF (6 µl, Millipore, Cat. No. 07-729); RNA Polymerase II (8 µg, Abcam #26721); H3K36me3 (5 µg, Abcam Cat. No. ab9050); anti-pan-ADP-ribose (5 µg, Millipore-Sigma MABE1016); XRCC1 (2.6 µg, Novus, Cat. No. NBP1-87154). For Western blotting, cells were collected and lysed in a buffer containing 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 5% Tween-20, 0.5% NP-40, 2 mM PMSF, 2.5 mM β-glycerophosphate (all from Sigma-Aldrich) and protease inhibitor cocktail tablet (complete Mini, Roche Diagnostics). Equal amounts of protein were loaded into precast mini-gels (Invitrogen) and resolved by SDS-PAGE. Proteins were blotted onto a nitrocellulose membrane, blocked with Intercept (TBS) blocking buffer (LI-COR Biosciences) and incubated with the corresponding primary/secondary antibodies: anti-DNA polymerase β (1:1000, Millipore), anti-Tubulin (1:10,000, Sigma-Aldrich). Fluorescent secondary antibodies were used at a dilution of 1:15,000 (Li-Cor Biosciences).

Selective chemical labeling and capture of 5hmC and 5fC

5hmC-Seal was performed as previously described⁵⁵ with modifications. Briefly, 80 µg genomic DNA was resuspended in TE buffer and sonicated to 200 bp fragments using a Covaris S220 sonicator. Fragmented DNA was precipitated by ethanol and sodium acetate and resuspended in TE buffer. The selective 5hmC chemical labeling was performed in 100 µl glucosylation buffer (50 mM HEPES buffer pH 8.0, 25 mM MgCl₂) containing above fragmented DNA, β-GT (NEB, Cat. No. M0357), UDP-Azide-Glucose (Active Motif, Cat. No. 55020), and incubated at 37°C for 1.5 hr. After the reaction, DNA was cleaned up with QIAquick Nucleotide Removal Kit. The

labeled DNA was eluted with ddH₂O, after which 1mM DBCO-PEG4-Biotin (Click Chemistry Tools) was added and incubated at 37°C for 2 hours. Then, the biotin-labeled DNA was pulled down by C1 Streptavidin beads (ThermoFisher, Cat. No. 650-01) for 30 minutes at room temperature. The captured DNA fragments were processed for library construction as described in SAR-seq (above).

For 5fC-SEAL, we modified the previously described protocol¹⁵, which reduces 5fC to 5hmC using NaBH₄⁵⁶. We labeled the newly generated 5hmC (derived from 5fC reduction with NaBH₄) with an azide-modified glucose as described above for 5hmC-Seal. In brief, 80 µg of fragmented i³Neurons genomic DNA was incubated in 100 µl glucosylation buffer (50 mM HEPES buffer pH 8.0, 25 mM MgCl₂) containing unmodified UDP-Glucose (NEB, Cat No. M0357), and β-GT for 1.5 hours at 37°C. The labeled DNA was cleaned up with QIAquick Nucleotide Removal Kit. Then, an equal volume of freshly prepared NaBH₄ (Aldrich, Cat. No. 213462) solution was added to the glucose-blocked DNA solution. The reaction mixture was vortexed and incubated in a Thermomixer for 15 minutes at room temperature. The chemical labeling and capture were performed as described above for 5hmC-SEAL.

In situ Hi-C from hiPSC-derived i³Neurons

Two *in situ* Hi-C libraries were generated from 10 million cultured hiPSC-derived i³Neurons as described in ⁵⁷. Briefly, *in situ* Hi-C consists of 7 steps: (1) crosslinking cells with formaldehyde, (2) DNA digestion using MboI, (3) filling in and marking ends with Biotin, (4) proximity ligation, (5) DNA shearing, (6) pulling down the biotinylated ligation junctions with streptavidin beads, and (7) paired end sequencing. As quality control (QC), we confirmed efficient restriction, ligation and DNA shearing with an agarose DNA gel and for appropriate size selection in using Agilent 4200 TapeStation system after steps (5) and (6). For the final QC, we

performed 100 paired end sequencing on the Illumina Nextseq to assess quality of the libraries based on percent of Intra-chromosomal reads, long Range (>20Kb) reads, and Library Complexity. The HiC libraries were sequenced on 150 paired end sequencing using Illumina Novaseq 6000.

RNA extraction, RNA-seq and quantitative real-time PCR (RT-PCR)

To extract RNA, cells were plated on 6-well dishes, washed with PBS, and then 500 μ L of tri-reagent (Zymo research corporation, Cat. No. R2050-1-200) was added directly to the cells. The lysed cells were collected using a cell scraper. To isolate RNA, we used a Direct-zol RNA miniprep kit (Zymo Research Corporation, Cat. No. R2052), following manufacturer's instructions including the optional DNase step. For QT-PCR, total RNA was reverse transcribed with iSCRIPT Advanced cDNA Synthesis Kit (Bio-Rad, Cat. No. 1725037) following manufacturer's instructions. The resulting cDNA was diluted 10-fold and used for quantitative real-time PCR (qRT-PCR) analyses with specific primer and probe sets (Bio-Rad, ACTB qHsaCEP0036280; PARP1 qHsaCEP0052423; XRCC1 qHsaCIP0033686; POL β qHsaCEP0057881; FEN1 qHsaCEP0039485; PNKP qHsaCEP0057803 in a final volume of 20 μ L, which contained 10 μ L of SsoAdvanced™ Universal Probes Supermix (Biorad, Cat. No. 1725280), 3 μ L of cDNA. qRT-PCR was performed in triplicate wells per sample on a CFX96 Real-Time System (Biorad). For RNA sequencing, six biological replicates were sequenced. Total RNA was enriched for polyA and sequenced 2x75 bp on a HiSeq machine.

ATAC-seq

ATAC-seq was performed as described previously⁵⁸. i³Neurons were grown on 96-well plates (0.05x10⁶ cells per well). Cells were washed gently with PBS. Then, 100 μ L of accutase per well

was added and removed, and the plate was incubated at 37°C for 5 minutes. 50 µL cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPAL CA-630) was added directly to the well for 10 minutes. Cells were then pipetted 10-20 times to break clumps and centrifuged at 500 × g for 10 minutes at 4°C. The nuclei pellet was resuspended in the Nextera transposition reaction mix (25 µL 2x TD Buffer, 2.5 µL Nextera Tn5 transposase (Illumina, Cat. No. FC-121-1030), and 22.5 µL nuclease free H₂O) on ice, then incubated for 30 minutes at 37°C. The tagmented DNA was purified using the Qiagen MinElute kit and eluted with 10 µL Elution Buffer. Following purification, library fragments were amplified using the Nextera index kit (Illumina, Cat. No. FC-121-1011) under the following cycling conditions: 72°C for 5 minutes, 98°C for 30 seconds, followed by thermocycling at 98°C for 10 seconds, 63°C for 30 seconds, and 72°C for 1 minute for a total of five cycles. To prevent saturation due to over-amplification, a 5 µL aliquot was then removed and subjected to qPCR for 20 cycles for calculation of the optimal number of cycles needed for the 45 µL reaction that remained. The number of additional cycles required was determined by plotting linear R_n vs. Cycle and calculating the cycle number corresponding to a quarter of the maximum fluorescence intensity. Adding seven cycles to this estimate was found to yield optimal libraries. PCR reactions were subsequently cleaned with Agencourt AMPure XP beads (Beckman Coulter), run on a 2% agarose gel and a smear of 200-800 bp was cut and gel purified using QIAquick Gel Extraction Kit (QIAGEN). Library concentration was determined with KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). Sequencing was performed on the Illumina Nextseq500 (75 bp paired-end reads).

Genome alignment

SAR-seq, END-seq, ChIP-seq and ATAC-seq reads were aligned to the reference genome (hg19 for human i³Neuron and iMuscle, mm10 for mouse pre B cells or rn6 for rat primary neurons) using bowtie (v1.1.2)⁵⁹ with parameters -n 3 -l 50 -k 1 for END-seq and -n 2 -l 50 -m 1 for the rest. RNA-seq reads were aligned by STAR (v2.7.6a)⁶⁰. Functions “view” and “sort” of samtools (v 1.6)⁶¹ were used to convert and sort the aligned sam files to sorted bam files. Bam files were further converted to bed files by the bedtools bamToBed command⁶². Mitochondrial reads were removed in SAR-seq for intensity comparisons.

Peak calling

We used MACS (v1.4.3)⁶³ to call SAR-seq, XRCC1 ChIP-seq and ATAC-seq peaks. SAR-seq and XRCC1 ChIP-seq peaks with >10 fold-enrichment over background were kept. H3K4me1 and H3K27ac ChIP-seq peaks were called by SICER⁶⁴ with default parameters. Peaks within blacklisted regions (<https://sites.google.com/site/anshulkundaje/projects/blacklists>) were filtered⁶⁵. Overlapped SAR-seq peaks from three non-treated replicates shown in Figure 1b were used for most of the analyses in this paper.

Quantification for sequencing data

For SAR-seq, reads per kilobase of transcript, per million mapped reads (RPKM) was calculated. For RNA-seq, fragments per kilobase of transcript per million mapped reads (FPKM) was calculated by cufflinks⁴⁴ based on the annotation from GENCODE v33⁶⁶. Mean values of replicates were used for analyses.

Enhancer enrichment and super enhancer identification

Bedtools shuffle command with parameter (-chrom) was used to generate 1,000 random sets for SAR-seq peaks and ATAC-seq peaks, respectively, in order to estimate enrichments at enhancers and promoters. 1 kb regions upstream of transcription start sites from GENCODE v33 annotation

were defined as promoters in analyses. Super enhancers were separated from conventional enhancers by identifying an inflection point of H3K27ac signal versus enhancer rank⁶⁷. iPSC enhancers were identified by H3K4me1 ChIP-seq from an ENCODE iPSC cell line (GSM2527632).

Gene Ontology (GO) analysis

We used the DAVID web-tool⁶⁸ to find the GO terms enriched for genes containing the top 2,000 SAR-seq peaks. The most significant Gene Ontology Biological Process terms and their associated FDR values were reported.

Motif analysis

MEME-ChIP⁶⁹ of MEME suite was used to identify the common sequence motif of the nucleotide sequences from ± 500 bp around the summits for the top 5,000 SAR-seq peaks.

Kinetics Model for EdU labeling

The rate of EdU labeling was estimated under the assumptions that the fraction of a synthesis event at a given site is fixed with different EdU incorporation times and 18 hours is enough for EdU to incorporate all synthesis events at a given site. The fraction of EdU labeled synthesis event at h hours at a given site was denoted as $[\text{EdU}]_h = 1 - e^{-kh}$. k represents the rate of EdU labeling at a given site and $[\text{EdU}]_h$ was estimated by SAR-seq intensity at h hours divided by SAR-seq intensity at 18 hours. The intensity values (RPKM) of SAR-seq at different time points (1, 2, 4, 8 and 18 hours of EdU) was fitted to the function $[\text{EdU}]_h = 1 - e^{-kh}$ using the nls package in R. The k value shown in Extended Data Fig. 2b was derived using the top 2,000 SAR-seq peaks.

Hi-C analysis

Using Juicer software ⁷⁰, .hic files were generated, and normalized contact matrices and observed over expected normalized contact matrices were obtained with a dump command. Loop or domain calls were also done by juicer software ⁷⁰, and interaction matrices were visualized by Juicebox software. After processing two replicates, final map is generated by merging these two replicates. The final bin resolution of the Hi-C map was 1 kb. From 2.1B raw reads generated by together with QC and data-generation runs, we obtained 1.5B final total contacts which are over Q30.

For compartment analysis, PC1 eigen vector values are extracted in 50k resolution by juicer software, then sign of eigen vector in each chromosome was adjusted according to the H3K27ac, H3K4me1 and ATAC-seq peaks distribution pattern in whole chromosome. Among H3K4me1 peaks which were not overlapped with TSS within 1kb, peaks overlapping with SARseq peaks were extracted for the analysis (SAR+H3K4me1). 10,000 random sequence was generated by bedtools random program with 2600 nt length which is the average length of H3K4me1 peaks. While 40% of random sequences are located in A compartment, 63% of SAR+H3K4me1 peaks are located in A compartment.

Visualization

BedGraph files were generated by bedtools genomecov, normalized by reads per million (RPM) and then converted to bigWig files using bedGraphToBigWig from UCSC pre-compiled utilities for visualization at UCSC genome browser ^{71,72}.

Data matrices for heatmaps were calculated by computeMatrix and plotted by plotHeatmap of deepTools suite⁷³. Venn diagrams were plotted by VennDiagram package in R. Confocal images

were processed using FIJI. Schematics were created using BioRender, and figures compiled in Adobe Illustrator.

Quantification of Confocal Images

Confocal images were quantified using Nikon software. Images were first background subtracted. Then nuclei or EdU-positive cells were identified and counted using bright spot detection. Corresponding PAR signal intensity was measured per cell and reported as a mean. Data compilation and statistical analyses were performed using PRISM software.

Statistical Analysis

Statistical analysis was performed using R version 3.6.2 (<http://www.r-project.org>). The statistical tests are reported in the figure legend and main text.

Acknowledgments

We thank Anjana Rao, Michael Kruhlak, Yilun Sun, Yves Pommier, Marek Adamowicz, Paul Meltzer, and Kai Ge for helpful discussions and reagents; Raj Chari for constructing the CAS9D10A inducible nickase, and the CCR genomics core for help with sequencing. KWC is supported by Programme Grants from the UK Medical Research Council (MR/P010121/1), Cancer Research-UK (C6563/A7322), and by ERC Advanced Investigator (SIDSCA 694996) and a Royal Society Wolfson Research Merit Award. The M.W. laboratory is supported by the NINDS Intramural Research Program, the Chan Zuckerberg Initiative, and the Packard ALS Center.. S.E.H. received funding from the BrightFocus Foundation. The A.N. laboratory is supported by the Intramural Research Program of the NIH, an Ellison Medical Foundation Senior Scholar in Aging Award (AG-SS-2633-11), the Department of Defense Awards (W81XWH-16-1-599 and W81XWH-19-1-0652), the Alex's Lemonade Stand Foundation Award, and an NIH Intramural FLEX Award.

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., E.C., D.W., performed most of the experiments with assistance from K.S., J.C-M., N.V.W., D.Z., R.P., D.W., H-Y. S., S.C., M.P., R.P. and C.C. 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. S. P. performed Hi-C, and S. K. J. and R.C. analyzed the Hi-C data; 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 Caldecott, K. W. Single-strand break repair and genetic disease. *Nat Rev Genet* **9**, 619-631, doi:10.1038/nrg2380 (2008).
- 2 McKinnon, P. J. Genome integrity and disease prevention in the nervous system. *Genes Dev* **31**, 1180-1194, doi:10.1101/gad.301325.117 (2017).
- 3 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).
- 4 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).
- 5 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).
- 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 Miller, M. R. & Chinault, D. N. The roles of DNA polymerases alpha, beta, and gamma in DNA repair synthesis induced in hamster and human cells by different DNA damaging agents. *J Biol Chem* **257**, 10204-10209 (1982).
- 8 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).
- 9 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).
- 10 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).
- 11 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).
- 12 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).
- 13 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).
- 14 Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* **35**, D88-92, doi:10.1093/nar/gkl822 (2007).
- 15 Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* **51**, 1252-1262, doi:10.1038/s41588-019-0472-1 (2019).
- 16 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).
- 17 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).

- 18 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).
- 19 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).
- 20 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).
- 21 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).
- 22 Gomez-Herreros, F. *et al.* TDP2 suppresses chromosomal translocations induced by DNA topoisomerase II during gene transcription. *Nat Commun* **8**, 233, doi:10.1038/s41467-017-00307-y (2017).
- 23 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).
- 24 Caldecott, K. W. XRCC1 protein; Form and function. *DNA Repair (Amst)* **81**, 102664, doi:10.1016/j.dnarep.2019.102664 (2019).
- 25 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).
- 26 Caldecott, K. W. DNA single-strand break repair. *Exp Cell Res* **329**, 2-8, doi:10.1016/j.yexcr.2014.08.027 (2014).
- 27 Caldecott, K. W. Mammalian DNA base excision repair: Dancing in the moonlight. *DNA Repair (Amst)* **93**, 102921, doi:10.1016/j.dnarep.2020.102921 (2020).
- 28 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).
- 29 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).
- 30 DiGiuseppe, J. A., Hunting, D. J. & Dresler, S. L. Aphidicolin-sensitive DNA repair synthesis in human fibroblasts damaged with bleomycin is distinct from UV-induced repair. *Carcinogenesis* **11**, 1021-1026, doi:10.1093/carcin/11.6.1021 (1990).
- 31 Poetsch, A. R. The genomics of oxidative DNA damage, repair, and resulting mutagenesis. *Comput Struct Biotechnol J* **18**, 207-219, doi:10.1016/j.csbj.2019.12.013 (2020).
- 32 Bansal, K., Yoshida, H., Benoist, C. & Mathis, D. The transcriptional regulator Aire binds to and activates super-enhancers. *Nat Immunol* **18**, 263-273, doi:10.1038/ni.3675 (2017).
- 33 Puc, J. *et al.* Ligand-dependent enhancer activation regulated by topoisomerase-I activity. *Cell* **160**, 367-380, doi:10.1016/j.cell.2014.12.023 (2015).
- 34 Kalasova, I. *et al.* Pathological mutations in PNKP trigger defects in DNA single-strand break repair but not DNA double-strand break repair. *Nucleic Acids Res* **48**, 6672-6684, doi:10.1093/nar/gkaa489 (2020).

- 35 Whitehouse, C. J. *et al.* XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* **104**, 107-117, doi:10.1016/s0092-8674(01)00195-7 (2001).
- 36 Lio, C. J. *et al.* TET methylcytosine oxidases: new insights from a decade of research. *J Biosci* **45** (2020).
- 37 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).
- 38 Steinacher, R. *et al.* SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation. *EMBO J* **38**, doi:10.15252/embj.201899242 (2019).
- 39 Song, C. X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678-691, doi:10.1016/j.cell.2013.04.001 (2013).
- 40 Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci* **14**, 1607-1616, doi:10.1038/nn.2959 (2011).
- 41 Hoch, N. C. *et al.* XRCC1 mutation is associated with PARP1 hyperactivation and cerebellar ataxia. *Nature* **541**, 87-91, doi:10.1038/nature20790 (2017).
- 42 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).
- 43 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).
- 44 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).
- 45 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).
- 46 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).
- 47 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).
- 48 Farias, G. G., Britt, D. J. & Bonifacio, 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).
- 49 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).
- 50 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).
- 51 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).

- 52 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).
- 53 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).
- 54 Noordermeer, S. M. *et al.* The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**, 117-121, doi:10.1038/s41586-018-0340-7 (2018).
- 55 Cui, X. L. *et al.* A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat Commun* **11**, 6161, doi:10.1038/s41467-020-20001-w (2020).
- 56 Dai, Q. & He, C. Syntheses of 5-formyl- and 5-carboxyl-dC containing DNA oligos as potential oxidation products of 5-hydroxymethylcytosine in DNA. *Org Lett* **13**, 3446-3449, doi:10.1021/ol201189n (2011).
- 57 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).
- 58 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).
- 59 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).
- 60 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 61 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 62 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).
- 63 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
- 64 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).
- 65 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).
- 66 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).
- 67 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).
- 68 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).
- 69 Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696-1697, doi:10.1093/bioinformatics/btr189 (2011).

- 70 Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99-101, doi:10.1016/j.cels.2015.07.012 (2016).
- 71 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).
- 72 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1101/gr.229102 (2002).
- 73 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

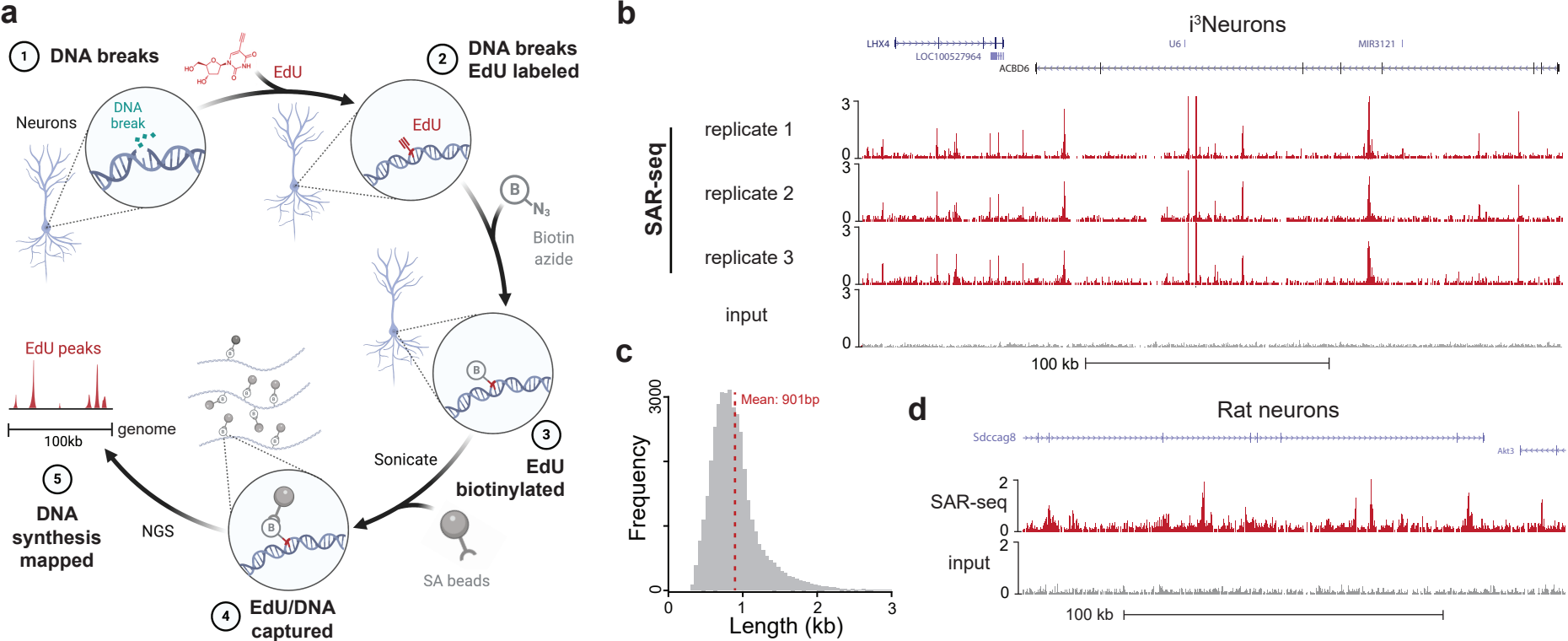


Figure 2

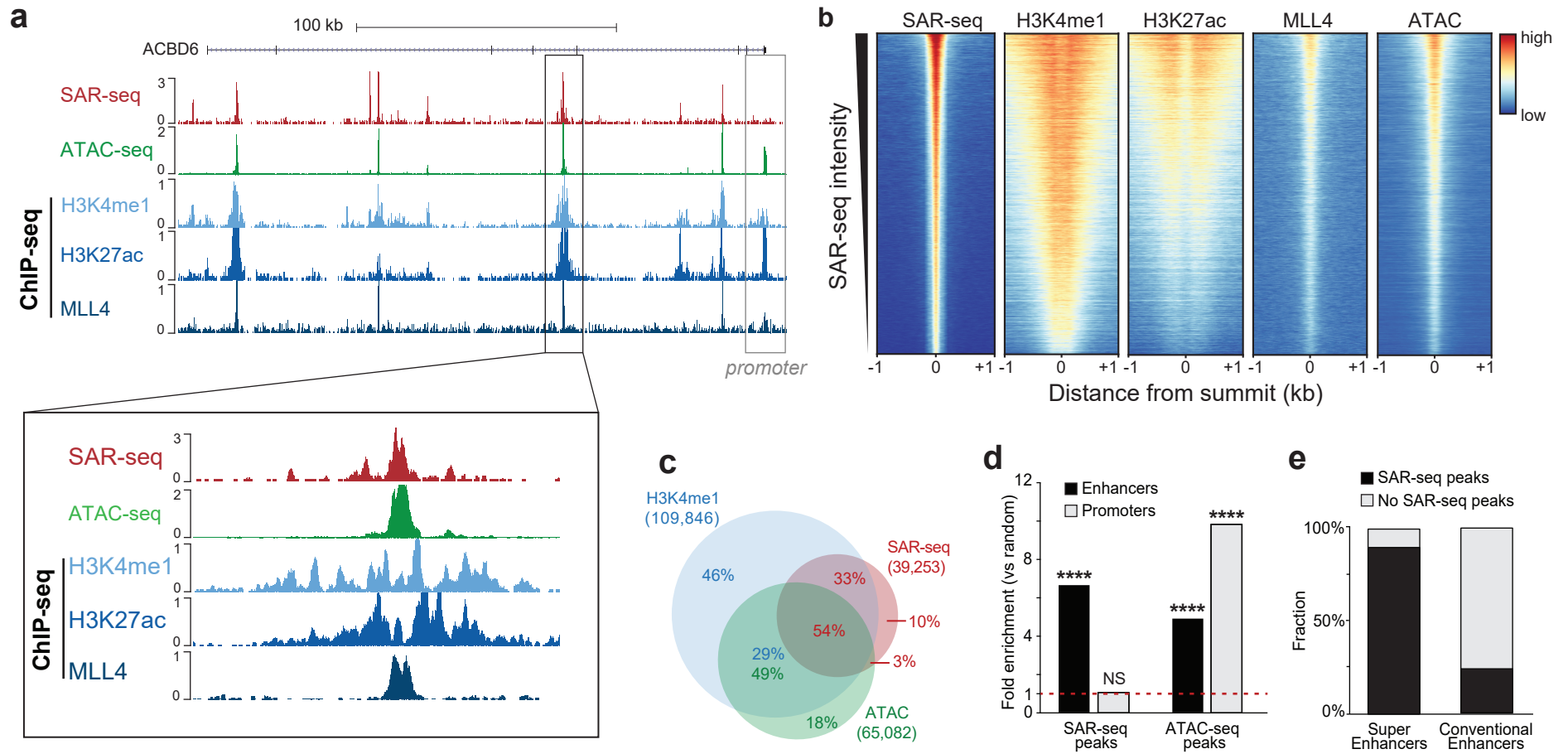


Figure 3

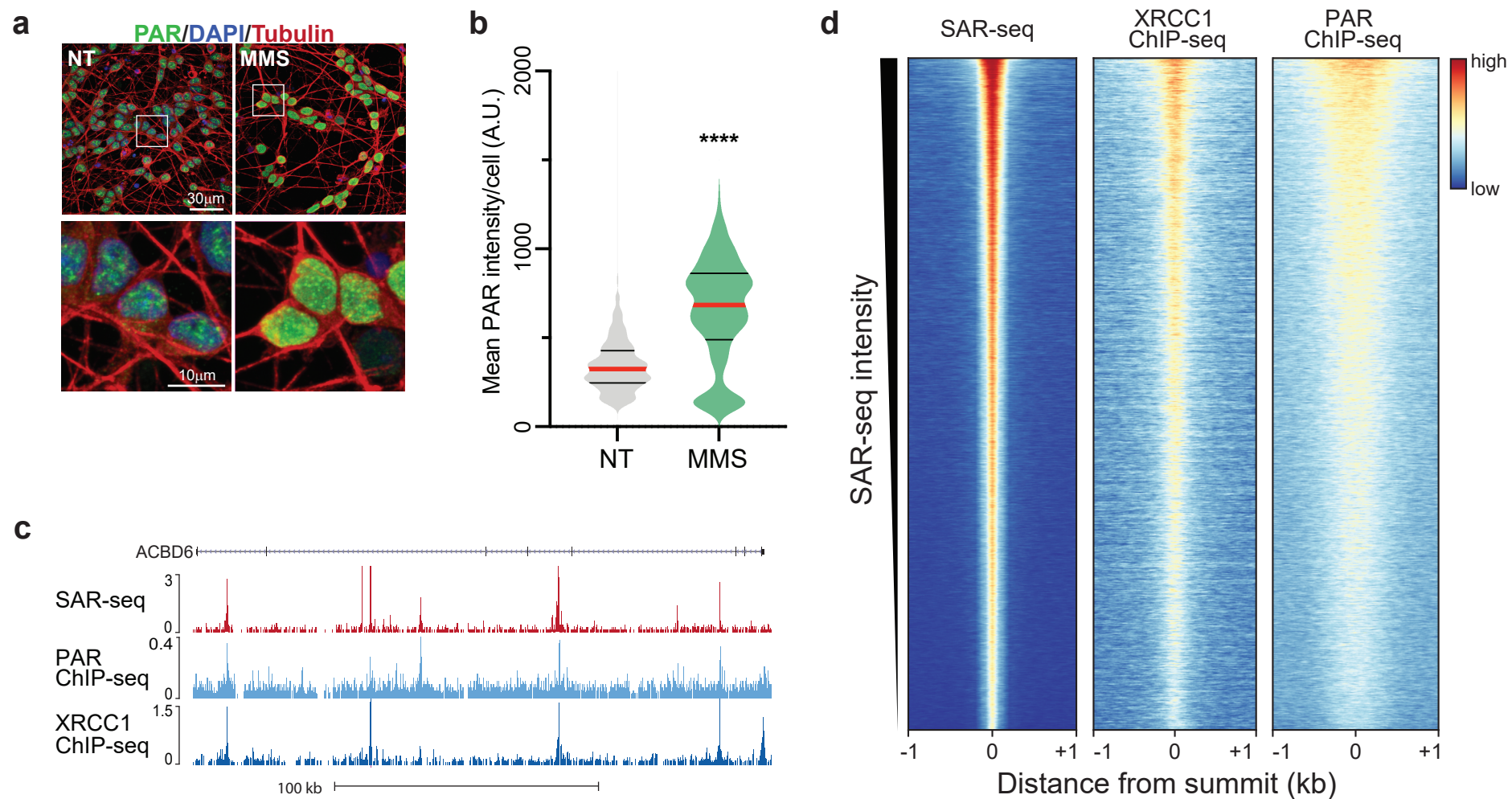


Figure 4

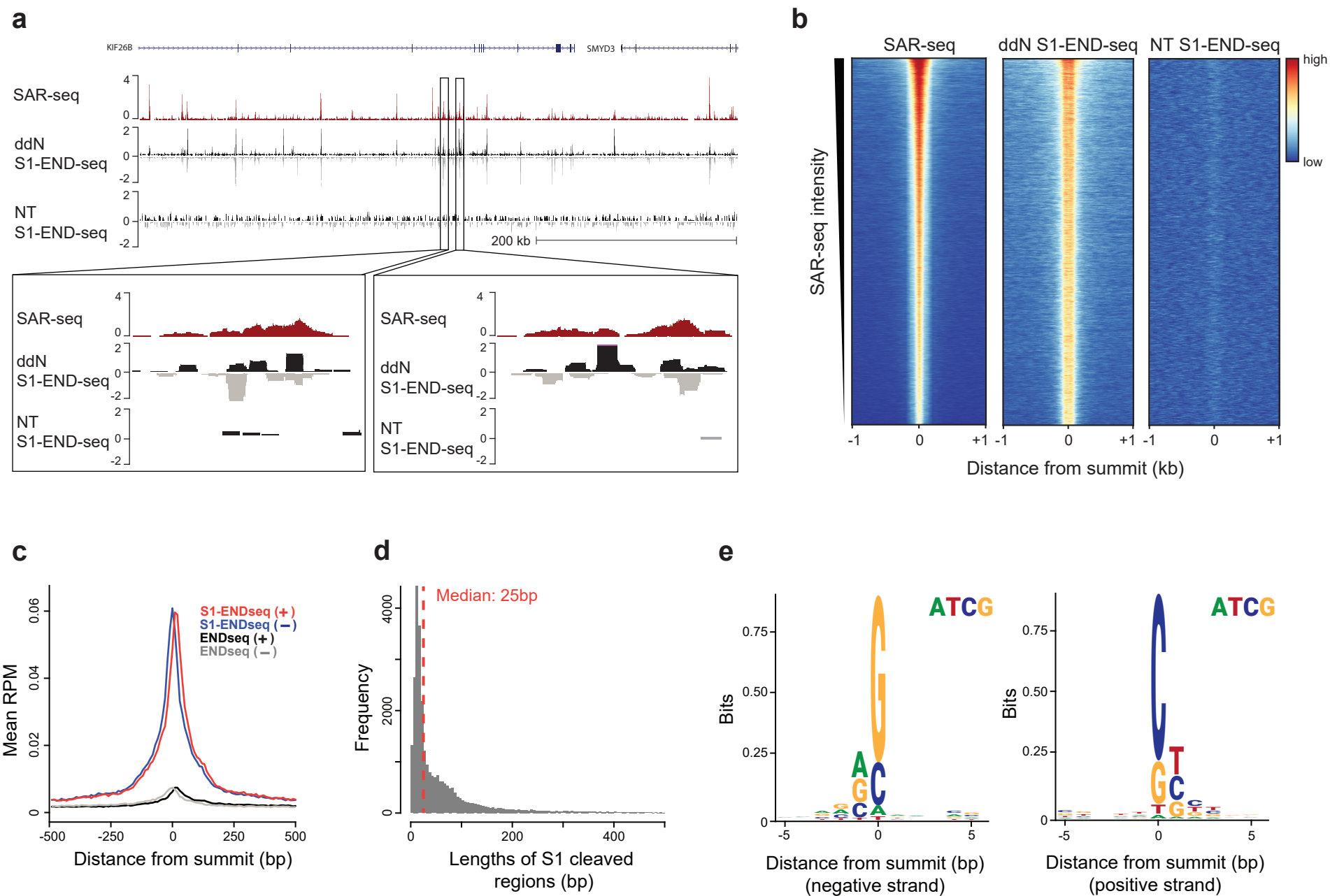


Figure 5

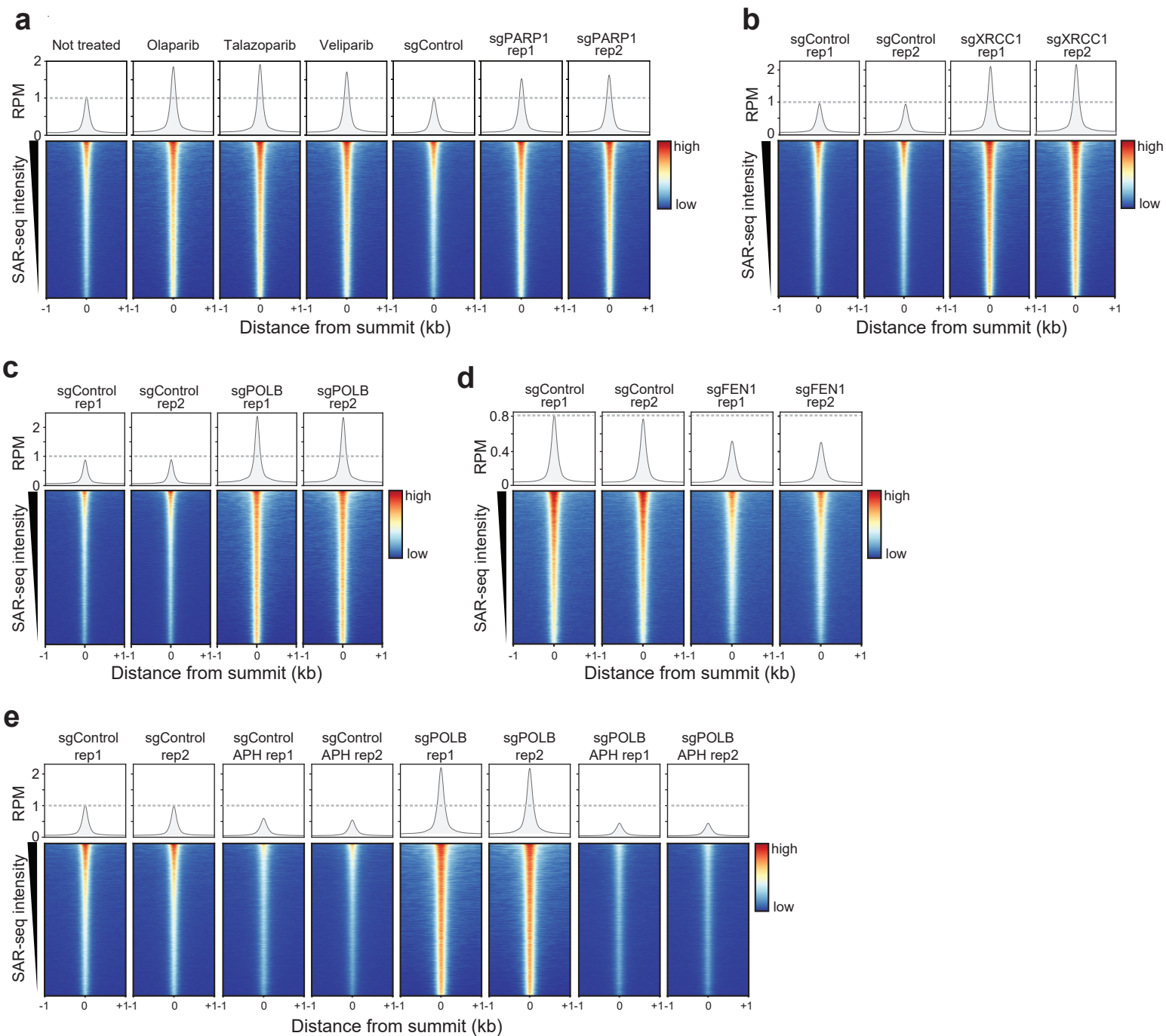
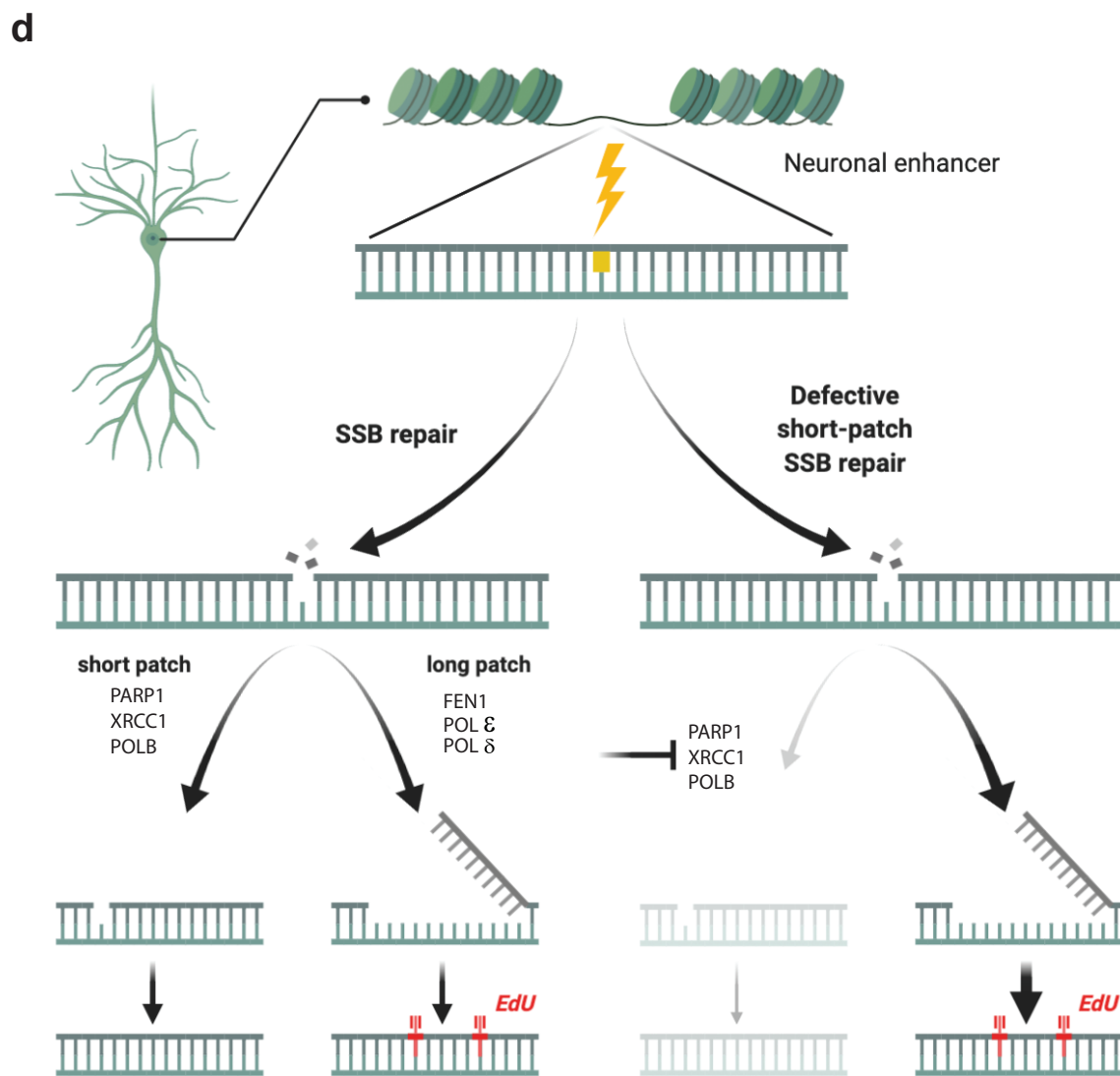
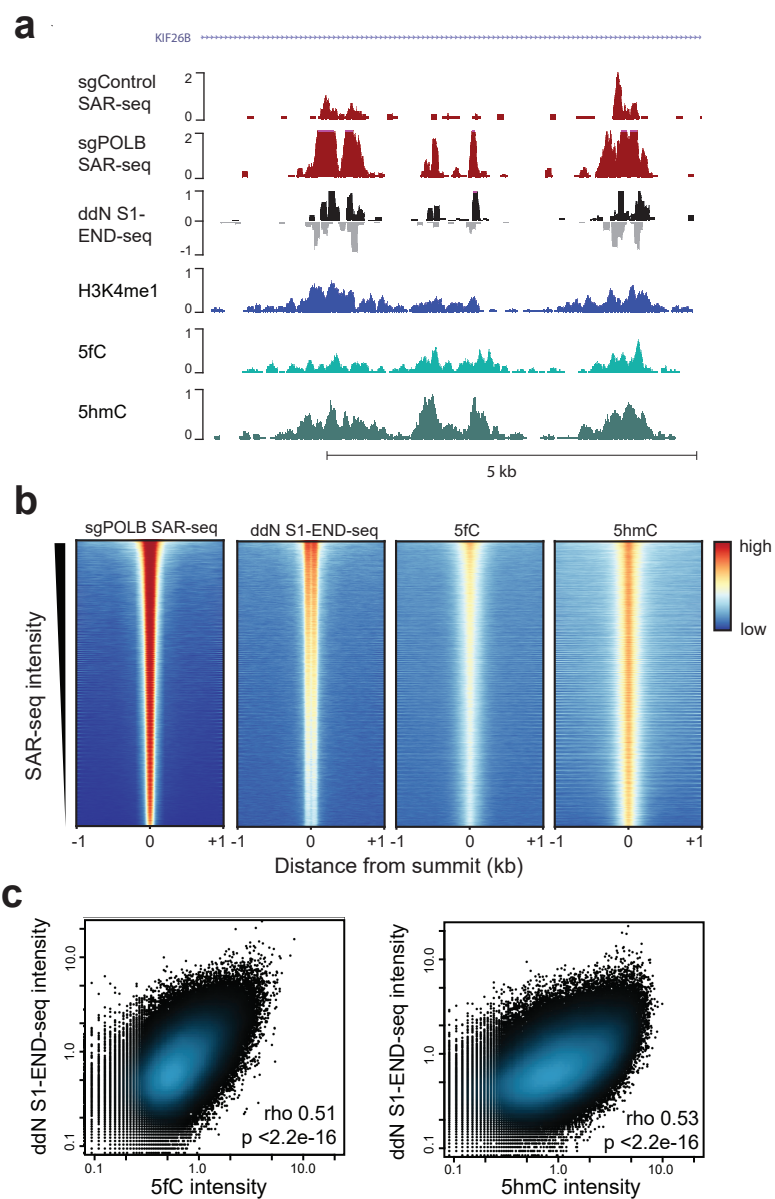
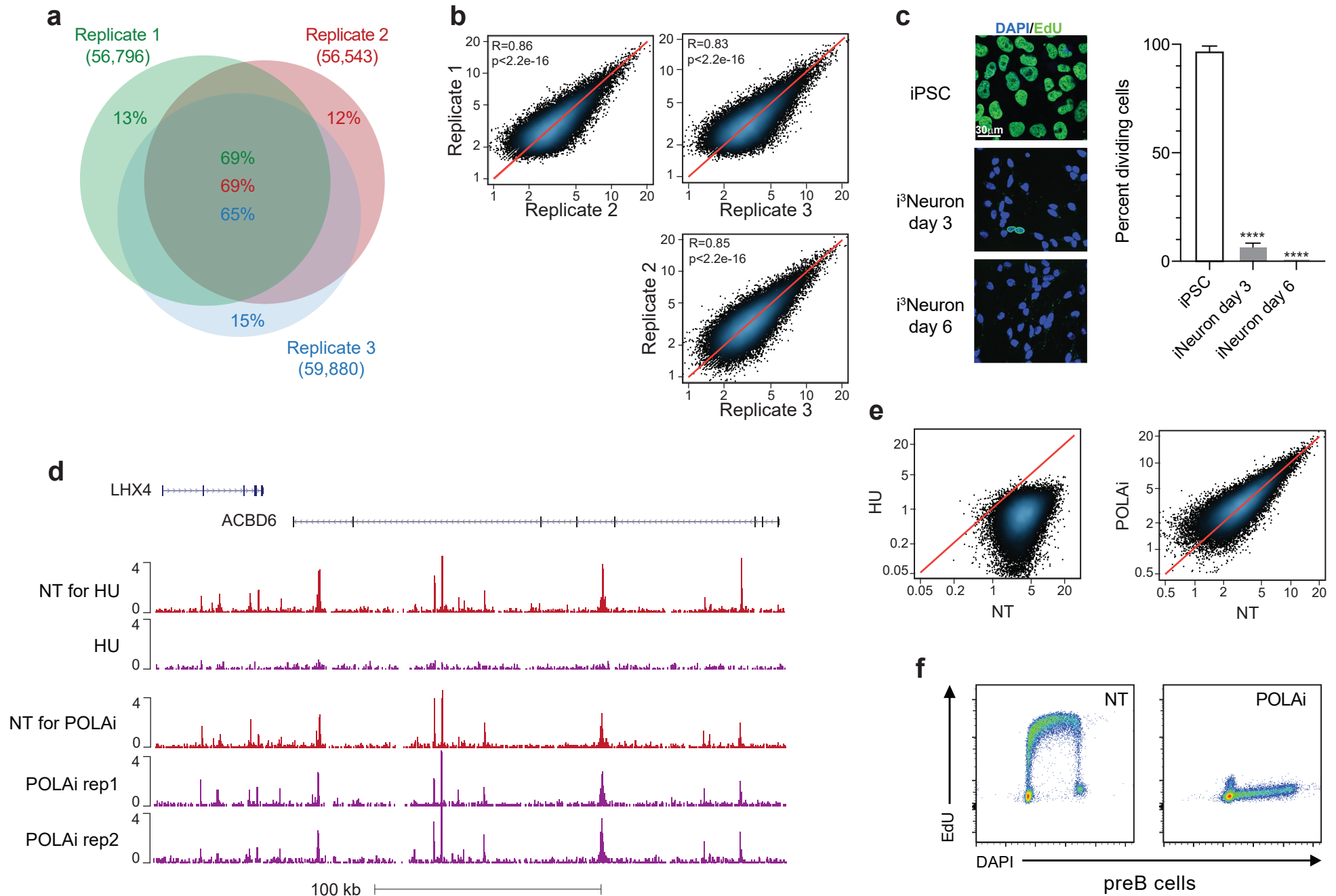


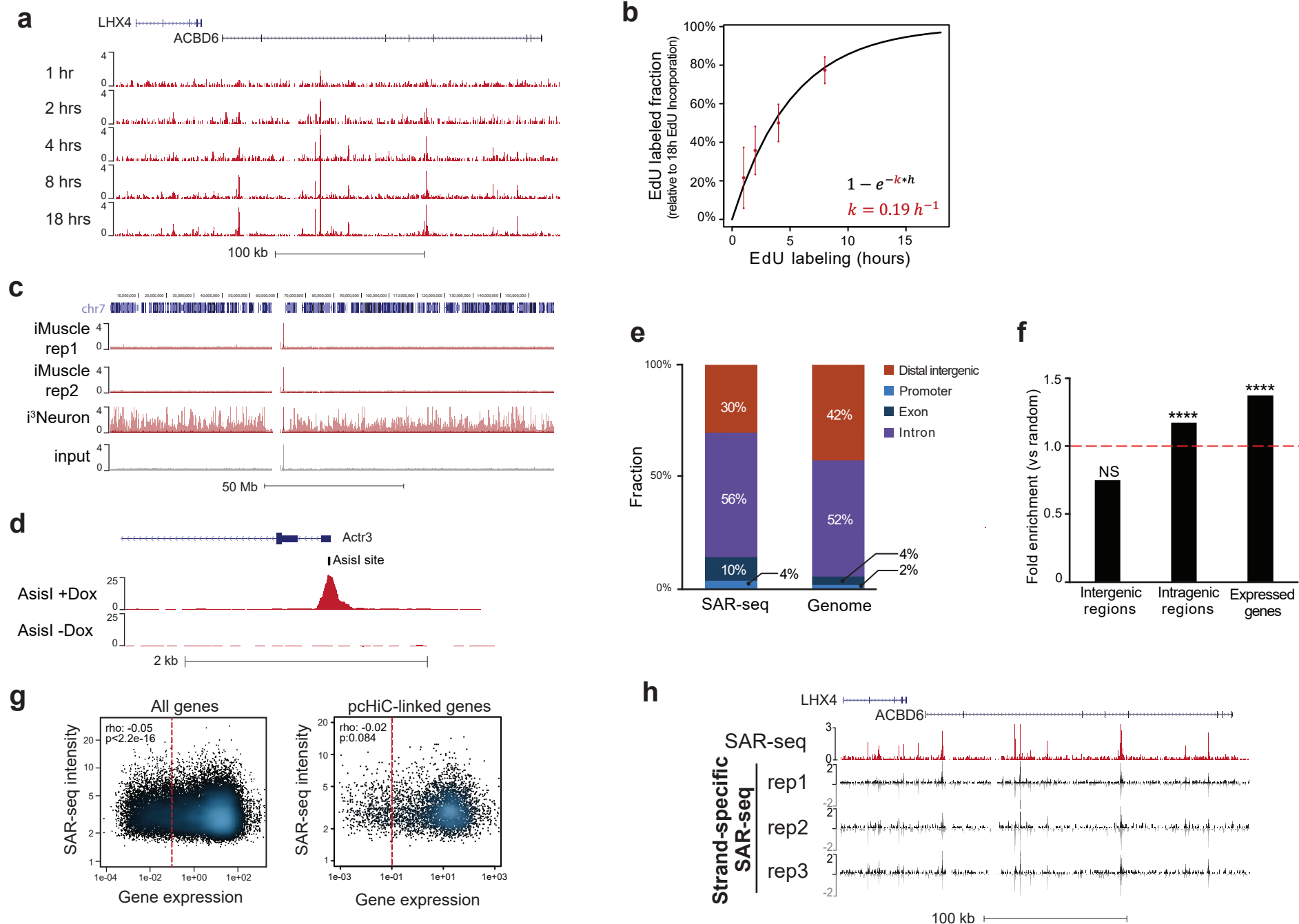
Figure 6



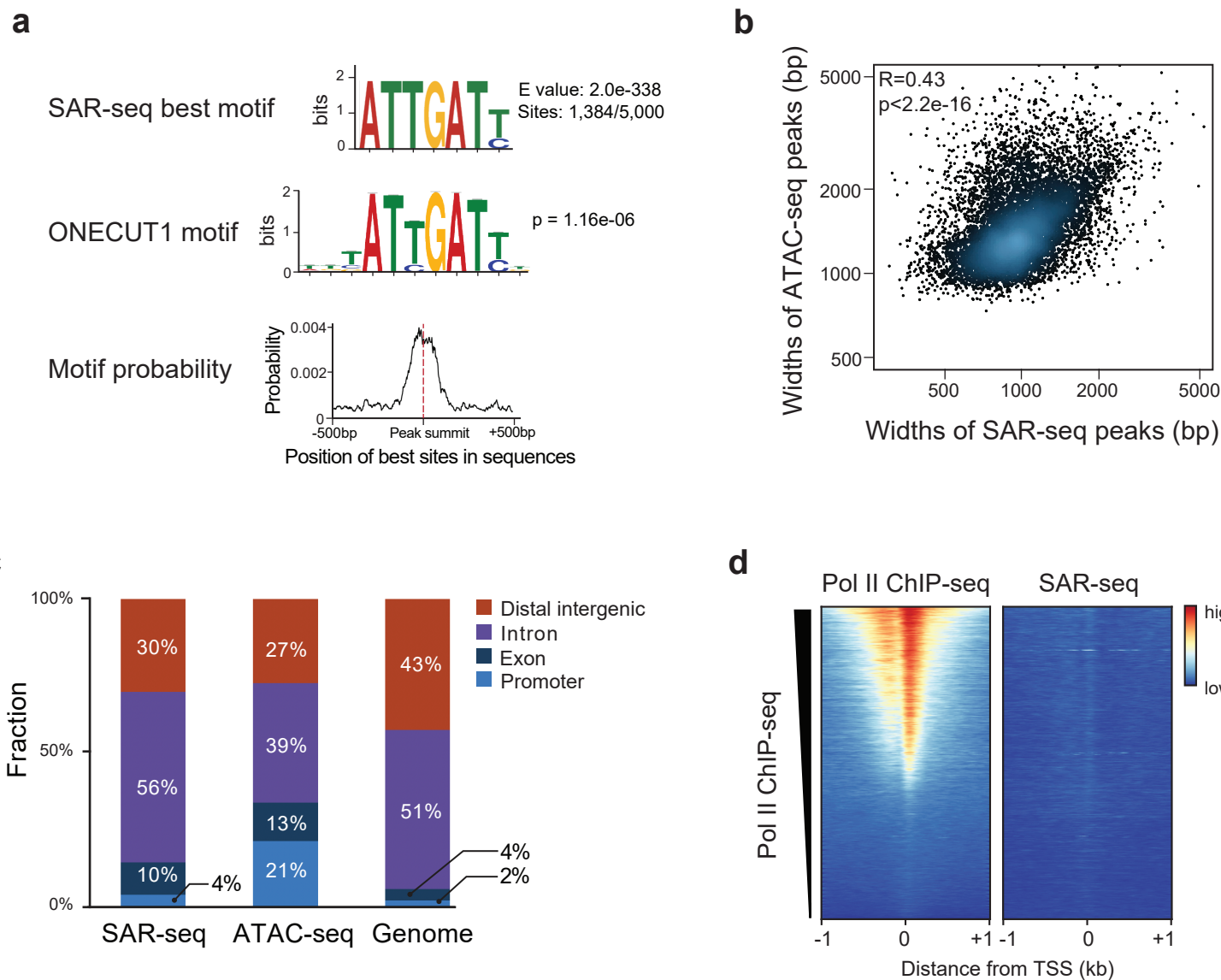
Extended Data Figure 1



Extended Data Figure 2

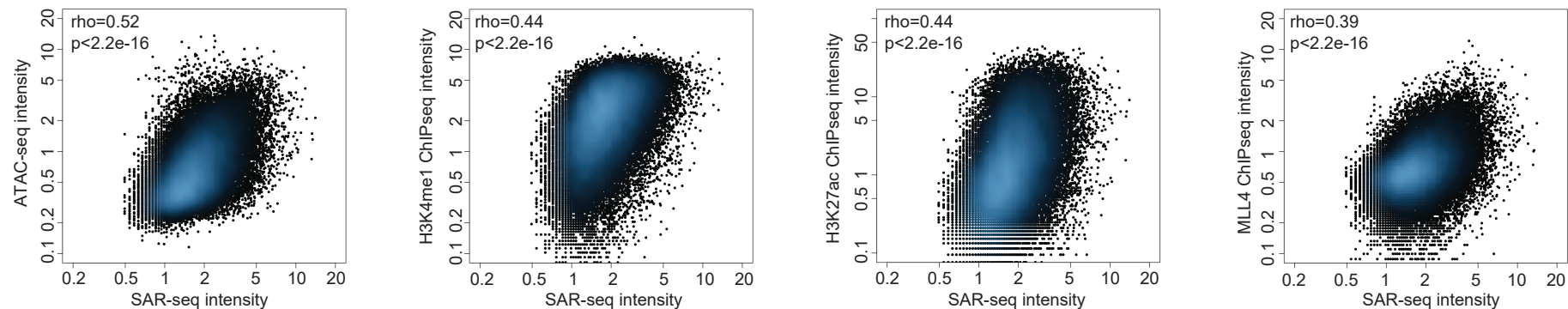


Extended Data Figure 3

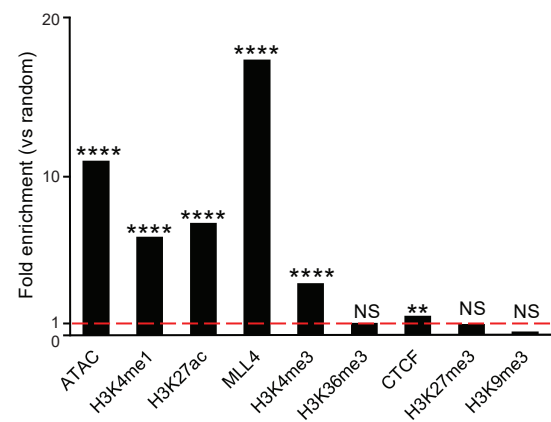


Extended Data Figure 4

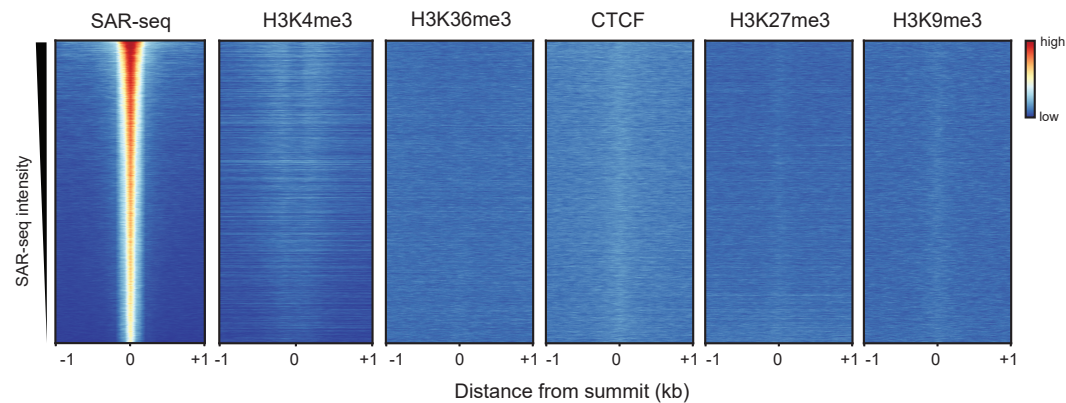
a



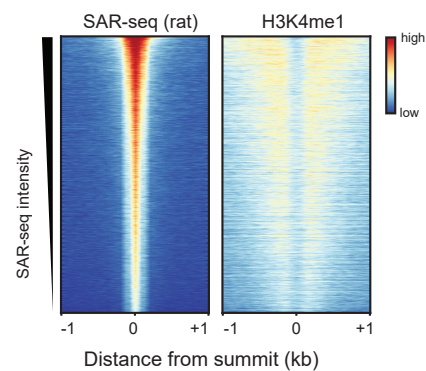
b



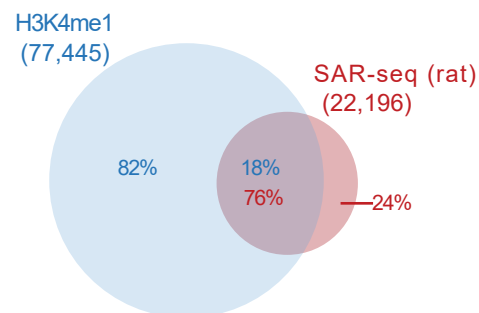
c



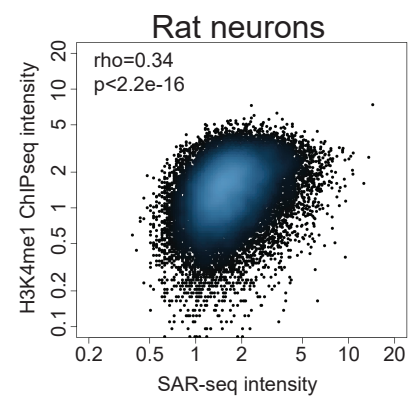
d



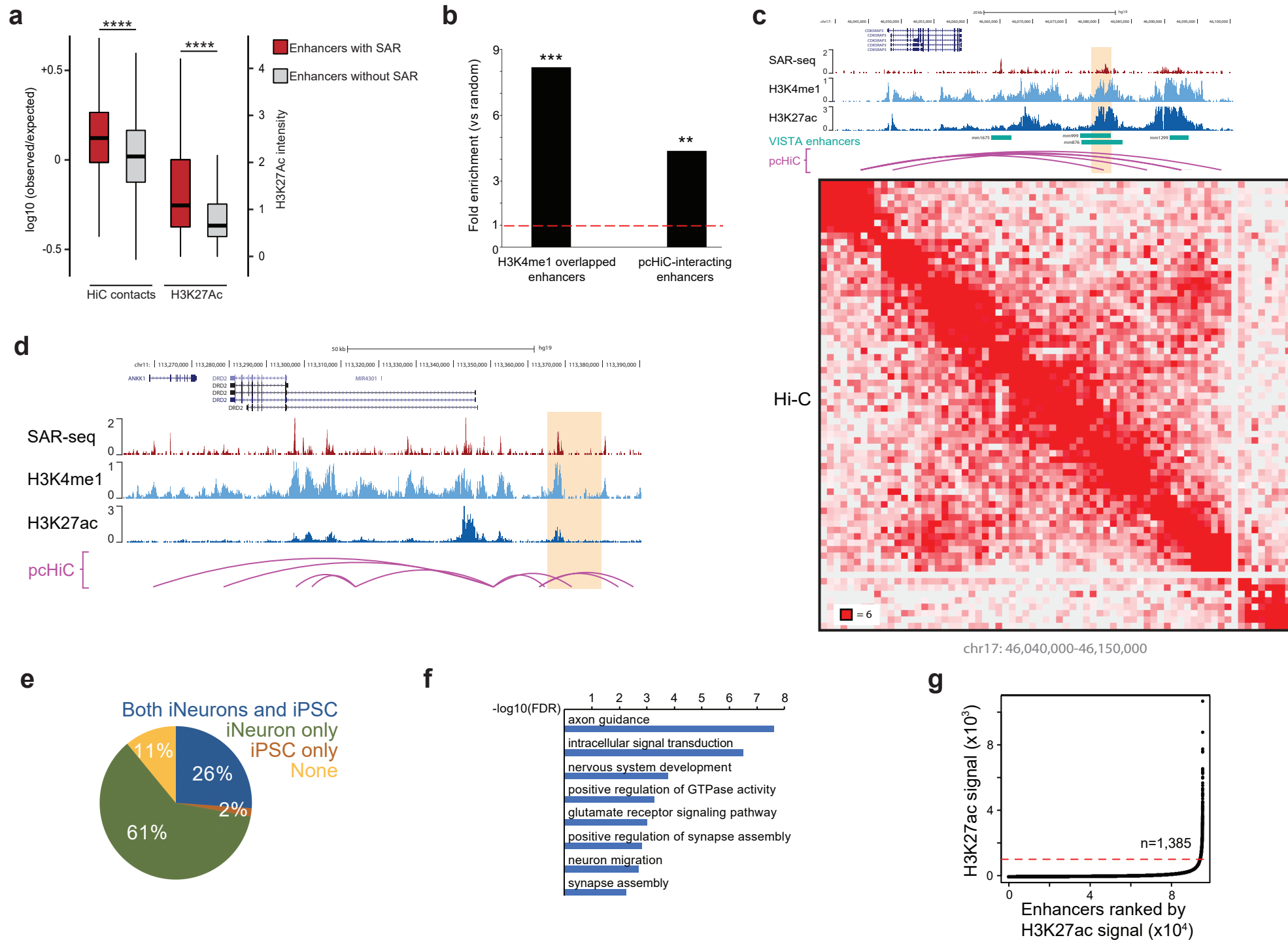
e



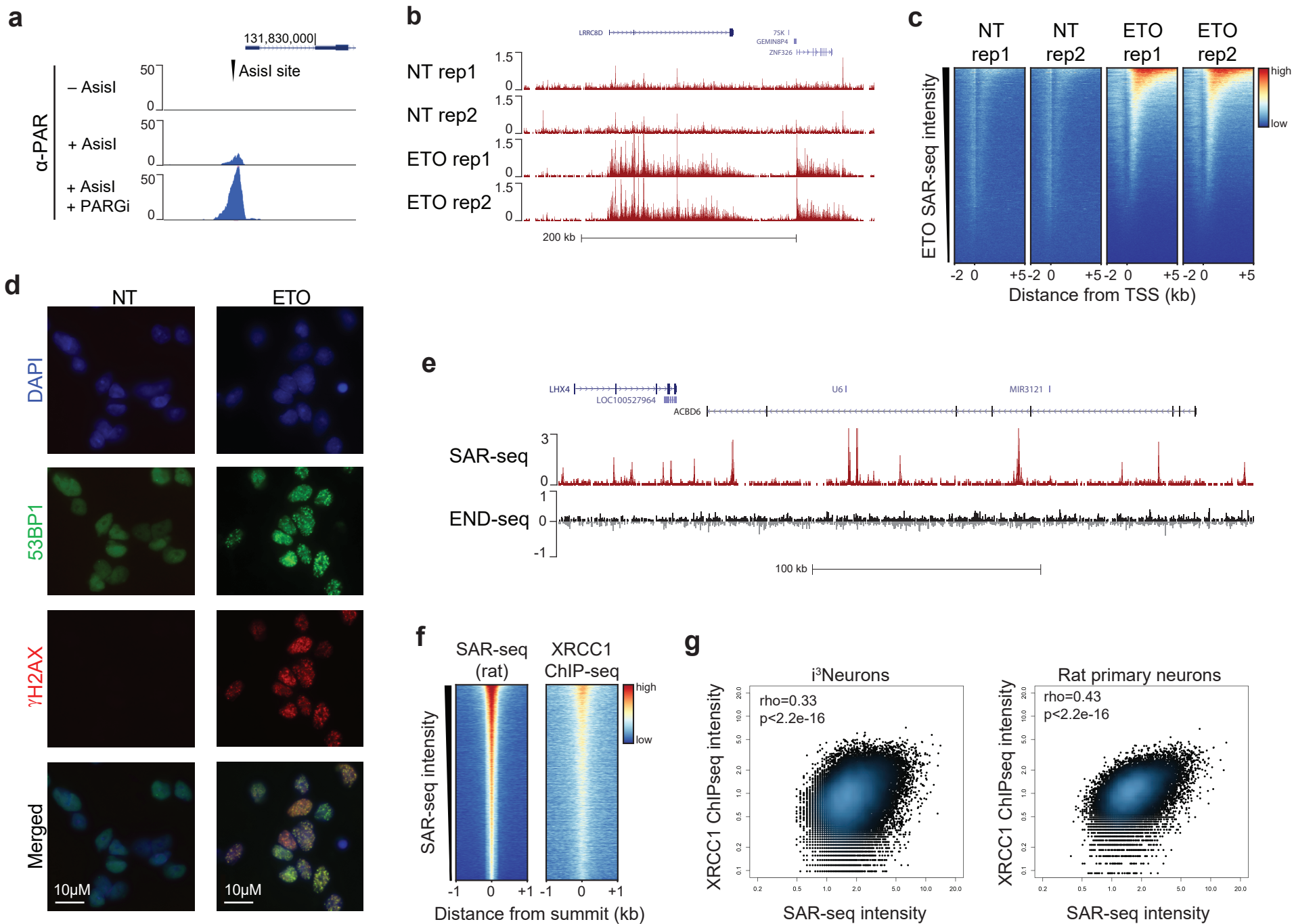
f



Extended Data Figure 5

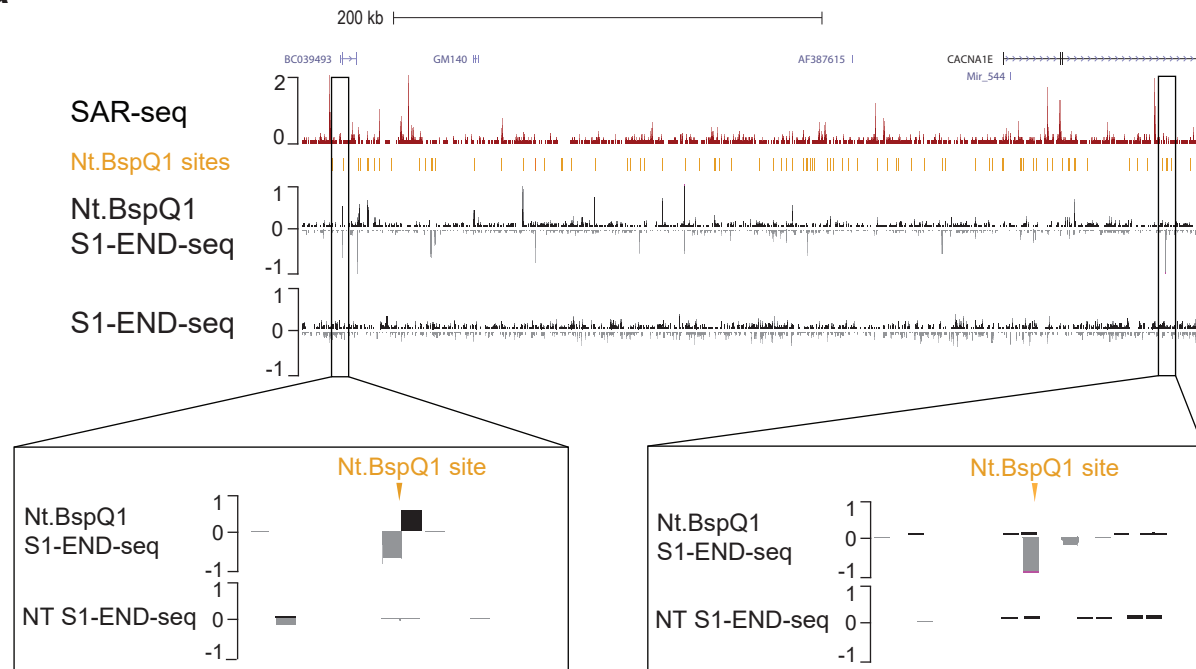


Extended Data Figure 6

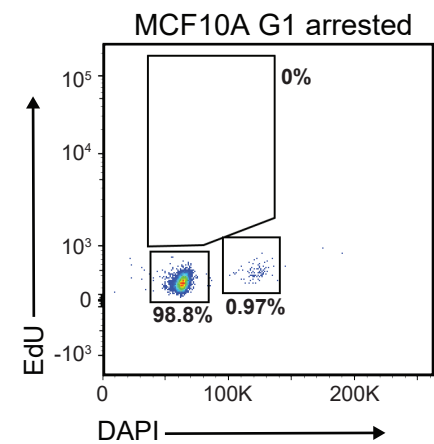


Extended Data Figure 7

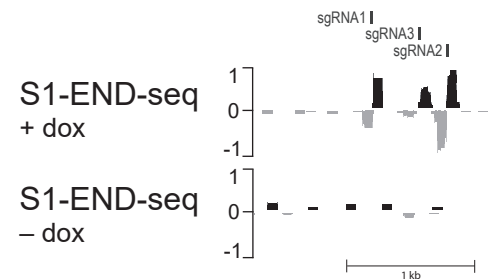
a



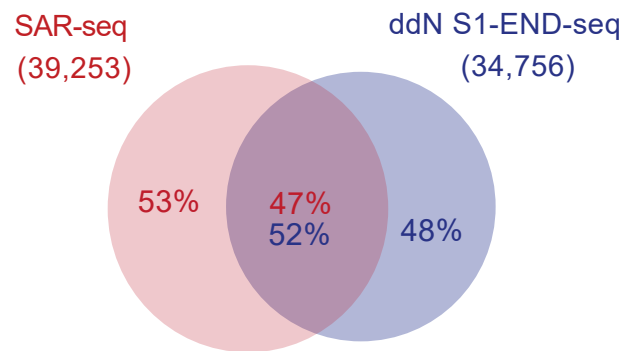
b



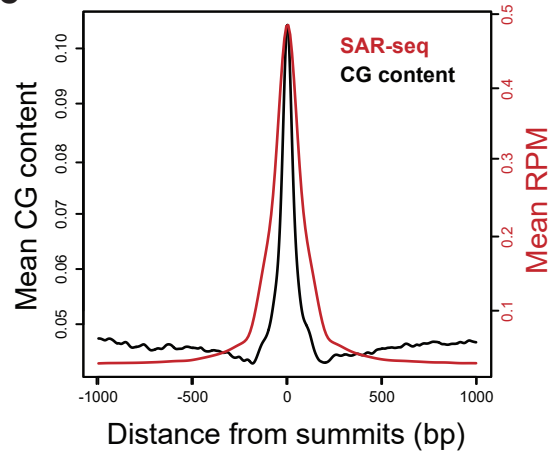
c



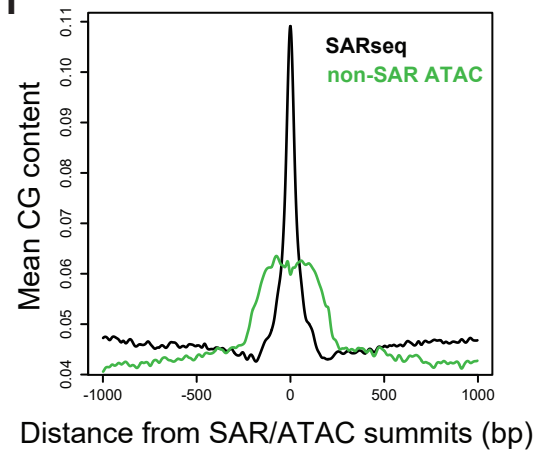
d



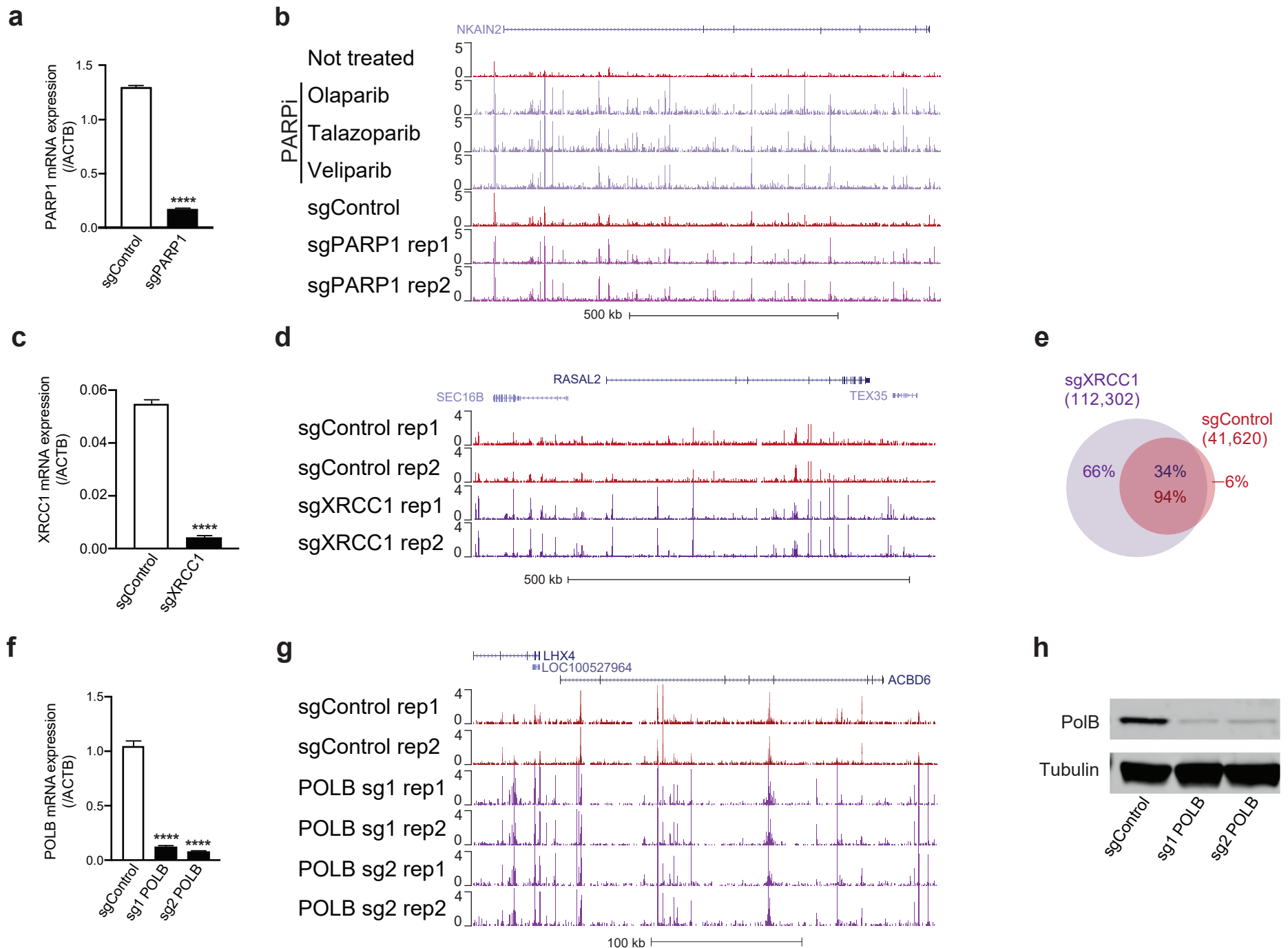
e



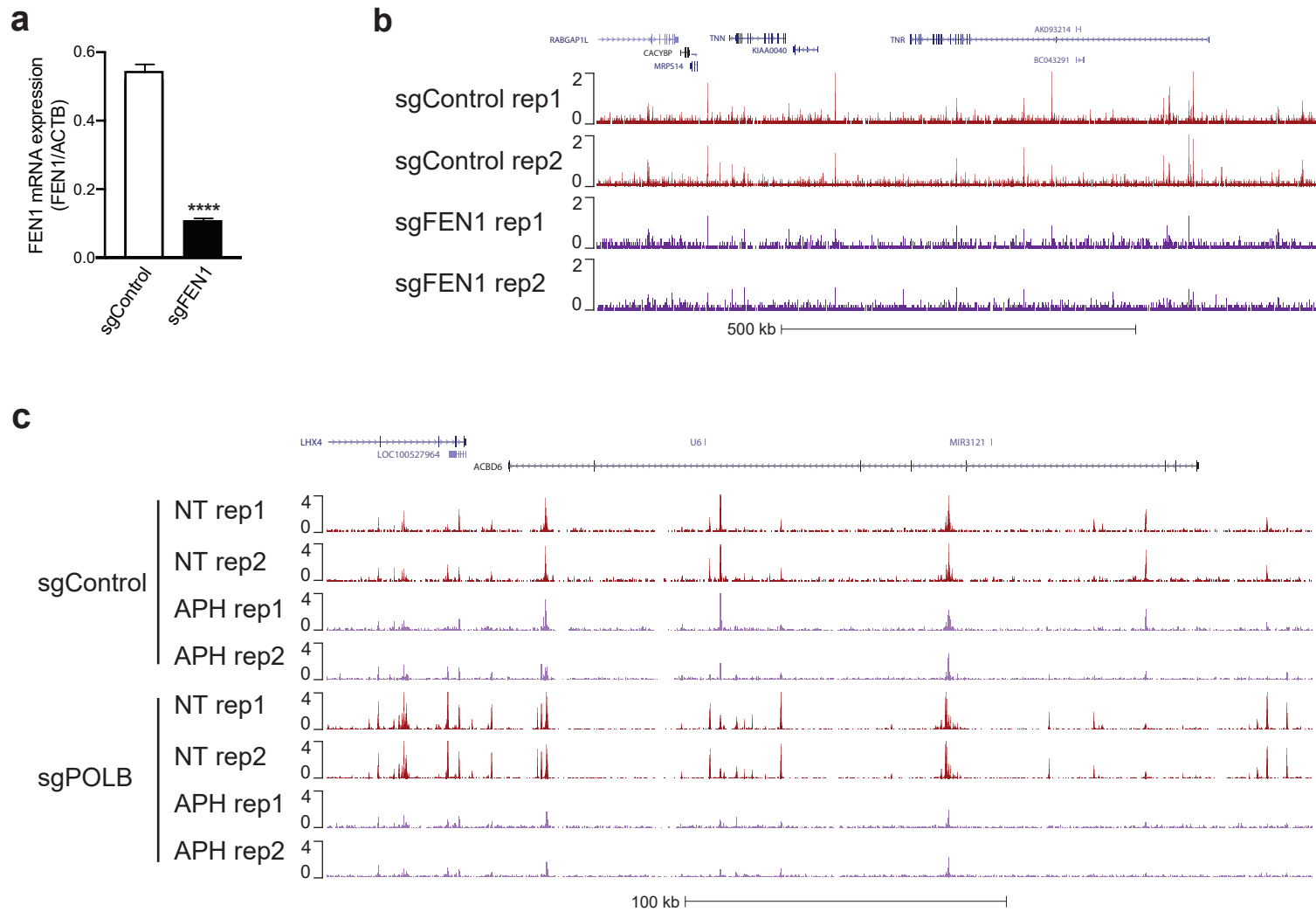
f



Extended Data Figure 8



Extended Data Figure 9



Extended Data Figure 10

