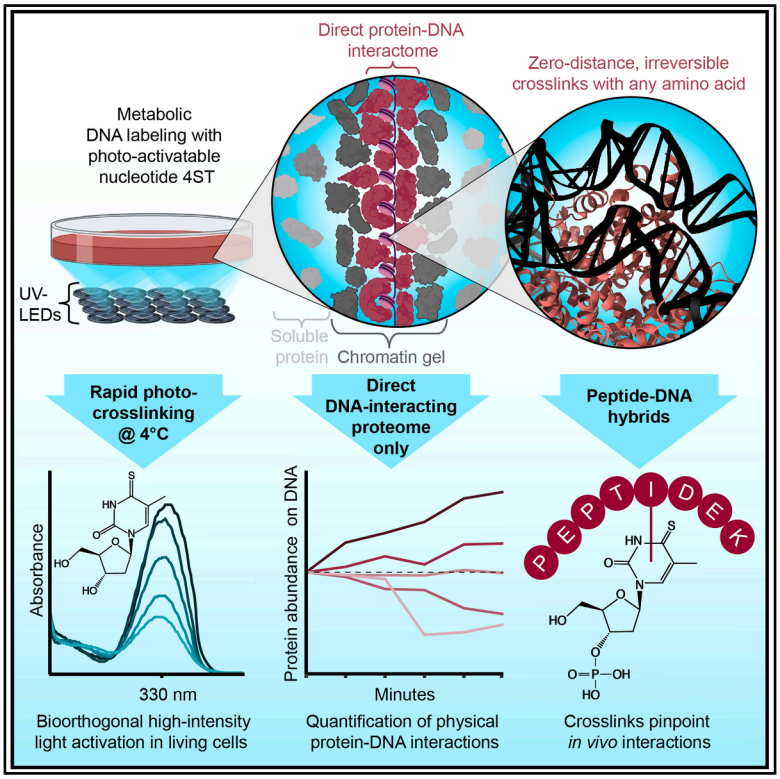
Cell

Resource

The human proteome with direct physical access to DNA

Graphical abstract



Highlights

- High-yield protein-DNA photo-crosslinking via metabolic labeling and UV-LED irradiation
- Widespread and extensive intrinsic disorder among proteins with physical access to DNA
- DNA-crosslinked peptides map *in vivo* DNA interactions with single-amino-acid resolution
- Differential quantification of DNA-bound transcription factors and chromatin organizers

Authors

Jakob Trendel, Simon Trendel, Shuyao Sha, ..., Stefan Kubicek, N. Henriette Uhlenhaut, Bernhard Kuster

Correspondence

kuster@tum.de

In brief

Zero-distance photo-crosslinking reveals direct protein-DNA interactions in living cells, enabling quantitative analysis of the DNA-interacting proteome on a timescale of minutes with single-amino-acid resolution.

Trendel et al., 2025, Cell *188*, 1–17
August 7, 2025
https://doi.org/10.1016/j.cell.2025.04.037

CellPress

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

Cell

CellPress
OPEN ACCESS

Resource

The human proteome with direct physical access to DNA

Jakob Trendel,[1] Simon Trendel,[6] Shuyao Sha,[1] Franziska Greulich,[2] Sandra Goll,[5] Susanne I. Wudy,[3] Karin Kleigrewe,[3] Stefan Kubicek,[5] N. Henriette Uhlenhaut,[2,4] and Bernhard Kuster[1,7,*]

[1]Proteomics and Bioanalytics, TUM School of Life Sciences, Technical University of Munich (TUM), Freising, Germany
[2]Metabolic Programming, TUM School of Life Sciences, ZIEL-Institute for Food & Health, Technical University of Munich (TUM), Freising, Germany
[3]Bavarian Center for Biomolecular Mass Spectrometry, TUM School of Life Sciences, Technical University of Munich (TUM), Freising, Germany
[4]Institute for Diabetes and Obesity (IDO) & Institute for Diabetes and Cancer (IDC), Helmholtz Center Munich (HMGU) and German Center for Diabetes Research (DZD), Neuherberg, Germany
[5]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria
[6]Munich, Bavaria, Germany
[7]Lead contact
*Correspondence: kuster@tum.de
https://doi.org/10.1016/j.cell.2025.04.037

SUMMARY

In a human cell, DNA is packed with histones, RNA, and chromatin-associated proteins, forming a cohesive gel. At any given moment, only a subset of the proteome has physical access to the DNA and organizes its structure, transcription, replication, repair, and other essential molecular functions. We have developed a "zero-distance" photo-crosslinking approach to quantify proteins in direct contact with DNA in living cells. Collecting DNA interactomes from human breast cancer cells, we present an atlas of over one thousand proteins with physical access to DNA and hundreds of peptide-nucleotide crosslinks pinpointing protein-DNA interfaces with single-amino-acid resolution. Quantitative comparisons of DNA interactomes from differentially treated cells recapitulate the recruitment of key transcription factors as well as DNA repair proteins and uncover fast-acting restrictors of chromatin accessibility on a timescale of minutes. This opens a direct way to explore genomic regulation in a hypothesis-free manner, applicable to many organisms and systems.

INTRODUCTION

What proteins interact with the genome determines the fate of a cell and often the fate of entire organisms.[1] DNA is embedded in a dense meshwork of protein and RNA generally referred to as chromatin, where protein-DNA interactions can occur directly, as, for example, in the case of histones, or indirectly, as in the case of histone-modifying complexes.[2] The overall composition of the subproteome with physical access to DNA is momentarily ill-defined. Proteomic assessments of the DNA interactome have so far been obscured by the gel-like nature of chromatin,[3] whose boundaries are fuzzy and usually defined by the method of fixation and isolation.[4] Previous studies interrogating chromatin composition in human cells have reported anywhere between 1,500 and 3,500 proteins.[5–8] Although condensed mitotic chromosomes can be purified for proteomic analysis because they form stable complexes,[9,10] covalent crosslinking is required to fix protein-DNA interactions of interphase chromosomes because their DNA is dispersed in the nucleoplasm and, therefore, cannot be purified as one physical entity. Proteins can be covalently crosslinked to nucleic acids, either by chemical crosslinking, typically with formaldehyde, or by ultraviolet (UV)-light-induced activation of the natural DNA bases. Formaldehyde crosslinking is highly effective for covalently fixing protein-DNA interactions; however, it also leads to indirect crosslinking of proteins with other proteins and RNA within the chromatin gel. This makes it impossible to differentiate direct from indirect interactions and creates a tripartite molecule that is notoriously hard to purify for proteomic analysis.[5–8] In contrast, photo-crosslinking only creates covalent connections between UV-activated nucleic acid bases and proteins physically interacting with them, and it has, therefore, been extensively applied to the proteomic interrogation of the RNA-interacting proteome.[11–16] UV-crosslinking of DNA, however, has been challenging due to its lower photo-reactivity.[17,18] Efforts to quantify proteins photo-crosslinked to DNA have thus struggled to obtain sufficient enrichment relative to non-crosslinked background proteins, limiting both sensitivity and dynamic range.[19,20] Similarly, studies focusing on nucleotide-peptide photo-crosslinks for pinpointing protein-DNA interaction sites have returned few successful identifications. Specifically, six photo-crosslinks mapping to ten highly abundant proteins have been reported from intact mouse

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

embryonic stem cells using a 258-nm laser[21] and 36 photo-crosslinks mapping to 27 proteins in HeLa nuclei using conventional 254 nm bulbs.[22] Thus, although the systematic quantification of direct protein interactions with the genome is highly desirable,[23] there is currently no adequate methodology to (1) effectively photo-crosslink proteins in physical contact with DNA and (2) to extract protein-crosslinked DNA with high purity.

Here, we present an optimized photo-crosslinking approach for interrogating direct protein-DNA interactions in living cells, addressing these challenges. We applied metabolic DNA labeling using the photo-activatable nucleotide 4-thiothymidine (4ST), followed by 365 nm photo-activation with a specialized high-intensity irradiation device based on light-emitting diodes (LEDs). For the proteomic analysis of photo-crosslinked protein-DNA complexes by liquid chromatography-mass spectrometry (LC-MS), we developed a multistep process of denaturing purifications, named XDNAX for "protein-crosslinked DNA extraction," which strongly enriched more than 1,800 proteins photo-crosslinked to DNA while eliminating most non-crosslinked proteins below the limit of detection. Serendipitously, we identified numerous nucleotide-crosslinked peptides in XDNAX samples without further enrichment, which we employed as additional evidence to characterize the protein-DNA interface. Across all presented experiments, we found 688 nt-crosslinked peptides mapping to 379 proteins from intact MCF7 cells—a 20-fold increase over previous reports.[21,22] By quantifying DNA interactomes differentially between treatments, we show that our photo-crosslinking approach can identify transcription factors responding to cellular perturbations in human cell lines, as well as primary mouse cells, and discover vulnerabilities in the DNA repair machinery of breast cancer cells treated with different genotoxic chemotherapeutics. Finally, we leverage the short crosslinking time of our approach to record a time course of DNA interactomes during the initial minutes of inhibiting the BRG1/BRM-associated factor (BAF) chromatin remodeling complex and discover ANP32 histone chaperones and other nucleosome remodelers rapidly responding. Overall, the presented DNA interactomes provide in-cell insights on direct protein-DNA interfaces, the mechanism of genotoxic drugs, and genomic regulation.

RESULTS

A high-powered UV irradiation system for the activation of photo-reactive nucleotides in living cells

To catalog and quantify human proteins accessing the genome, we devised a photo-crosslinking procedure that efficiently couples DNA of living cells to the proteins in its direct proximity. Previous studies using UV light in the range of 254 nm (UVC) for the photo-activation of natural DNA bases have detected very few DNA-crosslinked proteins, indicating that higher light intensities might be required. To increase crosslinking yields while avoiding unwanted photo-damage to DNA and protein, we decided to apply 365 nm light (UVA) in combination with metabolic labeling of the DNA with 4ST. Figure S1A shows that 365 nm light avoided much of the UV absorption in cells, presumably preventing photo-damage to proteins and fragmentation of natural DNA, which we intended to use as a purification handle. Because of their application in UV curing of resins and glues, very powerful 365 nm UV-LEDs have become commercially available, which rival light intensities of lasers despite much larger fields of irradiation. Although 365 nm light is not ideal for the photo-activation of 4ST, which has an absorption maximum of 330 nm (Figure S1A), LEDs with lower emission wavelengths currently exhibit much inferior intensities of less than 1% of the most powerful 365 nm LEDs. We constructed an irradiation device that employs an array of high-intensity 365 nm LEDs, able to evenly irradiate cell culture dishes of 15 cm in diameter (Figures 1A and 1B). The field of irradiation was more than 600 times larger than previous laser-based irradiation systems at very similar light intensities ($\sim$0.3 cm^2 for a 258-nm UV-laser system[24] vs. $\sim$175 cm^2 for our 365-nm UV-LED system, both at $\sim$2,000 mW/cm^2). In comparison to a conventional, bulb-based UV irradiation device, our system was able to emit three orders of magnitude more energy (Figure 1C). We named the device UVEN for "UV irradiation system for enhanced photo-activation in living cells" and benchmarked it for photo-crosslinking by irradiating aqueous solutions of 4ST or 4-thiouridine (4SU), which is commonly applied to enhance protein-RNA crosslinking in living cells.[12,25] Absorbance at 330 nm of either nucleotide decreased upon prolonged irradiation (Figure 1D), whereas no change was seen when they were irradiated with a conventional UV bulb (see also Figures S1B and S1C). MS indicated that, upon UVEN irradiation and with the loss of hydrogen sulfide, 4ST and 4SU were quickly converted to thymidine and uridine[26] (Figure 1E), respectively, which was also evident from the gain in 254 nm absorption and drop in pH (Figures S1C and S1D). Thus, UVEN accelerated the photo-activation approximately 100-fold in comparison to a standard device. Analogous to naturally occurring nucleotides, we found 4SU about 10 times more reactive than the deoxyribonucleotide 4ST. Metabolic labeling of expanding cancer cells with 4ST has been explored for photodynamic therapy in cell culture and a rat xenograft model, where UVA irradiation killed 4ST-labeled cells but not unlabeled cells.[27–29] We exploited this known photo-sensitization to demonstrate that UVEN could activate 4ST in living cells by monitoring their expansion with live-cell imaging (Figure 1F). After labeling with more than 100 μM 4ST, just 6 s of UVEN irradiation was enough to halt proliferation. Unlabeled control cells, however, continued proliferation even after 60 s of UVEN irradiation, indicating that (1) natural DNA remained mostly unharmed by light at this wavelength and (2) cells were alive throughout the irradiation. Comparing proliferation of expanding MCF7 (human female breast adenocarcinoma) or U2OS (human female osteosarcoma) cells, we found a half-inhibitory concentration (IC_{50}) between 1 and 2.5 mM for 4ST, which was much better tolerated than 4SU with an IC_{50} between 100 and 200 μM (Figure S1E). For our subsequent photo-crosslinking experiments, we chose a concentration of 100 μM 4ST for labeling cells before 1 min of UVEN irradiation in ice-cold PBS to "freeze" interactions and suppress responses to the emerging crosslinks.

Extraction of protein-crosslinked DNA from photo-sensitized cultured cells

To isolate photo-crosslinked protein-DNA complexes for LC-MS analysis, we developed a denaturing purification protocol that eliminates non-crosslinked proteins from DNA, enriching only

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

CellPress
OPEN ACCESS

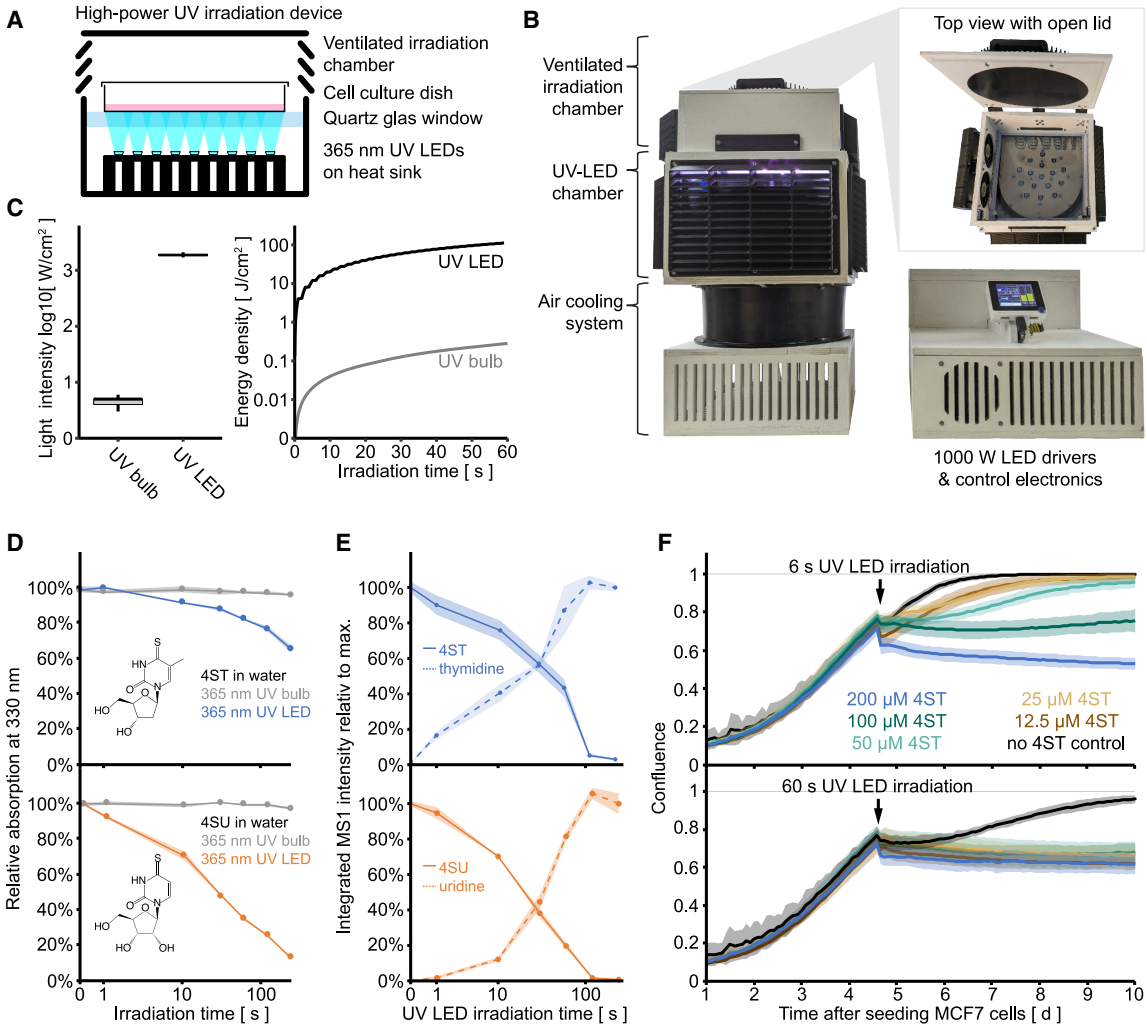


Figure 1. Photo-activation of the UV-reactive deoxyribonucleotide 4ST in living cells

(A) Schematic representation of the high-intensity UV-irradiation device UVEN.

(B) Photos of the actual device.

(C) Comparison of UV emissions by a conventional 365-nm UV-bulb-based device (UV bulb) and UVEN (UV LED).

(D) Time series displaying 330 nm absorption measurements of 4ST (top) or 4-thiouridine (4SU, bottom) irradiated in water for indicated time (note logarithmic scaling). Displayed are means of triplicates with shaded areas indicating one standard deviation. See also Figures S1B and S1C.

(E) Time series showing the formation of thymidine from 4ST (top) or uridine from 4SU (bottom) upon irradiation in water, as detected by LC-MS. Displayed are means of triplicates with shaded areas indicating one standard deviation.

(F) Line plot showing the confluence of MCF7 cells grown in the presence of the indicated 4ST concentration, irradiated for 6 s (top) or 60 s (bottom), and returned to culture in fresh medium. Displayed are means of quintuplicates with shaded areas indicating one standard error of the mean.

See also Figure S1.

DNA-crosslinked proteins with direct genome access at the time of irradiation (Figure 2A). During conventional thiocyanate-phenol-chloroform (TRIZOL) extraction, chromatin will collect in the interphase and can be recovered as a cohesive film, removing much of the non-crosslinked proteins, RNA, lipids, and other cellular debris from the other phases.[14,30] We treated the interphase with an additional iteration of the TRIZOL procedure to remove trapped, non-crosslinked protein[15] and then resolubilized it for RNase digestion. The resulting extract of protein-crosslinked DNA was sheared by ultrasonication, denatured by boiling in guanidinium thiocyanate, and vigorously washed on silica spin columns to remove non-crosslinked protein to completion. Photo-crosslinked protein was released from DNA by nuclease treatment and trypsin-digested for LC-MS analysis. We named this method "protein-crosslinked DNA extraction" or XDNAX (for details, see STAR Methods).

To evaluate XDNAX, stable isotope labeling in cell culture (SILAC) was applied[31] (Figure 2B). SILAC-labeled MCF7 cells were expanded for 5 days in the presence of 4ST. SILAC heavy cells were irradiated with UVEN, whereas SILAC light cells were left unirradiated. The cells were combined and subjected to XDNAX so that the MS signal in the heavy SILAC channel

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

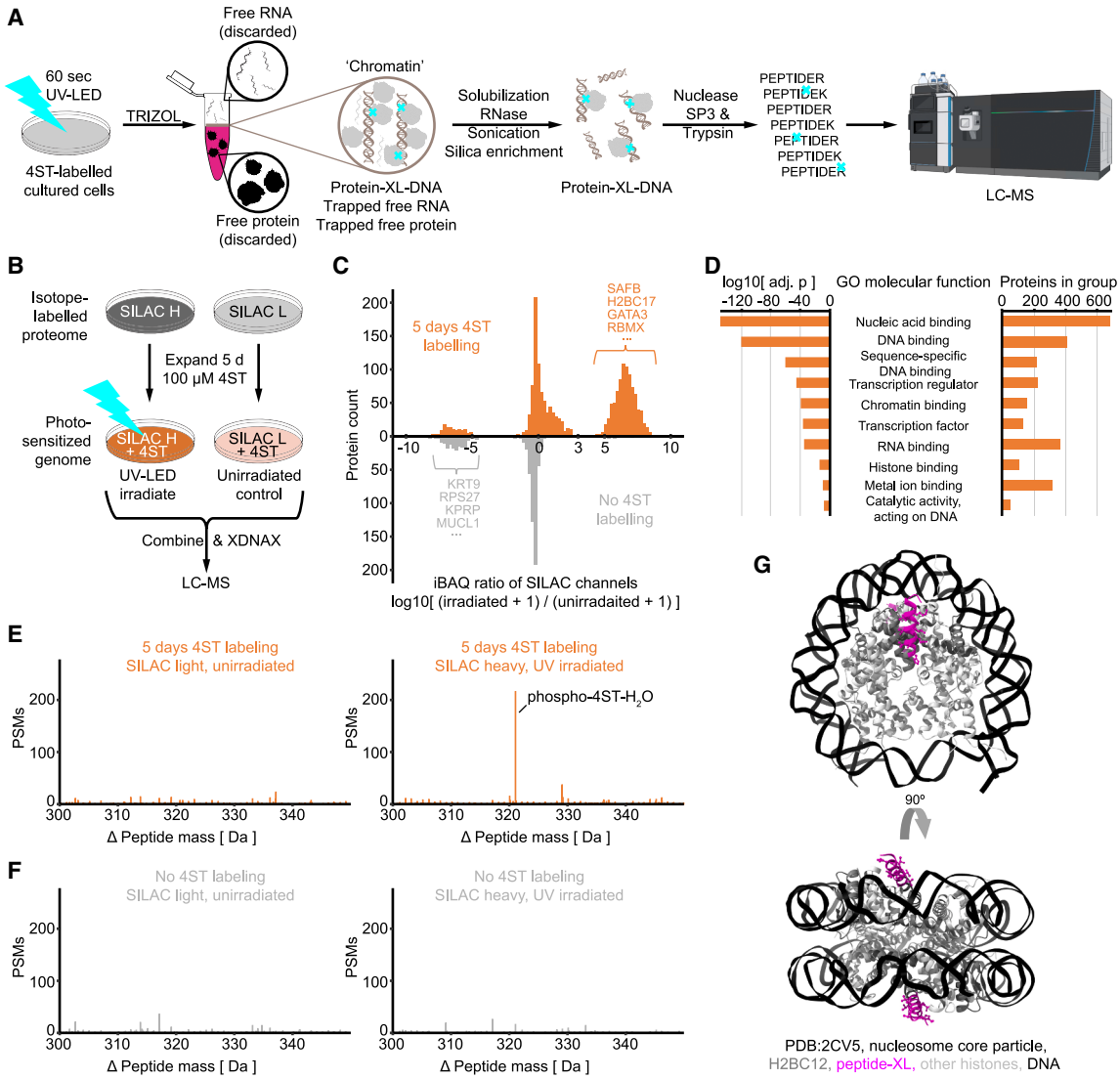


Figure 2. Extraction and proteomic analysis of protein photo-crosslinking to DNA of living cells

(A) Workflow for protein-crosslinked DNA extraction, or XDNAX.

(B) Experimental scheme for the validation of XDNAX using SILAC. For details see text.

(C) Histogram illustrating the enrichment of photo-crosslinked protein extracted by XDNAX from 4ST-labeled MCF7 cells (orange, top) or unlabeled control cells (gray, bottom). Most abundant proteins in indicated groups are given as examples.

(D) Bar plot showing enrichment and number of proteins under ten selected GO terms. All proteins in the heavy SILAC channel from (C) were compared with a deep MCF7 full proteome.

(E) Histogram of peptide-spectrum matches (PSMs) identified in an open search on the data (C) (orange, top) by the mass-tolerant search engine MSfragger. For visibility only a Δ mass window around the 4ST modification is shown (also see Figure S2D).

(F) Same as in (E) but for control XDNAX experiment omitting 4ST labeling.

(G) Crystal structure of the human nucleosome core particle (PDB: 2CV5) illustrating the position of a nucleotide-crosslinked peptide (magenta) relative to the DNA (black).

See also Figure S2.

indicated protein after photo-crosslinking, whereas the signal in the light SILAC channel indicated a non-crosslinked background. Indeed, our workflow resulted in strong enrichment of proteins from irradiated cells. Of the 1,803 quantified proteins, 1,026 were enriched by more than 10-fold, and 876 proteins were detected exclusively in the SILAC channel corresponding to irradiated cells (Figure 2C, upper). Moreover, when we omitted the addition of 4ST and performed the same experiment using unlabeled cells, no enrichment was observed (Figure 2C, lower). Gene Ontology (GO) analysis comparing proteins exclusively found in the irradiated SILAC channel after XDNAX against a deep, full proteome of the same cell line confirmed strong

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

enrichment for proteins annotated as nucleic acid binding (p = 1E−149) and, more specifically, DNA binding (p = 1E−120) (Figure 2D). As expected, intensity-based quantification of this SILAC channel showed especially histones being among the most abundant proteins, followed by splicing factors and chromatin organizers (Figure S2A). We performed XDNAX with and without RNase digestion and observed no systematic differences in the purification of proteins annotated as DNA or RNA binding (Figure S2B). This validated the specificity of our approach, which metabolically incorporates a photo-label only into DNA, preventing simultaneous crosslinking of RNA that is seen with 254 nm UV irradiation.[32] RNase-digested samples contained higher amounts of proteins annotated as glycosylated, which has been previously observed with other TRIZOL-based purification methods.[15] In the case of XDNAX, this indicated that glycoproteins were retained in the TRIZOL interphase and, in addition, better retained on silica spin columns in the absence of competing RNA. Importantly, our SILAC-controlled experiment demonstrated that the glycoprotein enrichment was independent of photo-activation (Figure S2C), indicating that glycoproteins constitute a considerable (∼10% of protein identifications, ∼10% of total intensity-based absolute quantification [iBAQ]) but constant background of XDNAX (see also Tables S3, S4, and S5). To investigate post-translational modifications on the crosslinked proteins we applied the mass-tolerant search engine MSfragger and found that phosphorylation as well as ubiquitination were well preserved after XDNAX (Figure S2D). To our great interest, we found an additional modification of 321 Da occurring only in the irradiated SILAC channel (Figures 2E and S2D). The modification did not occur when 4ST-labeling was omitted (Figures 2F and S2E) and corresponded exactly to the mass of phosphorylated 4ST minus a water molecule, a loss previously observed for uridine-crosslinked peptides in the context of protein-RNA crosslinking.[14,33] Two of the most frequently observed peptide-4ST hybrids were derived from the core histone H2BC12 (Figure 2G) and the structural protein HMGB1 (Figure S2F), for which co-crystal structures complexed with DNA were available that illustrated the interface between the nucleotide-crosslinked peptide and DNA.

An atlas of proteins with direct access to DNA

In order to take inventory of proteins directly interacting with DNA, we applied our photo-crosslinking approach to five replicates of MCF7 cells. We applied the SILAC-controlled setup displayed in Figure 2B and used a stringent, sample-specific enrichment cutoff to call DNA interactors (Figure S3A; for details, see STAR Methods). This discovered 1,805 candidate proteins, 1,191 of which were present in three or more replicates (Figure S3B; Table S1). We found a large overlap of our DNA interactome with a deep nuclear proteome from MCF7 cells (83%; Figure S3C; see STAR Methods for details) and annotations from ProteinAtlas or OpenCell for nuclear localization (Figure S3D). Figure 3A shows that 60% of the DNA interactome carried a GO annotation for nucleic acid binding. To understand the contribution of specific protein classes to the DNA interactome, we used iBAQ, which estimates protein copy numbers in a sample based on combined peptide MS intensity.[34] First, we compared our photo-crosslinked DNA interactome to formaldehyde-crosslinked chromatin isolated by chromatin enrichment for proteomics (CHEP[5]). This revealed a Pearson correlation of R^2 = 0.6 (median replicate correlation XDNAX R^2 = 0.90, CHEP R^2 = 0.91) indicating that UV and formaldehyde-crosslinked DNA interactomes contained distinct information (Figure S3E; Table S1). Notably, histones had higher absolute abundance in formaldehyde-crosslinked chromatin, even though in the UV-crosslinked DNA interactome they still constituted the single largest iBAQ contribution among proteins with a common molecular function (Figures 3A and S3F). Among proteins lacking an annotation for nucleic acid binding, the largest contribution came from lamins, followed by proteins involved in chromatin organization and cell division (Figures 3A and S3G). Additionally, we identified a group of 278 proteins that could not be assigned to any of the main GO terms. For example, we found proteins related to the ubiquitin-proteasome system such as PSM3, which promotes degradation of proteins involved in cell-cycle progression, apoptosis, and DNA repair.[35,36] In addition, the group contained a number of uncharacterized proteins, one of which was C5orf24, whose entire sequence constitutes a single domain of unknown function. DUF5568 is conserved across chordates and mostly found in single-domain proteins, yet, in rare cases, in combination with a DEAD-box helicase domain.[37] Much of C5orf24 is predicted to be disordered (88% of amino acids with IUPRED score > 0.5) and AlphaFold shows very little tertiary structure for the human protein or its homologs in mice, fish, or frogs (Figure S3H). This suggested that DUF5568 might constitute an unknown, highly disordered "DNA-binding domain"—in principle, a conserved DNA-binding intrinsically disordered region (IDR). Comparing relative protein abundances between the MCF7 DNA interactome and the nuclear proteome, some proteins ranked much higher than expected from their abundance in the nucleus (Figure 3B). For example, among the 117 transcription factors in our DNA interactome, NACC1, CUX2, RARA, SOX4, and THAP10 appeared to have particularly good access to DNA. Indeed, transcription factors with higher rank in the DNA interactome than expected from the nuclear proteome (Figure 3B) had much higher occurrence of their binding motifs in MCF7 DNA accessibility sequencing data from ENCODE (Figure 3C), supporting the notion that transcription factor activity is highly regulated and often cannot be predicted from protein expression.[38]

Structural properties of proteins in direct contact with DNA

Chromatin has been shown to be solid-like in living cells, providing a nucleosome scaffold for protein and RNA that creates an elastic gel.[3] We were interested to see whether there were common sequence features among proteins in the DNA interactome that might grant them direct DNA access within this gel. When comparing the occurrence of domains in the MCF7 DNA interactome to their occurrence in the MCF7 full proteome, we found strong enrichment for canonical DNA and histone-binding domains, such as the PHD-type and C2H2-type zinc finger, the SAP domain, and DEAD/DEAH box helicase, as well as the chromodomain, bromodomain, the SANT domain (Figure S3I). Interestingly, the RNA recognition motif (RRM) domain was also strongly enriched in the DNA interactome. We searched for

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

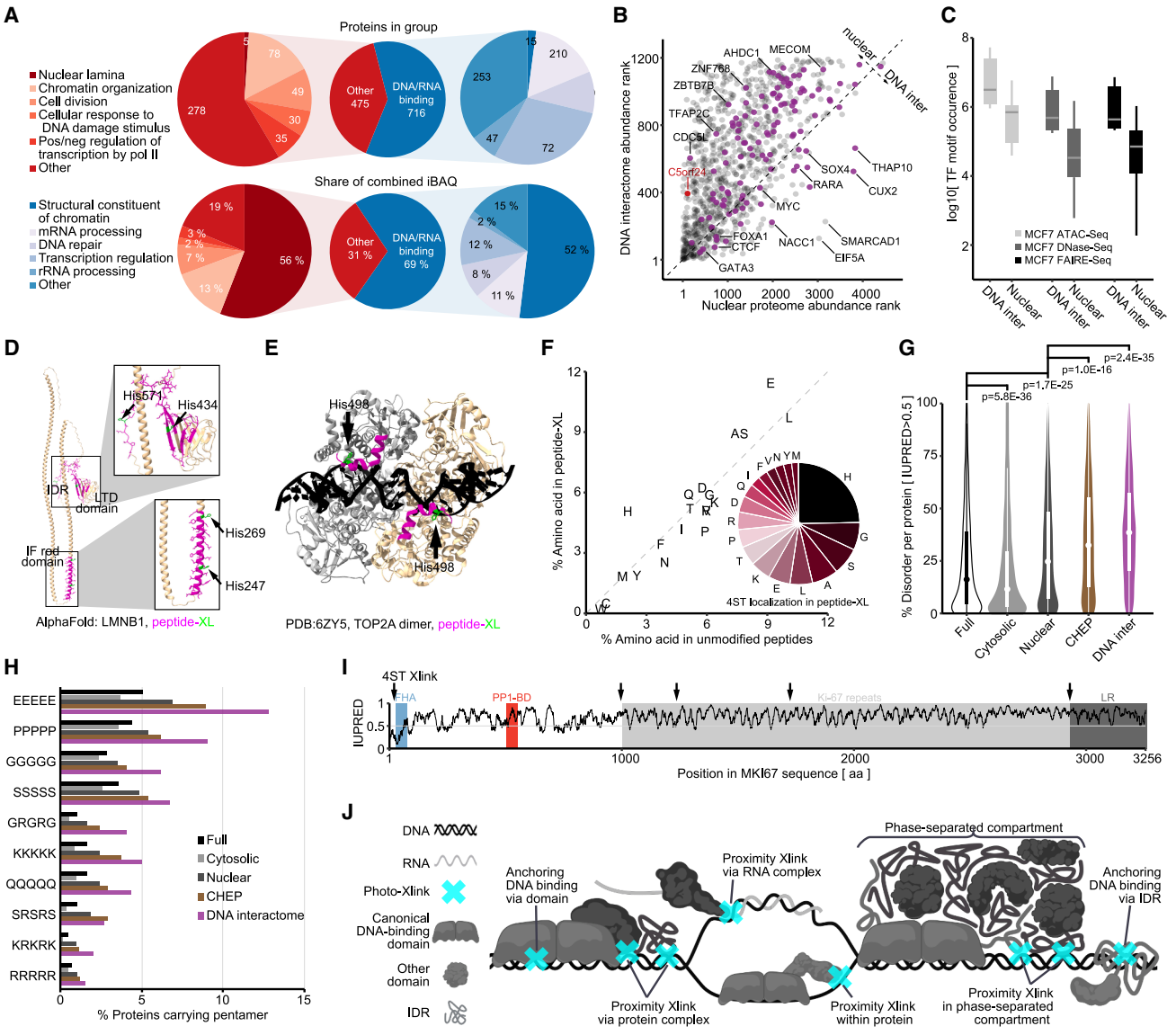


Figure 3. A census of proteins, domains, and IDRs with direct access to DNA

(A) Pie diagrams characterizing the constituents of the direct DNA interactome from MCF7 cells (see also Figures S3A and S3B).

(B) Scatterplot comparing abundance ranks of proteins within the MCF7 nuclear proteome to their relative abundance in the DNA interactome.

(C) Boxplots comparing occurrence of binding motifs in DNA accessibility sequencing from ENCODE. Compared are transcription factors with higher (DNA interactome) or lower (nuclear) abundance rank in the DNA interactome than expected from the nuclear proteome (see diagonal in B).

(D) AlphaFold structure prediction for LMNB1 displaying DNA crosslinks identified by 4ST-modified peptides (magenta, histidine crosslinking site in green).

(E) Cryo-electron microscopy structure of the TOP2A DNA-binding/cleavage domain dimer in complex with its DNA (black) substrate (PDB: 6ZY5).

(F) Scatterplot and pie chart comparing amino acid frequencies in unmodified to nucleotide-crosslinked peptides found in the MCF7 DNA interactome.

(G) Violin plots comparing percentage of protein disorder between different parts of the MCF7 proteome (IUPRED > 0.5). Testing occurred with a two-sided Kolmogorov-Smirnov test. See also Figure S3K and text for details.

(H) Bar plot comparing the occurrence of the most abundant disordered amino acid pentamers between the same groups as in (G). See also Figure S4M.

(I) Line plot showing IUPRED scores across MKI67. Arrows indicate DNA crosslinks (4ST Xlink).

(J) Schematic summary of protein-DNA interactions observed among proteins with direct access to DNA.

See also Figure S3.

4ST-crosslinked peptides and could confirm crosslinking sites in the RRM domains of MATR3, TRA2B, SRSF2, SRSF11, and RNPS1, implying that the RRM can, in fact, interact with DNA in living cells. Overall, we found one-third of the DNA interactome overlapping with an MCF7 RNA interactome derived in a previous study (Figure S3D).[14] Because many of these proteins are annotated as nucleolar, we suspect that some might interact with DNA circumstantially within phase-separated compartments.[39]

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

Overall, 162 4ST-crosslinked peptides from 83 proteins could be identified in the MCF7 DNA interactome (Table S2). The 4ST-modified peptides with the highest intensities mapped to LMNB1 (Figures 3D and S3J). We identified a 4ST-modified histidine in the DNA-binding domain of TOP2A, for which a co-crystal structure with DNA was available, visualizing the direct interface (Figure 3E). All 4ST modifications in LMNB1 were also localized to histidines (Figure 3D), as were 25% of the other identified crosslinks, followed by glycine, serine, and alanine (Figure 3F). Only 61 of the 162 4ST-modified peptides mapped to annotated domains, while many were located in IDRs (Table S2). To understand what role protein disorder played in the DNA interactome, we used the IUPRED score[40] and AlphaFold's confidence score pLDDT.[41] Both were in good agreement that proteins in a deep nuclear MCF7 proteome were significantly more disordered than the entire MCF7 full proteome (two-sided Kolmogorov-Smirnov test, $p < 7.1E{-}23$; Figures 3G and S3K), and proteins in the deep cytosolic proteome were significantly more folded ($p < 1.2E{-}34$). Compared with the nuclear proteome, proteins in formaldehyde-crosslinked chromatin were even more disordered (CHEP, $p = 1.0E{-}16$), and proteins within the photo-crosslinked DNA interactome were the most disordered ($p = 2.4E{-}35$). In the median, the combined length of disordered positions within one protein was 49 amino acids in the cytosol, 121 in the nucleus, 154 in formaldehyde-crosslinked chromatin, and 233 in our physical DNA interactome. Interestingly, the amino acid composition of proteins in the DNA interactome was virtually identical to the composition of proteins in the MCF7 full proteome (Figure S3L). We counted the occurrence of pentamer sequences containing any permutation of the amino acids G, S, D, Q, P, E, K, and R, typically associated with protein disorder.[42] Indeed, this resulted in the discovery of homopolymeric sequences of E, P, S, G, K, and Q, as well as RS and RG-repeats strongly enriched in the DNA interactome compared with the MCF7 full proteome (Figure S3M). The occurrence of these repeats again tracked with the proximity of proteins to DNA (Figure 3H). In combination, the ten repeat sequences displayed in Figure 3H occurred in 36% of proteins in our DNA interactome and only 13% of the MCF7 cytosolic proteome. By far the most frequent homorepeats in our DNA interactome were stretches of poly(E), whose occurrence in proteins involved with chromatin organization and transcription is conserved from yeast to humans.[43–45] Among disordered motifs unique to a single protein, the Ki-67 repeats of the highly disordered MKI67 stood out (85% of amino acids with IUPRED score > 0.5), to which several crosslinks with very high MS intensity could be mapped (Figures 3I and S3J). Figure 3J summarizes the protein-DNA interfaces implied by our DNA interactome and the nucleotide-crosslinked peptides it contained. This included DNA interaction due to (1) globular DNA-binding domains, (2) DNA-binding IDRs, and (3) circumstantial proximity because of interaction with other proteins (e.g., histone-binding proteins), RNA (e.g., splicing factors), or within phase-separated compartments (e.g., nucleolar proteins).

Quantification of transcription factor binding to DNA during cellular perturbations

The estrogen receptor alpha (ESR1) resides in the cytosol until it encounters its ligand, triggering translocation into the nucleus where it engages specific regions in the genome to activate a specific transcriptional program.[46] We were interested to see whether, by quantitatively comparing DNA interactomes between untreated and stimulated cells, we could recapitulate the activation of individual transcription factors. Hence, serum-starved MCF7 cells were exposed to a high dose of estrogen (10 nM 17β-estradiol) for 45 min and their DNA interactome compared with mock-treated control cells (Figure 4A). Strikingly, we observed a 30-fold increase of ESR1 in the DNA interactome after estrogen stimulation (Figure 4B; Table S3), making it the most significantly changed protein in the differential analysis (adj. $p = 2.6E{-}42$, negative binomial model [NBM] testing; see STAR Methods for details). In total, around 9% of the 1,814 quantified protein-DNA interactions showed significant changes exceeding 2-fold (adj. $p < 0.01$). We observed increased DNA binding of known ESR1 interactors, such as RNA polymerase II (POLR2A, adj. $p = 2.1E{-}6$; Figure 4B) and SMARCA4 (adj. $p = 2.7E{-}5$), as well as members of other transcriptional coactivator complexes, including BPTF (adj. $p = 0.004$), CHD1 (adj. $p = 0.004$), and CHD6 (adj. $p = 1.0E{-}5$). The transcription repressors SET, CBX3, and CBX5 showed the largest absolute loss in protein abundance on DNA (Figure 4C; Table S3). CBX3 and CBX5 promote heterochromatin formation,[47] whereas SET represses histone acetylation,[48] suggesting that their estrogen-induced release prepared decondensation of chromatin in the serum-starved cells. The most substantial absolute increase in protein abundance was observed for the nucleolar transcription factor UBTF, indicating strong stimulation of rRNA transcription.[49] Increases were also observed for SUMO2 and SUMO3, along with highly significant recruitment of the SUMO-specific protease SENP3 (Figure 4B). As sumoylation inhibits ESR1 binding to its target genes,[50] SENP3 at the DNA surface might counteract this inhibition to amplify the estrogen response.

We treated MCF7 cells with eleven additional concentrations of estrogen, ranging from 10 nM to 30 fM for 45 min before quantifying their photo-crosslinked DNA interactome.[51,52] Applying a log-logistic dose-response model, we were able to derive half-effective concentrations (EC_{50}) for the engagement of 1,483 proteins with DNA (Table S3). The best fit in the relevant concentration range was for ESR1 ($R^2 = 0.94$; Figure 4D), showing an EC_{50} of 35 pM for its engagement with DNA, which is close to the K_d of 50 pM reported from a cell-free radioligand assay.[53] We observed numerous proteins changing their interaction with DNA at similar or slightly higher effective concentrations than ESR1, at around 100 pM estrogen, such as the chromatin organizers TRIM33 and TRIM24 (Figure 4E). A second group of proteins with approximately 10-fold-higher effective concentrations (Figure 4F, class II) included genes known to be transactivated by ESR1, suggesting that estrogen had increased their expression, leading to increased binding of their protein product to DNA. Across all samples in the estrogen series, we identified 346 4ST-crosslinked peptides from 255 proteins (Table S2), some of which validated the estrogen-induced changes we had observed on the protein level (Figure 4G).

We asked whether this analysis was also possible in primary cells and expanded mouse bone-marrow-derived macrophages *ex vivo* for 3 days in the presence of 4ST before photo-crosslinking and XDNAX (Figure S4A). Indeed, comparison of the DNA interactome from untreated cells to cells challenged for 3 h with

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

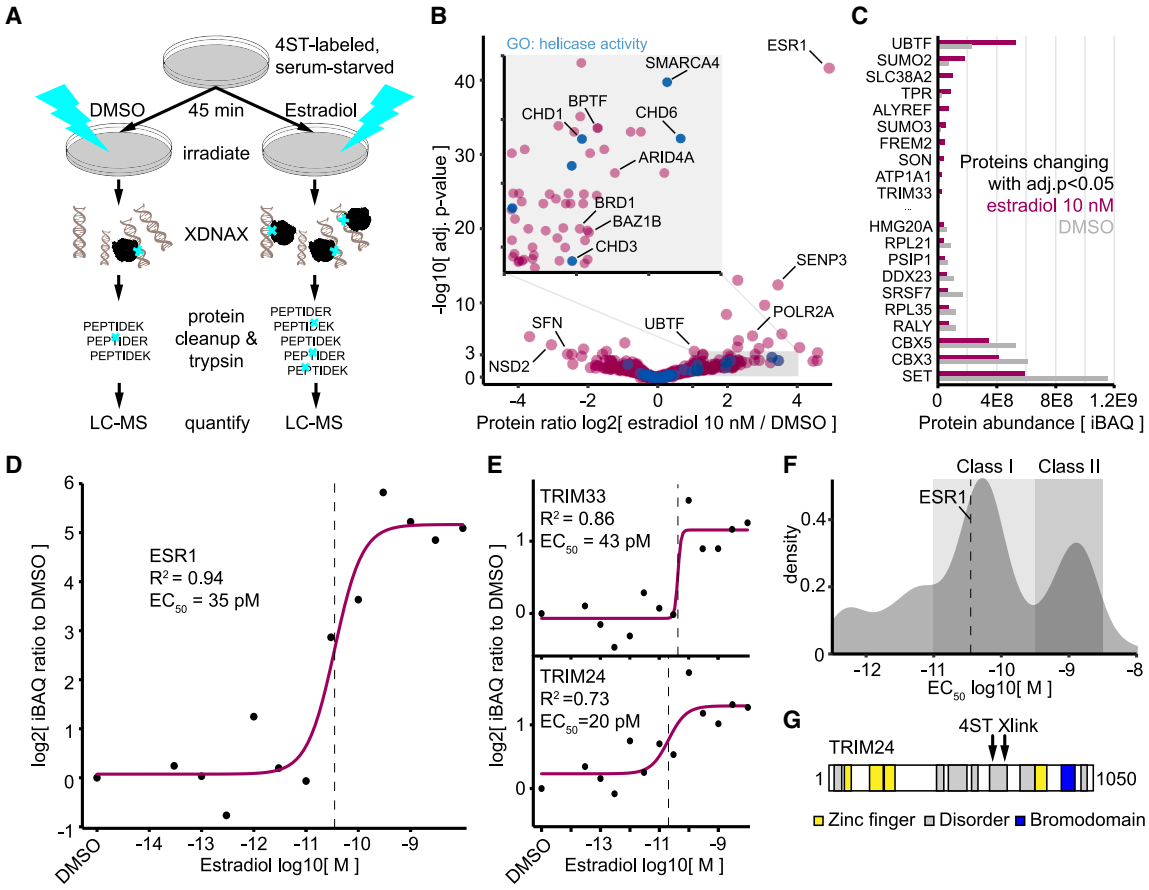


Figure 4. Quantitative comparisons of DNA interactomes from breast cancer cells treated with estrogen

(A) Experimental workflow for the comparison of proteins interacting with DNA after 45 min of estrogen treatment in 48-h serum-starved MCF7 cells.

(B) Volcano plot illustrating changes in the DNA interactome after 45 min of high-dose estrogen exposure (10 nM 17β-estradiol). Testing was performed with a negative binomial model; see STAR Methods for details.

(C) Bar plot comparing the most extreme absolute changes in the abundance of proteins interacting with DNA in the same experiment as (B).

(D and E) Exemplary dose-response curve derived from the relative protein abundance in the DNA interactome of MCF7 cells at increasing estrogen concentrations.

(F) Density plot of half-effective concentrations for the association of proteins with DNA during estrogen exposure.

(G) Schematic structure of TRIM24 indicating estrogen-induced DNA crosslinks.

See also Figure S4.

lipopolysaccharide (LPS) indicated strong engagement of Cebpb (Figure S4B; adj. p = 2.5E−10) and Cebpd (adj. p = 2.9E−6), key transcription factors for macrophage activation, as well as other inflammation modulators such as Tnfaip2 (adj. p = 3.63E−8) and the nuclear factor κB (NF-κB) subunit Rela (adj. p = 0.01). Chromatin immunoprecipitation sequencing (ChIP-seq) data of Cebpb, Cebpd, and Rela confirmed the expansion of their cistromes in proximity to pro-inflammatory genes upon LPS stimulation (Figures S4C–S4E).

Genomic regulation in response to three clinically relevant genotoxic drugs

In addition to understanding how the genome is read and regulated, there is great interest in understanding how proteins organize DNA maintenance and repair. Genotoxic drugs form an important pillar of clinical pharmacology aiming to halt the growth of fast-dividing cells by damaging their DNA.[54] We considered three commonly applied genotoxic compounds—etoposide, cisplatin, and oxaliplatin—which are well-characterized and known to act by distinct modes of action.[54,55] Live-cell imaging indicated that all drugs reached their maximal potential to halt proliferation and induce apoptosis in MCF7 cells 24 h after addition (Figure S5A). We derived DNA interactomes and quantified them against mock-treated control cells (Figure S5B). Common to all treatments were strongly increased DNA interactions of central cell-cycle regulators such as CDKN1A (also known as p21; Figure S5C) and DNA repair proteins such as TP53BP1, SIRT7, and, in particular, HINT1 (Figures 5A–5C; Table S4). The most significant loss in DNA interaction across all treatments was observed for NHP2, a component of the telomerase complex, which was only detected on DNA in untreated cells, indicating that any genotoxic drug treatment led to

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

CellPress
OPEN ACCESS

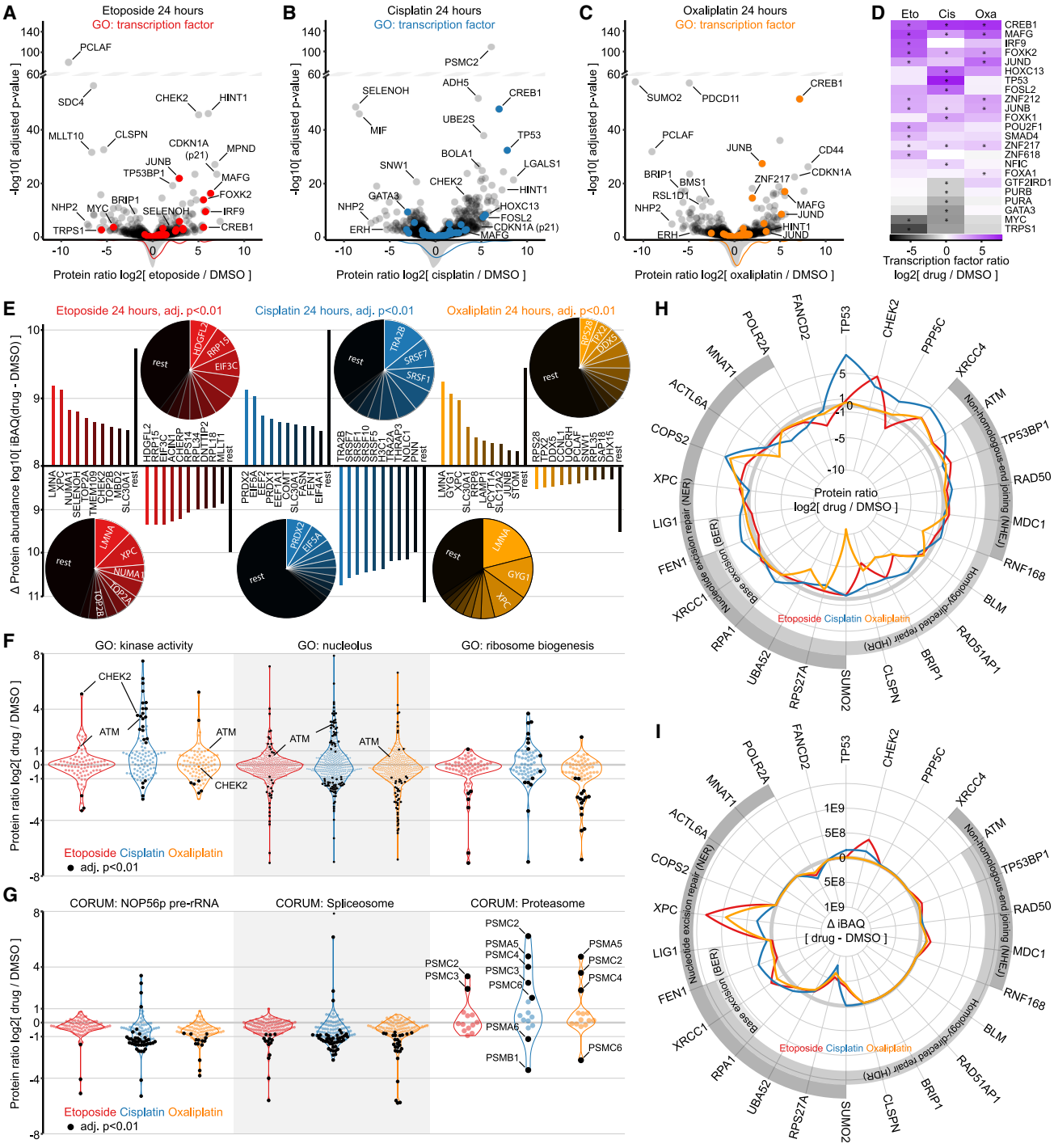


Figure 5. Changes in the direct DNA interactomes of breast cancer cells under genotoxic drug treatment

(A–C) Volcano plots showing differences in the DNA interactomes of 24-h mock-treated MCF7 cells (DMSO) to cells treated with etoposide, cisplatin, or oxaliplatin at 100 μM concentration. Testing was performed with a negative binomial model; see STAR Methods for details.

(D) Heatmap comparing significant (adj. $p < 0.01$ indicated by *) fold changes of transcription factor-DNA interactions between treatments.

(E) Bar plots showing the top ten absolute abundance differences of significantly changing proteins (adj. $p < 0.01$). Pie diagrams illustrate relative contributions to the total iBAQ change.

(F) Beeswarm plot comparing changes in the DNA interactomes for selected GO terms.

(legend continued on next page)

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

decreased telomerase activity and contributed to damage-induced senescence.[56] The strongest decreases in DNA interaction for transcription factors were observed for TRPS1 and MYC (Figure 5D), whose expression is known to be reduced upon genotoxic stress in MCF7 and other cell lines.[57] Under cisplatin, we also observed strongly reduced DNA interaction of the metastasis marker GATA3[58] and increased interaction of TP53. In mock-treated cells, TP53 was undetectable on DNA, ranked in the bottom 30% of protein abundances under etoposide or oxaliplatin treatment but in the top 20% of protein abundances under cisplatin stress (Table S4). We observed large amounts of peroxiredoxins accumulating on DNA, specifically in cisplatin-treated cells, where PRDX1 and PRDX2 combined accounted for more than 10% of the entire gain in iBAQ relative to the mock-treated control (Figure 5E). Intriguingly, under conditions of oxidative stress, PRDX2 forms a decameric toroid,[59,60] which, by proportion, would be large enough to accommodate the DNA double helix (Figure S5D). Principal-component analysis (PCA) indicated that protein abundances in the DNA interactomes from cells treated with etoposide or oxaliplatin were more similar than from cells treated with cisplatin (Figure S5E). However, the recruitment of DNA damage response proteins was quite distinct between all treatments. For instance, the central damage-relaying kinases ataxia telangiectasia mutated (ATM) and CHEK2 were strongly recruited to DNA in response to cisplatin, only CHEK2 in response to etoposide, and neither in response to oxaliplatin (Figure 5F). Short-term oxaliplatin treatment in the range of several hours has been shown to cause ribosome biogenesis stress,[55] whereas longer treatment causes DNA damage and disintegration of nucleoli.[61,62] Indeed, after 24 h of oxaliplatin, we observed extensive loss of nucleolar proteins from the DNA interactome (adj. p = 6.4E−9, two-sided Kolmogorov-Smirnov test with Bonferroni-Holm correction), including many proteins involved in ribosome biogenesis (adj. p = 4.3E−4). Under both platinum drugs, the largest CORUM-annotated protein complex with reduced DNA association was the nucleolar pre-ribosomal NOP56p complex (Figure 5G). Moreover, various components of the proteasome complex showed increased association with DNA, supporting previous reports about the removal of drug-induced protein-DNA adducts.[63] We focused on the best-populated DNA repair pathways in the DNA interactomes and compared them directly (Figures 5G and 5H). In line with previous reports on the repair of etoposide-induced damage,[64,65] we observed signatures for DNA repair via non-homologous end joining (NHEJ), base-excision repair (BER), and nucleotide-excision repair (NER) pathways (Figure 5H). TOP2A and TOP2B were among the ten proteins with the largest absolute gains in protein abundance (Figure 5E). Active removal of topoisomerases trapped on DNA has been shown to be triggered in MCF7 cells by the SUMO-ligase PIAS4, followed by ubiquitination and proteasomal degradation.[66] We derived additional DNA interactomes from cells treated for 4 h with etoposide that showed 6-fold TOP2A and TOP2B accumulation (Figure S5F), strong TOP2A phosphorylation (Figure S5G), and significant recruitment of PIAS4 (adj. $p < 0.002$; Figure S5H), suggesting that active topoisomerase removal might be stronger during early etoposide treatment while cells are still proliferating (Figure S5A). Cisplatin treatment led to a marked increase in DNA interaction of the NHEJ marker XRCC4, the ATM-activating phosphatase PPP5C, ATM itself, CHEK2, and several other kinases (Figure 5F). Next to PPP5C, two other druggable DNA repair enzymes showed notable enrichment on DNA under cisplatin: FEN1 and the exonuclease REXO4 (Table S4). Knockdown of FEN1 and PPP5C strongly increased apoptosis (Figures S5J and S5L) and cell death (Figures 5K and S5M) at the cisplatin concentration used to derive our DNA interactomes, whereas REXO4 depletion led to desensitization at higher cisplatin concentrations (Figure S5M). This confirmed that DNA interactomes can be leveraged to compare DNA-damaging drugs mechanistically and identify potential targets for genotoxic combination treatments.

Acute changes in the DNA interactome during restriction of DNA accessibility by BAF inhibition

Chromatin accessibility is determined by the topology of nucleosomes, which strongly influences the association of chromatin-binding proteins and the RNA around them.[67] The mammalian genome is organized by four nucleosome-remodeler families, ISWI, CHD, INO80, and SWI/SNF, also called BAF.[68] Perturbation of BAF with the small-molecule ATPase inhibitor BRM014 leads to a rapid and extensive restriction in DNA accessibility within minutes[69,70]; however, the mechanisms underlying this widespread chromatin compaction remain incompletely understood. We recorded a 20-min time course upon BRM014 treatment of MCF7 cells and compared photo-crosslinked DNA interactomes to matched nuclear and total proteomes as well as two time points of formaldehyde-crosslinked chromatin (Figure 6A; Table S5). The abundance of BRG1/BRM (encoded by SMARCA4/SMARCA2) or any other detected component of the BAF complex remained constant in the series of photo-crosslinked DNA interactomes. Yet, we observed highly significant protein abundance changes for other DNA interactors, including numerous chromatin organizers (Figure S6A). Figures 6A and 6B highlight a selection of proteins with particularly striking changes in the photo-crosslinked DNA interactome, which were basically unchanged in the other datasets. We speculate that the much larger fold changes observed upon XDNAX resulted from more efficient removal of nuclear background proteins, which under formaldehyde crosslinking remained trapped in the chromatin gel and led to ratio suppression (Figure 6C). For abundant proteins such as HMGB1 or LMNB1, highly significant changes in DNA interaction could also be corroborated with nucleotide crosslinks. In the case of HMGB1, peptide-DNA hybrids mapping to its DNA-binding domains were detected with increasing frequency over the course

(G) Beeswarm plot for best-populated complexes annotated in CORUM.
(H) Radar plot comparing relative abundance changes in the DNA interactome for proteins participating in specific repair pathways (Reactome annotation). Only proteins with significantly changed interactions under any of the three drugs are displayed (adj. $p < 0.01$).
(I) Same as in (H), but comparing absolute protein abundance changes.
See also Figure S5.

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

CellPress
OPEN ACCESS

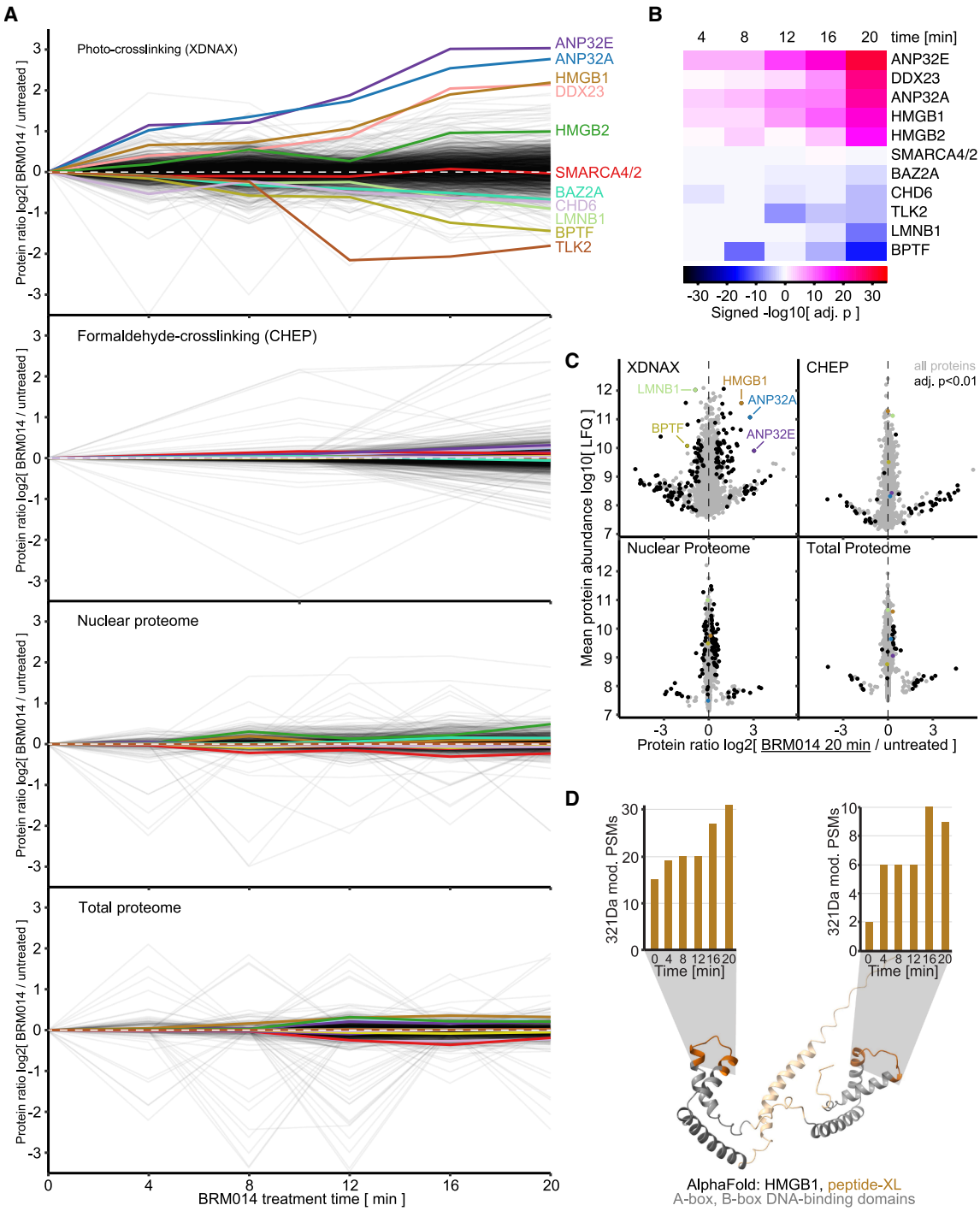


Figure 6. Identification of chromatin accessibility factors during small-molecule BAF inhibition

(A) Line plots comparing four proteomic analyses on MCF7 cells treated with the BAF inhibitor BRM014 in a 20-min time course. Each line represents one protein. Highlighted are the drug target SMARCA4/2 (BRG1/BRM) and selected chromatin organizers, with particularly significant changes in the photo-crosslinked DNA interactome.

(B) Heatmap of signed *p* values for changes in the photo-crosslinked DNA interactome of proteins highlighted in (A). For all proteins with GO annotation "chromatin organization," see Figure S6A. Testing occurred for each time point individually against the untreated condition with a negative binomial model; see STAR Methods for details.

(legend continued on next page)

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

of the treatment (Figure 6D), whereas decreasing detection frequency was observed for crosslinks in LMNB1 within its IF rod domain (Figure S6B). Interestingly, in LMNA, nucleotide crosslinks indicated unchanged interaction in this region but decreasing interaction in the tail domain at a location associated with Emery-Dreifuss muscular dystrophy (Figure S6C).[71] As early as 4 min into the treatment—and consistently across all other time points—three proteins showed significantly increased DNA interaction: HMGB1, ANP32A, and ANP32E (adj. $p < 0.01$, NBM testing). HMGB proteins are known to facilitate DNA accessibility, implying that their rapid recruitment to DNA might compensate BAF inhibition.[72] ANP32A is part of the INHAT complex (inhibitor of acetyltransferase), which masks histones to prevent their acetylation.[48,73] ANP32E is a histone chaperone that removes H2A.Z, which has been shown to restrict DNA accessibility at promoters and enhancers genome wide.[74] Both ANP32A and ANP32E were among the proteins with the strongest increase in DNA interaction throughout the 20-min treatment (Figure S6D), suggesting that their combined action may contribute to the progressing restriction of DNA accessibility and antagonize BAF activity. Significantly decreased protein abundances in the photo-crosslinked DNA interactomes were observed at later time points for members of other nucleosome-remodeling complexes, including CHD6, BAZ2A, and, especially, BPTF (Figures 6A and 6B), which interacts with H3K4me3 and H2A.Z as part of the NURF complex at transcription start sites of active genes.[75,76] Although this was again consistent with increased chromatin compaction and transcriptional repression, we did not observe similar changes in formaldehyde-crosslinked chromatin, the nuclear proteome, or the total proteome, indicating that photo-crosslinking could capture interaction changes that were previously intractable.

DISCUSSION

A photo-crosslinking system for the study of direct protein-DNA interactions in living cells

Investigating proteins accessing the genome to understand how it is read and put into practice has been a long-standing quest. The high protein content of formaldehyde-crosslinked chromatin has posed challenges for its purification for proteomics, as standard isolation procedures for protein or DNA struggle to purify a cross-linked hybrid molecule sharing physicochemical features of both. During metabolic labeling with the photo-activatable 4ST, only about 0.1% of thymidine bases are replaced.[77] We capitalized on this sparse crosslinking and low protein content by using a series of standard DNA isolation procedures during XDNAX that allowed for very effective removal of non-crosslinked protein. Because it is based on the robust TRIZOL extraction, we expect XDNAX to be applicable to many organisms and tissues.[78] The recent discovery of histone-like proteins in bacteria indicates that there is much to explore about the ways in which proteins from other organisms interact with their genomes.[79]

Limitations of the study

Obvious drawbacks of our photo-crosslinking approach include the requirement for cellular proliferation to incorporate 4ST into the genome, as well as intense photo-activation. As outlined in Figure 3J, photo-crosslinking can be utilized to screen for direct protein-DNA contact, but a DNA crosslink does not necessarily imply that the identified sequence feature anchors a protein to the nucleic acid molecule. If a DNA crosslink is observed in a protein sequence feature not known as a dedicated DNA-binding domain or IDR, additional follow-up experiments such as gel shift assays may be needed to confirm a DNA-anchoring function. However, not all protein-DNA interactions serve anchorage. For instance, the proteasome is known to remove drug-induced protein adducts from DNA without actively binding to DNA,[63] and the proteasome components identified in our interactomes provide insightful evidence for their direct contact with DNA in this process (Figure 5G). A limitation concerning absolute protein quantification is that some proteins photo-crosslink much better than others,[25] which is further biased by the use of 4ST. Indeed, various members of the KLF transcription factor family with a preference for GC-rich motifs were detected across all abundance levels in the MCF7 nuclear proteome (KLF3, 4, 5, 9, 10, 13, and 16) yet none of them in the photo-crosslinked DNA interactome (Table S1). Although the number of transcription factors entirely lacking thymidine in their motif is actually small,[80] this limitation might be addressed by using 6-thioguanosine as an alternative photo-sensitizer.[19,20] Finally, the irradiation-induced 4ST modification we observed can serve as powerful, direct evidence for protein-DNA crosslinking, yet it is rare and only found on highly abundant proteins.

Physical access to the genome from a proteomic perspective

Genome accessibility has primarily been investigated from a DNA-centric perspective, where sequencing techniques such as DNase I hypersensitive site sequencing (DNase-seq) or assay for transposase-accessible chromatin using sequencing (ATAC-seq) have been applied to assess what loci in the genome are open for protein binding. Various proteomic approaches have explored the overall composition of chromatin; however, what proteins within the chromatin gel physically interact with DNA has so far remained unclear.[6–10,23,81] Using zero-distance photo-crosslinking, we found that the majority of proteins in the physical DNA interactome carried canonical DNA, RNA, or histone-binding domains, underscoring the importance of globular domains for DNA access by permanently anchoring proteins to the DNA molecule. In addition, we found that proteins with direct DNA access often carried large IDRs. Specifically for transcription factors, this has been recognized before, and their IDRs have been shown to modulate the DNA sequence the proteins bind to or their ability to activate transcription.[82–85] We hypothesize that, in the competition for a direct interface with the DNA molecule, proteins with certain IDRs have an advantage: they

(C) Scatterplots comparing protein abundance changes upon BAF inhibition to mean absolute protein abundances. The four proteomic analyses from A are displayed at the 20-min time point.
(D) Bar plots comparing the detection of DNA-crosslinked peptides mapping to the DNA-binding domains of HMGB1 between time points.
See also Figure S6.

can participate in phase-separated compartments that already exist around DNA to gain access and, subsequently, use their DNA-binding or histone-binding domains to anchor and stay in place.[83,85] Conversely, once a protein has bound its target DNA locus with a dedicated DNA-binding domain, its IDRs can establish a phase-separated compartment that facilitates access of specific auxiliary proteins in its vicinity.[82]

Quantifying differential DNA binding with single-minute and single-amino-acid resolution

Chromatin processes are often rapid and occur without changes in protein expression. For instance, the removal of H2A.Z by ANP32E at DNA damage sites has been shown to complete within as little as 10 min.[86] The proteomic analysis of chromatin using formaldehyde crosslinking requires 10 min of fixation at 37°C,[4] illustrating that such fast processes are not easily tractable with chemical crosslinking. Photo-crosslinking occurs at 4°C and requires less than a minute of irradiation, allowing for time series with minute resolution, while biochemical processes are halted. We demonstrate that XDNAX outperformed conventional proteomic approaches in resolving DNA interaction changes, capturing recruitment of the histone chaperones ANP32E and ANP32A only 4 min into BAF inhibition. In addition, DNA-crosslinked peptides revealed key sites within abundant proteins that provided additional insight into their changing interaction on the level of protein domains and reinforced the notion that the monitored protein-DNA interactions were direct. Of the 379 proteins identified with nucleotide crosslinks, 142 carried disease annotations in OMIM (Table S2), and we found DNA-crosslinked peptides mapping to variant sites associated with numerous conditions, including Emery-Dreifuss muscular dystrophy (LMNA: R401), childhood-onset neurodegeneration (UBTF: E210), and Zinsser-Engman-Cole syndrome (DKC1: F36, L37, and P40). Overall, this demonstrates that photo-crosslinking enables analyses that were previously impossible and introduces XDNAX as a powerful tool in the study of protein-DNA interactions.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the lead contact, Bernhard Küster (kuster@tum.de).

Materials availability

The high-intensity UV irradiation system UVEN is developed as an open science project, which will be described in detail in an upcoming publication. Plans, construction manuals, and software for the device can be found at www.uven.org.

Data and code availability

Proteomic data and search results have been deposited in the MassIVE database under the identifier MassIVE: MSV000094079. This paper does not report original code.

ACKNOWLEDGMENTS

We thank all members of the Kuster laboratory for continuous support and discussion, in particular Karl Kramer, Polina Prokofeva, Johanna Tueshaus, and Guillaume Medard. We thank Sara Brajkovic, Nicola Berner, Xintong Sui, and Julian Müller for help with the MS measurement, as well as Amirhossein Sakhteman for computational support. We thank Jiri Reinis (CeMM) for computational support. We thank Patroklos Samaras and Tobias Schmidt for help with the initial LED crosslinking tests. We thank Nicholas Kuhn (UCSF) and Christian Frese (Bayer) for advice and discussion. We gratefully acknowledge funding by the German Research Council (DFG) supporting this work (DFG project number 492625837 and SFB1309-B02 project number 325871075). We also acknowledge funding from the Bavarian Research Foundation for construction of the UVEN irradiation system (BFS project number AZ-1444-20C).

AUTHOR CONTRIBUTIONS

J.T. and B.K. conceived and directed the project and wrote the manuscript. All authors edited the revised manuscript. J.T. and S.T. conceived the irradiation device UVEN. S.T. designed and constructed UVEN. J.T. performed *in vitro* benchmarking of UVEN. J.T. performed the XDNAX experiments. J.T. and S.S. performed other proteomic analyses. J.T., S.S., and B.K. measured the proteomic samples. S.S. and J.T. performed knockdown experiments and live-cell imaging. F.G. and N.H.U. provided 4ST-labeled mouse macrophages and analyzed macrophage ChIP-seq data. S.I.W. and K.K. analyzed 4ST and 4SU upon irradiation. S.G. and S.K. performed experiments in the context of BAF inhibition.

DECLARATION OF INTERESTS

B.K. is a co-founder and shareholder of OmicScouts and MSAID. He has no operational role in either company. S.T. is employed by Mynaric Lasercom, a company unaffiliated to the present study. Mynaric Lasercom was not involved in the development of UVEN or the present study and has disclaimed any interest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell Lines
  - Animals
- METHOD DETAILS
  - UV measurements and monitoring of 4SU/4ST photo-conversion by spectrophotometry or LC-MS
  - Live cell imaging for monitoring proliferation, apoptosis and cell death
  - Knockdown of candidate proteins using siRNA pools
  - Hypotonic nuclear fractionation and protein cleanup for proteomics using single-pot-solid-phase-enhanced-sample-preparation (SP3)
  - Treatment of photo-sensitized cells and UV irradiation
  - Protein-crosslinked DNA extraction (XDNAX) for proteomic quantification of the DNA interactome
  - Chromatin enrichment for protoemics (CHEP) from formaldehyde-crosslinked cells
  - HPLC-MS for the detection and quantification of proteomic samples
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - MS database search
  - Processing and analysis of proteomic data
  - Processing and analysis of DNA sequencing data
  - Analysis of live cell imaging data
- ADDITIONAL RESOURCES
  - Functional annotation of proteins and data visualization

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.04.037.

Received: April 5, 2024
Revised: January 10, 2025
Accepted: April 27, 2025
Published: May 22, 2025

REFERENCES

1. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.
2. Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395. https://doi.org/10.1038/cr.2011.22.
3. Strickfaden, H., Tolsma, T.O., Sharma, A., Underhill, D.A., Hansen, J.C., and Hendzel, M.J. (2020). Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells. Cell *183*, 1772–1784.e13. https://doi.org/10.1016/j.cell.2020.11.027.
4. Kustatscher, G., Hégarat, N., Wills, K.L.H., Furlan, C., Bukowski-Wills, J. C., Hochegger, H., and Rappsilber, J. (2014). Proteomics of a fuzzy organelle: Interphase chromatin. EMBO J. *33*, 648–664. https://doi.org/10.1002/embj.201387614.
5. Kustatscher, G., Wills, K.L.H., Furlan, C., and Rappsilber, J. (2014). Chromatin enrichment for proteomics. Nat. Protoc. *9*, 2090–2099. https://doi.org/10.1038/nprot.2014.142.
6. Ginno, P.A., Burger, L., Seebacher, J., Iesmantavicius, V., and Schübeler, D. (2018). Cell cycle-resolved chromatin proteomics reveals the extent of mitotic preservation of the genomic regulatory landscape. Nat. Commun. *9*, 4048. https://doi.org/10.1038/s41467-018-06007-5.
7. Rafiee, M.R., Zagalak, J.A., Sidorov, S., Steinhauser, S., Davey, K., Ule, J., and Luscombe, N.M. (2021). Chromatin-contact atlas reveals disorder-mediated protein interactions and moonlighting chromatin-associated RBPs. Nucleic Acids Res. *49*, 13092–13107. https://doi.org/10.1093/nar/gkab1180.
8. Sigismondo, G., Arseni, L., Palacio-Escat, N., Hofmann, T.G., Seiffert, M., and Krijgsveld, J. (2023). Multi-layered chromatin proteomics identifies cell vulnerabilities in DNA repair. Nucleic Acids Res. *51*, 687–711. https://doi.org/10.1093/nar/gkac1264.
9. Lewis, C.D., and Laemmli, U.K. (1982). Higher order metaphase chromosome structure: Evidence for metalloprotein interactions. Cell *29*, 171–181. https://doi.org/10.1016/0092-8674(82)90101-5.
10. Ohta, S., Bukowski-Wills, J.-C., Sanchez-Pulido, L., Alves, F.D.L., Wood, L., Chen, Z.A., Platani, M., Fischer, L., Hudson, D.F., Ponting, C.P., et al. (2010). The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. Cell *142*, 810–821. https://doi.org/10.1016/j.cell.2010.07.047.
11. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell *149*, 1393–1406. https://doi.org/10.1016/j.cell.2012.04.031.
12. Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Mol. Cell *46*, 674–690. https://doi.org/10.1016/j.molcel.2012.05.021.
13. Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. Nat. Rev. Mol. Cell Biol. *19*, 327–341. https://doi.org/10.1038/nrm.2017.130.
14. Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M. W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell *176*, 391–403.e19. https://doi.org/10.1016/j.cell.2018.11.004.
15. Queiroz, R.M.L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R.F., Dezi, V., et al. (2019). Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). Nat. Biotechnol. *37*, 169–178. https://doi.org/10.1038/s41587-018-0001-2.
16. Gebauer, F., Schwarzl, T., Valcárcel, J., and Hentze, M.W. (2021). RNA-binding proteins in human genetic disease. Nat. Rev. Genet. *22*, 185–198. https://doi.org/10.1038/s41576-020-00302-y.
17. Smith, K.C., and Meun, D.H.C. (1968). Kinetics of the photochemical addition of [35S] cysteine to polynucleotides and nucleic acids. Biochemistry *7*, 1033–1037. https://doi.org/10.1021/bi00843a023.
18. Smith, K.C. (1969). Photochemical addition of amino acids to 14C-uracil. Biochem. Biophys. Res. Commun. *34*, 354–357. https://doi.org/10.1016/0006-291X(69)90840-7.
19. Guven, M., Barnouin, K., Snijders, A.P., and Karran, P. (2016). Photosensitized UVA-Induced Cross-Linking between Human DNA Repair and Replication Proteins and DNA Revealed by Proteomic Analysis. J. Proteome Res. *15*, 4612–4623. https://doi.org/10.1021/acs.jproteome.6b00717.
20. Gueranger, Q., Kia, A., Frith, D., and Karran, P. (2011). Crosslinking of DNA repair and replication proteins to DNA in cells treated with 6-thioguanine and UVA. Nucleic Acids Res. *39*, 5057–5066. https://doi.org/10.1093/nar/gkr112.
21. Reim, A., Ackermann, R., Font-Mateu, J., Kammel, R., Beato, M., Nolte, S., Mann, M., Russmann, C., and Wierer, M. (2020). Atomic-resolution mapping of transcription factor-DNA interactions by femtosecond laser crosslinking and mass spectrometry. Nat. Commun. *11*, 3019. https://doi.org/10.1038/s41467-020-16837-x.
22. Stützer, A., Welp, L.M., Raabe, M., Sachsenberg, T., Kappert, C., Wulf, A., Lau, A.M., David, S.S., Chernev, A., Kramer, K., et al. (2020). Analysis of protein-DNA interactions in chromatin by UV induced cross-linking and mass spectrometry. Nat. Commun. *11*, 5250. https://doi.org/10.1038/s41467-020-19047-7.
23. Van Mierlo, G., and Vermeulen, M. (2021). Chromatin proteomics to study epigenetics - Challenges and opportunities. Mol. Cell. Proteomics *20*, 100056. https://doi.org/10.1074/mcp.R120.002208.
24. Steube, A., Schenk, T., Tretyakov, A., and Saluz, H.P. (2017). High-intensity UV laser ChIP-seq for the study of protein-DNA interactions in living cells. Nat. Commun. *8*, 1303. https://doi.org/10.1038/s41467-017-01251-7.
25. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell *141*, 129–141. https://doi.org/10.1016/j.cell.2010.03.009.
26. Pleiss, M., Ochiai, H., and Cerutti, P.A. (1969). Photochemically induced transition of 4-thiouridine to uridine or uridine and cytidine in E. coli transfer ribonucleic acid. Biochem. Biophys. Res. Commun. *34*, 70–76. https://doi.org/10.1016/0006-291X(69)90530-0.
27. Massey, A., Xu, Y.Z., and Karran, P. (2001). Photoactivation of DNA thiobases as a potential novel therapeutic option. Curr. Biol. *11*, 1142–1146. https://doi.org/10.1016/S0960-9822(01)00272-X.
28. Reelfs, O., Xu, Y.Z., Massey, A., Karran, P., and Storey, A. (2007). Thiothymidine plus low-dose UVA kills hyperproliferative human skin cells independently of their human papilloma virus status. Mol. Cancer Ther. *6*, 2487–2495. https://doi.org/10.1158/1535-7163.MCT-07-0166.
29. Pridgeon, S.W., Heer, R., Taylor, G.A., Newell, D.R., O'Toole, K., Robinson, M., Xu, Y.Z., Karran, P., and Boddy, A.V. (2011). Thiothymidine combined with UVA as a potential novel therapy for bladder cancer. Br. J. Cancer *104*, 1869–1876. https://doi.org/10.1038/bjc.2011.180.
30. Chomczynski, P., and Sacchi, N. (2006). The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. Nat. Protoc. *1*, 581–585. https://doi.org/10.1038/nprot.2006.83.

31. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Mol. Cell. Proteomics *1*, 376–386. https://doi.org/10.1074/mcp.M200025-MCP200.

32. Sinha, N.K., McKenney, C., Yeow, Z.Y., Li, J.J., Nam, K.H., Yaron-Barir, T.M., Johnson, J.L., Huntsman, E.M., Cantley, L.C., Ordureau, A., et al. (2024). The ribotoxic stress response drives UV-mediated cell death. Cell *187*, 3652–3670.e40. https://doi.org/10.1016/j.cell.2024.05.018.

33. Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.-L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat. Methods *11*, 1064–1070. https://doi.org/10.1038/nmeth.3092.

34. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342. https://doi.org/10.1038/nature10098.

35. Zhang, Z., and Zhang, R. (2008). Proteasome activator PA28γ regulates p53 by enhancing its MDM2-mediated degradation. EMBO J. *27*, 852–864. https://doi.org/10.1038/emboj.2008.25.

36. Thomas, T.A., and Smith, D.M. (2022). Proteasome activator 28γ (PA28γ) allosterically activates trypsin-like proteolysis by binding to the α-ring of the 20S proteasome. J. Biol. Chem. *298*, 102140. https://doi.org/10.1016/j.jbc.2022.102140.

37. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al. (2023). InterPro in 2022. Nucleic Acids Res. *51*, D418–D427. https://doi.org/10.1093/nar/gkac993.

38. Zaborowski, A.B., and Walther, D. (2020). Determinants of correlated expression of transcription factors and their target genes. Nucleic Acids Res. *48*, 11347–11369. https://doi.org/10.1093/nar/gkaa927.

39. Shao, W., Bi, X., Pan, Y., Gao, B., Wu, J., Yin, Y., Liu, Z., Peng, M., Zhang, W., Jiang, X., et al. (2022). Phase separation of RNA-binding protein promotes polymerase binding and transcription. Nat. Chem. Biol. *18*, 70–80. https://doi.org/10.1038/s41589-021-00904-5.

40. Erdős, G., Pajkos, M., and Dosztányi, Z. (2021). IUPred3: Prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Res. *49*, W297–W303. https://doi.org/10.1093/nar/gkab408.

41. Akdel, M., Pires, D.E.V., Pardo, E.P., Jänes, J., Zalevsky, A.O., Mészáros, B., Bryant, P., Good, L.L., Laskowski, R.A., Pozzati, G., et al. (2022). A structural biology community assessment of AlphaFold2 applications. Nat. Struct. Mol. Biol. *29*, 1056–1067. https://doi.org/10.1038/s41594-022-00849-w.

42. Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014). Classification of intrinsically disordered regions and proteins. Chem. Rev. *114*, 6589–6631. https://doi.org/10.1021/cr400525m.

43. Soto, L.F., Li, Z., Santoso, C.S., Berenson, A., Ho, I., Shen, V.X., Yuan, S., and Fuxman Bass, J.I. (2022). Compendium of human transcription factor effector domains. Mol. Cell *82*, 514–526. https://doi.org/10.1016/j.molcel.2021.11.007.

44. Chavali, S., Singh, A.K., Santhanam, B., and Babu, M.M. (2020). Amino acid homorepeats in proteins. Nat. Rev. Chem. *4*, 420–434. https://doi.org/10.1038/s41570-020-0204-1.

45. Lee, B., Jaberi-Lashkari, N., and Calo, E. (2022). A unified view of low complexity region (LCRs) across species. eLife *11*, e77058. https://doi.org/10.7554/eLife.77058.

46. Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622–634. https://doi.org/10.1016/j.cell.2011.03.042.

47. Zeng, W., Ball, A.R., and Yokomori, K. (2010). HP1: Heterochromatin binding proteins working the genome. Epigenetics *5*, 287–292. https://doi.org/10.4161/epi.5.4.11683.

48. Seo, S.B., McNamara, P., Heo, S., Turner, A., Lane, W.S., and Chakravarti, D. (2001). Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. Cell *104*, 119–130. https://doi.org/10.1016/S0092-8674(01)00196-9.

49. Jantzen, H.M., Admon, A., Bell, S.P., and Tjian, R. (1990). Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature *344*, 830–836. https://doi.org/10.1038/344830a0.

50. Wang, L., and Han, T. (2024). Pharmacologic Induction of ERα SUMOylation Disrupts Its Chromatin Binding. ACS Chem. Biol. *19*, 2383–2392. https://doi.org/10.1021/acschembio.4c00606.

51. Zecha, J., Bayer, F.P., Wiechmann, S., Woortman, J., Berner, N., Müller, J., Schneider, A., Kramer, K., Abril-Gil, M., Hopf, T., et al. (2023). Decrypting drug actions and protein modifications by dose- and time-resolved proteomics. Science *380*, 93–101. https://doi.org/10.1126/science.ade3925.

52. Eckert, S., Berner, N., Kramer, K., Schneider, A., Müller, J., Lechner, S., Brajkovic, S., Sakhteman, A., Graetz, C., Fackler, J., et al. (2025). Decrypting the molecular basis of cellular drug phenotypes by dose-resolved expression proteomics. Nat. Biotechnol. *43*, 406–415. https://doi.org/10.1038/s41587-024-02218-y.

53. Kuiper, G.G.J.M., Lemmen, J.G., Carlsson, B., Corton, J.C., Safe, S.H., Van Der Saag, P.T., Van Der Burg, B., and Gustafsson, J.A. (1998). Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor β. Endocrinology *139*, 4252–4263. https://doi.org/10.1210/endo.139.10.6216.

54. Swift, L.H., and Golsteyn, R.M. (2014). Genotoxic anti-cancer agents and their relationship to DNA damage, mitosis, and checkpoint adaptation in proliferating cancer cells. Int. J. Mol. Sci. *15*, 3403–3431. https://doi.org/10.3390/ijms15033403.

55. Bruno, P.M., Liu, Y., Park, G.Y., Murai, J., Koch, C.E., Eisen, T.J., Pritchard, J.R., Pommier, Y., Lippard, S.J., and Hemann, M.T. (2017). A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress. Nat. Med. *23*, 461–471. https://doi.org/10.1038/nm.4291.

56. Zhang, R.G., Zhang, R.P., Wang, X.W., and Xie, H. (2002). Effects of cisplatin on telomerase activity and telomere length in BEL-7404 human hepatoma cells. Cell Res. *12*, 55–62. https://doi.org/10.1038/sj.cr.7290110.

57. Castell, A., Yan, Q., Fawkner, K., Bazzar, W., Zhang, F., Wickström, M., Alzrigat, M., Franco, M., Krona, C., Cameron, D.P., et al. (2022). MYCMI-7: A Small MYC-Binding Compound that Inhibits MYC: MAX Interaction and Tumor Growth in a MYC-Dependent Manner. Cancer Res. Commun. *2*, 182–201. https://doi.org/10.1158/2767-9764.CRC-21-0019.

58. Adomas, A.B., Grimm, S.A., Malone, C., Takaku, M., Sims, J.K., and Wade, P.A. (2014). Breast tumor specific mutation in GATA3 affects physiological mechanisms regulating transcription factor turnover. BMC Cancer *14*, 278. https://doi.org/10.1186/1471-2407-14-278.

59. Pastor-Flores, D., Talwar, D., Pedre, B., and Dick, T.P. (2020). Real-time monitoring of peroxiredoxin oligomerization dynamics in living cells. Proc. Natl. Acad. Sci. USA *117*, 16313–16323. https://doi.org/10.1073/pnas.1915275117.

60. Somyajit, K., Gupta, R., Sedlackova, H., Neelsen, K.J., Ochs, F., Rask, M.B., Choudhary, C., and Lukas, J. (2017). Redox-sensitive alteration of replisome architecture safeguards genome integrity. Science *358*, 797–802. https://doi.org/10.1126/science.aao3172.

61. Schmidt, H.B., Jaafar, Z.A., Wulff, B.E., Rodencal, J.J., Hong, K., Aziz-Zanjani, M.O., Jackson, P.K., Leonetti, M.D., Dixon, S.J., Rohatgi, R., et al. (2022). Oxaliplatin disrupts nucleolar function through biophysical disintegration. Cell Rep. *41*, 111629. https://doi.org/10.1016/j.celrep.2022.111629.

62. Sutton, E.C., and DeRose, V.J. (2021). Early nucleolar responses differentiate mechanisms of cell death induced by oxaliplatin and cisplatin. J. Biol. Chem. *296*, 100633. https://doi.org/10.1016/j.jbc.2021.100633.

63. Essawy, M., Chesner, L., Alshareef, D., Ji, S., Tretyakova, N., and Campbell, C. (2023). Ubiquitin signaling and the proteasome drive human DNA–protein crosslink repair. Nucleic Acids Res. *51*, 12174–12184. https://doi.org/10.1093/nar/gkad860.

64. Biard, D.S.F. (2007). Untangling the relationships between DNA repair pathways by silencing more than 20 DNA repair genes in human stable clones. Nucleic Acids Res. *35*, 3535–3550. https://doi.org/10.1093/nar/gkm195.

65. Olivieri, M., Cho, T., Álvarez-Quilón, A., Li, K., Schellenberg, M.J., Zimmermann, M., Hustedt, N., Rossi, S.E., Adam, S., Melo, H., et al. (2020). A Genetic Map of the Response to DNA Damage in Human Cells. Cell *182*, 481–496.e21. https://doi.org/10.1016/j.cell.2020.05.040.

66. Sun, Y., Miller Jenkins, L.M.M., Su, Y.P., Nitiss, K.C., Nitiss, J.L., and Pommier, Y. (2020). A conserved SUMO pathway repairs topoisomerase DNA-protein cross-links by engaging ubiquitin-mediated proteasomal degradation. Sci. Adv. *6*, eaba6290. https://doi.org/10.1126/sciadv.aba6290.

67. Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220. https://doi.org/10.1038/s41576-018-0089-8.

68. Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell Biol. *18*, 407–422. https://doi.org/10.1038/nrm.2017.26.

69. Schick, S., Grosche, S., Kohl, K.E., Drpic, D., Jaeger, M.G., Marella, N.C., Imrichova, H., Lin, J.G., Hofstätter, G., Schuster, M., et al. (2021). Acute BAF perturbation causes immediate changes in chromatin accessibility. Nat. Genet. *53*, 269–278. https://doi.org/10.1038/s41588-021-00777-3.

70. Iurlaro, M., Stadler, M.B., Masoni, F., Jagani, Z., Galli, G.G., and Schübeler, D. (2021). Mammalian SWI/SNF continuously restores local accessibility to chromatin. Nat. Genet. *53*, 279–287. https://doi.org/10.1038/s41588-020-00768-w.

71. Vytopil, M., Ricci, E., Dello Russo, A., Hanisch, F., Neudecker, S., Zierz, S., Ricotti, R., Demay, L., Richard, P., Wehnert, M., et al. (2002). Frequent low penetrance mutations in the Lamin A/C gene, causing Emery Dreifuss muscular dystrophy. Neuromuscul. Disord. *12*, 958–963. https://doi.org/10.1016/S0960-8966(02)00178-5.

72. Saunders, H.S., Chio, U.S., Moore, C.M., Ramani, V., Cheng, Y., and Narlikar, G.J. (2024). HMGB1 restores a dynamic chromatin environment in the presence of linker histone by deforming nucleosomal DNA. Preprint at bioRxiv. https://doi.org/10.1101/2024.08.23.609244.

73. Seo, S.B., Macfarlan, T., Mcnamara, P., Hong, R., Mukai, Y., Heo, S., and Chakravarti, D. (2002). Regulation of Histone Acetylation and Transcription by Nuclear Protein pp32, a Subunit of the INHAT Complex. J. Biol. Chem. *277*, 14005–14010. https://doi.org/10.1074/jbc.M112455200.

74. Murphy, K.E., Meng, F.W., Makowski, C.E., and Murphy, P.J. (2020). Genome-wide chromatin accessibility is restricted by ANP32E. Nat. Commun. *11*, 5063. https://doi.org/10.1038/s41467-020-18821-x.

75. Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86–90. https://doi.org/10.1038/nature04815.

76. Kim, K., Punj, V., Choi, J., Heo, K., Kim, J.M., Laird, P.W., and An, W. (2013). Gene dysregulation by histone variant H2A.Z in bladder cancer. Epigenetics Chromatin *6*, 34. https://doi.org/10.1186/1756-8935-6-34.

77. Reelfs, O., MacPherson, P., Ren, X., Xu, Y.Z., Karran, P., and Young, A.R. (2011). Identification of potentially cytotoxic lesions induced by UVA photoactivation of DNA 4-thiothymidine in human cells. Nucleic Acids Res. *39*, 9620–9632. https://doi.org/10.1093/nar/gkr674.

78. Threlfall, J., and Blaxter, M. (2021). Launching the Tree of Life Gateway. Wellcome Open Res. *6*, 125. https://doi.org/10.12688/wellcomeopenres.16913.1.

79. Hocher, A., Laursen, S.P., Radford, P., Tyson, J., Lambert, C., Stevens, K.M., Montoya, A., Shliaha, P.V., Picardeau, M., Sockett, R.E., et al. (2023). Histones with an unconventional DNA-binding mode in vitro are major chromatin constituents in the bacterium Bdellovibrio bacteriovorus. Nat. Microbiol. *8*, 2006–2019. https://doi.org/10.1038/s41564-023-01492-x.

80. Rauluseviciute, I., Riudavets-Puig, R., Blanc-Mathieu, R., Castro-Mondragon, J.A., Ferenc, K., Kumar, V., Lemma, R.B., Lucas, J., Chèneby, J., Baranasic, D., et al. (2024). JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *52*, D174–D182. https://doi.org/10.1093/nar/gkad1059.

81. Nakamura, K., Kustatscher, G., Alabert, C., Hödl, M., Forne, I., Völker-Albert, M., Satpathy, S., Beyer, T.E., Mailand, N., Choudhary, C., et al. (2021). Proteome dynamics at broken replication forks reveal a distinct ATM-directed repair response suppressing DNA double-strand break ubiquitination. Mol. Cell *81*, 1084–1099.e6. https://doi.org/10.1016/j.molcel.2020.12.025.

82. Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842–1855.e16. https://doi.org/10.1016/j.cell.2018.10.042.

83. Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D.K., Carmi, M., and Barkai, N. (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. Mol. Cell *79*, 459–471.e4. https://doi.org/10.1016/j.molcel.2020.05.032.

84. Ahn, J.H., Davis, E.S., Daugird, T.A., Zhao, S., Quiroga, I.Y., Uryu, H., Li, J., Storey, A.J., Tsai, Y.H., Keeley, D.P., et al. (2021). Phase separation drives aberrant chromatin looping and cancer development. Nature *595*, 591–595. https://doi.org/10.1038/s41586-021-03662-5.

85. Xing, Y.-H., Dong, R., Lee, L., Rengarajan, S., Riggi, N., Boulay, G., and Rivera, M.N. (2024). DisP-seq reveals the genome-wide functional organization of DNA-associated disordered proteins. Nat. Biotechnol. *42*, 52–64. https://doi.org/10.1038/s41587-023-01737-4.

86. Gursoy-Yuzugullu, O., Ayrapetov, M.K., and Price, B.D. (2015). Histone chaperone Anp32e removes H2A.Z from DNA double-strand breaks and promotes nucleosome reorganization and DNA repair. Proc. Natl. Acad. Sci. USA *112*, 7507–7512. https://doi.org/10.1073/pnas.1504868112.

87. Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46*, D754–D761. https://doi.org/10.1093/nar/gkx1098.

88. Nguyen, H.C.B., Adlanmerini, M., Hauck, A.K., and Lazar, M.A. (2020). Dichotomous engagement of HDAC3 activity governs inflammatory responses. Nature *584*, 286–290. https://doi.org/10.1038/s41586-020-2576-2.

89. Tsunaka, Y., Kajimura, N., Tate, S.I., and Morikawa, K. (2005). Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. Nucleic Acids Res. *33*, 3424–3434. https://doi.org/10.1093/nar/gki663.

90. Kim, M.S., Chuenchor, W., Chen, X., Cui, Y., Zhang, X., Zhou, Z.H., Gellert, M., and Yang, W. (2018). Cracking the DNA Code for V(D)J Recombination. Mol. Cell *70*, 358–370.e4. https://doi.org/10.1016/j.molcel.2018.03.008.

91. Vanden Broeck, A., Lotz, C., Drillien, R., Haas, L., Bedez, C., and Lamour, V. (2021). Structural basis for allosteric regulation of Human Topoisomerase IIα. Nat. Commun. *12*, 2962. https://doi.org/10.1038/s41467-021-23136-6.

92. Schröder, E., Littlechild, J.A., Lebedev, A.A., Errington, N., Vagin, A.A., and Isupov, M.N. (2000). Crystal structure of decameric 2-Cys

peroxiredoxin from human erythrocytes at 1.7 Å resolution. Structure *8*, 605–615. https://doi.org/10.1016/S0969-2126(00)00147-7.

93. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R.E. (1981). Structure of a B-DNA dodecamer: conformation and dynamics. Proc. Natl. Acad. Sci. USA *78*, 2179–2183. https://doi.org/10.1073/pnas.78.4.2179.
94. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
95. Pratt, H.E., Andrews, G.R., Phalke, N., Purcaro, M.J., Van Der Velde, A., Moore, J.E., and Weng, Z. (2022). Factorbook: An updated catalog of transcription factor motifs and candidate regulatory motif sites. Nucleic Acids Res. *50*, D141–D149. https://doi.org/10.1093/nar/gkab1039.
96. Tsitsiridis, G., Steinkamp, R., Giurgiu, M., Brauner, B., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2023). CORUM: the comprehensive resource of mammalian protein complexes-2022. Nucleic Acids Res. *51*, D539–D545. https://doi.org/10.1093/nar/gkac1015.
97. Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*, eaal3321. https://doi.org/10.1126/science.aal3321.
98. Cho, N.H., Cheveralls, K.C., Brunner, A.D., Kim, K., Michaelis, A.C., Raghavan, P., Kobayashi, H., Savy, L., Li, J.Y., Canaj, H., et al. (2022). OpenCell: Endogenous tagging for the cartography of human cellular organization. Science *375*, eabi6983. https://doi.org/10.1126/science.abi6983.
99. Amberger, J., Bocchini, C.A., Scott, A.F., and Hamosh, A. (2009). McKusick's Online Mendelian Inheritance in Man (OMIM®). Nucleic Acids Res. *37*, D793–D796. https://doi.org/10.1093/nar/gkn665.
100. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372. https://doi.org/10.1038/nbt.1511.
101. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. Nat. Methods *14*, 513–520. https://doi.org/10.1038/nmeth.4256.
102. Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., Vandergheynst, J., Fiehn, O., and Arita, M. (2015). MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat. Methods *12*, 523–526. https://doi.org/10.1038/nmeth.3393.
103. Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48. https://doi.org/10.1186/1471-2105-10-48.
104. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
105. Wickham, H. (2016). ggplot2 (Springer International Publishing) https://doi.org/10.1007/978-3-319-24277-4.
106. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
107. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv.
108. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/NAR/GKW257.
109. Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A., and Loraine, A.E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics *25*, 2730–2731. https://doi.org/10.1093/bioinformatics/btp472.
110. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.
111. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. *9*, e1003118. https://doi.org/10.1371/journal.pcbi.1003118.
112. Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237. https://doi.org/10.1186/1471-2105-11-237.
113. Xu, S., Hu, E., Cai, Y., Xie, Z., Luo, X., Zhan, L., Tang, W., Wang, Q., Liu, B., Wang, R., et al. (2024). Using clusterProfiler to characterize multiomics data. Nat. Protoc. *19*, 3292–3320. https://doi.org/10.1038/s41596-024-01020-z.
114. Barish, G.D., Downes, M., Alaynick, W.A., Yu, R.T., Ocampo, C.B., Bookout, A.L., Mangelsdorf, D.J., and Evans, R.M. (2005). A nuclear receptor atlas: Macrophage activation. Mol. Endocrinol. *19*, 2466–2477. https://doi.org/10.1210/me.2004-0529.
115. Hall, M.D., Telma, K.A., Chang, K.E., Lee, T.D., Madigan, J.P., Lloyd, J.R., Goldlust, I.S., Hoeschele, J.D., and Gottesman, M.M. (2014). Say no to DMSO: Dimethylsulfoxide inactivates cisplatin, carboplatin, and other platinum complexes. Cancer Res. *74*, 3913–3922. https://doi.org/10.1158/0008-5472.CAN-14-0247.
116. Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., Steinmetz, L.M., and Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. Mol. Syst. Biol. *10*, 757. https://doi.org/10.15252/msb.20145625.
117. Bian, Y., Zheng, R., Bayer, F.P., Wong, C., Chang, Y.C., Meng, C., Zolg, D.P., Reinecke, M., Zecha, J., Wiechmann, S., et al. (2020). Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC–MS/MS. Nat. Commun. *11*, 157. https://doi.org/10.1038/s41467-019-13973-x.
118. Gessulat, S., Schmidt, T., Zolg, D.P., Samaras, P., Schnatbaum, K., Zerweck, J., Knaute, T., Rechenberger, J., Delanghe, B., Huhmer, A., et al. (2019). Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. Nat. Methods *16*, 509–518. https://doi.org/10.1038/s41592-019-0426-7.
119. The, M., Samaras, P., Kuster, B., and Wilhelm, M. (2022). Reanalysis of ProteomicsDB Using an Accurate, Sensitive, and Scalable False Discovery Rate Estimation Approach for Protein Groups. Mol. Cell. Proteomics *21*, 100437. https://doi.org/10.1016/j.mcpro.2022.100437.
120. Benhalevy, D., Anastasakis, D.G., and Hafner, M. (2018). Proximity-CLIP provides a snapshot of protein-occupied RNA elements in subcellular compartments. Nat. Methods *15*, 1074–1082. https://doi.org/10.1038/s41592-018-0220-y.
121. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740. https://doi.org/10.1038/nmeth.3901.
122. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. *48*, D882–D889. https://doi.org/10.1093/nar/gkz1062.

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Dialysed FBS	Silantes	281000900
Pen-Strep	Gibco	15140-122
DMEM for SILAC	Silantes	280001300
Opti-MEM reduced serum medium	Gibco	31985062
Lipofectamine RNAiMAX	Invitrogen	13778075
SILAC heavy L-lysine ($^{13}C_6$,$^{15}N_2$-L-lysine HCl)	Silantes	211604102
SILAC heavy L-arginine ($^{13}C_6$,$^{15}N_4$-L-Arginine HCl)	Silantes	201604102
Trypsin	Roche	3708985001
SP3 beads	GE	44152105050250
TRI reagent (TRIZOL)	Sigma	T9424
4-Thiothymidine (4ST)	Biosynth	NT06341
4-Thiothyuridine (4SU)	Biosynth	NT06186
RNase A	Thermo	EN0531
RNase T1	Thermo	EN0541
Covaris plate 96 AFA-TUBE TPX Plate	Covaris	520291
Zymo-Spin IIICG Columns	Zymo	C1006-50-G
Nuclease P1	NEB	M0660S
Benzonase	Santa Cruz	sc-202391
SCX StageTip material	Sigma	225166889-U
C18 StageTip material	Sigma	66883-U
Sep-Pak C18 Cartridges	Waters	WAT054960
17ß-estradiol	Sigma	E8875-1G
LPS E.coli O111:B4	Sigma	LPS25
Cisplatin	MedChemExpress	HY-17394
Oxaliplatin	Sigma	PHR1528
BRM014	MedChemExpress	HY-119374
Critical commercial assays		
Oxazole Yellow Homodimer	Biotum	40090
Annexin V CF 594	Biotum	29085R-5ug
siPOOL REXO4 (NCBI Gene ID 57109)	siTOOLs Biotech	NCBI 57109
siPOOL PPP5C (NCBI Gene ID 5536)	siTOOLs Biotech	NCBI 5536
siPOOL FEN1 (NCBI Gene ID 2237)	siTOOLs Biotech	NCBI 2237
Deposited data		
Proteomics data	This paper, MassIVE	MassIVE: MSV000094079
Human proteome (search term: 'reviewed:yes AND proteome:up000005640', 20350 entries, retrieved 5.19.2020)	Uniprot	Uniprot: UP000005640
Human genome (hg38)	GENCODE	GENCODE: GRCh38
Human genome (hg19)	GENCODE	GENCODE: GRCh37.p13
Mouse genome (mm10)	Ensembl	Ensembl: GRCm38.p6
BioMart (Ensembl release 96)	Zerbino et al.[87]	RRID:SCR_002344
ATAC-Seq in MCF7	ENCODE	ENCODE: ENCFF821OEF
DNase-Seq in MCF7	ENCODE	ENCODE: ENCFF835KCG

(*Continued on next page*)

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FAIRE-Seq in MCF7	ENCODE	ENCODE: ENCFF001UYW
Cebpb ChIP-Seq in LPS stimulated mouse macrophages	NA	GEO: GSE99895
Cebpd ChIP-Seq in LPS stimulated mouse macrophages	NA	GEO: GSE99895
Rela ChIP-Seq in LPS stimulated mouse macrophages	Nguyen et al.[88]	GEO: GSE140611
Crystal structure of human nucleosome core particle	Tsunaka et al.[89]	PDB: 2CV5
Pre-Reaction Complex, RAG1(E962Q)/2-intact/nicked 12/23RSS complex in Mn2+	Kim et al.[90]	PDB: 6CIK
Cryo-EM structure of the Human topoisomerase II alpha DNA-binding/cleavage domain in State 1	Vanden Broeck et al.[91]	PDB: 6ZY5
thioredoxin peroxidase B from red blood cells	Schröder et al.[92]	PDB: 1QMV
Structure of a B-DNA dodecamer: conformation and dynamics	Drew et al.[93]	PDB: 1BNA
AlphaFold structure predictions for LMNB1, C5orf24 homologues	Jumper et al.[94]	https://alphafold.ebi.ac.uk/
Factorbook TF Motif Catalog	Pratt et al.[95]	https://www.factorbook.org/downloads
CORUM: The comprehensive resource of mammalian protein complexes	Tsitsiridis et al.[96]	https://mips.helmholtz-muenchen.de/corum/
The Human Protein Atlas	Thul et al.[97]	https://www.proteinatlas.org/
OpenCell	Cho et al.[98]	https://opencell.czbiohub.org/
OMIM	Amberger et al.[99]	https://omim.org/
Experimental models: Cell lines		
Human (female): MCF7 cells	ATCC	RRID:CVCL_0031
Human (female): U2OS cells	ATCC	RRID:CVCL_0042
Software and algorithms		
MaxQuant (2.1.2.0)	Cox et al.[100]	https://www.maxquant.org/, RRID:SCR_014485
MSFragger 3.8	Kong et al.[101]	http://www.nesvilab.org/software.html
FragPipe 18.0	Kong et al.[101]	http://www.nesvilab.org/software.html, RRID:SCR_022864
MS-Dial 4.90	Tsugawa et al.[102]	https://systemsomicslab.github.io/compms/msdial/main.html, RRID:SCR_023076
GOrilla	Eden et al.[103]	http://cbl-gorilla.cs.technion.ac.il/, RRID:SCR_006848
DESeq2	Love et al.[104]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html, RRID:SCR_000154
RStudio (1.1.463)	RStudio: Integrated Development for R. RStudio, Inc., Boston, MA	http://www.rstudio.com/, RRID:SCR_000432
ggplot2 (R)	Wickham et al.[105]	https://ggplot2.tidyverse.org/, RRID:SCR_014601
samtools	Li et al.[106]	http://www.htslib.org/, RRID:SCR_002105
BWA-MEM (0.7.13)	Li et al.[107]	https://github.com/lh3/bwa, RRID: SCR_010910
Picard Tools (2.0.1)	Broad Institute	https://broadinstitute.github.io/picard/, RRID: SCR_006525
deepTools (3.5.0)	Ramírez et al.[108]	https://deeptools.readthedocs.io/en/develop/, RRID: SCR_016366

(*Continued on next page*)

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
The Integrated Genome Browser (IGB)	Nicol et al.[109]	http://igb.bioviz.org and http://, RRID:SCR_011792
MACS (3.0.0)	Zhang et al.[110]	https://pypi.org/project/MACS3/, RRID:SCR_01329
Biostrings (R)	Pagès et al.	https://bioconductor.org/packages/release/bioc/html/Biostrings.html, RRID:SCR_016949
GenomicRanges (R)	Lawrence et al.[111]	https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html, RRID:SCR_000025
ChIPpeakAnno (R)	Zhu et al.[112]	https://bioconductor.org/packages/release/bioc/html/ChIPpeakAnno.html RRID: SCR_012828,
clusterProfiler (R)	Xu et al.[113]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html, RRID:SCR_016884
Additional resources		
UVEN 365 nm irradiation system	This paper	www.uven.org

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell Lines

MCF7 (human [Homo sapiens], female breast adenocacrinoma) and U2OS (human [Homo sapiens], female bone sarcoma) were obtained from ATCC. Cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10 % dialysed FBS and Pen-Strep (100 U / ml penicillin, 100 mg / ml streptomycin) at 37 °C, 5 % CO_2. DMEM for SILAC was supplemented with 1 mM L-lysine and 0.5 mM L-arginine of the individual SILAC labels as well as 1.7 mM light L-proline and 2 mM L-glutamine. For full labeling, the SILAC label was introduced during six cell passages in DMEM for SILAC. For the dose-dependent analysis of the DNA interactome after estrogen treatment, cells were expanded and treated in DMEM without phenol red as described below. If not further specified labeling with 4-thiothymidine (4ST) occurred over three days at 100 μM concentration in the culture medium added from a 100 mM stock in DMSO.

Animals

Mouse experiments were performed according to the rules and guidelines established by the Institutional Animal Committee at Helmholtz Center Munich. Bone marrow-derived macrophages were derived from 6-12 week old male mice as reported previously.[114] Briefly, bone marrow was harvested into RPMI, erythrocytes lysed in AKC lysis buffer (1 M NH_4Cl, 1 M $KHCO_3$, 0.5 M EDTA) and mononucleated cells purified with a Ficoll Paque gradient. Differentiation occurred in DMEM containing 20% FBS, PenStrep and 30% supernatant from L929 cells for 3 days, after which 200 μM 4ST was added and differentiation continued for 3 additional days.

METHOD DETAILS

UV measurements and monitoring of 4SU/4ST photo-conversion by spectrophotometry or LC-MS

Light intensities of the UVEN irradiation system and the reference device VILBER Bio-Link were measured using an LS-128 UV energy meter (Linshang). The sensor was placed behind a conventional polystyrene cell culture dish (Greiner CELLSTAR) at a distance of 3.5 cm from the bulb or LED. Photo-conversion of 4ST and 4-thiouridine (4SU) was monitored in 100 μl of a 1 mM solution irradiated in glass vials at 3.5 cm distance from the bulb or LED. The absorbance spectrum of the irradiated solutions was recorded on a Nanodrop spectrophotometer (Thermo Scientific). For LC-MS analysis samples were diluted 1:100 with methanol. The untargeted analysis was performed using a Nexera UHPLC system (Shimadzu, Duisburg, Germany) coupled to a Q-TOF mass spectrometer (TripleTOF 6600, AB Sciex, Darmstadt, Germany). Separation was achieved with a UPLC ACQUITY Premier BEH Amide 2.1 × 100 mm, 1.7 μm analytical column (Waters, Eschborn, Germany) with a flow rate of 400 μL/min. The mobile phase consisted of water/acetonitrile (50/50, v/v) with 5 mM ammonium acetate at pH 9.5, adjusted with 25% ammonia solution (eluent A), and acetonitrile/water (95/5, v/v) with 5 mM ammonium acetate at pH 9.5 (eluent B). The gradient profile was 100% B from 0 to 2 min, decreasing to 20% B at 12 min. The samples were analyzed in Information Dependent Acquisition (IDA) mode in the negative electrospray ionization (ESI) mode. MS settings were as follows: Gas 1, 55; Gas 2, 65; Curtain gas, 35; Temperature, 500 °C; Ion Spray Voltage, -4500; declustering potential, -80. The mass range of the TOF MS and MS/MS scans was 50–2000 m/z, and the collision energy was set to -35 V with a 15 V spread. Raw data were processed as profile data using MS-Dial. Parameters for peak picking, deconvolution, and peak alignment were set to

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

default values, and the adduct type was [M-H]-. For metabolite annotation, the exact masses and MS2 fragmentation patterns of the most abundant signals were manually annotated, and MS2 spectra for uridine and thymidine were confirmed by comparison to in-house library spectra. For relative quantitative comparison, the MS1 area under the curve of annotated precursors was used.

Live cell imaging for monitoring proliferation, apoptosis and cell death

Cells were monitored in culture by time-lapse microscopy with an IncuCyte S3 automated microscopy system (Essen Bioscience) in a conventional cell culture incubator. For monitoring proliferation in the presence of 4ST or 4SU, 5000 MCF7 cells or 10000 U2OS cells were seeded onto a 96-well tissue culture plate (testplate 96F, TPP) in 200 µl DMEM and cultured at normal conditions. For investigating the UV sensitization of MCF7 cells after 4ST exposure the 96-well plate was irradiated with UVEN after 4 days of culture. In order to control for potential toxic effects of 4ST or 4SU photo-products in the medium, cells were washed with PBS, and the medium replaced with fresh DMEM without nucleotide before the incubation was continued.

For monitoring the effect of genotoxic drugs 5000 MCF7 cells were seeded and expanded for three days in 100 µl DMEM. Subsequently, 100 µl genotoxic drugs were added at 200 µM concentration (100 µM final concentration) in medium containing 63 ng/ml of the apoptosis fluorescence reporter annexin-V CF 594 as well as 100 nM of the cell death fluorescence reporter oxazole yellow homodimer. Because of the incompatibility of platinum compounds with DMSO,[115] cisplatin and oxaliplatin were directly dissolved to a final concentration of 100 µM in DMEM and etoposide added from a 1:1000 stock in DMSO.

For the monitoring of dose-dependent genotoxic drug effects during siRNA-mediated knockdown a 1:2 dilution series starting from 3.2 mM in 50 µl DMEM was prepared and combined with 50 µl DMEM containing 25 ng annexin-V CF 594 as well as 400 nM oxazole yellow homodimer (final starting concentration 1.6 mM genotoxic drug). This mix was added onto the 100 µl culture medium to yield a final concentration of 800 µM – 12.5 µM of the genotoxic drugs on the cell culture plate.

IncuCyte scans were acquired every two hours using a 4 x objective with phase and additionally green (300 ms) and red (400 ms) channel acquisition if required. For analysis a standard IncuCyte definition was applied using a segmentation adjustment of 1.1 for phase, Top-Hat segmentation for the green channel (100 µM radius, 70 GCU threshold), as well as the red channel (100 µM radius, 1 RCU threshold). For the quantification of proliferation (phase) the confluence (%) and for the quantification of apoptosis (red) or cell death (green) total areas (μM^2 / image) were exported.

Knockdown of candidate proteins using siRNA pools

Candidate proteins were knocked down in MCF7 cells with siRNA pools containing 30 distinct siRNAs directed against the target transcript designed and provided by the vendor (siTools biotech). Knockdown was performed according to the manufacturer's instructions and efficiency validated by LC-MS. In brief, solution A was prepared by combining 1230 µl of Opti-MEM with 20 µl Lipofectamine RNAiMAX and vortexed vigorously. Solution B was prepared by combining 1050 µl Opti-MEM with 200 µl siRNA dilution (150 nM in ultrapure water). Solution A and B were combined, vortexed and lipid-nucleic acid (LNP) particles allowed to form for 15 minutes. The LNP solution was combined with 1.2 million singularized MCF7 cells in 10 ml DMEM and mixed by pipetting. Subsequently, 10000 cells in 100 µl medium were distributed onto a 96-well plate and allowed to adhere for 15 minutes on the bench before transfer into the IncuCyte. Drug treatment occurred 24 hours after transfection by addition of 100 µl drug concentrate as described above.

Hypotonic nuclear fractionation and protein cleanup for proteomics using single-pot-solid-phase-enhanced-sample-preparation (SP3)

Reference deep proteomic analysis of the MCF7 total proteome as well as their nuclear and cytosolic subproteomes was performed in quadruplicates. Hypotonic nuclear extraction was applied to generate the cytosolic and nuclear fractions. Per replicate a 15 cm diameter dish of 70 % confluent MCF7 cells was harvested by scraping into ice-cold PBS and pelleted with 500 g centrifugation for 5 minutes at 4 °C. All subsequent steps were performed on ice. The supernatant was discarded and cells spun down to remove all residual PBS. The pellet was suspended in 2 ml ice-cold nuclear isolation buffer (Tris-Cl 10 mM, KCl 60 mM, MgCl2 1.5 mM, NP40 0.1 %), and pipetted ten times with a 1 ml tip to break the plasma membrane. Cells were incubated for 10 minutes on ice and pipetted again ten times. Nuclei were then pelleted by centrifugation with 2000 g, for 10 minutes at 4 °C. The supernatant containing the cytosolic fraction was transferred to a fresh tube and the nuclei pellet suspended in another 2 ml nuclear isolation buffer for washing. Nuclei were again pelleted with 2000 g, for 10 minutes at 4 °C and taken up in 2 ml nuclear isolation buffer and their integrity checked by microscopy. To create equal conditions between all samples, MCF7 cells for total proteomes were lysed in 2 ml nuclear isolation buffer. Protein content of the samples was determined by BCA and 250 µg protein used for further processing. Volumes were adjusted to 300 µl using nuclear isolation buffer. For protein cleanup a modified version of the SP3 protocol was used.[116] To avoid SDS precipitation by potassium, proteins were completely denatured with guanidinium thiocyanate (GuTC) and nucleic acids fragmented with trifluoracetic acid (TFA). Therefore, samples were combined with 300 µl SP3 lysis buffer (GuTC 5 M, TFA 200 mM), mixed by vortexing, incubated at 90 °C, 700 rpm shaking for 5 minutes. After addition of 100 µl Tris-Cl pH=7.5 1M samples were homogenized by pipetting and allowed to reach room temperature. For protein aggregation 50 µl of SP3 beads (original slurry) were added, samples vortexed, combined with 1 ml ethanol 100 % and again mixed by inversion. After 15 minutes beads were captured on a magnetic rack, supernatants discarded and beads washed four times with 2 ml EtOH 70%, while attached to the magnet. Protein was digested off the beads overnight in 200 µl trypsin digestion buffer (EPPS 50 mM pH=8, 5 mM DTT, 10 µg trypsin per sample)

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

37 °C, 700 rpm shaking. Cysteines were alkylated by addition of 6 µl chloroacetamide 550 mM while the incubation was continued for 60 minutes. The beads were collected on a magnet for 5 minutes and the supernatant transferred to a fresh vial. Samples were topped off with 800 µl formic acid 1 % and peptides cleaned up with Sep-Pak C18 cartridges according to the manufacturer's instructions. Peptides were fractionated at high pH into 96 fractions and pooled to 48 samples as reported before.[117]

Treatment of photo-sensitized cells and UV irradiation

For the SILAC-based cataloguing of DNA interactors 1.5 million MCF7 cells were expanded on a 15 cm diameter culture dish for three days in the presence of 100 µM 4ST until approximately 70 % confluent. SILAC light cells were used as unirradiated control and SILAC heavy cells were UV irradiated for protein-DNA crosslinking. Therefore, the medium was decanted, cells washed once with 50 ml ice-cold PBS, covered with 20 ml ice-cold PBS and kept on ice. All subsequent steps were carried out in a cold-room at 4 °C. Photo-crosslinking occurred with the UV irradiation device UVEN for 60 seconds amounting to a combined irradiation energy of 125 J/cm^2 at 365 nm wavelength. Cells were scraped into the 20 ml ice-cold PBS they were irradiated in, pelleted by centrifugation, and stored at -80 °C until further use.

For the dose-dependent analysis of the DNA interactome in response to estrogen, DMEM without phenol red was used. After three days of expansion in the presence of 4ST MCF7 cells were washed twice with PBS, switched to serum-free DMEM supplemented with 4ST and cultivated for another 48 hours before addition of estrogen (17ß-estradiol) for 45 minutes from a 1:1000 stock in DMSO. Final estrogen concentrations were 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM, 1 pM, 300 fM, 100 fM, 30 fM and DMSO control. UV irradiation occurred as described above.

For comparing the DNA interactome of cells treated with genotoxic drugs MCF7 cell were expanded in the presence of 4ST for three days. Consequently, the medium was discarded and replaced by fresh media containing the genotoxic drugs or a DMSO mock control alongside 4ST for 24 hours. Because of the incompatibility of platinum compounds with DMSO,[115] cisplatin and oxaliplatin were directly dissolved to a final concentration of 100 µM in DMEM supplemented with 4ST and etoposide added from a 1:1000 stock in DMSO.

For the comparison of DNA interactomes in primary mouse cells, 4ST-labelled bone marrow-derived macrophages were challenged for 3 hours with 100 ng/ml LPS from a 1:1000 stock in PBS or only PBS as vehicle control. The medium was discarded and immediately 50 ml of ice-cold PBS added. UVEN irradiation was performed, and macrophages were harvested by scraping.

For the time course of DNA interactomes upon BAF inhibition MCF7 cells were expanded in the presence of 4ST for three days. To control for a consistent medium composition during the treatment, the medium was refreshed and culture continued for one day. BRM014 was added from a 1:1000 stock in DMSO to a final concentration of 1 µM and cells were returned to normal culture for 0, 4, 8, 12, 16 and 20 minutes. The medium was discarded, immediately 50 ml of ice-cold PBS added, and UVEN irradiation performed. For formaldehyde crosslinking cells were treated identically but 4ST was omitted. Because formaldehyde-crosslinking in the CHEP protocol requires 10 minutes (see below), only three time points were taken, 0, 10, and 20 minutes.

Protein-crosslinked DNA extraction (XDNAX) for proteomic quantification of the DNA interactome

Cell pellets from one confluent 15 cm diameter culture dish of MCF7 cells were lysed in 1 ml TRIZOL by pipetting until completely homogenous. Lysates were combined with 200 µl chloroform, mixed by inversion and incubated for 5 minutes at room temperature. Tubes were centrifuged with 12000 g for 10 minutes at 4 °C. The aqueous phase containing RNA was discarded and the interphase containing chromatin transferred to a fresh tube avoiding transfer of organic phase. The interphase was disintegrated in 1 ml recovery buffer (Tris-Cl pH=7.5 50 mM, EDTA 1 mM, SDS 1 %) until completely dissolved. DNA was precipitated by addition of 100 µl NaCl 5 M and 1 ml isopropanol, mixed by inversion, incubated for 15 minutes at -20 °C and spun down for 15 minutes with 20000 g at -11 °C. The supernatant was discarded and the pellet washed with 2 ml ethanol 70 %, centrifuged again for 5 minutes and ethanol removed to completion. For hydration 900 µl ultrapure water (ELGA) was added, the pellet detached from the wall of the tube by vortexing, incubated on a rotating wheel for 30 minutes at 4 °C, and further dissolved by pipetting. For RNA digestion 50 µl Tris-Cl pH=7.5 1 M was added along with 20 µl RNase A (0.5 µg/µl) and 20 µl RNase T1 (0.5 µg/µl) and incubated for 30 minutes at 37 °C, 700 rpm shaking. After addition of 5 µl SDS 20 % clumps were dissolved by pipetting and incubation continued for another 30 minutes. For DNA fragmentation one sample was distributed onto one column (8 wells) of a Covaris plate in 120 µl portions and sonicated using the parameters 50 cycles, Scan Speed: 1.0, PIP: 300, CPB: 50, AIP: 75, Dithering: Y=1 Speed=10. The sonicated sample was collected in a fresh tube and centrifuged with 5000 g for 5 minutes at room temperature. The supernatant was transferred to a fresh falcon tube, and the remaining pellet dissolved in sonication buffer (Tris-Cl 50 mM, SDS 0.1 %). Sonication and centrifuging was repeated until no pellet remained. The SDS concentration in the sonicated supernatants was adjusted to 2 % and the sample incubated at 95 °C for 5 minutes, 700 rpm shaking. The sample volume was doubled by addition of guanidinium thiocyanate (GuTC) 5 M, the sample mixed by vortexing and again incubated at 95 °C for 5 minutes, 700 rpm shaking. The volume was doubled again by addition of ethanol 100 %, the sample thoroughly mixed by vortexing and allowed to reach room temperature. The sample was successively applied to a silica spin column (Zymo-Spin IIICG Column) in 800 µl increments using a table top mini centrifuge with approximately 2000 g. We note here that overloading will impair removal of non-crosslinked protein and recommend to stop loading as soon as the column flow decreases notably. Columns were washed twice with 800 µl wash solution (GuTC 2.5 M, ethanol 50 %), and three times with 800 µl ethanol 70 % each time centrifuging 2 minutes with 10000 g. For elution of protein-DNA complexes column was transferred to a fresh tube and the drain hole blocked with parafilm. After addition of 200 µl nuclease elution mix (NEB nuclease

P1 buffer 1 x, $MgCl_2$ 5 mM, 0.5 µl NEB nuclease P1, 0.5 µg benzonase) samples were incubated over night at 37 °C, 700 rpm shaking. After addition of 20 µl SDS 20 % the incubation was continued for 30 minutes. The sample was recovered into a fresh tube by centrifugation with 16000 g for 5 minutes. Remaining DNA was eluted two more time with 200 µl elution buffer (Tris-Cl 50 mM, SDS 2 %) and the eluates combined. For protein cleanup 10 µl SP3 beads (original slurry) were added and vortexed, resulting in an approximately 600-640 µl of total sample volume. For protein aggregation 1 ml ethanol 100 % was added, the sample mixed by inversion and incubated for 15 minutes without shaking. The beads were collected on a magnet for 5 minutes and the supernatant discarded. Remaining on the magnet, the beads were washed four times with 2 ml ethanol 70 % and subsequently spun down to remove all residual ethanol. Protein was digested off the beads overnight in 200 µl trypsin digestion buffer (EPPS 50 mM pH=8, 5 mM DTT, 0.5 µg trypsin per sample) 37 °C, 700 rpm shaking. Cysteines were alkylated by addition of 6 µl chloroacetamide 550 mM while the incubation was continued for 60 minutes. The beads were collected on a magnet for 5 minutes and the supernatant transferred to a fresh vial. Resulting peptides are heavily contaminated with nucleotides and a polymer of unknown origin (probably PEG from spin columns), which can be removed by SCX StageTip but not C18. We therefore first cleaned up samples by SCX StageTip and then used an additional C18 cleanup for high-pH fractionation. For SCX StageTip 10 µl formic acid 10 % was added along with 100 µl acetonitrile (approximately 30 % final). SCX StageTips were preconditioned with 200 µl acetonitrile and 200 µl SCX wash solution (formic acid 0.1 %, ACN 30). Samples were loaded onto SCX StageTips and washed five times with 200 µl SCX wash solution. Peptides were eluted twice with 50 µl high-pH elution buffer (ammonium formate pH=10 250 mM, 50 % ACN) and dried by SpeedVac. For high-pH fractionation C18 StageTips were preconditioned with 200 µl ACN and 200 µl formic acid 0.1 %. Dried samples were applied in 200 µl formic acid 0.1 % and washed twice with 200 µl formic acid 0.1 %. Fractionation occurred with high-pH buffer (ammonium formate pH=10 50 mM) supplemented with 0, 5, 10, 15, 20, and 50 % acetonitrile. Following fractions were combined and dried down before LC-MS analysis, F1 5 + 50 %, F2 10 %, F3 15 %, F4 0+20 % acetonitrile. In case of the BRM014 time course, fractions were combined to three samples, F1 10 % + 50 %, F2 15 %, F3 0 % + 20 %.

Chromatin enrichment for proteomics (CHEP) from formaldehyde-crosslinked cells

Formaldehyde-crosslinked DNA interactomes were derived as described by Kustatscher et al.[5] Briefly, a 15 cm cell culture dish of MCF7 cells was rinsed once with PBS and fixed with 10 ml of 1 % formaldehyde in PBS for 10 minutes in the cell culture incubator. The solution was discarded and crosslinking quenched with 10 ml of 0.25 M glycine in PBS for 5 minutes. Cells were harvested by scraping into PBS. Nuclei were extracted by pipetting cell pellet in 1 ml ice-cold cell lysis buffer (Tris-Cl 25 mM, pH=7.4, 0.1 % Triton X-100, 85 mM KCl) to homogeneity, followed by centrifugation with 2300 g at 4 °C for 5 minutes. Nuclei were RNase digested in 500 µl cell lysis buffer with 100 µg RNase A for 15 minutes at 37 °C, 700 rpm shaking, followed by centrifugation with 2300 g at 4 °C for 10 minutes. The pellet was solubilized in 500 µl SDS buffer (Tris-Cl 50 mM, pH=7.4, 4 % SDS, 10 mM EDTA), before addition of 1.5 ml urea buffer (Tris-Cl 10 mM, pH=7.4, 8 M urea, 1 mM EDTA), thorough mixing, and centrifugation with 16000 g at 25 °C for 30 minutes. The step was repeated. The pellet was then solubilized in 2 ml SDS buffer and again centrifuged with 16000 g at 25 °C for 30 minutes. Finally, the pellet was covered with 1.5 ml 70 % EtOH, mixed by inversion, centrifuged with 20000 g at 25 °C for 5 minutes and washed once with 2 ml 70 % EtOH. The resulting chromatin pellet was allowed to expand in 500 µl ultrapure water before addition of 20 µl Tris-Cl 1 M and 2 µl $MgCl_2$ 1 M and decrosslinking for 1 hour at 95 °C, 700 rpm shaking. For DNA digestion 1 µg benzonase was added and incubation continued for 1 hour at 37 °C, 700 rpm shaking. Protein concentration was determined by BCA and 10 ug of protein used as input for SP3 protein cleanup, trypsin digestion, SCX StageTip, and high-pH fractionation on C18 StageTips into 4 fractions as described above for XDNAX samples.

HPLC-MS for the detection and quantification of proteomic samples

Peptide analysis occurred on an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Scientific) or an Orbitrap Lumos Tribrid mass spectrometer (Thermo Scientific), connected to a Dionex UltiMate 3000 RSLCnano system (Thermo Scientific) for nanoflow detection of XDNAX and CHEP samples. Microflow detection occurred on an Orbitrap Eclipse Tribrid mass spectrometer connected to a Vanquish Neo UHPLC system (Thermo Scientific) for the 48-fraction deep reference proteomes, and single shot full and nuclear proteomes during BRM014 treatment.

For nanoflow analysis, the sample was injected onto a trap column (75 µm x 2 cm) packed with 5 µm C18 resin (Dr. Maisch Reprosil PUR AQ) in solvent C (formic acid 0.1 %). Peptides were washed with solvent C at 5 µl/min for 10 minutes and subsequently transferred on to an analytical column (75 µm x 48 cm, heated to 55 °C) packed with 3 µm C18 resin (Dr. Maisch Reprosil PUR AQ) using a gradient of solvent A (formic acid 0.1 % in ultrapure water, DMSO 5%) and solvent B (formic acid 0.1 % in acetonitrile, DMSO 5%). Elution occurred across a 60-minute gradient at a flow rate of 300 nl/min starting from 4 % B followed by a linear increase to 32 % B. Nanosource voltage was 2000 V, ion transfer tube temperature 275 °C. Detection occurred with data-dependent acquisition using an OT-OT method and a cycle time of 2 seconds. MS1 resolution was 60000, scan range 360-1300, RF lens 40 %, AGC target 100 % and maximum injection time 50 ms. MS2 isolation occurred with a quadrupole window of 1.2 m/z and fragmentation with 30 % HCD energy. MS2 resolution was 30000, first mass 100 m/z, AGC 200 % and maximum injection time 54 ms.

For microflow analysis, the sample was directly injected onto an Acclaim PepMap 100 analytical column (2 µm particle size, 1 mm x 150 mm, heated to 55 °C) for 1.4 s at 100 µl/s flow of solvent A (formic acid 0.1 % in ultrapure water, DMSO 3 %) combined with 5.2 % of solvent B (formic acid 0.1 % in acetonitrile, DMSO 5%). Elution occurred at a flow rate of 50 µl /min starting from 5.2 % B, followed by a linear increase to 23 % B until 12.4 minutes, followed by a linear increase to 28% B until 13.6 minutes. Source voltage was

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

3500 V, ion transfer tube temperature 325 °C, vaporizer temperature 125 °C. Detection occurred with data-dependent acquisition using an OT-OT method and a cycle time of 0.9 s. MS1 resolution was 120000, scan range 360-1300, RF lens 40 %, AGC target 100 % and maximum injection time 50 ms. MS2 isolation occurred with a quadrupole window of 1.3 m/z and fragmentation with 28 % HCD energy. MS2 resolution was 15000, first mass 100 m/z, AGC 200 % and maximum injection time 22 ms.

QUANTIFICATION AND STATISTICAL ANALYSIS

MS database search

MS raw files were searched using MaxQuant. In addition, MSfragger was used to identify 4ST-modified peptides.

All MaxQuant settings were used at their default value, except for specifying SILAC configurations and indicating the appropriate number of fractions per sample. Label-free quantification with iBAQ was activated and used at default parameters. In case of the differential quantification of DNA interactomes from MCF7 cells treated with estrogen or genotoxic drugs the 'match between runs' option was activated and used at default parameters. In case of the time courses of BAF inhibition the 'match between runs' option as well as LFQ quantification was activated and used at default parameters. In case of the differential quantification of DNA interactomes from MCF7 cells treated with estrogen, as well as the SILAC-based quantification of four-hour etoposide treatment 'phospho STY' was added as additional variable modification. In these two cases the MaxQuant search was performed with default FDR settings and standard proteinGroups.txt and Phospho(STY)Sites.txt tables used for further analysis. In case of the MCF7 DNA interactome, the deep reference proteomes and the quantification of DNA interactomes in response to genotoxic drugs the MaxQuant search was performed 100 % FDR followed by PROSIT rescoring.[118] Therefore, 'protein FDR' and 'peptide FDR' were set to 1 and the search results rescored by an in-house pipeline of PROSIT followed by picked-FDR thresholding.[119]

MSfragger was used as part of FragPipe. Raw files for the SILAC-controlled MCF7 DNA interactome were first surveyed with an open search using parameters preset in the software. Precursor mass tolerance was set to -150 – 1000 Da, add_C_cystein was set to 57.021464, add_R_arginine was set to 10.0083, add_K_lysine was set to 8.0142, and the export format set to TSV_PEPXML_PIN. In the validation section Crystal-C was activated, and PSM validation set to the recommended open search parameters. In the FDR filter and report section 'Generate peptide-level summary' was selected. All DNA interactome data was also searched for the 321 Da phospho-4ST-H_2O modification in an offset search with parameters at their default values. Mass offsets were set to '0 321.0316', and the modification add_C_cystein to 57.021464, as well as add_R_arginine to 10.0083, add_K_lysine to 8.0142 for the SILAC-controlled experiment. In the validation section Crystal-C was de-activated, and PSM validation set to the recommended offset search parameters.

Processing and analysis of proteomic data

All proteomic data was processed in R(4.2.2) using RStudio. For the analysis of proteomic data rescored with PROSIT the PickedFDR proteinGroups_fdr0.01.txt file was used. MaxQuant contaminants were filtered by removing all 'CON__' entries. In case of the standard MaxQuant searches including the phospho-STY modification the proteinGroups.txt table was used and filtered to remove 'Potential contaminants', 'Reverse' matches to the decoy database, as well as proteins 'Only identified by site'.

In order to determine the MCF7 DNA interactome each SILAC-controlled replicate of was processed independently. The aim of our enrichment strategy was to eliminate as much non-crosslinked protein (SILAC light) and recover as much DNA-crosslinked protein as possible (SILAC heavy). As a proxy for protein quantities in each SILAC channel iBAQ values were used.[34] Proteins with only iBAQ intensity in the irradiated SILAC channel were immediately called as DNA interacting in this replicate and proteins with only iBAQ intensity in the unirradiated SILAC channel as non-crosslinked protein. For all other proteins iBAQ rations of $\log_2$(irradiated (SILAC heavy) / unirradiated (SILAC light)) were calculated and the apex of the resulting distribution determined to account for mixing errors between the SILAC channels (Figure S3A). The left side of this distribution was then mirrored to create a symmetrical lognormal distribution estimating iBAQ ratios among non-crosslinked proteins. Three standard deviations from the apex of this mirrored distribution were used a replicate-specific cutoff to call DNA interacting proteins. Proteins called DNA interacting in at least three out of five replicates were included in the MCF7 DNA interactome (Table S1).

Deep reference proteomes were analysed to determine the MCF7 nuclear and cytosolic proteome using iBAQ quantification, which were normalized by median centering between the four replicates of one experiment. The imperfect hypotonic nuclear fractionation and our exhaustive peptide fractionation strategy led to the detection of most MCF7 proteins contained in the full proteome also in the cytosolic fraction, albeit at very different abundance (Table S1). Because we were focused on the proteomic surrounding of DNA we defined the nuclear proteome using a quantitative cutoff oriented on transcription factors. Figure S3C shows that the iBAQ ratio nucleus / cytosol for transcription factors was in 89 % of cases higher 0.1. The remaining 11 % contained mostly transcription factors know to be naturally localized to the cytosol, such as STAT3, IRF3, AR, NFKB1, RELA etc., so that we used a cutoff of 0.1 to call proteins present in the nucleus. Vice versa, proteins with an iBAQ ratio smaller 0.1 were defined as cytosolic.

For the differential analysis of the single-dose estrogen treatment as well as the genotoxic drug treatments we used iBAQ quantification. Under the prerequisite that iBAQ values represent counts of protein molecules we applied NBM-based differential analysis with DESeq2.[104] Foldchanges were corrected using the apeglm package.[120] Because of the extreme changes in the DNA interactome of cells treated with genotoxic drugs, where many interesting proteins such PSMC2, SELENOH, CREB1, TP53, CHEK2 etc. (Table S4) were undetected in one of the conditions, we applied imputation of missing values in order to include these proteins in

Please cite this article in press as: Trendel et al., The human proteome with direct physical access to DNA, Cell (2025), https://doi.org/10.1016/j.cell.2025.04.037

our differential analysis. Therefore, we used an adaptation of the imputation function described for PERSEUS,[121] which draws random numbers from a down-shifted normal distribution with shrunken standard deviation using the parameters width 0.3 and downshift 1.8. For the differential analysis of the BRM014 time course we used MaxQuant's LFQ, which performed more consistently across time points than iBAQ.

For dose-response analysis of DNA interactomes in response to estrogen iBAQ quantification was used (Table S3). Therefore, protein abundances from all DNA interactomes in the analysis were first normalized by median centering. To avoid overfitting quadruplicates of the highest estrogen concentration and the DMSO control were collapsed into one replicate each using their median. For fitting a log-logistic model the R package 'drc' was used.

Processing and analysis of DNA sequencing data

For the comparison of transcription factor binding sites in ENCODE DNA accessibility sequencing data transcription factors motifs were acquired from Factorbook[95] and filtered for MCF7 and a MEME p-value<0.01. MCF7 ATAC-Seq, DNase-Seq and FAIRE-Seq bed files were downloaded from ENCODE,[122] accessible regions extracted from the hg38 or hg19 genome using samtools and Factorbook consensus motifs counted across all sequences allowing for two mismatches using the Bioconductor package Biostrings and the vcountPattern function.

ChIP-seq data for Atf3 and Cebpb in bone marrow-derive macrophages (BMDMs) upon vehicle or LPS stimulation were obtained from the gene expression omnibus (GEO) and mapped to the murine reference genome with BWA-MEM. PCR duplicates were removed using Picard Tools. For visualization, bam files were filtered for properly paired and mapped reads, and multimappers were removed with samtools. Alignments were converted to bigwig files, merging 10 bp per bin using 'bamCoverage' from the deepTools and scaled for sequencing depth. Peaks were called with MACS with a FDR cut-off of 0.05 over matched input controls. Blacklisted regions (http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse/mm10.blacklist.bed.gz) were removed from analyses. Peak annotation was performed with the Bioconductor package ChIPpeakAnno. Tracks were visualized with the Integrative Genome Browser (IGV). For peak overlaps, reproducible peaks in at least two replicates were extracted with the Bioconductor package GenomicRanges. Peak regions were defined as overlapping when sharing at least 10 bp. GO enrichment analysis was performed using the Bioconductor package clusterProfiler. The top 10 enriched GO terms under biological process are displayed, sorted by gene ratio (proportion of set genes enriched in GO term).

Analysis of live cell imaging data

Growth inhibition by the photo-activatable nucleotides 4ST and 4SU towards MCF7 or U2OS cells was analyzed by summing up the confluence data recorded over five days for each individual replicate of each concentration. The dose-dependent induction of apoptosis or cell death by cisplatin (Figures 6C and 6D) was analysed in the same way by summing up fluorescence measurements for annexin-V or oxodazol yellow (see above) over the 3 days of the knockdown experiment. The R package 'drc' was used to fit a log-logistic dose-reponse model and derive effective concentrations.

ADDITIONAL RESOURCES

Functional annotation of proteins and data visualization

For all cross-references with protein databases UniProt identifiers were used, except for OMIM where gene names were used. Proteins were annotated with their gene ontology (GO) terms via ENSEMBL BioMart accessed via the R package 'biomaRt'. We note here that for clarity GO was used as only source to annotate proteins as 'DNA binding' or 'RNA binding'. In some cases this varies from their annotation in UniProt, which integrates annotation from different sources. GO enrichment analysis was performed using the GOrilla web interface. Protein structures were visualized in UCSF Chimera. Data was plotted using the R package 'ggplot2'. Schemes in Figures 2A and 3J were created in part with Biorender.

CellPress
OPEN ACCESS

Supplemental figures

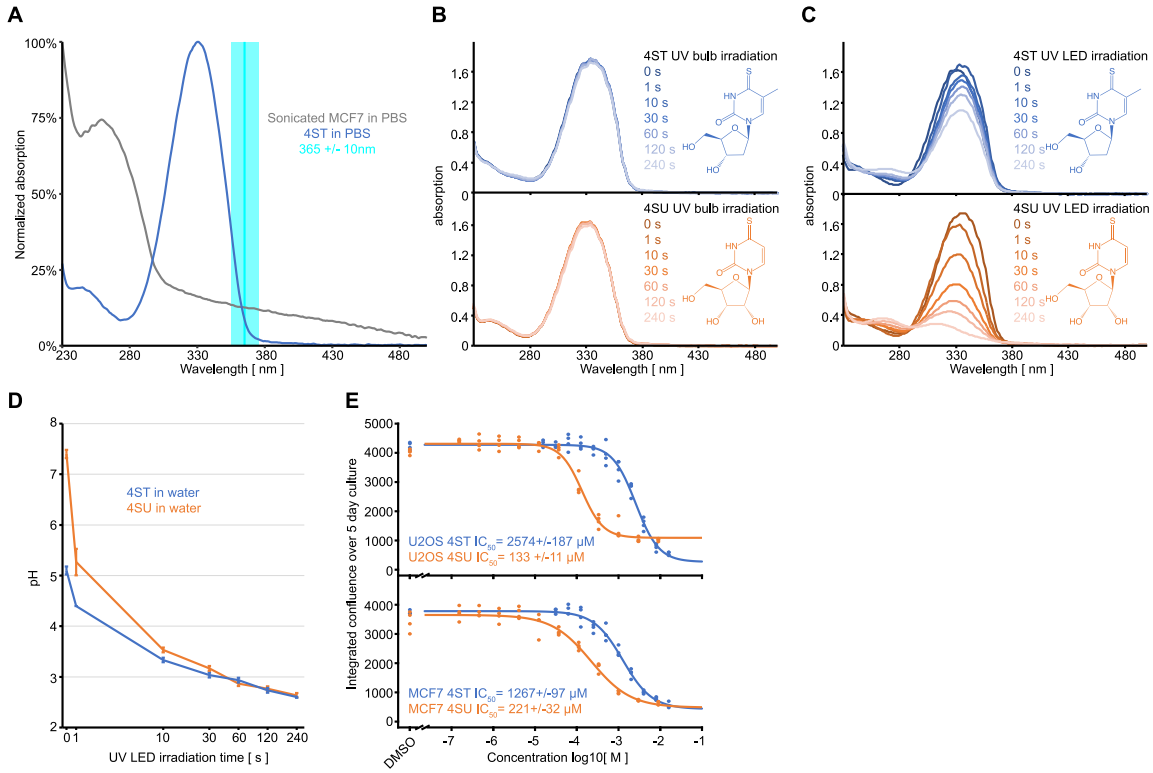


Figure S1. Photo-activation of 4ST and metabolic labeling of living cells, related to Figure 1

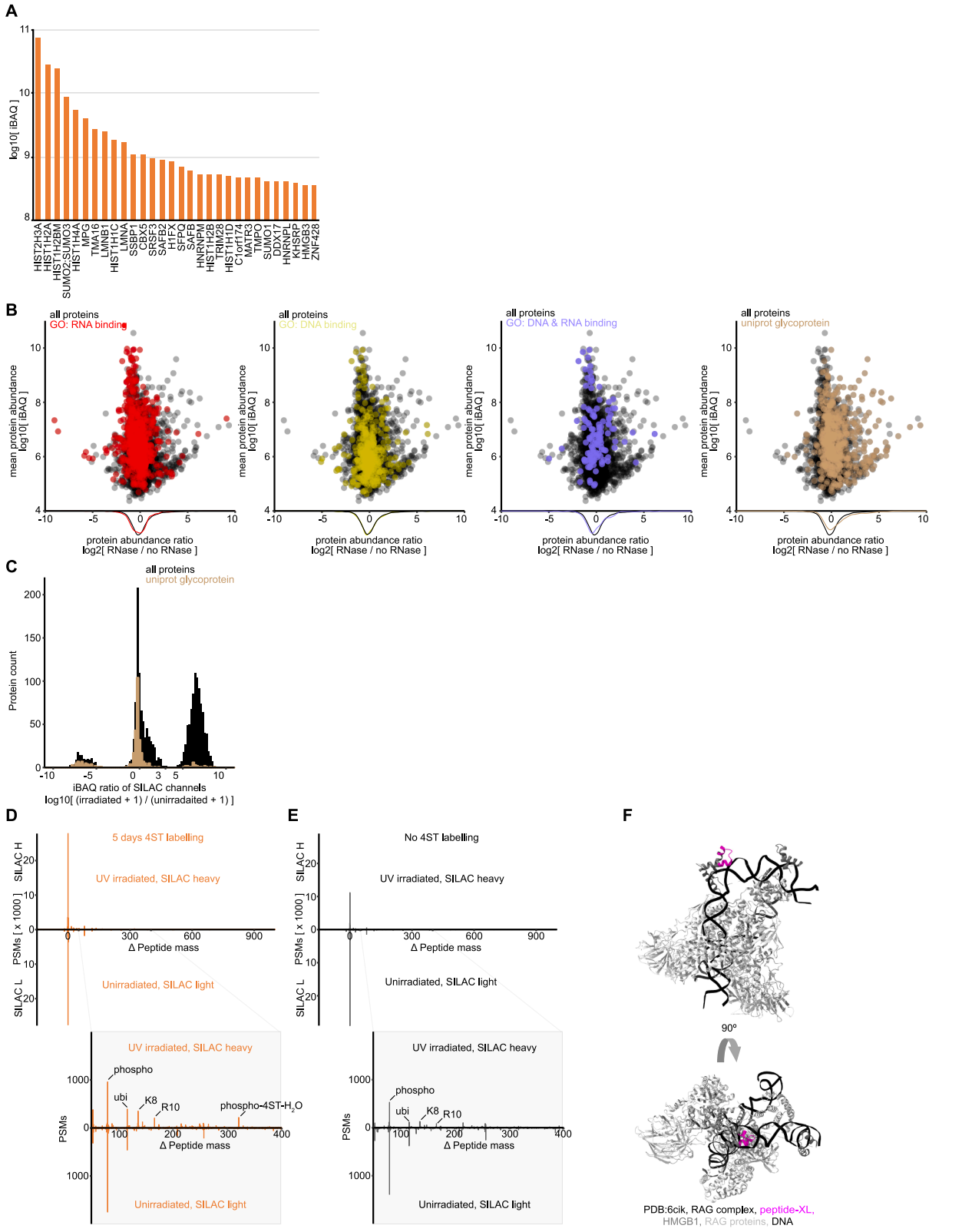
(A) Line plot comparing the light absorption of 4ST in PBS to MCF7 cells sonicated to homogeneity in PBS.

(B) Absorption spectra of 4ST (top) and 4SU (bottom) irradiated for indicated time, with a conventional UV bulb irradiation device. See also Figure 1D.

(C) Same as in (B) but irradiation with UVEN. See also Figure 1D.

(D) Line plot showing pH change of 4ST or 4SU solutions upon UVEN irradiation.

(E) Dose-response analysis for the growth inhibition of U2OS (top) or MCF7 cells (bottom) by 4ST (blue) or 4SU (orange) added to the culture medium. Cell confluence was monitored by live-cell imaging and integrated over 5 days of culture. Points indicate individual replicates and the line a log-logistic model.



(legend on next page)

Figure S2. Proteins and nucleotide-peptide hybrids enriched by XDNAX, related to Figure 2

(A) Bar plot illustrating the most abundant proteins identified in protein-crosslinked DNA enriched by XDNAX. Displayed are the top 30 proteins from the irradiated SILAC channel of Figure 2C ranked by iBAQ.

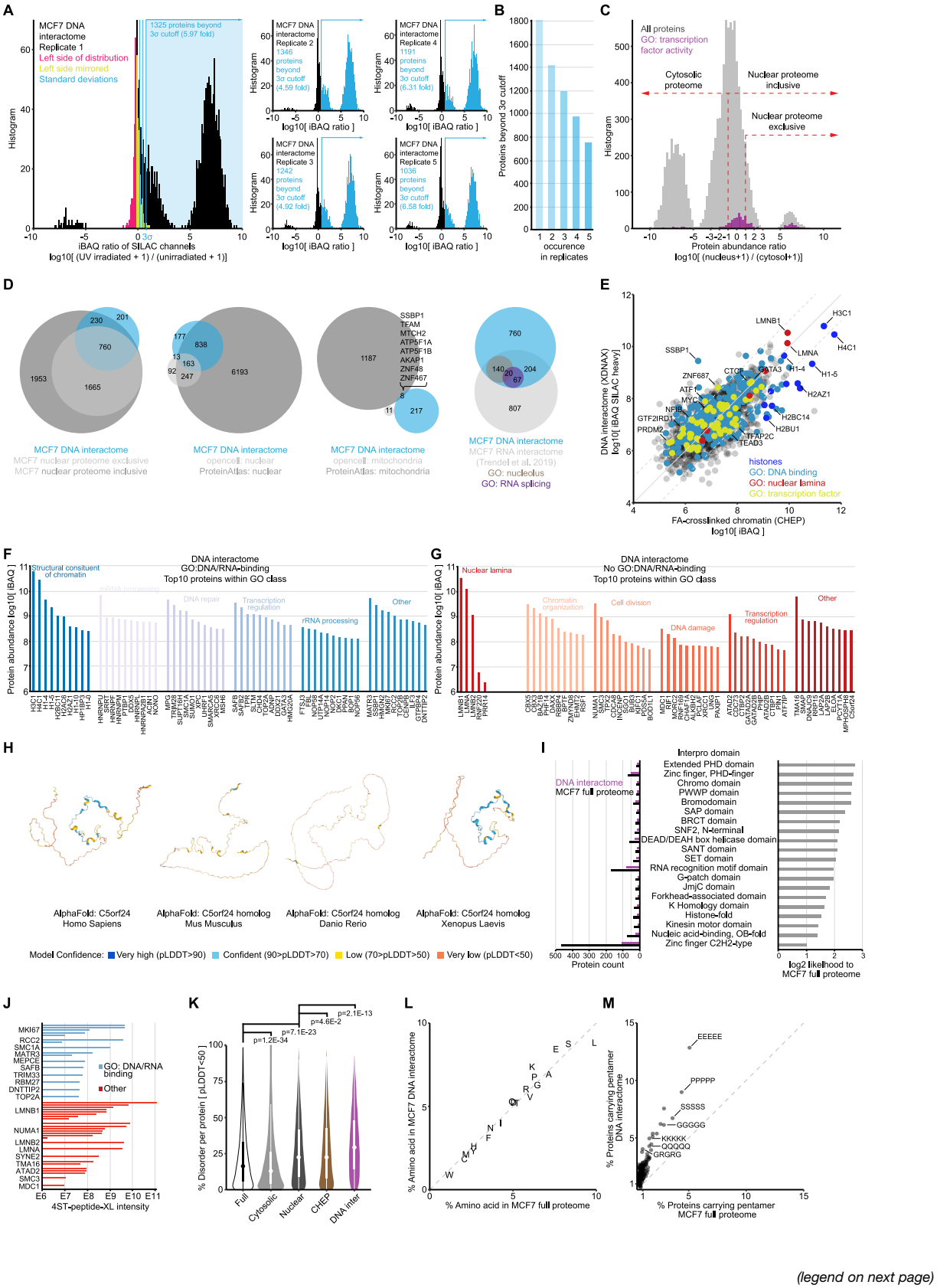
(B) Scatterplot comparing protein abundances in XDNAX samples, with and without RNase treatment. Samples were prepared from MCF7 cells (no SILAC label) according to Figure 2A, with or without RNase digestion in the second step, and compared by label-free quantification. Four panels show the identical data highlighting proteins with the indicated GO annotations or proteins annotated as glycosylated.

(C) Histogram illustrating glycoproteins as a constant background of XDNAX indifferent to irradiation. Same as in Figure 2C but only data from 4ST-labeled cells are shown (Figure 2C top, orange). Glycoproteins occur with very similar abundance in the SILAC light (unirradiated) and SILAC heavy (irradiated) channel, indicating that their enrichment occurs independent of photo-crosslinking.

(D) Histogram of peptide-spectrum matches (PSMs) identified in an open search by the mass-tolerant search engine MSfragger. The magnified area highlights the occurrence of mass adducts corresponding to protein phosphorylation (phospho), ubiquitination (ubi), heavy lysine and arginine (K8, R10, and SILAC artifacts), and a new modification corresponding to 4ST with a water loss (4ST-H_2O).

(E) Same as (B) but for cells without 4ST labeling. See Figure 2D.

(F) Crystal structure of the RAG1 pre-reaction complex containing HMGB1 (PDB: 6CIK). The nucleotide-crosslinked peptide from HMGB1 identified in our DNA interactome (magenta) localizes to the direct vicinity of the DNA (black).



(legend on next page)

Figure S3. Defining the direct DNA interactome as well as the nuclear, cytosolic, and full proteome of MCF7 cells using quantitative proteomics, related to Figure 3

(A) Histograms showing the filtering process for the generation of the direct DNA interactome from five replicates of photo-crosslinked MCF7 cells subjected to XDNAX. For each replicate, a background distribution was estimated by mirroring negative log fold changes under the assumption that non-crosslinked proteins exhibit the same intensity variance in both SILAC channels. We used three standard deviations (3σ) of this distribution as fold-change threshold for calling candidate DNA-interacting proteins in each replicate.

(B) Bar plot comparing occurrence of proteins across replicates. For the final DNA interactome shown in Figure 3A and used throughout the rest of our analysis, only candidate DNA-interacting proteins occurring in three or more replicates were considered.

(C) Histogram of protein abundance ratios between the MCF7 nuclear and cytosolic proteome. An arbitrary abundance cut-off was chosen to include as many transcription factors as possible in the nuclear proteome while excluding transcription factors known to be constitutively localized to the cytosol. For details see STAR Methods.

(D) Venn diagrams comparing the MCF7 DNA interactome to various subproteomes. Left: comparison to the MCF7 nuclear proteomes from (E). Middle: comparison to the groups of proteins annotated as "nuclear" or "mitochondrial" in the two databases ProteinAtlas[97] and OpenCell.[98] Right: comparison to MCF7 RNA interactome[14] in the overlap the two largest groups sharing a common GO term are highlighted (nucleolus, RNA splicing).

(E) Histogram of protein abundance ratios between the MCF7 nuclear and cytosolic proteome. An arbitrary abundance cutoff was chosen to include as many transcription factors in the nuclear proteome as possible (inclusive), while excluding transcription factors know to be constitutively located in the cytosol (STAT3, IRF3, AR, NR3C1, RELB, RELA, NFKB2, NFKB1, etc.). For comparison in (D) a second more stringent cutoff was chosen, which includes only proteins with higher abundance in the nucleus than the cytosol (exclusive). For details see STAR Methods.

(F and G) Bar plot showing the top ten proteins within each GO term in Figure 3A with the highest protein abundances. Shown are means of the iBAQ from the SILAC channel of photo-crosslinked cells (SILAC heavy).

(H) Protein structure predictions by AlphaFold for the human C5orf24 and its homologs in mice, zebra fish, and frogs.

(I) Bar plots comparing the occurrence (left) and relative likelihood (right) of InterPro domains in the DNA interactome relative to the MCF7 full proteome.

(J) Bar plot comparing MS intensities of crosslinked DNA-peptide hybrids identified in the MCF7 DNA interactome. Displayed are peptides from the top ten proteins with the highest intensity crosslinks, with and without prior GO annotation for nucleic acid binding.

(K) Violin plots comparing the percentages of amino acid positions predicted as disordered according to their AlphaFold prediction confidence (pLLDT < 50).[41] Testing occurred with a two-sided Kolmogorov-Smirnov test. Proteins in the MCF7 full proteome (full) are compared with the cytosolic and nuclear proteome (see F) and the DNA interactome (DNA interactome), and its parts annotated as nucleic acid binding (DNA/RNA) or lacking this annotation (other).

(L) Scatterplot comparing the amino acid frequency in proteins of the MCF full proteome to the DNA interactome.

(M) Scatterplot comparing the occurrence of disordered amino acid pentamers in proteins of the MCF7 full proteome to the DNA interactome. All possible pentameric permutations of the amino acids G, S, D, Q, P, E, K, and R were counted if they occurred at least once in a protein in the two groups.

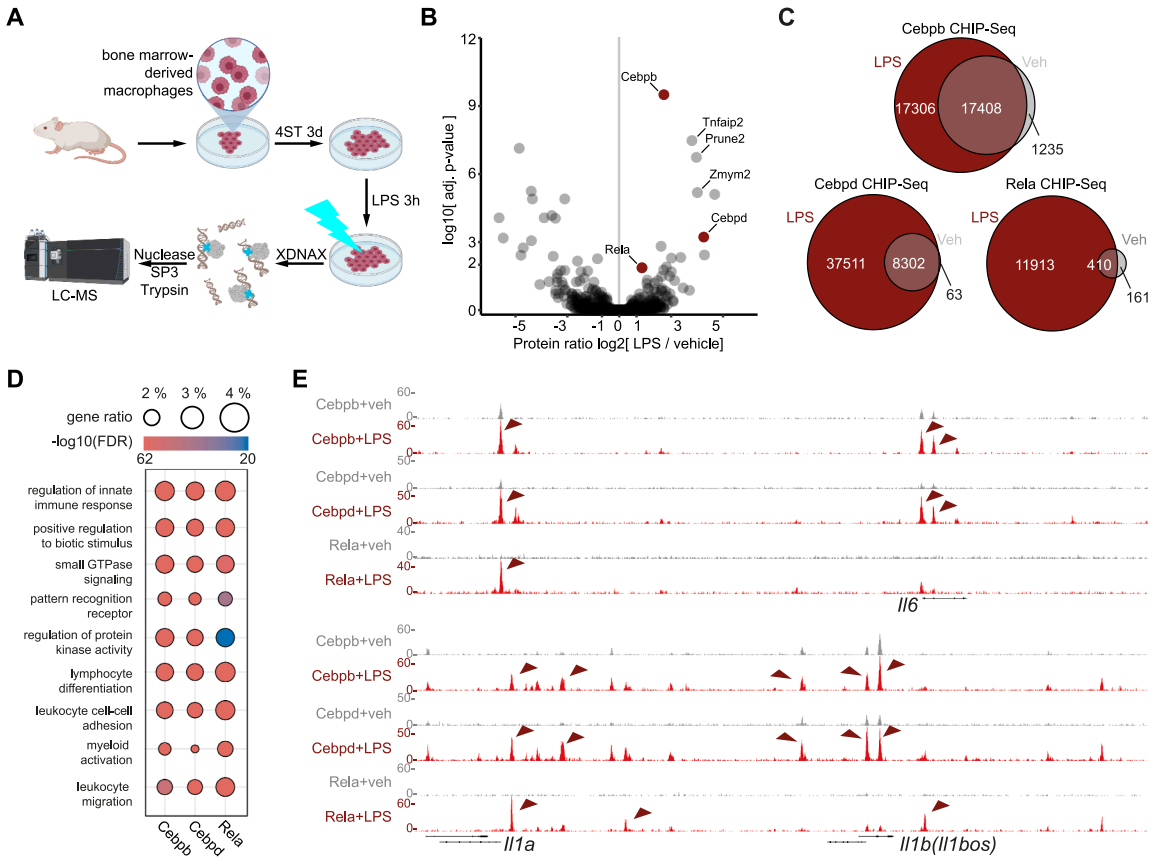


Figure S4. Quantitative comparisons of DNA interactomes in primary mouse cells, related to Figure 4

(A) Experimental scheme for the quantitative comparison of DNA interactomes from mouse bone-marrow-derived macrophages upon LPS challenge. Mononucleated cells are harvested from mouse bone marrow and differentiated *ex vivo* alongside 4ST-labeling. Resulting macrophages are challenged with lipopolysaccharide (LPS) or mock-treated (PBS) before photo-crosslinking, XDNAX, and LC-MS quantification.

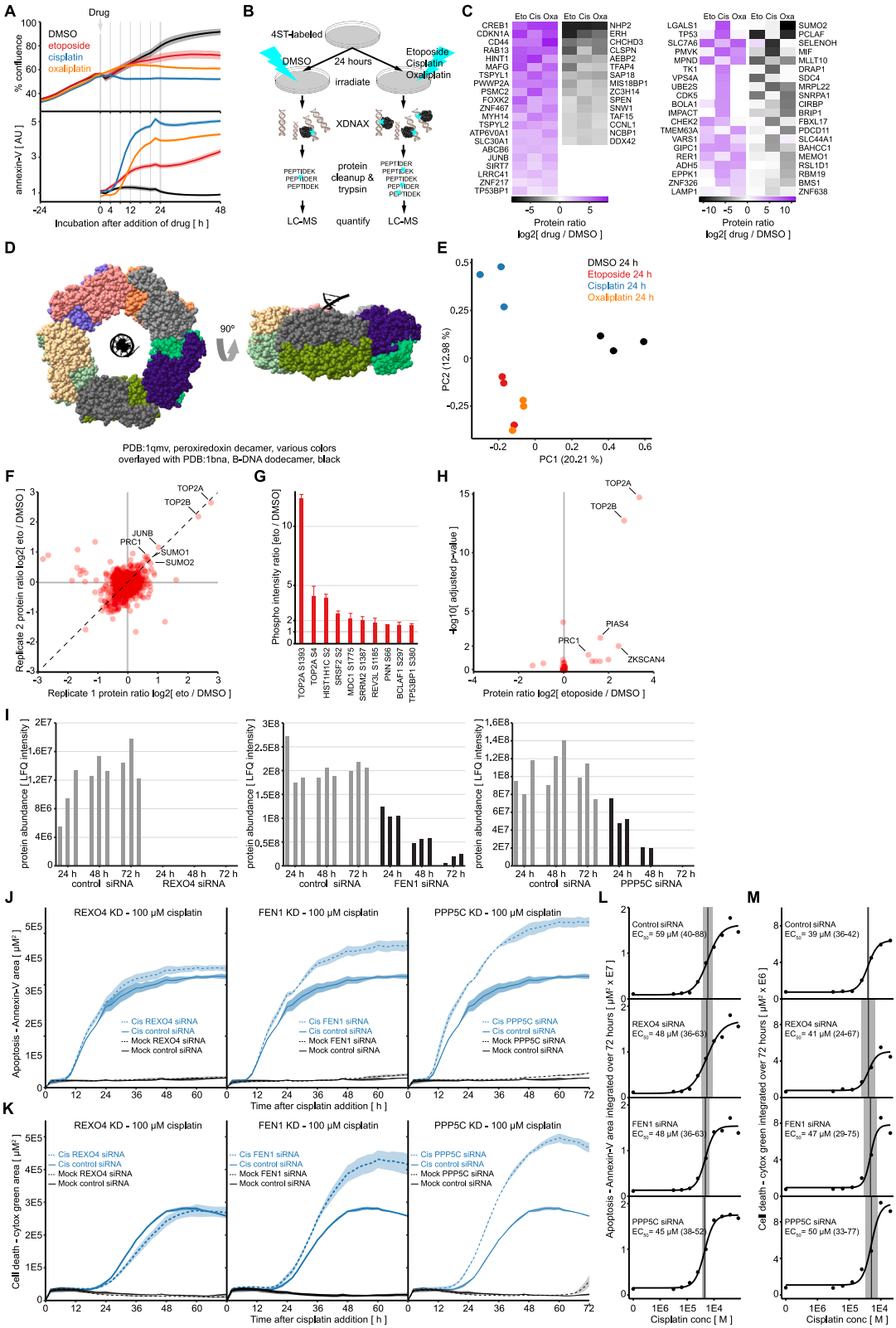
(B) Volcano plot illustrating changes in the DNA interactome after 3 h of LPS exposure. Highlighted are the key inflammation mediators Cebpb, Cebpd, and Rela.

(C) Venn diagram of peaks in ChIP-seq data from mouse bone-marrow-derived macrophages before and after LPS challenge for proteins highlighted in (B).

(D) GO enrichment (biological process) for LPS-specific peaks from (C).

(E) Exemplary genome browser tracks for two pro-inflammatory genes from (C) (Il6, Il1a, and Il1b) gaining interaction upon LPS stimulation. Arrows indicate peaks with FDR < 0.05.

CellPress
OPEN ACCESS



(legend on next page)

Figure S5. Changes in the direct DNA interactomes from breast cancer cells exposed to different genotoxic drugs, related to Figure 5

(A) Line plot showing proliferation (top, confluence) and induction of apoptosis (bottom, annexin-V fluorescence) monitored by live-cell imaging. After 24 h of normal expansion, genotoxic drugs were added at 100 μM concentration. Shown are the mean of ten replicates as bold line and their standard error as shading.

(B) Experimental outline for the comparison of proteins interacting with DNA after 24 h of treatment with a genotoxic drug or mock control (DMSO).

(C) Left: heatmap displaying relative abundance ratios of all proteins with significantly changed DNA interaction across all three genotoxic drugs compared with the mock control (adj. $p < 0.01$ in all three treatments). Right: heatmap displaying relative abundance ratios of proteins with significantly changed DNA interaction in maximal two treatments (adj. $p < 0.01$ not in all three treatments). The top 20 proteins with the most extreme increase (left) and decrease (right) are displayed.

(D) Protein structure overlay of the peroxiredoxin II decamer (PDB: 1qmv) and a DNA twelve-mer (PDB: 1bna) visualizing that, by proportion, the DNA helix could be encompassed by a PRDX2 toroid.

(E) PCA analysis of the triplicate DNA interactomes from MCF7 cells using all protein abundances (iBAQ) as input. Missing values were imputed to include extreme changes (e.g., TP53, CREB1, PCLAF, etc., see Table S4).

(F) Scatterplot comparing normalized SILAC ratios between replicates of DNA interactomes from MCF7 cells treated with 100 μM etoposide for 4 h or a DMSO control. SILAC labels were swapped between the replicates.

(G) Bar plot showing intensity ratios for the ten phospho-peptides quantified in the experiment in (F), with the most extreme fold changes of etoposide over untreated. Error bars indicate composite standard deviations from four replicates.

(H) Volcano plot illustrating changes in the DNA interactome after 4 h of etoposide treatment. Protein abundances in each SILAC channel (iBAQ) from four biological replicates were used.

(I) Bar plot comparing the expression of target proteins in MCF7 cells during small interfering RNA (siRNA)-mediated knockdown.

(J) Line plots following the induction of apoptosis in MCF7 cells with knockdowns for REXO4, FEN1, or PPP5C using fluorescence-labeled annexin-V in live-cell imaging. One uniform pool of cells was transfected with siRNAs 24 h before addition of 100 μM cisplatin (cis) or drug-free media (mock). Bold lines indicate cells transfected with control siRNAs, dotted line for one specific siRNA target, shading one standard deviation.

(K) Same as in (A) but for the quantification of dying cells using the fluorescence dye cytotox green.

(L and M) Dose-response curves using the integrated florescence signal over 72 h to determine half-effective concentrations for cisplatin to induce apoptosis (J) or cell death (K) during the duration of the indicated knockdown in MCF7 cells.

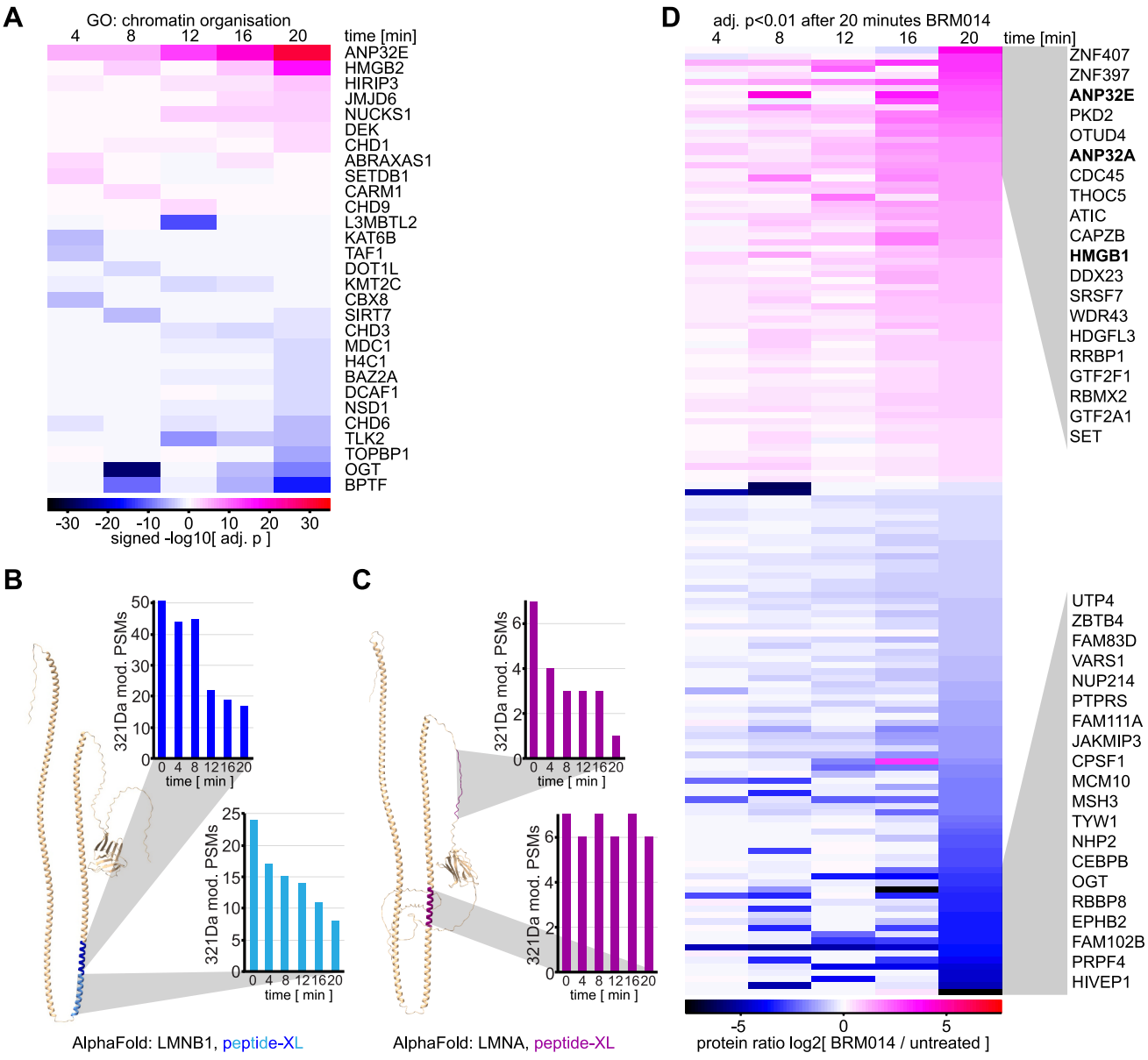


Figure S6. Changes in the photo-crosslinked DNA interactome of MCF7 cells during small-molecule BAF inhibition, related to Figure 6

(A) Heatmap of signed *p* values for changes in the photo-crosslinked DNA interactome. Displayed are proteins with the GO annotation "chromatin organization" with significantly (adj. $p < 0.01$) changed abundance in the DNA interactome at any time point.

(B) Bar plots comparing the occurrence of 4ST-crosslinked peptides (321 Da modification) mapping to the IF rod domain of LMNB1 (highlighted blue in AlphaFold structure) between time points.

(C) Same as in (B) but for LMNA. A second peptide in the tail domain is highlighted, mapping to a region that has been associated with Emery-Dreifuss muscular dystrophy (R401).

(D) Heatmap of abundance ratios in the DNA interactome at each time point of BRM014 treatment for proteins with significantly (adj. $p < 0.01$) changed abundance on DNA at the 20-min time point.