




ARTICLE

<https://doi.org/10.1038/s41467-019-13394-w>

OPEN

Alkyladenine DNA glycosylase associates with transcription elongation to coordinate DNA repair with gene expression

Nicola P. Montaldo^{1,2}, Diana L. Bordin^{1,9}, Alessandro Brambilla^{1,9}, Marcel Rösinger², Sarah L. Fordyce Martin¹, Karine Øian Bjørås¹, Stefano Bradamante¹, Per Arne Aas¹, Antonia Furrer^{2,8}, Lene C. Olsen ^{1,3,4}, Nicolas Kunath¹, Marit Otterlei¹, Pål Sætrum ^{1,3,4,5}, Magnar Bjørås^{1,6}, Leona D. Samson⁷ & Barbara van Loon ^{1,2*}

Base excision repair (BER) initiated by alkyladenine DNA glycosylase (AAG) is essential for removal of aberrantly methylated DNA bases. Genome instability and accumulation of aberrant bases accompany multiple diseases, including cancer and neurological disorders. While BER is well studied on naked DNA, it remains unclear how BER efficiently operates on chromatin. Here, we show that AAG binds to chromatin and forms complex with RNA polymerase (pol) II. This occurs through direct interaction with Elongator and results in transcriptional co-regulation. Importantly, at co-regulated genes, aberrantly methylated bases accumulate towards the 3' end in regions enriched for BER enzymes AAG and APE1, Elongator and active RNA pol II. Active transcription and functional Elongator are further crucial to ensure efficient BER, by promoting AAG and APE1 chromatin recruitment. Our findings provide insights into genome stability maintenance in actively transcribing chromatin and reveal roles of aberrantly methylated bases in regulation of gene expression.

¹Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. ²Department of Molecular Mechanisms of Disease, University of Zurich, 8057 Zurich, Switzerland. ³Bioinformatics core facility - BioCore, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. ⁴K.G. Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. ⁵Department of Computer Science, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. ⁶Department of Microbiology, Oslo University Hospital, University of Oslo, 0372 Oslo, Norway. ⁷Department of Biological Engineering, Department of Biology, David H. Koch Institute of Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ⁸Present address: Paul Scherrer Institute, 5232 Villigen, Switzerland. ⁹These authors contributed equally: Diana L. Bordin, Alessandro Brambilla. *email: barbara.v.loon@ntnu.no

Thousands of DNA base lesions are generated in each cell of our body every day as a consequence of exposure to endogenous and exogenous DNA damaging agents¹. Aberrantly methylated bases account for a large proportion of those lesions. Estimated steady-state levels of the most frequent aberrantly methylated bases: 7-methylguanine (7meG) and 3-methyladenine (3meA), are 6,000 and 1,200 lesions per mammalian cell per day respectively². Several lines of evidence strongly indicate that accumulation of aberrant DNA bases and genome instability accompany major human diseases like cancer, inflammatory diseases and neurological conditions^{3,4}.

Base excision repair (BER) is a highly efficient mechanism for the removal of aberrant DNA bases. This multi-step repair process is initiated by substrate-specific DNA glycosylases that recognize and remove aberrant bases^{3,5}. The resulting abasic (AP) site is processed by the apurinic/apyrimidinic endonuclease 1 (APE1) that generates a single-strand break (SSB), which is subsequently filled by DNA polymerase (pol) β , and the resulting nick sealed by DNA ligase III/XRCC1 (ref. 6). The BER specificity is determined by the type of DNA glycosylase that initiates the pathway⁷. Alkyladenine DNA glycosylase (AAG; also known as N-methylpurine-DNA glycosylase, MPG) is the major mammalian DNA glycosylase responsible for the removal of aberrantly methylated bases 7meG and 3meA^{6,7}. Besides abundant 3meA and 7meG, AAG is able to act on several other aberrant bases including hypoxanthine (Hx) and 1,N⁶-ethenoadenine³. AAG participates in both nuclear and mitochondrial DNA repair⁸. Much of our current knowledge about the different BER steps derives from *in vitro* studies using naked DNA as a substrate. However, eukaryotic cells need to repair DNA in the context of chromatin. Recent work demonstrated that the activity of human AAG, and of several other DNA glycosylases, is strongly impaired in the context of nucleosomes⁹. AAG activity is dramatically inhibited (84–100%) when the aberrant base is positioned mid-way or directly faces the surface of the histone octamer. Furthermore, the presence of nucleosomes was shown to significantly impair the activity of the downstream BER proteins APE1, DNA pol β and DNA ligase III/XRCC1 (refs^{10–12}). It was proposed that BER could overcome the inhibitory effect of nucleosomes by associating with processes that involve chromatin reorganization, such as replication and transcription^{13–15}. This idea is further supported by the notion that BER preferentially occurs on the transcribed strand^{16–18}. However, firm evidence demonstrating that BER associates with the transcription is still lacking. AAG could potentially be associated with the transcription through interaction with different transcription factors. It was shown that AAG forms a complex with a transcriptional repressor methylated DNA-binding domain 1 (MBD1) and that this complex could inhibit the transcription¹⁹. Furthermore, AAG was demonstrated to bind to the estrogen receptor (ER) α , which in turn stimulates AAG activity²⁰ and could potentially result in repression of the ER α -responsive genes. Interestingly, recent work in mice focusing on DNA glycosylases responsible for the repair of oxidized bases, similarly, suggested the role of glycosylases in regulation of ER α -gene expression²¹. Taken together, these findings indicate that coupling of AAG-initiated BER to transcription could enable more efficient repair and that AAG potentially influences gene expression.

Recent extensive protein interaction study²² predicted that AAG forms a complex with the active transcription machinery through a possible interaction with Elongator complex. Elongator is a highly conserved complex that participates in several pathways, including facilitation of transcription elongation by interacting with hyperphosphorylated RNA pol II^{23,24}. It is composed of six subunits ELP1 to ELP6 organized in a dodecamer^{25,26}. Every subunit is required for the complex to function. ELP1 (also known

as IKBKAP) is the largest subunit of the complex and provides a scaffolding function²⁷. ELP3 has histone acetyltransferase (HAT) activity and acetylates histone H3 (refs^{28,29}). Mutations in human *ELP1*, and consequently impaired Elongator function cause neurodevelopmental disorder, and lead to reduced expression of Elongator-dependent genes in patient-derived cells³⁰. Interestingly, loss of Elp1 in yeast causes hypersensitivity to methyl methanesulfonate (MMS) and hydroxyurea³¹. While the predicted association between AAG and Elongator complex suggests that AAG-initiated BER may be coordinated with transcription elongation, the interaction remains yet to be confirmed and its relevance explored. Taken together, several studies suggest that AAG could form a complex with different transcriptional components, however whether and where these interactions take place in the context of chromatin, as well as the extent to which they facilitate AAG-initiated BER and regulate gene expression remains unknown.

In this study we show that the majority of cellular AAG localizes at chromatin and is engaged in a complex with actively transcribing RNA pol II. This occurs primarily through direct interaction of the unstructured N-terminal AAG region with the ELP1 subunit of transcriptional Elongator complex. To our knowledge this is the first evidence of BER association with active transcription. RNA-sequencing experiments further show that AAG and ELP1 co-regulate genes, which are primarily repressed by AAG and stimulated by ELP1. Notably, at co-regulated genes, the endogenous aberrantly methylated DNA bases accumulate towards the 3' end in regions co-occupied by ELP1, elongating RNA pol II, and BER enzymes AAG and APE1. Chromatin recruitment of the BER enzymes is strongly dependent on Elongator presence and active transcription, since ELP1 loss as well as transcription inhibition cause globally reduced AAG and APE1 binding. Importantly, ability to interact with ELP1, presence of functional Elongator and active transcription are needed to ensure efficient AAG-initiated BER, and their inactivation results in impaired repair and significant accumulation of AAG substrates in the genome. Based on our findings, we propose that AAG, in concert with Elongator complex and active transcription, coordinates repair of aberrantly methylated DNA bases with regulation of gene expression.

Results

AAG associates with transcription to regulate expression. AAG predominantly localizes to the nucleus where it recognizes and removes aberrant DNA bases. However, it remains unknown how AAG initiates repair in the context of chromatin, and in which specific regions of the genome the repair takes place. To determine which portion of nuclear AAG is bound to the chromatin and could actively participate in the repair of genomic DNA, HEK293T cells were fractionated. The amount of AAG detected in the chromatin fraction (CF) was comparable to the total AAG fraction (TF), while only low levels of AAG were detected in the soluble fraction (SF) (Fig. 1a, b), thus indicating that the majority of cellular AAG is chromatin bound. In order to efficiently repair aberrant bases in chromatin context, DNA glycosylases were suggested to associate with processes that involve chromatin reorganization, such as transcription¹³. To determine if AAG engages with the active transcription, immunoprecipitation (IP) was performed and AAG-bound cellular complexes analyzed. Interestingly, the IP analysis indicated that AAG forms a complex with active RNA pol II phosphorylated at Serine 2 of the C-terminal domain (RNA pol II S2P) (Fig. 1c). Presence of different nucleases (DNaseI, MNase and RNaseI) did not majorly affect the level of IP-ed RNA pol II S2P (Fig. 1c, compare lanes 6, 7 to 5), thus suggesting that complex formation between AAG and RNA

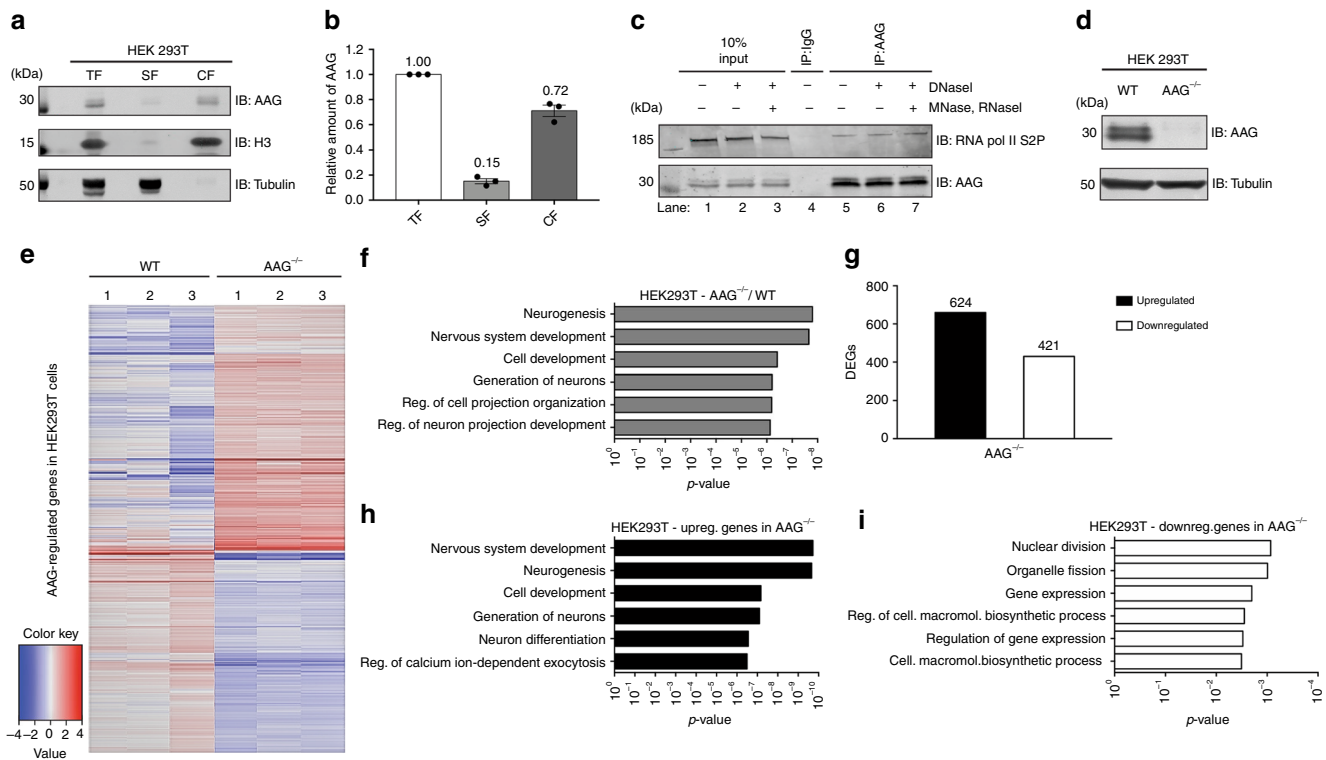


Fig. 1 AAG associates with active transcription and regulates gene expression. **a** Immunoblot analysis of chromatin fractionation assay indicating AAG distribution in total fraction (TF), soluble supernatant fraction (SF) and chromatin fraction (CF). Histone H3 and α -tubulin served as controls. **b** Quantification of three independent experiments as the one depicted in **a**; error bars represent mean \pm SEM ($n = 3$). **c** Immunoprecipitation of AAG from HEK293T whole-cell extracts (WCEs) untreated or treated with DNaseI, Mnase, and RNaseI, showing the interaction with RNA polymerase II phosphorylated at Serine 2 (S2P) of CTD repeat. **d** Immunoblot analysis of WCEs from HEK293T WT and AAG^{-/-} cells generated by CRISPR-Cas9 technology. **e** Heat map of the expression levels of AAG-regulated genes in HEK293T cells. Color scale is representing change in the gene expression depicting log₂ fold change relative to the mean. **f** Top six biological processes (BP) gene ontology (GO) terms as determined by the Database for Annotation, Visualization and Integrated Discovery (DAVID) for genes dysregulated in HEK293T AAG^{-/-} cells when compared to WT. **g** Up- and down-regulated differentially expressed genes (DEGs) in HEK293T AAG^{-/-} cells. **h, i** Top six BP GO terms as determined by DAVID for genes upregulated (**h**) and downregulated (**i**) genes. Source data are provided as Source Data file.

pol II is DNA and RNA independent, and likely mediated through protein-protein interactions. To determine regions in which transcription is influenced by AAG presence, we generated HEK293T cells lacking AAG (AAG^{-/-}) using the CRISPR-Cas9 engineering approach (Fig. 1d), and subjected them to RNA sequencing (Fig. 1e). Comparison of wild type (WT) and AAG^{-/-} transcriptomes revealed that, at ≥ 1.5 -fold change and an FDR ≤ 0.1 , 1,045 genes were differentially expressed in AAG^{-/-} cells. The subsequent DAVID gene ontology (GO) enrichment analysis³² revealed that the majority of differentially expressed genes (DEGs) in AAG^{-/-} belonged to neurogenesis and nervous system development processes, 12.3 and 16.8% respectively (Fig. 1f, Supplementary Data 1). These findings are further supported by analysis of DEGs in HAPI AAG^{-/-} cells, which determined the neurodevelopmental processes as the most significantly enriched (Supplementary Fig. 1, Supplementary Data 1). Subsequent separation of DEGs based on the expression change, revealed that the large portion of genes affected by the loss of AAG are upregulated (Fig. 1g). These upregulated genes were the main drivers of the GO, segregating in the nervous system development and neurogenesis processes (Fig. 1h, Supplementary Data 1), while the downregulated genes produced no significant GO terms (Fig. 1i, Supplementary Data 1). Taken together these findings suggest that AAG binds to chromatin, forms a complex with the active transcription and regulates gene expression.

AAG directly binds the transcriptional Elongator complex. AAG-initiated BER can be linked to transcription through direct association with RNA pol II or by binding other components of the transcription machinery. To identify proteins that form a complex with AAG, we expressed and affinity-purified the FLAG-tagged AAG from HEK293T cells, either untreated or exposed to the alkylating agent MMS. A control mock purification was performed in parallel from cells transfected with empty FLAG vector. Proteins in all samples were visualized by silver staining, indicating AAG presence in appropriate fractions (Fig. 2a). The affinity-purified samples were next subjected to liquid chromatography-tandem mass spectrometry (LC/MS-MS) analysis (Supplementary Data 2). This approach indicated ELP1 as the most enriched AAG interacting partner. In addition, ELP2 and 3 were also detected in AAG samples, thus suggesting that whole holo-Elongator complex is present. MMS exposure did not appear to affect AAG-Elongator complex formation, since a comparable number of peptide spectra were detected in the untreated and treated samples (Supplementary Data 2). Subsequent IP experiments showed that, similar to the exogenous FLAG-AAG, the endogenous AAG forms a complex with ELP1 independently of the MMS treatment (Fig. 2b). Further, the separation of protein complexes present in the HeLa whole cell extract (WCE) revealed that the main Elongator subunits co-elute with AAG (Supplementary Fig. 2a). Since both AAG and Elongator

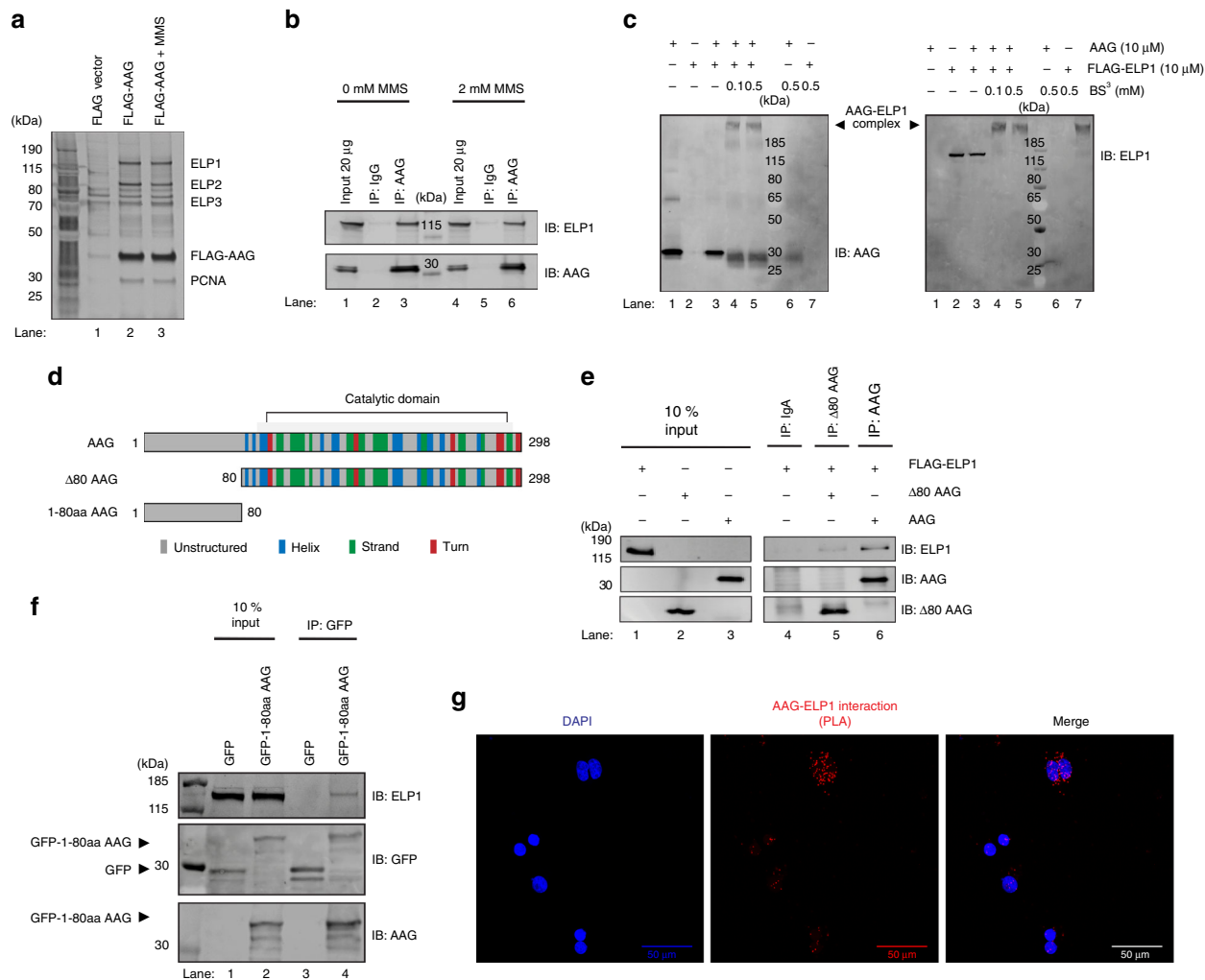


Fig. 2 AAG directly interacts with ELP1 subunit of transcriptional Elongator complex. **a** Silver staining analysis of untreated or MMS (methyl methanesulfonate) treated FLAG-AAG immunoprecipitation (IP) samples subjected to LC/MS-MS analysis. IP sample from HEK293T cells transfected with empty FLAG vector served as negative control. LC/MS-MS analysis of FLAG-AAG IP samples identified ELP1 as the most specifically enriched AAG interacting partner. Remaining subunits of core Elongator (ELP2 and 3), and known interacting partner PCNA are depicted. See also Supplementary Data 2 for the complete list of proteins identified in all samples. **b** IP of AAG from HEK293T whole cell extracts (WCEs) treated or untreated with MMS. **c** Immunoblot analysis of gel shift experiments with purified recombinant AAG and FLAG-ELP1 proteins, and chemical crosslinking with indicated amounts of BS³ (bissulfosuccinimidyl suberate). Control samples - proteins in the absence of crosslinking agent (lanes 1-3); and single proteins with highest amount of BS³ (lanes 6, 7). **d** Schematic representation of full-length AAG, AAG lacking 80 N-terminal amino acids (Δ80 AAG), and first 80 N-terminal amino acids of AAG (1-80aa AAG); numbers indicate amino acids. **e** Co-IP of purified FLAG-ELP1 with AAG full length, or AAG lacking 80 N-terminal amino acids (Δ80 AAG). **f** IP of GFP-tagged first 80 N-terminal amino acids of AAG (1-80aa AAG) (GFP-1-80aa AAG) expressed in HEK293T AAG^{-/-} cells. GFP-tag positioned N-terminally. **g** Proximity ligation assay (PLA) showing the AAG-ELP1 interaction in HEK293T cells. Scale bar: 50 μm. Source data are provided as a Source Data file.

are DNA-binding proteins, we next examined whether the complex formation is dependent on the presence of nucleic acids. Addition of DNaseI, MNase or RNaseI during either AAG or ELP1 IP did not influence the levels of co-precipitated proteins (Supplementary Fig. 2b, c). These findings thus suggest that in human cells endogenous AAG forms a complex with Elongator, independently of the presence of nucleic acids and exposure to genotoxic stress. To examine if AAG binds directly to Elongator, gel shift experiments were performed using purified recombinant AAG and the FLAG-tagged ELP1 (Supplementary Fig. 2d, e), and the chemical crosslinking with bissulfosuccinimidyl suberate (BS³). ELP1 was targeted, since it was the predominant Elongator subunit detected by LC/MS-MS in the complex with AAG (Supplementary Data 2). Interestingly, the gel shift experiments combined with chemical

crosslinking revealed that AAG efficiently forms a complex with the ELP1 dimer, indicating direct binding of these proteins (Fig. 2c, lanes 4, 5) and further supporting earlier reports of ELP1 dimerization^{33,34} (Fig. 2c, lane 7). No complex formation was observed in control reactions with crosslinker and AAG alone (Fig. 2c, lane 6). Since AAG is composed of an unstructured N-terminal region and a catalytic DNA glycosylase domain (Fig. 2d), we next addressed which AAG region is important for the interaction with ELP1. Incubation of recombinant FLAG-ELP1 with full-length AAG or AAG lacking 80 N-terminal amino acids (Δ80 AAG) (Supplementary Fig. 2f) and subsequent IP analysis, indicated that both AAG forms interact with FLAG-ELP1, although the amount of precipitated Elongator was markedly reduced with Δ80 AAG (Fig. 2e). This result suggested that the unstructured N-terminal region might

be important for AAG-ELP1 interaction. To further test this HEK293T AAG^{-/-} cells were complemented with the 80 N-terminal amino acids (1-80aa) of AAG fused to a GFP tag, positioned either N- or C-terminally (GFP-1-80aa AAG and 1-80aa AAG-GFP, respectively). Subsequent IP indicated that 1-80aa AAG is sufficient to bring down ELP1 (Fig. 2f and Supplementary Fig. 2g), indicating that the unstructured N-terminal AAG region successfully binds ELP1. Next, to visualize the AAG-ELP1 interaction within cells proximity ligation assay (PLA) was performed, showing that the AAG-ELP1 interaction mainly occurs in the nucleus (Fig. 2g and Supplementary Fig. 2h, i). Interestingly, a small fraction of the signal was detected surrounding the nucleus. Taken together, these findings indicate that in the nucleus AAG binds directly to the ELP1 subunit of the transcriptional Elongator complex, predominantly through the unstructured N-terminal domain of AAG.

AAG and Elongator co-regulate gene expression. To identify genes and regions at which AAG and Elongator have the most significant impact we generated, in addition to AAG^{-/-} cells, the HEK293T cells lacking ELP1 (ELP1^{-/-}), and both AAG and ELP1 (AAG^{-/-}ELP1^{-/-}) (Fig. 3a), via the CRISPR-Cas9 approach. Subsequent RNA sequencing analysis of ELP1^{-/-} cells showed that loss of ELP1 results in ≥2-fold altered expression of 489 genes, with FDR ≤ 0.1. Similar to AAG^{-/-} cells, 21.2% of all DEGs in ELP1^{-/-} cells belonged to the nervous system

development processes (Fig. 3b, Supplementary Data 1). The majority of the DEGs were downregulated in ELP1^{-/-} cells (343 out of 489) (Fig. 3c), thus supporting previous observation that Elongator promotes transcription³⁰. Comparison of genes differentially expressed in AAG^{-/-} and ELP1^{-/-} cells resulted in identification of 113 co-regulated genes (Fig. 3d) that predominantly clustered into processes involving generation of neurons and neurogenesis (Fig. 3e, Supplementary Data 1). Interestingly, the majority of co-regulated genes were upregulated in AAG^{-/-} and downregulated in ELP1^{-/-} cells (Fig. 3f), thus following the general direction of all DEGs in these two cell lines (Figs. 1g and 3c). The RNA sequencing results were confirmed by qPCR measurements of mRNA levels for multiple co-regulated genes belonging to the top identified biological processes presented in Fig. 3e. The mRNA levels of ALDH1A2, CRMP1, CDH23, SYT9, CDH4, NPTX2, NOVA2, and CDH22 were, as expected, significantly increased in AAG^{-/-}, and decreased in ELP1^{-/-} cells (Fig. 3g-i and Supplementary Fig. 3), while the level of non-affected control YTHDC1 was unchanged in all tested cell lines (Fig. 3j). The observed increase in the expression of co-regulated genes was similar in two independent AAG^{-/-} clones (Supplementary Fig. 4). Notably, in AAG^{-/-}ELP1^{-/-} cells, the co-regulated genes were significantly downregulated in comparison to WT levels, thus following the direction of ELP1^{-/-} cells (Fig. 3g-i and Supplementary Fig. 3). This suggests that at the transcriptional level Elongator acts upstream of AAG. In summary, these results provide evidence of genes co-regulated

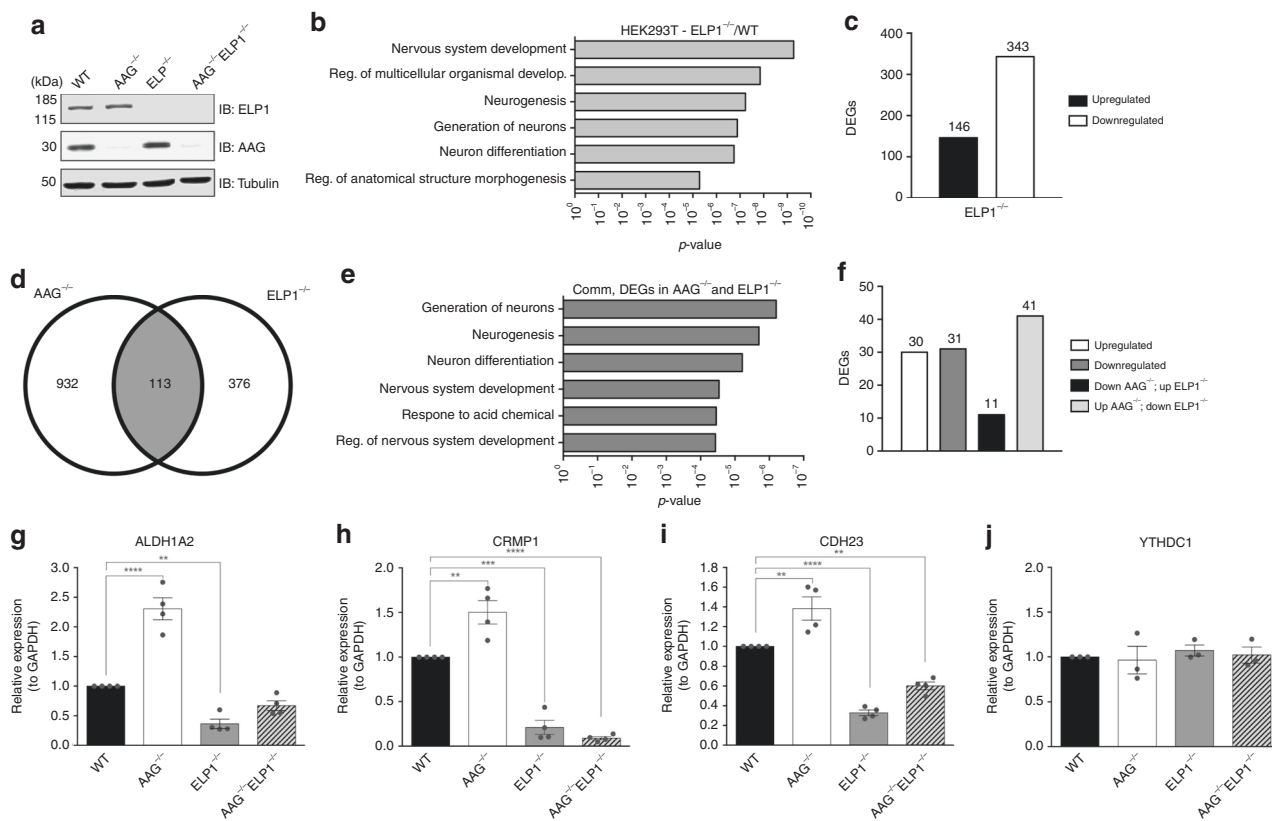


Fig. 3 Elongator and AAG co-regulate gene expression. **a** Immunoblot of HEK293T whole-cell extracts (WCEs) from WT, AAG^{-/-}, ELP1^{-/-}, and AAG^{-/-}ELP1^{-/-} cell lines generated via CRISPR-Cas9 technology. **b** Top six biological processes (BP) gene ontology (GO) terms as determined by DAVID for genes dysregulated in HEK293T ELP1^{-/-} cells when compared to WT cells. **c** Downregulated and upregulated differentially expressed genes (DEGs) in HEK293T ELP1^{-/-} cells. **d** Venn diagrams of AAG- and ELP1-regulated genes in HEK293T cells. **e** Top six BP GO terms as determined by DAVID of genes co-regulated by AAG and ELP1 in HEK293T cells. **f** Directionality of DEGs regulated by AAG and ELP1 in HEK293T cells. **g-j** Relative mRNA levels of *ALDH1A2* (**g**), *CRMP1* (**h**), *CDH23* (**i**) and *YTHDC1* (negative control) (**j**) genes in HEK293T WT, AAG^{-/-}, ELP1^{-/-} and AAG^{-/-}ELP1^{-/-} cells. Error bars represent mean ± SEM (n ≥ 3). **p ≤ 0.01; ***p ≤ 0.001; ****p ≤ 0.0001, one-way ANOVA. Source data are provided as Source Data file.

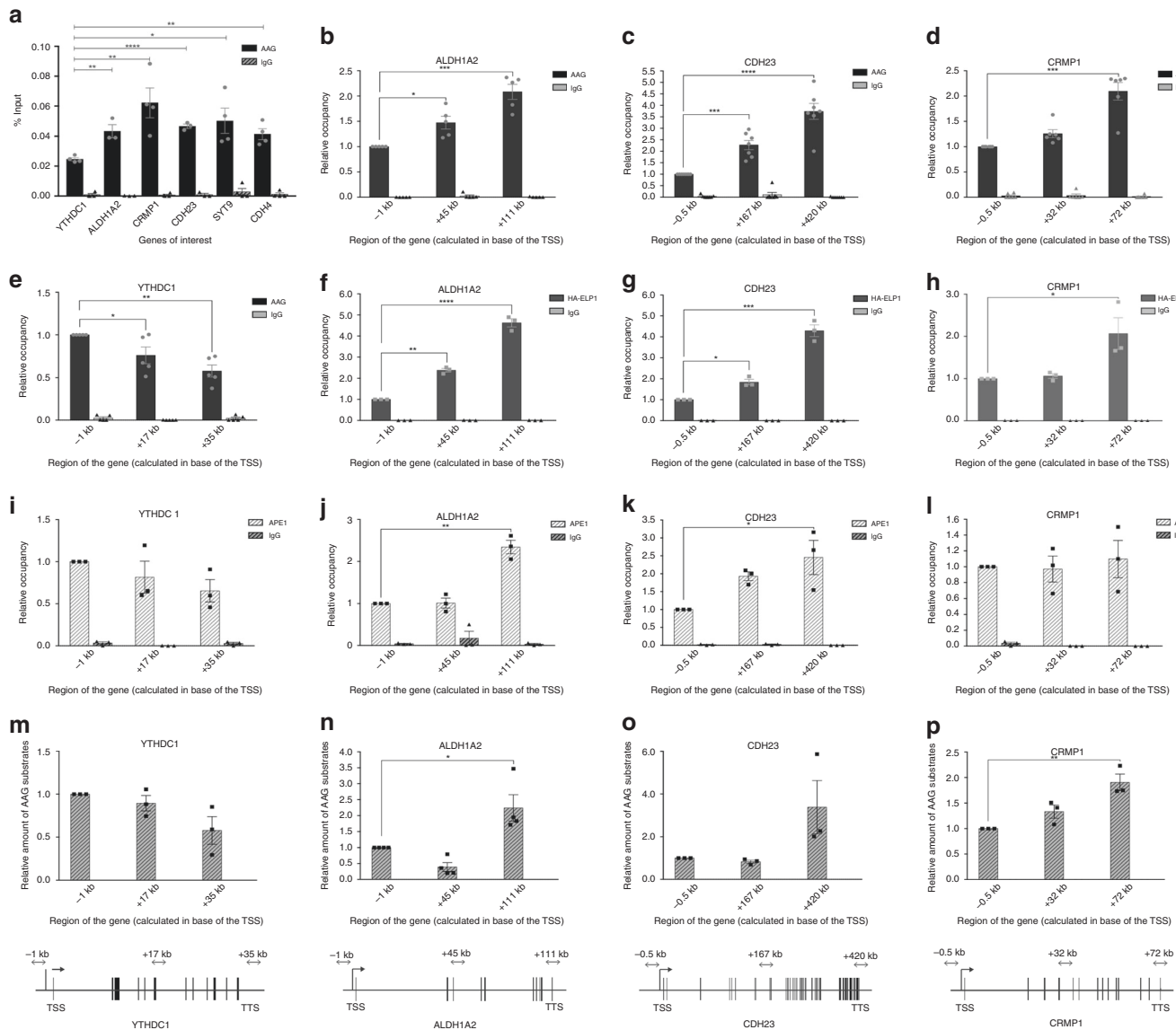


Fig. 4 Elongator, components of AAG-initiated BER and AAG substrates accumulate towards the 3' end of co-regulated genes. **a** AAG ChIP-qPCR assays in HEK293T WT cells comparing percentage of input at gene bodies of unaffected gene (*YTHDC1*) and differentially expressed genes (*ALDH1A2*, *CRMP1*, *CDH23*, *SYT9*, and *CDH4*). **b–e** ChIP-qPCR assays showing relative AAG occupancy in AAG- and ELP1- dependent genes *ALDH1A2* (**b**), *CDH23* (**c**), *CRMP1* (**d**), and unaffected gene *YTHDC1* (**e**) in HEK293T WT cells. **f–h** ChIP-qPCR assays showing relative occupancy of HA-ELP1 in AAG- and ELP1- dependent *ALDH1A2* (**f**), *CDH23* (**g**), and *CRMP1* (**h**) genes in HEK293T HA-ELP1 cells. **i–l** ChIP-qPCR assays showing relative APE1 occupancy in negative control *YTHDC1* gene (**i**), and AAG- and ELP1-dependent *ALDH1A2* (**j**), *CDH23* (**k**), *CRMP1* (**l**) genes in HEK293T WT cells. **m–p** ChIP DNA damage assay showing differences in distribution pattern of aberrant AAG substrates in unaffected gene *YTHDC1* (**m**) and genes regulated by AAG and ELP1: *ALDH1A2* (**n**), *CDH23* (**o**), *CRMP1* (**p**) in HEK293T WT cells. Values are shown as relative occupancy: % input of specific gene region, relative to percentage input of promoter region. Error bars indicate mean \pm SEM ($n \geq 3$). Two-tailed Student's *t*-test in **a**; one-way ANOVA in **b–p**; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$. Source data are provided as Source Data file.

by the AAG DNA repair glycosylase and its interaction partner the Elongator complex.

AAG and Elongator accumulate in aberrant base-rich regions.

To determine whether AAG and ELP1 localize at the co-regulated genes identified in Fig. 2, chromatin immunoprecipitation (ChIP) experiments were performed. AAG ChIP showed significant enrichment at the co-regulated genes (*ALDH1A2*, *CRMP1*, *CDH23*, *SYT9*, and *CHD4*), when compared to *YTHDC1* unaffected control (Fig. 4a). Notably, AAG binding was most significantly increased towards the 3' end of the co-regulated genes (*ALDH1A2*, *CRMP1*, *CDH23*) (Fig. 4b–d). The defined 3' end

enrichment was characteristic to co-regulated genes and was not present at *YTHDC1* (Fig. 4e). To determine the relation between ELP1 and AAG distribution, HEK293T cells with endogenously HA-tagged ELP1 were generated, using homologous recombination dependent CRISPR-Cas9 gene editing (Supplementary Fig. 5). Subsequent HA-ChIP experiments revealed that similar to the AAG distribution, and in line with its role in the transcription elongation, HA-ELP1 was significantly enriched towards the 3' end of the co-regulated genes (Fig. 4f–h). Importantly, the same distribution pattern was observed for RNA pol II phosphorylated at Serine 2 of the C-terminal domain (CTD) (RNA pol II S2P), which is the predominant form during transcription elongation (Supplementary Fig. 6). Further, to test if other BER enzymes co-

occupy the same regions as AAG and ELP1, APE1 ChIP was performed. APE1 distribution closely resembled AAG and ELP1 localization at the co-regulated genes, with the highest levels detected towards the 3' end (Fig. 4i–l). Collectively, these results suggest that AAG-initiated BER associates with Elongator and transcription elongation, predominantly at the 3' end of the co-regulated genes. Since the main AAG function is to initiate BER, we next analyzed levels of aberrantly methylated AAG substrates along the co-regulated genes, using real-time qPCR based approach for quantification of aberrant DNA bases³⁵. Interestingly, the distribution of endogenous aberrantly methylated AAG substrates closely followed AAG, APE1, ELP1 and RNA pol II pattern, with the levels of aberrant bases being highest towards the 3' end of the co-regulated genes (Fig. 4m–p). Taken together these findings suggest that regions of co-regulated genes that are co-occupied by AAG-initiated BER and Elongator have high levels of aberrant bases, thus indicating an interplay between the repair of DNA base lesions and transcription regulation.

Loss of functional Elongator impairs AAG-initiated repair. Since AAG interacts directly with ELP1 and the two proteins accumulate in the same gene regions, we next addressed the importance of ELP1 for the recruitment of AAG-initiated BER to

the chromatin. AAG ChIP in WT and ELP1^{-/-} cells revealed that loss of ELP1 causes reduction in AAG binding to the chromatin at all genes tested (Fig. 5a–d), supporting that ELP1 is essential for AAG recruitment during transcription (Fig. 3g–j and Supplementary Fig. 3). Reduced AAG binding was not a consequence of perturbed chromatin organization in ELP1^{-/-} cells, since histone occupancy was comparable in WT and ELP1^{-/-} cells, as indicated by histone 3 ChIP (Supplementary Fig. 7a). Analysis of chromatin fractions isolated from WT and ELP1^{-/-} cells demonstrated that ELP1 loss leads to global reduction of AAG on chromatin (Fig. 5e, f), while the total AAG levels were unchanged (Fig. 3a), further supporting the role of ELP1 in facilitating AAG recruitment. Similar to the impact on AAG localization, loss of ELP1 resulted in a notable reduction of APE1 recruitment to all tested genes (Fig. 5g). Since ELP1 promotes AAG chromatin recruitment, we next tested if ELP1 status could influence the complex formation between AAG and hyperphosphorylated RNA pol II. IP of AAG in WT and ELP1^{-/-} HEK293T cells showed that loss of ELP1 diminishes the interaction between AAG and active RNA pol II S2P (Fig. 5h, compare lane 5 to 6). These findings suggest that ELP1 plays an important role in associating AAG with active transcription. To further evaluate the extent to which impaired AAG recruitment and reduced association with active transcription influence repair of aberrant bases in ELP1^{-/-}

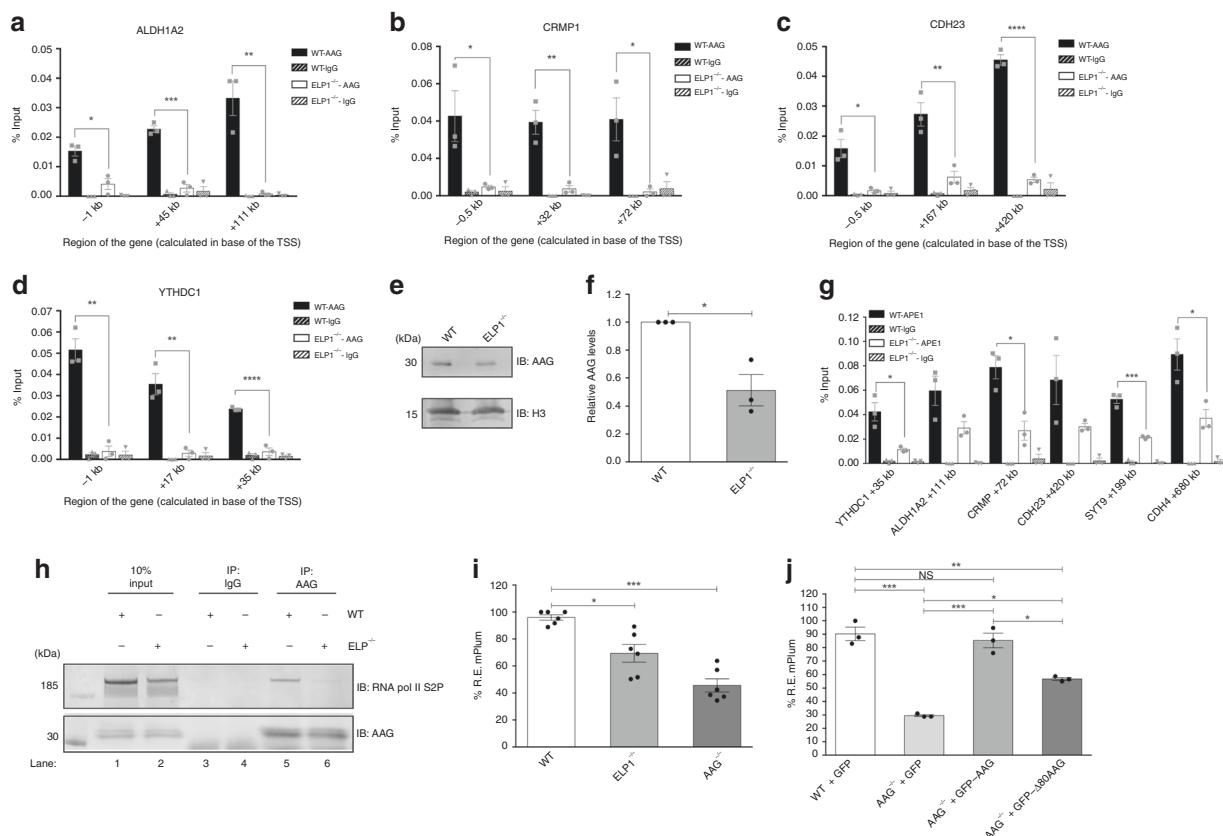


Fig. 5 Lack of functional Elongator hampers AAG and APE1 chromatin recruitment, and AAG-initiated repair. **a–d** ChIP-qPCR experiments comparing AAG occupancy at promoters and gene bodies of *ALDH1A2* (**a**), *CRMP1* (**b**), *CDH23* (**c**), and *YTHDC1* (**d**) in HEK293T WT and ELP1^{-/-} cells. **e** Immunoblot analysis of AAG levels in chromatin fraction of HEK293T WT and ELP1^{-/-} cells. Histone H3 served as control. **f** Quantification of experiments as the one depicted in **e**. **g** ChIP-qPCR experiments comparing APE1 occupancy at gene bodies of *YTHDC1*, *ALDH1A2*, *CRMP1*, *CDH23*, *SYT9*, and *CDH4* in HEK293T WT and ELP1^{-/-} cells. Error bars represent the SEM calculated from at least three independent experiments. **h** Immunoprecipitation of AAG from HEK293T WT and ELP1^{-/-} whole cell extracts showing the interaction with RNA polymerase II phosphorylated at Serine 2 (S2P) of CTD repeat. **i** Measurement of AAG DNA glycosylase activity in HEK293T WT, AAG^{-/-} and ELP1^{-/-} on hypoxanthine (Hx)-containing plasmid by FM-HCR assays. **j** AAG DNA glycosylase activity determined by FM-HCR assays on hypoxanthine (Hx)-containing plasmid in HEK293T WT cells complemented with GFP and AAG^{-/-} cells complemented with GFP, GFP-AAG or GFP-Δ80 AAG. Error bars represent mean ± SEM ($n \geq 3$). Two-tailed Student's *t*-test (**a–d**, **f**, **g**); one-way ANOVA (**i**, **j**), NS – non significant, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$. Source data are provided as Source Data file.

cells, AAG-initiated BER capacity was analyzed by flow cytometry host cell reactivation assay (FC-HCR)³⁶. Importantly, cells lacking ELP1 showed significantly decreased capacity to repair AAG substrate hypoxanthine (Hx), when compared to WT cells (Fig. 5i). Hx was chosen over 3meA and 7meG, since these aberrantly methylated AAG substrates are unstable, and thus cannot be readily incorporated at the specific site in the DNA repair construct. While our findings suggest that global loss of ELP1 impairs repair (Fig. 5i), the necessity of ELP1 binding for efficient AAG-initiated BER remains unclear. To address this we compared BER capacity by FC-HCR analysis, in AAG^{-/-} cells complemented with full-length or Δ 80 AAG, lacking the unstructured region important for ELP1 binding (Fig. 2). Importantly, previous work showed that the catalytic activities of full-length AAG and Δ 80 AAG are very similar^{37,38}. Both full-length and Δ 80 AAG were GFP-tagged, allowing to directly relate BER capacity to the amount of protein present in the cells, and were expressed at level comparable to endogenous AAG in WT cells (Supplementary Fig. 7b). Interestingly, while complementation of AAG^{-/-} cells with GFP-AAG successfully rescued repair capacity, the AAG-initiated BER remained significantly reduced in AAG^{-/-} cells expressing GFP- Δ 80, when compared to WT cells (Fig. 5j). Taken together, our findings suggest that ELP1 binding as well as functional Elongator complex, are prerequisite for effective AAG chromatin recruitment, association with actively transcribing RNA pol II and AAG-initiated BER.

Active transcription promotes repair of aberrant bases. Our findings suggest that AAG-initiated BER is associated with transcription through direct interaction with ELP1, and that Elongator promotes AAG-initiated BER in actively transcribing genes (Figs. 2 and 5). It remains however unclear to which extent active transcription directly influences AAG occupancy, and what is its importance in the repair of aberrant bases. To determine if transcription inhibition affects AAG distribution, ChIP experiments were performed in untreated cells and cells exposed to 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB). Treatment with transcription inhibitor DRB results in repression of transcription elongation, as consequence of impaired RNA pol II CTD phosphorylation at Serine 2 (refs^{39,40}), thus causing reduced

RNA pol II occupancy (Supplementary Fig. 8). Similar to the ELP1 loss (Fig. 5), transcription inhibition resulted in globally reduced AAG binding at all analyzed chromatin regions in DRB treated cells (Fig. 6a–d). This further suggests that transcription is an important modulator of AAG chromatin occupancy. To next determine if active transcription, besides promoting AAG occupancy, also has a role in AAG-initiated repair a single-cell AAG-Comet FLARE analysis was performed. Interestingly, inhibition of transcription elongation caused significant accumulation of AAG-specific aberrant bases, while it did not affect global DNA damage levels (Fig. 6e). In summary, these results strongly suggest that active transcription is required to promote AAG recruitment to the chromatin and to facilitate AAG-initiated BER of aberrant bases.

Discussion

Removal and repair of aberrant bases are dramatically impaired in the context of chromatinized DNA^{9–12,41}. It has been suggested that for efficient repair within chromatin to take place, BER needs to be associated with essential nuclear processes, such as transcription⁶. Besides the potential importance of transcription in promoting efficient BER, several studies indicated that BER enzymes, in particular DNA glycosylases, could influence transcription and play an important role in modulation of gene expression^{19–21,42}. However, direct evidence for the existence of transcription associated BER has not been provided so far. In this work we show that AAG-initiated BER associates with the transcription machinery, primarily by binding to the Elongator complex (Fig. 2). Set of IP experiments indicates that AAG forms a complex with active RNA pol II (Fig. 1c), predominantly through direct association of its unstructured N-terminal region with the ELP1 subunit of the Elongator (Figs. 2 and 5h). Identification of direct interaction between AAG and the ELP1 confirms recent predictions arising from a large proteomics screen²². Results of PLA experiments further indicate that this interaction primarily takes place in the nucleus (Fig. 2g). The relevance of AAG interaction with Elongator is demonstrated by the RNA sequencing, presented in Fig. 3d, which revealed that AAG and ELP1 co-regulate a specific set of genes. On the chromatin level, the occupancy of the BER components AAG and APE1 accompanies the progressive enrichment of ELP1 and elongating RNA pol II at

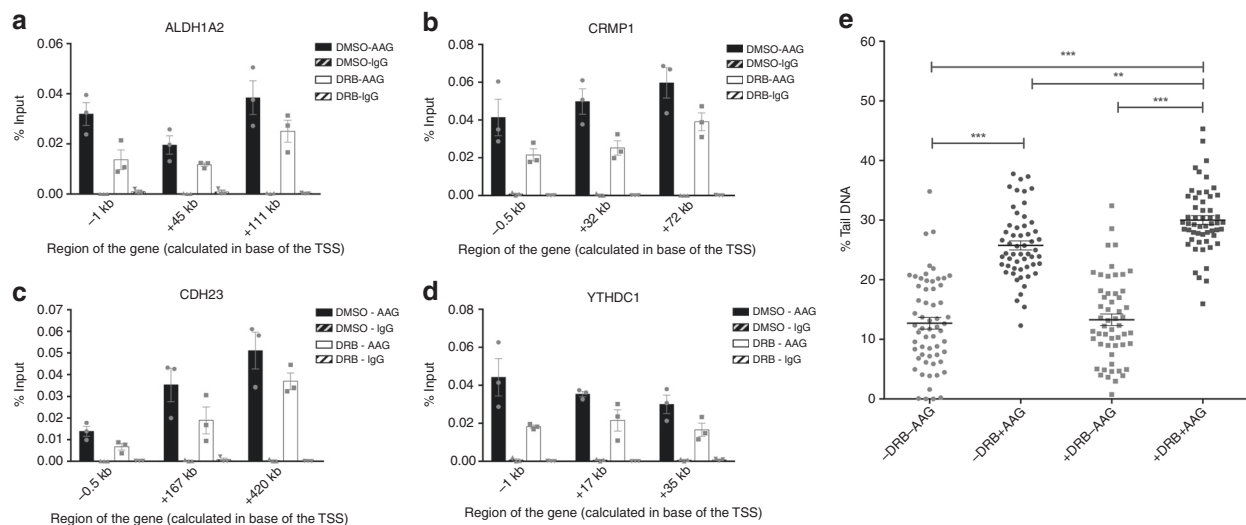


Fig. 6 Transcription inhibition reduces AAG occupancy and impairs removal of AAG substrates. **a–d** ChIP-qPCR experiments comparing AAG occupancy in DMSO and DRB treated HEK293T WT cells at *ALDH1A2* (**a**), *CRMP1* (**b**), *CDH23* (**c**), and *YTHDC1* (**d**) genes. Error bars, SEM from three independent experiments. **e** AAG Comet-FLARE assay in DMSO or DRB treated HEK293T cells. Error bars indicate mean \pm SEM ($n \geq 3$). $**p < 0.01$; $***p < 0.001$, one-way ANOVA. Source data are provided as Source Data file.

the 3' end of the co-regulated genes (Fig. 4). The ELP1 and RNA pol II distribution patterns are directly in line with the previous observations³⁰. Importantly, regions of the co-regulated genes, enriched for AAG, APE1, and ELP1 present with the highest level of endogenous aberrantly methylated AAG substrates (Fig. 4m–p). Thus suggesting potential role of aberrantly methylated bases in regulation of gene expression. This is particularly interesting in light of recent work, which indicated coevolution of epigenetic DNA methylation with repair of aberrantly methylated DNA bases, including AAG-initiated BER⁴³. Further, accumulation of aberrantly methylated AAG substrates coincides with accumulation of elongating RNA pol II S2P (Supplementary Fig. 6), which was shown to be indicative of a reduced elongation rate due to increased nucleosome density⁴⁴. Indeed, the findings provided by genome-wide mapping of aberrantly methylated bases in yeast, suggested that repair of AAG substrates is slower in the regions with strongly positioned nucleosomes, in comparison to the nucleosome depleted regions⁴⁵.

As in the case of RNA pol II complex formation, the recruitment of AAG-initiated BER to the chromatin is strongly dependent on functional Elongator, and results in dramatic decrease of AAG and APE1 occupancy, both in specific gene regions, as well as in chromatin bound fraction of ELP1^{-/-} cells (Fig. 5). Besides impact on the chromatin recruitment, analysis of AAG-repair capacity revealed that ELP1 binding is necessary for efficient AAG-initiated repair (Fig. 5i, j). Accordingly, both loss of functional Elongator in ELP1^{-/-} cells, as well as lack of the ELP1 binding region within AAG ($\Delta 80$ AAG), resulted in significantly impaired AAG-initiated repair. In line with Elongator importance for efficient repair, AAG-initiated BER is further directly dependent on active transcription as indicated by experiments involving transcription elongation inhibitor DRB. Notably, DRB treatment resulted in reduced AAG occupancy and impaired AAG-initiated repair (Fig. 6), thus suggesting that active transcription has an important role in the maintenance of genome stability by facilitating BER. Very recently Menoni et al. similarly indicated that the inhibition of transcription elongation by DRB results in reduced recruitment of the BER scaffold protein XRCC1 to the sites of initiated repair⁴⁶. Interestingly, early work in mouse cells did not detect impact of transcription coupled repair in clearance of methylation damage⁴⁷. However this study largely focused on the removal of aberrantly methylated bases from a single specific gene. The fact that inhibition of transcription elongation impairs AAG-initiated BER and results in global accumulation of aberrantly methylated AAG substrates, strongly supports the idea that one key functional consequence of Elongator-mediated AAG association with the active transcription is ensuring efficient repair. Accumulation of AAG substrates genome wide could directly have harmful effects on transcription, since 3meA was shown to affect the fidelity of incorporation by RNA pol II⁴⁸. While the 3meA analog 3-deaza-3meA can successfully be bypassed by RNA pol II, its presence causes 10-fold greater misincorporation of CMP, thus resulting in synthesis of mRNA's bearing mutations. In contrast to 3meA, AP site intermediates, generated by APE1 during BER, impair RNA pol II progression⁴⁹. Finding that AAG, similarly to other DNA glycosylases⁴², primarily inhibits gene expression (Fig. 1g) suggests that by coordinating BER initiation with transcription, AAG could potentially inhibit RNA pol II progression and therefore repress generation of faulty transcripts. Imbalanced AAG-initiated BER and consequent accumulation of aberrantly methylated bases can thus, besides causing genome instability and increased mutation rate, also impact gene expression.

Loss of AAG results in increased expression of genes that primarily segregate in neurodevelopmental processes (Fig. 1 and Supplementary Fig. 1). It has been demonstrated recently that the

DNA glycosylases *Ogg1* and *Mutyh*, involved in repair of oxidized bases, play an important role in neurodevelopment through impact on transcription. Lack of *Ogg1* and *Mutyh* was reported to upregulate ER α target-genes, which in turn modulated cognition and anxiety-like behavior in mice²¹. It will be thus interesting to determine the importance of transcription associated AAG-initiated BER in regulation of neurodevelopment, and test its role in brain functioning.

Taken together, based on our findings we propose a model of AAG-initiated DNA repair coordinated with gene expression (Fig. 7). During transcription Elongator associates with RNA pol II and promotes transcription elongation, accumulating towards the 3' end of regulated genes. Through its unstructured N-terminal region AAG associates with the ELP1 subunit of Elongator and forms a complex with the active transcription machinery. As a consequence of active transcription chromatin is suggested to be locally decondensed, which allows AAG to efficiently initiate BER by recognizing and removing aberrant bases. BER initiation likely temporarily inhibits RNA pol II progression, resulting in reduced expression of co-regulated genes. In the absence of Elongator, transcription of target genes is impaired due to reduced elongation, while AAG chromatin recruitment and BER initiation are hampered. In summary our results suggest that the transcriptional status, in addition to the type of aberrant bases and DNA glycosylases involved, is an essential determinant of BER efficiency and its impact on genome and chromatin organization. Association of AAG with transcription elongation could thus, in addition to enabling efficient removal of aberrantly methylated bases, serve as an important layer of gene expression control.

Methods

Cells and culturing. Human embryonic kidney cells expressing a mutant version of the SV40 large T antigen (HEK293T, American Tissue Culture Collection, CRL-3216™) were maintained under 5% CO₂ and 37 °C in DMEM high glucose supplemented with 10% fetal bovine serum, and 1% penicillin/streptomycin. HAP1 cells knockout and parental WT (Horizon Genomics, HZGHC001537c002 and HZGHC001537c003) were maintained under 5% CO₂ and 37 °C in IMDM supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, and 1% L-Glutamine (L-Glut).

Fractionation. To obtain total fraction (TF) 2.5×10^6 HEK293T cells were resuspended in 200 μ L of MNase buffer (50 mM Tris-HCl pH 7.5, 30 mM KCl, 7.5 mM NaCl, 4 mM MgCl₂, 1 mM CaCl₂, 0.125% (v/v) NP-40, 0.25% Na-deoxycholate, 0.3 M sucrose, 1x Halt™ Protease Inhibitor Cocktail) with 1U MNase, and the samples incubated at 37 °C for 30 min. For the preparation of the soluble fraction (SF) and chromatin fraction (CF) 2.5×10^6 cells were resuspended in 200 μ L of chromatin extraction buffer (10 mM HEPES pH 7.6, 3 mM MgCl₂, 1 mM DTT, 1x Halt™ Protease Inhibitor Cocktail) and samples rotated 30 min at RT. SF was separated from nuclei by low-speed centrifugation (1300 \times g for 10 min) and collected for further analysis. Pelleted nuclei were lysed in 200 μ L of MNase buffer with 1U MNase, for 30 min at 37 °C. TF, SF, and CF were boiled in Laemmli buffer, centrifuged at 16,000 \times g for 5 min at RT, and equal volumes of each fraction subjected to immunoblot analysis.

Generation of HEK293T knockout and HA-ELP1 cell lines. CRISPR short guide RNAs (sgRNAs) were designed using the Optimized CRISPR Design tool (<http://tools.genome-engineering.org>). The oligo pairs encoding the sgRNAs (Supplementary Table 1) were annealed, ligated into pSpCas9(BB)-2A-GFP (PX458) (Addgene plasmid # 48138; a gift from Feng Zhang) at BbsI site. Next, HEK293T WT cells were transiently transfected using calcium phosphate and seeded as single clones into a 96-well plate via several dilution. Mutations and loss of target protein expression was confirmed in the clonal KO cell lines by sequencing and immunoblot analysis, respectively. The presence of off-target mutations in KO cells was excluded by sequencing of top-candidate off-target sites predicted by Optimized CRISPR Design tool. For generation of HEK293T HA-ELP1 cells plasmid containing the sgRNA for ELP1 C-terminus was transfected together with the HA-ELP1 repair oligo containing BspEI specific cut-site (Supplementary Table 1) into HEK293T cells using nucleofector (Lonza), according to the manufacturer's protocol. The cells were tested for HA insertion by: PCR followed by specific BspEI digestion, sequencing, and immunoblot analysis.

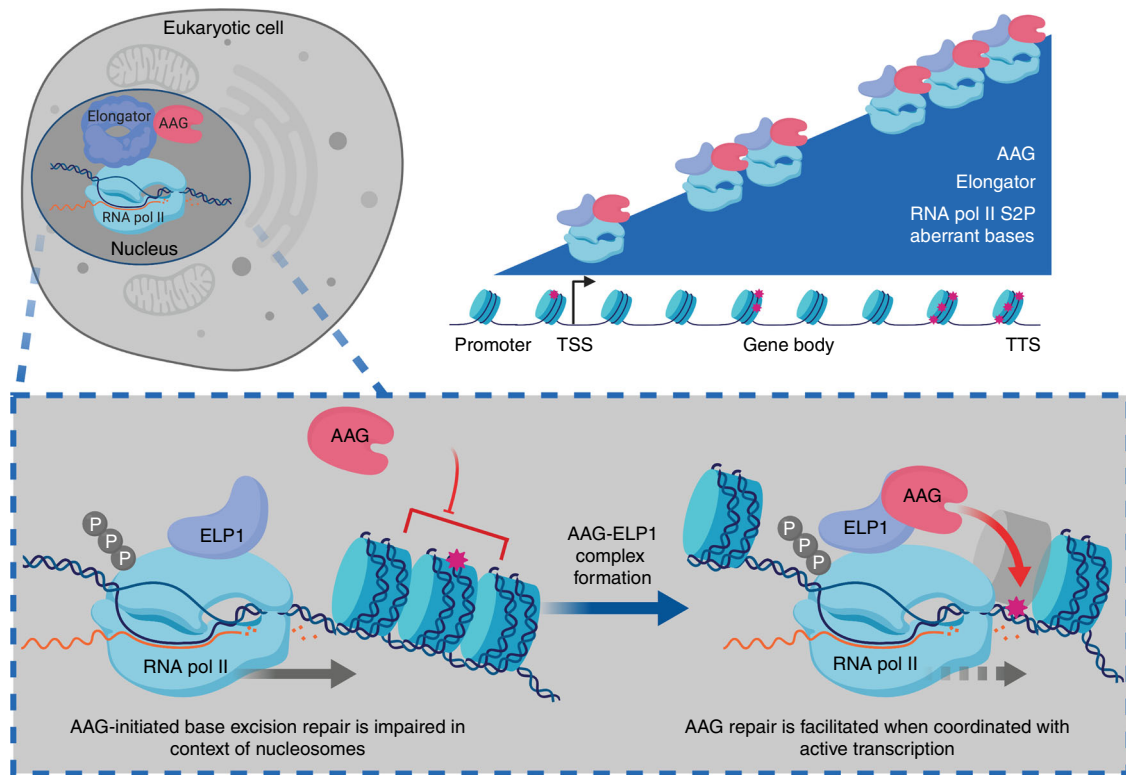


Fig. 7 Model of AAG-initiated DNA repair coordinated with gene expression. Elongator complex associates with RNA pol II to promote transcription elongation, accumulating towards 3' end of regulated genes. AAG through its unstructured N-terminal region associates with ELP1 subunit of Elongator, thus forming complex with active transcription machinery. During active transcription chromatin is locally decondensed, which allows AAG to efficiently initiate BER by recognizing and removing aberrant bases. AAG-initiated BER likely temporarily inhibits RNA pol II progression, thus resulting in reduced expression of co-regulated genes. In the absence of Elongator, transcription of target genes is repressed, while AAG recruitment to chromatin and initiation of BER is impaired. For more details see text. Schematic representation was created with Biorender.com.

RNA isolation, library preparation, and RNA sequencing. Total RNA was purified with RNeasy kit (QIAGEN), and DNaseI digested (QIAGEN) according to the manufacturer's protocol. RNA library preparation, and sequencing of HEK293T samples, were carried out by the Functional Genomic Centre, Zurich (FGCZ). The libraries were screened on the Bioanalyzer (Agilent), pooled in equimolar concentrations, and sequenced on one lane of a flow cell on an Illumina HiSeq2000 using single-read TruSeq v3 chemistry with 125-cycles. PolyA RNA sequencing of HAP1 was performed by the Beijing Genomics Institute (BGI). Briefly, polyA RNA was isolated and generated into single read, 50 bp libraries following standard Complete Genomics protocols.

Bioinformatic processing. FASTQ files of HEK293T samples were processed to remove adapter sequences and low quality reads prior to alignment to the GRCh37 Ensembl release 80 of the human genome using STAR v2.4.2a aligner⁵⁰. Reads within exon features were counted using featureCounts in Bioconductor (3.2) package Rsubread (1.20.6). Differential expression analysis was performed in R using the Bioconductor package DESeq2 (ref. 51). Pairwise comparisons were made between the HEK293T WT and the knockout cell-lines, AAG^{-/-} or ELP1^{-/-} samples. For the WT vs AAG^{-/-} comparisons, statistically significant differentially expressed genes were defined as having a ≥ 1.5 -fold change and $FDR \leq 0.1$. For the WT vs ELP1^{-/-} comparisons, statistically significant differentially expressed genes were defined as having a ≥ 2 -fold change and $FDR \leq 0.1$. DAVID v6.8³² gene-ontology (GO) enrichment analysis was performed to identify enriched Biological Processes (BP).

FASTQ files from HAP1 samples were processed to remove adapter sequences and low quality reads prior to alignment to the hg18 genome. Specifically, reads in which mismatched bases are more than 10% and low quality reads (the percentage of low quality bases is over 50% in a read) are removed using a custom BGI pipeline. Reads passing filtering were mapped to the hg18 reference transcriptome using Bowtie2 (ref. 52) and BWA⁵³ was used for mapping to the reference genome. Gene Quantification done using RSEM⁵⁴, which calculates FPKM. DEGs screened using Poisson Distribution Method, performed using custom BGI pipelines.

Whole cell extracts. HEK293T cells were trypsinized and washed twice with ice-cold PBS and the pellet flash frozen in liquid N₂. The cell pellets were next resuspended in 2 PCV of hypotonic lysis buffer (20 mM Hepes pH 7.9, 2 mM MgCl₂, 0.2 mM EGTA, 10% glycerol, 0.1 mM PMSF, 2 mM DTT, 1x Halt™ Protease

Inhibitor Cocktail), left on ice for 5 min and subjected to three freeze and thaw cycles. Subsequently, NaCl and NP-40 were added to final concentration 0.5 M and 0.5% (v/v) respectively, and samples incubated 20 min on ice. The samples were diluted by addition of 8 PCV hypotonic lysis buffer containing 50 mM NaCl and subjected to sonication. In case of nuclease treatment either 1.2 μ l CaCl₂ (2.5 M) with 25U of DNase, or 2.1 μ g RNaseI, and 50U MNase and samples incubated for 1 h, at 4 °C. Samples were centrifuged and the WCE supernatant collected.

Protein mass spectrometry. Using calcium phosphate precipitation empty pcDNA3.1-FLAG or pcDNA3.1-FLAG-AAG expression constructs (0.1 μ g ml⁻¹) were transiently transfected into HEK293T cells. Before harvesting, cells were mock or MMS (2 mM) treated for 1 h and WCEs prepared. Next, 1 mg WCE proteins was incubated with 20 μ l of FLAG M2 resin (Sigma) in buffer A1 (20 mM HEPES at pH 7.9, 2 mM MgCl₂, 0.2 mM EGTA, 10% (v/v) glycerol, 0.1 mM PMSF, 2 mM DTT, 140 mM NaCl, 0.01% (v/v) Nonidet-P40) for 2 h at 4 °C. The resin was washed four times using the same buffer, bound samples eluted in two steps with buffer A1 containing 100 μ g ml⁻¹ 3 \times FLAG peptide (Sigma), and subjected to trichloroacetic acid precipitation. Identification of proteins in FLAG-AAG mock or MMS treated, and control FLAG samples was performed by the Biopolymers and Proteomics Core Facility of the David H. Koch Institute for Integrative Cancer Research. Briefly, protein samples were reduced, alkylated, and digested with trypsin, followed by purification and desalting on an analytical C18 column tip. The processed peptides were then analyzed on an Agilent Model 1100 Nanoflow high-pressure liquid chromatography (HPLC) system coupled with electrospray ionization on a Thermo Electron Model LTQ Ion Trap mass spectrometer (MS). All MS/MS samples were analyzed using Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version SRF v. 3). Sequest was set up to search the merged_human_90T.fasta.hdr database assuming the digestion enzyme trypsin. Sequest was searched with a fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 2.0 Da. Dehydro of serine and threonine, oxidation of methionine, iodoacetamide derivative of cysteine and phosphorylation of serine, threonine, and tyrosine were specified in Sequest as variable modifications. Scaffold (version Scaffold_4.3.0, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted, if they could be established at $>95.0\%$ probability (as specified by the Peptide Prophet algorithm⁵⁵). Peptide identifications were also required to exceed specific database search engine thresholds. Sequest identifications required at least deltaCn scores of greater than

0.10 and XCorr scores of greater than 2.0, 2.5, 3.5, and 3.5 for singly, doubly, triply and quadruply charged peptides. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm⁵⁶. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. All proteins identified in FLAG-AAG and control Flag samples, of cells untreated or treated with MMS are indicated in Supplementary Data 2 and available in PRIDE under accession PXD013508.

Co-immunoprecipitation (CoIP) assays. WCEs (0.5 mg) were incubated with 2 µg of antibody (anti-AAG antibody;⁸ or anti-ELP1, Abcam ab56362) or IgG (rabbit IgG, Diagenode, C15410206; or mouse IgG, Diagenode, C15400001) at 4 °C overnight. In case of purified proteins (FLAG-ELP1, AAG or Δ80 AAG) 1 µg of protein was mixed with 1 µg of antibody and incubated overnight at 4 °C in 200 µl final volume of IP buffer (20 mM HEPES pH 7.9, 2 mM MgCl₂, 0.2 mM EGTA, 10% (v/v) glycerol, 0.1 mM PMSF, 2 mM DTT, 140 mM NaCl, 0.01% (v/v) Nonidet-P40). Protein-A Dynabeads, or Protein-A sepharose beads (in case of purified proteins), were equilibrated in the IP buffer at 4 °C overnight. After three consecutive washes the beads were added to samples and incubated for 4 h at 4 °C. The supernatant was removed and beads washed three times with cold wash buffer. Beads were boiled in Laemmli buffer and subjected to immunoblot analysis.

Immunoblot analysis. Samples were separated on 4–12% Bis-Tris polyacrylamide gel (Invitrogen) followed by transfer to Amersham™ Hybond™ P 0.2 PVDF (GE Healthcare) for immunoblotting. Primary antibodies: anti-anti-ELP1 (1:250, Abcam, ab56362), anti-ELP3 (1:1000, Abcam, ab96781), anti-AAG (1:1000, custom rabbit polyclonal antibody, Covance, raised against Δ80AAG) or anti-AAG (1:1000, LSBio LS-C133325), anti-RNA pol II S2P (1:1000, Abcam, ab5095), anti-HA (1:1000, Abcam, ab9110), anti-α-Tubulin (1:10000, Cell Signaling, 2144), anti-H3 (1:1000, Abcam, ab1791), anti-GFP (1:1000, Abcam, ab290); were detected using infrared (IR) Dye-conjugated secondary antibodies (1:15000, Li-COR Biosciences, 827-11081 and 925-32210). The signal was visualized by using direct IR fluorescence via the Odyssey Scanner, Li-COR Biosciences.

Column purification of cellular complexes. To isolate cellular complexes 1 ml, HeLa cell pellet was washed twice with 1 packed cell volume (PCV) of ice-cold PBS, and resuspended in 4PCV of hypotonic buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 5 mM DTT) and incubated at 0 °C for 20 min. After addition of 0.5 mM PMSF and 0.5 µg ml⁻¹ Leupeptin, Pepstatin, Chymostatin, cells were homogenized using a Douncer and 4PCV of ice-cold buffer 2 (50 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 2 mM DTT, 25% sucrose, 50% (v/v) glycerol) were slowly added while stirring. Next, 1 PCV of neutralized saturated ammonium sulfate solution was added during stirring for 30 min at 0 °C, followed by ultracentrifugation 3 h at 2 °C, 45'000 rpm. The supernatant was collected and dialyzed into buffer B (20 mM Tris Acetate pH 7.8, 20% (v/v) glycerol, 1 mM EDTA, 1 mM DTT, 0.01% (v/v) NP-40, 1 mM PMSF, 2 µM Pepstatin A, 2 µg ml⁻¹ Chymostatin, 0.6 µM Leupeptin, 2 mM Benzamide) with 50 mM KCl; 3 times for 1 h at 4 °C. Dialyzed samples were load onto heparin-sepharose column overnight. After washing with 10 column volumes of buffer B supplemented with 50 mM KCl, bound proteins were eluted as 1 ml fractions using buffer B with 150 mM, 300 mM, 450 mM, 700 mM and 1000 mM KCl, sequentially. The different elutions were subjected to immunoblot analysis.

Protein purification. Human full length AAG and Δ80AAG lacking first 80 N-terminal amino acids were expressed and purified by using the established expression system⁵⁷. ELP1 cDNA was excised from pCMV6-XL5-ELP1 (OriGene) by NotI (NEB) and inserted into pcDNA3.1(+)-3xFLAG vector generated as described in⁸. HEK293T cells were transfected by calcium phosphate, harvested after 48 h and WCE prepared as described above. 0.5 mg WCE were incubated with 75 µl anti-FLAG M2 affinity beads (Sigma Aldrich) for 2 h at 4 °C, the beads were washed three times with washing buffer (20 mM Hepes pH 7.9, 2 mM MgCl₂, 0.2 mM EGTA, 10% (v/v) glycerol, 0.1 mM PMSF, 2 mM DTT, 140 mM NaCl, 1x Halt™ Protease Inhibitor Cocktail). FLAG-ELP1 was eluted twice by addition of 80 µl Elution buffer (washing buffer with 0.5% (v/v) NP-40 and 0.15 µg µl⁻¹ 3x FLAG peptide (Sigma Aldrich)) for 30 min. The supernatant was applied on an Amicon Ultra-0.5 Centrifugal Filter Unit (Merck) and buffer exchanged to storage buffer (washing buffer with 0.1% (v/v) NP-40).

Gel shift experiments combined with chemical crosslinking. To evaluate affinity between AAG and ELP1 10 µM of purified proteins were mixed in a final volume of 50 µl reaction buffer (0.1 M sodium phosphate, 0.15 M NaCl, pH 7.4) and equilibrated 30 min at RT. Crosslinking agent BS³ (Thermo Fisher Scientific) was next added in different amounts: 0 (control), 0.1 mM and 0.5 mM and samples incubated 1 h at 4 °C, shaking. The reactions were quenched with 2.5 µl of 1 M Tris (pH 7.5) during 5 min at RT. As an additional control, samples containing single proteins alone were also evaluated in the presence of BS³. Samples were analyzed via immunoblot analysis with anti-AAG (1:1000, LSBio, LS-C133325), anti-ELP1 (1:1000, Genosphere custom antibody, raised against 1199-1218aa of ELP1 NP_0013052891) and appropriate secondary antibodies.

GFP-AAG and AAG-GFP constructs. The cDNA of AAG transcript variant 1 (NM_002434.4) was inserted into the pEGFP-C1 or pEGFP-N1 (Clontech) using BamHI, XbaI and NotI, XbaI respectively, to create pEGFP-C1-AAG and pEGFP-N1-AAG, which were validated by restriction enzyme digestion and sequencing. To generate AAG truncated forms: containing first 80 N-terminal AAG amino acids (1–80aa AAG) or AAG lacking 80 N-terminal amino acids (Δ80 AAG), pEGFP-C1-AAG plasmid was subjected to site directed mutagenesis using specific primer sets (Supplementary Table 2). Generated pEGFP-C1-(1–80aa AAG), pEGFP-N1-(1–80aa) and pEGFP-C1-(Δ80 AAG) were confirmed by sequencing.

GFP-Trap co-immunoprecipitation. In all, 0.3 × 10⁷ HEK293T cells were transfected by calcium phosphate, harvested after 48 h, resuspended in 200 µl ice-cold lysis buffer (10 mM Tris HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA 0.5% NP-40, 1 mM PMSF, 1X protease inhibitors) and incubated at 4 °C for 30 min. WCE was isolated as supernatant after centrifugation 20.000 × g for 10 min at 4 °C. 25 µl GFP-Trap™_MA (Chromotek) beads prewashed in wash buffer (10 mM Tris HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA) was added to 0.5 mg of WCE and the mixture was incubated for 2 h at 4 °C with constant rotation. The beads were washed three times with wash buffer. Proteins were eluted off the beads by boiling the beads in Laemmli sample buffer for 10 min and subjected to SDS-PAGE electrophoresis and immunoblot analysis.

Duolinkproximity ligation assay. To visualize close proximity between AAG and the ELP1, HEK293T cells were fixed in 4% paraformaldehyde for 15 min at RT, washed three times with PBS and permeabilized with PBS containing 0.1% Triton X-100 for 40 min at RT in Poly-D-Lysine and laminin double-coated plates. PLA was performed using Duolink™ detection kit according to manufacturer's instructions (Sigma-Aldrich) for analysis of dual PLA receptor recognition (AAG-ELP1 proximity) using AAG (1:80, LSBio LS-C133325) and ELP1 (1:50, Genosphere custom rabbit polyclonal antibody, raised against 1199-1218aa of ELP1 NP_0013052891) primary antibodies. Cells were washed, and the cell nuclei were counterstained using DAPI. Z-stacks were collected by Zeiss LSM 880 confocal microscope, with an Plan-Apochromat 63 × /1.4 oil DICM27 objective and ZEN v2.5 software. The images were obtained as a compilation of confocal Z-stacks comprising 15 optical splices (0.5 µm intervals) into 2D using maximum intensity projection using ImageJ.

Gene expression analysis. RNA was purified with RNeasy kit (QIAGEN) according to the manufacturer's protocol and reverse transcribed using MultiScribe Reverse Transcriptase (QIAGEN) according to the manufacturers protocol. qPCR was performed with Power SYBR Green PCR Master Mix (Applied Biosystems) on a StepOnePlus v2.3 Real-Time PCR System. Relative transcription levels were determined by normalizing to GAPDH mRNA levels. All primer sequences are listed in Supplementary Table 3.

Chromatin immunoprecipitation. HEK293T cells were crosslinked with 1% formaldehyde and quenched with 0.110 mM glycine. Cells were washed with ice-cold PBS, and harvested. Cell pellets were resuspended in cell lysis buffer (100 mM Tris-HCl pH 8, 10 mM DTT) and incubated 15 min on ice, and 15 min at 30 °C shacking. Next, nuclei were pelleted and washed with buffer A (10 mM EDTA pH 8, 10 mM EGTA, 10 mM HEPES pH 8, 0.25% Triton X-100), followed by buffer B (10 mM EDTA pH 8, 0.5 mM EGTA, 10 mM HEPES pH 8, 200 mM NaCl). Nuclei were lysed in lysis buffer (50 mM Tris-HCl pH 8, 10 mM EDTA and 1% SDS) and chromatin sheered to 200–250 bp DNA fragments by sonication with Bioruptor (Diagenode) for 30 cycles of 30 s. In all, 40 µg of chromatin was next precleared 2 h at 4 °C, and then incubated with 2 µg antibody in ChIP buffer (16.7 mM Tris-HCl pH 8, 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS, and 1.1% Triton X-100) overnight at 4 °C. ChIP antibodies included: anti-AAG (LSBio, LS-C133325-100); ant-HA (Abcam, ab9110); anti-APE1 (Abcam, ab194); anti-RNA pol II S2P (Abcam, ab5095); anti-H3 (Abcam, ab1791); anti-RNA pol II (MBL, MABI0601). The DNA-protein-antibody complexes were isolated using A/G dynabeads (ThermoFischer) and washed using sequentially: low salt wash buffer (16.7 mM Tris-HCl pH 8, 167 mM NaCl, 0.1% SDS, 1% Triton X), high salt wash buffer (16.7 mM Tris-HCl pH 8, 500 mM NaCl 0.1% SDS, 1% Triton X) and LiCl wash buffer (250 mM LiCl, 0.5% NP40, 0.5% Na-deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8). Proteinase K treatment was performed for 1 h at 50 °C with 10 mM EDTA, 40 mM Tris-HCl pH 6.5 and 20 µg proteinase K. The DNA was purified with phenol-chloroform, ethanol precipitated and analyzed by qPCR. Levels of ChIPed DNA is expressed as percentage of input, or relative abundance = (% input of specific gene region)/(% input of promoter region). Primer sequences are listed in the Supplementary Table 4.

Region-specific quantification of aberrant AAG substrates. The levels of aberrant bases substrates of AAG were determined in sheared HEK293T WT genomic DNA with an average fragment size of 200 bp. DNA was incubated with either a combination of 10U of AAG (NEB, M0313) and 10U of APE1 (NEB, M0282), or only with 10U of APE1 for 1 h at 37 °C in 1X ThermoPol Reaction Buffer (NEB, B9004). Next, level of DNA damage was assessed by qPCR (StepOne Software version v2.3) using primers targeting regions corresponding to promoter, middle and end of the gene of interest (Supplementary Table 4). Level of AAG

substrates in the targeted regions was inferred by calculating: $\Delta Ct = Ct(+AAG + APE1) - Ct(-AAG + APE1)$. The relative amount of AAG substrates for each gene region was calculated with respect to the promoter.

Flow cytometry host cell reactivation assay. Flow cytometry host cell reactivation assay (FM-HCR) assay was performed using plasmids for expression of the fluorescent proteins EGFP and mPlum subcloned into the pmaxCloning. EGFP plasmid served as control. The repair mPlum plasmid (mPlum.Hx) was engineered by placing the AAG-specific base lesion hypoxanthine (Hx) lesion into the open reading frame of mPlum. mPlum.Hx Substrate Containing a Site-Specific Hypoxanthine (Hx) was generated as follows; single-stranded DNA was obtained by nicking the plasmid pmax:mPlum with Nb.BtsI (NEB R0707S) for 4 h at 37 °C in NEB Buffer 4 (NEB B7004). After phenol chloroform extraction the DNA was incubated with 5 U of ExoIII (NEB M0206S) per μg of DNA at 37 °C for 1 h in NEB buffer 1 (NEB B7001) to digest the nicked strand. 25 μg of the resulting single-strand DNA (ssDNA) were combined with 0.5 μM phosphorylated Hx-containing oligonucleotide (5'-CACG-TAGGCCTTGGXGCCGTACATGATCTG-3', where X = Hx) in annealing buffer (1 mM Tris pH 7.5, 10 mM MgCl_2 , 250 mM NaCl). The mixture was heated to 95 °C in a thermal cycler for 10 min, and then allowed to anneal by cooling to 27 °C. After phenol chloroform extraction, 25 μg of the annealing product was combined with 30U of T7 polymerase (NEB 0274 L) and 0.8 mM dNTPs, and incubated for 30 min at 37 °C to allow primer extension. The resulting product was then phenol chloroform extracted. In all, 10 μg of the extended nicked dsDNA was incubated with T4 DNA ligase buffer supplemented with 0.8 mM dNTPs, 0.8 mM ATP, 3U T4 DNA polymerase (NEB M0203S), and 400U T4 DNA ligase (NEB M0202S) for 10 min at 37 °C and 120 min at 16 °C for a total of three cycles to yield closed circular plasmid. Finally, the product was purified via gel extraction. In case of inefficient repair, Hx is maintained in the plasmid and due to transcriptional mutagenesis at the site of lesion results in the expression of a non-fluorescent reporter protein. Fluorescent mPlum is generated only upon efficient BER and removal of the base lesion. The mPlum fluorescence is thus proportional to the level of AAG-initiated BER. Transfected cells were analyzed for fluorescence on a BD LSR II cytometer (BD biosciences), the cell debris, doublets, and aggregates were excluded based on their side-scatter and forward-scatter properties using BD FACSDiva v6.1. To exclude dead cells Zombie NIR (BioLegend) dye was added to cells 15–30 min before the analysis. The following fluorophores and corresponding detectors (in parentheses) were used: EGFP (Alexa Fluor 488), mPlum (PE-Cy5-5), and Zombie NIR (Alexa Fluor 700). Results were computed using FlowJo™ v10.6.1 and Percent Fluorescent Reporter Expression %R.E. calculated after gating (Supplementary Fig. 7c–f). Fluorescent signal was (F) was calculated using Eq. 1:

$$F = \frac{(N \times \text{MFI})}{S}$$

Where N is the total number of live cells, MFI is the mean fluorescence intensity of the N cells, and S is the total number of the live cells. The fluorescence signal from the reporter EGFP expressed from the undamaged plasmid, included in all transfections as transfection efficiency control, was designated as F^E . The normalized fluorescence signal for EGFP reporter F^O was calculated using Eq. 2:

$$F^O = \frac{F}{F^E}$$

Normalized reporter mPlum expression from the damaged reporter plasmid bearing Hx F^O_{dam} , and that from the same reporter plasmid in the absence of damage F^O_{un} , were used to compute the percent reporter expression (%R.E.) using Eq. 3:

$$\%R.E. = \left(\frac{F^O_{\text{dam}}}{F^O_{\text{un}}} \right) \times 100$$

Comet FLARE. The level of AAG-specific substrates in actively transcribing and inhibited cells was assessed by single-cell AAG-Comet FLARE (Fragment Length Analysis using Repair Enzymes) assay. Briefly, transcriptional stalling was induced by treatment of HEK293T WT cells with 150 μM DRB for 4 h at 37 °C, DMSO served as control. Next cells were collected and mixed with 1% of low melting point agarose, spread onto glass slides (Trevigen) and lysed overnight in lysis buffer (2.5 M NaCl, 100 mM EDTA, 10 mM Tris, 1% Triton, 10% DMSO, 1% sodium lauryl sarcosinate). Next, detergent was removed by washing and the slides were incubated for 1 h at 37 °C in the presence of 2U of AAG (NEB, M0313) in reaction buffer (2.5 mM HEPES-KOH pH 7.4, 100 mM KCl, 25 mM EDTA, 2 mM BSA). The DNA was subjected to alkaline electrophoresis (1 mM EDTA, 200 mM NaOH, pH > 13) and labeled with SYBR® Green (Sigma Aldrich). In total, four replicates from two independent experiments were performed, and on average one hundred cells (fifty cells per replicate) analyzed per experiment. Slides were imaged on a Zeiss Axiovert 200 M epifluorescent microscope using a $\times 20$ objective and the percentage of the DNA in the tail was evaluated for each data point with Comet IV, v4.2 (Perceptive Instruments, Suffolk, U.K.).

Quantification and statistical analysis. Immunoblots were quantified by GelEval v1.35 scientific imaging software (FrogDance Software, UK). Analysis of data was performed using GraphPad Prism v5 (GraphPad Software, Inc., La Jolla, CA). Statistical significance was determined primarily by one-way ANOVA with follow-

up Dunnett's multiple comparison test (Figs. 3g–j, 4b–o, and 5i, j) and two-tailed unpaired Student's t -test. Comet FLARE was analyzed using one-way ANOVA and Bonferroni post hoc test. All data represent mean values \pm SEM. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$ vs. control; NS, not significant.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

A reporting summary is available as a Supplementary Information file. The source data underlying Figs. 1a–d, 2a–f, 3a, g–j, 4, 5, 6, and Supplementary Figs. 2a, b, g, 4, 5b, c, 6, 7a, b, and 8 are provided as a Source Data file. The RNA sequencing data reported in this paper are available in GEO under accession [GSE129009](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129009) for HEK293T cells and [GSE129010](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129010) for HAP1 cells. The proteomics data associated with Supplementary Data 2 are available in PRIDE under accession [PX013508](https://www.ebi.ac.uk/pride/archive/projects/PXD013508). All data is available from the corresponding authors upon reasonable request.

Received: 25 April 2019; Accepted: 5 November 2019;

Published online: 29 November 2019

References

- Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
- Friedberg, E. C. et al. *DNA Repair And Mutagenesis* (ASM Press, 2006).
- Fu, D., Calvo, J. A. & Samson, L. D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* **12**, 104–120 (2012).
- Bosshard, M., Markkanen, E. & van Loon, B. Base excision repair in physiology and pathology of the central nervous system. *Int. J. Mol. Sci.* **13**, 16172–16222 (2012).
- Wallace, S. S. Base excision repair: a critical player in many games. *DNA Repair* **19**, 14–26 (2014).
- Krokan, H. E. & Bjoras, M. Base excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012583 (2013).
- Jacobs, A. L. & Schar, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1–20 (2011).
- van Loon, B. & Samson, L. D. Alkyladenine DNA glycosylase (AAG) localizes to mitochondria and interacts with mitochondrial single-stranded binding protein (mtSSB). *DNA Repair* **12**, 177–187 (2013).
- Olmon, E. D. & Delaney, S. Differential ability of five DNA glycosylases to recognize and repair damage on nucleosomal DNA. *ACS Chem. Biol.* **12**, 692–701 (2017).
- Odell, I. D. et al. Nucleosome disruption by DNA ligase III-XRCC1 promotes efficient base excision repair. *Mol. Cell Biol.* **31**, 4623–4632 (2011).
- Rodriguez, Y. & Smerdon, M. J. The structural location of DNA lesions in nucleosome core particles determines accessibility by base excision repair enzymes. *J. Biol. Chem.* **288**, 13863–13875 (2013).
- Rodriguez, Y., Howard, M. J., Cuneo, M. J., Prasad, R. & Wilson, S. H. Unnumbered Pol beta lyase activity in nucleosome core particles. *Nucleic Acids Res.* **45**, 8901–8915 (2017).
- Odell, I. D., Wallace, S. S. & Pederson, D. S. Rules of engagement for base excision repair in chromatin. *J. Cell Physiol.* **128**, 258–266 (2013).
- Hegde, M. L. et al. Pre-replicative repair of oxidized bases in the human genome is mediated by NEIL1 DNA glycosylase together with replication proteins. *Proc. Natl Acad. Sci. USA* **110**, E3090–E3099 (2013).
- Bj Ras, K. O. et al. Monitoring of the spatial and temporal dynamics of BER/SSBR pathway proteins, including MYH, UNG2, MPG, NTH1 and NEIL1-3, during DNA replication. *Nucleic Acids Res.* **45**, 8291–8301 (2017).
- Reis, A. M. et al. Targeted detection of in vivo endogenous DNA base damage reveals preferential base excision repair in the transcribed strand. *Nucleic Acids Res.* **40**, 206–219 (2012).
- Izumi, T. et al. Mammalian DNA base excision repair proteins: their interactions and role in repair of oxidative DNA damage. *Toxicology* **193**, 43–65 (2003).
- Guo, J., Hanawalt, P. C. & Spivak, G. Comet-FISH with strand-specific probes reveals transcription-coupled repair of 8-oxoGuanine in human cells. *Nucleic Acids Res.* **41**, 7700–7712 (2013).
- Watanabe, S. et al. Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin. *Proc. Natl Acad. Sci. USA* **100**, 12859–12864 (2003).
- Likhite, V. S., Cass, E. L., Anderson, S. D., Yates, J. R. & Nardulli, A. M. Interaction of estrogen receptor alpha with 3-methyladenine DNA glycosylase modulates transcription and DNA repair. *J. Biol. Chem.* **279**, 16875–16882 (2004).
- Bjorge, M. D. et al. Synergistic actions of Ogg1 and Mutyh DNA glycosylases modulate anxiety-like behavior in mice. *Cell Rep.* **13**, 2671–2678 (2015).

22. Huttlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).
23. Hawkes, N. A. et al. Purification and characterization of the human elongator complex. *J. Biol. Chem.* **277**, 3047–3052 (2002).
24. Kim, J. H., Lane, W. S. & Reinberg, D. Human elongator facilitates RNA polymerase II transcription through chromatin. *Proc. Natl Acad. Sci. USA* **99**, 1241–1246 (2002).
25. Glatt, S. et al. The Elongator subcomplex Elp456 is a hexameric RecA-like ATPase. *Nat. Struct. Mol. Biol.* **19**, 314–320 (2012).
26. Lin, Z. et al. Crystal structure of elongator subcomplex Elp4-6. *J. Biol. Chem.* **287**, 21501–21508 (2012).
27. Frohloff, F., Jablonowski, D., Fichtner, L. & Schaffrath, R. Subunit communications crucial for the functional integrity of the yeast RNA polymerase II elongator (gamma-toxin target (TOT)) complex. *J. Biol. Chem.* **278**, 956–961 (2003).
28. Winkler, G. S., Kristjuhan, A., Erdjument-Bromage, H., Tempst, P. & Svejstrup, J. Q. Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo. *Proc. Natl Acad. Sci. USA* **99**, 3517–3522 (2002).
29. Wittschleben, B. O., Fellows, J., Du, W., Stillman, D. J. & Svejstrup, J. Q. Overlapping roles for the histone acetyltransferase activities of SAGA and elongator in vivo. *EMBO J.* **19**, 3060–3068 (2000).
30. Close, P. et al. Transcription impairment and cell migration defects in elongator-depleted cells: implication for familial dysautonomia. *Mol. Cell* **22**, 521–531 (2006).
31. Li, Q. et al. The elongator complex interacts with PCNA and modulates transcriptional silencing and sensitivity to DNA damage agents. *PLoS Genet.* **5**, e1000684 (2009).
32. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
33. Xu, H. et al. Dimerization of elongator protein 1 is essential for Elongator complex assembly. *Proc. Natl Acad. Sci. USA* **112**, 10697–10702 (2015).
34. Dauden, M. I. et al. Architecture of the yeast Elongator complex. *EMBO Rep.* **18**, 264–279 (2017).
35. Wang, W., Scheffler, K., Esbensen, Y. & Eide, L. Quantification of DNA damage by real-time qPCR. *Methods Mol. Biol.* **1351**, 27–32 (2016).
36. Nagel, Z. D. et al. Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis. *Proc. Natl Acad. Sci. USA* **111**, E1823–E1832 (2014).
37. O'Connor, T. R. Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic Acids Res.* **21**, 5561–5569 (1993).
38. O'Brien, P. J. & Ellenberger, T. Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines. *Biochemistry* **42**, 12418–12429 (2003).
39. Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D. & Handa, H. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO J.* **17**, 7395–7403 (1998).
40. Yamaguchi, Y. et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51 (1999).
41. Menoni, H. et al. ATP-dependent chromatin remodeling is required for base excision repair in conventional but not in variant H2A.Bbd nucleosomes. *Mol. Cell Biol.* **27**, 5949–5956 (2007).
42. Luhsdorf, B., Epe, B. & Khobta, A. Excision of uracil from transcribed DNA negatively affects gene expression. *J. Biol. Chem.* **289**, 22008–22018 (2014).
43. Rosic, S. et al. Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nat. Genet.* **50**, 452–459 (2018).
44. Jimeno-Gonzalez, S. et al. Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing. *Proc. Natl Acad. Sci. USA* **112**, 14840–14845 (2015).
45. 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).
46. Menoni, H. et al. The transcription-coupled DNA repair-initiating protein CSB promotes XRCC1 recruitment to oxidative DNA damage. *Nucleic Acids Res.* **46**, 7747–7756 (2018).
47. Plosky, B. et al. Base excision repair and nucleotide excision repair contribute to the removal of N-methylpurines from active genes. *DNA Repair* **1**, 683–696 (2002).
48. Malvezzi, S. et al. Mechanism of RNA polymerase II stalling by DNA alkylation. *Proc. Natl Acad. Sci. USA* **114**, 12172–12177 (2017).
49. Yu, S. L., Lee, S. K., Johnson, R. E., Prakash, L. & Prakash, S. The stalling of transcription at abasic sites is highly mutagenic. *Mol. Cell Biol.* **23**, 382–388 (2003).
50. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
51. 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).
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
54. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
55. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392 (2002).
56. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).
57. Lingaraju, G. M., Davis, C. A., Setser, J. W., Samson, L. D. & Drennan, C. L. Structural basis for the inhibition of human alkyladenine DNA glycosylase (AAG) by 3,N4-ethenocytosine-containing DNA. *J. Biol. Chem.* **286**, 13205–13213 (2011).

Acknowledgements

This work has been supported by the Research Council of Norway grant (ID: 263152/F20) and Swiss National Science Foundation grant (ID: 31003A_152621) to B.v.L, N.P.M. and A.B.; Onsager Fellowship to B.v.L; NTNU to D.L.B., S.B., P.A.A., L.C.O., N.K., M.O., P.S. and M.B.; University of Zurich to M.R. and A.F. We thank Raffaella Santoro, Eva Vollenweider, and Sergio Leone for help with initial ChIP experiments and RNA-seq analysis; to Hilde Nilsen, Lisa Lirussi, and Alexander Chaim for suggestions and shearing protocols and reagents; to Cathrine Broberg Vågbo and PROMEC for analysis of base modifications; to Hans E. Krokan, Katja Scheffler and Cathrine Broberg Vågbo, for discussions and comments to the manuscript; and to FGCZ for RNA-sequencing.

Author contributions

B.v.L. and L.D.S. had the original idea. B.v.L. supervised the project and together with N. P.M. and M.B. interpreted the experiments. B.v.L., N.P.M., and D.L.B. wrote the manuscript and generated figures. N.P.M. performed most of the experiments. D.L.B. performed gel-shift experiments, qPCR analysis of AAG substrates and Comet-FLARE. A.B. contributed to PLA, ChIP and IP experiments. M.R. purified proteins and contributed to IP studies. S.B. contributed to qPCR analysis. K.Ø.B. and M.O. contributed with RNA sequencing of HAP1 cells. P.A.A. contributed to generation of HA-ELP1 cells. A.F. performed isolation of complexes. N.K. contributed to imaging. S.F.M., L.C.O., and P.S. contributed to bioinformatics analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-13394-w>.

Correspondence and requests for materials should be addressed to B.v.L.

Peer review information *Nature Communications* thanks Natalia Tretyakova, Bennett Van Houten and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019