A mechanism for oxidative damage repair at gene regulatory elements

Swagat Ray[1,2,3 a], Arwa A Abugable[1,2a], Jacob Parker[1], Kirsty Liversidge[1], Nelma M Palminha[1,2], Chunyan Liao[1,2], Adelina E Acosta-Martin[4], Cleide DS Souza[5], Mateusz Jurga[6], Ian Sudbery[1], Sherif F El-Khamisy[1,2,6*]

[1]School of Biosciences, Firth Court, University of Sheffield, Sheffield, UK

[2]The Healthy Lifespan and Neuroscience Institutes, University of Sheffield, Sheffield, UK

[3]School of Life and Environmental Sciences, Department of Life Sciences, Joseph Banks Laboratories, University of Lincoln, Lincoln, UK

[4]biOMICS Facility, Faculty of Science Mass Spectrometry Centre, University of Sheffield, Sheffield, UK

[5]Sheffield Institute of Translational Neuroscience, University of Sheffield, Sheffield, UK

[6]Institute of Cancer Therapeutics, School of Pharmacy and Medical Sciences, University of Bradford, Bradford, UK

[*]Correspondence: s.el-khamisy@sheffield.ac.uk, Tel. 00 (44) 114 2222 791, Fax. 00 (44) 114 222 2850

[a]These authors contributed equally to the work

Summary

Oxidative genome damage is an unavoidable consequence of cellular metabolism. It arises at gene regulatory elements by epigenetic demethylation during transcriptional activation[1,2]. Here we show promoters are protected from oxidative damage via a process mediated by the nuclear mitotic apparatus protein (NuMA). NuMA exhibits genomic occupancy ~100 bp around transcription start sites. It binds the initiating form of RNA polymerase II, pause release factors and single-strand break repair (SSBR) components such as TDP1. The binding is increased on chromatin following oxidative damage and TDP1 enrichment at damaged chromatin is facilitated by NuMA. Depletion of NuMA increases oxidative damage at promoters. NuMA promotes transcription by limiting the PARylation of RNA polymerase II, increasing its availability and release from pausing at promoters. Metabolic labelling of nascent RNA identifies genes that depend on NuMA for transcription including immediate early response genes. Complementation of NuMA deficient cells with a mutant that mediates binding to SSBR, or a mitotic separation-of-function mutant, restores SSBR defects. These findings underscore the importance of oxidative DNA repair at gene regulatory elements and describe a process fulfilling this function.

Significance to the lay

Oxidative genome damage is an unavoidable consequence of cellular metabolism. Promoters and enhancers are regions in the genome that are critical for converting the information stored in DNA to proteins essential for life. The formation and repair of oxidative DNA damage at promoters is important for this essential decoding process; however, we don't currently understand how it takes place. This study identifies a process that fulfils this function and describes its implication in disease.

Introduction

Reactive oxygen species (ROS) generated as by-products of metabolism can induce stochastic damage to cellular components including DNA. Localised oxidative damage arises during transcriptional activation at gene regulatory elements such as promoters and enhancers. ROS are generated during oxidative demethylation of histones and 5-methyl cytosine in CpG clusters[1,2]. Guanine base is the most frequently oxidised base among canonical DNA bases[3]. The resulting 7,8-dihydro-8-oxoguanine (8-oxoG) is abundant at guanine-rich promoter regions and regulates the transcription of proinflammatory genes[4,5]. 8-oxoG is converted to apurinic/apyrimidinic sites (AP sites) by 8-oxoguanine DNA glycosylase (OGG1), which is repaired by single-strand break repair (SSBR) components such as tyrosyl-DNA phosphodiesterase 1 (TDP1) and X-Ray repair cross complementing 1 (XRCC1). The primary substrate for TDP1 is the product of abortive topoisomerase 1 (TOP1) activity, which leaves TOP1 covalently linked to the 3'-terminus of DNA[6]. The phosphodiesterase activity of TDP1 also cleans other blocked 3'-ends and AP sites[7–9]. XRCC1 is a SSBR scaffold protein that facilitates most SSBR events[10]. Unrepaired SSBs perturb gene transcription and hyperactivate the SSB sensor poly(ADP-ribose) polymerase 1 (PARP1), which catalyses the formation of poly(ADP-ribose) polymers on itself, SSBR components, histones and transcription machines. PARylation facilitates the formation of SSBR complexes and promotes the recruitment of TDP1 and XRCC1 to damaged chromatin[11,12]. PARylation also regulates the promoter proximal pause-release of RNA polymerase II (Pol2)[13]. Deficiency of XRCC1 or TDP1 causes cerebellar atrophy and peripheral neuropathy[14,15]. Although the molecular architecture of oxidative damage repair at gene bodies is well characterised, how oxidative breaks at gene regulatory regions are repaired and their impact on transcription remain poorly understood.

The spatiotemporal coordination of DNA repair and transcription involves genome organisation driven by structural and cytoskeletal proteins such as actin and lamins[16,17]. The nuclear mitotic apparatus protein (NuMA) is another structural protein, related to lamins, which localises at spindle poles during mitosis, thus gaining its name[18]. It was found to associate with cytoplasmic dynein and dynactin to tether microtubules at the spindle poles, promoting mitotic spindle assembly and stabilization[19,20]. NuMA was later reported in the interphase nucleus and it was suggested this acts as a reservoir for use during mitosis[21]. Subsequent work reported involvement in chromatin remodelling,

homology directed repair and nuclear shape[22–24]. Here, we report that NuMA executes an important new role during interphase by promoting gene transcription and repairing oxidative DNA breaks at regulatory genome elements.

NuMA promotes SSBR and transcription recovery

The cerebellum is vulnerable to defects in DNA repair proteins that safeguard the genome from oxidative damage[6,14,15]. To explore the unique properties of the cerebellum we examined the GTEx RNA-seq dataset, which contains large variety of human brain regions. We observed higher expression of the SSBR components, TDP1 and XRCC1 in the cerebellum compared to other brain tissues (**Extended Data Fig.1**). We also noticed high expression of NuMA (**Fig.1a**) which prompted us to study its possible role in oxidative damage repair. We first examined NuMA's function in a range of non-neuronal cells and then validated key findings in neural cells. NuMA depletion led to a delay in repairing oxidative DNA breaks, induced by H_2O_2 treatment, which are predominantly SSBs[6] (**Fig.1b**). The inefficient repair of SSBs in NuMA-depleted cells led to hypersensitivity to H_2O_2 (**Fig.1c**). Pre-incubation with the pause-release transcription inhibitor 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) abrogated the NuMA-dependent defect in SSBR rates, suggesting a transcription associated role for NuMA during SSBR (**Fig.1d, Extended Data Fig.2**). The reduced ability to repair SSBs could delay transcription recovery following DNA damage. To test this, newly synthesized RNA was labelled with an alkyne-modified nucleoside, 5-ethynyl uridine (EU). Control cells recovered transcription to background levels during a 90 min recovery period whereas NuMA-deficient cells did not (**Fig.1e**). These data suggest a role for NuMA during SSBR and transcription recovery following oxidative stress.

Genes that need to respond quickly following oxidative stress are predicted to display higher demand for efficient DNA repair and transcription recovery. Those genes often exhibit Pol2 pausing at their promoters, which may help poising them for a rapid transcription burst following stress[25]. To identify NuMA regulated genes, we compared nascent transcripts in control and NuMA-deficient cells. Here we utilised cells in which NuMA depletion is achieved by a doxycycline-inducible CRISPR-Cas9 cassette[19] (**Extended Data Fig.3a**). Nascent RNA transcribed during the recovery period was labelled with 4-thiouridine (4sU), biotinylated and purified using streptavidin beads

(**Extended Data Fig.3b**). During the recovery from H_2O_2 treatment, 4579 genes were differentially upregulated, and 1696 genes were downregulated (**Extended Data Fig.3c**). Out of the 1023 paused genes (Pausing Index ≥ 2), 355 were upregulated whereas 151 were downregulated (**Extended Data Fig.3d**). Upon NuMA depletion we observed an almost reversal of the expression profile, with 1209 genes differentially upregulated and 4996 genes downregulated (**Extended Data Fig.3e**). Strikingly, the profile of paused genes (pausing index ≥2) was also reversed upon NuMA depletion. In contrast to ~35% paused genes (355 out of 1023) that were differentially upregulated by H_2O_2 treatment in presence of NuMA in **Extended Data Fig.3d**, almost the same proportion (397 genes) was instead downregulated upon NuMA depletion (**Extended Data Fig.3f**). We conclude that NuMA promotes transcription of a subset of genes during recovery from oxidative stress.

NuMA promotes transcription of NRGs

Further inspection of the differentially expressed gene lists revealed 2358 genes that were upregulated in NuMA-containing cells following oxidative stress but downregulated upon NuMA depletion (**Fig.1f**), hereafter called NuMA-Regulated Genes (NRGs). Approximately 36% of the reported Immediate Early Response Genes (IERGs) were enriched among NRGs (**Fig.1g**). Of the 2358 NRGs, 16 IERGs require NuMA for transcription after H_2O_2 treatment (**Fig.1h**). Consistently, qPCR validation showed ~ 220-fold increase of *Fos* transcripts, an IERG that is highly paused, during recovery from oxidative stress, which was reduced to only 100-fold in the absence of NuMA (**Fig.1i**). In addition to IERGs, the coordinated DNA repair response to oxidative damage at promoters promotes the transcription of proinflammatory genes[26,27]. Consistently, NuMA promotes transcription of several proinflammatory genes including *CXCL2* and *NFKB* (**Fig.1j**). We next tested if pausing is a determining factor for transcription regulation by NuMA. By setting the pausing index at 2 and above, we identified ~18% of paused genes amongst NRGs (**Fig.1k, left**). Furthermore, ~20% of genes with a pausing index above 20 were found to be NRGs, which again constituted a significant enrichment (**Fig.1k, right**). However, the extent of pausing alone was not sufficient to distinguish NRGs from non-NRGs (**Fig.1l**).

To determine other features that define NRGs, we considered fragile genomic sequences. Although oxidative damage to DNA predominately generates SSBs, it can also generate double-strand breaks (DSBs). DSB mapping identified genomic sequences that are fragile and thus more likely to accumulate DSBs[28]. NRGs were not enriched in genes with fragile promoters, arguing against a role for NuMA in protecting promoters from DSBs (**Extended Data Fig.3g**). Notably, the higher the transcription level of the gene, the more reliance on NuMA (**Extended Data Fig.3h**). Moreover, the dependence of transcription on NuMA following oxidative stress is correlated with gene length, with longer genes showing lower transcription in absence of NuMA (**Extended Data Fig.3i**). NuMA dependent genes contained those implicated in SSBR such as *XRCC1*, *Lig3α*, *TDP1* and *PARP1* but not *bona fide* DSBR genes such as *XRCC4*, *Lig4*, *53BP1* and *MRE11* (**Extended Data Fig.3j**). Together, these data demonstrate that NuMA promotes the transcription of a well-defined cohort of genes following oxidative damage.

NuMA interacts with SSBR components and Pol2

We next tested if NuMA physically interacts with SSBR components using an unbiased proteomic approach. Whole cell extract following oxidative damage was isolated from GFP-NuMA-expressing cells and the GFP immunoprecipitates were subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS). NuMA was found to bind components of both RNA polymerase II (Pol2) and SSBR machineries (**Extended Data Fig.4a,b, Supplementary Data 1**). Purification of GFP-NuMA complexes further confirmed the interaction with the SSBR components TDP1, PARP1 and XRCC1, but not DSB repair components such as Ku70/80, Lig4 and 53BP1 (**Fig.2a**). We also noted that NuMA interacts with the initiating p-Ser5(Pol2) but not p-Ser2(Pol2), suggesting a role during transcription initiation. The reciprocal immunoprecipitation confirmed that endogenous NuMA interacts with TDP1 (**Fig.2b**), p-Ser5(Pol2), but not p-Ser2(Pol2) (**Fig.2c**). Cellular fractionation followed by IP of endogenous p-Ser5(Pol2) from chromatin pellets and soluble nucleoplasm showed interaction of NuMA with PARP1, p-XRCC1 and TDP1 in both compartments (**Fig.2d**). Notably, the interaction of p-Ser5(Pol2) with NuMA increased ~8-fold on chromatin during a 10-min recovery from oxidative stress. A similar 5-fold, 10-fold and 3-fold increase was observed for PARP1, TDP1 and p-XRCC1, respectively. This interaction was not mediated by DNA as it persisted following nuclease digestion. Moreover, the endogenous interaction was

confirmed by an orthogonal method using proximity ligation assays following triton extraction to remove nucleoplasmic interactions and enrich chromatin-associated interactions. Although the interaction of NuMA with p-Ser2(Pol2) under unperturbed conditions was detectable using this method, it declined during a 10 min recovery from oxidative damage (**Fig.2e**). In contrast, the interaction of NuMA with p-Ser5(Pol2) increased following recovery from oxidative damage, confirming the chromatin-IP observations (**Fig.2e**). Together, these data demonstrate that NuMA physically interacts with p-Ser5(Pol2) and the TDP1-SSBR machinery and that this interaction is enriched on chromatin during the recovery from oxidative stress.

Since NuMA interacts with TDP1 and PARP1, we next tested if this interaction is Poly(ADP-ribose) "PAR" dependent. PARP3 and tankyrase 1 have been implicated in the mitotic and chromatin remodelling functions of NuMA yet the role of PARP1 remains unknown[23,29,30]. We compared TDP1 immunoprecipitates with and without prior incubation with Poly(ADP-ribose) glycohydrolase (PARG) inhibitor, to enrich for PARylation dependent interactions (**Extended Data Fig.4c**). PARG inhibition led to enrichment of both NuMA and PARP1 in TDP1 immunoprecipitates (**Fig.2f**). Incubation of recombinant TDP1 protein with *in vitro* transcribed-translated ^{35}S-methionine labelled NuMA in presence or absence of synthetic PAR led to an increase in NuMA binding to TDP1 (**Fig.2g,h**). To confirm binding using an orthogonal assay, His-TDP1 or control His-eIF4a were immobilised on Ni-beads and incubated with ^{35}S-methionine labelled NuMA. PARylation was performed *in vitro* by incubation with recombinant PARP1, free DNA and NAD^+. PARylation increased NuMA binding to TDP1, but not eIF4a, confirming the specificity of the PARylation mediated interaction between NuMA and TDP1 (**Fig.2i,j**).

To dissect the respective roles of PARP1 versus PARP3 in NuMA's function, we employed CRISPR mediated gRNA- against PARP1 and PARP3. Depletion of PARP1, but not PARP3, reduced NuMA interaction with TDP1 (**Fig.2k,l**). Purification of GFP tagged PARP1 and PARP3 complexes from cells confirmed that PARP1, but not PARP3, exists in a complex with TDP1 (**Fig.2m**). This data also shows that both PARP1 and PARP3 bind NuMA, consistent with previous reports[22,29]. Thus, NuMA exists in two distinct pools, a previously reported and characterised PARP3-bound

fraction implicated in mitosis and chromatin remodelling, and a novel PARP1-bound fraction involved in SSBR and transcription following oxidative stress.

NuMA-L promotes SSBR and transcription

The full length NuMA isoform is a 238,260 Da protein of 2115 amino acids (NuMA-L) (**Extended Data Fig.5a**). An exon skipping alternative splicing generates a NuMA short (NuMA-S) isoform, missing 14 amino acids. To determine which NuMA isoform is implicated in SSBR associated transcriptional responses, NuMA complexes were purified from cells expressing GFP-tagged NuMA-L or NuMA-S and subjected to on-beads mass spectrometry analysis. We observed two distinct signature peptides, LTAQVEQLEVFQR and LTAQVEELSK, which corresponds to NuMA-L and NuMA-S, respectively (**Extended Data Fig.5b**). Hierarchical clustering revealed more proteins enriched in NuMA-L complexes than NuMA-S (**Extended Data Fig.4d, Supplementary Data 2**). A preferential enrichment of PARP1 and Pol2 complexes was observed in NuMA-L, but not NuMA-S (**Extended Data Fig.5c**). Purification of GFP-tagged NuMA complexes showed that TDP1, XRCC1 and PARP1 are primarily enriched with NuMA-L (**Extended Data Fig.5d**). Complementation of NuMA-deficient cells with NuMA-L restored the delayed SSBR rates (**Extended Data Fig.5e-g**), defects in clonogenic survival (**Extended Data Fig.5h**) and transcription recovery (**Extended Data Fig.5i,j)**. Together these data suggest an isoform specific role of NuMA during SSBR and transcription responses.

NuMA C-terminal tail promotes SSBR

We next mapped TDP1 and PARP1 binding sites on NuMA. Four truncations of GFP-tagged-NuMA-L were generated (**Extended Data Fig.6a**). Co-immunoprecipitation showed that the globular C-terminal domain ($NuMA^{GD}$) alone is sufficient to pull down both TDP1 and PARP1 (**Extended Data Fig.6b**). Consistently, alkaline comet assays showed that $NuMA^{GD}$ is sufficient to restore SSBR defects in NuMA-deficient cells (**Extended Data Fig.6c-e**). Given the reported roles of NuMA in mitosis, we next set out to generate a separation-of-function mutant. We deleted the microtubule-binding domain spanning amino acids 1866-1936 and compared this mutant ($NuMA^{MD}$)[31] with full-length NuMA ($NuMA^{FL}$) for their abilities to bind and support SSBR. Whereas $NuMA^{MD}$ was unable to bind microtubules after metaphase arrest with nocodazole (**Extended Data Fig.6f, g**), it was able to associate with TDP1 and PARP1 (**Extended**

Data Fig.6h). Complementation of NuMA-deficient cells with siRNA targeting-resistant NuMA^{MD} was able to restore SSBR defects (**Extended Data Fig.6i-k**). We conclude that NuMA exhibits at least two distinct cellular roles, an established mitotic function and a novel SSBR function mediated by the interaction of its C-terminal globular domain with the TDP1/SSBR machinery.

NuMA increases TDP1 at damaged chromatin

The increased chromatin binding of NuMA to TDP1 and PARP1 following oxidative stress suggests an adaptor function through which NuMA promotes the enrichment of TDP1 at sites of damaged chromatin. To test this, we compared the accumulation of GFP-tagged TDP1 at sites of oxidative laser damage in control and NuMA-deficient cells using live cell imaging. TDP1 is recruited to DNA damage sites within 10 sec of microirradiation, peaked at ~20 sec and the signal remained stable through 80 sec of live cell imaging. However, NuMA depletion led to a marked reduction of TDP1 accumulation, and so did PARP1 inhibition which is consistent with published work[12]. Simultaneous PARP1 inhibition and NuMA depletion did not reduce TDP1 accumulation further, suggesting that NuMA and PARP1 work together to facilitate the recruitment of TDP1 to sites of damaged chromatin (**Extended Data Fig.7a-c**). TDP1 or NuMA depletion led to a similar SSBR defect following oxidative stress, and depletion of both did not lead to an additive increase in SSBR defect (**Extended Data Fig.7d-f**). NuMA depletion alone led to a detectable reduction in TDP1 protein levels, consistent with a role for NuMA in promoting *TDP1* transcription (**Extended Data Fig.3j**). Clonogenic survival assays supported the epistatic functional interaction between TDP1 and NuMA (**Extended Data Fig.7g**). Together, these data demonstrate that NuMA works together with TDP1 to promote its availability for SSBR and cell survival following oxidative stress.

NuMA regulates Pol2 pausing and PARylation

The association of NuMA with the initiating form of Pol2 and with promoter-proximal pause-release factors, and the enrichment of paused genes among NRGs prompted us to test if NuMA promotes Pol2 pause release. Pol2 occupancy was compared genome-wide in control and NuMA-deficient cells using ChIP-seq. Under endogenous physiological conditions, NuMA depletion led to piling of Pol2 molecules around Transcription Start Sites (TSS) and a corresponding reduction of Pol2 occupancy in

gene bodies, resulting in a significant increase in pausing ratio (**Fig.3a,b**). The increase in pausing ratio was observed in all genes and was highly increased at NRGs (**Fig.3c**). Next, we examined Pol2 occupancy following exposure to exogenous oxidative stress. Here, NuMA depletion led to a reduction of Pol2 occupancy at promoters (**Fig.3d,e**). These observations suggest that NuMA increases the availability of Pol2 complexes at promoters, which is required for transcription recovery following oxidative damage. To understand this further we considered the possibility that NuMA may regulate Pol2 PARylation and therefore its availability at promoters.

Recent proteomic screens identified 16 PARylation sites on NuMA following oxidative stress, 12 of which are serine residues located in the globular tail domain that we show is critical for SSBR[32,33] . Since PARP1 dependent PARylation has been shown to delay Pol2 release from promoter proximal pausing[13,34], we considered the possibility that NuMA may regulate PARylation at promoter regions. To test this, we compared the extent of p-Ser5(Pol2) PARylation in the presence and absence of NuMA. NuMA depletion resulted in ~2-fold increase of Pol2 PARylation following oxidative stress (**Fig.3f**), which was specific to Pol2 since we did not observe a similar increase in PARP1 or p-XRCC1 PARylation. To exclude non-covalent binding mediated interaction, we conducted a sequential co-immunoprecipitation with p-Ser5(Pol2) antibodies, followed by anti-PAR antibodies (**Fig.3g**). This experiment further confirmed that NuMA depletion increased the levels of Pol2 PARylation compared to control cells. They show that NuMA limits the PARylation of the initiating Pol2 complexes, increasing their availability at promoters to facilitate release from pausing. Consistently, mass spectrometry showed NuMA binding to all six components of the Pol2 pausing-associated factor-1 complex (PAF1C), including PAF1, CTR9, LEO1, CDC73, WDR61 and RTF1 (**Extended Data Fig.8a**). Co-immunoprecipitation confirmed NuMA binding to p-Ser-5(Pol2) and PAF1 (**Extended Data Fig.8b**). Furthermore, NuMA-deficiency led to a slower rate of removal of the silencing chromatin mark $H2A^{K119}$ monoubiquitination during recovery from oxidative stress (**Fig.3h**), which has been shown to be regulated by PARylation[35]. It also led to a decrease in the levels of the transcription activating chromatin mark, $H2B^{K120}$ monoubiquitination (**Fig.3i**). Together, we conclude that NuMA facilitates transcription by increasing Pol2 availability at promoters and its release from pausing.

NuMA protects promoters and enhancers

Two predictions arose from the above findings; an increase of oxidative damage in promoters in NuMA-deficient cells and an increase of NuMA occupancy at promoters. NuMA-deficient cells exhibited more genomic 8-oxoG than wild-type cells under unperturbed conditions and following recovery from oxidative damage (**Extended Data Fig.9a**). 8-oxoG is recognized and excised by OGG1, leaving an AP-site. AP-sites are primarily removed by AP-endonuclease (APE1) with contributions from other enzymes such as TDP1[9,36]. Mapping oxidative breaks by chemical enrichment and immunoprecipitation will not detect oxidative damage that has been processed into AP-sites, which are more persistent. Therefore, we mapped oxidative breaks using a recently developed method employing biotin-labelled aldehyde-reactive probe[37].

To assess oxidative damage as the sum of AP-sites and 8-oxoG, we applied recombinant OGG1 *in vitro* to the extracted DNA (**Fig.4a**). Under these conditions, any remaining 8-oxoG is excised in sequence-independent manner after DNA extraction resulting in secondary AP-sites[37]. We will be referring to this method as OGG1-enriched AP-Seq (OGAP-Seq). Biotin-tagged DNA was pulled down using streptavidin beads from cells containing or lacking NuMA and subjected to high throughput sequencing (**Fig.4b**). In wild-type cells, we observed ~40% reduction of oxidative damage at TSS of all reported genes (**Fig.4c**), which is consistent with published work[38,39]. However, in striking contrast to wild-type cells, NuMA-deficient cells possessed ~25% more oxidative breaks at promoters (**Fig.4c,d**), which is in line with DNA repair hotspots (DRHs) reported at promoters[40]. We also observed a weaker (~10%) protection of NuMA at gene bodies (**Fig.4d**). The promoters of NRGs exhibited higher protection from oxidative damage compared to non-NRGs (**Fig.4e,f**). We also observed an increase in oxidative breaks at enhancers in NuMA-deficient cells (**Fig.4g**), which is consistent with enhancers being hotspots of SSB accumulation[41]. Together, these findings demonstrate a protective role for NuMA from oxidative damage that is more prominent at gene regulatory elements.

NuMA is enriched at promoters and enhancers

We next tested if NuMA binds SSBs *in vitro* and if it occupies promoter regions in cells. NuMA^{GD} which interacts with TDP1/SSBR and can restore SSBR defects in NuMA-deficient cells was incubated with oligonucleotide substrates containing an

intact duplex DNA, a nick or a 1-nucleotide gap, and binding analysed using electrophoretic mobility shift assays. Incubation with NuMA^{GD} led to the appearance of slower migrating bands, suggesting it binds SSBs *in vitro* (**Fig.5a**). We then mapped the occupancy of endogenous NuMA genome-wide using ChIP-seq. The metagene profile revealed a prominent increase of NuMA at the beginning of genes, which then declined across gene bodies (**Fig.5b**). Close inspection shows the peak of NuMA occupancy taking place ~100 bp around the TSS (**Fig.5c**). Of note, the maximum NuMA occupancy at TSS matches well with the maximum level of oxidative breaks in NuMA-deficient cells at TSS (**Fig.4c**). Since the promoters of NRGs showed higher protection from oxidative damage than non-NRGs, we reasoned that this might be explained in part by increased NuMA occupancy and thus protection at these regions. Indeed, the promoter occupancy of NuMA was enriched at NRGs compared to non-NRGs (**Fig.5d**). Consistent with a role of NuMA in controlling Pol2 release from pausing, NuMA was more enriched at promoters of paused genes (**Fig.5e**). The increased NuMA occupancy at promoters of IERGs, pro-inflammatory genes and SSBR genes goes well with the reliance of those genes on NuMA for transcription following oxidative damage (**Fig.5f-h**). The increased NuMA occupancy at promoters was not cell type specific since it was also observed in RPE-1 cells using an orthogonal method, CUT&RUN (**Extended Data Fig.9b**). The enrichment of NuMA at promoters was further enhanced following oxidative stress (**Extended Data Fig.9b**). We also observed a similar enrichment of the NuMA interacting partner, TDP1, at promoters (**Extended Data Fig.9c**). TDP1 occupancy at the promoters of *FOS, CCN2* and *SRF* decreased upon NuMA depletion (**Extended Data Fig.9d**). Moreover, we observed an increase in NuMA occupancy at enhancers, which is consistent with the increased oxidative damage at enhancers in NuMA-deficient cells (**Extended Data Fig.9e**). Together, these data show increased NuMA occupancy at gene regulatory elements, which is prominent at the promoters of NRGs, IERGs and paused genes.

Neuronal NuMA facilitates SSBR and transcription

Collectively, using multiple cell types and experimental approaches we describe a process for safeguarding gene regulatory elements and facilitating transcription mediated by NuMA. Consistent with this role of NuMA in neural cells, we show that NuMA-deficient differentiated non-cycling SH-SY5Y neuroblastoma cells (**Extended Data Fig.10a-c, Fig.6a)** and iPSC-derived neurons **(Extended Data Fig.10e-h**) display

defective SSBR (**Fig.6b, Extended Data Fig.10d,i,j**) and reduced transcription recovery following oxidative stress (**Fig.6c, Extended Data Fig.10k**). In the neural cells, NuMA also binds p-Ser5(Pol2) and SSBR components (**Fig.6d**) and the binding with p-Ser5(Pol2) is further increased following oxidative damage (**Fig.6e, Extended Data Fig.10l**). Moreover, NuMA-deficient neural cells possessed more oxidative breaks at their promoter and enhancer regions, in unperturbed conditions and after H_2O_2-treatment (**Fig.6f,h**). NuMA was also found to be enriched at the TSS and enhancers under both conditions (**Fig.6g,i**).

Discussion

The finding that NuMA promotes SSBR and gene transcription (**Fig.6j**) is consistent with the importance of SSBR to maintain neurological function[6,14,15], the localisation of NuMA in post mitotic Purkinjie cells[42], the high expression of NuMA in the cerebellum - the tissue most affected by SSBR defects - and the structural role of NuMA in maintaining nuclear mechanics[24]. It is also consistent with the role of lamins in protecting from oxidative genome damage[17]. Moreover, the role of NuMA at promoters is consistent with its reported role in facilitating the transcription initiation of *p21*[43] and rDNA transcription initiation[44].

The high expression of NuMA in the cerebellum likely reflects a high demand for the effective activation of transcription following oxidative stress. Consistently, we show that NuMA facilitates transcription of genes that need to respond promptly to oxidative damage. However, this comes at cost if not coupled with efficient SSBR to repair the transcription-induced oxidative damage at promoters. We noticed that the expression of NuMA, TDP1 and XRCC1 in the human cerebellum is higher compared to other brain regions, whereas PARP1 expression relative to NuMA is lower (**Extended Data Fig.11a**). This suggests that the cerebellum is more "primed" for transcriptional activation, but less able to repair promoter damage caused by transcriptional activation due to low PARP1/NuMA ratio. This would be particularly problematic if SSBR proteins are limiting, either physiologically or pathologically such as in the case of their inherited deficiency in cerebellar ataxia[14,15]. Consequently, this may put the cerebellum at higher risk of SSBR defects, potentially explaining its selective vulnerability in SSBR defective disorders. However, this is a purely speculative model inspired by our

findings in neural cell culture and awaits testing in animal models by genetically altering cerebellar PARP1/NuMA ratio and examine if it leads to neuroprotection. NuMA, like XRCC1, is an essential protein and germline deficiency causes embryonic lethality. Only recently, a hypomorphic *XRCC1* mutation has been linked with cerebellar degeneration[14]. Whether NuMA hypomorphic mutations in human or brain specific deletion in mice also cause neurodegeneration warrants examination.

NuMA is a substrate for PARylation by tankyrase in a DNA independent manner and PARP3 in a DNA dependent manner[29,45]. Here, we report a previously unidentified role of PARP1, whereby PARylation stimulates NuMA interaction with SSBR proteins. We identify two physically and functionally distinct pools of NuMA, a previously characterised PARP3 bound pool which has been shown to promote mitotic function and DSBR during cell division, and a new PARP1 bound pool promoting SSBR during gene transcription. The mechanism governing the choice to assemble one or the other of these mutually exclusive complexes (NuMA/PARP3 versus NuMA/PARP1/TDP1) remains unclear. It is likely that NuMA participates in these molecular decisions based on the type of genomic insult (SSB versus DSB), the cell cycle status or the structural context of the nucleus.

NuMA is PARylated following oxidative damage at 12 Ser-PARylation sites on the globular domain that we show is critical for SSBR[32,33]. NuMA depletion increases PARylation of p-Ser5(Pol2), which provides a plausible mechanism for how NuMA promotes gene transcription. By acting as a PAR "sink" or an enrichment factor for PAR degrading enzymes, NuMA limits Pol2 PARylation, thereby promoting its availability and release from pausing (**Extended Data Fig.11b,c**). PARP1 facilitates the recruitment of the pause stabilising factor, negative elongation factor (NELF) to Pol2, thereby delaying the release from pausing and promoting transcriptional silencing[13,34]. The presence of NuMA therefore may counteract this effect, promoting the release of paused Pol2. Consistent with this model is the increased Pol2 pausing ratio in NuMA-deficient cells, enrichment of paused genes among NRGs, increased NuMA promoter occupancy at paused genes, and interaction of NuMA with pausing factors, p-Ser5(Pol2), but not p-Ser2(Pol2).

PAR polymers are markedly similar to RNA and through liquid-liquid phase separation (LLPS)[46] they can elicit mechanical forces to bring together distant genetic elements, as is the case for promoter-enhancer interactions[47]. By controlling the extent of PARylation, NuMA may regulate the dynamics of LLPS, which could exert roles beyond controlling pause release. In favour of this idea is the enrichment of NuMA at promoters and enhancers, and the observation that NuMA deficiency increases oxidative damage at both regions. It is also consistent with two recent studies in NuMA-proficient wild-type cells, reporting DRHs at promoters and enhancers[40,48]. We now show that a subset of these breaks that are induced by oxidative damage require NuMA/SSBR for repair and transcriptional regulation. Another possibility is that NuMA facilitates promoter-enhancer interactions independently of PARylation and through its long-range structural roles[24], which could bridge long distances and regulate chromosome packaging.

In addition to neurodegeneration, these findings have implications in cancer. NuMA expression and localisation is disrupted in several cancer types[49]. Inherited or acquired changes of NuMA isoform expression may lead to increased genomic instability or favour DNA repair in cancer cells resistant to DNA repair targeting therapeutics. Of note is the association of promoter specific DNA repair with cancer. Cancers that are dependent on nucleotide excision repair (NER) such as melanomas possess high rate of NER somatic mutation hotspots at active gene promoters[50] unlike cancers that are not dependent on NER such as colon cancer. Whether cancers that exhibit perturbed NuMA expression possess promoter specific signatures for SSBR somatic mutations such as GC>TA mutations warrants testing.

In summary, these findings underscore the importance of DNA repair at gene regulatory elements and describe a NuMA mediated SSBR process that promotes this function.

Acknowledgements

We thank Laura Ferraiuolo for advice on iPSC derived neurons, Sophie Dumont for providing the hTERT (RPE1) cells containing the stably integrated spCas9 (Tet-On promoter) and NuMA sgRNA, the Sheffield Wolfson Light Microscopy facility for the imaging experiments and the Sheffield biOMICS Facility for the protein mass

spectrometry experiments. This work was funded by a Wellcome Trust Investigator Award (103844) and a Lister Institute of Preventative Medicine Fellowship (137661) to S.E.K.

Author contribution

S.R performed the protein interaction and *in vitro* transcription experiments. A.A. performed the PARylation experiments. S.R and A.A. performed the imaging, DNA strand break repair and genome-wide occupancy and oxidative damage experiments. J.P performed all the bioinformatics analyses with advice and support from I.S. K.L., N.P, C.L. conducted comet assays, *in vitro* binding and ubiquitination. A.E.A.M. performed and analysed the protein mass spectrometry experiments. C.D.S.Z generated the iPSC derived neurons. M.J. assisted with the comet assays and performed the cell cycle analysis. S.E.K, S.R. and A.A wrote the manuscript. All authors edited the manuscript. S.E.K conceived the study, led and managed the project.

Competing interests

The authors declare no competing interests.

Additional Information

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to S.E.K.

Peer review information

Reprints and permissions information is available at http://www.nature.com/reprints

References

1. Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biology* **13**, 297–311 (2012).

2. Xu, G.-L. & Wong, J. Oxidative DNA demethylation mediated by Tet enzymes. *Natl Sci Rev* **2**, 318–328 (2015).

3. Margolin, Y., Cloutier, J.-F., Shafirovich, V., Geacintov, N. E. & Dedon, P. C. Paradoxical hotspots for guanine oxidation by a chemical mediator of inflammation. *Nat Chem Biol* **2**, 365–366 (2006).

4. Pan, L. *et al.* Oxidized Guanine Base Lesions Function in 8-Oxoguanine DNA Glycosylase-1-mediated Epigenetic Regulation of Nuclear Factor κB-driven Gene Expression. *J Biol Chem* **291**, 25553–25566 (2016).

5. Fleming, A. M., Ding, Y. & Burrows, C. J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc National Acad Sci* **114**, 2604–2609 (2017).

6. El-Khamisy, S. F. *et al.* Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. *Nature* **434**, 108–113 (2005).

7. Nilsen, L., Forstrøm, R. J., Bjørås, M. & Alseth, I. AP endonuclease independent repair of abasic sites in Schizosaccharomyces pombe. *Nucleic acids research* **40**, 2000–2009 (2012).

8. Interthal, H., Chen, H. J. & Champoux, J. J. Human Tdp1 cleaves a broad spectrum of substrates, including phosphoamide linkages. *The Journal of biological chemistry* **280**, 36518–36528 (2005).

9. Lebedeva, N. A., Rechkunova, N. I., El-Khamisy, S. F. & Lavrik, O. I. Tyrosyl-DNA phosphodiesterase 1 initiates repair of apurinic/apyrimidinic sites. *Biochimie* **94**, 1749–1753 (2012).

10. 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 (2001).

11. El-Khamisy, S. F., Masutani, M., Suzuki, H. & Caldecott, K. W. A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. *Nucleic acids research* **31**, 5526–5533 (2003).

12. Das, B. B. *et al.* PARP1-TDP1 coupling for the repair of topoisomerase I-induced DNA damage. *Nucleic acids research* **42**, 4435–4449 (2014).

13. Awwad, S. W., Abu-Zhayia, E. R., Guttmann-Raviv, N. & Ayoub, N. NELF-E is recruited to DNA double-strand break sites to promote transcriptional repression and repair. *Embo Rep* **18**, 745–764 (2017).

14. Hoch, N. C. *et al.* XRCC1 mutation is associated with PARP1 hyperactivation and cerebellar ataxia. *Nature* **541**, 87–91 (2017).

15. Takashima, H. *et al.* Mutation of TDP1, encoding a topoisomerase I–dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. *Nature Genetics* **32**, 267–272 (2002).

16. Giampetruzzi, A. *et al.* Modulation of actin polymerization affects nucleocytoplasmic transport in multiple forms of amyotrophic lateral sclerosis. *Nat Commun* **10**, 3827 (2019).

17. Maynard, S. *et al.* Lamin A/C promotes DNA base excision repair. *Nucleic Acids Res* **47**, 11709–11728 (2019).

18. Lydersen, B. K. & Pettijohn, D. E. Human-specific nuclear protein that associates with the polar region of the mitotic apparatus: Distribution in a human/hamster hybrid cell. *Cell* **22**, 489–499 (1980).

19. Hueschen, C. L., Kenny, S. J., Xu, K. & Dumont, S. NuMA recruits dynein activity to microtubule minus-ends at mitosis. *eLife* **6**, R162 (2017).

20. Merdes, A., Ramyar, K., Vechio, J. D. & Cleveland, D. W. A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assembly. *Cell* **87**, 447–458 (1996).

21. Merdes, A. & Cleveland, D. W. The role of NuMA in the interphase nucleus. *J Cell Sci* **111 (Pt 1)**, 71–79 (1998).

22. Vidi, P.-A. *et al.* NuMA promotes homologous recombination repair by regulating the accumulation of the ISWI ATPase SNF2h at DNA breaks. *Nucleic acids research* **42**, 6365–6379 (2014).

23. Salvador-Moreno, N. *et al.* The nuclear structural protein NuMA is a negative regulator of 53BP1 in DNA double-strand break repair. *Nucleic acids research* **47**, 2703–2715 (2019).

24. Serra-Marques, A. *et al.* The mitotic protein NuMA plays a spindle-independent role in nuclear formation and mechanics. *J Cell Biol* **219**, (2020).

25. Bahrami, S. & Drabløs, F. Gene regulation in the immediate-early response process. *Adv Biological Regul* **62**, 37–49 (2016).

26. Samson, L. D. A target to suppress inflammation. *Science* **362**, 748–749 (2018).

27. Visnes, T. *et al.* Small-molecule inhibitor of OGG1 suppresses proinflammatory gene expression and inflammation. *Science* **362**, 834–839 (2018).

28. Dellino, G. I. *et al.* Release of paused RNA polymerase II at specific loci favors DNA double-strand-break formation and promotes cancer translocations. *Nature Genetics* **47**, 1–21 (2019).

29. Boehler, C. *et al.* Poly(ADP-ribose) polymerase 3 (PARP3), a newcomer in cellular response to DNA damage and mitotic progression. *Proc Natl Acad Sci U S A* **108**, 2783–2788 (2011).

30. Chang, P., Coughlin, M. & Mitchison, T. J. Tankyrase-1 polymerization of poly(ADP-ribose) is required for spindle structure and function. *nature cell biology* **7**, 1133–1139 (2005).

31. Haren, L. & Merdes, A. Direct binding of NuMA to tubulin is mediated by a novel sequence motif in the tail domain that bundles and stabilizes microtubules. *J Cell Sci* **115**, 1815–1824 (2002).

32. Hendriks, I. A., Larsen, S. C. & Nielsen, M. L. An advanced strategy for comprehensive profiling of ADP-ribosylation sites using mass spectrometry-based proteomics. *Mol Cell Proteomics* **18**, 1010 –1026 (2019).

33. Palazzo, L. *et al.* Serine is the major residue for ADP-ribosylation upon DNA damage. *eLife* **7**, 935 (2018).

34. Páhi, Z. G., Borsos, B. N., Pantazi, V., Ujfaludi, Z. & Pankotai, T. PARylation During Transcription: Insights into the Fine-Tuning Mechanism and Regulation. *Cancers* **12**, 183 (2020).

35. Ting, X. *et al.* USP11 acts as a histone deubiquitinase functioning in chromatin reorganization during DNA repair. *Nucleic Acids Res* **47**, 9721–9740 (2019).

36. Kuznetsov, N. A. *et al.* Pre-steady state kinetics of DNA binding and abasic site hydrolysis by tyrosyl-DNA phosphodiesterase 1. *J Biomol Struct Dyn* **35**, 2314–2327 (2017).

37. Poetsch, A. R., Boulton, S. J. & Luscombe, N. M. Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. *Genome biology* **19**, 215 (2018).

38. Wu, J., McKeague, M. & Sturla, S. J. Nucleotide-Resolution Genome-Wide Mapping of Oxidative DNA Damage by Click-Code-Seq. *J Am Chem Soc* **140**, 9783–9787 (2018).

39. Mao, P. *et al.* Genome-wide maps of alkylation damage, repair, and mutagenesis in yeast reveal mechanisms of mutational heterogeneity. *Genome Res* **27**, 1674–1684 (2017).

40. Reid, D. A. *et al.* Incorporation of a nucleoside analog maps genome repair sites in postmitotic human neurons. *Science* **372**, 91–94 (2021).

41. Wu, W. *et al.* Neuronal Enhancers are Hotspots For DNA Single-Strand Break Repair. *Biorxiv* 2020.12.16.423085 (2020) doi:10.1101/2020.12.16.423085.

42. Radulescu, A. E. & Cleveland, D. W. NuMA after 30 years: the matrix revisited. *Trends Cell Biol* **20**, 214–222 (2010).

43. Ohata, H. *et al.* NuMA is required for the selective induction of p53 target genes. *Molecular and cellular biology* **33**, 2447–2457 (2013).

44. Jayaraman, S. *et al.* The nuclear mitotic apparatus protein NuMA controls rDNA transcription and mediates the nucleolar stress response in a p53-independent manner. *Nucleic acids research* **45**, 11725–11742 (2017).

45. Chang, W., Dynek, J. N. & Smith, S. NuMA is a major acceptor of poly(ADP-ribosyl)ation by tankyrase 1 in mitosis. *Biochem J* **391**, 177–184 (2005).

46. Altmeyer, M. *et al.* Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat Commun* **6**, 8088 (2015).

47. Nair, S. J. *et al.* Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat Struct Mol Biol* **26**, 193–203 (2019).

48. Wu, W. *et al.* Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* **593**, 440–444 (2021).

49. Endo, A., Moyori, A., Kobayashi, A. & Wong, R. W. Nuclear mitotic apparatus protein, NuMA, modulates p53-mediated transcription in cancer cells. *Cell death & disease* **4**, e713–e713 (2013).

50. Perera, D. *et al.* Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. *Nature* **532**, 259–263 (2016).

Figure legends

Figure 1. NuMA deficiency perturbs SSBR and transcription recovery. (**a**) NuMA expression across brain tissues of the GTEx v8 dataset. (**b**) *Left;* Violin plots showing comet tail moments from 250 MRC5 cells immediately after or during recovery from H_2O_2. (two-sided, unpaired t-test). *Right*; Bars represent percentage breaks remaining. Error bars represent ±s.e.m., n=4, (two-sided, unpaired t-test). *Inset;* Representative immunoblot of lysates used for comet assays. (**c**) Clonogenic survival of H_2O_2-treated MRC5 cells. Error bars represent ±s.e.m, n=3, (two-sided, unpaired t-test). (**d**) MRC5 cells pre-treated with DMSO/DRB before H_2O_2 for alkaline comet. Bars represent percentage breaks remaining. Error bars represent ±s.e.m., n=3, (two-sided, unpaired t-test). (**e**) Nascent transcripts labelled in MRC5 cells with EU. Scatter plots showing fold change of average fluorescence. Error bars represent ±s.e.m., n=3, (two-sided, unpaired t-test). (**f**) Overlap of genes differentially upregulated in WT(+H_2O_2) cells with genes differentially downregulated in KD(+H_2O_2) cells. Background gene set is all genes expressed in 4sU-seq data. *OR,* odds ratio, (two-sided Fisher's exact test). (**g**) Overlap of IERGs with NRGs. OR, odds ratio, (two-sided Fisher's exact test). (**h** and **j**) Heat map showing $\log_2$ fold change in expression levels of (**h**) IERGs and (**j**) pro-inflammatory genes. (**i**) Relative transcript level of *FOS*, during a 90 min recovery after H_2O_2 treatment. Error bars represent the range, n=2 biological repeats done in triplicates. (**k**) Overlap of (*left*) paused and (*right*) highly paused genes with NRGs. OR, odds ratio, (two-sided Fisher's exact test). (**l**) Distribution of pausing indices for lowly-paused (*left*), moderately-paused (*middle*) and highly-paused (*right*) NRGs and Non-NRGs. Boxplot summary statistics: Internal line = median, upper and lower hinges = upper and lower quartiles, top and bottom whiskers = largest and smallest values no further than 1.5 times interquartile range from the upper and lower hinges respectively. n.s.=non-significant (two-sided Wilcoxon rank sum test).

Figure 2. NuMA interacts with SSBR components and the initiating form of Pol2. (**a**) Immunoblotting of GFP-immunoprecipitates from GFP-empty- or GFP-NuMA-expressing HEK293 cells after H_2O_2 treatment, n=3. (**b**) Immunoblotting of myc-immunoprecipitates from myc-TDP1-transfected HEK293 cells treated as in (a), n=3. (**c**) Immunoblotting of Pol2- immunoprecipitates from HEK293 cells transfected and treated as in (b). Pol2(O) and Pol2(A); hyper- and hypo-phosphorylated Pol2,

respectively. (**d**) *Left;* Immunoblotting of Ser5(Pol2)- immunoprecipitates from HEK293 cell lysates transfected and treated as in (b). *Right*; quantification of the SSBR complex from insoluble chromatin, normalised to p-Ser5(Pol2) and presented as a ratio of $+H_2O_2/-H_2O_2$. Error bars represent the range, n=2. (**e**) *Left*; a representative image of PLA on HEK293 cells treated as in (a). *Right*; Number of p-Ser2/NuMA and p-Ser5/NuMA foci/nuclei (n=30) (two-sided, unpaired t-test). Scale bar-5μm (**f**) *Left*; Immunoblotting of myc-immunoprecipitates from myc-TDP1-transfected HEK293 cells treated with DMSO/PARGi. *Right*; quantification of the increase in pulldown of NuMA and PARP1, normalised to myc-TDP1 and presented as a ratio of PARGi/DMSO. Error bars represent the range, n=2. (**g**) Recombinant His-tagged TDP1/BSA (control) incubated with *in vitro* transcribed/translated ^{35}S-Methionine NuMA in presence/absence of purified PAR. (**h**) Quantification of ^{35}S from (g). Error bars represent ± s.e.m., n=3, (two-sided, unpaired t-test). (**i**) Recombinant His-tagged TDP1/eIF4a (control) incubated with *in vitro* transcribed/translated ^{35}S-Methionine NuMA in presence/absence of PARylation buffer. ^{35}S signal (*Top*), Ponceau-S staining (*Middle*) and anti-PAR immunoblotting (*Bottom*). (**j**) Quantification of ^{35}S from (i) normalised to TDP1 or eIF4a and presented as fold increase compared to non-PARylated conditions. Error bars represent ±s.d., n=4, (One-way ANOVA). (**k**) Immunoblotting of Myc- immunoprecipitates from wild-type or PARP1-knockout TK6 cells electroporated with myc-TDP1. (**l**) Immunoblotting of myc- immunoprecipitates from wild-type or PARP3-knockout HEK293 cells. n=3. (**m**) Immunoblotting of GFP-immunoprecipitates from HEK293 cells co-transfected with empty-GFP or GFP-PARP1, GFP-PARP3 alongside myc-TDP1, n=3.

Figure 3. NuMA regulates Pol2 pausing and reduces Pol2 PARylation (**a**) Metagene profile of fold change in counts per million relative to input over TSS in RPE1 cells under unperturbed conditions. (**b**) Distribution of Pol2 pausing ratios for (*left*) all genes and (*right*) paused genes. Pausing ratio for each gene was averaged across replicates. Boxplot summary statistics: Internal line = median, upper and lower hinges = upper and lower quartiles, top and bottom whiskers = largest and smallest values no further than 1.5 times interquartile range from the upper and lower hinges respectively, (two-sided Wilcoxon rank sum test). (**c**) IGV snapshots of representative track of Pol2 ChIP-Seq at promoters of NRGs from data in (a). (**d**) Metagene profile of fold change in counts

per million relative to input over TSS after H_2O_2 treatment. (**e**) IGV snapshots of representative track of Pol2 ChIP-Seq at promoters of NRGs from data in (d). (**f**) *Left*; RPE-1 cells were treated with H_2O_2 and PAR-immunoprecipitants analysed by immunoblotting. *Right*; PARylated p-Ser5(Pol2), PARP1 and p-XRCC1 normalised to total PAR signal (bait IP) and presented as fold change in NuMA-deficient cells compared to wild-type cells. Error bars represent ±s.e.m., n=3, (two-sided, unpaired t-test). **(g)** *Left;* Schematic diagram for sequential immunoprecipitation. *Middle*; H_2O_2-treated RPE-1 cells subjected to sequential immunoprecipitation of the soluble nucleoplasm fraction, and analysed by immunoblotting. *Right*; The total PAR signal was normalized to the bait, p-Ser5 Pol2 and presented as a fold change in NuMA-deficient cells compared to wild-type cells. Error bars represent ±s.e.m., n=3, (two-sided, unpaired t-test). (**h** and **i**) RPE1 cells were treated with H_2O_2, and (**h**) H2A-K119 or (**i**) H2B-K120 monoubiquitination were measured by immunofluorescence. Error bars represent ± s.e.m, where the mean intensity for each cell is normalized to the average mean intensity at R0, n=3, (two-sided, unpaired t-test), Scale bar-10μm.

Figure 4. NuMA depletion increases oxidative damage at promoters and enhancers. (**a**) Schematic diagram of OGAP-Seq capturing AP-sites and 8-oxoG damaged bases at approximately 250-bp resolution on a genome-wide scale. (**b**) Representative immunoblotting to confirm NuMA knockdown in RPE-1 cells. n=2. (**c**) Metagene profile of OGAP-Seq showing fold change in counts per million relative to input over the TSS for wild-type (*left*) and NuMA-depleted cells (*right*). (**d**) Metagene profile of percent change in normalised OGAP-Seq signal upon NuMA depletion relative to wild-type cells. 0-1000: Upstream of TSS, 1000-2000: gene body (exons), 2000-3000: Downstream of TTS. The regions flanking the gene body are 2500bps. (**e**) The distribution of counts per million for OGAP-Seq reads overlapping the promoters of protein coding NRGs and Non-NRGs. Genes with no OGAP-Seq signal in their promoter region were filtered out prior to analysis. Boxplot summary statistics: Internal line = median, upper and lower hinges = upper and lower quartiles, top and bottom whiskers = largest and smallest values no further than 1.5 times interquartile range from the upper and lower hinges respectively, (two-sided Wilcoxon rank sum test). (**f**) IGV snapshots of representative tracks of OGAP-Seq at selected IERGs. (**g**) Metagene profiles of fold change in counts per million relative to input over enhancer regions for OGAP-Seq. The regions flanking the enhancers are 2500 bp.

Figure 5. NuMA occupancy increases at promoters. (**a**) EMSA showing the binding of $NuMA^{GD}$ to a 3'-Cy5 labelled intact oligonucleotide (*top left*), an oligonucleotide harboring a single-strand gap (*top middle*) or a nick (*top right*). *Bottom*; The fraction of NuMA-DNA complex was calculated as a proportion of total DNA in each condition and presented as fold change relative to dsDNA. Error bars represent ± s.e.m, n=3. (**b**) Metagene profile of fold change in counts per million relative to input over gene windows for NuMA and IgG ChIP-seq. 0–1000: Upstream of TSS, 1000-2000: gene body (exons), 2000-3000: Downstream of TTS. The regions flanking the gene body are 2500 bps (**c**) Metagene profile of fold change in counts per million relative to input over TSS windows for NuMA and IgG ChIP-seq. The TSS is situated at base 3000. (**d**) Distribution of counts per million for NuMA ChIP-seq reads overlapping the promoter regions of protein coding NRGs and non-NRGs. Boxplot summary statistics: Internal line = median, upper and lower hinges = upper and lower quartiles, top and bottom whiskers = largest and smallest values no further than 1.5 times interquartile range from the upper and lower hinges respectively, (two-sided Wilcoxon rank sum test). (**e**) Distribution of counts per million for NuMA ChIP-seq reads overlapping the promoter regions of protein coding paused and non-paused genes. Boxplot summary statistics as in (d), (two-sided Wilcoxon rank sum test). (**f**) Distribution of counts per million for NuMA ChIP-seq reads overlapping the promoter regions of protein coding IERGs and Non-IERGs. Boxplot summary statistics as in (d). (two-sided Wilcoxon rank sum test). (**g and h**) IGV snapshots of a representative track of the NuMA ChIP-Seq at promoters of (**g**) IERGs and pro-inflammatory genes and (**h**) SSBR and DSBR genes.

Figure 6. NuMA promotes SSBR and gene transcription in differentiated SH-SY5Y neuroblastoma cells. (**a**) A representative immunoblot of lysates from control and NuMA-deficient differentiated SH-SY5Y cells, n=4. (**b**) Bars represent percentage breaks remaining after removal of H_2O_2 and recovery for 30 and 60 minutes. Error bars represent ± s.e.m., n=4, (two-sided, unpaired t-test). (**c**) Nascent transcripts labelled in SH-SY5Y cells with EU. Data are scatter plots of fold change of average fluorescence. Error bars represent ± s.e.m., n=3, (two-sided, unpaired t-test) (**d**) Pol2 immunoprecipitates from Myc-TDP1-transfected SH-SY5Y were analysed by immunoblotting, n=3. (**e**) PLA was performed on SH-SY5Y cells. *Top*; A representative image, which is a merge of the red and blue channels, is shown. *Bottom*; Violin plots

showing the number of p-Ser5 Pol2/NuMA foci/nuclei (n=30) from 3 biological replicates, (two-sided, unpaired t-test). Scale bar-5μm (**f and g**) Metagene profile fold change in counts per million over the TSS relative to (**f**) input for OGAP-Seq and (**g**) relative to IgG for NuMA CUT&RUN. (**h and i**) Metagene profiles of fold change in counts per million over enhancer regions relative to (**h**) input for OGAP-Seq and (**i**) IgG for NuMA CUT&RUN. The regions flanking the enhancers are 2500 bp. (**j**) A model showing NuMA interaction with the initiating form of Pol2 and SSBR components to promote gene transcription. IERGs, immediate early response genes; PIGs, proinflammatory genes.

Extended Data Figure 1. Expression of SSBR proteins in different brain regions. Expression of TDP1, XRCC1 and PARP1 across different brain regions in the GTEx v8 dataset.

Extended Data Figure 2. DRB treatment abolishes NuMA-dependent defect in SSBR. siSCR and siNuMA-transfected MRC5 cells were pre-treated with DMSO or DRB (10 μM) for 2 hours before H_2O_2 treatment and recovered for 30 and 60 minutes. Representative violin plots show the spread of comet tail moments from 250 nuclei, n=5, (two-sided, unpaired t-test)

Extended Data Figure 3. NuMA promotes transcription following oxidative damage. (**a**) Immunoblotting of lysates from RPE-1 cells for 4sU-Seq, n=2. (**b**) A schematic for 4sU-Seq. (**c**) 4sU-Seq showing $-Log_{10}$ plotted against Log_2 fold change for genes differentially expressed in WT ($+H_2O_2$) versus ($-H_2O_2$) cells. (DEseq2, two-sided Wald test, Benjamini-Hochberg procedure adjusted p-values≤0.05). Log_2 fold changes with respect to WT ($+H_2O_2$) cells. (**d**) Overlap of paused genes with genes upregulated (*left*) and downregulated (*right*) in WT ($+H_2O_2$) versus ($-H_2O_2$) cells. (**e**) 4sU-Seq showing $-Log_{10}$ against Log_2 fold change for genes differentially expressed in KD versus WT ($+H_2O_2$) cells. (DEseq2, two-sided Wald test, Benjamini-Hochberg adjusted p-values≤0.05). Log_2 fold changes to KD cells. (**f**) Overlap of paused genes with genes upregulated (*left*) and downregulated (*right*) in KD versus WT ($+H_2O_2$) cells. (**g**) Overlap of genes containing fragile promoters with NRGs. The background set is the total set of differentially expressed genes. (**h**) Distribution of average transcription

levels in Log_2 transcripts per million (TPM) for NRGs and genes differentially upregulated in WT(+H_2O_2) and are averaged across replicates. Only genes with TPM values ≥1 across WT(+H_2O_2) replicates were included in the analysis. Boxplot summary statistics: Internal line = median, upper and lower hinges = upper and lower quartiles, top and bottom whiskers = largest and smallest values no further than 1.5 times interquartile range from the upper and lower hinges respectively. (two-sided Wilcoxon rank sum test) (**i**) Scatterplot of Log_2 fold changes in transcription level in WT versus KD(+H_2O_2) cells against gene length in kilobases (KB) on a log_{10} scale. Correlation coefficient (R) by Spearman's rank and significance using the asymptotic t approximation (two-sided). (**j**) *Left*; Fold change in nascent RNA transcripts for the indicated SSBR genes. *Right*; IGV snapshots of a representative track of the 4sU-Seq profiles of SSBR and DSBR genes.

Extended Data Figure 4. Clustered proteins identified by mass spectrometry in GFP pull downs. (**a**) HEK-293 cells transfected with GFP-empty or GFP-NuMA were treated with H_2O_2 and recovered for 10 min. Cell lysates were subjected to GFP-trap immunoprecipitation, on-bead trypsin digestion and subsequent analysis by mass spectrometry. Volcano plot showing quantitative enrichment of NuMA interacting proteins identified from statistical analysis of data sets using Student t-test. Solid lines indicate significant enrichment of interacting proteins after filtering for a false discovery rate of 0.05 and an artificial within groups variance S0=1. Components of the Pol2 and SSBR machineries are highlighted as red squares. (**b**) Heat map of clustered proteins identified by mass spectrometry in GFP pull downs. Hierarchical clustering of LFQ protein intensity shows two dominant sample clusters, GFP and GFP-NuMA-L (LTR), according to sample conditions. (**c**) HEK293 cells were treated with either DMSO or inhibitors of PARG (PDD00017273), PARP3 (ME0328) or PARP1 (Olaparib). Western immunoblotting was carried out to show poly-ADP ribosylation (pADPR) levels. n=3. (**d**) Heat map of clustered proteins identified by mass spectrometry in GFP pull downs. Hierarchical clustering of LFQ protein intensity shows three dominant sample clusters, GFP, GFP-NuMA-S (STR) and GFP-NuMA-L (LTR), according to sample conditions. On the right-hand side, zoomed heat map of NuMA-S (STR) and NuMA-L (LTR) intensity profiles across samples.

Extended Data Figure 5. NuMA-L promotes SSBR and transcription recovery functions. (**a**) A schematic of human NuMA isoforms. (**b**) GFP- immunoprecipitates from GFP-NuMA-L and NuMA-S-transfected HEK-293 cells, treated with H_2O_2 and subjected to mass spectrometric analysis. Annotated tandem mass spectra of NuMA-L (*top*) and NuMA-S (*bottom*) specific peptides were identified in the immunoprecipitates. Precursor mass deviation was 2.07 ppm for the peptide LTAQVEQLEVFQR and 0.66 ppm for the peptide LTAQVEELSK. (**c**) Quantitative enrichment of interacting proteins between NuMA-L and NuMA-S from (b). NuMA-interacting proteins were identified from statistical analysis using Student t-test. Solid lines indicate enriched interacting proteins after filtering for FDR of 0.05 and an artificial within groups variance, S0 = 1. Annotated blue and red dots indicate NuMA-S (STR), NuMA-L (LTR) and additional proteins of interest. Non-annotated pale red dots represent proteins significantly enriched in each NuMA pull down compared to GFP controls. Green dots are proteins enriched in both NuMA-S and NuMA-L pull-downs and red dots are proteins enriched only in NuMA-L pull-downs. (**d**) Immunoblotting of GFP- immunoprecipitates from GFP-empty, GFP-NuMA-L and GFP-NuMA-S-transfected HEK-293 cells, treated with H_2O_2, n=3. (**e**) Immunoblotting of transfected MRC5 cells, n=5. (**f**) MRC5 cells were transfected as in (e), treated with H_2O_2 and recovered in media. A representative plot showing spread of comet tail moments from 250 nuclei. (two-sided, unpaired, t-test) (**g**) Bars represent percentage DNA strand breaks remaining. Error bars represent ±s.e.m, n=5, (two-sided, unpaired t-test) (**h**) Clonogenic survival of MRC5 cells transfected as in (e) and presented on a semi-log scale, n=3, (two-sided, unpaired t-test). (**i**) Immunoblotting of transfected MRC5 cells (**j**) MRC5 cells were transfected as in (i) and nascent transcripts labelled with EU. Scatter plots show fold change of average fluorescence. Error bars represent ± s.e.m., n=3, (two-sided, unpaired t-test).

Extended Data Figure 6. NuMA C-terminal tail or a mitotic separation-of-function mutant promotes SSBR. (**a**) Schematic of the truncation mutants generated from full-length NuMA with the C-terminal tail (1700-2115) designated as globular domain (GD) and the microtubule-binding domain (1866-1936) as MD. (**b**) HEK293 cells were co-transfected with either full-length or different truncated versions of NuMA as in (a), alongside myc-TDP1. GFP immunoprecipitants were analysed by immunoblotting, n=3. (**c**) Immunoblotting of cell lysates from MRC5 cells transfected with siSCR or siNuMA followed by complementation with targeting resistant GFP-tagged NuMA-FL, NuMA-

GD, or EV. (**d**) Representative violin plot showing spread of comet tail moments from 250 MRC5 cells transfected as in (c), treated with H_2O_2 and recovered at indicated time points, (two-sided, unpaired t-test) (**e**) Bars represent percentage breaks remaining after removal of H_2O_2 and recovery in complete medium from (d). Error bars represent ±s.e.m, n=3, (two-sided, unpaired t-test). (**f**) HEK293 cells were transfected with either NuMA-FL or NuMA-MD followed by nocodazole treatment before harvest. Representative image showing NuMA (green), alpha-tubulin (red), nuclei (DAPI/blue), n=6, Scale bar-10μm. (**g**) The percentage of cells with NuMA at the mitotic poles was quantified. Error bars represent ±s.e.m, n=6 (100 cells counted), (two-sided, unpaired t-test). (**h**) Immunoblotting of GFP-immunoprecipitates from HEK-293 cells transfected with either GFP-empty, NuMA-FL or NuMA-MD, alongside myc-TDP1, and treated with H_2O_2, n=3. (**i**) Immunoblotting of lysates from MRC5 cells transfected with siSCR or siNuMA and complemented with targeting-resistant GFP-tagged NuMA-FL, NuMA-MD, or EV. (**j**) MRC5 cells were transfected as in (i), treated with H_2O_2 and recovered in media at the indicated time points. Representative violin plot showing spread of comet tail moments from 150 cells. (two-sided, unpaired t-test) (**k**) Bars represent percentage strand breaks remaining after removal of H_2O_2 and recovery in complete medium. Error bars represent ±s.e.m, n=3, (two-sided, unpaired t-test).

Extended Data Figure 7. NuMA promotes the enrichment of TDP1 at damaged chromatin. (**a**) *Top;* The knockdown efficiency of NuMA was analysed by immunoblotting using anti-NuMA antibodies. *Bottom;* The expression levels of GFP–TDP1 was measured in cells used in photo-bleaching experiments at time 0 (sec). Bars represent average ± s.e.m., n=3. (**b**) MRC5 cells were plated onto glass-bottom dishes and co-transfected with GFP-TDP1 and siSCR or siNuMA. Cells were pre-incubated with DMSO or olaparib. Cells expressing similar total GFP signal were locally irradiated with an ultraviolet A laser (405 nm), and GFP–TDP1 accumulation at the site of damage was quantified for the indicated time points. Scale bar-5μm. (**c**) Data are plotted as the average percentage fluorescence (arbitrary units) in micro-irradiated tracks ± s.e.m from ∼30 cells measured, n=3, (two-sided, unpaired t-test) (**d**) Lysates of MRC5 cells transfected with siSCR or siNuMA and siTDP1 either individually or in combination was analysed by immunoblotting, n=3. (**e**) MRC5 cells transfected as in (d), treated with H_2O_2 and recovered for 0, 30, and 60 min. Representative violin plot showing spread of comet tail moments from 150 cells, n=3. (**f**) Bars represent

percentage breaks remaining after removal of H_2O_2 and recovery in complete medium for 30 and 60 min. Error bars represent ± s.e.m, n=3. (two-sided, unpaired t-test) (**g**) Survival of MRC5 cells transfected as in (d) was compared using the indicated doses of H_2O_2. Results are presented on a semi-log scale and represent the average of three biological replicates ± s.e.m. ns - not significant (two-sided, unpaired t-test)

Extended Data Figure 8. NuMA interacts with PAF1. (**a**) HEK-293 cells were transfected with plasmids encoding GFP or GFP-NuMA-L, treated with H_2O_2 and cell lysates analysed by mass spectrometry. Volcano plot showing quantitative enrichment of NuMA interacting proteins identified from statistical analysis of data sets using Student t-test. Solid lines indicate significant enrichment of interacting proteins after filtering for a false discovery rate of 0.05 and an artificial within groups variance S0=1. Components of the PAF1 complex are highlighted as blue squares. (**b**) HEK-293 cells were treated with H_2O_2 and fractionated into insoluble chromatin bound fraction and soluble nucleoplasmic fraction. Endogenous RNA Pol2 was immunoprecipitated using p-Ser5 Pol2 antibody. Immunoprecipitated complexes were analysed by immunoblotting using antibodies against p-Ser5 Pol2, NuMA and PAF1, n=3.

Extended Data 9. NuMA deficiency increases 8-oxoG levels, damage at enhancers and decreases TDP1 occupancy at promoters, while oxidative stress increases NuMA and TDP1 availability at promoters. (**a**) RPE-1 cells were either untreated, treated with H_2O_2 or $KBrO_3$. Genomic DNA was extracted and the levels of 8-oxoG per 10,000,000 deoxyguanosine (dG) was quantified by HPLC-QQQ mass spectrometry. The error bars represent ± s.e.m., n=3, (two-sided, unpaired t-test) (**b** and **c**) Metaprofiles of fold change in counts per million relative to IgG for (**b**) NuMA and (**c**) TDP1 CUT&RUN in RPE-1 cells. (**d**) RPE-1 cells were transfected with Flag-TDP1 and ChIP-qPCR was conducted at the promoters of *FOS*, *CCN2* and *SRF*. Error bars represent ± s.e.m., n=3. (two-sided, unpaired t-test) (**e**) Metaprofiles of fold change in counts per million relative to input over enhancer regions for OGAP-Seq in WT and NuMA KD RPE-1 cells under unperturbed conditions. The regions flanking the enhancers are 2500 bp.

Extended Data Figure 10. NuMA-deficient neuronal cells exhibit defective SSBR and transcription recovery. (**a**) Brightfield image showing differentiation of SH-

SY5Y cells to neuronal morphology (similar to primary neurons with long processes) after growth in low-serum media supplemented with retinoic acid (RA). Scale bar-100 µm. (**b**) Violin plots showing quantification of neurite outgrowth following RA differentiation, quantified using Fiji (ImageJ) software with NeuronJ plugin, n=16/condition, (two-sided, unpaired t-test) (**c**) MAP2 immunofluorescence images of undifferentiated and differentiated SH-SY5Y cells, n=2, scale-bar 10µm. (**d**) Representative violin plots showing spread of comet tail moments from 200 differentiated SH-SY5Y cells immediately after or during recovery from H_2O_2 (two-sided, unpaired t-test) (**e**) Brightfield image showing human iPSC-derived motor neurons, n=3, scale bar-100µm. (**f**) Representative immunofluorescence images of human shSCR or shNuMA-transduced iPSC-derived motor neurons showing relative levels of NuMA (green), n=2, scale bar-10µm. (**g**)Violin plot showing NuMA fluorescence intensity n=20 cells/condition (two-sided, unpaired t-test). (**h**) Immunoblot of lysates from shSCR and shNuMA-transduced neurons, n=2. (**i**) Bars represent percentage breaks remaining after removal of H_2O_2 and recovery in complete medium, as quantified by the alkaline comet assay. Error bars represent the range, n=2. (**j**) Representative violin plots showing spread of comet tail moments from 100 cells immediately after or during recovery from H_2O_2 (two-tailed, unpaired t-test) (**k**) shSCR or shNuMA-transduced motor neurons were treated with H_2O_2 and nascent transcripts were labelled by EU. Scatter plots shows fold change of average fluorescence. Error bars represent ± s.e.m, n=4 (average 100 nuclei/coverslip) (two-sided, unpaired t-test). (**l**) PLA was performed on motor neurons that were either mock-treated or H_2O_2-treated. For each sample, a representative image, which is a merge of the red and blue channels, is shown. Violin plots show mean signal intensity of p-Ser5 Pol2-NuMA complex/nuclei, n=125 cells from three separate wells, (two-sided, unpaired t-test), scale bar-10µm.

Extended Data Figure 11. Depicting the function of NuMA in response to oxidative stress. (**a**) Distribution of PARP1/NuMA expression across brain tissues of the GTEx v8 dataset. (**b**) A schematic diagram depicting NuMA's role in response to oxidative stress. NuMA is enriched at the promoters and interacts with RNA Polymerase 2 (Pol2) and the SSBR proteins, TDP1 and PARP1. As a result of the enrichment of this complex, cells are efficient at repairing oxidative damage at promoters. NuMA acts as a PAR sink, ensuring the physiological levels of Pol2 PARylation that takes place in the

nucleoplasm and thus increases the availability of Pol2 at promoters and enhances its release from pausing. This results in increasing the transcription of NuMA regulated genes (NRGs) (**c**) Decreased levels of NuMA increases the oxidative damage at promoters. Moreover, the absence of NuMA increases Pol2 PARylation in the nucleoplasm, thereby decreasing Pol2 availability at promoters, resulting in a reduction of the transcription of NRGs.

Materials and methods

Reagents

All reagents are purchased from Sigma Aldrich or Fisher Scientific, unless otherwise specified.

Cell Culture

HEK-293 and MRC5 cells were cultured in Minimum Essential Medium supplemented with 10% fetal bovine serum, 2 mM L -Glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin. TK6 cells were cultured in RPMI-1640 medium supplemented with 15% fetal bovine serum and 2 mM L-Glutamine. RPE-1 cells were cultured in Dulbecco's Modified Eagle Medium F-12 Nutrient Mixture (DMEM/F-12) (1:1) (1X) + GlutaMAX. The medium was supplemented with 10% tetracycline-free fetal bovine serum and 1% penicillin/streptomycin. A final concentration of 5 µg/ml puromycin was added to ensure selection of the cells containing the spCas9 and NuMA sgRNA. 1 µg/ml Doxycycline was added to induce NuMA knock out. SH-SY5Y cells were cultured in DMEM/F12 (1:1) (1X) + GlutaMAX supplemented with 1X non-essential amino acids, 100 U/ml penicillin and 100 µg/ml streptomycin and 10% fetal bovine serum. Differentiation was performed by culturing in 2.5% fetal bovine serum supplemented with 20µM retinoic acid for 6 days. All cells tested negative for mycoplasma and grown at 37°C in a humidified atmosphere containing 5% carbon dioxide.

Differentiation of induced pluripotent stem cells into motor neurons

Induced pluripotent stem cells (iPSCs) are plated onto a 6-well matrigel-coated plates to reach 100% confluency 24 hours prior to neuralisation in iPSC culture medium (mTeSR Plus media supplemented with 10 µM Rock inhibitor, Y27632). The neuralisation process is initiated after a 24 hour incubation. The media is discarded and replaced with warm mTeSR Plus media for 2 hours. This is then followed by daily media change with warm Day 1-6 medium (Basal Media containing 50% KnockOut DMEM/F-12, 50% Neurobasal Media, 0.5x N2 Supplement, 0.5x B27 Supplement, 1X GlutaMAX, 1% Penicillin/Streptomycin supplemented with 2 µM DMH1, 2 µM SB431542, 3 µM CHIR and SMAD inhibitors) every 24 hours until Day 6. On Day 6, the media is discarded and daily media changes are done using the Day 7-12 medium

(Basal Media supplemented with 1 µM CHIR, 2 µM DMH1, 2 µM SB431542, 0.1 µM All-Trans Retinoic Acid and 0.5 µM Purmorphamine) until Day 12. The cells are then treated with 10 µM Y27632 and 2 µl ROCK inhibitor /well for 1 hour, rinsed with warm HBSS, Modified (without Ca++ and Mg++) and then lifted with 1 ml Accutase after incubation for 5-7 minutes at 37°C. The cells are then centrifuged and the pellet is resuspended in 1 ml warm Day 7-12 medium supplemented with 10 µM ROCK inhibitor and the cells are plated at the required density in 6-well Matrigel-coated plates containing Day 7-12 medium. Following a 24 hour incubation, the medium is replaced daily with Day 13-18 medium (Basal Media supplemented with 0.1 µM All-Trans Retinoic Acid and 0.5 µM Purmorphamine) until Day 18, with cells being passaged between days 19-21. On Day 19, the media was discarded and daily media change with Day 19-28 medium (Basal media supplemented with 0.5 µM All-Trans Retinoic Acid, 0.1 µM Purmorphamine, 0.1 µM Compound-E, 10 ng/ml BDNF, 10 ng/ml CNTF and 10 ng/ml IGF-1) was conducted until Day 28. On Day 28, the media is discarded and media is changed with Day 28-40 medium (Neurobasal medium supplemented with 1x B27 supplement, 1% Penicillin/Streptomycin, 10 ng/mL BDNF, 10 ng/mL CNTF and 10 ng/mL IGF-1) every 3 days until Day 40 when motor neurons can be observed.

Hydrogen peroxide (H_2O_2) treatment

Cells are washed with PBS, then either left untreated or treated with 10 µM H_2O_2 in cold PBS (unless otherwise stated) for 10 minutes on ice, in the dark. Cells are then washed with cold PBS and allowed to recover for the specified time points or harvested immediately.

Drug Treatments

Cells were treated with 10 µM DRB for 2 hours, 1µM PARG inhibitor (PDD00017273) overnight for the co-IP and for 1 hour for assessing PAR levels, 1 hr with 10µM PARP1 inhibitor olaparib and 10µM PARP3 inhibitor ME0328, 1 hr with 10 µM nocodazole and 1 hr with 20 mM $KBrO_3$.

Gene silencing

siRNA silencing was carried out using DharmaFECT 1 or Metafectene transfection reagent and either a scrambled (siSCR) or a pool of specific siRNA sequences

(**Supplementary Table 1**). 20 nM siRNA and 2 µl DharmaFECT 1 or 50 nM siRNA and 3 µl Metafectene were each incubated in 100 µl Opti-MEM reduced serum media for 5 min before combining and incubating further for 20 min at RT. Transfection mix was then added to cells seeded at $1.5x10^5$ drop wise in complete media. After 48 hr incubation, cells were harvested, and Western immunoblotting was performed to check the efficiency of knockdown. For ***PARP3*** knockdown, gRNA was designed against exon-1 of PARP3 gene (**Supplementary Table 1**) and cloned into pSpCas9(BB)-2A-GFP_PX458 plasmid. Cells were transfected with the gRNA containing plasmid, and single cells were isolated and allowed to form colonies. Western immunoblotting was carried out to screen for colonies with the least expression of PARP3. To ***transduce iPSC-derived motor neurons*** with control (Cat. No.: sc-108080, Santa Cruz Biotechnology) or NuMA shRNA (Cat. No.: sc043978-V, Santa Cruz Biotechnology) lentiviral particles, complete medium containing 5 µg/ml Polybrene (Cat. No.: sc-134220, Santa Cruz Biotechnology) was added, followed by the addition of the lentiviral particles at a MOI of 5 to the cells (100 µl for 24-well plates and 500 µl for a 6-well plate). The media was changed after 24 hours and the cells were harvested after 48 hours.

Western immunoblotting

Cells were lysed in NP-40 buffer (50 mM Tris HCl pH 8.0, 150 mM NaCl, 1% NP-40) supplemented with 1x cOmplete EDTA-free Protease inhibitor Cocktail (Sigma Aldrich), phosphatase inhibitor, PhosSTOP (Sigma Aldrich), 1 mM DTT and 10 units/ml of Base Muncher Endonuclease (Expedeon) for 30 min on ice. Cellular debris was removed by centrifugation at 13,200 rpm for 20 min at 4°C. Bradford Assay was used to determine total protein concentration. Equal amounts of protein were resolved on 4-15% pre-cast gradient SDS–PAGE (Bio-Rad) gels before transfer to nitrocellulose membrane (Bio-Rad). Membranes were blocked with 5% skimmed milk powder reconstituted in TBS-Tween-20 (TBS-T) for 1 hr at RT before incubating with primary antibodies overnight at 4°C. Membranes were washed three times with TBS-T before incubation with HRP-conjugated secondary antibodies for 1 hr at RT. Membranes were developed using ECL reagent (Bio-Rad) and visualised using a Chemi Doc MP system (Bio-Rad). A list of antibodies used in this study are described in **Supplementary Table 2**.

Alkaline Comet Assays

siRNA transfected MRC5 and SH-SY5Y cells were treated with either 10 or 20 μM H_2O_2 (as described in figure legends) while iPSC-derived motor neurons were treated with 20 μM H_2O_2 in pre-chilled PBS, on ice, for 10 min. H_2O_2 treated cells were left to recover in complete media for varying periods of time at 37°C. A thin layer of 0.6% agarose was laid onto frosted slides. Cells were resuspended in ice cold PBS before being mixed with an equal volume of 1.2% low gelling temperature agarose, maintained at 42°C. Slides were then placed at 4°C to set. Cells were lysed in a pre-chilled lysis buffer (2.5 M NaCl, 10 mM Tris HCl, 100 mM EDTA pH 8.0, 1% Triton X-100, 1% DMSO; pH 10) for 1 hr at 4°C, before submerging in pre-chilled alkaline electrophoresis buffer (50 mM NaOH, 1 mM EDTA, 1% DMSO) for 45 min at 4°C. Electrophoresis was performed at 12 V for 25 min in the dark at 4°C, followed by the addition of 400 mM Tris HCl pH 7 to neutralise. DNA was stained with SYBR Green (1:10000 in PBS) before measuring the average tail moments using Comet Assay IV software (Perceptive Instruments, UK).

Clonogenic Survival Assay

MRC5 cells were plated at varying densities on 10 cm dishes in triplicate. Cells were left to adhere for 24 hrs, before being treated with the specified doses of H_2O_2 in ice cold PBS for 10 min on ice, or overnight with Olaparib. Plates were washed once with PBS before replenishing with fresh media. Colonies were left to form for 7 to 10 days at 37ºC. Colonies were fixed with 80% ethanol, left to air-dry and stained with 1% Methylene Blue for 1 hr before washing with distilled water. Colonies were counted on a colony counter (Stuart) and adjusted for plating density, before being normalised against control samples.

Measurement of nascent RNA transcription:

MRC5 and SH-SY5Y cells were grown on coverslips and transfected with either siSCR or siNuMA for 72 hrs. iPSC-derived motor neurons were transduced with shSCR or shNuMA lentiviral particles for 48 hrs. The Click-iT® RNA Alaxa Flour 488 or 594 Imaging Kit (Invitrogen) was employed according to manufacturer's instructions to quantify nascent RNA transcription. Briefly, cells were serum starved for 48 hrs before

H_2O_2 treatment and labelled either immediately or after a 30 min recovery for 60 min in serum-containing media supplemented with 10 µM 5-ethynyl-2'-deoxyuridine (EdU) to label newly synthesized RNA. Cells were washed with PBS, fixed with 3.7% formaldehyde for 15 min, and permeabilized with 0.5% Triton X-100 for 15 min at RT. Cells were then incubated with Click-iT® reaction cocktail containing Click-iT® additive and Alexa Flour azide for 30 min. Following washing with PBS, cells were mounted on glass slides using a DAPI-containing antifade media (Vectashield, Vector Laboratories). Cells were alternatively incubated with 1:1000 DAPI in PBS for 15 minutes, rinsed with PBS then mounted on a glass slide with Immu-Mount. RNA labelled with EdU was then subjected to immuofluorescence analyses using Leica DM5000B automated upright microscope or the Nikon Widefield Live-Cell System. Average fluorescence signal was quantified from 150 cells using ImageJ software.

Nascent 4sU RNA sequencing

Cells were treated with 700 µM 4-thiouridine and RNA isolated using TRIzol. 75-100 µg RNA was biotinylated in a volume of 150 µl containing 10 mM HEPES pH7.5, 1 mM EDTA pH 8, 5 µg MTSEA Biotin-XX (Iris Biotech, dissolved in DMF). After incubation in the dark for 60 min, biotinylated RNA was twice chloroform extracted, phenol chloroform extracted and then ethanol precipitated. µMACS beads (Miltenyi Biotech) were used for selecting the biotinylated RNA. 50 µl of µMACS beads was washed with 1x wash buffer (10 mM Tris HCl pH 7.4, 50 mM NaCl, 1 mM EDTA) and 2 µl yeast tRNA (10 mg/ml) for 20 min at RT. Columns were washed with nucleic acid equilibration buffer (provided in kit with beads) and thrice with 1x wash buffer. Beads were applied to the column and washed 5 times with 1x wash buffer. Beads were eluted off the column by running 100 µl 1x wash buffer through, twice. 200 µl bead suspension was then incubated with labelled RNA (re-suspended in 1x wash buffer) and incubated at RT for 20 min. Bead-bound RNA suspension was applied to pre-washed column. Columns were washed thrice with wash buffer 1 (10 mM Tris HCl pH 7.4, 6 M Urea, 10 mM EDTA) warmed to 60°C, and then thrice with wash buffer 2 (10 mM Tris HCl pH 7.4, 1 M NaCl, 10 mM EDTA) warmed to 60°C. Labelled RNA was eluted by adding 400 µl 0.1 M DTT in 4x100 µl aliquots. RNA was precipitated in ethanol and subjected to library preparation and RNA sequencing.

Quantitative PCR (qPCR)

$1.5x10^5$ MRC5 cells were seeded in a 6-well plate and transfected the following day with 50 nM scrambled (siScr) or pooled siRNA against NuMA (siNuMA) using Metafectene Pro (Biontex). Cells were incubated at 37°C for 24 hrs and a second transfection hit was conducted, however, serum-free media was used instead, to serum starve the cells for 48 hrs. Cells were treated with 10 µM H_2O_2 in cold PBS, on ice in the dark for 10 min then washed with PBS and recovered in serum-containing media for 90 min. Cells were lysed using RLT buffer from the RNeasy Mini Kit (Qiagen), homogenized using QIAshredder (Qiagen) and then the total RNA was extracted using the RNeasy Mini Kit and eluted in 30 µl RNase-free water. 400 ng of the RNA was used for cDNA synthesis using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Pooled cDNA was used to construct a standard curve using 5-point 10-fold serial dilution. The cDNA samples were diluted 10-fold and SensiMIX SYBR No-ROX Mix (Bioline) was used. To calculate the fold change, relative quantification using the standard curve was conducted and normalized over the scrambled-transfected untreated samples. The primer sequences used are listed in **Supplementary Table 3**.

Whole cell lysate and cell fractionation

Cells were lysed for 10 min in hypotonic buffer (20 mM HEPES pH 8.0, 10 mM KCl, 1 mM $MgCl_2$, 20% glycerol, 0.1% TritonX-100) and spun for 4 min at 6400 rpm in a pre-cooled microcentrifuge. The supernatant was collected as cytoplasmic fraction. The nuclear pellet was washed once with hypotonic buffer and was left to lyse in 50 µl hypertonic buffer (20 mM HEPES pH 8.0, 1 mM EDTA, 20% glycerol, 0.1% TritonX-100, 400 mM NaCl) for 20 min on ice, with periodic vortexing. Samples were centrifuged at 13,200 rpm for 5 min and supernatant was collected as soluble nuclear fraction. The insoluble nuclear pellet was washed once with hypertonic buffer then extracted in 50 µL of insoluble buffer (20 mM Tris HCl pH 8.0, 150 mM NaCl, 1% SDS, 1% NP-40, 10 mM iodoacetate) for 50 min at 4°C on a ThermoMixer®, followed by a 15 min incubation at 25°C with Base Muncher (Expedeon). Tubes were centrifuged at 13,200 rpm for 5 min and supernatant was collected as insoluble nuclear fraction. All buffers were supplemented with cOmplete EDTA-free Protease inhibitor

Cocktail and phosphatase inhibitor, PhosSTOP. For PAR co-immunoprecipitation, the buffers additionally contained 1 µM PAR glycohydrolase inhibitor, ADP-HPD.

Co-Immunoprecipitation

Cells were transfected with polyethylenimine (PEI) (2µl per µg of DNA) or GeneJuice (3µl per µg of DNA). After 48-hrs incubation, cells were harvested in an appropriate volume of NP-40 (whole cell extract) or fractionation buffers. A list of plasmids used in this study are described in **Supplementary Table 4**. ***For myc-IPs*** Myc antibody or mouse IgG were immobilised on 30 µl of Protein G Dynabeads (Thermo Fisher) for 1 hr at RT. Lysates were added to immobilised beads on a rotator at 4°C for 2 hrs. Beads were washed thrice with NP-40 buffer before boiling in SDS loading buffer. ***For GFP-IPs***, lysates were directly added to 25 µl of GFP-Trap beads (Chromotek), incubated for 2 hrs at 4°C, washed thrice with GFP wash buffer (10 mM Tris HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA) before boiling in SDS loading buffer. ***For PAR IPs***, RPE-1 NuMA KO cells were seeded in 15 cm plates and allowed to grow until 80% confluency. Doxycycline was added to induce NuMA KO as previously described. The cells were then treated with 10 µM H_2O_2 in cold PBS for 10 min in the dark then washed with PBS. Cells were lysed and fractionated as described above. Anti-pADPr or anti-mouse IgG were immobilized on 30 µl Protein G Dynabeads and incubated for 1 hr at RT. The beads were then washed twice in 200 µl PBS-T and once in 200 µl NP-40 buffer. The lysates were diluted 4-folds with NP-40 buffer and incubated with the antibody-conjugated beads for 2 hrs at 4°C at 20 rpm. They were then washed twice in NP-40 buffer and twice in NP-40 wash buffer (50 mM Tris HCl pH 8.0, 150 mM NaCl, 0.2% NP-40) before being eluted in 1x SDS loading buffer and boiled at 95°C for 5 min with vortexing. All immunoprecipitated eluates and inputs were then run on a 4-15% Mini-Protean TGX Precast Protein Gels (BioRad) then analysed by immunoblotting.

Sequential Immunoprecipitation

RPE-1 NuMA wild-type and deficient cells were seeded in 15 cm plates and allowed to grow until 80-90% confluency. Doxycycline was added to induce NuMA KO as previously described. The cells were then treated with 10 µM H_2O_2 in cold PBS for 10 minutes in the dark, on ice and then washed with PBS. Cells were lysed and fractionated

as described above. 4.5 µg of p-Ser5 (Pol2) or rabbit IgG antibodies were immobilized on 30 µl Protein A Dynabeads for 1 hr at RT. 400 µg of the lysates (soluble nuclear fraction) were diluted 6.67-fold with NP-40 buffer (50 mM Tris HCl pH 8.0, 150 mM NaCl, 1% NP-40) and incubated with the washed antibody-conjugated beads for 2 hours at 4°C. The complexes were washed twice in NP-40 buffer (50 mM Tris HCl pH 8.0, 150 mM NaCl, 1% NP-40) and once in NP-40 wash buffer (50 mM Tris HCl pH 8.0, 150 mM NaCl, 0.2% NP-40) and then eluted in 50 µl 1% SDS in 50 mM Tris-HCl pH 7.4, at 100°C for 10 minutes with vortexing. 10% of the eluate was kept aside for immunoblotting and the rest was diluted 10-fold in IP dilution buffer (16.7 mM Tris-HCl pH 8, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100) and incubated with 4 µg of PAR or mouse IgG antibodies overnight at 4°C. The following day, 30 µl of washed Protein G Dynabeads are added and incubated for 2 hours at 4°C. The complexes are then washed 3 times in NP-40 wash buffer (50 mM Tris HCl pH 8.0, 150 mM NaCl, 0.2% NP-40) and eluted in 50 µl 1X SDS protein loading buffer. All the lysis and dilution buffers were supplemented with cOmplete EDTA-free Protease Inhibitor Cocktail, phosphatase inhibitor, PhosSTOP and 1 µM PARG inhibitor, ADP-HPD.

Cloning

NuMA truncation mutants were created using Gibson Assembly® Cloning Kit (NEB) following the manufacturer's protocol. The template used for PCRs was pEGFP-C1-NuMA (Addgene) which codes for short isoform of NuMA. Long isoform of NuMA was designed by cloning a synthetic oligonucleotide encompassing the missing 14 amino acids into the pEGFP-C1-NuMA plasmid by Gibson cloning. siRNA targeting resistant NuMA was made by silent mutations within the 1701-1707 stretch of the protein sequence against which the siRNA was designed previously. Truncation mutations were created by PCR of respective fragments of the pEGFP-C1-NuMA targeting resistant plasmid and assembly using the Gibson cloning. TDP1 ORF from pCI-myc-neo-TDP1 plasmid was cloned into a pET16b backbone using restriction cloning using NdeI (NEB) and BamHI (NEB) to generate pET16b-TDP1 construct. PARP1 ORF from pCMV-Sport6-PARP1 (I.M.A.G.E, Source Bioscience) was cloned into the pET16b backbone using Gibson cloning to generate pET16b-PARP1 construct. GFP-PARP1 and GFP-PARP3 plasmids were created by Gibson cloning using pEGFP-C1 as backbone and PARP1 and PARP3 ORF (I.M.A.G.E clone, Source Bioscience)

as inserts. For cloning NuMA$^{1700-2115}$ into pET16b vector, primers containing the restriction sites NdeI and BamHI were used to amplify NuMA$^{1700-2115}$ from a pEGFP-C1-NuMA plasmid, using KOD hot start DNA polymerase protocol. 1 µg of amplified NuMA and pET16b vector were digested with NdeI and BamHI enzymes (NEB), followed by PCR purification (QIAquick PCR purification kit - Qiagen) and ligation using a T4 DNA ligase (NEB) in a 1:3 molar ratio (vector:insert) for 1 h at 25°C. 5 µL of the reaction was used to transform competent DH5α cells (NEB) according to manufacturer's protocol. DNA was extracted and purified from cells using a QIAprep spin miniprep kit (Qiagen) and the sequences were confirmed by Sanger sequencing (Eurofins GATC LightRun).

Recombinant protein expression

pET16b-NuMA$^{1700-2115}$, pET16b-TDP1 or pET16b-PARP1 were transformed separately into BL21 (DE3) *E.coli* strain by heat shock method on ice. After the outgrowth period in SOC media, the transformed cells were plated on ampicillin plates overnight at 37°C. For each plasmid, a single colony was picked and grown in a 5 ml starting LB-culture for 4 hrs at 37°C. Once the culture became slightly turbid, it was diluted 200 times in fresh LB-culture with 100 µg/ml Ampicillin and allowed to grow till OD=0.6. The cultures were then induced with 0.5 mM IPTG either overnight at 20°C (NuMA) or for 4 h at 30°C (TDP1 and PARP1). After induction, the cells were centrifuged at 5000 rpm for 10 min at 4°C. Cell pellets were frozen at -20°C for downstream protein harvest and purification.

Protein A/G-fused MNase construct was transformed in BL21 cells (Cat. No.: C2530H, NEB),cultured in 100 µg/ml Kanamycin and induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside following standard protein expression and purification protocols.

Recombinant protein purification

For NuMA$^{1700-2115}$, cells harvested from 1.6 L cultures were lysed in buffer U (8 M urea, 0.5 M NaCl, 50 mM Tris HCl pH 8.0), sonicated thrice for 20 sec each and centrifuged at 72000 x *g* for 30 min. After centrifugation the supernatant was applied on a 5 mL HisTrap HP column (GE Healthcare) on an AKTA purifier and ran at a flow rate of 5 mL/min. Proteins were eluted at a 50 mL gradient of 0-0.25 M imidazole in buffer U. Protein concentration of the eluted fractions was quantified by Bradford assay

and assessed by SDS-PAGE. The fractions were combined and concentrated to 0.5 mL using Vivaspin MWCO 50,000 columns (Sartorius). Finally, the protein was refolded by dialysis against buffer A (0.5 M NaCl, 50 mM Tris HCl pH 8.0) for 2.5 hrs at RT and stored at -80°C. ***For TDP1 and PARP1***, cells harvested from 1.6 L cultures were lysed in buffer A (0.5 M NaCl, 50 mM Tris HCl pH 8.0), sonicated thrice for 20 sec each and centrifuged at 72000 x *g* for 10 min. Cell free extract (CFE) was applied on a 5 mL HisTrap HP column (GE Healthcare) on an AKTA purifier and chromatography was ran a flow rate of 5 mL/min. Elution from the column occurred at a 50 mL gradient of 0-0.3 M imidazole in buffer A. The eluates were then quantified by Bradford assay and assessed by SDS-PAGE. The fractions containing the expressed fragments were pulled and concentrated using Vivaspin MWCO 50,000 columns. Concentrated samples were then subjected to gel filtration (GF) using an AKTA purifier system. PARP1 purification by GF was conducted on a 1.6x60cm Hi-Load Superdex200 column (GE Healthcare), while TDP1 was purified using a 1x30cm Superdex200 Increase column (GE Healthcare) (TDP1), both equilibrated in buffer A, at a flow rate of 1.5 mL/min (PARP1) or 0.5 mL/min (TDP1). Selected fractions were concentrated and stored at -80°C. For ***Protein A/G-fused MNase***, cells were lysed in lysis buffer (50 mM Tris HCl pH 7.5, 500 mM NaCl and EDTA-free protease inhibitor), sonicated with a Branson Sonifier blunt-end adapter at output level 4, with 45 seconds intervals for 5-10 rounds or until the turbidity was reduced. The lysate was cleared by high-speed centrifugation and purified by Ni-NTA magnetic beads (Cat. No.: 88831, Thermo Scientific), washed thrice in three bed volumes of wash buffer (Lysis buffer supplemented with 0.05% Tween and 5 mM imidazole). Elution was performed using Elution Buffer (Lysis buffer supplemented with 500 mM imidazole). The eluate was dialyzed twice against a 750 ml volume of 10 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM PMSF to remove the imidazole. Glycerol was then added to a final concentration of 50%, and aliquots were stored at -80°C.

In vitro NuMA binding assays

50 µl of Ni-NTA magnetic agarose beads (Qiagen) were washed twice in H100 buffer (20 mM HEPES pH 7.5, 100 mM KCl, 2 mM DTT). 10 µg of recombinant His-tagged TDP1 (pET16b-TDP1) or BSA was immobilised on washed beads in 200 µl H100 buffer for 1 hr at 4°C. Beads were spun briefly and washed twice with 200 µl H100 buffer. NuMA was *in vitro* transcribed and translated from a T7 promoter containing

plasmid (pcDNA3-NuMA-full length) in rabbit reticulocyte lysate using a TnT® Quick Coupled Transcription/Translation kit (Promega) supplemented with 1 mM [^{35}S] Methionine (PerkinElmer), using the manufacturer's protocol. 10 µl of TnT reaction supplemented with either 0.25 µM Poly(ADP-ribose) (PAR) polymer (Trevigen) or PAR buffer (10mM Tris HCl pH 8.0, 1mM EDTA) was added to washed beads and incubated for 1 hr at 30°C in a water bath. Beads were washed twice in H100 buffer and boiled with SDS loading buffer and then run on an 8% polyacrylamide gel. The gel was dried and [^{35}S] signal was developed on an image plate for 24 hrs. The image plate was scanned using Typhoon FLA-7000 laser scanner (GE Healthcare Life Sciences). **For the *in vitro* PARylation reactions**, 30 µL of Ni-NTA magnetic beads (Thermo Scientific) were washed twice in H100 buffer and conjugated with equimolar quantities (~145 pmol) of either His-eIF4a or His-TDP1 for 1 hr at 4°C. The beads were washed with H100 buffer and further incubated with 10 µL of *in vitro* synthesised ^{35}S-NuMA in H100 buffer for 1 hr at 4°C. Non-PARylated samples were then incubated with 200 µL of 50 mM Tris HCl pH 7.5 for 25 min at room temperature (RT) with rotation. For the PARylated samples, 1 mL of *in vitro* PARylation reaction (50 mM Tris HCl pH 7.5, 1 mM DTT, 10 mM $MgCl_2$, 50 µg NAD^+, 1 µg DNA, 1 µg PARP1) was prepared. The samples were then incubated with 200 µL of *in vitro* PARylation reaction for 25 min at RT with rotation. Ni-NTA beads were washed 4 times in H100 buffer and eluted in 1x SDS loading buffer at 95°C for 5 min with periodic vortexing. The eluates were separated by SDS-PAGE and blotted into a nitrocellulose membrane. The membrane was dried at 37°C for 2 min, exposed to an imaging plate overnight and imaged using Typhoon FLA 7000. To check PARylation, nitrocellulose membranes were rehydrated in TBS-T and blocked in 5% milk for 30 min, followed by incubation with anti-PAR antibody overnight at 4°C.

Immunofluorescence

Cells were grown on coverslips, treated with 10 µM H_2O_2 for 10 min on ice. Coverslips were rinsed in PBS and fixed with methanol–acetone (1:1, v/v) for 10 min at 4°C. Fixed cells were rinsed in PBS and permeabilized for 5 min at 4°C in 20 mM HEPES pH 7.4, 50 mM NaCl, 3 mM $MgCl_2$, 300 mM sucrose, 0.5% Triton X-100. Permeabilized cells were rinsed in PBS, blocked in 2% bovine serum albumin and incubated for 30-60 min at RT with the primary antibody of choice in PBS supplemented with 2% bovine serum albumin. After rinsing in PBS, coverslips were incubated with the relevant Alexa Fluor

secondary antibodies in PBS supplemented with 2% bovine serum albumin for 30-60 min at RT. Nuclei were counterstained with 0.000025% 4′,6′-diamidino-2-phenylindole (DAPI). For ***H2AK119mUb*** and ***NuMA***, cells were fixed with 4% paraformaldehyde for 15 minutes at RT. The fixed cells were then rinsed in PBS and permeabilized for 15 minutes at RT in 0.2% Triton X-100 in PBS. Cells were then rinsed with PBS and blocked in 3% BSA in PBS for 1 hour at RT. All other steps were the same. For ***MAP2***, cells were fixed with 4% paraformaldehyde for 10 minutes, rinsed with PBS, permeabilized with 0.3% Triton X-100 for 5 minutes, rinsed with PBS and blocked in 5% donkey serum. Primary and secondary antibodies were diluted in 1% donkey serum and all other steps were the same. For ***H2BK120mUb***, cells were fixed in 100% methanol for 10 minutes, rinsed with PBS then permeabilized in 0.2% Triton X-100 in PBS for 10 minutes. All other steps were the same. Cells were photographed with a Leica DM5000B automated upright microscope or a Nikon Widefield Live-cell system and analyzed with FW4000 (Leica) imaging software or ImageJ.

Proximity Ligation Assay (PLA)

PLA was carried out using the Duolink® PLA Green starter kit (Merck). Cells grown on chamber slides and soluble proteins were pre-extracted by incubating cells twice for 3 min at RT in CSK Buffer (10 mM Pipes pH 7.0, 100 mM NaCl, 300 mM Sucrose, 3 mM $MgCl_2$, 0.7% Triton X-100) and then were fixed with 4% PFA in PBS for 10 min at RT and permeabilised with ice-cold methanol for 10 min at -20°C. Cells were washed with PBS and blocked with Duolink® block solution for 1 hr at RT. Primary antibodies were diluted in Duolink® antibody dilution buffer. Cells were incubated with antibodies for 1 hr at RT and washed thrice in 5% BSA in PBS for 10 min each. 20 µL of secondary antibody mixure was added to each sample, incubated at 37°C for 1 hr and washed twice for 5 min with propitiatory 1x Wash Buffer A. 20 µL of Ligation Mix prepared according to manufacturer's protocol was added to each sample, and further incubated at 37°C for 30 min. Cells were washed 1x with Wash Buffer A. 20 µL of amplification mixture was then added to each sample and incubated at 37°C for 100 min in the dark. Cells were washed twice for 10 min in propitiatory 1x Wash Buffer B and once with 0.01x Wash Buffer B for 1 min. The chamber was removed from the slide, and cells mounted with a coverslip using a minimal volume of Duolink® In Situ

Mounting Medium with DAPI. The slide was imaged with an automated upright fluorescent microscope (Leica DM5000 B).

Laser micro-irradiation

Exponentially growing MRC5 cells were plated onto 35-mm glass-bottom dishes (MatTek) and transfected with pMCEGFP–TDP1 using Lipofectamine 2000 (Invitrogen) and siRNA against NuMA using DharmaFECT 1 (Thermo Fisher Scientific) sequentially, according to the manufacturer's protocol. 48 hrs following transfection, cells were incubated with 10 μg/ml Hoechst 33285 (Sigma Aldrich) for 1 hr at 37 °C before irradiation. Cells expressing similar total GFP signal were irradiated with a 405-nm ultraviolet A laser channelled through a ×63 objective using a ZEISS LSM 880 with Airyscan confocal microscope (Zeiss). Ultraviolet A was focused to an area of approximately 20 μm × 0.1 μm, and images were captured at 2.5-sec intervals following laser damage and quantified by ImageJ and ZEN (Zeiss) software.

Electrophoretic Mobility Shift Assay (EMSA)

1 μM 3'Cy5-labelled oligonucleotides were annealed in 2X Annealing Buffer (10 mM Tris HCl pH 7.5, 50 mM NaCl, 1 mM EDTA) with non-labelled oligonucleotides to form double-stranded DNA and single-strand break substrates. The sequence of the oligonucleotides used are listed in **Supplementary Table 5**. 125 and 250 nM recombinant C-terminal $NuMA^{GD}$ was incubated for 3 hrs at 37°C with 50 nM of the Cy5-labelled oligonucleotides in Binding Buffer (40 mM Tris HCl pH 8.0, 100 nM NaCl, 50 nM KCl, 1 mM MgCl2, 5 mM DTT, 0.1 mg/ml BSA, 0.1% NP-40) in a total volume of 10μl. Samples were mixed with 10X Loading Buffer (40% glycerol) and loaded on an 8% native polyacrylamide gel with 1x TBE as the running buffer. The gel was degassed for 15 minutes at RT, then left to polymerise at RT for 2 hours. It was then pre-run at 60 V for 60 min, then samples were loaded, and it was run at 300 V for 15 min at 4°C, followed by 100V for 60 min. The gel was then imaged using the ChemiDoc (BioRad) and bands were analyzed using ImageJ.

OGG1-enriched AP-Seq (OGAP-Seq)

7-10 μg genomic DNA from cells was extracted using the Qiagen Blood and Tissue Kit and eluted in 100 μL mQ H_2O. DNA was digested with 1:1000 dilution of 8-oxoguanine DNA glycosylase (OGG1) or Fpg (NEB) in 1x NEB-buffer 2, 1x bovine serum albumin

and incubated at 37°C for 1 hr. DNA was precipitated using cold 100% ethanol and reconstituted in 90 µL PBS. DNA was labelled with 5 mM biotin labelled aldehyde-reactive probe (ARP). Labelled DNA was transferred to 1.5ml tube and precipitated with ice-cold ethanol (100%), washed with 70% ethanol, and reconstituted in 130 µL TE buffer pH 8. DNA was subsequently sheared to an average peak size of 300 bp on a Bioruptor Pico (Diagenode). 30 µL sheared DNA was kept aside as inputs. 100 µL MyOne Dynabeads (Invitrogen) was washed twice with 1M NaCl in TE buffer, reconstituted in 100 µL 2M NaCl in TE buffer and added to 100 µL of labelled DNA from above. Samples were rotated at RT for 10 hrs. DNA was eluted twice from the beads using 95% formamide and 10 mM EDTA for 10 min at 65°C in a total 100 µL volume. MinElute Clean Up kit (Qiagen) was used for DNA extraction and DNA was eluted in 30 µL TE (3 x 10 µL elution). DNA was repaired using the PreCR Repair Mix (NEB) as per the manufacturer's protocol. Repaired DNA was extracted using MinElute Clean Up Kit and eluted in 13 µL mQ H_2O and subjected to library preparation and sequencing. Library prep was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB).

Chromatin Immunoprecipitation

Cells were either untreated or treated with 10 µM H_2O_2 for 10 min and left to recover for 90 min. Cells were then washed with PBS and crosslinked in a final concentration of 1% paraformaldehyde in 10-20 ml PBS for 10 min at RT. The crosslinking was quenched with a final concentration of 0.125 M glycine for 5 min at RT. Cells were then washed twice in cold PBS then collected by scraping. Cell pellets were resuspended in 5 pellet volumes of ChIP Lysis Buffer 1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100) and incubated for 5 min in 4°C at 20 rpm rotation. Cells were then spun down at 3000 x *g* for 5 min at 4°C. The pellet (nuclei) was resuspended in 5 pellet volumes of ChIP Buffer 2 (10 mM Tris HCl pH 7.4, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) and rotated at 20 rpm for 10 min at RT. The lysate was then spun down at 1500 x *g* for 5 min at 4°C and the nuclear pellet was then lysed by resuspending in a suitable volume of ChIP Lysis Buffer 3 (10 mM Tris HCl pH 7.4, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Sodium deoxycholate, 0.5% Sodium lauroylsarcosine). The lysates were then sonicated using Bioruptor Pico to yield DNA fragments of the size 100-300 bp and cleared by spinning down at 20000 x *g*, at 4°C for 15 min. 1 or 10% of the lysate used

for immunoprecipitation was reserved as an input. Lysates containing an equal quantity of protein were incubated with either anti-NuMA/Pan Pol2/Flag antibody or the corresponding Rabbit/Mouse IgG overnight at 4°C. 30 µl Protein A or Protein G Dynabeads were added to each sample and incubated for 2 hrs at 4°C. The beads were washed thrice in 500 µl RIPA Wash Buffer and once in 1 ml ChIP final wash buffer then eluted in 200 µl ChIP Elution Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1% SDS) at 65°C for 30 min for the NuMA ChIP in MRC5 cells. The beads were washed once in Low Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0 and 150 mM NaCl), once in High Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 500 mM NaCl) and once in LiCl Wash Buffer (0.25M LiCl, 1% NP-40, 1% Sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0) before being eluted twice in 100 µl ChIP Elution Buffer at 65°C for 30 min at 1000 rpm for the Pan Pol2 and Flag-TDP1 ChIP. The eluted DNA and the input were reverse crosslinked by incubating with up to 0.2 M NaCl at 65°C for 16 hrs. Treatment with 0.2 mg/ml RNase A was done at 37°C for 30 min at 800 rpm and 0.2 mg/ml Proteinase K at 55°C for 2 hrs at 8 rpm. DNA was purified by phenol chloroform extraction followed by ethanol precipitation and processed for sequencing or qPCR. For ChIP-Seq, library prep was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). For ChIP-qPCR, the eluted DNA (10 µl) was diluted 1:10 then subjected to qPCR using primers listed in **Supplementary Table 3** and the % input was calculated.

Cleavage Under Target & Release Using Nuclease (CUT&RUN)

CUT&RUN was conducted in SH-SY5Y and RPE-1 cells following the protocol described by[51]. 500,000 cells were used for each experiment and antibodies were left to incubate overnight at 4°C. For SH-SY5Y cells, 0.75 µg of the antibodies were used, and 3 biological repeats were conducted. For RPE-1 cells, 1 µg of the antibodies was used, and 2 biological repeats were conducted. Library prep was performed following the protocol described by[52] using the NEBNext Ultra II DNA Library Prep Kit for Illumina (Cat. No.: E7103, NEB).

Mass Spectrometry Analyses: On-bead reduction, alkylation and digestion

Triplicate GFP-trap bead samples were resuspended in 100 µL of 100 mM Pierce™

AB, pH 8. Reduction was performed by the addition of 1 µL of 50 mM TCEP and incubation for 10 min at 70°C. After cooling to RT, 2 µL of 50 mM iodoacetamide were added and the samples were incubated in the dark with shaking for 30 min at RT. Lyophilized trypsin was dissolved in 100 mM AB to a final concentration of 0.1 µg/µL and 20 µL of enzyme solution were added to the samples. Digestion was performed at 37°C for 6h. The digestion was stopped by lowering the pH to 3.0 with 10% FA. Digested peptides were desalted with C18 spin columns (P/N 89870, Pierce, Thermo Scientific) according to manufacturer's instructions except that 0.5% tri-flour acetic acid was replaced by 0.5% FA. Digested and cleaned samples were dried in a vacuum concentrator and resuspended in 40 µl of 0.5% FA before LC-MS/MS analysis.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS)

Peptides were separated in a Dionex Ultimate 3000 uHPLC (Thermo Scientific) with an Acclaim™ PepMap™ 100 C18 trap column (3 µm particle size, 75 µm X 150 mm) and an EASY-Spray™ C18 column (2 µm particle size, 50 µm X 150 mm) using an 80 min gradient at 0.25 µl/min with 0.1% FA in water (mobile phase A) and 0.1% FA in 80% ACN (mobile phase B) as follows: 0-5 min, hold at 3% B; 5-60 min, from 3% B to 40% B; 60-65 min, from 40% B to 90% B; 65-70 min, hold at 90% B; 70-71 min, from 90% B to 3% B; 71-80 min, re-equilibrate at 3% B.

MS/MS analysis was performed on a LTQ Orbitrap Elite (Thermo Scientific) hybrid ion trap-orbitrap mass spectrometer equipped with an EASY-Spray^{TM} Ion Source. MS survey scans in positive ion mode were acquired in the FT-orbitrap analyzer using an m/z window from 375 to 1600, a resolution of 60,000, and an automatic gain control target setting of 1 × 10E6. The 20 most intense precursor ions were selected for the acquisition of MS/MS spectra in the ion trap (Normal Scan Rate) using collision-induced dissociation with normalized collision energy of 35%, activation time of 10 ms, activation Q of 0.25, isolation width of 2 Th, and automatic gain control target value of 1 × 10E4. Ions with charge state 1+ were excluded from precursor selection. Monoisotopic precursor selection was activated. Dynamic exclusion for precursor ions was applied for 45 sec after 1 fragmentation count and a repeat duration of 30 sec.

LC-MS/MS Data Analysis

MS/MS data files were used for protein identification using MaxQuant software[53] ersion 1.5.5.1. Default MaxQuant parameters were used except for the following: protein database was the Human Uniprot proteome (downloaded on the 28th September, 2017) and a fasta file containing the theoretical sequence of GFP-NuMA-L and GFP-NuMA-S; LFQ quantification was selected, unique peptides were used for quantification, minimum ration count was set to 1, calculation of iBAQ values was selected; variable modifications, Oxidation(M); fixed modifications, Carbamydomethylation©. The statistical analyses of MaxQuant LFQ intensities were performed by Perseus[54] as follows. The dataset was filtered to remove proteins with less than three values in at least one group (GFP, LTR-NuMA "NuMA-L" or STR-NuMA "NuMA-S"). LFQ intensities were Log2-transformed and missing values were replaced using a downshifted normal distribution of the total matrix (width 3, downshift 1.8). The quantitative differences between conditions were evaluated after hierarchical clustering of the data and graphical representation in a volcano plot (Student t-test with false discovery rate of 0.05 and artificial within groups variance S0=1). NuMA-L and NuMA-S signature peptide spectra were manually inspected after annotation with pLabel software[55,56], with relative intensity threshold set to 2% and tolerance to 0.5 Da for fragment matching.

8-oxoG Mass Spectrometry

Genomic DNA was extracted from RPE-1 cells using the DNeasy Blood & Tissue Kit (Cat. No: 69506, QIAgen). The RNA was degraded by incubating the extracted DNA with 10 µg RNase A in 2X RNase A buffer (10 mM ammonium bicarbonate pH 7.0, 1 mM $MgCl_2$ and 0.1 mM deferoxamine mesylate) at 37°C for 30 minutes at 600 rpm. The samples were then passed through the Microcon-30kDa Centrifugal Filter Unit with Ultracel-30 membrane (Cat. No.: MRCF0R030, Merck Millipore) to remove free nucleosides and nucleotides and eluted in Water for UHPLC Gradient Grade Analysis (Cat. No.: 11307090, Fisher Chemical).

1 µg of DNA in 100 µL of water was added to 50 µL of hydrolysis solution (100 mM NaCl, 20 mM MgCl2, 20 mM Tris pH 7.9, 1000 U/ml Benzonase, 600 mU/ml Phosphodiesterase I, 80 U/ml Alkaline phosphatase, 36 µg/ml EHNA hydrochloride,

2.7 mM deferoxamine). The mixture was incubated for two hours and then lyophylised by SpeedVac.

The lyophylisate was resuspended in 30 µL of buffer A, spiked with 12.5 nM 8-oxo-deoxyguanosine +3 Da (8OG +3). 25 µL were transferred into an LC-MS vial for analysis.

For the analysis by HPLC– QQQ mass spectrometry, a 1290 Infinity UHPLC was fitted with a Zorbax Eclipse plus C18 column, (1.8 µm, 2.1 mm 150 mm; Agilent) and coupled to a 6495a triple quadrupole mass spectrometer (Agilent Technologies) equipped with a Jetstream ESI-AJS source. The data were acquired in dMRM mode using positive electrospray ionisation (ESI1). 8OG was quantified by mass spectrometry, whereas dC, dG, dT and dA was quantified solely by HPLC-UV (λ=254 nm).

For UV analysis, 2 µL of the sample were injected, whereas 20 uL were injected for MS analysis. The gradient used to elute the nucleosides started with a 5-min isocratic gradient composed with 100% buffer A (10 mM ammonium acetate, pH 6) and 0% buffer B (100% methanol) with a flow rate of 0.400 ml min-1 and was followed by the subsequent steps: 5-8 min, 94.4% A; 8–9 min, 94.4% A; 9–16min 86.3% A; 16–17 min 0% A; 17– 21 min 0% A; 21–24.3 min 100% A; 24.3–25min 100%A. The gradient was followed by a 5 minute post time to re-equilibrate the column.

The AJS ESI settings were as follows: drying gas temperature 230 °C, the drying gas flow 14 lmin-1, nebulizer 20 psi, sheath gas temperature 400 °C, sheath gas flow 11 lmin-1,Vcap 2,000 V and nozzle voltage 0 V. The iFunnel parameters were as follows: high pressure RF 110 V, low pressure RF 80 V. The fragmentor of the QQQ mass spectrometer was set to 380 V and the delta EMV set to +200.

The raw mass spectrometry data was analysed using the MassHunter Quant Software package (Agilent Technologies, version B.08.01). The transitions and retention times used for the characterization of nucleosides and their adducts are summarized in **Supplementary Table 6**. For the identification of compounds, raw mass spectrometry data was processed using the dMRM extraction function in the MassHunter software.

RNA 4sU-Seq Processing and Analyses

Quality of samples was assessed by examining FastQC reports provided by Novogene Plc[57]. The data was mapped to the human reference genome using STAR by 172 Novogene Plc[58]. Differential expression analyses were performed in R using DESeq2. The data was normalised using DESeq and false discovery rate calculation after model fitting was performed using the Benjamini-Hochberg procedure, all of which was done by Novogene Plc[59,60,61]. Less than or equal to an adjusted p-value of 0.05 was taken as the cut off for differential expression. Prior to downstream analyses, unexpressed genes were filtered out using transcripts per million (TPM) data for each gene derived from mapping the data using salmon (index building parameters: --type=-fmd kmer=31, mapping parameters: quant libtype A)[62]. The resulting quant files were analysed in R studio and imported into R studio using the package tximport[63,64]. Any genes that did not have a TPM value of 1 or greater in both replicates of at least one condition were considered to be unexpressed and were discarded. Statistical enrichment for gene categories was calculated using fishers exact test through the R function fisher.test. Additional information on genes, such as length, was found using biomaRt[65].

Identification of gene categories and enhancers

Paused genes analysed in the 4sU-seq analysis were determined using publicly available RNA polymerase II ChIP-seq data from RPE cells (Available under the accession GSE60024), whereas the paused genes used in the NuMA ChIP-seq analysis were determined using publicly available RNA polymerase II ChIP-seq data from MRC-5 cells (Available under the accession GSE55171[66]) as the 4sU-seq was performed in RPE-1 cells whereas the NuMA ChIP-seq was performed in MRC-5 cells. Pausing indices were calculated from RNA polymerase II ChIP-seq by dividing the average number of reads overlapping 100 bases pairs upstream and 300 base pairs downstream of the transcription start site by the average number of reads overlapping 500 to 2000 base pairs downstream of the transcription start site. Genes less than 1000bp and genes with less than 400bp of coverage were ignored. Genes with a pausing index ≥ 2 were considered to be paused and were carried forward to subsequent analyses. Immediate early response genes were collected from Immediate-early and delayed primary response genes are distinct in function and genomic architecture[67]. Gene promoter windows were generated using biomaRt for R to identify the chromosome and transcription start site (TSS) for each gene. The TSS was then extended by +/- 2500 base pairs to approximate the promoter region, as described in the paper listed under

the BLISS data accession[68], and a promoter BED file created. An Intron BED file was downloaded from the UCSC table browser (https://genome.ucsc.edu/cgi-bin/hgTables) (options: clade = Mammal, genome = Human, assembly = Dec.2013 (GRCh38/hg38), group = Genes and Gene Predictions, track = NCBI RefSeq, table = knownGene, region = Genome, output format = BED – browser extensible data, Create one BED record per = Introns plus (0)). Overlaps between the promoter and intron BED files and publicly available data detailing endogenous DSB sites identified using BLISS data (available under the accession GSE93038[68], was performed using bedtools intersect (parameters: -wa) in order to identify fragile promoters/introns[69]. Enhancers in RPE1 cells were defined as H3K27ac peaks that overlapped with H3K4me2 peaks but not with promoter windows (+/- 2.5kb either side of TSS regions), introns or H3K9me3 peaks. Enhancers in SH-SY5Y cells were defined as H3K27ac peaks that overlapped with H3K4me1 peaks but not with promoter windows (+/- 2.5kb either side of TSS regions) or introns. Peak calling for all the histone ChIP-seq was undertaken according to the protocols described under **ChIP-seq processing and analyses** except broad rather than narrow peaks were called. Overlaps between genomic regions were determined using bedtools intersect, using either the parameter -wa to retain, or the paramter -v to exlude overlapping regions, and finally bedtools merge was used to merge overlapping enhancer regions within the bedfile[69]. Both the RPE1 and SH-SY5Y histone ChIP-seq data are publicly available datasets (available under the accessions GSE113399[70] and GSE80197[71] respectively).

ChIP-Seq Processing and Analyses

The ChIP-seq data was mapped using BWA-mem (parameters: -M -k 25) to the hg38 human reference genome via the mapping pipeline from CGAT pipelines[57,72,73]. Prior to peak calling and the generation of metagene profiles, duplicates were first removed from the BAM files using Picard MarkDuplicates in conjunction with samtools view to remove PCR or optical duplicates (parameters: -F 1024)[74]. Reads with a quality score of less than 30 were also removed using samtools view (parameters: -q 30). Peak calling was performed using MACS2 (narrow peaks, parameters: default aside from -g hs). Fragment pileup and control lambda were output to BEDgraph files using the MACS2[75] peak calling option -B –SPMR. The sample and control lambda (derived from inputs) BEDgraph signal tracks were then compared using MACS2 bdgcomp with the option -m FE in order to generate sample BEDgraph files showing fold change relative to input,

which were then converted into bigwig files. Metagene profiles were generated using bam2geneprofile from CGAT scripts using gene annotations derived from Ensembl 85, with reporter set to genes and no normalisation[76]. Both gene profiles, transcription start site, and transcription termination site profiles were generated. ChIP-seq reads overlapping promoter regions were determined by creating a window 1000 base pairs upstream and downstream of the TSSs of protein coding genes from Ensembl 85 using bedtools slop along with a pseudo contigs file generated from the GTF, converting the resulting BED file to a GTF file using CGAT scripts bed2gff and running featureCounts (parameters: -t exons -g gene) on the ChIP-seq bam files over the promoter GTF[77]. The metagene profile pipeline excluded overlapping genes. This was done by converting the GTF(s) to BED files using gff2bed (parameters: --is-gtf) from CGAT scripts, extending resulting regions by 1250 base pairs upstream and downstream using bedtools slop (parameters: -s), counting the resulting gene overlaps using bedtools merge (parameters: -c -o count), identifying overlapping genes as having an overlap count of at least one and removing these from the original GTF(s) using an exclusionary bedtools intersect (parameters: -v). The genome file passed to bedtools slop was a pseudo-contigs files generated from the provided GTF. In all metagene plots, the trace was averaged across replicates. Metagenes were normalised by the number of mapped reads, determined using samtools (parameters: view -c -F 4), and divided by the input metagene. Pausing ratios were determined, as described above in the identification of gene categories section.

OGAP-Seq Processing and Analyses

The OGAP-Seq data was mapped using BWA-mem (parameters: -M -k 25) to the hg38 human reference as part of the mapping pipeline from CGAT pipelines. Metagenes, fold change bigwig files and OGAP-Seq reads overlapping the promoters of genes were determined as described above for the ChIP-seq analysis.

CUT&RUN Processing and Analyses

The CUT&RUN data was mapped using BWA-mem (parameters: -M -k 25) to the hg38 human reference as part of the mapping pipeline from CGAT pipelines. Metagenes, fold change BEDgraph files and peaks were determined as described in the **ChIP-seq procesisng and analyses** section except the CUT&RUN data was not deduplicated and

the IgG samples were used for normalisation in peak calling and metagene analyses in place of inputs.

Data processing and plots in R

Within R, data manipulation was performed using the dplyr[78] and tidyr[79] packages. Plots were generated using base R[61] and the ggplot2[80] package.

Statistical Analyses

Graphs and statistical analysis for all figures was generated using GraphPad Prism (GraphPad Software). Data is presented as the mean ± s.e.m, with Student t-test or One-way ANOVA used as indicated in the figure legends. The exact p-values are listed in **Supplementary Data 3**.

Data availability

Mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[81]. The identifier number of the dataset is PXD018002. Next generation sequencing data have been deposited in the GEO database under the SuperSeries accession codeGSE147015 at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147015. The SubSeries are: RPE-1 4sU-seq (GSE147012), RPE-1 OGAP-seq (GSE147013), MRC5 NuMA ChIP-seq (GSE147014), SH-SY5Y OGAP-seq (GSE201981), RPE1 Pol2 ChIP-seq (GSE201982), SH-SY5Y NuMA CUT&RUN (GSE201983), RPE1 NuMA and TDP1 CUT&RUN (GSE201984). The existing publicly available datasets that were used are as follows: RNA Pol2 ChIP-seq data from RPE-1 cells (GSE60024, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60024, sample IDs:GSM1463920), RNA Pol2 ChIP-seq data from MRC-5 cells (GSE5517114, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55171, sample IDs: GSM1330727), BLISS data (GSE93038, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93038, filename: GSE93038_Dellino_BLISS_Processed_Data_Tier1_DSBs.txt.gz), RPE1 histone ChIP-seq (GSE113399, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113399, sample IDs: GSM3105085,GSM3105086, GSM3105087, GSM3105091), SH-SY5Y histone ChIP-

Seq (GSE80197, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80197, sample IDs: GSM2120705, GSM2120706, GSM2120708).

Code Availability

The code for the pausing ratio calculator is available at https://github.com/jdparker101/Pausing_Ratio_Notebook

References for Methods

51. CUT&RUN: Targeted in situ genome-wide profiling with high efficiency for low cell numbers. https://www.protocols.io/view/cut-amp-run-targeted-in-situ-genome-wide-profiling-zcpf2vn.

52. Library Prep for CUT&RUN with NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (E7645). https://www.protocols.io/view/library-prep-for-cut-amp-run-with-nebnext-ultra-ii-wvgfe3w?version_warning=no.

53. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372 (2008).

54. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731–740 (2016).

55. Li, D. *et al.* pFind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry. *Bioinformatics* **21**, 3049–3050 (2005).

56. Wang, L. *et al.* pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Commun Mass Sp* **21**, 2985–2991 (2007).

57. Andrews & S. FASTQC A Quality Control tool for High Throughput Sequence Data. *Babraham Institute* (2015).

58. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

59. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

60. Love, I., M., Anders, Huber, S. & & W. Differential analysis of count data - the DESeq2 package. *Genome Biol* **15**, 550 (2014).

61. R Development Core Team. R: A Language and Environment for Statistical Computing. (2015).

62. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).

63. Team, Rs. RStudio: Integrated Development for R. (2015).

64. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000research* **4**, 1521 (2016).

65. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184–1191 (2009).

66. Giannakakis, A. *et al.* Contrasting expression patterns of coding and noncoding parts of the human genome upon oxidative stress. *Sci Rep-uk* **5**, 9737 (2015).

67. Tullai, J. W. *et al.* Immediate-Early and Delayed Primary Response Genes Are Distinct in Function and Genomic Architecture*. *J Biol Chem* **282**, 23981–23995 (2007).

68. Dellino, G. I. *et al.* Release of paused RNA polymerase II at specific loci favors DNA double-strand-break formation and promotes cancer translocations. *Nat. Genet.* **51**, 1011–1023 (2019).

69. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

70. Sun, Z. & Chadwick, B. P. Loss of SETDB1 decompacts the inactive X chromosome in part through reactivation of an enhancer in the IL1RAPL1 gene. *Epigenet Chromatin* **11**, 45 (2018).

71. Henrich, K.-O. *et al.* Integrative Genome-Scale Analysis Identifies Epigenetic Mechanisms of Transcriptional Deregulation in Unfavorable Neuroblastomas. *Cancer Res* **76**, 5523–5537 (2016).

72. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

73. Sims, D. *et al.* CGAT: computational genomics analysis toolkit. *Bioinform Oxf Engl* **30**, 1290–1 (2014).

74. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

75. Gaspar, J. M. Improved peak-calling with MACS2. *Biorxiv* 496521 (2018) doi:10.1101/496521.

76. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res* **46**, D754–D761 (2017).

77. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

78. Hadley *et al.* dplyr: A Grammar of Data Manipulation. (2019).

79. Hadley, W., Lionel H. & RStudio, Tidyr: Tidy Messy Data. (2017).

80. Hadley, W. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

81. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442–D450 (2019).