Computational and Structural Biotechnology Journal 21 (2023) 614–629

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Histone H2Bub dynamics in the 5′ region of active genes are tightly linked to the UV-induced transcriptional response

Smaragda Fanourgakis [1], Anna-Chloe Synacheri [1], Matthieu D. Lavigne [2], Dimitris Konstantopoulos *, Maria Fousteri *

Institute for Fundamental Biomedical Research, BSRC 'Alexander Fleming', 16672 Vari, Attica, Greece

ARTICLE INFO

Article history:
Received 15 August 2022
Received in revised form 22 November 2022
Accepted 7 December 2022
Available online 13 December 2022

Keywords:
Transcription elongation
DNA damage response
SAGA DUB module
Genome-wide H2Bub distribution
Post-UV RNA polymerase II dynamics

ABSTRACT

The timing and location of writing and erasing of histone modifications determine gene expression programs and are tightly controlled processes. One such modification is the monoubiquitination of histone H2B (H2Bub), whose precise level during transcription elongation is dynamically regulated by the synergistic action of RNF20/40 ubiquitin-ligase and the de-ubiquitinase (DUB) of the ATXN7L3-containing DUB modules. Here, we characterize the dynamics of H2Bub in transcription and explore its role in perspective with the recently updated model of UV damage-induced transcription reorganization. Employing integrative analysis of genome-wide high-throughput approaches, transcription inhibitors and ATXN7L3-DUB knockdown cells, we find that H2Bub levels and patterns depend on intron-exon architecture both in steady state and upon UV. Importantly, our analysis reveals a widespread redistribution of this histone mark, rather than a uniform loss as previously suggested, which closely mirrors the post-UV dynamics of elongating RNA Polymerase II (RNAPII) at transcribed loci. The observed effects are due to a direct inter-dependence on RNAPII local concentration and speed, and we show that deficient ATXN7L3-mediated DUB activity leads to increased elongation rates in both non-irradiated and irradiated conditions. Our data and the implementation of a high-resolution computational framework reveal that the H2Bub pattern follows that of RNAPII, both in the ATXNL3 knockdown and in response to UV guaranteeing faithful elongation speed, especially in the context of the transcription-driven DNA damage response.

1. Introduction

Genotoxic stressors damage DNA and obstruct the proper progression of crucial cellular processes occurring on chromatin. In particular, faulty DNA repair or damage overload impair healthy gene expression and accurate DNA replication programs leading to severe consequences on cellular function and might promote disease pathogenesis [1–3]. To maintain genomic integrity and protect stressed cells from disturbed homeostasis and mutagenesis, cells activate DNA-damage responses (DDR) and repair mechanisms, such as the nucleotide excision repair (NER) pathway. NER operates via two sub-pathways, the global genome repair (GG-NER) and the transcription-coupled repair (TC-NER), to repair bulky helix-distorting DNA lesions induced by ultraviolet (UV) irradiation, intercalating agents, like chemotherapeutic drugs, and other mutagens (e.g. benzo[*a*]pyrene guanine adducts in cigarette smoke) [2,4–6].

In recent years, a number of labs have shed light on a multi-layered transcription-driven DDR mechanism that is activated in response to UV-C-induced genotoxic stress in human cells [7–13]). This mechanism involves a prompt and synchronous release of promoter-proximal paused (PPP) RNA polymerase II (RNAPII) elongation complexes at early steps following damage induction. The so-called transcription wave functions to maximize sensing and accelerate repair of DNA damages by TC-NER on the transcribed strand of all actively expressed genes and regulatory elements [8,13] resulting in limited mutagenesis throughout the whole transcribed regions regardless of gene activity or length as evidenced in various genotoxins exposed tissues [8].

Histone tails are subject to numerous reversible post-translational modifications (PTM), altering the chromatin environment and providing extra layers of regulation of cellular processes. PTMs such

* Corresponding authors.
E-mail addresses: konstantopoulos@fleming.gr (D. Konstantopoulos), fousteri@fleming.gr (M. Fousteri).
[1] These authors contributed equally.
[2] Present address: Institute of Molecular Biology & Biotechnology, FORTH, Heraklion, Crete, Greece.

https://doi.org/10.1016/j.csbj.2022.12.013

as acetylation, methylation, ubiquitination and their crosstalk, have been associated with the regulation of transcription at various stages including initiation, PPP and elongation. The role of histone H2B monoubiquitination (H2Bub) is unique in that both the ubiquitinated and the deubiquitinated states of H2B are equally important and robustly regulated during transcription [14–17]. Genome-wide studies have begun to provide insights into H2Bub distribution during transcription revealing their specific localization on active gene bodies [18–23] and point to an intriguing enrichment in the first intron [24]. Despite these advances, a debate still exists relating to the exact role of H2Bub in transcription, as well as the temporal and spatial sequence of events ruling deposition and removal of H2Bub during elongation (for review see [25]).

In mammals, H2Bub involves the addition of a single ubiquitin protein to the lysine residue K120 by the heterodimeric ubiquitin ligase complex, RNF20/40 [26,27]. H2Bub is reversible by virtue of a class of cysteine-type deubiquitinases (DUB) known as ubiquitin-specific proteases (USPs), which remove ubiquitin from target substrates [28]. USP22 is ubiquitously expressed in human tissues and is the most abundant and best-characterised DUB enzyme of H2Bub [29–32]. USP22 is part of the conserved SAGA (Spt-Ada-Gcn5 Acetyltransferase) co-activator complex that includes catalytically active histone acetyltransferase (HAT) and DUB modules. USP22, in association with the three adaptor proteins ATXN7, ATXN7L3 and ENY2, comprise the DUB module of SAGA, whose activity depends on a properly structured module within SAGA [33–36]. Interestingly, ATXN7L3 downregulation specifically and exclusively impairs the DUB module, without affecting the integrity of the rest of the SAGA complex [35,37,38]. In addition to USP22, USP27X and USP51, have been reported to form DUB modules with both ATXN7L3 and ENY2 and to deubiquitylate H2Bub in a SAGA-independent manner [39].

Chromatin remodelling has been shown to directly affect the efficiency of repair by either of the two NER sub-pathways during the recovery of UV irradiation [40–43]. Increased chromatin accessibility [13,44] at the transcription start sites (TSSs) of active regulatory and gene regions has been reported to occur in parallel with the maintenance of active histone marks (such as histone H3K27acetylation) and the lack of deposition of transcriptional silencing modifications (H3K27me3) at the same loci in response to UV-induced genotoxic stress. Interestingly, chromatin accessibility gain is correlated with the enhanced PPP release of RNAPII into productive elongation, the continuous initiation of RNAPII from all actively transcribed regions and the accelerated repair of transcription-blocking lesions in response to UV irradiation [13].

In this study, we aimed to examine further aspects of the transcription elongation-dependent epigenome, which may dynamically participate in the aforementioned damage-induced re-organization of transcription, and for which limited information is currently available. We were intrigued by the fact that, although ubiquitination of histone H2B requires and is firmly coupled to active elongation [23,24], the global levels of H2Bub have been reported to rapidly diminish in response to UV stress [45]. In light of this, we analysed a set of high throughput next-generation sequencing (NGS) approaches in human cells and demonstrated that H2Bub undergoes an extensive and dynamic redistribution upon UV treatment. Further looking into the location of H2Bub, we noticed that the maximum H2Bub levels (peak summits) were observed within 5 kb of genes TSS, most often in intron 1, and that repeated sharp drops in signals were observed at consecutive intron-exon junctions. Although this pattern is not changed upon stress, we highlight a uniform gain at 5′ proximal regions on essentially all active genes, which is accompanied by a loss from distal gene body regions in all cell types tested. This UV-induced co-transcriptional H2Bub reshuffling reflects the widespread PPP release of transcription elongation waves that was reported previously [7–13] and may justify the rapid loss of H2Bub detected in extracts from UV-irradiated cells. Interestingly, we found that the extent of gain in H2Bub was uniform across all active genes and that the length of the 5′ regions associated with gain in H2Bub was equal, independently of gene or intron 1 length, indicating that changes in local RNAPII density and speed have a strong effect on H2Bub levels in chromatin.

To address the post-UV role and dynamics of H2Bub de-ubiquitination, we used previously generated ATXN7L3 knockdown HeLa cells [35] and we compared them with HeLa control cells [23,35]. We took advantage of our setup involving chemical inhibition and controlled RNAPII release from PPP sites in non-treated and UV-irradiated cells in combination with ChIP-sequencing (ChIP-seq) experiments. We found a striking overlap in RNAPII and H2Bub profiles, suggesting that there is a physical requirement for RNAPII to maintain H2Bub levels. Next, we performed an integrative analysis of high-resolution ChIP-seq and nascent RNA (nRNA) sequencing in our cells deficient in DUB activity to monitor *de novo* RNAPII wave progression during transcription and in response to UV. Custom estimation by Hidden Markov Model (HMM) and noise reduction allowed accurate wavefront position estimation and provided novel insights into the positive role of H2Bub in the regulation of transcriptional elongation [24]. We also revealed that the elongation rate is regulated by ATXN7L3-containing DUBs, presumably indirectly by limiting H2Bub levels, especially in the beginning of transcribed regions where it is most needed, in both steady-state and UV-stress conditions.

2. Materials and methods

2.1. Cell culture of human cell lines

Experiments were conducted using human skin fibroblasts (VH10) [46], immortalised by the human telomerase reverse transcriptase (hTERT) method [47] and supplemented with G418 (25 μg/ml final, AppliChem) for selection. For the investigation of the role of SAGA DUB, HeLa cells (malignant human epithelial adenocarcinoma cells) [48] transfected and stably expressing short hairpin RNA against ATXN7L3 (shATXN7L3) and non-targeting control (shCTL) were used [35], kindly provided to us by the Tora Lab, IGBMC. Cells were cultured under standard conditions in Dulbecco's Modified Eagle Medium (DMEM, Thermo Scientific) supplemented with 10 % v/v fetal bovine serum (FBS, Thermo Scientific) and 1 % v/v penicillin-streptomycin (Thermo Scientific). Previously established protocols to synchronize cells in G1 were applied to limit cell-cycle heterogeneity and achieve steady-state levels of RNAPII, histone modifications, and nascent transcripts across the transcribed regions [8,13]. Briefly, serum-starvation (DMEM + 0.5 % FBS) for 72 h of cells at confluency was followed by a release in complete medium for 3 h, allowing for rapid recovery of steady-state levels of transcription and the examined factors to take place before treatment.

2.1.1. UV-C irradiation

The cells were exposed to UVC radiation (254 nm, TUV Lamp, Philips). A dose of 20 J/m^2 20 J/m^2 was applied unless otherwise stated. Cells were left to recover in a complete medium at 37 °C for indicated periods of time.

2.1.2. DRB treatment

5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB, Calbiochem) was added directly to the growth media at a final concentration of 10 μM. For preDRB-nRNA-seq, cells were pre-incubated with DRB (100 μM) for 3 h [8,49]. Cells were mock (NO UV) or UV-irradiated (20 J/m^2) and recovered in fresh complete medium for indicated time points. Nascent RNA was labelled with EU for 10 min before harvesting (described below). For DRB inhibition (DRB-ChIP-seq), cells were incubated with DRB (100 μM) 10 min before UV irradiation (20 J/m^2).

After irradiation, cells were allowed to recover for 30 min in the presence of DRB before crosslinking and ChIP.

2.2. Acetic histone extraction

Cells were placed on ice and washed twice with cold PBS, collected in PBS 1 × solution containing 1 mM EDTA, 0.5 mM EGTA and 1 mM PMSF followed by centrifugation. The cell pellet was washed in PBS resuspended in 10 volumes of Lysis Buffer (10 mM HEPES PH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 M Dithiothreitol (DTT), 1.5 mM PMSF). Sulfuric acid was added to a final concentration of 0.2 M for 30 min, and the supernatant was collected and trichloroacetic acid (TCA) was added at a final concentration of 20 %. Samples were incubated for one hour on ice and 1 ml of cold acetone was added to the pellet. The pellet was thoroughly dried and resuspended in a suitable volume of TE solution (10 mM Tris, 1 mM EDTA).

2.3. ChIP for protein analysis (Western Blot)

Crosslinked cell pellets of $2 * 10^7$ cells were lysed and chromatin was immunoprecipitated as described previously [8]. Briefly, following lysis, fragmentation of chromatin was performed by sonication on the Bioruptor Sonicator apparatus (Diagenode) in an ice bath. Immunoprecipitation was performed overnight at 4 °C, by incubating equal amounts of chromatin per condition with the desired antibody. Samples were then rotated with the appropriate magnetic (Dynabeads Protein A or G for Immunoprecipitation, Thermo Scientific) or agarose (Protein A or G Agarose Fast Flow, Millipore) beads, for 3 h at 4 °C. An equal volume of 2 × Loading Buffer (100 mM Tris PH 6.8, 4 % (w/v) SDS, 0.2 % (w/v) bromophenol blue, 20 %(v/v) glycerol, 200 mM DTT) was added to the samples, which were then incubated at 95 °C for 50 min for reversal of crosslinking, and denaturation of proteins.

2.4. Polyacrylamide Gel Electrophoresis (PAGE) for Western Blot

For Western Blot analysis, either homemade PAGE gels were prepared (12 %, for histone extracts 8 % for everything else) or precast gradient gels (gradient gel, NuPAGE 4–12 % Bis-Tris Protein Gels, Thermo Fisher). Wet electrotransfer of separated proteins from the gel onto a Polyvinylidene Fluoride (PVDF) membrane (Millipore, IPFL00010), was carried out at 4 °C overnight. Membranes were blocked in a 50 % Li-cor Odyssey Blocking Buffer solution and incubated overnight at 4 °C with the primary antibodies. Wash steps were performed according to the Li-cor Odyssey Clx protocol before and after incubation with the appropriate secondary antibodies (Li-cor Odyssey IRDye, Cat. #926). Membranes were developed digitally, on the Odyssey CLx Imaging system, and analysed using the Image Studio software (Li-cor). Antibodies used for blotting: RNAPII CTD (phospho-S2) (ab5095, Abcam), H2Bub (05-1312-I clone 56, Millipore), H4 (ab10158), ATXN7L3 (PA5-80365, Invitrogen), β-tubulin (SC-58884, Santa Cruz).

2.5. ChIP for DNA analysis (qPCR & Next Generation Sequencing (NGS))

ChIP was performed, as described above. The immunoprecipitated material was eluted from the beads by two 15-min incubations in Elution Buffer (100 mM $NaHCO_3$, 1 % SDS) at 65 °C. The supernatant was collected after each incubation accordingly, and 200 mM NaCl was added. Crosslinks were reversed overnight at 65 °C, followed by Proteinase K (Invitrogen, Cat # 25530-015) treatment (0.1 μg/μl) in 0.5 % SDS. DNA was purified by Agencourt AMPure XP paramagnetic beads (Beckman Coulter Life Sciences, Cat no. A63881) for input samples, or by overnight ethanol precipitation at − 80 °C followed by purification by spin columns (MinElute Reaction Cleanup Kit, QIAGEN, Cat no. 28204) for ChIP DNA, according to manufacturer's protocol. Quality control of ChIP DNA was performed as described [8].

Antibodies used for ChIP: RNAPII CTD (phospho-S2) (ab5095, Abcam) and ubiquitinated H2B (NRO3) (MM-0029, Medimabs).

2.5.1. qPCR

qPCR analyses were performed with 100–200 pg of ChIP and Input DNA in duplicate reactions with qPCRBIO SyGreen mix (PCR Biosystems) on a Roche Light Cycler 96 instrument. Analysis of qPCR was done using the δδCt method between positive and negative loci normalised with inputs. Primers used are listed in Table 1.

2.6. Total RNA extraction (for KD check)

Briefly, cells were grown to a confluency of about 80 %. After removal of the medium, cells were harvested with TRIzol reagent (Invitrogen #15596026) followed by chloroform (Applichem #A1935 chloroform-isoamyl 24: 1) phase separation and RNA precipitated with ethanol overnight. Nucleic acids were quantified on a NanoDrop spectrophotometer ND-1000 (ThermoFisher). RNA was treated with DNase I, according to the manufacturer's instructions (Turbo DNase, Ambion, Life Technologies AM2238) and RNA was purified by acetic phenol extraction (Acid-Phenol:Chloroform pH 4.5 with IAA, 125:24:1 ThermoFisher Scientific #AM9722) and chloroform phase separation (chloroform-isoamyl 24: 1, Applichem A1935) followed by overnight ethanol precipitation. RNA was quantified by NanoDrop and stored at − 80 °C.

2.7. Nascent RNA (nRNA) extraction

Nascent RNA was labelled by the addition of 5-ethynyl uridine (EU, 0.1 mM) 10 min before the extraction of total RNA with Trizol (as above). After DNAse I treatment, only EU-labelled molecules were isolated using the Click-iT™ package (Nascent RNA Capture Kit, C10365, ThermoFisher Scientific) according to the manufacturer's protocol. rRNA was depleted using the Ribo-Zero Magnetic kit (Epicentre, #MRZH11124) as per the manufacturer's instructions. For cDNA synthesis, the SuperScript™ II Reverse Transcriptase Kit (Cat # 18064014) was employed. For mRNA, the manufacturer protocol was followed. For cDNA synthesis of nRNA, a slightly modified protocol was employed whereby nRNA was fragmented and synthesis was performed on the streptavidin beads. Eluted DNA was purified with AMPure XP beads and measured on Qubit 2.0 before library preparation.

2.8. Library preparation for NGS

Double-stranded DNA fragments derived from either ChIP (including input) or RNA isolation (cDNA) were modified to be compatible with NGS. Library preparation was performed according to standard Illumina protocols with minor modifications. Briefly, dsDNA was end-repaired, 3′-end adenylated and ligated to barcoded Truseq (synthesised by IDT) or NEBNext adaptors (NEB) according to the manufacturer's guidelines. Samples were PCR amplified by ensuring reduced amplification bias (optimal cycle number was determined by qPCR using the Illumina P5 and P7 primers) and size selected (200–500 bp) through double-selection by AMPure XP beads. Libraries were assayed on a 2100 BioAnalyzer instrument using the High Sensitivity DNA kit (Agilent Technologies 5067-4626) and multiplexed before sequencing, adjusting for concentration and fragment size of each sample, ensuring equal representation in the multiplexed libraries. Next Generation Sequencing was performed at Genecore-EMBL, using the Illumina HiSeq 2000 or NextSeq 5000 platforms for single-end 50 bp and 75 bp reads, respectively.

2.9. Bioinformatics analysis

2.9.1. Next Generation Sequencing (NGS) data preprocessing and alignment

DNase-seq (ENCSR000EJT) and H3K27ac (ENCSR000AOC), H3K4me1 (ENCSR000APW), H3K27me3 (ENCSR000DTY) Chip-seq processed data (alignments and peak sets) for HeLa S3 were obtained from the Encyclopedia of DNA Elements (ENCODE).

Quality control, preprocessing, mapping against the UCSC hg19 human reference genome, and alignment filtering were performed as previously described in [8] with minor modifications. Biological replicates were examined regarding their genomic bins similarity with Pearson's R correlation metric and PCA with batch effect correction using Limma [50] R library and were consequently downsampled to the same alignment depth and merged using samtools for visualization and downstream analysis.

2.9.2. Gene and enhancer transcriptional activity classification

RefSeq Gene and FANTOM5 enhancer activity for VH10 cells was assessed as previously described in [13] using ChIP-seq data from the same study. For HeLa cells, RefSeq hg19 release 86 annotation of all known protein-coding and non-coding RNA was retrieved from UCSC table browser (http://genome-euro.ucsc.edu/cgi-bin/hgTables). Genes whose transcripts are of different length were excluded from further analysis. HeLa cells' transcriptional activity was assessed separately for shCTL and shATXN7L3 regarding Active genes. Genes' TSSs were extended − 500 bp and + 2 kb and the RNAPII-ser2P noUV Reads Per million (RPM) were calculated for each of these regions. RPMs were visualized as a kernel density plot, forming a bimodal distribution, and a threshold was arbitrarily set to separate genes with sufficient RNAPII-ser2P noUV binding and mark them as Active (Fig. S1A). The bimodal distribution was calculated and visualized using Python scripts with sklearn's [51] GaussianMixture and matplotlib's (Hunter, 2007) pyplot packages. The intersection of shCTL and shATXN7L3 Active genes (n = 4532) sets was extracted and used as common active genes for visualization and downstream analysis. Common genes with insufficient RNAPII-ser2P noUV binding were split into Repressed and Inactive, according to the following rules. Repressed genes must have an overlap with an H3K27me3 peak (ENCODE data) within their + /- 2 kb around TSS flanking region, while Inactive genes' ± 2 kb flanking region should not overlap with any H3K27me3 peaks.

Human enhancers were retrieved from phases 1 and 2 of the FANTOM5 project from http://fantom.gsc.riken.jp/5/datafiles/phase2.2/extra/Enhancers/human_permissive_enhancers_phase_1_and_2.bed.gz, and the center of each annotation was considered as the corresponding transcription start site (eTSS). Enhancers which overlap with any gene of the hg19 RefSeq annotation (including a − 250 bp from the TSS promoter region) were filtered out, to only keep intergenic elements. Furthermore, enhancers that do not overlap with each other within ± 500 bp around the eTSS region were extracted. On this filtered subsample of FANTOM5 enhancers, a flanking region of ± 100 bp around eTSS was generated, and transcriptional activity was assessed according to the following rules. Active enhancers must have their flanking region overlap with both an H3K27ac and a DNase peak, Repressed enhancers must have their flanking region overlap with an H3K27me3 peak while not overlapping with any H3K27ac or DNase peaks, and Inactive enhancers must not overlap with any H3K27ac, H3K27me3 or DNase peaks. Any enhancers with other combinations of peak overlapping were removed from downstream analysis.

2.9.3. Read count density analysis and visualization

Gene regions used in average profile and heatmap bin counts visualizations were cut to a fixed length and split into equally sized genomic bins using custom scripts. Counts on bins were calculated with pybedtools [52,53] intersect and normalized for region size and library depth using the Reads Per Kilobase of transcript per Million mapped reads (RPKM) formula using custom Python scripts and pandas [54] package. Average profiles were produced using custom Python scripts using matplotlib's pyplot, single-sample heatmaps were produced by SeqMiner [55] version 8.4, while log2FC heatmaps were produced by deeptools [56] or custom python scripts using matplotlib package. Boxplot represented data were likewise normalized (RPKM), log2 transformed using numpy [57] and visualized with seaborn [58] package. Statistical annotation over boxplots was added with the statannot package. UCSC Genome Browser tracks were produced from filtered samples, as mentioned above, and normalized for library density using bedtools and the UCSC toolkit [59].

2.9.4. Genomic distribution of H2Bub signal

2.9.4.1. Annotation. The hg19 canonical genome was categorized into Gene, Enhancer, and intergenic regions. Furthermore, Active, Inactive and Repressed genes were split into promoter (− 250 bp/ + 100 bp around TSS) and gene body region (TSS + 100 bp to TTS). Only intergenic enhancers were used (see above for more details). "Intergenic" region category was generated by subtracting all enhancers and genes (original annotations) from canonical chromosomal coordinates. Active region bps were subtracted from Inactive and Repressed, and Repressed bps were subtracted from Inactive. Promoter region bps were subtracted from gene body regions. Subtraction was applied by the pybedtools subtract command. Consequently, the final annotation consists of non-overlapping regions. Read centers were used to count densities on these annotations, in order to ensure that a read would fall on a single region. Read densities were normalized to the total genomic region size (e.g. Active promoters count was normalized to the total length of active promoter regions).

2.9.4.2. Visualization. Donut plots that summarize the read density distribution on final annotations were generated by custom Python scripts using pandas, matplotlib and numpy packages.

2.9.5. Genic distribution of H2Bub signal

2.9.5.1. Annotation. For this particular type of analysis, genes with multiple annotated alternative splicing regions, 3′ and 5′ UTRs were excluded to minimize the confounding effect of annotating overlapping regions multiple times. RefSeq exon coordinates were extracted from UCSC Table Browser and intron coordinates were computed based on the particular data, as the complementary genic elements. Gene regions were split into promoters (− 250 bp/+ 100 bp around TSS), introns, exons, 3′ and 5′ UTRs. To have a non-overlapping annotation, as mentioned above, 3′ and 5′ UTR bps were subtracted from the rest of the region categories, and promoter region bps were subtracted from intronic and exonic regions. Normalization was applied as mentioned above, and reads' centers were used likewise (see Section 2.9.4).

2.9.5.2. Visualization. As mentioned above for genomic donut plots (see Section 2.9.4).

2.9.6. No UV vs plus UV H2Bub distribution

Using Deeptools with genomic bins of 1 kb of the first 100 kb for genes larger than 100 kb, sorted on intron 1 size, a log2 FC heatmap was produced. This indicated that the redistribution of H2Bub signal upon UV is independent of intron 1 signal, within an 18 kb region away from TSS. To further evaluate whether there is no correlation between the two characteristics, a 5th-degree polynomial was fit on each gene using the log2 FC matrix produced by Deeptools "compute matrix". This determined the 1 kb bin in which the FC becomes negative for each gene. The scatter plot produced regarding this data

shows no correlation between intron 1 length and the location of the sign change.

2.9.7. Intron-Exon H2Bub density distribution

2.9.7.1. Annotation. For the particular type of analysis, Active genes with a single known transcript (based on UCSC Table Browser RefSeq data, as previously described) were selected. Moreover, to observe the gradient decrease of H2Bub density on successive introns, genes whose first 3 introns are smaller than 1 kb were discarded. To determine whether this decrease is relevant to exon sites, a flanking region +/− 2 kb around exon start sites was generated from a subset of genes whose first 6 introns are larger than 2 kb. These regions were split into genomic bins of 1 bp and averaged for each of the 5 first exons. Counts of samples on these genomic regions and bins were acquired using pybedtools intersect command and normalized for size and library depth (RPKM).

2.9.7.2. Visualization. To visualize the levels of H2Bub signal on the first 3 introns, log2 RPKM boxplots were utilized, while to illustrate the average gradient decrease around exons, lasagna plots were employed. Both graphs were generated by custom Python scripts using pandas, matplotlib and numpy packages.

2.10. Statistical analysis

Statistical comparisons between the first 3 introns' H2Bub signal level were assessed through 1000 random samplings of 100 data points with replacement. In each sampling, log2 normalized RPKM values of genes' introns were extracted and the significance of their difference was computed by t-test and 95% confidence intervals of mean log2 differences between introns of the same gene. A total t.test meta p-value was calculated as the ratio of the total non-significant t-tests (p-value ≥ 0.05) and total samplings. The particular operations were applied by custom R scripting, using boot [60,61] and t.test functions.

2.11. H2Bub accumulation characterization analysis

2.11.1. Annotation & Visualization

To study the relation between intron 1 and H2Bub peak accumulation, active genes with a unique transcript larger than 30 kb were split into 5 equally populated groups based on their inton1 size, and their first 30 kb were partitioned into 1 kb genomic bins. The groups contained 227 genes each and their median intron 1 lengths were 66,906 bps, 31,062 bps, 17,057 bps, 7772 bps and 23,99 bps, respectively. For each bin, no UV and 3 h UV samples' RPKMs were computed. Using seaborn's lmplot, a regression model was fit on each group's binned genes with an order parameter equal to 5, and the numpy's mean function as an estimator. As signal trends coincide up to 5 kb throughout intron 1 length groups, and the peak's width correlates with the median intron 1 size of each group, dynamic programming change point detection was applied on each group's average profile. Python's ruptures package [62]. was used for the particular analysis. Specifically, the binary change point detection algorithm with the least squared deviation cost function was applied with 1, 1, 2, 2, and 2 breakpoints for each group, respectively. Additionally, as the peak seems to be contained within intron 1, we looked into the correlation between the introns signal and total genes signal. RefSeq exons coordinates were extracted from the Table Browser and intron coordinates were computed based on that data as described above. As Pearson correlation showed a slightly higher correlation between the intron 1 signal and total gene signal, we applied sklearn's lasso regression model whose parameters were fine-tuned with sklearn's GridSearchCV and RepeatedKFold with neg_mean_absolute_error function. Introns and gene counts were normalized in the 0–1 interval before fitting. Barplots representing the coefficient of each intron were generated using matplotlib.

2.12. H2Bub distribution index analysis

2.12.1. Annotation

Active genes whose length is over 60 kb were used. These genes were split into two regions: TSS to + 30 kb (proximal region) and + 30 kb to + 60 kb (distal region); for each region category, read counts of H2Bub samples were computed and transformed to RPKM values. Then, the ratio of the proximal/distal regions was calculated and transformed through the log2 function with custom python scripts using pandas, numpy and pybedtools packages.

2.12.2. Visualization

Scatter plots were employed to visualize the between-samples correlation. Graphs were produced by custom python scripts using the matplotlib package.

2.13. RNAPII escape index analysis

2.13.1. Annotation

Active genes larger than 2 kb were selected and were filtered to only include non-overlapping instances of this region. Promoter (− 250 bp/+ 100 bp around TSS) and gene body (TSS + 101 bp to + 2 kb) regions were extracted for these genes, to compute RNAPII read counts. Counts were normalized (RPKM), and the ratio of gene body/promoter was calculated and transformed through the log2 function. This analysis was performed with custom python scripts, likewise for distribution index analysis.

2.13.2. Visualization

As previously mentioned in Section 2.12.

2.14. Pri-elongating RNAPII wave analysis

2.14.1. Annotation and visualization

The first 60 kb of active genes longer than 60 kb were used to observe pri-elongating RNAPII along with levels of H2Bub levels under the same conditions. Genes were binned to 600 bp genomic regions and RPKM normalized counts on each bin were calculated. Average profiles were produced to compare H2Bub and RNAPII patterns between shCTL and shATXN7L3 cells. Alongside lasagna plots (heatmap of the average of each bin) were employed, depicting the bin signal of each sample subtracted by the corresponding input bin signal. Analysis was performed using custom python scripts.

2.15. Pre-DRB nascent RNA-seq elongation wave progression estimation

The basic idea is to define a discrete two-state left-to-right (only forward state transitions are allowed) Hidden Markov Model [63] (HMM) to decode gene bin counts as transcriptionally active or inactive using the Viterbi algorithm. In order to do so, the following procedure was followed to define the emission probabilities of each of the two states of the model.

2.15.1. Annotation

Two gene length thresholds were used depending on the time point and condition shared between shCTL and shATXN7L3. To generate elongation-wave progression estimates on all time points (20 min, 30 min, 60 min) of the plus UV condition, as also for no UV 20 min and 30 min, the first 100 kbs of genes larger than 100 kb (n = 614) were split into 100 bp bins. For 60 min no UV time point, the first 230 kb of genes larger than 230 kb were binned into 230 bp regions. Counts on bins were extracted using the intersect command of pybedtools and the corresponding samples for each annotation.

2.15.2. Normalization and Interpolation

Genomic bin counts per gene were normalized to be in the 0–1 interval using sklearn's MinMaxScaler in order to minimize the effect of varying transcription levels. As samples preprocessing (see Section 2.15) results in only intronic reads, for bins that overlap with over 40 % of an exon site, the corresponding normalized counts were interpolated using all other bins that do not overlap with exonic sites. Linear interpolation was performed with scipy's [64] interpolate function, and a moving average filter was applied to the results to smooth the signal (Fig. S1B).

2.15.3. Change point detection

Dynamic programming change point detection was applied to determine the location where the average signal equals zero. A change point detection model was fit on the average bin count profile of each sample, and depending on the time point and condition, 2 or 4 breakpoints were predicted. Based on the rightmost predicted change point, the last bin that corresponds to the elongation wave signal was inferred, and all the consecutive bins are considered "no-signal regions". For the particular analysis, Python's ruptures package was used. Results were visualized through rupture's show function (Fig. S1C).

2.15.4. States distribution

For each time point and condition, using previously inferred signal/no-signal bins, the emission probabilities of the two HMM states were computed. Bins contain discrete integer values ranging from 1 to 100, therefore for each state, 100 probabilities (one probability per possible value) are estimated based on the frequency of the value in the signal/no signal region correspondingly (Fig. S1D).

2.16. Hidden Markov Model and sequence decoding

An HMM is defined for each time point and condition using hmmlearn's MultinomialHMM.

Emission probabilities were assigned with the values computed as described in Section 2.15.4. Initial probabilities were set to 0.1 for the No-Signal state, and 0.9 for the signal state as we used active genes that are mostly expected to have an active elongation signal. Transition probabilities were set as follows for all samples but 60 min no UV: Signal to Signal 0.85, no Signal to Signal 0, and the rest as complements. For 60 min no UV samples, they were set: Signal to Signal 0.99, no Signal to Signal 0 as the signal tends to have some gaps for large genes (Fig. S1E, F).

After decoding, genomic bins that were mapped to "signal" state, were converted to kbs of the respective elongation wave progression for each time point and condition, by multiplying the number of bins with the genomic region they represent.

2.16.1. Filtering and visualization

Estimates of 0 kbs are discarded and a two-component gaussian mixture model is fit on the remaining estimations. It is expected that the component with the higher mean is the one containing the best estimates. Therefore, estimates that fall in a region ± 1.5 standard deviation from the selected component's mean value are accepted (Fig. S1F). The accepted estimates are visualized with boxplots of wave progression (kbs), and as kb bars over single genes plots.

2.16.2. Statistical analysis

The difference between kb distributions of estimates of the same time point between cell lines was evaluated with the two-sample, two-sided Kolmogorov-Smirnov test using scipy's ks_2samp function. Estimations distributions were fit with a gaussian kernel using sklearn's KernelDensity with a bandwidth of 2 and a sample of 150 points from each kernel was fed to the test.

The whole analysis was performed using custom python scripts with pandas, matplotlib, numpy, pybedtools, ruptures, hmmlearn, scipy, seaborn and sklearn packages.

2.17. H2Bub density on enhancers

2.17.1. High H2Bub enhancers group

A flanking region ± 1 kb around eTSS (as defined above) was selected for all intergenic non-overlapping enhancers from the FANTOM5 annotation. Counts of the shCTL and shATXN7L3 no UV samples were computed for these regions and enhancers were sorted based on these values and split into equally sized groups containing 2468 enhancers each. The intersection of the groups containing the highest levels of H2Bub density between the two cell lines was considered the high H2Bub enhancer group and was further studied regarding the activity of its components. The High H2bub group, with a total of 1531 enhancers, consists of 481 active, 1046 inactive and 4 repressed enhancers.

2.17.2. Enhancers' H2Bub level and chromatin state

As observed in the high H2Bub group, chromatin activity seems to be more intense as compared to the complete active enhancer set average profile. Subsequently, active intergenic non-overlapping enhancers were sorted in ascending order with respect to their shCTL no UV H2Bub sample count (± 1 kb around eTSS) and split into 3 equally populated groups containing 868 enhancers each (low, mid, high). To evaluate the relation between the increase of H2Bub density and other (RNAPII, nascentRNA, H3K27ac, DNase I) assays' signal levels, corresponding enhancer counts were converted to categorical variables according to their sorting and splitting in equally sized groups. A custom python script for statistical analysis implementing chi-square followed by a post hoc test with Bonferroni correction was applied using scipy's chi2_contingency and pandas.

2.18. H2Bub levels and Enhancer-Gene associations

Using human enhancer–gene pairs that were extracted from the correlation of CAGE signal at enhancers and CAGE-derived TSSs (http://enhancer.binf.ku.dk/presets/human.associations.hdr.txt.gz), a subgroup of active enhancers that have known gene associations was generated. This subgroup was divided into three equally sized shCTL H2Bub count ordered groups, likewise above (low, mid, high) containing 202 enhancers each. Three gene groups were created, containing the genes associated with the enhancers of each corresponding enhancer group. Although enhancer groups were equally populated, their corresponding gene groups follow an incremental trend according to the levels of H2Bub in the respective enhancers' groups. Therefore, unique enhancer-gene associations were grouped by enhancer id, and the number of gene associations per enhancer was calculated.

3. Results

3.1. Gene structure and transcriptional stress determine distributions of H2Bub

To characterize histone PTMs that are dynamically coupled to the UV-induced reorganization of transcription elongation, we initially analysed bulk histone levels from extracts of UV-irradiated cells. We used human skin fibroblasts, which we synchronised by serum starvation and left to recover in normal serum-containing medium for three hours (h) (Section 2; [8]), before UV-C irradiation (20 J/m^2). Western Blot (WB) analysis of histones PTMs levels at the indicated recovery time intervals between 0.5 h and 48 h confirmed the previously suggested transient, rapid and significant reduction of H2Bub in the early and middle phase of recovery, (Fig. S2A–B) [45]. In

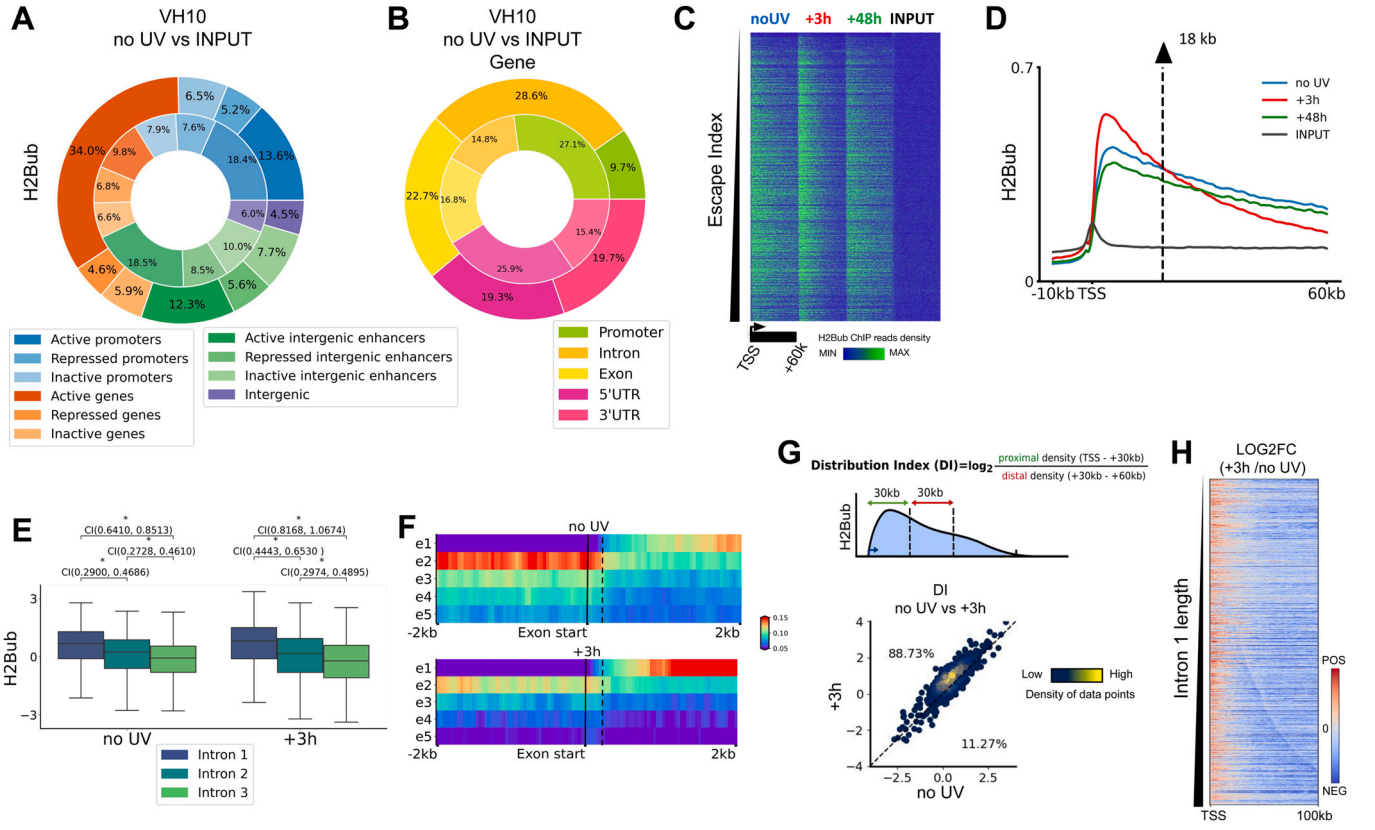

Fig. 1. H2Bub is globally redistributed along actively transcribed genes of skin fibroblasts after UVC stress (A) Donut plot depicting the relative (normalised for genomic size) genome-wide distribution of H2Bub ChIP-seq reads in non-irradiated VH10 cells (outer) compared to input (inner) using custom-defined genomic annotation (Section 2). (B) Same as in (A) focusing on genic regions of active genes. (C) Heatmaps showing read densities for H2Bub before, 3 h and 48 h after UV irradiation for genes larger than 60 kb, from TSS to TSS + 60 kb, for (n = 3332) active genes (rows) ranked by increasing RNAPII escape index (EI, Section 2; comparison of the density of RNAPII at promoters (− 250 bp to + 100 bp around the TSS) to that in gene bodies (TSS + 100 bp to + 2 kb)) (D) Average profiles of read distributions shown in (C) for (n = 2279) non-overlapping 10 kbs from TSS. The intersection point of H2Bub signal + 3 h with no UV is shown (dotted line) (E) Box plot depicting the distribution of H2Bub ChIP-seq reads intron density in the first 3 introns (length > 1 kb, line is median) of active genes before and 3 h after UV irradiation. Quantification of read densities in each intron shows stepwise reduction of H2Bub signal in consecutive introns in both conditions (Iterative Student's t-tests ($p < 0.05$) and 95 % confidence interval (CI) excluding 0, see Section 2)) (F) Lasagna plots (heatmap of genomic bins average) of H2Bub ChIP-seq reads density around the start position of the first five exons (− /+ 2 kb) for active genes before UV (upper) and 3 h post UV (lower). Exons with flanking introns < 2 kb were excluded (see Section 2) (G) Schematic showing the Distribution index (DI) calculation method (upper panel), and scatter plot (lower panel) illustrating the global increase of DI in active genes upon UV irradiation (3 h post-treatment). Percentages represent the proportion of genes with DI higher with (top) or without (bottom) UV. (H) Heatmap of $\log_2$ fold change ($\log_2$ FC) of H2Bub read densities (+ 3 h UV/NO UV). Genes are sorted from top to bottom based on increasing length of the first intron.

contrast, H3K27me3 (Fig. S2C), as well as H3K27ac [13] remained stable under similar UV conditions maintaining a permissive chromatin environment throughout the active transcriptome.

To interpret the genome-wide distribution of H2Bub and find what mechanism regulates the observed loss in global H2Bub in response to UV, we performed ChIP-seq against H2Bub. Taking advantage of a custom script to curate gene and regulatory elements annotation in our cell line, we established precisely that in steady-state conditions reads distribution of H2Bub is mainly restricted to active transcription units (see Section 2, Fig. 1**A**), in accord with prior studies [14,65–67]. In particular, within active genes, 90 % of H2Bub reads were found to be enriched over introns and exons, with a lower enrichment in the latter, and to a lower extent on 3′ UTR regions. On the other hand, H2Bub was depleted from promoters (defined as − 250 to + 100 bp around the transcription start site (TSS)) and 5′UTR regions (Fig. 1**B**, Fig. S3A) [18,23,24]. H2Bub average profiles were computed along long active genes (gene length ≥ 60 kb) and we confirmed that in steady-state fibroblasts, this histone mark is virtually absent from TSS, increases sharply within the first kilobases (kbs) of the gene body, and gradually decreases towards the 3′ end (Fig. 1**C–D**).

In line with Fuchs et al. data [24], we also note that H2Bub levels drop significantly in consecutive introns of active genes in our cells (Iterative Student's t-test $p < 0.05$ and 95 % confidence interval (CI) excludes 0) (Fig. 1**E**). Next, we determined the summit position of H2Bub signal and the effect that intron 1 length might have on H2Bub distribution patterns. We separated active genes over 30 kb into five categories (based on their first intron length) and plotted the corresponding average H2Bub densities (using 1 kb bin-counts; Section 2). H2Bub maximum signal was found consistently within the first 5 kb from TSS for all introns larger than 5 kbs, even if intron 1 length increased, and was mainly located within intron 1 sequence (Fig. S3B–C). Interestingly, we note that although peak summit position is detected at the beginning of the genes independently of intron 1 length, intron 1 length positively correlates with the total level of H2Bub in genes as well as the width of the H2Bub peak (Fig. S3D–E). Taken together, these results show that i) the majority of H2Bub levels is accumulated in intron 1, ii) H2Bub levels culminate after 5 kb of intron 1, and iii) H2Bub levels seem to be saturated (reaching a plateau regardless of intron length) in long first introns.

Focusing around intron-exon junctions, we counted H2Bub reads in 2 kb windows around the start position of the first 5 exons (for active genes with at least five exons with flanking introns of length > 2 kb). Our analysis (Fig. 1**F,** upper panel) highlights significant losses in H2Bub signals at about 250 bp following exon starts (probably within the following intron) at these exon-intron

junctions except from the first corroborating the observation that the majority of H2Bub signal is found in intron 1 and suggesting that intron-exon architecture constitutes a major determinant of H2Bub levels in chromatin.

Next, we examined the effects of UV irradiation on the H2Bub landscape. We detected a major redistribution of this histone modification throughout active genes at 3 h after exposure to UV that returned to the steady state pattern at 48 h of recovery (Fig. 1**C–D**). In particular, UV triggered an increase in H2Bub density at 5′ proximal regions of all active genes, which contrasted a decrease at distal regions- a phenomenon reminiscent of elongating RNAPII profile in similar conditions [8]. This finding is supported by the observation that H2Bub is substantially enhanced 3 h post UV, especially in intron 1 (Fig. 1**E**, Fig. S3F) and H2Bub levels show increased accumulation at a shorter distance to TSS in comparison with the non-UV condition (Fig. 1**F).** We then quantified the redistribution of H2Bub around the estimated position of the transcriptional elongation wave that is released from PPP sites upon UV-C [8] by calculating the ratio of H2Bub reads between proximal (TSS to + 30 kb) and distal (+ 30 to + 60 kb from TSS) regions in both conditions (Distribution Index; DI, Section 2). This analysis confirmed a higher DI in over 88 % of active genes 3 h post-irradiation in comparison to NO UV (Fig. 1**G**). Interestingly, while increased accumulation is detected immediately after TSS in regions located at the beginning of intron 1 at most genes, H2Bub density decreases upon UV around the last 2 kb of intron 1 sequences, indicating that the redistribution of H2Bub upon UV does not occur uniformly along the length of intron 1 (Fig. 1**F,** lower panel). Sorting transcripts based on increasing intron 1 length and plotting the $\log_2$ fold change (LFC) of read densities before and after UV-irradiation, we found that H2Bub gain was confined uniformly within about 18 kb in all genes (Fig. 1**H,** Fig. S3G**;** see also intersection point in Fig. 1**D**). We also confirmed that UV-induced redistribution of H2Bub is independent of the 1st intron length (Fig. S3H). We note that the depletion of H2Bub signal in the remaining gene body covers longer genomic regions and justifies the global loss of H2Bub observed by total histone extracts in response to UV (Fig. 1**H**, see Fig. S2A–C). Overall, these results demonstrate a uniform and significant accumulation of H2Bub at the 5′ proximal region of active genes in response to UV stress.

3.2. ATXN7L3-DUBs regulate of H2Bub dynamics in the 5′ proximal regions of actively transcribed genes

The synergy between ubiquitination and deubiquitination of H2B is implicated in the proper function of transcription elongation [18,65,66] and affects the observed H2B levels in the genome [23,24,32]. Given the reduction of global H2Bub levels in chromatin extracts in response to UV (Fig. S2) and the redistribution pattern around 5′ genic regions similar to the pattern reported earlier for the elongating RNAPII at the early steps after UV [8] (Fig. 1**C–D**), we investigated the potential role of H2B deubiquitination in the UV-triggered transcription re-organisation response. To this end, we used control (shCTL) and ATXN7L3 knockdown (KD, shATXN7L3) HeLa cells, in which the ATXN7L3-DUB modules (containing USPs) are unable to assemble and be activated, neither within the context of SAGA nor independently [23,35].

We first verified by WB analysis of histone extracts isolated at various time points after UV irradiation in control and ATXN7L3 KD cells that downregulation of ATXN7L3 protein levels results in substantially increased H2Bub levels (Fig. S4A). Next, we performed ChIP-seq experiments for H2Bub and we confirmed the essential role of ATXN7L-DUB in controlling dynamics of H2Bub turnover in genes by comparing the distributions of H2Bub in shCTL and shATXN7L3 cells treated or non-treated with UV irradiation. As expected, under NO UV conditions a higher number of reads mapped to active gene bodies in shATXN7L3 cells compared to shCTL cells (Fig. 2A–C, Fig. S5A, [35]). We note that H2Bub levels were elevated especially around the beginning of active genes with gains localised in intron 1 regions (Fig. S5B–C). To quantify this observation, we calculated Distribution Indices (see above: DI = ratio of reads within the first and the second 30 kb of gene bodies). The increase in DI between shATXN7L3 and shCTL cells for over 90 % of active genes ((ΔDI > 1, Fig. 2**D** (left)) illustrates that increase in H2Bub caused by loss of ATXN7L3-containing DUB activity preferentially occurs in the first part of the genes. In other words, the higher DI in shATXN7L3 cells results from a higher ratio of proximal to distal H2Bub densities upon KD, positing a role for ATXN7L3-containing DUBs in controlling H2Bub turnover in the first kbs of all active genes (Fig. S5D).

Investigation of the redistribution patterns of H2Bub upon UV irradiation in shCTL and shATXN7L3 cell lines revealed that the vast majority of active genes (approximately 96 % in both cell lines) displayed an increase in DI upon UV (Fig. S6A). We then compared the extent of redistribution in response to UV in the two cell lines by comparing the changes in DI upon UV and found that in about 61 % of genes in shATXN7L3 cells and 38 % in shCTL cells this difference is greater (Fig. 2**D,** right). This could be attributed directly to the increase of H2Bub signal in the first kb of genes upon UV (Fig. 2**B,** bottom and Fig. S5C). Taken together, these results also argue for an essential role for ATXN7L3-DUBs in the regulation of H2Bub level in the proximal regions of actively transcribed genes.

To track the concomitant changes in RNAPII across the same conditions, we also performed ChIP-seq for RNAPII in shCTL and shATXN7L3 cells prior to and upon UV irradiation. Interestingly, although we did detect a lower level of RNAPII in gene bodies in shATXN7L3 cells (Fig. 2**B**), RNAPII distribution was not majorly affected in shATXN7L3 cells **(**Fig. 2**A–B)**. We compared the density of RNAPII at promoters (− 250 bp to + 100 bp around the TSS) to that in gene bodies (TSS +100 bp to + 2 kb) (Escape Index; EI, see Section 2; Fig. 2**E**). In non-irradiated conditions, ATXN7L3 knockdown did not significantly impact EI, denoting that higher H2Bub density in the gene body is not related to an altered PPP-release of RNAPII into active elongation (Fig. 2**E**, left). Furthermore, our data suggest that the presence of the UV-triggered RNAPII elongation wave is hardly affected by impaired DUB activity at the early times (2 h) after UV (increased EI (UV vs NO UV) for 87 % of genes in shCTL and 86 % in shATXN7L3 upon UV treatment, (Fig. S6B), although there might be a tendency to detect decreased DEI in some genes in shATXN7L3 (Fig. 2**E** right), suggesting that wave characteristics may be fine-tuned by proper H2Bub dynamics.

3.3. ATXN7L3- DUB activity determines the transient sate of H2Bub levels in the wake of elongating RNAPII

To determine the causality in the relationship between active transcription and H2Bub, we inhibited transcription using the reversible inhibitor 5,6-Dichloro-1-*beta*-D-ribofuranosylbenzimidazole (DRB). DRB prevents *de novo* release of elongating RNAPII from PPP sites in response to UV, whilst allowing already engaged elongating RNAPII molecules to proceed in transcription [8]. Analysis of total histone extracts demonstrated a global and uniform decrease in H2Bub levels over time in both shCTL and shATXN7L3, confirming that active transcription is necessary for the maintenance of H2Bub level (Fig. S7A). Notably, the reduction observed was slower and less pronounced in the shATXN7L3 cells, implicating ATXN7L3-containing DUBs, and the process of deubiquitination, but not just reduced H2B ubiquitination activity, in the observed reduction of global H2Bub levels in response to transcriptional inhibition.

We then performed DRB treatment followed by ChIP-seq to monitor the progression of the RNAPII already engaged in elongation prior to inhibition (pri-elongating) and the corresponding H2Bub distribution genome-wide. Incubation of shCTL cells with DRB for 40 min ensured that some elongating RNAPII remained engaged in

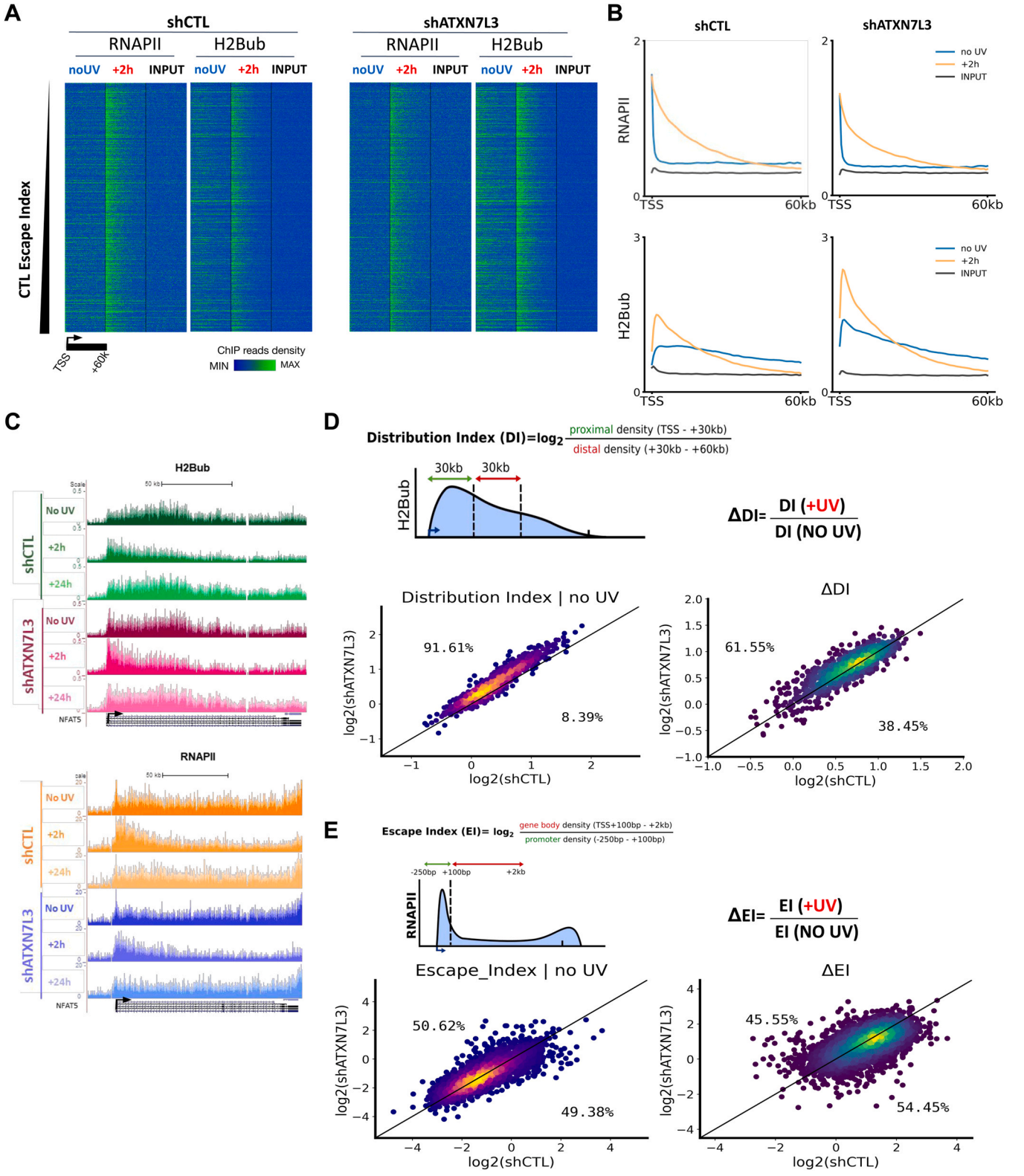

Fig. 2. H2Bub redistribution upon UV irradiation is mediated by ATXN7L3 (A) Heatmaps depicting read densities for elongating RNAPII and H2Bub for genes larger than 60 kb (n = 1131) from TSS to TSS + 60 kb, aligned at individual (rows) active genes and ranked by increasing escape index (see Section 2) of the shCTL cells in non-irradiated conditions. (B) Average profiles illustrating the same annotation as (A). (C) Genome browser snapshots showing H2Bub (top) and RNAPII (bottom) in a representative gene in shCTL (top) and shATXN7L3 (bottom) cells in non-irradiated, 2 h and 24 h after UV irradiation conditions. (D) Density scatter plots illustrating DI for H2Bub in shCTL VS shATXN7L3 cells in non-irradiated conditions (left) depicting the accumulation of H2Bub signal in shATXN7L3 cells in proximal gene regions. A comparison of the change in DI (ΔDI = log2(+ UV/no UV)) 2 h post UV in the two cell lines (right), demonstrating a more extensive H2Bub redistribution in shATXN7L3 cells compared to shCTL cells upon UV. (E) Schematic illustrating the EI calculation (top) (EI, Section 2; comparison of the density of RNAPII at promoters (− 250 bp to + 100 bp around the TSS) to that in gene bodies (TSS + 100 bp to + 2 kb)). Density scatter plots showing the correlation between constitutive EI (NO UV) for RNAPII for all common active genes in shCTL VS shATXN7L3 (left) and a comparison of the change in EI (ΔEI = log2(+ UV/no UV)) 2 h post-UV (right). H2Bub and RNAPII samples shown in this figure are the result of two replicates merging.

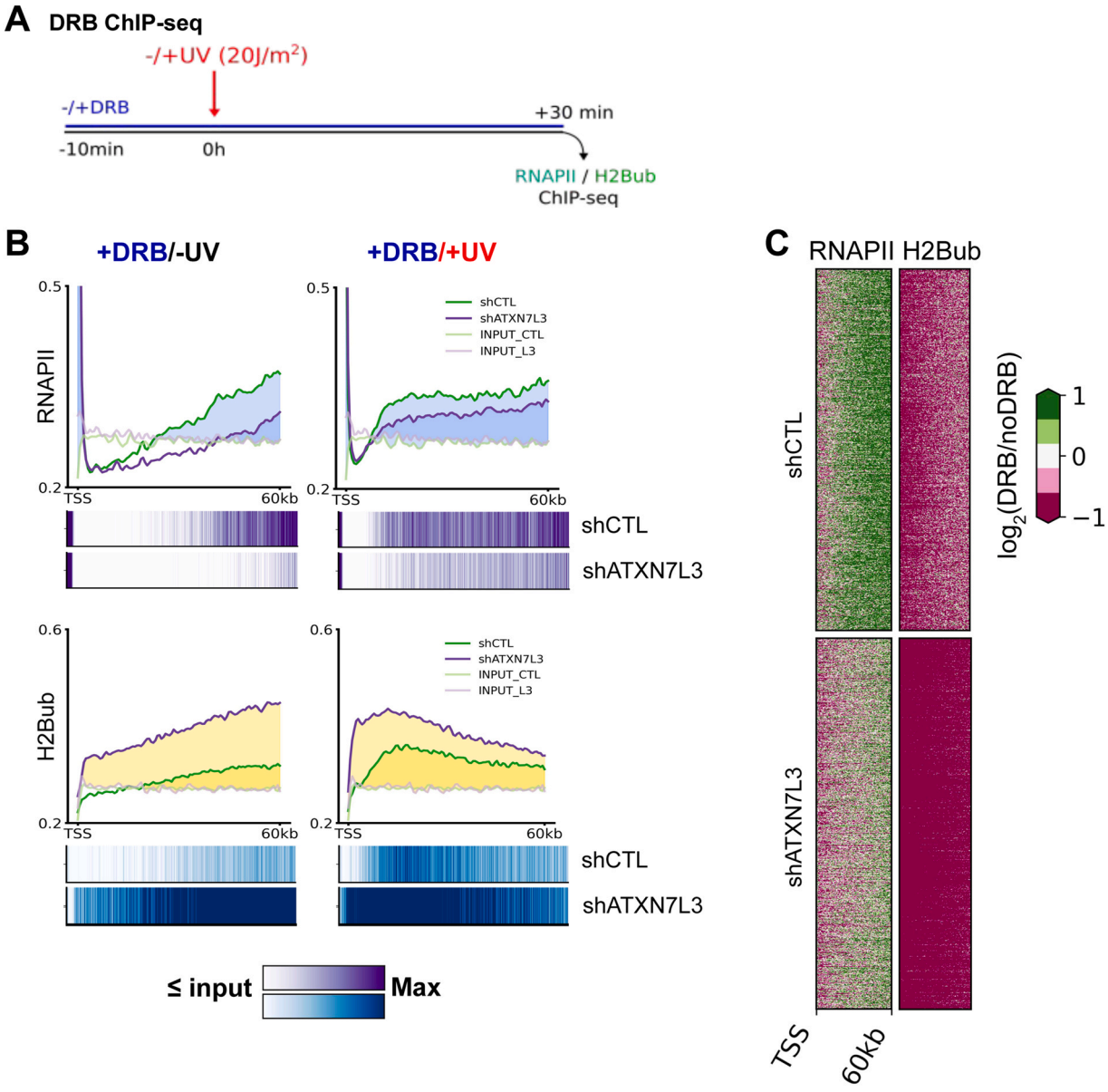

Fig. 3. Aberrant ATXN7L3-related H2Bub distribution affects pri-elongating RNAPII wave backend position. (A) Schematic of the experimental setup. Briefly, cells were either treated with DRB or mock treated for 40 min to induce pausing of RNAPII at PPP in order to monitor the progression of the wave-backend of RNAPII already engaged in elongation (pri-elongating). 10 min after DRB or mock treatment, cells were either irradiated or not, and all cells were left to recover for 30 min in the same conditions as before irradiation. (see Section 2) (B) Average profiles illustrating read densities for RNAPII (top) and H2Bub (bottom) in the presence of DRB on active genes over 60 kb, from TSS to TSS + 60 kb in non-irradiated (left panel) and the same for irradiated (right panel, 30 min) conditions. Gene region where sample density > input density is coloured. Below each, lasagna plots depicting the signal for each bin of each sample minus the corresponding input signal are shown. (C) Log2 FC heatmap of H2Bub and RNAPII read densities (DRB sample/no DRB sample).

distal regions on long genes of 40 kb or longer [8]. The extent of progression of the advancing wave of RNAPII was determined by the 5′position of the receding RNAPII wave, or "wave backend" (Fig. 3**A**; Section 2). Treatment with DRB in non-irradiated conditions resulted in the depletion of RNAPII signal at 5′ proximal gene regions and the concomitant loss of H2Bub from those regions (Fig. 3**B** left panel). Conversely, H2Bub was retained at more distal regions, mirroring the profile of the receding pri-elongating wave of RNAPII, and illustrating that in active genes the level of H2Bub on gene bodies is dynamically regulated and closely correlated to the local presence of elongating RNAPII. We confirmed this by showing that H2Bub is more depleted from areas where RNAPII is lost upon DRB (Fig. 3**C**, purple areas on heatmap). Together, these results strongly support a model whereby H2Bub is removed in the wake of RNAPII during elongation.

We next determined the role of ATXN7L3-associated DUBs in this process. We found that in shATXN7L3 cells, H2Bub distribution shift upon DRB treatment followed the same pattern as in control cells, but with higher absolute residual levels of H2Bub throughout the RNAPII-depleted region (Fig. 3**B,** left, bottom). Our data indicate that increased H2Bub half-life in the shATXN7L3 cells most likely results from the inability of these cells to deubiquitinate H2B after RNAPII passage. Independently of the initial level of H2Bub or RNAPII on chromatin before DRB treatment, our experimental setup allowed us to clearly determine how ATXN7L3-mediated DUB activity regulates H2Bub levels in the absence of new transcription elongation events. Interestingly, in KD cells RNAPII wave backend was located more distally upon DRB treatment (Fig. 3**B**, top). We also note that the levels of pri-elongating RNAPII are lower in shATXN7L3 cells through gene bodies, similar to the picture seen above in the average profile

of the actively elongating RNAPII under steady-state conditions (Fig. 2**B,** shATXN7L3 cells).

We then cumulated DRB treatment and UV irradiation and performed H2Bub and RNAPII ChIP-seq to monitor the impact of UV lesions and the damage-induced deceleration of elongating RNAPII [8,9] on H2Bub dynamics. Following the same experimental set-up, as described above, with a UV-irradiation step 10 min after the addition of DRB (Fig. 3**A**, Section 2), we confirmed that RNAPII wave backend position was closer to the promoter in UV conditions in both cell lines (Fig. 3**B**, right panel, top). In response to damage induction, the difference in the wave backend position between the two cell lines observed under normal conditions is lost, demonstrating that transcription-blocking UV lesions override the effect observed on RNAPII elongation rate caused by impaired H2Bub dynamics in KD cells. Interestingly, while this delay is accompanied by a loss of H2Bub signal in the regions upstream of the wave backend in shCTL cells, in shATXN7L3 cells a higher H2Bub signal is retained (Fig. 3**C**, bottom).

Together, these data provide valuable mechanistic insights into the temporal and spatial organisation of deposition and removal of H2Bub during transcription elongation and reveal the determinant role of RNAPII and ATXN7L3-containing DUBs status during steady-state transcription as well as in the process of transcriptional reorganisation in response to genotoxic stress.

3.4. ATXN7L3-DUB activity modulates transcription elongation rate

To further characterise the differences seen in RNAPII binding in shATXN7L3 cells, we investigated the dynamics of nascent RNA synthesis. We performed nascent RNA-seq experiments in cells that had been initially treated with DRB for three hours (pre-DRB nRNA-seq) and then we monitored RNAPII progression in the two cell lines (shCTL and shATXN7L3) at different times upon release from DRB. We isolated EU-labelled nascent RNA at 20, 30 and 60 min after DRB removal (with or without a UV irradiation step just before release from DRB, Fig. 4**A**). Wavefront position in active gene bodies was estimated from RNA reads progression patterns using a two-state left-to-right Hidden Markov Model [63] (Section 2 Fig. S1E). We found that the RNA synthesis rate is significantly higher in shATXN7L3 cells both in steady-state and in irradiated conditions, as illustrated by the more distal position of the wavefront in shATXN7L3 cells at 30 and 60 min after release from DRB in No UV conditions, and at all time points after UV irradiation (Fig. 4**B–D**). As previously shown [7–9], obstacles posed by UV-lesions on gene bodies significantly reduced wavefront progression at all time points after UV in both cell lines (Fig. 4 **B–D**). Strikingly, the difference between the two cell lines was reduced upon UV, but not completely abolished; RNAPII elongation stalling, as a result of UV-lesions, uncovered differences in the elongation rate upon ATXN7L3 KD sooner (20 min) upon UV than in non-irradiated conditions (30 min). These results indicate that impairment of H2Bub de-ubiquitination perturbs RNA synthesis rates. While this effect is less prominent in the context of the decreased RNAPII progression by UV lesions, we note that a proportion of polymerases not blocked at UV lesions might still elongate faster in shATXN7L3 cells, as higher H2Bub levels will remain on adjacent nucleosomes.

3.5. H2Bub correlates with activity at enhancers

We used our data to examine the distribution pattern of this modification around regulatory regions. Using ENCODE data from HeLa cells and PHANTOM5 annotation data, we defined intergenic enhancers and categorized them based on their activity status, as determined by the chromatin state 10 kb around eTSS positions (Section 2). Presence of H3K27ac and DNase peaks defined active loci, whilst repressed elements corresponded to the presence of H3K27me3 and the absence of H3K27ac and DNase I signal. Enhancers with absence of H3K27ac, DNase and H3K27me3 signal were classified as inactive. We detected an enrichment of H2Bub in the vicinity of active enhancers in UV-irradiated and non-irradiated conditions (Fig. 5**A**). Although the signal was low relative to H3K27ac and H3K4me1, it was specific as demonstrated by the bimodal profile around the eTSS. These results therefore, confirm an association between enhancer activity and H2Bub presence in our setup. To investigate the function of H2Bub in these regions, we focused on H2Bub-enriched enhancers, which were commonly detected in shCTL and shATXN7L3 cells. We found that the H2Bub signal was marginally higher in shATXN7L3 cells, suggesting that ATXN7L3-mediated DUB is involved in the regulation of H2Bub levels during non-coding eRNA transcription (Fig. 5**B**).

Exploiting our experimental setup, we confirmed that upon UV irradiation, loss of RNAPII reads from eTSS regions **(**Fig. 5**A–B),** in both shCTL and shATXN7L3 cells, was compensated by the transition of RNAPII into elongation as highlighted by increased EI in these areas (Fig. S8A), in line with a previous report [13]. In agreement with the results presented above at genic TSSs, we did not find significant differences in EI upon UV between shCTL and shATXN7L3 cells. We thus conclude that transcriptional wave release from PPP sites upon genotoxic stress is only marginally impacted by potential changes in H2Bub levels both at TSSs and at eTSSs.

To examine the potential correlation between H2Bub level and enhancer features such as chromatin landscape and transcriptional activity, we next divided active intergenic enhancers into 3 groups (sorted on their H2Bub level in CTL no UV condition) and compared the levels of RNAPII, nascent RNA (nRNA), H3K27ac, DNase at these loci. We found that enhancers with increasing H2Bub levels were also associated with greater enhancer activity as depicted by nRNA levels, higher RNAPII density, as well as with a more open chromatin state (H3K27ac, DNase) (chi-squared test) (Fig. 5**C,** Fig. S8B). Such a relationship is somewhat unsurprising considering the established role of H2Bub in transcriptional elongation, together with the fact that the vast majority of active enhancers are indeed transcribed.

Finally, we examined the impact of enhancer H2Bub level on the activity of associated genes with CAGE enhancer-gene associations. Using the three categories generated above, we looked at levels of H2Bub, RNAPII, nRNA and EI in non-irradiated conditions at genes associated with enhancers (see Section 2) in each category. Although, we did not find any significant correlations between the enhancer levels of H2Bub and RNAPII EI of the associated genes (Fig. S8C), artificial impairment of the H2Bub turnover rate brought about by ATXN7L3 KD resulted in a reduced level of both nascent RNA and RNAPII binding in the same group of genes (Fig. 5**D**). These data indicate that excess of H2Bub could prevent induction of the target genes, in line with previous reports [68,69]. Notably, our analysis also revealed a greater number of genes to be associated with enhancers showing higher levels of H2Bub (Fig. 5**E–F**). Taken together, these observations are compatible with the possibility of H2Bub increased levels to grant more eRNAs and stabilise larger transcription factories that involve more genes. They further substantiate its role in transcription initiation at eTSSs and extend the H2Bub role previously reported at inducible enhancers [68,69].

4. Discussion

Elucidating the state of the epigenome during the cellular response to UV stress will broaden our understanding of how cells modulate DDR and transcription-coupled damage-sensing to ensure the maintenance of genome integrity. In this study, we characterized novel features of the genome-wide distribution of H2Bub in a steady state and its reorganisation patterns detected after DNA damage. We used to our advantage the UV-triggered transcriptional response to reveal details of the mechanism that regulates proper writing/

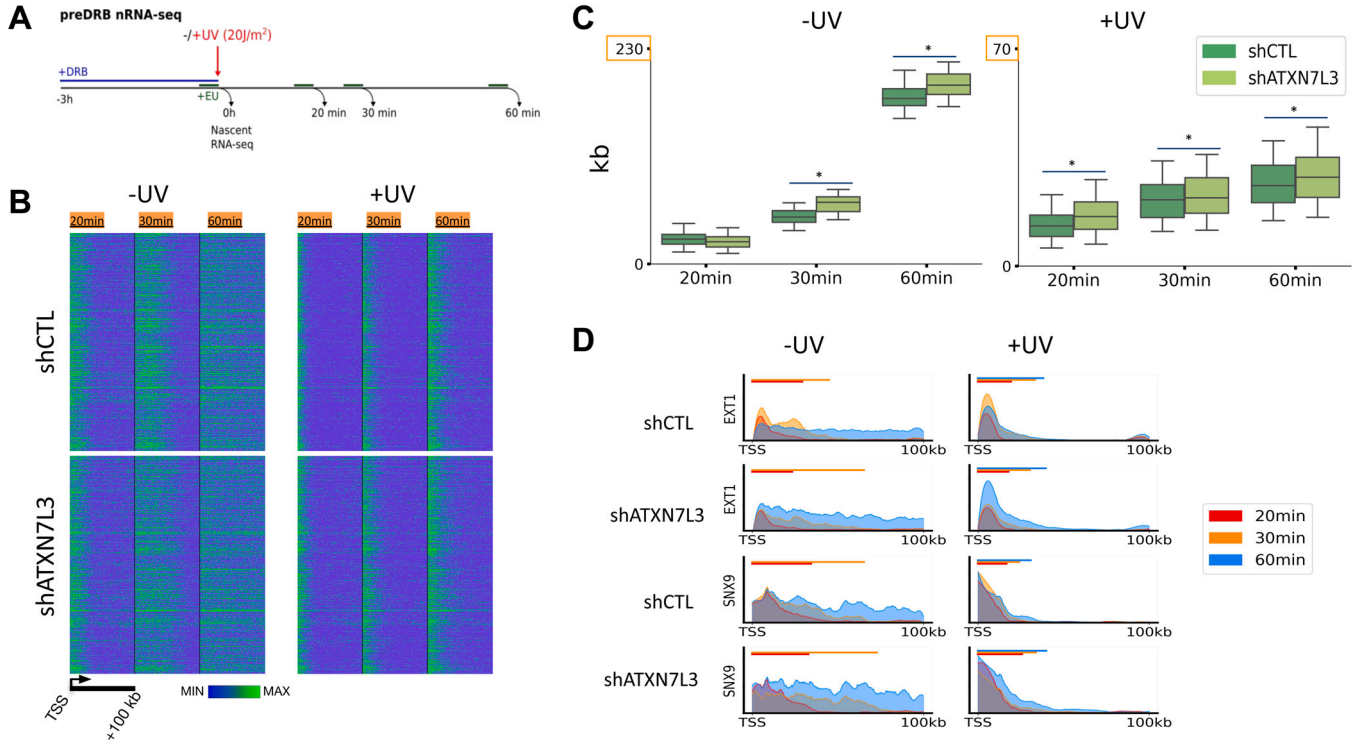

Fig. 4. Deficient H2Bub DUB activity leads to a faster rate of *de novo* elongation wave in non-irradiated and + UV conditions. (A) Scheme of DRB-release methodology (pre-DRB-nRNA-seq). Briefly, cells are treated with DRB for 3 h, UV-irradiated, and released from DRB inhibition during recovery for various amounts of time as indicated. Cells are then treated with EU for 10 min prior to harvesting at the indicated time points by RNA isolation with Trizol. (B) Heatmaps depicting read density for nRNA at indicated time points after DRB release in non-irradiated (left) and irradiated (right) conditions, in shCTL (top) and shATXN7L3 (bottom) cells, illustrating the progression of the *de novo* transcriptional wave, on active genes over 100 kb, from TSS to TSS + 100 kb (n = 614). (C) Boxplots showing predicted wavefront position, in kb, at the above time points for both cell types in non-irradiated (left) and irradiated (right) conditions, illustrating the impaired progression of the *de novo* RNAPII wave upon UV in both cell types, and highlighting a greater wave advancement in shATXN7L3 cells. Bars with a star over boxplots represent significantly different distributions (two-sided Kolmogorov-Smirnov p-Val < 0.05). (D) Overlapping single-gene nRNA signal profiles of EXT1 and SNX9 genes for shCTL (top) and shATXN7L3 (bottom) in unirradiated (left) and irradiated (right) conditions. Bars over profiles indicate the active elongation signal estimation for the time point of the corresponding colour.

erasing of H2Bub during active transcription and shed light on the role of H2Bub on these chromatin-associated processes. Our H2Bub ChIP-seq data provided high-resolution details on the Spatio-temporal distribution of H2Bub in relation to gene architecture. Maps established in response to transcriptional inhibition and genotoxic stress as well as in cells with impaired deubiquitination activity revealed functions of H2Bub essential to transcription elongation rate regulation. These results highlight the importance of controlling H2Bub turnover activity and reveal SAGA and ATXN7L3-DUB modules as essential components of this process in the 5′ proximal part of gene bodies and at the early stages of the UV response.

In non-damaged cells, the presented data reinforced findings from recent literature [21,23,24,70] showing that H2Bub rapidly accumulates at the beginning of genes, and most often within intron 1 sequences. We located the summit of the H2Bub peak within 5 kb from the TSS regardless of the gene or intron 1 size. We note that H2B levels remain high throughout intron 1 and are sharply and incrementally reduced at successive intron-exon junctions. Furthermore, our data suggest that it is not intron length per se but intron-exon architecture that constitutes a major determinant of H2Bub levels on chromatin. Our analysis provides positional insights into genomic H2Bub distribution and explains why H2Bub levels remain higher in the first 10 kb for genes with fewer exons in this window [24]. Interestingly, it has been shown that FACT recruits PAF1C to couple RNF20/40 H2B ubiquitinating complex to elongating RNAPII [66,71–73]and this mechanism may act primarily on the first 30 kb of active genes [74]. Taken together, these data suggest a requirement for the regulation of elongation, by modification of chromatin, in the regions closest to the TSS.

Recent reports indicate how chromatin influences the efficiency of repair [42,43], but there is still little information regarding the state of the epigenome during the early stages of the UV response. Global levels of transcription-associated histone modifications, such as H3K27ac and K3K27me3, were recently found to remain stable in the early recovery times (1–4 h) after a moderate UV dose (15 J/m^2) [13]. On the contrary, we show here that H2Bub undergoes an extensive and dynamic reorganization over actively transcribed regions upon UV where a uniform gain at 5′ regions across all active genes is accompanied by a loss of signal from distal gene body regions in a gene and intron 1 length independent manner (Fig. 1**H**). This pattern mirrors the reshuffling of elongating RNAPII observed during the transcriptional reorganization response to UV [8–10]. and we speculated that our experimental setup could highlight the interdependency of the H2Bub mark and elongating RNAPII (see below).

Impairment of ATXN7L3-DUB activity in human cells cannot be fully compensated for by other DUB enzymes, as evidenced by the significant increase in overall H2Bub levels in shATXN7L3 cells (Fig. S3) and in agreement with previous reports [23,32,35,39]. Assessment of H2Bub genomewide distribution in these cells, showed the most increased levels proximal to the TSS, confirming that SAGA and ATXN7L33-DUB activities regulate H2Bub levels in these regions (Fig. 2 **and** Fig. S5). On the other hand, whereas a greater accumulation of H2Bub in 5′ proximal regions was also observed upon UV, the levels of this modification were not less reduced in more distal regions. We, therefore, propose that preferential targeting of SAGA and ATXN7L3-DUB activity to 5′ proximal gene regions, which is more pronounced upon stress, could underly ubiquitin removal specifically from those areas. Our findings are compatible with studies in *S. cerevisiae* showing that UBP8 (hUSP22, the native

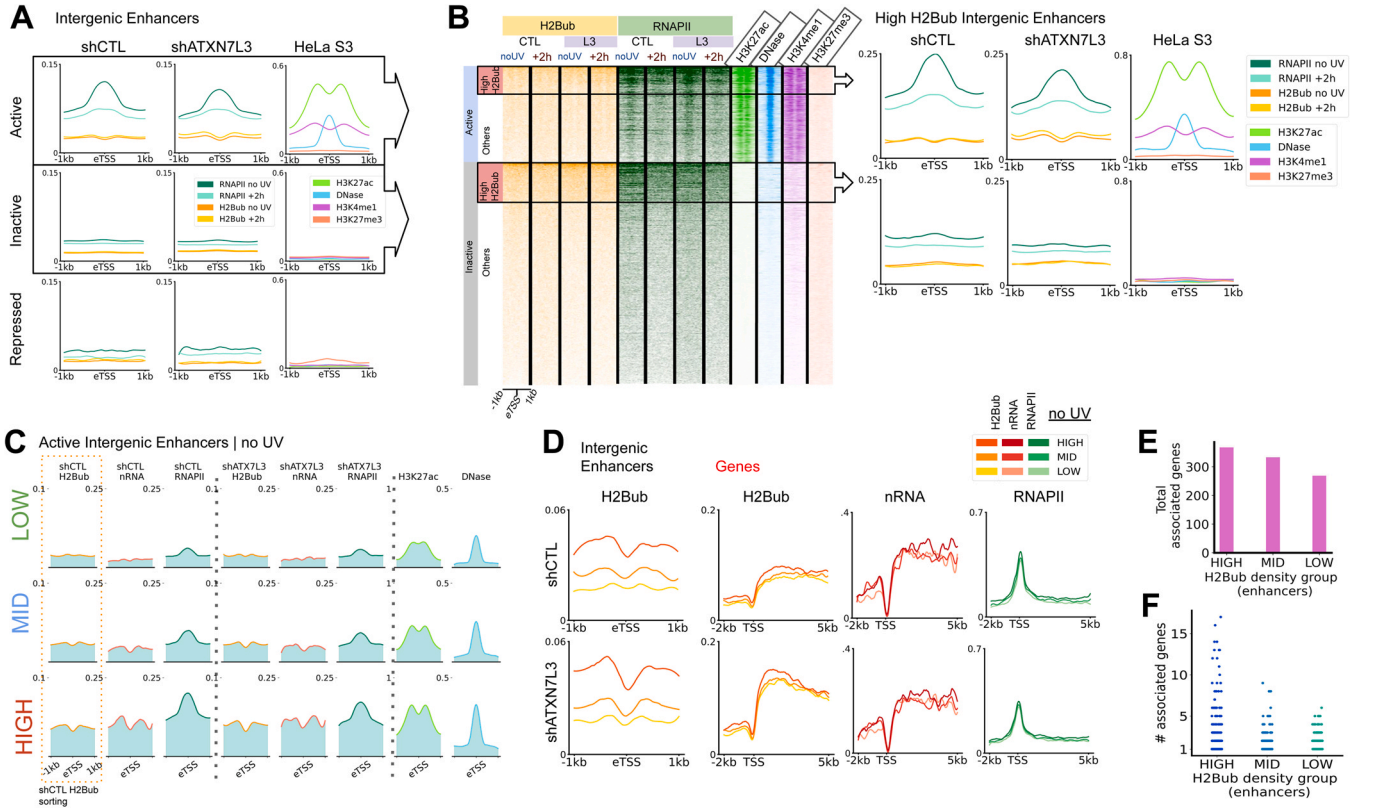

Fig. 5. H2Bub associates with actively transcribed intergenic enhancers. (A) Average profiles showing read densities from H2Bub and RNAPII ChIP-seq in irradiated (+2 h UV) and non-irradiated conditions, in shCTL and shATXN7L3, for genomic ± 1 kb region around the eTSSs regions of active (n = 2798), inactive(n = 20,718) and repressed (n = 218) intergenic enhancers, illustrating the presence of H2Bub specifically in active enhancers. (B) Heatmap showing high H2Bub density groups of active and inactive intergenic enhancers (marked with the black frame) in contrast to the rest of active and inactive (n = 5000 sample) intergenic enhancers on the same data as in (A). Intergenic enhancers, regardless of their activity status, were sorted and split into ten, equally populated, groups based on their no UV H2Bub density (each cell line independently) and the intersection of the shCTL and shATXN7L3 highest density groups formatted the high H2Bub group (n = 1531). In average profiles, the active (n = 481) and inactive (n = 1046) intergenic enhancers that constitute the group are shown separately. (C) Active intergenic enhancers were sorted on their H2Bub level (shCTL no UV condition dataset, highlighted) and split into 3 equally populated groups (n = 868 each) and on a + /- 1 kb region around the eTSSs levels of H2Bub and RNAPII in no UV conditions, nRNA (no UV, 60 min release from DRB), as well as Hela S3 H3K27ac, DNase and H3K4me1 from ENCODE, were plotted. (D) Active intergenic enhancers with known gene associations (CAGE data) (n = 606) were sorted and split as in (C). Average profile showing levels of H2Bub for each group (left) on a ± 1 kb region around the eTSSs and for the genes (> 5 kb) associated with the enhancers of each group (right) alongside with corresponding levels of nRNA (no UV, 60 min release from DRB) and RNAPII on a − 2 kb to + 5 kb from TSS region. (E) Bar plot indicating the total number of associated genes for each group of enhancers shown in (D). (F) Categorical scatter plot showing that higher H2Bub density groups (as in (D)) are more enriched in enhancers with many gene associations. H2Bub and RNAPII samples shown in this figure are the results of two replicates merging.

deubiquitinase of SAGA) and UBP10 (hUSP36) target different regions of the genome. Specifically, UBP8 activity is localized at promoter and TSS regions whereas UBP10 has a more global role [75]. Furthermore, although some DUBs may be partially redundant, it is commonly accepted that they operate within specific complexes, target different genomic regions, and function in distinct pathways and cellular processes [39]. As such, rescuing USP22 activity in shATXN7L3 in a limited amount of time (early recovery phase) by other DUBs may be difficult under stress as opposed to steady-state conditions.

The increased transcriptional activity (more traveling RNAPII) due to the synchronous and widespread entry of elongating RNAPII in gene bodies after stress probably rises indirectly H2Bub levels via a feed-forward mechanism where a higher RNAPII occupancy in proximal gene regions following UV stress provokes the recruitment of more RNF20/40. In addition, since the elongation rate of RNAPII is decreased because of UV lesions [8] together with the fact that blocked RNAPII molecules are removed from chromatin [76], might influence the rate of dynamic conversion between H2B and H2Bub. As such, the local speed/stability of RNAPII could affect the time lapse between ubiquitin and de-ubiquitin activities, momentarily increasing H2Bub levels.

Related directly to this, we note that knockdown of ATXN7L3 and UV irradiation both result in the accumulation of H2Bub at the 5′ region of active genes (Figs. 1**D and** 2**B**). Exposure of ATXN7L3 KD cells to UV leads to an even greater level of H2Bub than in either of the individual conditions, pointing to a cumulative effect on H2Bub dynamics in these regions. Strikingly, this is not reflected in RNAPII dynamics; UV irradiation leads to an increase of RNAPII at the beginning of genes (also seen by the increase in EI upon UV), whereas knockdown of ATXN7L3 and the resulting H2Bub redistribution and density-increase does not majorly affect RNAPII. We can speculate that, since H2Bub levels influence RNAPII rate, but RNAPII rate also can impact the H2Bub/H2B exchange rate (feedback), these processes could thus be contributing independently to the combined UV and KD condition to increase H2Bub levels at the beginning of active genes.

Sequence and position of H2B ubiquitination and deubiquitination during transcriptional elongation create a dynamic H2B-H2Bub turnover via repeated cycles of histone H2B ubiquitination and deubiquitination as RNA Polymerases progress through genes [18,65,66,77,78]). However, the ambiguity of the evidence has resulted in inconsistent models regarding the position and timing of H2Bub deposition and removal, relative to RNAPII during transcription and the mechanism through which it conveysits effect is still elusive [66,79]. To address this question, and based on previous results suggesting how a reduction in pre-mRNA levels on a set of genes correlates with H2Bub levels in response to DRB inhibition of

transcription elongation [24], we demonstrate that the default state of H2B is non-ubiquitinated in active genes and that detection of H2Bub depends on RNAPII. Indeed, when we tracked the wave backend of an advancing RNAPII complex on active genes, we showed how the presence of RNAPII locally induces ubiquitination of H2Bub and that loss of RNAPII from specific regions within active genes results in loss of H2Bub from the same regions (Fig. 3**B, C**). Importantly, the DRB-dependent loss of H2Bub (in both UV irradiated and normal conditions) from the areas upstream of the wave backend was lower in knockdown cells, implicating DUB by SAGA in transcription elongation, and perhaps more interestingly, locates SAGA-DUB activity again at the beginning of genes (see above).

Interestingly, we show that although this loss of H2Bub is a result of the absence of H2Bub deposition in an RNAPII-dependent manner, it also requires its active removal by the de-ubiquitinase activity of SAGA and ATXN7L3-mediated DUB modules (Fig. 3). This data provides reasonable evidence to suggest that during ongoing transcription, H2B is ubiquitinated ahead of a given RNAPII complex, creating, either directly or indirectly, a permissive environment for the incoming transcriptional machinery. In particular, H2Bub is known to give fluidity to RNAPII transcription elongation through nucleosomes [66,80,81] because it stabilizes hexasomes and impacts H2A/H2B exchange dynamics ruled by histone chaperones such as FACT [66,82]. Indeed, H2Bub is required for FACT activity in genic regions [83], which tethers the proximal H2A-H2B dimer removed during nucleosome traversal. Their concerted action facilitates nucleosome reassembly in the wake of RNAPII [84], as transcribed DNA rewraps around a conserved nucleosome (same histones molecules) to guarantee that epigenetic marks are retained during transcription [85]. In this way, while the RNAPII complex moves forward, the ubiquitin moiety is removed from H2B, by SAGA DUB or other ATXN7L3-mediated DUBs, after nucleosome clearance. Restoration of non-modified H2B level creates a restrictive chromatin state, which is importantfor the advancement of the next RNAPII complex, which will operate in the same way, as well as in preventing events, such as cryptic transcription or transcriptional machinery collisions [86].

5. Conclusions

Taken together, our data support a model where a transient ubiquitination of H2B on the nucleosome ahead of a given RNAPII occurs to facilitate nucleosome traversal and is subsequently removed by an ATXN7L3-containing DUB following RNAPII exit from the nucleosome. We speculate that increased stability of H2Bub in KD cells might positively affect RNAPII elongation rate, as indicated by the downstream shift of RNAPII wave back end in the KD cells (Fig. 3). This result is compatible with prior observations linking H2Bub levels with RNAPII rates [24,74]. Similarly, marginal acceleration of *de novo* transcription waves following DRB release in steady state and to a lesser extent in irradiated conditions was observed in ATXN7L3 KD cells (Fig. 4). Although the effect is less prominent in the context of the decreased RNAPII progression by UV lesions, we note that a proportion of polymerases not blocked at UV-lesions might still elongate faster in shATXN7L3 cells, as higher H2Bub levels will remain on adjacent nucleosomes. Conceivably, the increased levels of H2Bub observed upon UV (Figs. 1, 3) could help in the maintenance of genomic integrity following UV stress, as the increased lesion sensing and proper stalling of RNAPII at lesions is a crucial step in triggering TC-NER and therefore removing harmful transcription-blocking lesions. It would be of interest, to assess the effect, if any, of elongation rate changes due to the over-representation of H2Bub on the mutational landscape following UV.

Our findings highlight the importance of the activity of ATXN7L3-containing DUBs to fine-tune levels of H2Bub in a co-transcriptional manner and to regulate RNAPII elongation rate, offering evidence for SAGA's and ATXN7L-DUB modules' essential role in performing de-ubiquitination in the wake of elongating RNAPII during active transcription, a synergy that is possibly disturbed under transcriptional stress.

CRediT authorship contribution statement

The authors state the following authors' contributions. M.F. conceived, supervised and obtained financial support for the study. M.F., D.K., S.F. and M.D.L designed the experiments and together with A.C.S. were responsible for the interpretation of the results. S.F. performed the experiments and analysed the data. A.C.S. performed the statistical and bioinformatics analyses. D.K. supervised A.C.S. with significant contribution of M.D.L. S.F. wrote the manuscript with significant contributions of M.F., M.D.L., D.K. and A.C.S. All authors discussed the results, reviewed, commented on and approved the final version of the manuscript.

Data Availability

The primary sequencing data have been deposited to the BioSample database with accession ID PRJNA869350.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

We are deeply indebted to Prof. László Tora, and Dr. Didier Devys, Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France. Université de Strasbourg, 67404 Illkirch, France, for kindly providing us with the shCTL and shATXN7L3 transfected HeLa cells. We also thank Prof. George Kollias and Christos Tzaferis, Medical School, NKUA and BSRC 'Alexander Fleming' for their help with the Graphical Abstract and the past and current members of the Fousteri lab for their help and critical input during this study. We thank Genecore facility (EMBL, Germany) for sequencing our NGS libraries. This work was funded by an ERC grant to M.F., Agreement-309612 (TransArrest), < Matching Funds > to MF funded by National Sources. and H.F.R.I. 'Research Projects to support Faculty Members and Researchers" Project No:3199 to MF.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2022.12.013.

References

[1] Jackson SP, Bartek J. The DNA-damage response in human biology and disease. Nature 2009;461:1071–8.
[2] Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. Understanding nucleotide excision repair and its roles in cancer and ageing. Nat Rev Mol Cell Biol 2014. https://doi.org/10.1038/nrm3822
[3] Liakos A, Lavigne MD, Fousteri M. Nucleotide excision repair: from neurodegeneration to cancer. Adv Exp Med Biol 2017;1007:17–39.
[4] Hanawalt PC, Spivak G. Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol 2008. https://doi.org/10.1038/nrm2549
[5] Hendriks G, Jansen JG, Mullenders LHF, de Wind N. Transcription-coupled repair and apoptosis provide specific protection against transcription-associated mutagenesis by ultraviolet light. Transcription 2010;1:95–8.
[6] Vermeulen W, Fousteri M. Mammalian transcription-coupled excision repair. Cold Spring Harb Perspect Biol 2013;5.
[7] Andrade-Lima LC, Veloso A, Paulsen MT, Menck CFM, Ljungman M. DNA repair and recovery of RNA synthesis following exposure to ultraviolet light are delayed in long genes. Nucleic Acids Res 2015. https://doi.org/10.1093/nar/gkv148
[8] Lavigne MD, Konstantopoulos D, Ntakou-Zamplara KZ, Liakos A, Fousteri M. Global unleashing of transcription elongation waves in response to genotoxic stress restricts somatic mutation rate. Nat Commun 2017;8.

[9] Williamson L, et al. UV irradiation induces a non-coding RNA that functionally opposes the protein encoded by the same gene. Cell 2017. https://doi.org/10.1016/j.cell.2017.01.019
[10] Borisova ME, et al. P38-MK2 signaling axis regulates RNA metabolism after UV-light-induced DNA damage. Nat Commun 2018;9.
[11] Bugai A, et al. P-TEFb activation by RBM7 shapes a pro-survival transcriptional response to genotoxic stress. Mol Cell 2019;74:254–67. [.e10].
[12] Heilbrun EE, Merav M, Parnas A, Adar S. The hardwired transcriptional response to DNA damage. Curr Opin Syst Biol 2020;19:1–7.
[13] Liakos A, Konstantopoulos D, Lavigne MD, Fousteri M. Continuous transcription initiation guarantees robust repair of all transcribed genes and regulatory regions. Nat Commun 2020;11.
[14] Henry KW, et al. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev 2003;17:2648–63.
[15] Daniel JA, et al. Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. J Biol Chem 2004;279:1867–71.
[16] Kao CF, et al. Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. Genes Dev 2004;18:184–95.
[17] Weake VM, Workman JL. Inducible gene expression: diverse regulatory mechanisms. Nat Rev Genet 2010;11:426–37.
[18] Minsky N, et al. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nat Cell Biol 2008. https://doi.org/10.1038/ncb1712
[19] Shema E, et al. The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. Genes Dev 2008;22:2664–76.
[20] Vethantham V, et al. Dynamic loss of H2B ubiquitylation without corresponding changes in H3K4 trimethylation during myogenic differentiation. Mol Cell Biol 2012;32:1044–55.
[21] Huff JT, Plocik AM, Guthrie C, Yamamoto KR. Reciprocal intronic and exonic histone modification regions in humans. Nat Struct Mol Biol 2010;17:1495–9.
[22] Fuchs G, Oren M. Writing and reading H2B monoubiquitylation. Biochim Biophys Acta - Gene Regul Mech 2014;1839:694–701.
[23] Bonnet J, et al. The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes Dev 2014;28:1999–2012.
[24] Fuchs G, Hollander D, Voichek Y, Ast G, Oren M. Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. Genome Res 2014;24:1572–83.
[25] Wright DE, Kao CF. Ubi)quitin' the h2bit: recent insights into the roles of H2B ubiquitylation in DNA replication and transcription. Epigenetics 2015;10:122–6.
[26] Zhu B, et al. Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Mol Cell 2005;20:601–11.
[27] Kim J, Hake SB, Roeder RG. The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. Mol Cell 2005;20:759–70.
[28] Nijman SMB, et al. A genomic and functional inventory of deubiquitinating enzymes. Cell 2005;123:773–86.
[29] Uhlén M, et al. Tissue-based map of the human proteome. Science (80-) 2015;347.
[30] Melo-Cardenas J, Zhang Y, Zhang DD, Fang D. Ubiquitin-specific peptidase 22 functions and its involvement in disease. Oncotarget 2016;7:44848–56.
[31] Jeusset LMP, McManus KJ. Ubiquitin specific peptidase 22 regulates histone H2B mono-ubiquitination and exhibits both oncogenic and tumor suppressor roles in cancer. Cancers 2017;9.
[32] El-Saafin F, Devys D, Johnsen SA, Vincent SD, Tora L. SAGA-dependent histone H2Bub1 deubiquitination is essential for cellular ubiquitin balance during embryonic development; 2022. ⟨https://doi.org/10.3390/ijms23137459⟩.
[33] Köhler A, Zimmerman E, Schneider M, Hurt E, Zheng N. Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. Cell 2010;141:606–17.
[34] Samara NL, et al. Structural insights into the assembly and function of the SAGA deubiquitinating module. Science 2010;328:1025–9.
[35] Lang G, et al. The tightly controlled deubiquitination activity of the human SAGA complex differentially modifies distinct gene regulatory elements. Mol Cell Biol 2011;31:3734–44.
[36] Morgan MT, et al. Structural basis for histone H2B deubiquitination by the SAGA DUB module. Science 2016;351:725–8.
[37] Lee KK, et al. Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. Mol Syst Biol 2011;7:503.
[38] Helmlinger D, Tora L. Sharing the SAGA; 2017. ⟨https://doi.org/10.1016/j.tibs.2017.09.001⟩.
[39] Atanassov BS, et al. ATXN7L3 and ENY2 coordinate activity of multiple H2B deubiquitinases important for cellular proliferation and tumor growth. Mol Cell 2016;62:558–71.
[40] Smerdon MJ. DNA repair and the role of chromatin structure. Curr Opin Cell Biol 1991;3:422–8.
[41] Adam S, Polo SE. Chromatin dynamics during nucleotide excision repair: histones on the move. Int J Mol Sci 2012;13:11895–911.
[42] Hu J, Adar S, Selby CP, Lieb JD, Sancar A. Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. Genes Dev 2015. https://doi.org/10.1101/gad.261271.115
[43] Adar S, Hu J, Lieb JD, Sancar A. Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. Proc Natl Acad Sci USA 2016;113:E2124–33.
[44] Liu J, Liu L, He J, Xu Y, Wang Y. Multi-omic analysis of altered transcriptome and epigenetic signatures in the UV-induced DNA damage response. DNA Repair 2021;106.
[45] Mao P, Meas R, Dorgan KM, Smerdon MJ. UV damage-induced RNA polymerase II stalling stimulates H2B deubiquitylation. Proc Natl Acad Sci USA 2014;111:12811–6.
[46] Kolman A, Bohusova T, Lambert B, Simons JWIM. Induction of 6-thioguanine-resistant mutants in human diploid fibroblasts in vitro with ethylene oxide. Environ Mol Mutagen 1992;19:93–7.
[47] Lee KM, Choi KH, Ouellette MM. Use of exogenous hTERT to immortalize primary human cells. Cytotechnology 2004. doi:10.1007/10.1007/s10616-004-5123-3.
[48] Scherer WF, Syverton JT, Gey GO. Studies on the propagation in vitro of poliomyelitis viruses iv. Viral multiplication in a stable strain of human malignant epithelial cells (strain hela) derived from an epidermoid carcinoma of the cervix. J Exp Med 1953;97:695–710.
[49] Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol 2015. https://doi.org/10.1038/nrm3953
[50] Ritchie ME, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43.
[51] Pedregosa F, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12.
[52] Dale RK, Pedersen BS, Quinlan AR. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics 2011;27.
[53] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26.
[54] McKinney, W. Data structures for statistical computing in python. In: Proceedings of the 9th python in science conference; 2010. ⟨https://doi.org/10.25080/majora-92bf1922-00a⟩.
[55] Ye Tao, et al. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res 2011;39(6):e35-e35.
[56] Ramírez F, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 2016;44.
[57] Harris CR, et al. Array programming with NumPy. Nature 2020;585.
[58] Waskom M. seaborn: statistical data visualization. J Open Source Softw 2021;6.
[59] Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 2010;26.
[60] Canty A, Ripley B. boot: bootstrap R (S-Plus) functions. R package version 1.3-28; 2021. ⟨Http://Cran.R-Project.Org/Doc/Packages/⟩.
[61] Davison AC, Hinkley DV. Bootstrap methods and their application. Bootstrap methods and their application. 1997. https://doi.org/10.1017/cbo9780511802843
[62] Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. Signal Process 2020;167.
[63] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 1989;77(2):257–86. https://doi.org/10.1109/5.18626
[64] Virtanen P, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17.
[65] Davie JR, Murphy LC. Level of ubiquitinated histone H2B in chromatin is coupled to ongoing transcription. Biochemistry 1990. https://doi.org/10.1021/bi00472a002
[66] Pavri R, et al. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell 2006;125:703–17.
[67] Weinberger L, et al. Expression noise and acetylation profiles distinguish HDAC functions. Mol Cell 2012;47:193–202.
[68] Segala G, Bennesch MA, Pandey DP, Hulo N, Picard D. Monoubiquitination of histone H2B blocks eviction of histone variant H2A.Z from inducible enhancers. Mol Cell 2016;64:334–46.
[69] Segala G, Picard D. H2B monoubiquitination: t′ub or not t′ub for inducible enhancers. Transcription 2017;8:126–32.
[70] Jung I, et al. H2B monoubiquitylation is a 5′-enriched active transcription mark and correlates with exon-intron structure in human cells. Genome Res 2012;22:1026–35.
[71] Orphanides G, LeRoy G, Chang CH, Luse DS, Reinberg D. FACT, a factor that facilitates transcript elongation through nucleosomes. Cell 1998;92:105–16.
[72] Kim J, et al. RAD6-mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell 2009;137:459–71.
[73] Van Oss SB, et al. The histone modification domain of Paf1 complex subunit Rtf1 directly stimulates H2B ubiquitylation through an interaction with Rad6. Mol Cell 2016;64:815–25.
[74] Hou L, et al. Paf1C regulates RNA polymerase II progression by modulating elongation rate. Proc Natl Acad Sci USA 2019;116:14583–92.
[75] Nune M, et al. FACT and Ubp10 collaborate to modulate H2B deubiquitination and nucleosome dynamics. Elife 2019;8.
[76] Tufegd Zi C, Vidakovi A, et al. Regulation of the RNAPII pool is integral to the dna damage response. Cell 2020;180.
[77] Shema-Yaacoby E, et al. Systematic identification of proteins binding to chromatin-embedded ubiquitylated H2B reveals recruitment of SWI/SNF to regulate transcription. Cell Rep 2013;4:601–8.
[78] Henry KW, et al. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev 2003;17:2648–63.
[79] Smolle M, Workman JL. Transcription-associated histone modifications and cryptic transcription. Biochim Biophys Acta - Gene Regul Mech 2013;1829:84–97.

[80] Worden EJ, Hoffmann NA, Hicks CW, Wolberger C. Mechanism of cross-talk between H2B ubiquitination and H3 methylation by Dot1L. Cell 2019;176:1490–501. [.e12].
[81] Tanny JC, Erdjument-Bromage H, Tempst P, Allis CD. Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. Genes Dev 2007;21:835–47.
[82] Krajewski WA, Li J, Dou Y. Effects of histone H2B ubiquitylation on the nucleosome structure and dynamics. Nucleic Acids Res 2018;46:7631–42.
[83] Murawska M, et al. The chaperone FACT and histone H2B ubiquitination maintain S. pombe genome architecture through genic and subtelomeric functions. Mol Cell 2020;77:501–13. [.e7].
[84] Liu Y, et al. FACT caught in the act of manipulating the nucleosome. Nature 2020;577:426–31.
[85] Filipovski M, Soffers JHM, Vos SM, Farnung L. Structural basis of nucleosome retention during transcription elongation. Science (80-) 2022;376:1313–6.
[86] Kaplan CD, Laprade L, Winston F. Transcription elongation factors repress transcription initiation from cryptic sites. Science (80-) 2003;301:1096–9.