Title

Chromatin-contact atlas reveals disorder-mediated protein interactions

Short title: Disordered regions mediate chromatin interactions

Mahmoud-Reza Rafiee[1], Julian A Zagalak[1,2]†, Sviatoslav Sidorov[1]†, Jernej Ule[1,2]*, Nicholas M Luscombe[1,2,3,4]*

[1]The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. [2] Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK. [3] UCL Genetics Institute, University College London, Gower Street, London WC1E 6BT, UK. [4] Okinawa Institute of Science & Technology Graduate University, Okinawa 904-0495, Japan

† Authors contributed equally

* Corresponding authors

Abstract

RNA-binding proteins (RBPs) play diverse roles in regulating co-transcriptional RNA-processing and chromatin functions, but our knowledge of the repertoire of chromatin-associated RBPs (caRBPs) and their interactions with chromatin remains limited. Here, we developed SPACE (Silica Particle Assisted Chromatin Enrichment) to isolate global and regional chromatin components with high specificity and sensitivity, and SPACEmap to identify the chromatin-contact regions in proteins. Applied to mouse embryonic stem cells, SPACE identified 1,825 chromatin-associated proteins, >40% of which are annotated as RBPs, indicating their dual roles in chromatin and RNA-binding. SPACEmap showed that about half of chromatin contacts of proteins are mediated by intrinsically disordered regions (IDRs). Additionally, SPACEmap strictly verified chromatin-binding of 422 RBPs by locating their chromatin-contact regions. As an example, we show Dazl co-localization with PRC2 on transcription start sites revealing an independent function from Dazl 3'-UTR-binding. Together, our approach provides an unprecedented structural insight into IDR-mediated interactions in native chromatin.

Teaser: Most chromatin contacts are mediated by intrinsically disordered regions, especially in RBPs.

Introduction

RBPs participate in regulating transcription as well as other aspects of co-transcriptional RNA regulation (*1, 2*). Indeed, it is known that transcriptional and post-transcriptional processes are integrated to coordinate alternative splicing and polyadenylation (*3, 4*), RNA stability (*5, 6*) and subsequent translation in the cytoplasm (*7*). Furthermore, RBPs promote biomolecular condensate formation, and were reported to contribute to the functionality of enhancers, transcription factors and RNA Pol II (*8-10*). **Considering all these potential RBP-chromatin interactions**, **the question is which RBPs join the repertoire of chromatin-associated proteins.** This is particularly important as changes in the dynamics of such RBPs are implicated in cancer and neurodegenerative diseases (*11, 12*).

Global UV-crosslinkable RNA interactome capture based on oligo-dT capture, click chemistry or organic phase separation have identified over ~2,300 candidate RBPs (*13*). However, these methods are not able to distinguish those RBPs that associate with chromatin (chromatin-associated RBPs, caRBPs). ChIP-seq has been used to assess the association of dozens of RBPs with chromatin (*2, 14*), but its application is limited by the availability and specificity of antibodies. Thus, **methods are needed that provide a global view of caRBPs** with high specificity and throughput.

Traditionally, chromatin is isolated by cellular fractionation and precipitation (*15*). However, the results are ambiguous due to the abundant cytoplasmic contaminations that remain in the nuclear fraction, and precipitate together with chromatin. To make it more specific, DNA-labelling by ethynyl deoxy-uridine (EdU) was implemented to isolate chromatin fragments by click-chemistry and streptavidin beads (*16, 17*). However, incorporation of modified nucleotides into DNA can't preserve the natural conditions of chromatin. Additionally, current chromatome methods are

unable to determine the chromatin-protein contact sites, which is essential to reliably understand how proteins are integrated to the chromatin network.

Here, we present SPACE (Silica Particle Assisted Chromatin Enrichment), a straightforward and highly sensitive method that relies on silica magnetic beads for chromatin purification. To demonstrate the power of the method, we evaluated SPACE by studying the global chromatin composition of mES cells. We successfully identified previously reported DNA- and chromatin-binding proteins, as well as many caRBPs. **Surprisingly, RBPs comprise >40% of the proteins obtained from chromatome.** To understand how RBPs bind to chromatin, we developed SPACEmap. We found that intrinsically disordered regions (IDRs) are frequently used by chromatin proteins, including caRBPs, for chromatin-binding. Together, we demonstrate that the various applications of SPACE provide flexible, highly sensitive and accurate approaches for studying dynamics of chromatin-associated proteins, which has proven particularly valuable to expand the knowledge of RBP-chromatin interactions.

Results

SPACE is a rapid, flexible and stringent method to enrich for chromatin-associated proteins

Silica matrices (columns or beads) are widely used to purify DNA in diverse contexts, but they have not been applied to chromatin purification yet. We reasoned that regions of DNA are likely to remain accessible even after formaldehyde crosslinking of proteins. Initially, we tried to purify crosslinked chromatin by silica columns, however, the yield was almost zero (data not shown); therefore, we used silica magnetic beads instead of columns. **SPACE** - which stands for Silica Particle Assisted Chromatin Enrichment - **exploits the capacity of silica magnetic beads to**

purify formaldehyde-crosslinked chromatin in the presence of chaotropic salts (Figure 1A). We prepared non-crosslinked negative controls in a similar way to routine DNA purification, which is normally free of contaminating proteins. By applying SILAC-labelling to the crosslinked (heavy SILAC) and non-crosslinked (light SILAC) samples before adding silica magnetic beads, we were able to determine whether a protein is isolated due to crosslinking or non-specific association to the beads and other proteins.

SPACE is stringent, yet fast and flexible, and requires little starting material. Starting with as few as 20,000 cells, SPACE takes approximately 1h from the cell lysis to the start of protein digestion; it employs denaturing reagents to efficiently remove contaminants (4M guanidinium isothiocyanate, 2% Sarkosyl, 80% ethanol and 100% acetonitrile) and extensive RNase treatment (RNase A, 100ug, 30min at 37 C) to remove RNA-dependent interactors. The method is readily extended to identify chromatin-binding sites of proteins by a two-step digestion strategy (SPACEmap, Figure 1B). Additionally, SPACE can be combined with SICAP (*18*) as a double purification and highly stringent variation of the method (Supplementary Figure 1), or with ChIP to identify co-localized protein on chromatin (ChIP-SPACE, Supplementary Figure 5).

SPACE reveals RBPs as major chromatin components

We first applied SPACE to mouse embryonic stem (mES) cells cultured in 2iL (2i + LIF) and identified 1,825 significantly enriched proteins (Log2FC > 1 and adj. p-value < 0.1, moderated t-test adjusted by Benjamini-Hochberg (BH) method) compared with the negative control (Figure 2A, Table S1_SPACE). We then rigorously characterised the enriched proteins using keyword searching in gene ontology terms and protein information obtained from Uniprot and DAVID databases (Figure 2B and Supplementary Figure 2A). Based on relative abundances, proteins involved in chromatin remodelling, chromatin modification, DNA replication and repair, transcription factor activity, transcription, cell cycle, splicing, RNA processing and RNA

transport comprise 88% of the isolated proteins. Apart from those, proteins involved in translation, metabolic process, cell adhesion, protein folding, and protein transport make up 8% of the enriched proteins. Thus, using SPACE potential true positive biological processes are enriched 11-fold more than potential false positive terms. Remarkably we identified 54 proteins that are involved in pluripotency or ES cell processes, including Oct4, Sox2 and Nanog as the core circuitry of pluripotency. In addition, we identified 16 proteins that are part of the polycomb group proteins (Supplementary Figure 2A).

We grouped proteins into three categories based on their gene ontology annotations (Figure 2C): 1) 645 (36%) known DNA or chromatin-binding proteins; 2) 883 (48%) proteins present in the nucleus but not annotated as DNA- or chromatin-binders; 3) and 297 (16%) other "unexpected" proteins, a large proportion of which are involved in translation. Weighted by relative protein abundance, it is apparent that known chromatin-binding proteins and proteins present in the nucleus are most abundant in the enriched proteins (57% and 40% respectively; Figure 2B, right bars), and that unexpected proteins have relatively low abundances (3%). Compared with the 6,009 proteins detected in the full proteome of whole-cell lysates, **SPACE clearly enriches for canonical chromatin proteins, with additional representation of nuclear proteins that are not yet known to bind chromatin**.

Interestingly, RBPs comprise a large proportion of the enriched proteins (Supplementary Fig. 2B). Among the known chromatin-binders (the first category), RBPs comprise 32% of the protein count (207 out of 645) and 72% by relative abundance. Focusing on the "proteins present in nucleus" (newly identified chromatin-binding proteins), we find that RBPs comprise 52% and 68% by counts and abundance respectively. In total, 743 RBPs are found in the SPACE dataset (41% of identified proteins), which comprise 70% of the chromatome weighted by abundance (Supplementary Fig 2C). In other words, our SPACE data reveals 536 new caRBPs in addition to 207 previously characterised caRBPs.

To understand if the RBPs in our dataset are associated with newly transcribed RNA, we compared our results with RICK (*19*) and CARIC results (*20*). Both of these methods work by incorporating Ethynyl Uridine (EU) into the newly synthesized RNA. Then UV-crosslinking is applied to crosslink the RBPs to RNA, and nascent RNAs are captured using click-chemistry. Interestingly, ~67% of the RBPs observed in our SPACE data overlap with RICK and CARIC (Figure 2D). There are 313 additional RBPs in our data that were not associated with nascent RNA using RICK and CARIC. We then ranked the SPACE proteins based on their abundance and compared the accumulation of the RBPs overlapping with RICK and CARIC, other RBPs and known chromatin proteins (Figure 2E). **Interestingly, the RBPs associated with newly synthesized RNA (RICK and CARIC) are accumulated faster than the other two groups, and they are more abundant in chromatin composition.** While this result was surprising, we sought to verify our observation. As a control we compared nuclear RBPs in the full proteome dataset with the RBPs identified by SPACE. We identified 1031 RBPs potentially present in nucleus in the full proteome dataset, while SPACE detected 646 of these RBPs (63% of 1031, Supplementary Figure 2D). Thus, not all nuclear RBPs bind chromatin.

To be conservative, we established an extremely stringent SPACE-SICAP double purification strategy: the initial SPACE purification is followed by SICAP in which DNA is biotinylated with terminal deoxynucleotidyl transferase and captured by protease-resistant streptavidin magnetic beads (Supplementary Figure 1). SPACE-SICAP identified 1,275 enriched proteins, about ~30% less than SPACE alone (Supplementary Figure 2E-H and Table S1_SPACE-SICAP). We considered 990 proteins which are common to SPACE and SPACE-SICAP datasets (Supplementary Figure 2G), and of these, nearly half belong to the proteins present in the nucleus. **Again, we observed a strong enrichment of RBPs among chromatin-associated proteins, as 66% of the proteins 'present in the nucleus' category are RBPs** (Supplementary Figure 2G, the bar). A DNase-treated control confirmed that the identification of

chromatin-associated proteins depends on the presence of DNA: just 138 proteins were found, of which 101 were RBPs (Supplementary Figure 2H). **Altogether this result indicates dual DNA- and RNA-binding functionality in chromatin-associated proteins.**

SPACE shows increased specificity and sensitivity in comparison to other methods

The traditional method for chromatin isolation is subcellular fraction and centrifuge-assisted chromatin sedimentation. A recent method established to obtain global view of chromatin composition is Dm-ChP, which is based on prolonged EdU labelling to pull down DNA using click chemistry. We compared chromatin pelleting(*21*), Dm-ChP(*17*), SPACE and SPACE-SICAP to evaluate their specificity for isolating chromatin proteins from mES cells (Figure 2E). The proteins were categorized to known DNA or chromatin-binding proteins; 2) proteins known to be present in the nucleus but not annotated as DNA- or chromatin-binders; 3) and "unexpected" proteins. We made the comparison based on protein counts and relative abundance of the proteins. Both of these criteria indicate SPACE and SPACE-SICAP have better performance in isolating relevant proteins and removing unexpected proteins.

Limitation of input material is a burden for many chromatin proteomic studies, especially those using primary tissue samples or cell sorting. The above-mentioned SPACE results were obtained using 2-3 million cells and fractionated peptides. While according to the published datasets tens of million cells were used for chromatin pelleting and DmChP, we aimed to assess the sensitivity of SPACE by progressively decreasing the number of input cells to 500,000, 100,000 and finally 20,000. In addition, samples were injected to the mass spec as single shots to reduce the mass spec time to 2h per replicate. Similar numbers of proteins were detected with 500,000 and 100,000 cells (1,446 and 1,452 respectively); with 20,000 cells, we identified a reduced, but still

substantial, number of proteins (Figure 2F). The distribution of enriched proteins between 'known chromatin proteins', 'present in the nucleus' and 'unexpected' categories are very similar among these samples (Supplementary Figure 2I). **Thus, SPACE is both more accurate and sensitive than the other methods for chromatome studies.**

SPACEmap locates the specific chromatin-binding regions of proteins

To understand how proteins are integrated into chromatin, we took an approach similar to RBDmap that identifies peptides crosslinked to RNA (*22*). However, instead of digesting the proteins with LysC or ArgC and then trypsin, we treated them twice with trypsin. Trypsin cleaves at both argininyl and lysinyl residues, so more peptides are digested and released in the first step, allowing us to identify crosslinked sites at higher resolution. Further, we used formaldehyde crosslinking, which is reversible (instead of UV-crosslinking used in RBDmap) which allowed for straightforward mass spec analysis.

First, we digested proteins using large amounts of trypsin, without reversing the crosslinking (Figure 1B). **Most peptides were released from the digested proteins, except those crosslinked to chromatin.** We then carried out another round of SPACE, in order to separate DNA crosslinked to the peptides (crosslinked fraction) from the released peptides (released fraction). We heated the samples to reverse the crosslinking, and to detach the peptides from DNA in the crosslinked fraction. Both fractions were digested again by trypsin and compared with each other to identify the peptides that were significantly enriched in each fraction. Peptides enriched in the crosslinked fraction are either crosslinked directly to DNA, or indirectly via another peptide to DNA (Supplementary Figure 3A). Peptides indirectly crosslinked to DNA remain in the crosslinked fraction if the bridging peptides are long enough to connect DNA to the other peptides. In addition, 2 crosslinking sites are needed to build the bridge. Therefore, we anticipate the chance of enriching indirectly crosslinked peptides to DNA is lower than directly

crosslinked peptides to DNA. In both cases, the peptides enriched in the crosslinked fraction are considered as the contact sites of the proteins with chromatin.

We identified 20,896 peptides, of which 6,158 were enriched in the crosslinked fraction and 6,312 in the released fraction (moderated t-test BH adj. p-value < 0.1 and logFC > 0.4, Figure 3A). 4,363 peptides from 1,270 proteins were captured by the original SPACE method and in the crosslinked fraction of SPACEmap (Figure 3B, Table S2_SPACEmap peptides). Of these, ~90% (3,895 peptides) mapped to a known protein domain or predicted intrinsically disordered region (IDR).

We compared the peptides from Oct4 (Pou5f1), Sox2 and Nanog with annotations of their DNA-binding regions (Figure 3C). The POU-specific domain of Oct4 extends from residues 131-205 (Uniprot coordinates), and the precise DNA-binding residues are at positions 150, 157, 173-179 and 186-189 (Esch et al., 2013). Two peptides corresponding to positions 151-170 and 180-188 containing almost all the DNA-binding residues are enriched in the cross-linked fraction. Seven other peptides from the non-DNA-binding regions of Oct4 were not enriched. **Thus, the Oct4 peptides in the crosslinked fraction accurately match with Oct4's known DNA-binding sites** (Figure 3C, left).

Nanog harbours a Homeobox domain that extends from residues 96 to 155. We identified three enriched peptides corresponding to positions 51-66, 76-87 and 76-89 (Figure 3C, middle)**. All three peptides are located in the IDR adjacent to the homeodomain at the N-terminal region of Nanog (**Supplementary Figure 3B**).** The crystal structure of the Nanog homeodomain suggests protein-DNA interface is located between residues 136-152-Helix H3 (*23*); here, we lack tryptic peptides encompassing this region owing to the large number of lysine and arginine residues. Our result suggests there is a protein-chromatin interface in the IDR close to the homeodomain. Thus, whereas crystal structures provide detailed information about interactions

involving ordered protein regions, SPACEmap complements with insights into chromatin interactions from IDRs which might otherwise be missed.

Finally, Sox2 contains an HMG box domain located at residues 43-111. We identified six Sox2 peptides, two of which were enriched in the crosslinked fraction. **The peptide encompassing residues 83-97 is located within the HMG box, whereas the peptide from residues 274-293 is located in the IDR of Sox2 near the C-terminus of the protein** (Figure 3C, right, and Supplementary Figure 3C). Our result predicts an additional chromatin-interacting element near the C-terminal domain of Sox2 (274-293).

Subsequently, we examined crosslinked fraction at peptide and protein levels to understand how RBPs bind to chromatin. **We found that ~42% of RBPs have at least one crosslinked peptide that maps to IDRs** (Figure 3D, Supplementary Figure 3D, Table S2_peptides mapped to a region). Strikingly, ~56% of 'known chromatin proteins' have at least one crosslinked peptide that maps to IDRs (Figure 3E, Supplementary Figure 3E). A recent study (*24*) has indicated that IDRs interact with DNA using low-affinity interactions with DNA or histones. Initially, IDR-guided weak interactions may allow accelerated recognition of broad DNA regions. Subsequently, DNA-binding domains could stably bind to specific DNA motifs (*24*).

We also observed p-loop domains among the top 5 enriched domains (Figure 3D-E). Although p-loop domains are associated with phosphate-binding such as nucleotide-triphosphates (NTPs), they emerged as avid RNA-binding and ssDNA-binding domains (*25*). As such, **our result confirms p-loop interactions with chromatin in living cells.** In addition, classical RNA or DNA-binding domains such as RRM, helicases, helix-turn-helix and zinc-fingers (C2H2) are highly enriched in the crosslinked fraction.

To further understand the general characteristics of crosslinked fraction peptides, we compared their amino acid composition with the released fraction peptides, as well as the peptides from the

full proteome. Negatively charged residues glutamate and aspartate are depleted in the crosslinked fraction peptides that map to the domains**,** whereas **hydrophobic residues such as leucine, valine, alanine and isoleucine are enriched** (Figure 3F). **The crosslinked fraction peptides that map to IDRs are enriched in glutamate, as well as proline** (Figure 3G), which agrees with the fact that proline and glutamate are the most disorder-promoting residues (*26*). It is surprising that glutamate is depleted from crosslinked peptides mapped to the domains but enriched in those mapped to the IDRs. It is likely that the glutamate residues in the IDRs are involved in protein-protein interactions on chromatin. Alternatively, glutamate residues may destabilize the interactions between the proteins and the target binding sites on DNA in order to accelerate target recognition. Yet, the precise role of glutamate or proline in interactions between IDRs and DNA or chromatin remains to be understood.

During the SPACEmap procedure, the crosslinked protein complexes are broken down, and only peptides remain crosslinked to DNA. As a result, abundant proteins are removed more efficiently, and they are precluded from associating with DNA during the purification procedure. Therefore, SPACEmap is even more stringent than SPACE for identification of chromatin-binding proteins. Interecting SPACE and the crosslinked fraction hits yielded 1270 proteins (Figure 3B). Among them, we found 600 RBPs of which 178 proteins were previously as DNA/chromatin-binders. Thus, SPACEmap provides strong evidence of chromatin-binding for 422 RBPs (Table S2_SPACEmap-verified caRBPs). **Altogether, SPACEmap stringently verifies chromatin-binding proteins and faithfully detects their chromatin interface.**

SPACE elucidates features of mES cells in the ground and metastable states

To demonstrate the quantitative capacity of SPACE, we compared mES cells grown in 2iL (the ground-state) and serum medium (the metastable state) in order to identify caRBPs in different

pluripotency conditions. We identified 1,879 proteins in total (Figure 4A): 100 proteins were significantly more abundant in 2iL and 87 in serum (Log2FC > 1 and adj. p-value < 0.1, moderated t-test adjusted by BH method, Table S3_comparative SPACE). We also compared the SPACE results with the full proteome from the total cell lysate. We found 1738 proteins in the intersection of SPACE and full proteome, and there was a strong correlation in log2 fold-change values between them (Figure 4B and supplementary Figure 4A; R = 0.62). However, **SPACE displays on average ~1.5-fold larger dynamic range of significantly differential proteins than full proteome** (Supplementary Figure 4B). In other words, there are several proteins that are differentially regulated at the level of chromatin-binding, while their expression (total amount) does not change (Figure 4B, the yellow lane). As an example, b-Catenin binds to chromatin in 2iL medium ~ 3-fold higher than serum condition. While, in total b-Catenin is up-regulated ~1.5-fold. Thus, activation of Wnt pathway by inhibiting Gsk3b (CHIR99021) is significantly detectable by SPACE.

To understand how the global network of pluripotency is regulated in 2iL and serum conditions, we looked for proteins with known functions in maintaining embryonic stem cells or exiting from pluripotency. We identified 68 proteins that are positively or negatively involved in the self-renewal of pluripotent stem cells. The network in Figure 4C depicts previously known experimental interactions between a subset of them (Log2FC > 0.6 and adj. p-value < 0.1, moderated t-test adjusted by BH method). Among them are chromatin proteins that physically interact with the core circuitry of pluripotency (Nanog, Oct4, Sox2). Our data suggests that the network of protein interactions surrounding the core pluripotency circuitry shifts substantially between the 2iL and serum conditions. In agreement with previous studies, our results indicate that Tfcp2l1, Prdm14, Cbfa2t2, Zfp42 (Rex1), Klf4, Trim24 and Esrrb (*27, 28*) bind to chromatin preferentially in 2iL conditions, whereas Lin28a and Zfp281 bind more abundantly to chromatin in serum conditions. Our results are in line with the role of Lin28a and Zfp281 in transitioning

from naive to primed state of pluripotency (*29, 30*). Interestingly, differential regulation of Zfp281 is only detectable by SPACE but not full proteome (Figure 4B). Thus, SPACE reveals how the ES cells respond to the cellular conditions more thoroughly than a full proteome analysis. The reason is that SPACE measures both quantity of the proteins, and their binding to chromatin. While a full proteome analysis measures only the quantity of the proteins.

Among the differentially enriched proteins there are 70 RBPs (moderated t-test BH adj. p-value < 0.1 and log2FC > 1, Supplementary Figure 4D). Lin28a is a well-characterised RBP that prevents ES cell differentiation by suppressing let-7 (*31*). Together with Prdm14, they are also known for their roles in DNA-demethylation by recruiting Tet proteins in mouse ES cells; thus, their presence among chromatin-binders was expected (*32, 33*). Our data also indicates Dazl as a caRBPs with highly differential chromatin-binding (log2FC > 2) in 2iL condition. We, therefore, examined Dazl's chromatin-binding by other methods.

Dazl a 3'-UTR-binding protein is recruited to transcription start sites on chromatin

Dazl is best known for targeting the 3' untranslated regions (3' UTRs) of mRNAs to regulate their translation, especially in germ cells (*34, 35*). We first assessed Dazl's cellular localization by immunofluorescent staining using a validated antibody, which confirmed that it is present both in the nuclei and cytoplasm of mES cells (Supplementary Figure 5A). We then performed chromatin immunoprecipitation and sequencing (ChIP-seq) to investigate the genome-wide locations of Dazl binding sites Figure 5A), revealing ~1,300 reproducible peaks. Considering Dazl's known 3' UTR-binding properties, we were surprised to find that 75% of peaks are found within 1kb of transcription start sites (TSS); many target genes are developmental regulators, including Hox genes (Supplementary Figure 5B), several Wnt ligands and Frizzled receptors. As most of the Dazl target genes are involved in development and differentiation of mES cells, we

compared Dazl, Suz12, Aebp2 and H3K27me3 profiles (Figure 5A the heatmap). Interestingly, we observed very similar binding patterns, demonstrating that **Dazl co-localizes with PRC2 on chromatin, especially at the promoters of genes related to the differentiation programs and exiting from pluripotency**.

We also performed individual-nucleotide crosslinking and immunoprecipitation (iCLIP) to identify the RNA-binding sites of Dazl across the transcriptome (*36*). We identified 2,550 peaks in mRNAs, 2099 of which were found in 3' UTRs, and only 166 located within 3,000 nucleotides of the 5' end of mRNAs (Supplementary Table S4_Dazl ChIP, iCLIP and ChIP-SPACE). Thus, the RNA binding sites were positioned at different locations in genes compared to DNA-binding sites, which were located mainly in promoters (Figure 5B). Moreover, most of the genes containing DNA-binding sites of Dazl in their promoter or gene body did not overlap with the genes containing RNA-binding sites of Dazl within their transcripts; only 61 out of 1144 genes (5%) with a gene-proximal ChIP-seq-defined peak on their genes (gene body and 3kb upstream of the TSS) also have an iCLIP-defined peak on their transcript level. These results suggest that the **chromatin- and RNA-binding functions of Dazl are mechanistically independent**.

Next, we examined our SPACEmap data to understand how Dazl binds to chromatin. We observed that out of the 7 peptides that were present in SPACEmap data, only **one peptide was enriched in the crosslinked fraction, corresponding to Dazl's RRM domain** (Figure 5C). RRM domain is known to participate in RNA-binding and DNA-binding; therefore, it remains to be seen whether Dazl binds to chromatin via a bridging RNA, or if it directly binds to the DNA itself. The first option might be plausible, despite the harsh RNase treatment, if RNA is incorporated into a multi-protein Dazl-containing complex that can partly protect it from RNase.

To study Dazl complexes on chromatin, we took a regional approach to identify proteins co-localised on chromatin with Dazl . Here, we developed ChIP-SPACE (Supplementary Figure

5C), a more convenient and faster method than ChIP-SICAP (*18, 37*) because it excludes DNA end-labelling and streptavidin purification and used it to identify Dazl's chromatin partners. Following ChIP, we treated samples with and without RNase A, then purified chromatin fragments by SPACE. 442 proteins were enriched in comparison with the IgG control (moderated t-test BH adj. p-value < 0.1 and log2FC > 1, Figure 5D-E). Sorting the enriched proteins based on their abundance (iBAQ) revealed histones followed by Dazl as the most abundant proteins. In addition, we identified several histones H1, as well as three members of the PRC2 complex: Ezh2, Eed and Suz12. Moreover, we identified pluripotency transcription factors such as Oct4, Klf4, Trim28, Esrrb and 155 SPACEmap-verified caRBPs (~35% out of 442). These findings indicate that **Dazl is part of a conglomerate of caRBPs and transcription factors that are colocalized with PRC2 and the linker Histone H1 in the vicinity of TSSs (Figure 5F)**.

Discussion

Here, we presented SPACE, a robust, sensitive, and accurate method for purifying chromatin-associated proteins by silica magnetic beads for proteomic analysis. **Strikingly, SPACE revealed that >40% of the chromatome are potentially able to interact with RNA**. To identify the specific protein regions that participate in contacts with chromatin, we developed SPACEmap, which showed that ~42% of the potential RBPs bind to chromatin via their IDRs. Similarly, according to RBDmap nearly half of the RNA-binding sites map to the IDRs (*22*). Proteins enriched in IDRs are critical for many chromatin functions such as transcriptional regulation and RNA processing (*38*). IDRs are primary drivers of phase separation of proteins into biomolecular condensates (*12, 39*), which is important in organizing the local chromatin structure (*40, 41*). Also, the activation domains of transcription factors consist of IDRs which

enable transcription factors to phase separate with the Mediator co-activators (*42*). Recently, another study has shown IDRs can generate confinement states for transcription factors to increase the local concentration of transcription factors, and to alter transcriptional output (*43*). **Our findings demonstrate that RBPs directly interact with chromatin components, largely via their IDRs. Probably, RBPs contribute to the condensed or confined chromatin zone formation using their IDRs in order to recruit or trap transcription factors and other chromatin components (Figure 5F).**

We compared the global chromatin composition in 2iL and serum conditions of mES cells, and **we observed Dazl as one of the most differentially expressed caRBPs**, which is highly upregulated on chromatin in the 2iL condition. Dazl has been primarily studied in the context of germ cells due to its substantial roles in controlling the mRNA translation and stability; especially mRNA of the genes that are necessary for germ cell survival (*34, 44*). To identify Dazl binding sites on chromatin we used ChIP-seq, and we found that Dazl associates with the same chromatin sites as PRC2. Thus, in contrast to a recent study that has shown RBPs often interact with enhancers, promoters and transcriptionally active regions on chromatin (*2*), our result indicates Dazl mostly binds to the transcriptionally silenced genes in mES cells (e.g. developmental genes). SPACEmap data reassures Dazl chromatin-binding and reveals Dazl's RRM domain as the chromatin contact site. Our ChIP-SPACE result also indicates >1/3 of the proteins co-localized with Dazl on chromatin are RBPs; providing enormous IDRs to drive condensate formation. In addition, there are 5 Histone H1 in the dataset together with the core nucleosomes. It has been shown that the disordered histone H1 tail forms phase separated condensates and behaves like a liquid glue that clamps condensed clusters of nucleosomes together (*45, 46*). **Thus, our results suggest caRBPs can generate condensed chromatin zones which are transcriptionally silent**.

In addition to Dazl, we found Lire1 as a RBP which binds to chromatin preferentially in serum condition. Lire1 is a nucleicacid-binding chaperone that mobilizes LINE-1 elements in the genome, and its differential regulation in serum condition and primed state pluripotency is intriguing to be studied.

SPACE is broadly applicable due to its superior sensitivity, as only 100,000 cells are enough to enrich >1,400 chromatin associated proteins in a single-shot injection to the mass spec. We believe SPACE will be particularly valuable for quantitative comparisons in timepoint studies, or for analyses of microdissected or sorted cell types. Past studies required much larger amounts of material (*21, 47*), and they suffered from incorporation of modified nucletoides such as EdU or biotin-dUTP into DNA (*17, 48, 49*). Many cell types, such as mES cells, are particularly sensitive to modified nucleotides (*50*). In addition, incorporation of modified nucleotides to tissues such as patient samples is almost impossible. SPACE overcomes all these challenges, while also being more straightforward and sensitive.

All in all, our study demonstrated the capacity of SPACE for quantitative analyses of chromatin composition across conditions, and the capacity of SPACEmap to identify the regions of proteins that contact chromatin. Due to the ease of its application, its high sensitivity and specificity, these methods hold a great potential for further applications that could unravel the dynamics of gene regulation and genome maintenance in development and diseases. Specifically, studying neurodegeneration using SPACE and its variants will shed light on the mechanism of the disease, and reveal novel therapeutic approaches.

Acknowledgement

We would like to thank Jeroen Krijgsveld for revising the manuscript. We appreciate Flora Lee for sharing the iCLIP protocol, and the proteomics platform at the Francis Crick Institute, Bram Snyder, Steve Howel and Helen Flynn for sharing their technical expertise. We would like to thank Yongkai Tan for his help in handling next generation sequencing samples and data.

MR was supported by a postdoc fellowship from EMBO (long-term postdoc fellowship 1217-2017) and a postdoc fellowship from European Commission (Marie Curie Standard Individual postdoc fellowship 752075). This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK(FC010110; 215593/Z/19/Z), the UK Medical Research Council (FC010110), and the Wellcome Trust (FC010110). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. NML is a Winton Group Leader in recognition of the Winton Charitable Foundation's support towards the establishment of the Francis Crick Institute. NML and JU are additionally funded by a Wellcome Trust Joint Investigator Award (103760/Z/14/Z) and core funding from the Okinawa Institute of Science & Technology Graduate University.

Author contribution

MR, JU and NML designed the research. MR, JAZ performed the experiments. MR and SS analysed the data. MR, JU and NML wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests

Figure legends

Figure. 1: Overview of SPACE and SPACEmap.

(A) In SPACE, 1: Cells are crosslinked by 1% formaldehyde, and resuspended in the lysis buffer containing guanidinium, and iso-propanol. Then silica magnetic beads are added to the lysate. 2: Chromatin binds to the magnetic beads and is separated from the lysate. The beads are washed with lysis buffer and ethanol. 4: Chromatin is eluted by sonication and is treated with RNase A. 4: Chromatin is captured again on the beads to be washed again with ethanol and Acetonitrile. Then the crosslinkings are reversed, and trypsin/LysC are added to digest the chromatin-associated proteins on the beads. (B) In SPACEmap, chromatin is recaptured in step 4, however, the crosslinking is not reversed. 5: trypsin is added to digest the chromatin-associated proteins. 6: using another round of SPACE released peptides are separated from crosslinked peptides. Both crosslinked fractions and released fractions are injected to the mass spec to be compared quantitatively.

Figure. 2: Chromatin composition in mES cells reveals RBPs as a major component of chromatome.

(A) Proteins identified by SPACE procedure in comparison to the non-crosslinked control. The blue dots are significantly enriched in comparison to the non-crosslinked control (moderated t-test, BH adjusted p-value < 0.1 and log2FC>1). (B) The enriched proteins by SPACE were categorized based on their biological processes into potential true positive and potential false positive. (C) The enriched proteins were categorized into 3 groups: 1- 'known DNA or chromatin binding proteins' (dark green), 2- 'Other proteins present in Nucleus' (pale green), 3- Proteins that do not fall into the previous categories are so-called 'unexpected' (yellow). The left two bars compare protein counts between the full proteome of mES cells and SPACE. The right two bars compare the relative abundance of the proteins. The full proteome data was obtained from (*18*), and re-analysed. (D) Comparing proteins enriched by SPACE with RICK(*19*) and CARIC(*20*) datasets. The latter two datasets represent RBPs that interact with newly transcribed RNA. The bar show 1163 non-overlapping proteins in SPACE dataset and the RBPs among them. The RBPs are shown by purple. (E) Enriched proteins by SPACE were ranked by their abundance. The rates of accumulation in the dataset were compared among 1- the RBPs that overlap with RICK and CARIC (continuous purple line), 2- Other RBPs (dashed purple line), and 3- Known DNA/chromatin-binders (continuous green line). (F) Chromatin pelleting (*21*), DmChP (*17*), SPACE and SPACE-SICAP results of mES cells were compared based on the protein counts and relative abundance of the enriched proteins. The enriched proteins were categorized into 3 groups, as mentioned previously in panel B. (G) Titration of the input cell number for SPACE procedure. The blue stacks show the number of enriched proteins relative to the non-crosslinked proteins (log2FC >1 and moderated t-test BH adj. p-value < 0.1). The grey stacks show identified proteins that are not enriched relative to the non-crosslinked proteins.

Figure. 3: Locating chromatin-binding sites of the proteins. (A) The volcano-plot shows peptides enriched (moderated t-test BH adj. p-value < 0.1 and log2FC > 0.4) in the crosslinked fraction (red) and in the released fraction (dark blue). (B) The overlap of the proteins identified by crosslinked fraction (red) and SPACE (light blue) is shown in yellow. The bar shows the number of peptides corresponding to the overlapping proteins and the proportion of the peptides that are mapped to any regions (domains or intrinsically disordered regions). (C) The plots show crosslinked and released peptides in Oct4, Sox2 and Nanog. The peptides significantly enriched in the crosslinked and released fractions are red and blue, respectively. Non-significantly enriched peptides are grey. The pink bars indicate the aminoacid positions of the DNA-binding domains. The green bars denote IDRs. The red boxes show the enriched peptides in the crosslinked fraction. (D) Top 5 domains/regions by the proportion of RBPs that contain them. RBPs containing these domains/regions have peptides enriched in the crosslinked fraction and overlapping with these domains/regions or residing no farther than 10 amino acids from them. (E) Top 5 domains/regions by the proportion of known DNA/chromatin proteins that contain them. The proteins containing these domains/regions have peptides enriched in the crosslinked fraction and overlapping with these domains/regions or residing no farther than 10 amino acids from them. (F) Aminoacid composition of the peptides mapped to domains in the crosslinked fraction relative to the peptides mapped to domains in the released fraction (left) and full proteome (right). These peptides overlap with domains or reside no farther than 10 amino acids from them. (G) Aminoacid composition of the peptides mapped to IDRs in the crosslinked fraction relative to the peptides mapped to domains in the released fraction (left) and full proteome (right). These peptides overlap with domains or reside no farther than 10 amino acids from them.

Figure. 4: Chromatin composition in 2iL and serum conditions of mES cells.

(A) The volcano-plot shows proteins that are significantly more abundant in 2iL and serum by red and blue, respectively (adj. p-value <0.1 and log2FC >1). The rest of the proteins were depicted by grey. Proteins involved in pluripotency, mES cell self-renewal or differentiation were marked by black dots. (B) Comparing full proteome analysis with SPACE. The yellow lane indicates differentially regulated proteins detectable only by SPACE. The full proteome data was obtained from (*18*), and re-analysed. (C) Experimental interaction network of the proteins involved in pluripotency, mES cell self-renewal or differentiation. RBPs were marked by purple borders.

Figure 5: Dazl is recruited to silenced transcription start sites in mES cells. (A) Annotation of Dazl ChIP-seq peaks, and the profile of Dazl peaks on the genome in comparison with Suz12, Aebp2 and H3K27me3 peaks in mES cells. The last ChIP profiles of Suz12, Aebp2 and H3K27me3 were obtained from (*51*). (B) Annotation of Dazl iCLIP peaks (top bar), and the intersect of Dazl ChIP-seq and iCLIP-seq peaks at the gene level (bottom Venn diagram). (C) Dazl peptides identified using SPACEmap procedure are shown by red (enriched in crosslinked fraction), blue (enriched in released fraction) and grey (statistically non-significant). (D) Proteins enriched by Dazl ChIP-SPACE in comparison to the IgG control were sorted by the abundance of the proteins (iBAQ). Histones and PRC2 components are shown by yellow and orange dots, respectively. Dazl and transcription factors are shown by red dots. (E) The volcano plot shows proteins identified by ChIP-SPACE and their sensitivity to RNase A treatment. Proteins that are affected by RNase treatment are named in the plot. (F) The schematic model of RBP interactions with chromatin based on Dazl data. Chromatin-associated RBPs form a condensed or confined zone probably via interactions among their IDRs with other components of chromatins. The RBPs that are in the periphery of the zone are sensitive to RNase treatment. The RBPs in the centre of the zone are resistant to RNase treatment. Transcription factors and other components of chromatin are probably recruited or trapped by the RBPs.

Material & Methods

Mass spectrometry and proteomics data analysis

The details of sample preparation using SPACE, SPACE-SICAP and ChIP-SPACE procedures are provided at the end. Following sample preparation, peptides were separated on a 50 cm, 75 µm I.D. Pepmap column over a 120 min gradient for SPACE and SPACE-SICAP, or a 70min gradient for ChIP-SPACE. Peptides were then injected into the mass spectrometer (Orbitrap Fusion Lumos) running with a universal Thermo Scientific HCD-IT method. Xcalibur software was used to control the data acquisition. The instrument was run in data-dependent acquisition mode with the most abundant peptides selected for MS/MS by HCD fragmentation. RAW data were processed with MaxQuant (1.6.2.6) using default settings. MSMS spectra were searched against the Uniprot (Swissprot) database (Mus musculus and Homo sapiens) and database of contaminants. trypsin/P and LysC were chosen as enzyme specificity, allowing a maximum of two missed cleavages. Cysteine carbamidomethylation was chosen as the fixed modification, and methionine oxidation and protein N-terminal acetylation were used as variable modifications. Global false discovery rate for both protein and peptides was set to 1%. The match-from-and-to and re-quantify options were enabled, and Intensity-based quantification options (iBAQ) were calculated.

Quantitative proteomics, statistical and computational analysis

The protein groups were processed in RStudio using R version 4.0.0. The proteins only identified by site, Reverse, potential contaminants were filtered out. Protein Gene Ontology and other information were downloaded from Uniprot and DAVID Gene Ontology database. For the SPACE experiments (Fig2A,F), the crosslinked samples were compared with non-crosslinked samples by SILAC ratios calculated using MaxQuant. Moderated t-test p-values and BH adj. p-values were calculated by limma package. Log2(crosslinked/non-crosslinked) >1 and adj. p-value <0.1 were considered as significantly enriched proteins. According to the GO information, the enriched proteins were categorized to known DNA/chromatin-binders and proteins that are “present in the nucleus” (but not DNA/chromatin-binders). The rest of the proteins were considered as “unexpected”. For the SPACE-SICAP experiment (Fig2C), the crosslinked samples were compared with the non-crosslinked samples by SILAC iBAQ values. The crosslinked samples and non-crosslinked samples were normalized separately using quantile-normalization from preprocessCore package. Moderated t-test p-values and BH adj. p-values were calculated by limma package. Log2(crosslinked/non-crosslinked) >1 and adj. p-value <0.1 were considered as significantly enriched proteins. For the SPACEmap experiment (Fig3), the crosslinked fraction was compared with the released fraction by label-free iBAQ values. The samples were normalized using quantile-normalization from preprocessCore package. Moderated t-test p-values and BH adj. p-values were calculated by limma package. Log2(crosslinked/non-crosslinked) > 0.4 and adj. p-value <0.1 were considered as differentially enriched proteins. For the comparative SPACE experiment and full proteome analysis (Fig4), the 2iL samples were compared with serum samples by SILAC ratios calculated using MaxQuant. Moderated t-test and BH adj. p-values were calculated by limma package. Log2(2iL/serum) >1 and adj. p-value <0.1

were considered as significantly enriched proteins. Interaction network determined only by experiments was downloaded from String database, and visualized by Cytoscape 3.8. For the Dazl ChIP-SPACE experiment, the RNase-treated and non-treated samples were compared by label-free iBAQ values. Moderated t-test p-values and BH adj. p-values were calculated by limma package. Log2(crosslinked/non-crosslinked) > 1 and adj. p-value <0.1 were considered as differentially enriched proteins.

Dazl ChIP-seq experiment and data analysis

The ChIP procedure and analysis were carried out essentially as described previously(Rafiee et al., 2016). Briefly, mESCs were grown in 2iL medium. The cells were detached, and fixed by 1.5% formaldehyde in PBS for 15min. Chromatin was solubilized by sonication, and sheared to < 500 bp fragments. Dazl immunoprecipitation was carried out using CST antibody #8042 overnight at 4 °C. Following washing steps, chromatin was eluted, and DNA was purified by SPRI beads. Library was prepared for the Illumina platform as described previously. Sequencing was carried out using 75nt reads on paired-end mode by HiSeq4000. Reads were trimmed, aligned to the mouse genome (mm10) using Bowtie2, and duplicated reads were removed with samtools. Peak calling was performed using macs2. Narrow peaks called by macs2 were extended by 250bp around their middle (to a total width of 500bp) Dazl peaks annotation into genomic features was done using ChIPseeker R package with 3kb around TSS set for promoter region window. The ChIP-seq profiles of Dazl, Suz12, Aebp2 and H3K27me3 were compared by deeptools 2.

Dazl iCLIP and data analysis

iCLIP was carried out as previously described (Blazquez et al., 2018). Briefly, mESCs were grown in 2iL medium. Cells were UV cross linked, lysed and IP performed using 1:70 DAZL antibody (CST #8042) in IP. RNaseI was used at 0.4U/mg cell lysate per IP. Finalised libraries were sequenced as single end 100bp reads on Illumina HiSeq 4000. Processing of DAZL iCLIP raw data was carried out using iMaps (https://imaps.genialis.com/). The demultiplexed and quality controlled data was mapped to mm10 genome assembly using STAR (2.6.0) with default settings. Both PCR duplicates and reads that did not map uniquely to the genome were discarded.

Cell culture

The 46C mESC cells were cultured using either 2i + LIF (2iL) medium or standard mESC serum medium + LIF. The 2iL medium consists of DMEM:F12 for SILAC, Glutamax, N2 supplement, non-essential amino acids, B27 supplement, β-mercaptoethanol (all from Gibco), CHIR99021 3uM (Sigma-Aldrich), PD 0325901 1uM (Sigma-Aldrich) and LIF 100 ng/ul (proteintech). To label the cells with heavy amino acids, $^{13}C_6$ $^{15}N_4$ L-Arginine and $^{13}C_6$ $^{15}N_2$ L-Lysine were added to the 2iL medium. To label the cells with light amino acids, $^{12}C_6$ $^{14}N_4$ L-Arginine and $^{12}C_6$ $^{14}N_2$ L-Lysine were added to the medium.

Domain analysis

We obtained protein sequences from UniProt (*52*) using UniProt IDs and scanned the sequences for domains and intrinsically disordered regions (IDRs) with InterProScan v5.47-82.0 (*53*). From the scanning results, we selected matches with the status "True" and either a non-empty InterPro (*54*) accession number (for domains) or MobiDB-lite (*55*, *56*) accession number present (for IDRs). Next, we excluded signatures that do not represent domains or IDRs (ProSitePatterns, PRINTS, PIRSF, PANTHER) and calculated consensus coordinates for all matches (both domains and IDRs) as follows. For each InterPro entry or group of IDRs in each protein, we checked if any pair of its signatures overlapped by at least 70% of length of either of the signatures, and, if such signatures were found, we merged them. We continued this process until no merged regions or signatures met the overlapping condition. Then, we defined consensus coordinates of matches as the start and stop coordinates of each of the final merged regions. This allowed us to delineate possible separate domain matches within the same InterPro entry and also to merge highly overlapping predicted IDRs. Next, we postulated that an InterPro entry or an IDR matches a peptide if it overlaps with the peptide or resides no farther than 10 amino acids either side of the peptide. Using this definition and the consensus coordinates of matches, we selected InterPro entries and IDRs that matched peptides in the crosslinked fraction and clustered the entries to obtain more general domain types (all IDRs were grouped in the same cluster). We did the clustering as follows. If matches of two InterPro entries overlapped at least ones, and in ≥70% of their overlaps the overlap size was ≥70% of their lengths (reciprocally), then we put the two entries in the same cluster. If the entries were already allocated to different clusters, then we merged those clusters. Informally, this procedure allowed us to group InterPro entries that match similar sites and hence describe similar domains. We named the top matched clusters based on the InterPro entries that they contained.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023903. The accession numbers for the Dazl ChIP-seq and iCLIP reported in this paper are ArrayExpress: E-MTAB-9302 and E-MTAB-9332, respectively.

Step by step SPACE procedure

Required material:

- Formaldehyde (Methanol-free, Pierce 28906, or 28908)
- BCA protein assay kit (Thermo, 23225)
- Benzonase (Sigma, E8263)
- DNA Binding Beads (Thermo 4489112)
- Guanidinium thiocyanate (Sigma, G9277-100G)
- N-Lauroylsarcosine sodium salt (Sigma, L9150-100G)
- Tris HCl (Sigma, T2319-1L)
- EDTA (Sigma, 324504-500ML)
- 2-Propanol (Sigma, 278475-1L)
- Ammonium Bicarbonate (Sigma, 09830-500G)
- DTT (D9779-5G)
- Iodoacetamide (Sigma I1149-5G)
- RNase A (Thermo EN0531)
- Acetonitrile (271004-1L)
- Trypsin (Promega, V5280)
- LysC (Wako, 121-05063)
- Trifluroacetic acid (Sigma 302031)
- Formic acid (Sigma 5.43804)
- ZipTip with 0.6 µL C_{18} resin, (Merck, ZTC18S096)

Reagents:

- PBS-T(PBS + Tween 0.1%)
- PBS-SDS (PBS + SDS 1%)
- Elution buffer: Tris HCl 10mM pH 7.5

· Lysis buffer: Guanidinium thiocyanate 4M, Tris HCl 100mM, Sarkosyl 2%, EDTA 10mM

· Wash buffer 1: Lysis buffer + 2-propanol (1:1 v/v)

· Wash buffer 2: Ethanol 80% (v/v)

· AMBIC buffer: Ammonium Bicarbonate 50mM, DTT 10mM (prepare freshly)

· DTT solution: DTT 1M (keep the aliquots in -20 °C, thaw once)

· IAA solution: IAA 0.4M (keep the aliquots in -20 °C, thaw once)

Equipments:

· Diagenode Bioruptor Pico

· James Products Ultrasonic 7000S (ultrasonic cleaner)

· Magnetic stand for 15ml tubes

· Magnetic stand for 2ml tubes

· Magnetic stand for PCR tubes

Experimental procedure:

A) Formaldehyde crosslinking:

1. formaldehyde crosslinking can be carried out in the medium of the cells, by adding 16% formaldehyde in the medium of the cells to make 1% final concentration. You may add formaldehyde directly to the medium of the cells in their plates. Obviously, as the non-crosslinked control you don't need to add formaldehyde to the medium.

Note: formaldehyde is toxic, and you should work in a fume hood.

2. Keep the plates 10min in the fume hood.

3. Discard the medium of the cells, as appropriate. Wash the cells two times with PBS.

4. Pour PBS-T in the plates, half of the volume of the medium.

5. Lift the cells by a cell lifter.

6. Collect the cell suspension, and transfer them to a 15-ml/50-ml tube.

7. Repeat steps 4-6 once again to collect most of the cells.

8. Spin the tubes in 400g (2000 RPM in an Eppendorf/Thermo centrifuge) for 2min

9. Discard the supernatant, and freeze the cells in -80 °C

Note: the cells are stable in -80 °C for months.

B) Chromatin purification.

1. Thaw the cells, and resuspend them in PBS-T. The volume depends on the number of the cells. For example you may resuspend the cells in 10% of the original volume of the medium. (i.e. 10 cm dish: 1ml, 6w plate: 0.2ml). The idea is to measure protein concentrations, to start the experiment from equal amounts of input cells.

   Note: Don't use more 3 million cells as the starting material.

2. Take 10ul of the samples, and increase the volumes to 30ul by PBS-SDS. Heat the samples at 95 °C for 5 min. After heating the samples, they will become viscous due to the release of DNA. Add 0.5ul Benzonase, and wait a few minutes before proceeding to the next steps. Vortex the samples gently, and measure the protein concentrations by BCA assay. The BSA standards should be diluted in PBS-SDS, and should be heated like the unknown samples.

3. Pour equal amounts of samples in 15-ml tubes. Add 10ul RNase A. Vortex gently, and wait 5 min.

4. Add 2ml Lysis buffer, and vortex the tubes vigorously.

5. Add 2ml 2-propanol, and vortex the tubes vigorously.

6. You may pool SILAC labelled samples. For example, you may pool heavy SILAC crosslinked samples with light SILAC non-crosslinked samples at this step.

Note: if you see white precipitations spin the tube briefly to separate them. Normally it happens in non-crosslinked controls. Do not carry the precipitations to the pool.

7. Add 50ul to 100ul DNA-binding beads to the mixture. It's not necessary to wash or dilute the beads.

8. Vortex the tubes, and wait 10min.

9. Put the tubes on the magnet, and wait 10min to separate the beads.

10. Gently remove the supernatant.

Note: the supernatant is probably brownish even after 10min. It's OK if you lose some beads.

11. Resuspend the beads in 1ml of Wash buffer 1, and transfer them to 2-ml tubes.

12. Vortex the tubes, spin briefly, and put them on the magnet.

13. Wait 2min to separate the beads on the magnet. Discard the supernatant.

14. Remove the beads from the magnet, add 1ml Wash buffer 2.

15. Vortex the tubes, spin briefly, and put them on the magnet.

16. Wait 2min to separate the beads on the magnet. Discard the supernatant.

17. Vortex the tubes, spin briefly, and put them on the magnet.

18. Discard the residual Wash buffer 2.

19. Add 200ul of elution buffer. Do not pipette as the beads are very sticky at this point.

20. Put the tube in the ultrasonic cleaner, and turn on the sonicate them for 5min.

21. Transfer the tubes to the rotor of Diagenode Bioruptor.

22. Sonicate the tubes 3 cycles: 30 sec ON, 30 sec OFF

    Note: the beads should be resuspended at this step.

23. Add 10ul of RNase A (10 mg/ml) to the samples, and agitate them (1000 RPM) in a thermomixer at 37 °C for 15 min

24. Add 250ul of Lysis buffer, and vortex the tubes.

25. Add 300ul of 2-Propanol, and vortex the tubes.

26. Wait 10min.

27. Put the tubes on the magnet, and wait 10min to separate the beads.

28. Gently remove the supernatant.

29. Remove the beads from the magnet, add 1ml Wash buffer 2.

30. Vortex the tubes, spin briefly, and put them on the magnet.

31. Repeat the last 2 steps once again.

32. Remove the beads from the magnet, add 0.5ml Acetonitrile 100%

33. Vortex the tubes, spin briefly, and put them on the magnet.

34. Resuspend the beads in 80ul of Acetonitrile 100%, and transfer them to PCR tubes.

Note: you may need to cut the head of the 200-ul tips to be able to transfer the beads.

You may also need to repeat this step once again to transfer all the beads to a PCR tube.

35. Put the cells on a magnet, and discard the supernatant.

36. Add 18ul of AMBIC buffer (if you wish to use TMT labelling use TEAB buffer)

37. Heat the tubes in a PCR machine at 95 °C for 10min. Turn on the lid (105 °C) to avoid evaporation.

38. Chill the tubes on ice. Add 2ul of the IAA solution. Vortex gently, and keep in a drawer for 15min.

39. Add 1ul of the DTT solution, and vortex gently.

40. Add 400ng trypsin, 100ng LysC, and incubate the tubes 12-14 h at 37 °C. Turn on the lid (80 °C) to avoid evaporation.

41. Add 100ng trypsin, vortex and continue the digestion for another 2-h.

42. Remove the beads on the magnet, and transfer the supernatant to a new PCR tube

43. Clean-up the peptides using stage-tips or ZipTips.

Step by step SPACE-SICAP procedure

Additional required material:

· Biotin-ddUTP (Jenabioscience, NU-1619-BIOX-S or L)

· Biotin-dCTP (Jenabioscience, NU-809-BIOX-S)

· Streptavidin magnetic beads (NEB S1420S)

· Sodium cyanoborohydride (Sigma, 8180530025)

· TdT (Thermo Scientific, EP0162)

· T4 PNK (NEB, M0201S)

· Amicon Ultra-0.5 Centrifugal Filter Unit, 30 KDa (Millipore, UFC503096)

· SPRIselect (Beckman Coulter, B23317)

Reagent preparation:

· Protease-resistant streptavidin beads: Beads should be treated in fume hood. For 5ml beads, prepare 5ml Sodium cyanoborohydride 0.2M (Reagent A) and 5ml Formaldehyde 4% (Reagent B). Pour the beads in a 15-ml tube, put the tube on the magnet, remove the beads, and discard the supernatant. Wash the beads with 5ml PBS-T, put the tube on the magnet, remove the beads, and discard the supernatant. Add reagent A and reagent B to the beads. Resuspend the beads, and keep them in the hood for 2 hours with occasional mixing. Put the tube on the magnet, remove the beads, and discard the supernatant appropriately. Wash the beads with 5ml Tris HCl 0.1 M pH 7.5, twice. Finally, resuspend the beads in 5ml PBS-T, and keep them in the fridge. The beads are now resistant to Lys-C digestion, and they are stable for months. For more details please refer to(Rafiee et al., 2020)

Experimental procedure:

A) DNA labelling and chromatin purification

1. After RNase A treatment in SPACE procedure (Step 22, section B), remove the beads on the magnet. Transfer the supernatant to Amicon ultrafilteration tubes, and spin them in 12000g for 7 min at 8 °C. The volume of the liquid in the column should be < 100ul, otherwise, continue the centrifuge for a few more minutes.

2. Collect the liquid in the column, and transfer them to PCR tubes.

3. Add the following reagents:

    I. TdT buffer 20ul

    II. Biotin-ddUTP 5ul

    III. Biotin-dCTP 5ul

    IV. TdT 3ul

    V. T4 PNK 1ul

    VI. H2O fill up to 100ul

4. Vortex, and spin briefly

5. Incubate the tubes at 37 °C for 30min

6. Add 100ul SPRIselect beads, vortex, and spin briefly

7. Wait 10min

8. Separate the beads on the magnet

9. Wash the beads with ethanol 80% without disturbing the beads

10. Repeat the last step once again

11. Remove the residual ethanol

12. Resuspend the beads in 100ul elution buffer

13. Sonicate the tubes in the ultrasonic cleaner for 5min

14. Remove the beads on the magnet, and transfer the supernatant to 2ml tubes

15. Increase the volume to 1ml with PBS-SDS

16. Add 75ul protease-resistant streptavidin beads.

17. Rotate 45min at room temperature

18. Separate the beads on the magnet, and discard the supernatant

19. Wash the beads two times with PBS-SDS.

Note: during this step you will observe the beads are dispersed on the magnetic stand

20. Wash the beads with 2-propanol 20%

21. Wash the beads two times with Acetonitrile 40%

22. Separate the beads on the magnet, and discard the supernatant

23. Resuspend the beads in 80ul of Acetonitrile 40%, and transfer them to PCR tubes.

Note: You may also need to repeat this step once again to transfer all the beads to a PCR tube.

24. Put the cells on a magnet, and discard the supernatant.

25. Add 18ul of AMBIC buffer (if you wish to use TMT labelling use TEAB buffer)

26. Heat the tubes in a PCR machine at 50 °C for 15min. Turn on the lid (80 °C) to avoid evaporation.

27. Chill the tubes on ice. Add 2ul of the IAA solution. Vortex gently, and keep in a drawer for 15min.

28. Add 1ul of the DTT solution, and vortex gently.

29. Add 400ng LysC, and incubate the tubes 12-14 h at 37 °C. Turn on the lid (80 °C) to avoid evaporation.

30. Separate the beads on the magnet. Transfer the supernatant to new PCR tubes.

Note: this is a mixture of DNA- and chromatin-binders. Add 300ng trypsin, and continue the digestion for 4-hours.

31. Wash the beads with Acetonitrile 40%

32. Separate the beads on the magnet. Discard the supernatant.

33. Add 18ul of AMBIC buffer

34. Heat the tubes in a PCR machine at 95 °C for 5min. Turn on the lid (105 °C) to avoid evaporation.

35. Separate the beads on the magnet. Transfer the supernatant to new PCR tubes.

36. Add 100ng trypsin, vortex and continue the digestion for another 4-h.

Note: Peptides identified in this sample belong to the proteins directly crosslinked to DNA.

37. Clean-up the peptides using stage-tips or ZipTips.

Step by step SPACEmap

After step 36 of SPACE (Pour 18ul AMBIC on the beads)

1. Heat the tubes in a PCR machine at 50 °C for 10min. Turn on the lid (80 °C) to avoid evaporation.
2. Chill the tubes on ice. Add 2ul of the IAA solution. Vortex gently, and keep in a drawer for 15min.
3. Add 0.5ul of 1M DTT solution, and vortex gently.
4. Add 1000ng Trypsin, 200ng LysC, and incubate the tubes 12-16 h at 37 °C. Turn on the lid (80 °C) to avoid evaporation.
5. Add 500ng Trypsin, and incubate the tubes 4 h at 37 °C.
6. Add 80ul SPACE Lysis buffer, and vortex
7. Add 80ul 2-propanol 100%, and vortex
8. Wait 10min, and spin the tubes briefly
9. Separate the beads on magnet
10. Collect the supernatant, and transfer it to a 1.5ml tube. This is the released fraction to be compared with the crosslinked fraction.
11. Wash the DNA-binding beads with Wash buffer 1 once, and discard it
12. Wash the DNA-binding beads with Wash buffer 2 once, and discard it
13. Wash the DNA-binding beads with acetonitrile twice, and discard it
14. Briefly spin the beads, and discard the residues of acetonitrile in the bottom of the tube
15. Resuspend the beads in 20ul of AMBIC buffer
16. Heat the tubes in a PCR machine at 95 °C for 10min. Turn on the lid (105 °C) to avoid evaporation.
17. Chill the samples, and add 300ng trypsin, 50ng lysC, and incubate the tubes 12 h at 37 °C. Turn on the lid (80 °C) to avoid evaporation.
18. Remove the beads on the magnet, and transfer the supernatant to a new PCR tube. This is the Crosslinked fraction, and should be compared with the released fraction.
19. Clean-up the peptides using stage-tips or ZipTips.

Step by step ChIP-SPACE

After the ChIP protocol, and regular washing steps perform the following steps:

1. Separate the beads on the magnet, and discard the last wash buffer
2. Resuspend the beads in 100ul TE buffer (or Elution buffer), add 5ul RNase A (10mg/ml), and agitate at 750 RPM in a ThermoMixer at 37 °C for 10min.

   Note: You may skip this step if you wish to have a RNase negative control.
3. Resuspend the beads in 300ul Lysis buffer, vortex vigorously, and incubate at 37 °C for 2min in a Thermomixer with 1000 RPM agitation.
4. Remove the beads on the magnet, and transfer the supernatant to a new 2-ml tube
5. Add 300ul 2-propanol, and vortex
6. Add 30ul of DNA-binding beads, vortex, and spin
7. Wait 10min
8. Separate the beads on the magnet, and discard the supernatant
9. Wash the beads with 500ul Wash buffer 1, separate the beads on the magnet, and discard the supernatant.
10. Wash the beads with 500ul Wash buffer 2, separate the beads on the magnet, and discard the supernatant.
11. Repeat the last step once again.
12. Wash the beads with 500ul Acetonitrile 100%, separate the beads on the magnet, and discard the supernatant.
13. Resuspend the beads in 80ul of Acetonitrile 40%, and transfer them to PCR tubes.

    Note: You may also need to repeat this step once again to transfer all the beads to a PCR tube.
14. Put the cells on a magnet, and discard the supernatant.
15. Add 18ul of AMBIC buffer (if you wish to use TMT labelling use TEAB buffer)
16. Heat the tubes in a PCR machine at 95 °C for 10min. Turn on the lid (105 °C) to avoid evaporation.
17. Chill the tubes on ice. Add 2ul of the IAA solution. Vortex gently, and keep in a drawer for 15min.

18. Add 1ul of the DTT solution, and vortex gently.
19. Add 250ng trypsin, 100ng LysC, and incubate the tubes 14-16 h at 37 °C. Turn on the lid (80 °C) to avoid evaporation.
20. Remove the beads on the magnet, and transfer the supernatant to a new PCR tube
21. Clean-up the peptides using stage-tips or ZipTips.

Supplementary Figure legends

Supplementary Figure. 1: Overview of the SPACE-SICAP procedure.

Following step3, chromatin is eluted by sonication, and treated with RNase A. Then silica magnetic beads are removed, and DNA is end-labelled by TdT and biotin-ddUTP or biotin-dCTP. Step 4: chromatin is re-captured on protease-resistant streptavidin beads. Step 5: chromatin-associated proteins are digested to be identified by mass spec.

Supplementary Figure. 2: Characterization of proteins enriched by SPACE and SPACE-SICAP.

(A) The enriched proteins by SPACE were categorized based on their biological processes. The node colours indicate if the proteins are 'known DNA/chromatin binding proteins' (dark green), 'Present in Nucleus' (pale green) or 'unexpected' (yellow) in chromatin composition. The proteins that are involved in chromatin remodelling, chromatin modification, DNA replication & repair, transcription factor activity, transcription, cell cycle, splicing, RNA processing and RNA transport were considered as 'potential true positive'. The proteins that are involved in translation, metabolic process, cell adhesion, protein transport and protein folding are considered as potential false positive. The stacked bars compare the abundance of 'potential true positive' with 'potential false positive'. (B) The proportion of RBPs in DNA/chromatin-binding proteins (left bars) and unexpected proteins (right bars) were shown as count and relative abundance. (C) The frequency of RBPs in the entire proteins enriched by SPACE were shown as counts and relative abundance. (D) (E) The volcano-plot shows proteins enriched via SPACE-SICAP. The blue dots are significantly enriched in comparison to the non-crosslinked control (moderated t-test, BH adjusted p-value < 0.1 and log2FC>1). (F) Enriched proteins by SPACE-SICAP were

categorized into the three abovementioned groups: known DNA/chromatin binding proteins' (dark green), 'Present in Nucleus' (pale green) or 'unexpected' (yellow) in chromatin composition. (G) The Venn diagram shows the intersection of SPACE and SPACE-SICAP. The bar shows the proportion of RBPs in the PN group. (H) The Venn diagram indicates the number of identified proteins by SPACE-SICAP in comparison to the DNase-treated sample. (I) Titration of the input cell number for SPACE procedure. (J) The proportion of DNA-binding domains (DBD) and chromatin-associated domains (CAD) in DNA/chromatin-binding proteins (left bars), Present in Nucleus (middle bars) and unexpected proteins (right bars) were shown as counts and relative abundance.

Supplementary Figure. 3: Chromatin binding sites of the proteins revealed by SPACEmap. (A) Product of SPACEmap after the first tryptic digestion. Most of the proteins are digested to short peptides by trypsin. Only peptides directly crosslinked to DNA (red) or indirectly via another crosslinked peptide (orange) remain after the digestion. (B-C) Prediction of intrinsically disordered regions in Nanog and Sox2 using IUPred2 (*57*). The red boxes show peptides enriched in the crosslinked fraction. The pink boxes show DNA-binding domains. (D) Top 10 domains/regions by the proportion of RBPs that contain them (left) or by the proportion of crosslinked fraction peptides of RBP origin that contain them (right). Crosslinked fraction peptides containing these domains/regions overlap with them or reside no farther than 10 amino acids from them. (E) Top 10 domains/regions by the proportion of DNA/chromatin-binders that contain them (left) or by the proportion of crosslinked fraction peptides of DNA/chromatin-binders that contain them (right). Crosslinked fraction peptides containing these domains/regions overlap with them or reside no farther than 10 amino acids from them. (F) Aminoacid composition of the peptides mapped to domains (left) or IDRs (right) in the crosslinked fraction and full proteome. These peptides overlap with domains or IDRs or reside no farther than 10 amino acids from them.

Supplementary Figure. 4: Comparing 2iL and serum states of mES cells using SPACE.

(A) Comparing proteins identified by SPACE and full proteome analysis of mouse ES cells in 2iL and serum conditions. (B) The differentially regulated proteins in the intersection of the full proteome and SPACE were compared. The boxes indicate the inter-quantile regions (IQR). The line in the box is median. The whiskers show 1.5 IQR. The spots are outliers. (C) Comparing RBPs identified by mES cell interactome capture (*58*), and SPACE. (D) Comparative SPACE between mES cells in 2iL and serum condition. Purple points show RBPs with differential binding ratios to chromatin in 2i and serum conditions.

Supplementary Figure 5: Validation of Dazl as a caRBP in mES cells.

(A) Confocal fluorescence microscopy imaging of fixed mouse embryonic stem cells. Staining was carried out using DAPI, Suz12 and Dazl antibodies. Fluorophore exclusion from subnuclear structures (putative nucleoli) was observed in Dazl staining corroborating its nuclear presence (arrow shows one representative example). (B) ChIP-seq peaks indicate Dazl binds to the promoter of several Hox genes. (C) ChIP-SPACE procedure: after the ChIP procedure (1), samples are treated with and without RNase A (2), then SPACE beads are added to purify chromatin (3), following mass spectrometry (4) proteins co-localized with the target protein are identified. In addition, proteins sensitive to RNase treatment are determined.

Reference

1. D. L. Bentley, Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**, 163-175 (2014).
2. R. Xiao *et al.*, Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell* **178**, 107-121 e118 (2019).
3. K. Oktaba *et al.*, ELAV links paused Pol II to alternative polyadenylation in the Drosophila nervous system. *Mol Cell* **57**, 341-348 (2015).
4. L. Herzel, D. S. M. Ottoz, T. Alpert, K. M. Neugebauer, Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**, 637-650 (2017).
5. A. Bregman *et al.*, Promoter elements regulate cytoplasmic mRNA decay. *Cell* **147**, 1473-1483 (2011).
6. T. Trcek, D. R. Larson, A. Moldon, C. C. Query, R. H. Singer, Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell* **147**, 1484-1497 (2011).
7. B. M. Zid, E. K. O'Shea, Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. *Nature* **514**, 117-121 (2014).
8. J. E. Henninger *et al.*, RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225 e224 (2021).
9. K. Shrinivas *et al.*, Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol Cell* **75**, 549-561 e547 (2019).
10. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
11. S. Shukla, R. Parker, Hypo- and Hyper-Assembly Diseases of RNA-Protein Complexes. *Trends Mol Med* **22**, 615-628 (2016).
12. D. Tauber, G. Tauber, R. Parker, Mechanisms and Regulation of RNA Condensation in RNP Granule Formation. *Trends Biochem Sci* **45**, 764-778 (2020).
13. F. Gebauer, T. Schwarzl, J. Valcarcel, M. W. Hentze, RNA-binding proteins in human genetic disease. *Nat Rev Genet* **22**, 185-198 (2021).
14. E. L. Van Nostrand *et al.*, A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711-719 (2020).
15. Y. Shiio *et al.*, Quantitative proteomic analysis of chromatin-associated factors. *J Am Soc Mass Spectrom* **14**, 696-703 (2003).
16. B. M. Sirbu *et al.*, Analysis of protein dynamics at active, stalled, and collapsed replication forks. *Genes Dev* **25**, 1320-1327 (2011).
17. S. Aranda *et al.*, Chromatin capture links the metabolic enzyme AHCY to stem cell proliferation. *Sci Adv* **5**, eaav2448 (2019).
18. M. R. Rafiee, C. Girardot, G. Sigismondo, J. Krijgsveld, Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins. *Mol Cell* **64**, 624-635 (2016).
19. X. Bao *et al.*, Capturing the interactome of newly transcribed RNA. *Nat Methods* **15**, 213-220 (2018).
20. R. Huang, M. Han, L. Meng, X. Chen, Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *Proc Natl Acad Sci U S A* **115**, E3879-E3887 (2018).
21. G. van Mierlo *et al.*, Integrative Proteomic Profiling Reveals PRC2-Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency. *Cell Stem Cell* **24**, 123-137 e128 (2019).
22. A. Castello *et al.*, Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* **63**, 696-710 (2016).

23. Y. Hayashi *et al.*, Structure-based discovery of NANOG variant with enhanced properties to promote self-renewal and reprogramming of pluripotent stem cells. *Proc Natl Acad Sci U S A* **112**, 4666-4671 (2015).
24. S. Brodsky *et al.*, Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol Cell* **79**, 459-471 e454 (2020).
25. M. L. Romero Romero *et al.*, Simple yet functional phosphate-loop proteins. *Proc Natl Acad Sci U S A* **115**, E11943-E11950 (2018).
26. V. N. Uversky, The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord Proteins* **1**, e24684 (2013).
27. T. Kalkan, A. Smith, Mapping the route from naive pluripotency to lineage specification. *Philos Trans R Soc Lond B Biol Sci* **369**, (2014).
28. S. Tu *et al.*, Co-repressor CBFA2T2 regulates pluripotency and germline development. *Nature* **534**, 387-390 (2016).
29. J. Zhang *et al.*, LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. *Cell Stem Cell* **19**, 66-80 (2016).
30. D. Mayer *et al.*, Zfp281 orchestrates interconversion of pluripotent states by engaging Ehmt1 and Zic2. *EMBO J* **39**, e102591 (2020).
31. C. R. Faehnle, J. Walleshauser, L. Joshua-Tor, Multi-domain utilization by TUT4 and TUT7 in control of let-7 biogenesis. *Nat Struct Mol Biol* **24**, 658-665 (2017).
32. Y. Seki, PRDM14 Is a Unique Epigenetic Regulator Stabilizing Transcriptional Networks for Pluripotency. *Front Cell Dev Biol* **6**, 12 (2018).
33. Y. Zeng *et al.*, Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. *Mol Cell* **61**, 153-160 (2016).
34. L. L. Zagore *et al.*, DAZL Regulates Germ Cell Survival through a Network of PolyA-Proximal mRNA Interactions. *Cell Rep* **25**, 1225-1240 e1226 (2018).
35. H. T. Jenkins, B. Malkova, T. A. Edwards, Kinked beta-strands mediate high-affinity recognition of mRNA targets by the germ-cell regulator DAZL. *Proc Natl Acad Sci U S A* **108**, 18266-18271 (2011).
36. J. Konig *et al.*, iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**, 909-915 (2010).
37. M. R. Rafiee *et al.*, Protease-resistant streptavidin for interaction proteomics. *Mol Syst Biol* **16**, e9370 (2020).
38. G. Hu, Z. Wu, V. N. Uversky, L. Kurgan, Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions. *Int J Mol Sci* **18**, (2017).
39. L. Zhu, C. P. Brangwynne, Nuclear bodies: the emerging biophysics of nucleoplasmic phases. *Curr Opin Cell Biol* **34**, 23-30 (2015).
40. Y. Shin *et al.*, Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* **176**, 1518 (2019).
41. M. T. Wei *et al.*, Nucleated transcriptional condensates amplify gene expression. *Nat Cell Biol* **22**, 1187-1196 (2020).
42. A. Boija *et al.*, Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816 (2018).
43. D. A. Garcia *et al.*, An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors. *Mol Cell* **81**, 1484-1498 e1486 (2021).
44. H. Li *et al.*, DAZL is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* **6**, 455-468 (2019).

45. E. B. Gibbs, R. W. Kriwacki, Linker histones as liquid-like glue for chromatin. *Proc Natl Acad Sci U S A* **115**, 11868-11870 (2018).
46. A. L. Turner *et al.*, Highly disordered histone H1-DNA model complexes and their condensates. *Proc Natl Acad Sci U S A* **115**, 11964-11969 (2018).
47. P. A. Ginno, L. Burger, J. Seebacher, V. Iesmantavicius, D. Schubeler, Cell cycle-resolved chromatin proteomics reveals the extent of mitotic preservation of the genomic regulatory landscape. *Nat Commun* **9**, 4048 (2018).
48. C. Alabert *et al.*, Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nat Cell Biol* **16**, 281-293 (2014).
49. A. E. Kliszczak, M. D. Rainey, B. Harhen, F. M. Boisvert, C. Santocanale, DNA mediated chromatin pull-down for the study of chromatin replication. *Sci Rep* **1**, 95 (2011).
50. F. Kohlmeier, A. Maya-Mendoza, D. A. Jackson, EdU induces DNA damage response and cell death in mESC in culture. *Chromosome Res* **21**, 87-100 (2013).
51. E. Healy *et al.*, PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol Cell* **76**, 437-452 e436 (2019).
52. T. UniProt Consortium, UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **46**, 2699 (2018).
53. P. Jones *et al.*, InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
54. M. Blum *et al.*, The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* **49**, D344-D354 (2021).
55. M. Necci, D. Piovesan, Z. Dosztanyi, S. C. E. Tosatto, MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics* **33**, 1402-1404 (2017).
56. M. Necci, D. Piovesan, D. Clementel, Z. Dosztanyi, S. C. E. Tosatto, MobiDB-lite 3.0: fast consensus annotation of intrinsic disorder flavours in proteins. *Bioinformatics*, (2020).
57. B. Meszaros, G. Erdos, Z. Dosztanyi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* **46**, W329-W337 (2018).
58. S. C. Kwon *et al.*, The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* **20**, 1122-1130 (2013).

Figure 1

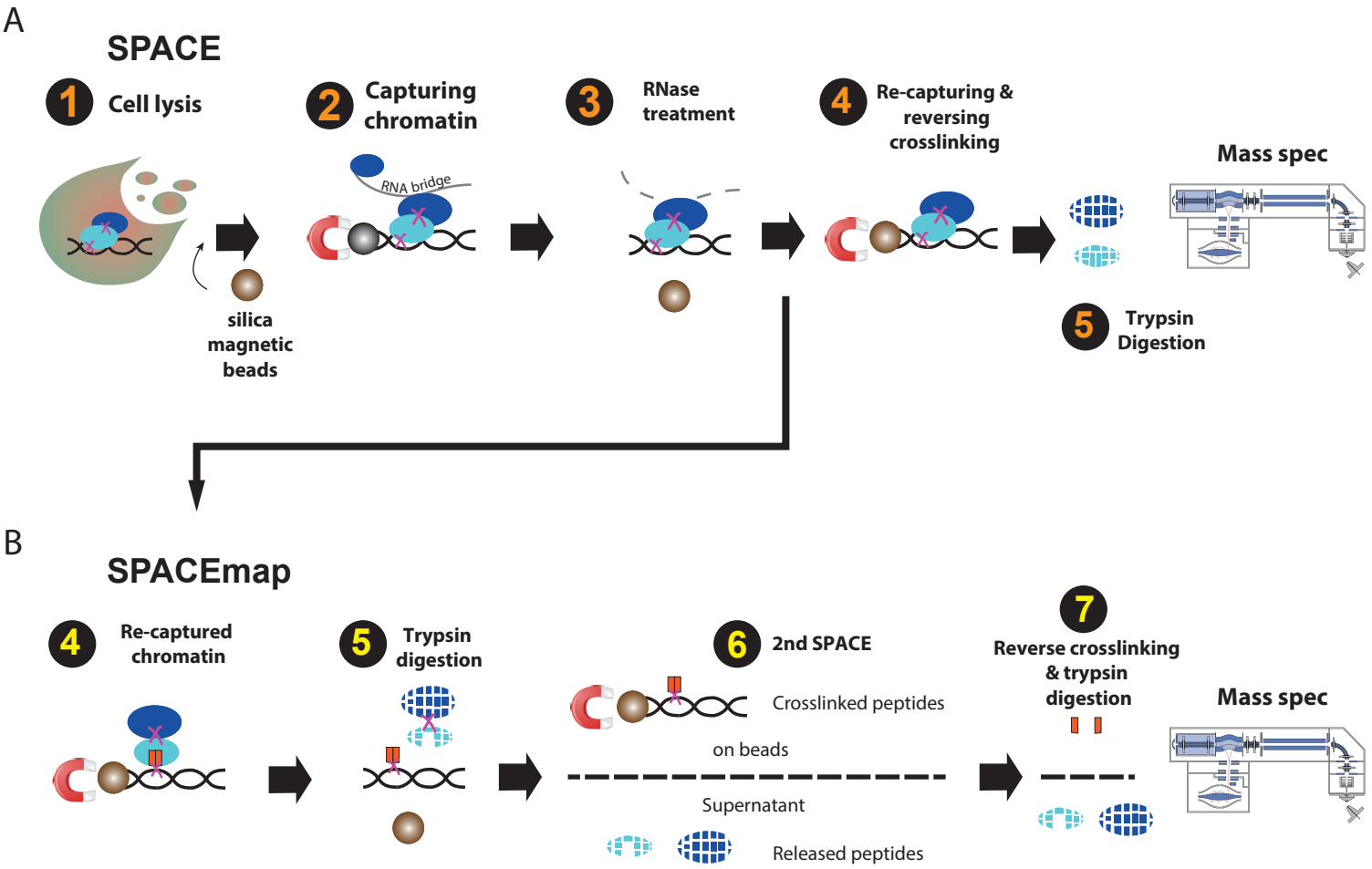

A
SPACE
1 Cell lysis
2 Capturing chromatin
3 RNase treatment
4 Re-capturing & reversing crosslinking
Mass spec
RNA bridge
silica magnetic beads
5 Trypsin Digestion
B
SPACEmap
4 Re-captured chromatin
5 Trypsin digestion
6 2nd SPACE
Crosslinked peptides
on beads
Supernatant
Released peptides
7 Reverse crosslinking & trypsin digestion
Mass spec

Figure 2

A
SPACE
Log2FC>1 & adj.p < 0.1
Non-significant
log2(Crosslinked/nonCrosslinked) R2
log2(Crosslinked/non-Crosslinked) R1
B
~11x
88%
8.4%
Relative abundance %
Potential False Positive
Potential True Positive
Chromatin remodelling
Chromatin modification
Replication & Repair
Transcription factor activity
Involved in transcription
Cell cycle
Splicing
RNA processing
RNA transport
Translation
Metabolic process
Cell adhesion
Protein transport
Protein folding
C
Counts
Abundance
6009
1825
Full proteome
SPACE
Known chromatin proteins
Other proteins present in nucleus
Unexpected proteins
D
RICK
SPACE
CARIC
132
252
1163
269
67
136
125
850
313
RBP 27%
RNA-binding proteins (RBPs)
E
SPACE
RICK+CARIC RBPs
Other RBPs
Chromatin proteins
protein counts
protein rank
F
Counts
Relative Abundance
4711
1811
1825
1279
Chromatin pellet
DmChP
SPACE
SPACE SICAP
G
LogFC>1 & adj.p < 0.1
Non-significant
Protein counts
Cell number
20,000
100,000
500,000
268
960
245
1452
352
1446

Figure 3

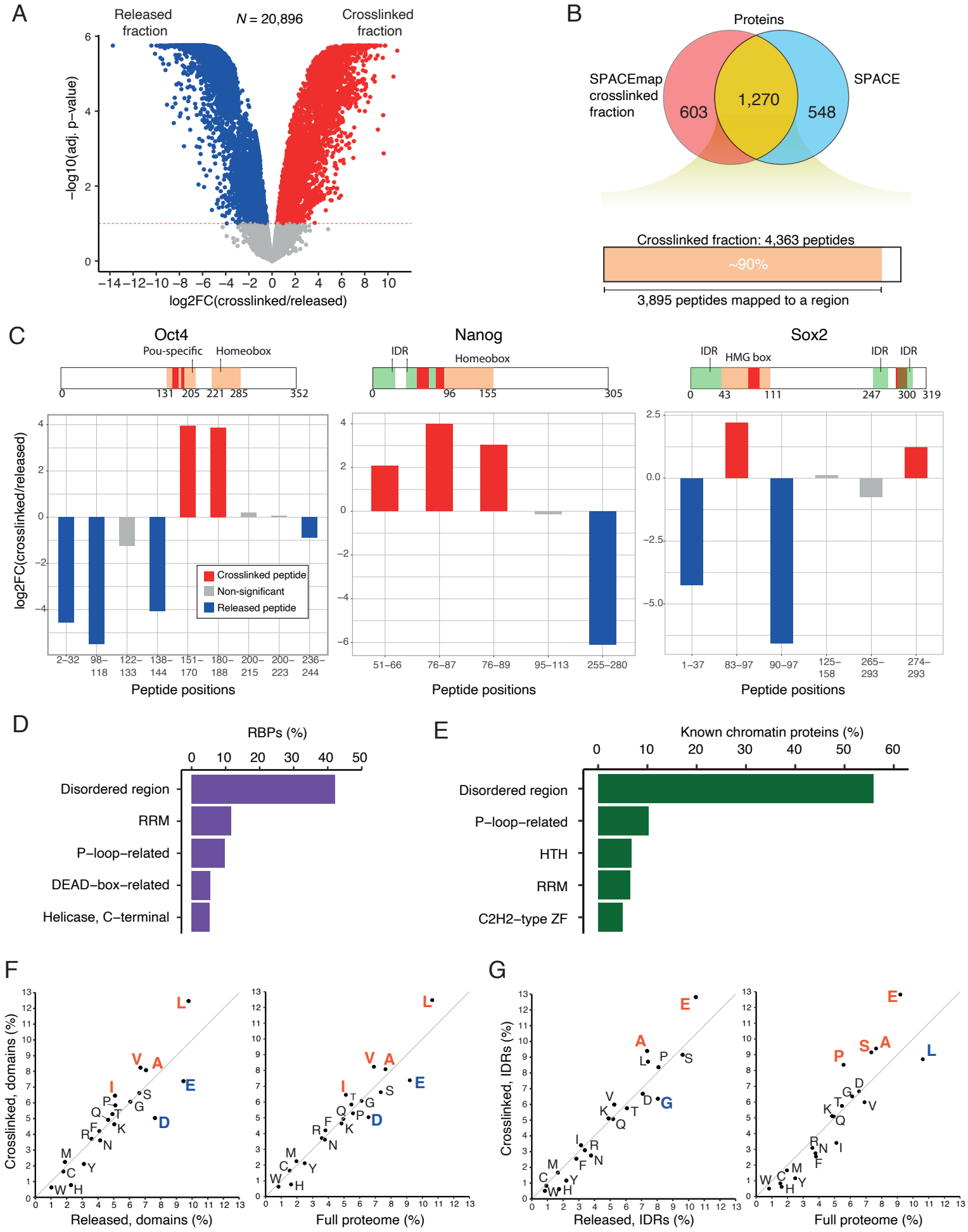

A
Released fraction
N = 20,896
Crosslinked fraction
−log10(adj. p−value)
log2FC(crosslinked/released)
B
Proteins
SPACEmap crosslinked fraction
SPACE
603
1,270
548
Crosslinked fraction: 4,363 peptides
~90%
3,895 peptides mapped to a region
C
Oct4
Pou-specific
Homeobox
Nanog
IDR
Homeobox
Sox2
IDR
HMG box
IDR
IDR
log2FC(crosslinked/released)
Crosslinked peptide
Non-significant
Released peptide
Peptide positions
D
RBPs (%)
Disordered region
RRM
P−loop−related
DEAD−box−related
Helicase, C−terminal
E
Known chromatin proteins (%)
Disordered region
P−loop−related
HTH
RRM
C2H2−type ZF
F
Crosslinked, domains (%)
Released, domains (%)
Full proteome (%)
G
Crosslinked, IDRs (%)
Released, IDRs (%)
Full proteome (%)

Figure 4

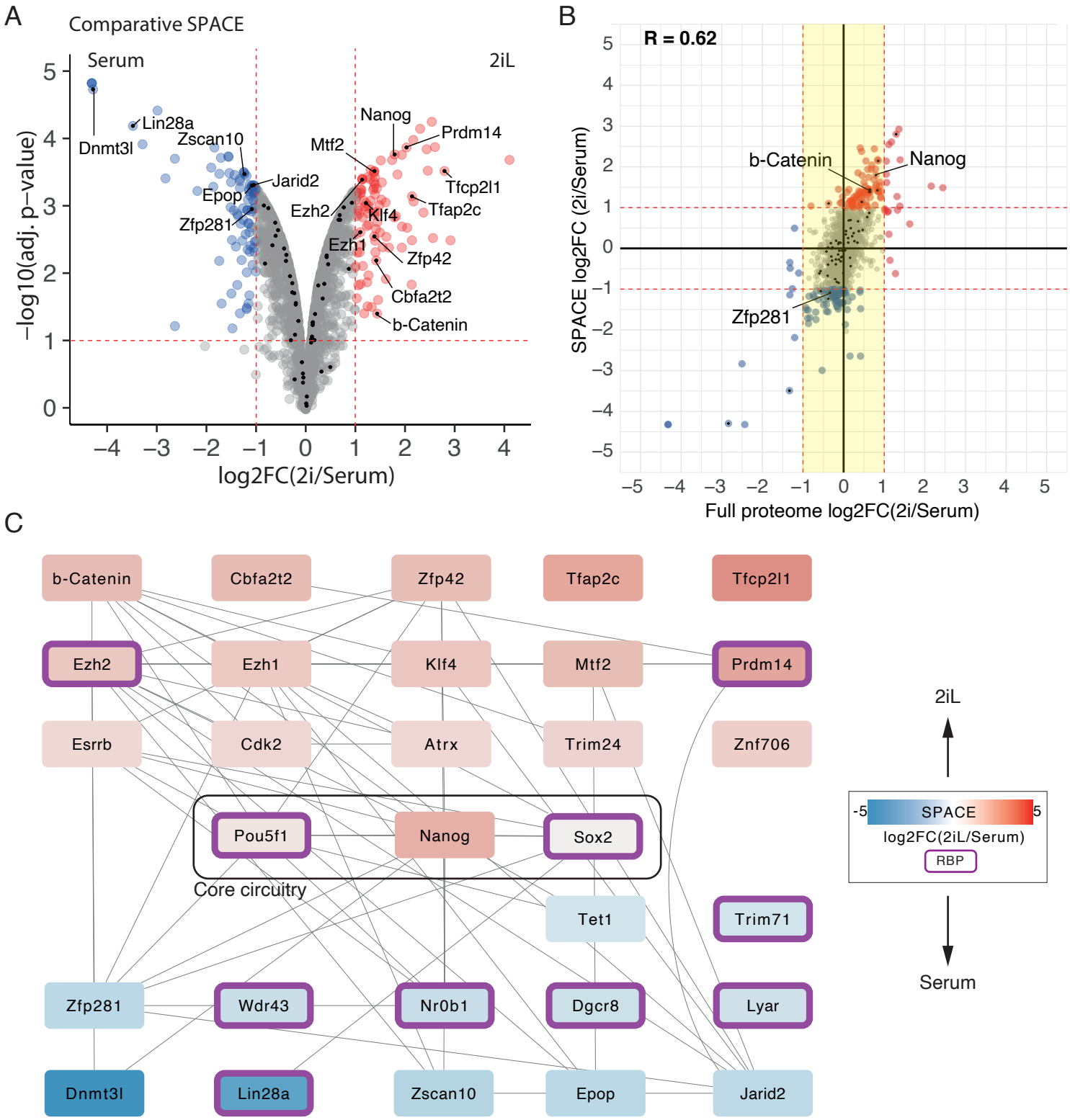
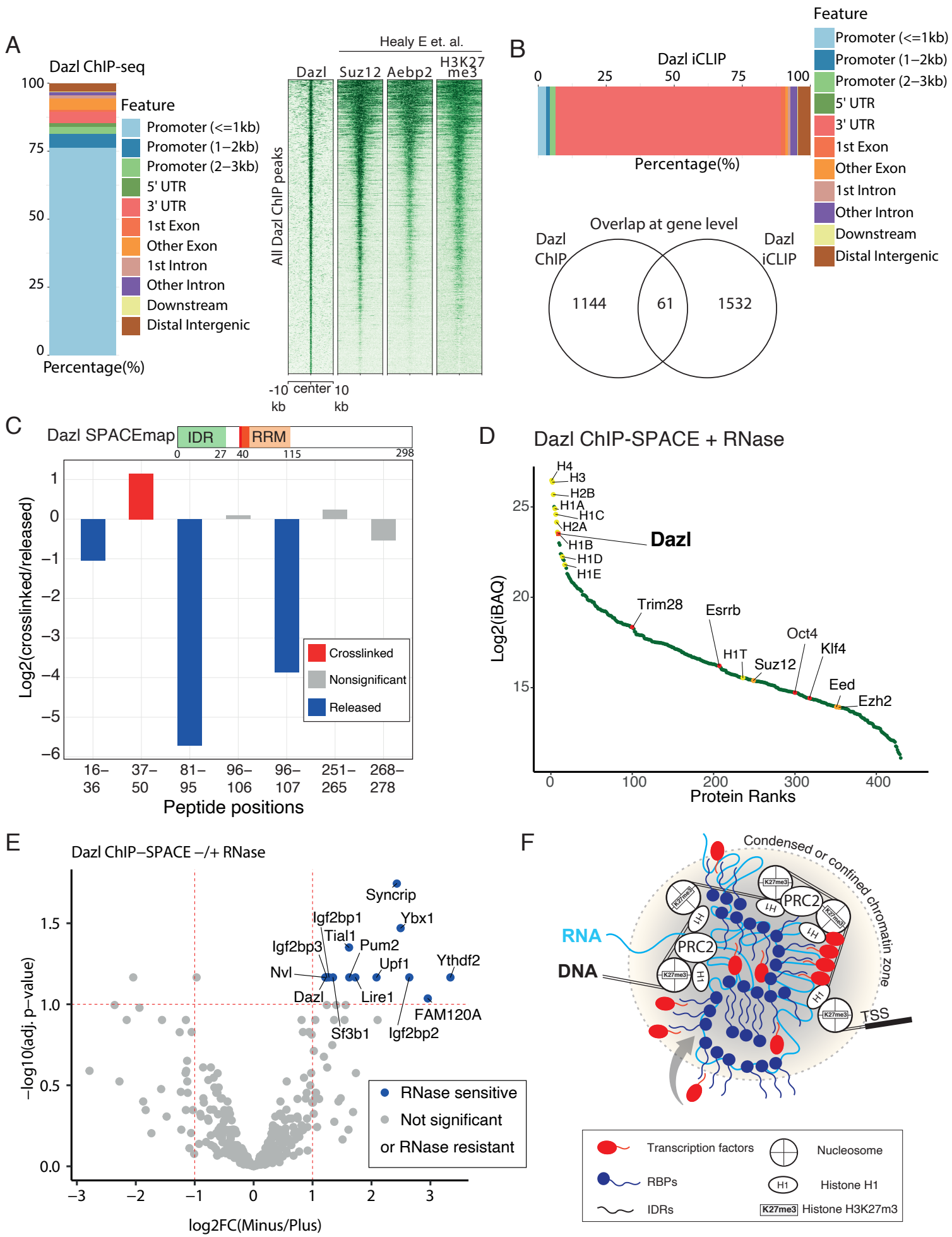
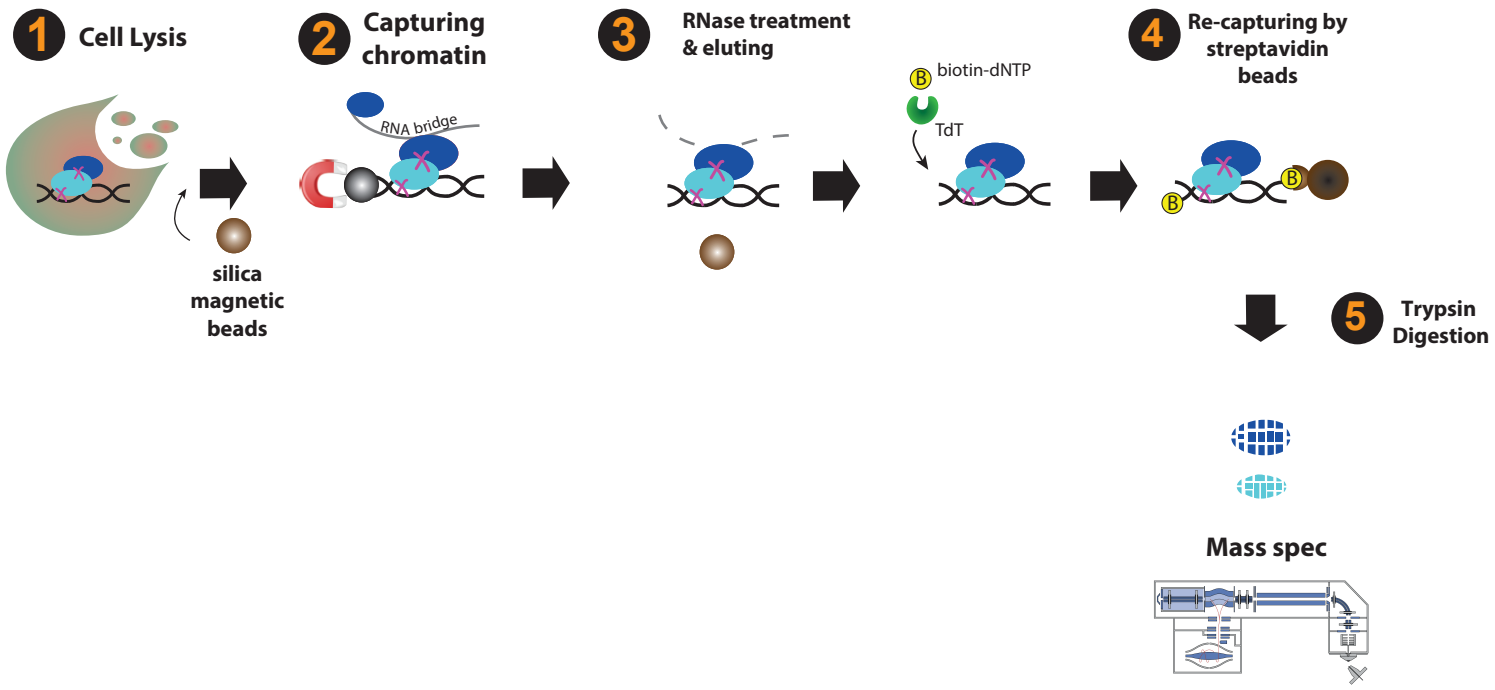

A
Comparative SPACE
Serum
2iL
Dnmt3l
Lin28a
Zscan10
Epop
Zfp281
Jarid2
Mtf2
Ezh2
Ezh1
Nanog
Klf4
Prdm14
Tfcp2l1
Tfap2c
Zfp42
Cbfa2t2
b-Catenin
−log10(adj. p−value)
log2FC(2i/Serum)
B
R = 0.62
b-Catenin
Nanog
Zfp281
SPACE log2FC (2i/Serum)
Full proteome log2FC(2i/Serum)
C
b-Catenin
Cbfa2t2
Zfp42
Tfap2c
Tfcp2l1
Ezh2
Ezh1
Klf4
Mtf2
Prdm14
Esrrb
Cdk2
Atrx
Trim24
Znf706
Pou5f1
Nanog
Sox2
Core circuitry
Tet1
Trim71
Zfp281
Wdr43
Nr0b1
Dgcr8
Lyar
Dnmt3l
Lin28a
Zscan10
Epop
Jarid2
2iL
-5
SPACE
5
log2FC(2iL/Serum)
RBP
Serum

Figure 5


A
Dazl ChIP-seq
Feature
Promoter (<=1kb)
Promoter (1–2kb)
Promoter (2–3kb)
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream
Distal Intergenic
Percentage(%)
Healy E et. al.
Dazl
Suz12
Aebp2
H3K27 me3
All Dazl ChIP peaks
-10 kb
center
10 kb
B
Dazl iCLIP
Percentage(%)
Overlap at gene level
Dazl ChIP
Dazl iCLIP
1144
61
1532
C
Dazl SPACEmap
IDR
RRM
Log2(crosslinked/released)
Crosslinked
Nonsignificant
Released
Peptide positions
D
Dazl ChIP-SPACE + RNase
Log2(iBAQ)
Protein Ranks
Dazl
Trim28
Esrb
H1T
Suz12
Oct4
Klf4
Eed
Ezh2
E
Dazl ChIP–SPACE –/+ RNase
−log10(adj. p−value)
log2FC(Minus/Plus)
RNase sensitive
Not significant
or RNase resistant
Syncrip
Ybx1
Igf2bp1
Tial1
Pum2
Igf2bp3
Upf1
Ythdf2
Nvl
Dazl
Lire1
FAM120A
Sf3b1
Igf2bp2
F
Condensed or confined chromatin zone
RNA
DNA
PRC2
TSS
Transcription factors
RBPs
IDRs
Nucleosome
Histone H1
Histone H3K27m3

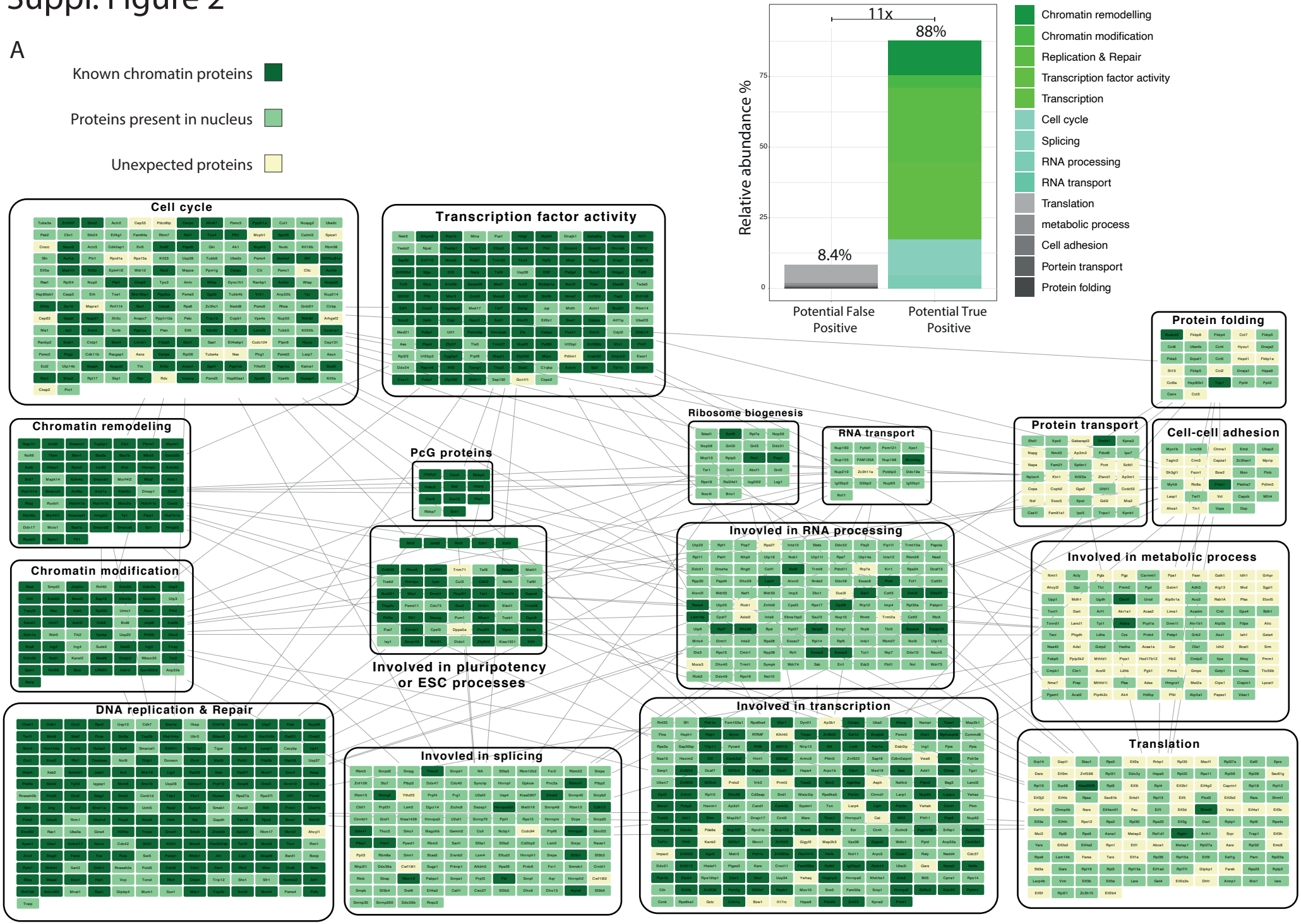
Suppl. Figure 1

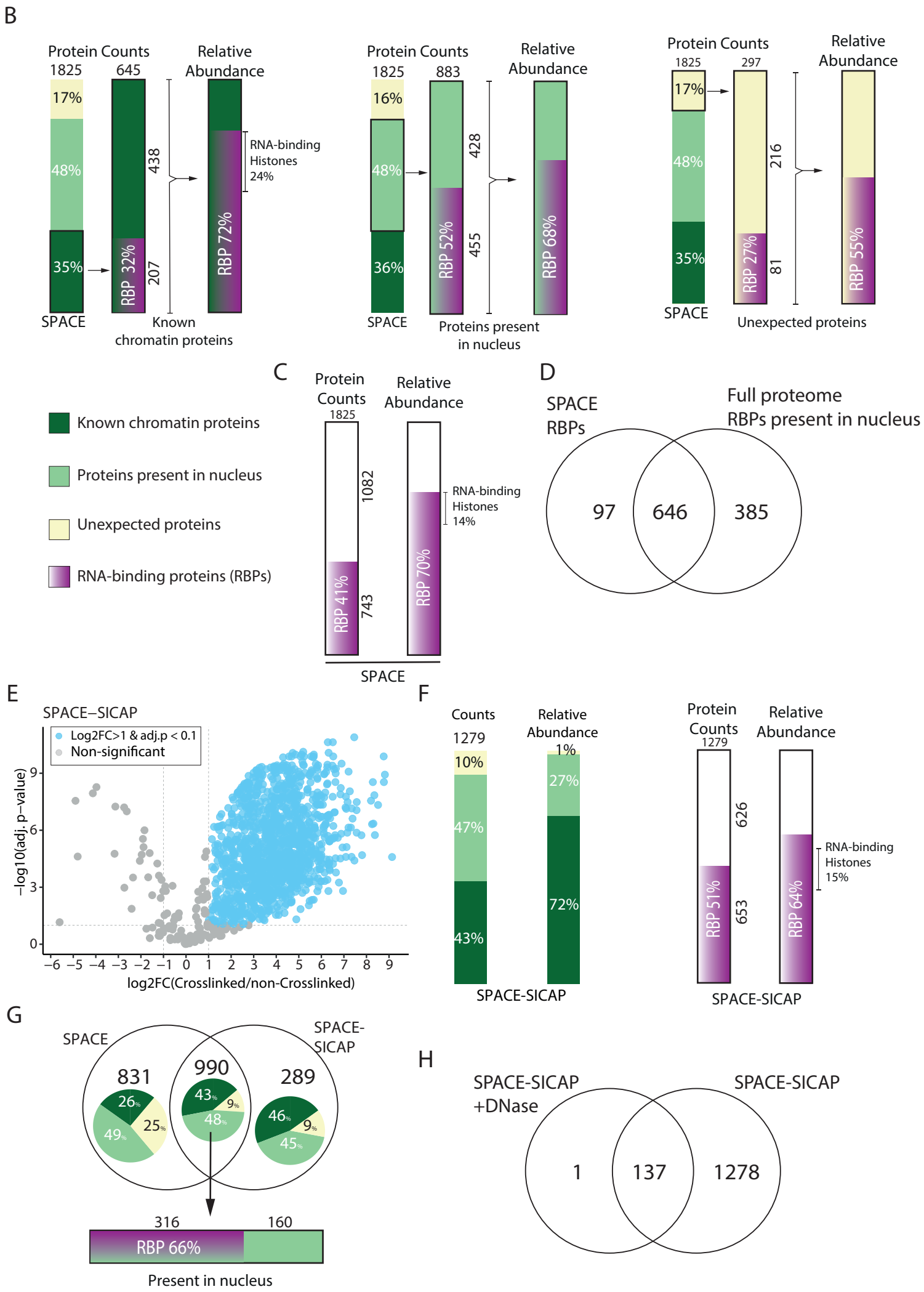
SPACE-SICAP



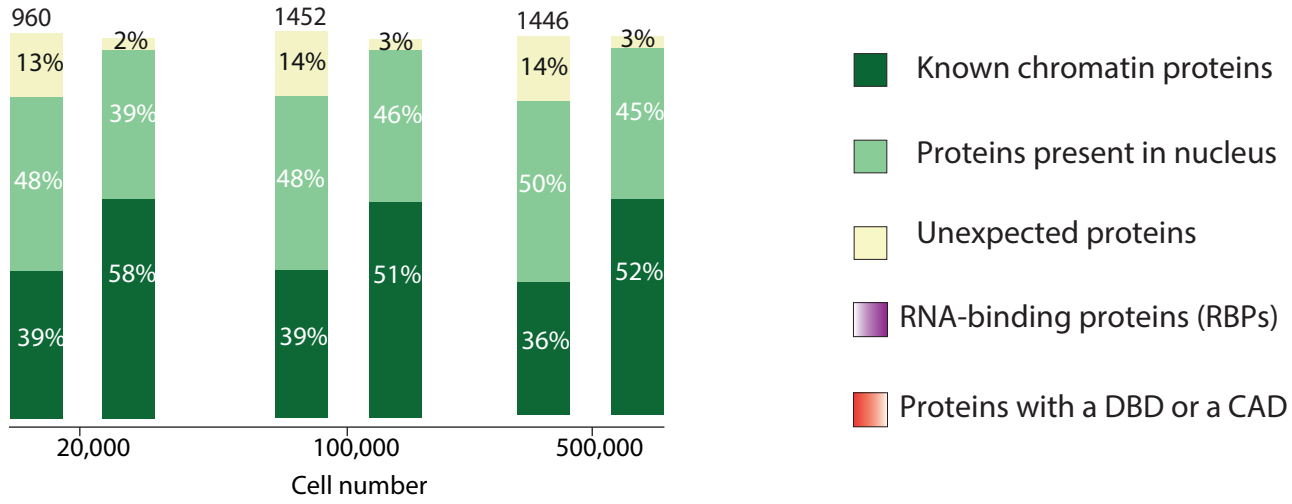
Suppl. Figure 2

A

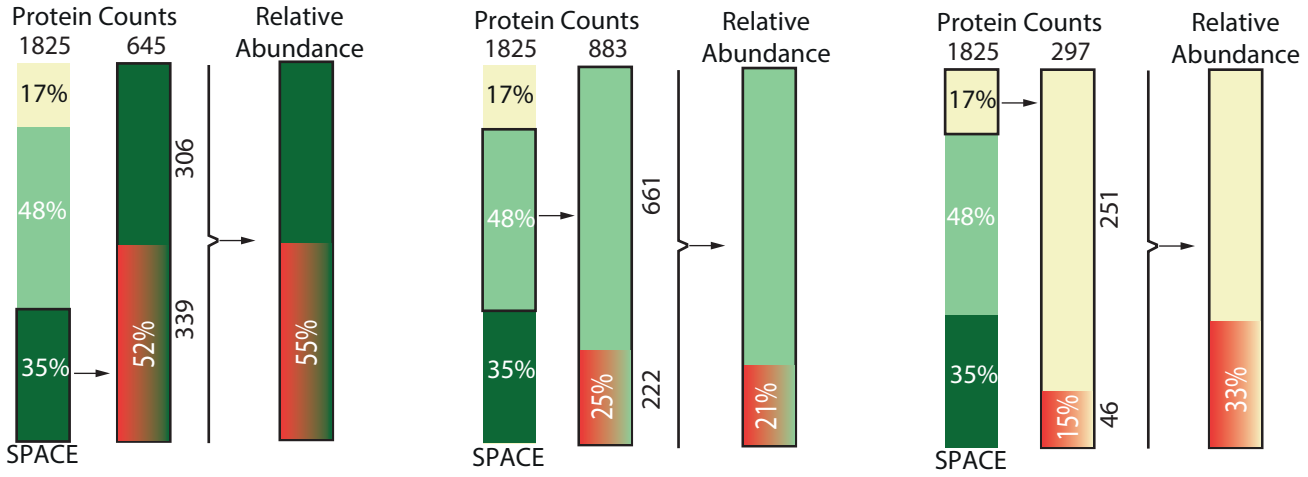



B
Protein Counts
1825
645
Relative Abundance
17%
48%
35%
438
207
RBP 32%
RBP 72%
RNA-binding Histones 24%
SPACE
Known chromatin proteins
Protein Counts
1825
883
Relative Abundance
16%
48%
36%
428
455
RBP 52%
RBP 68%
SPACE
Proteins present in nucleus
Protein Counts
1825
297
Relative Abundance
17%
48%
35%
216
81
RBP 27%
RBP 55%
SPACE
Unexpected proteins
Known chromatin proteins
Proteins present in nucleus
Unexpected proteins
RNA-binding proteins (RBPs)
C
Protein Counts
1825
Relative Abundance
1082
743
RBP 41%
RBP 70%
RNA-binding Histones 14%
SPACE
D
SPACE RBPs
Full proteome RBPs present in nucleus
97
646
385
E
SPACE–SICAP
Log2FC>1 & adj.p < 0.1
Non-significant
−log10(adj. p−value)
log2FC(Crosslinked/non-Crosslinked)
F
Counts
1279
Relative Abundance
10%
47%
43%
1%
27%
72%
SPACE-SICAP
Protein Counts
1279
Relative Abundance
626
653
RBP 51%
RBP 64%
RNA-binding Histones 15%
SPACE-SICAP
G
SPACE
SPACE-SICAP
831
990
289
26%
25%
49%
43%
9%
48%
46%
9%
45%
316
160
RBP 66%
Present in nucleus
H
SPACE-SICAP +DNase
SPACE-SICAP
1
137
1278

I

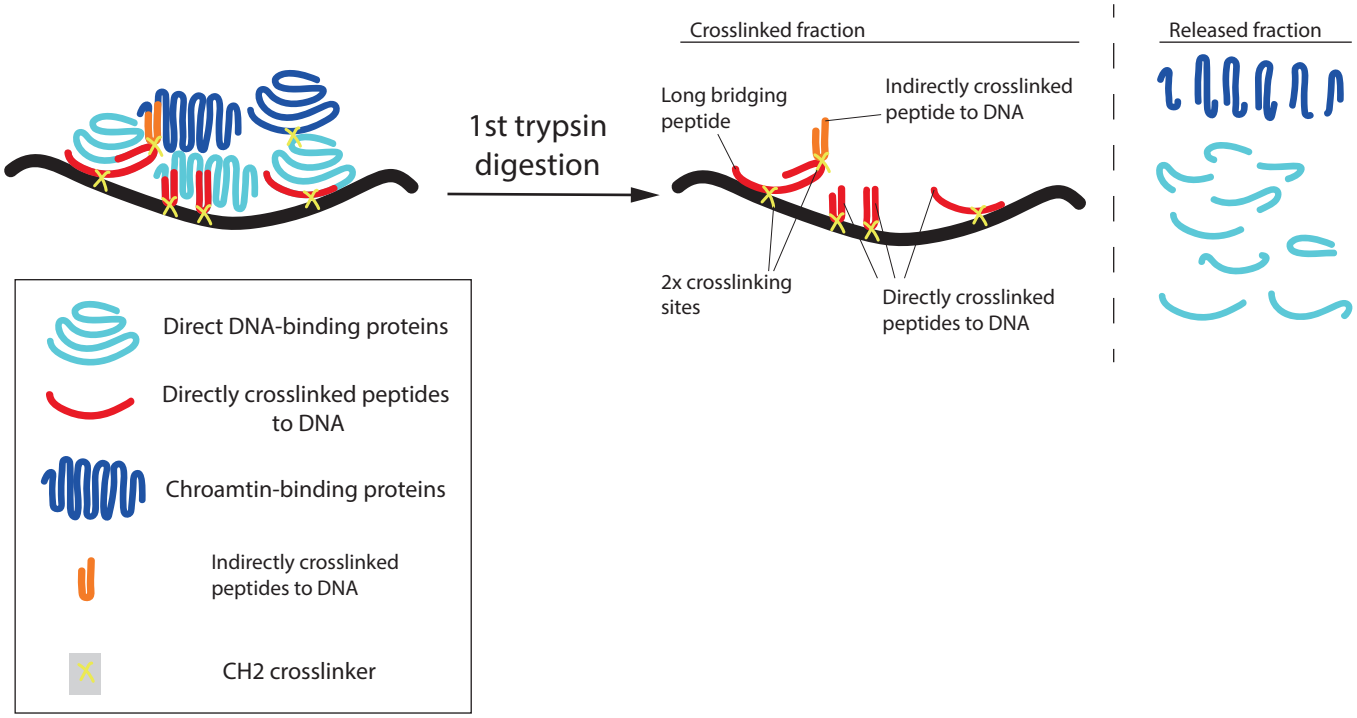

960
13%
48%
39%
2%
39%
58%
1452
14%
48%
39%
3%
46%
51%
1446
14%
50%
36%
3%
45%
52%
20,000
100,000
500,000
Cell number
Known chromatin proteins
Proteins present in nucleus
Unexpected proteins
RNA-binding proteins (RBPs)
Proteins with a DBD or a CAD

J


Protein Counts
1825
645
Relative Abundance
17%
48%
35%
SPACE
52%
306
339
55%
Protein Counts
1825
883
Relative Abundance
17%
48%
35%
SPACE
25%
661
222
21%
Protein Counts
1825
297
Relative Abundance
17%
48%
35%
SPACE
15%
251
46
33%

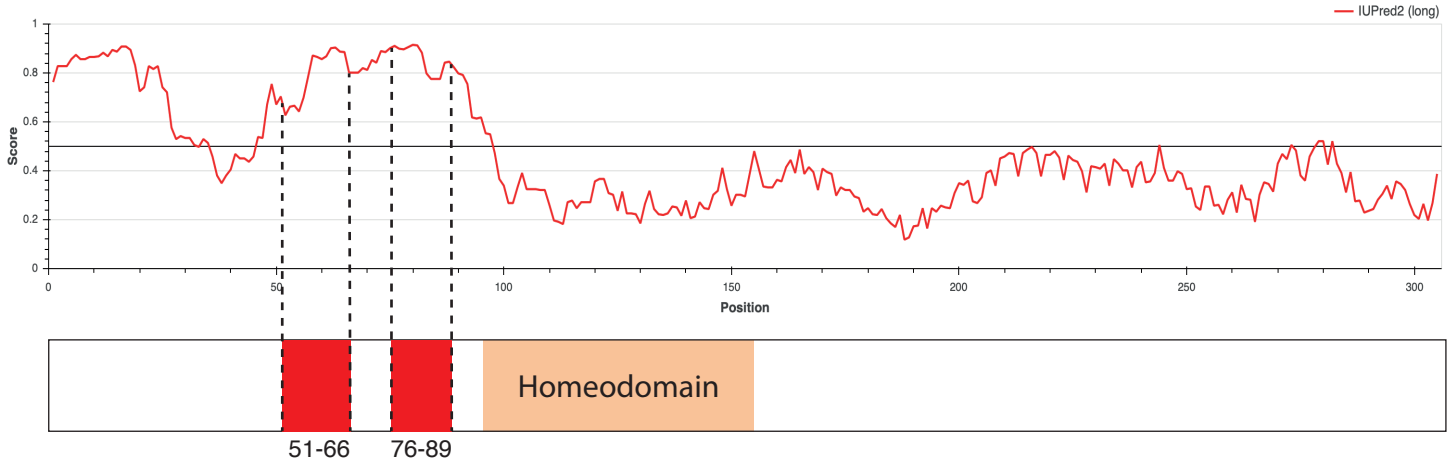
Suppl. Figure 3

A



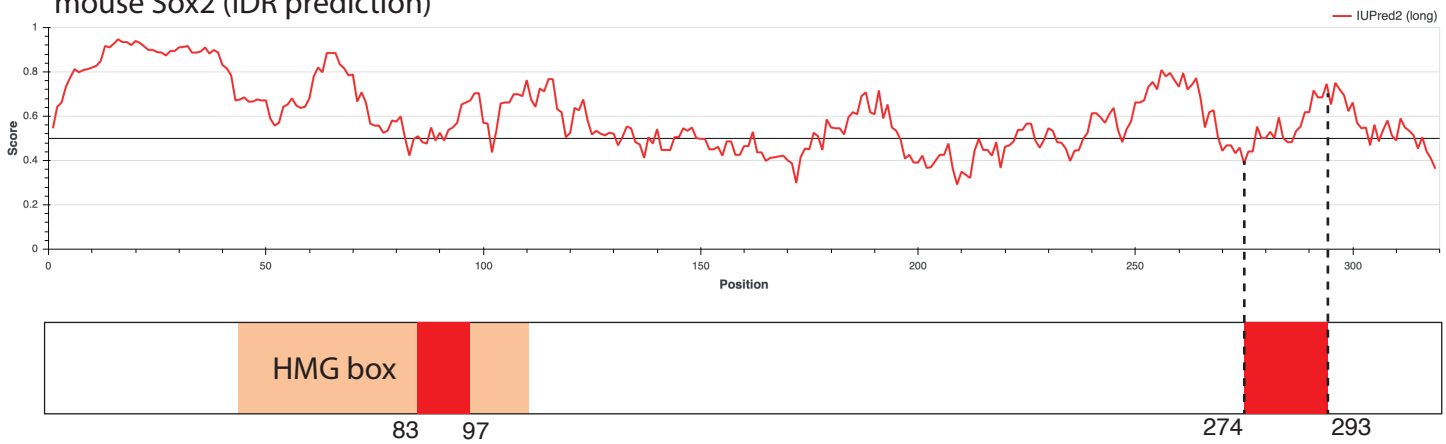
B

mouse Nanog (IDR prediction)



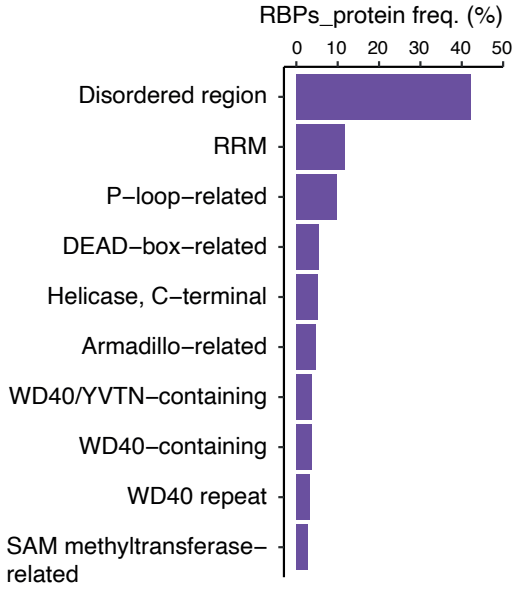
C

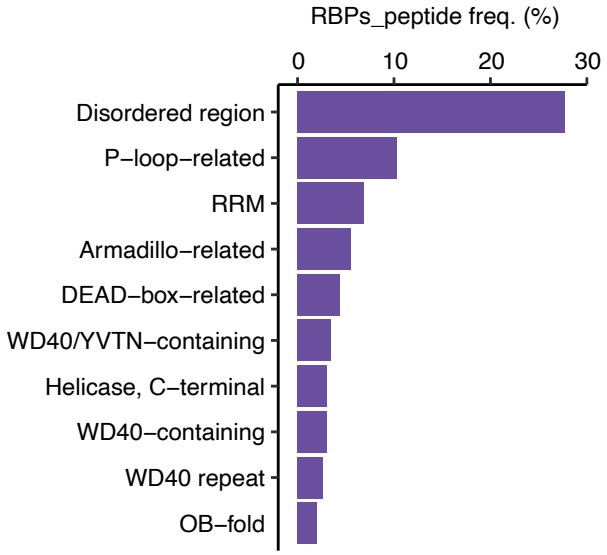
mouse Sox2 (IDR prediction)



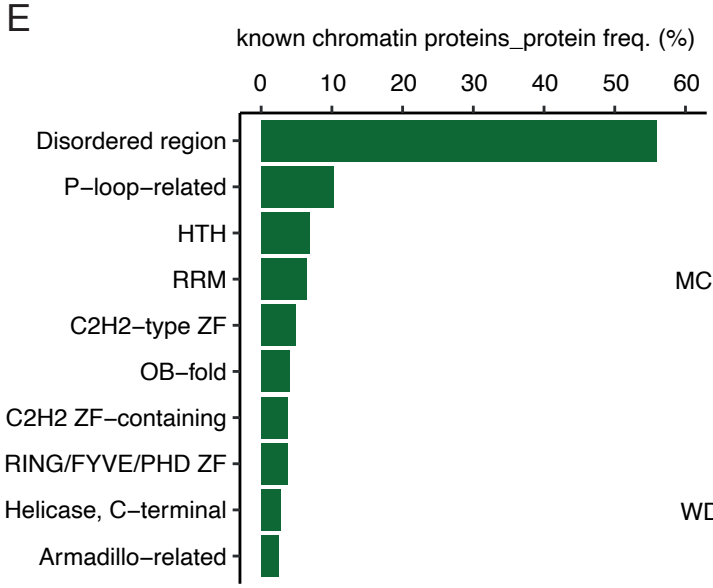
Suppl. Figure 3

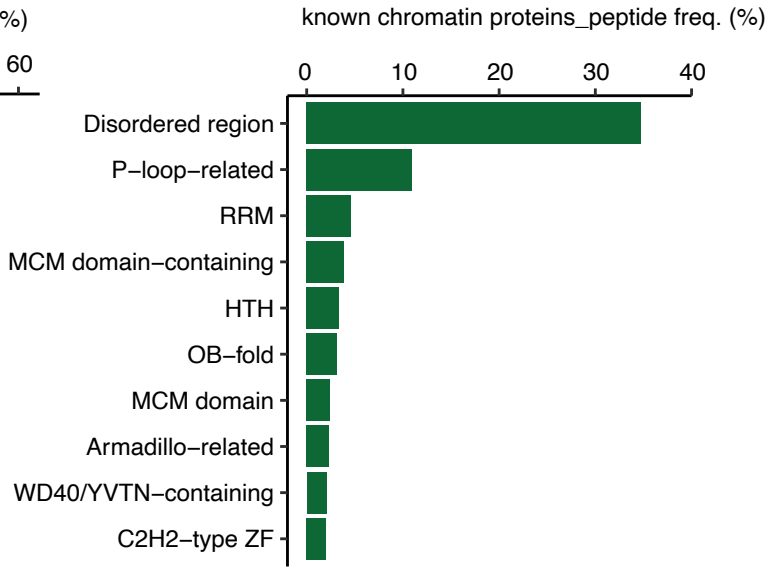
D





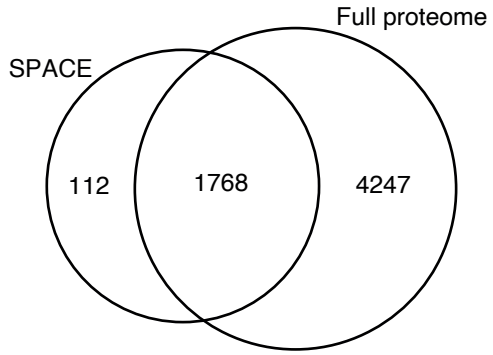
E



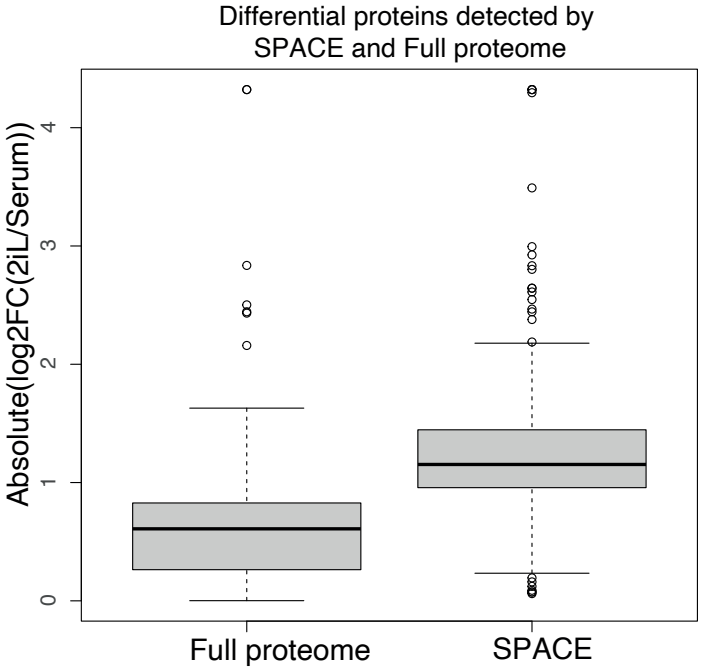


Suppl. Figure 4

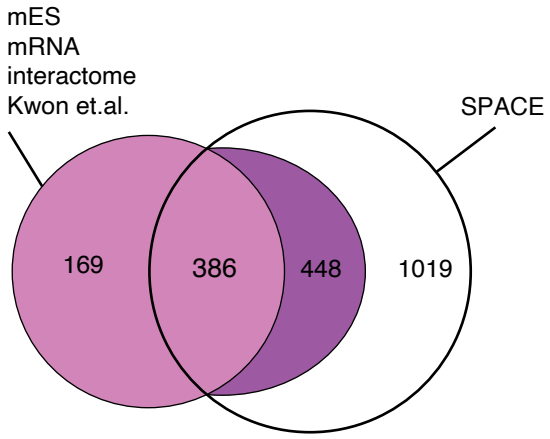
A



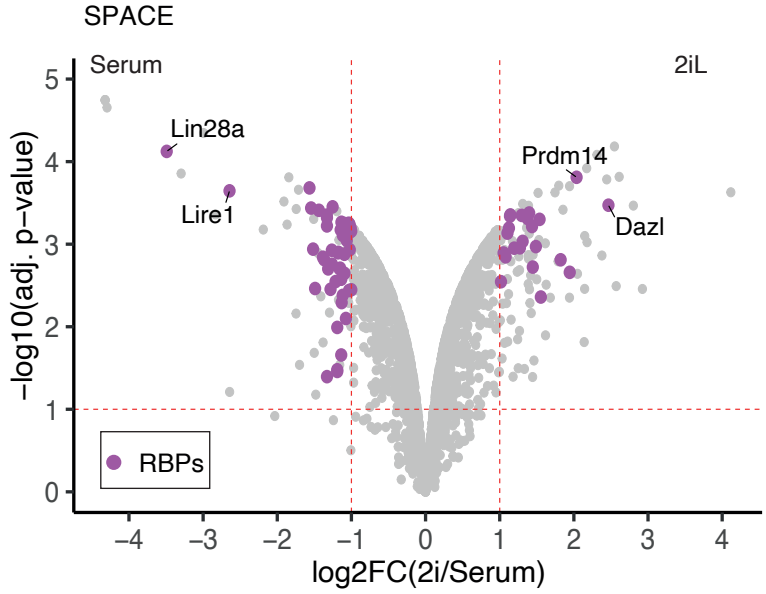
B



C

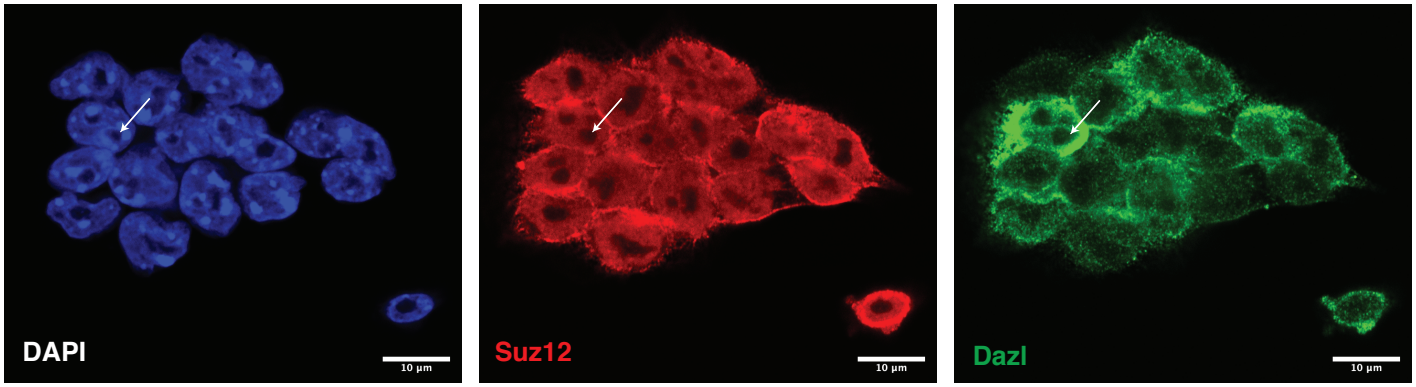


D

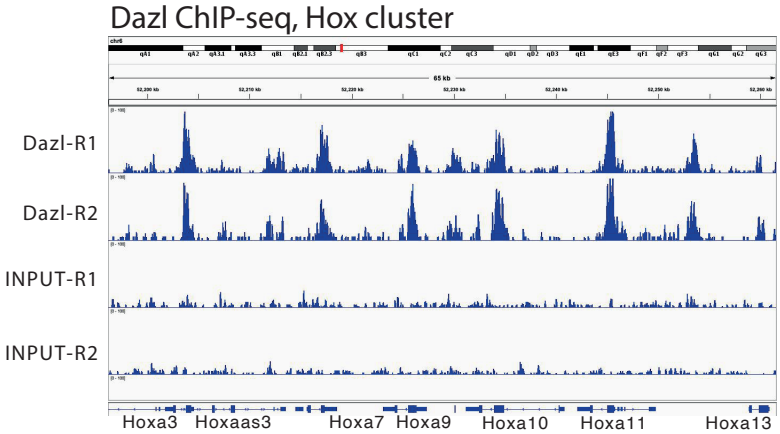
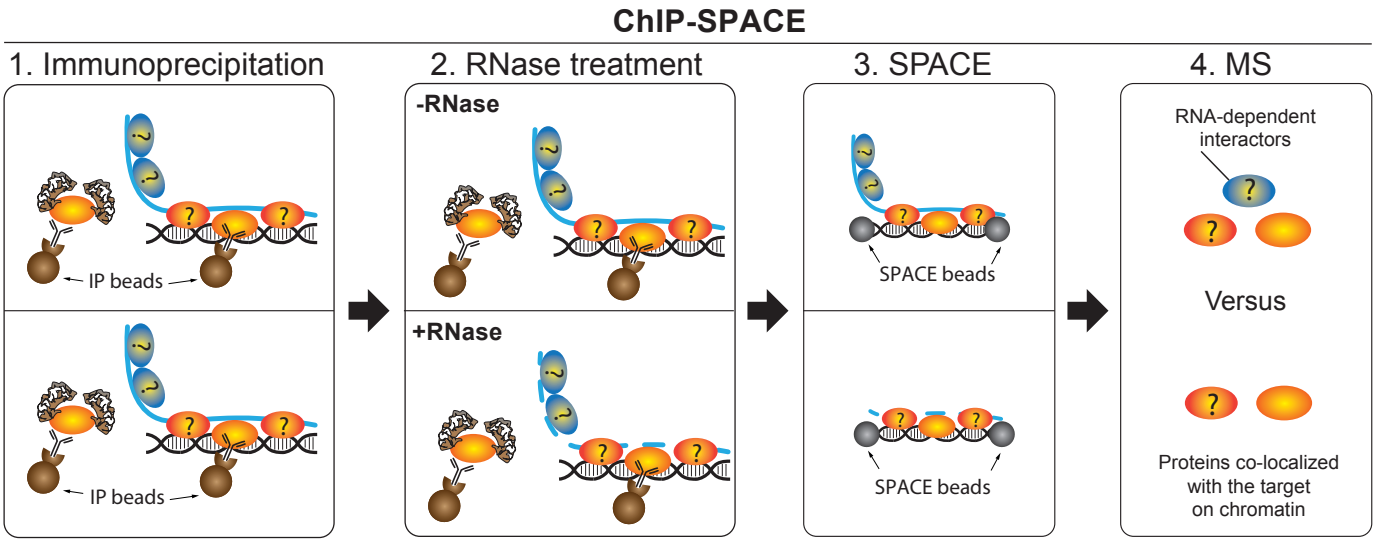


Suppl. Figure 5

A



B



C