SPACE exploration of chromatin proteome to reveal associated RNA-binding proteins

Mahmoud-Reza Rafiee[1*], Julian A Zagalak[1,2], Giulia Tyzack[1,3,4], Rickie Patani[1,3,4], Jernej Ule[1,2], Nicholas M Luscombe[1,5,6*]

[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] Sobell Department of Motor Neuroscience and Movement Disorders, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK. [4] Department of Neuroinflammation, UCL Institute of Neurology, Queen Square, London WC1N 1PJ, UK. [5] UCL Genetics Institute, University College London, Gower Street, London WC1E 6BT, UK. [6] Okinawa Institute of Science & Technology Graduate University, Okinawa 904-0495, Japan

* corresponding authors

Abstract

Chromatin is composed of many proteins that mediate intermolecular transactions with the genome. Comprehensive knowledge of these components and their interactions is necessary for insights into gene regulation and other activities; however, reliable identification of chromatin-associated proteins remains technically challenging. Here, we present SPACE (Silica Particle Assisted Chromatin Enrichment), a stringent and straightforward chromatin-purification method that helps identify direct DNA-binders separately from chromatin-associated proteins. We demonstrate SPACE's unique strengths in three experimental set-ups: the sensitivity to detect novel chromatin-associated proteins, the quantitative nature to measure dynamic protein use across distinct cellular conditions, and the ability to handle 10-25 times less starting material than competing methods. In doing so, we reveal an unforeseen scale of association between over 500 nuclear RNA-binding proteins (RBPs) with chromatin

and DNA, providing new insights into their roles as important regulators of genome maintenance and chromatin composition. Applied to iPSC-derived neural precursors, we discover a new role for the amyotrophic lateral sclerosis (ALS)-causing Valosin Containing Protein (VCP) in recruiting DNA-damage components to chromatin, thus paving the way for molecular mechanistic insights into the disease. SPACE is a fast and versatile technique with many applications.

Main

Gene expression and genome maintenance are enabled by dynamic interactions between diverse proteins, DNA and RNA. The repertoire of chromatin-associated proteins continues to be joined by non-canonical members; for example, metabolic enzymes were reported to perform moonlighting functions in transcriptional regulation and DNA repair[1]. Thus, methods that uncover novel associations of proteins with chromatin with high specificity are needed to fully understand the mechanisms of genome maintenance, transcriptional and post-transcriptional regulation.

Past studies relied on a density gradient[2] or sedimentation[3] for chromatin purification. These techniques are prone to contamination by the abundant cytoplasmic and other non-specific proteins that co-precipitate with the nuclear or chromatin fractions, partly through attraction to the charged DNA molecule. Moreover, the interconnecting membranes of the nucleus and endoplasmic reticulum are hardly separated by cell fractionation approaches. Existing methods also lack a clear negative control for sedimentation or precipitation, making it hard to distinguish genuine from non-specific interactors. One improvement for increased specificity is to label DNA by pre-incubating cells with ethynyl deoxyuridine (EdU) or biotin-dUTP[4-7]. This allows chromatin fragments to be captured with magnetic beads, thus

enabling stringent washing with denaturing reagents. DNA labelling after a pulse was successfully used to study nascent chromatin and DNA replication; however, prolonged treatment with these modified nucleotides is toxic to cells, so potentially distorting measurements. Sensitive cells such as mouse embryonic stem cells (mESC) are particularly affected[8]. Finally, none of the previous purification methods is able to distinguish between direct DNA-binders and chromatin-associated proteins. Therefore, a stringent and straightforward protocol to purify chromatin without manipulating cells is potentially of great use.

Here, we present SPACE (Silica Particle Assisted Chromatin Enrichment), a reproducible and specific method that relies on silica magnetic beads for parallel processing of large numbers of samples at low cost. It is fast to perform and it is readily combined with multiple protocols including ChIP, SICAP, mass spectrometry and sequencing. To demonstrate the power of the method, we evaluated SPACE in three different experimental settings. First, we studied the global chromatin composition of mESCs: we successfully identified the expected DNA- and chromatin-binding proteins, as well as over 500 RBPs that were not known to be associated with chromatin. One of these, Dazl - best known for targeting the 3' untranslated regions (3' UTRs) of mRNAs - binds primarily to the transcription start sites (TSSs) of developmental genes. Second, to demonstrate its versatility and quantitative nature, we combined SPACE with ChIP (chromatin immunoprecipitation) to study the dynamic changes in chromatin composition at Nanog-bound enhancers upon transition from serum to 2iL medium of mESCs. Finally, we applied SPACE to human induced pluripotent stem cell-derived (hiPSC) neural precursors to study changes in the chromatin proteome of hiPSC lines containing ALS-causing point mutations in the valosin-containing protein (VCP), as a chromatin-associated RBP[9, 10]. Our approach uncovered mutation-specific deregulation in

chromatin composition during neuronal specification, thus shedding light on the pathogenic mechanism linked to these mutations.

Results

SPACE reveals many RNA-binding proteins that bind chromatin and DNA

Silica matrices (columns or beads) are widely used to purify DNA from diverse samples; however, 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 chromatin and SPACE exploits the capacity of silica magnetic beads to purify formaldehyde-crosslinked chromatin in the presence of chaotropic salts, followed by chromatin digestion on the beads (Fig. 1, steps 1-4a). Unlike other methods, in SPACE chromatin is not precipitated and cells are not fed with modified nucleotides. Non-crosslinked negative controls are prepared 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 are able to determine whether a protein is isolated due to crosslinking or non-specific association to the beads and other proteins. SPACE is fast to perform (~ 1h from the cell lysis to the start of protein digestion), yet stringent: denaturing reagents include 4M guanidinium isothiocyanate, 2% Sarkosyl, 80% ethanol and 100% acetonitrile that efficiently remove contaminants. Chromatin fragments are also treated extensively with RNase remove RNA contaminations. Finally, the method is readily extended to identify direct DNA-binding proteins by a two-step protein digestion strategy (Fig. 1, steps 4b-6).

We first applied SPACE-MS to mESCs cultured in 2iL (2i+LIF) and we identified 1,965 significantly enriched proteins compared with the negative control (Fig. 2A, Table 1A). We grouped proteins into four categories based on their Gene Ontology annotations (Fig. 2B): 1)

461 (23%) known DNA-binding proteins (DB); 2) 219 (12%) known chromosome- or chromatin-binders (CB) but not DNA-binders; 3) 946 (48%) proteins known to be present in the nucleus (PN) but not annotated as DNA- or chromatin-binding; 4) and 339 (17%) that we labelled "unexpected" (UE) due to the large proportion of proteins annotated as having translation initiation activity and being located in the cytoplasm. Weighted by protein intensities (iBAQ), it is apparent that DB and PN proteins are the most abundant (76% and 20% respectively; Fig. 2B, right bars), indicating the relatively low abundance of unexpected proteins. Compared with the 6,007 proteins detected in the full proteome of whole-cell lysates, SPACE clearly enriches for canonical DNA-binders, with additional representation of the PN proteins that are not yet known to bind chromatin (Fig. 2B, right bars). Among PN proteins, there is strong enrichment for RNA-binding proteins (RBPs) and splicing factors (Fig. 2B, PN bars and Supplementary Fig. 2A-B). We identified 1,618 RBPs in our full proteome dataset. While 827 out of the 1618 RBPs are localized to the nucleus, interestingly, 522 (63%) of these were detected also by SPACE. This demonstrates that RBPs and splicing factors are the most common previously unannotated chromatin-associated proteins.

Next, to increase the stringency of purification and to distinguish direct DNA binders, we combined SPACE with SICAP (Selective Isolation of Chromatin-Associated Proteins[11]). Here, DNA is biotinylated using a terminal deoxynucleotidyl transferase and captured by protease-resistant streptavidin magnetic beads (Fig. 1, steps 4b-6; SPACE-SICAP). In addition to the double stringent purification, SPACE-SICAP also allows to identify some of the direct DNA-binders using the same samples prep. For the latter aim, we use streptavidin beads[12] with a two-step protein digestion strategy[13]. Following SPACE-SICAP purification, LysC is applied to break down proteins, leaving only small fragments of direct DNA-binding proteins crosslinked with DNA; we then reverse the crosslinks by heating, to identify the DNA-bound peptides. Thus, SPACE-SICAP yields two fractions: the supernatant of the LysC

digestion, containing indirect DNA-binding proteins, and the DNA-bound fraction (referred to as SPACE-SICAP I and II, respectively). Both fractions are further digested by Trypsin to improve protein identification (Supplementary Fig. 1).

We identified 1,567 enriched proteins in SPACE-SICAP I, (~25% less than SPACE alone; Fig. 2C and Table 1C) and identified proteins are distributed into the four categories similarly (Fig. 2D). The higher enrichment fold change compared with non-crosslinked controls reflects the increased stringency of SPACE-SICAP (compare Fig. 2A and C). 1,159 proteins are identified by both SPACE and SPACE-SICAP (Fig. 2E); of these, nearly half belong to the PN category, of which two-thirds are RBPs and a fifth are involved in RNA splicing (Fig. 2E, the bar). Comparing SPACE-SICAP results with a DNase-treated control indicates that identification of RBPs depends on the presence of DNA, as the number of enriched proteins dramatically diminishes to just 34 proteins (Fig. 2E).

Considering the vast parts of the proteins that are removed during the LysC digestion, the number of reproducibly identified proteins (at least 2 out of 4 replicates) in SPACE-SICAP II is just 55 proteins (Fig. 2F and Table 1D). This includes 31 known DNA-binding proteins including 7 histones, whereas most of the remaining enriched proteins are RBPs (14/24, Fig. 2G) that are primarily involved in mRNA splicing (10/14). 52% of the peptides we identified using this approach are inside or within 50 amino acids of annotated DNA/RNA-binding domains. As such, H15 (the linker histone) determined to be the most frequent domain that we identified using the enriched peptides (Supplementary Fig. 2D-E). Together, we show that SPACE successfully enriches for chromatin-associated proteins, with both SPACE and SPACE-SICAP identifying an unexpectedly large number of nuclear RBPs, some of which directly bind DNA.

Dazl co-localizes with PRC2 and H3K27me3 in the ground-state of mESC

To demonstrate the quantitative capacity of SPACE, we next compared mESC grown in 2iL (heavy SILAC) and serum conditions (light SILAC), which correspond to the ground and primed states of pluripotency in the mouse, respectively. Although previous studies compared these two pluripotency states[14], it is still not clear how pluripotency gene regulatory networks are reinforced by chromatin-associated deregulations in different states. We identified 1,879 proteins in total (Fig. 3A): 100 proteins are statistically significantly more abundant in 2iL and 87 in serum (fold-change > 2 and adj. p-value < 0.1, moderated t-test, Table 2). We also compared the SPACE results with the full proteome from the total cell lysate, which identified 6007 proteins, 1767 proteins overlapping with the SPACE dataset, while 112 proteins were only identified using SPACE, including pluripotency transcription factors such as Prdm14 and Klf4 (Supplementary Fig. 3A-B). There is a strong correlation in log2 fold-change values between SPACE and the full proteome (Fig 3B; R^2 = 0.64), but SPACE displays a larger dynamic range (on average ~2-fold higher than full proteome ratios, Supplementary Fig. 3C); demonstrating that it is more sensitive to changes in the abundance of chromatin-associated proteins.

98 proteins with known functions in the maintenance of pluripotency or exit from it or blastocyst development were identified in the 2iL and/or serum conditions (Fig. 3C). Among them are chromatin proteins that physically interact with the core pluripotency circuitry (Nanog, Oct4, Sox2): Tfcp2l1, Prdm14, Cbfa2t2, b-Catenin, Zfp42 (Rex1), Klf4, Trim24 and Esrrb bind to chromatin more abundantly in 2iL condition, whereas Lin28a, Zscan10, Znf281 and Nr0b1 bind more abundantly in serum condition (Fig. 3C). Although members of both groups are known for roles in preserving pluripotency, our results suggest that the core circuitry of pluripotency is supported by different sets of proteins in the 2iL and serum conditions. Among these proteins, Prdm14, Lin28a and Nr0b1 are known for their RNA-

binding functions. Notably, we identified two additional RBPs that have not been previously implicated in pluripotency, but had highly differential chromatin-binding in 2iL and serum (moderated t-test adj. p-value < 0.1 and fold-change > 4, Supplementary Fig. 3E): Dazl and Lire1. When clustered with the 98 known pluripotency proteins according to their log2 fold-changes (2iL/Serum) and protein abundances (iBAQ), Dazl clusters with the 2iL-related factors such as Nanog, Tfcp2l1, Prdm14 (cluster 6), whereas Lire1 clusters with serum-related factors such as Lin28a and Dnmt3l (cluster 5) (Fig. 3D).

We investigated the genome-wide locations of Dazl binding sites by chromatin immunoprecipitation and sequencing (ChIP-seq) using a validated antibody (Fig. 3E and S3F), revealing ~1,300 reproducible peaks (IDR < 0.01). Surprisingly, 75% of peaks are found close (< 1kb) to transcription start sites (TSS); many target genes are pluripotency regulators, including Esrrb, Sox2, several Wnt ligands and Frizzled receptors (Supplementary Fig. 3G, Supplementary Table 2 C-D). As most of the Dazl target genes are involved in development and differentiation of mESC, and considering the role of PRC2 in regulating differentiation genes in mESCs, we compared the pattern of Dazl, Suz12, Aebp2 and H3K27me3 profiles on the mESC genome (Fig. 3F). Interestingly, we observed a very similar pattern, 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[15]. We identified 2,550 peaks, 2099 of which locate to 3' UTRs, and only 166 locate within 3,000 nucleotides from the 5' end of mRNAs (Supplementary Table 2E). Thus, the RNA binding sites were positioned at different locations in genes compared to DNA-binding sites, which were located mainly in promoters (Fig. 3G and Supplementary Fig. 3H). 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 ChIP-seq-defined peak on their genes (gene body and 3kb upstream of the TSS) also have an iCLIP-defined peak on their transcript (Supplementary Fig. 3I). These results suggest that the DNA- and RNA-binding functions of Dazl are mechanistically independent.

Dynamics of proteins at Nanog-binding enhancers during the pluripotency transition

In addition to understanding the global dynamics of chromatin, we combined SPACE with ChIP to gain a detailed view of chromatin-associated proteins that co-localise with a target of interest. In doing so, we overcome the challenge faced by ChIP-MS, which is limited in the rigour of washing due to the sensitivity of the antibodies. We previously developed and recently updated ChIP-SICAP that enables very stringent washes[11], which revealed proteins co-localized with the core circuitry of pluripotency (OSN). Here, we compare ChIP-SICAP and ChIP-SPACE by immunoprecipitating Nanog (Supplementary Fig. 4A-B). Both methods enriched for potential true positives (PTP: known Nanog interactors and other chromatin-binders) while depleting for potential false positives (PFP: ribosomal proteins and other cytoplasmic proteins). ChIP-SICAP gives the largest relative difference between the positive and negative controls (PTP:PFP ratio of 28:1), ChIP-SPACE specificity is still much higher than ChIP-MS (7:1 compared with 2:1; Supplementary Fig. 4B). Moreover, ChIP-SPACE has the advantage over ChIP-SICAP that it doesn't need DNA end-labelling by biotinylated nucleotides, and streptavidin purification of chromatin. Thus ChIP-SPACE is significantly more convenient for studies with many samples to be processed simultaneously such as chromatin dynamics across multiple time points.

To understand the protein composition bound to the cis-regulatory elements (CREs) during the transition from the primed and ground-state pluripotency conditions, we applied ChIP-SPACE to study Nanog-bound CREs upon switching the medium from serum to 2iL (0h,

12h, 24h and 48h). We immunoprecipitated Nanog followed by H3K27ac to enrich for Nanog-bound CREs. We chose H3K27ac as a marker for active enhancers[16]. Finally, we used SPACE to purify chromatin binders (Fig. 4A, Supplementary Table 3). Like the global view of chromatin-associated proteins, among the 720 enriched proteins are many RBPs and RNA splicing factors (Fig. 4B).

We then identified 48 proteins involved in pluripotency and blastocyst development and clustered them according to their abundances (iBAQ intensities) across the time points. We observed 4 clusters (Fig. 4C): proteins that are present and equally abundant from the beginning to the end (Cluster 1), proteins that are similarly upregulated following the transition to the 2iL medium between 12h-48h period (Cluster 2), proteins that are upregulated at 12-24h, but then decline at 48h (Cluster 3), and proteins with a fluctuating pattern of binding (Cluster 4). Nanog and Esrrb belong to the cluster 1: proteins are consistently abundant throughout the time course, indicating their essential roles in keeping the CREs active. In contrast, most other proteins are present at enhancers from 12h after the transition to 2iL onwards, including the master pluripotency regulators Oct4 and Sox2, 2iL-associated factors such as Klf4 and Tfcp2l1 (Cluster 2), serum-associated factors such as Lin28a and Zfp281, and even suppressors of enhancer activity such as Mtf2, Rnf2 and Dnmt1 (Cluster 3); this means that positive regulators remain on the CREs (Clusters 1 and 2), while negative regulators and serum-associated factors decline and abandon the CREs around 24-48 hours (Cluster 3). Altogether, our results indicate that upon the transition towards naive pluripotency, both activators and inhibitors initially occupy the Nanog-regulated CREs, but until 48h the inhibitors decline and activators take over to consolidate the activity of Nanog-regulated enhancers.

VCP mutations globally affect chromatin composition in neural precursor cells

In our final experimental set-up, we tested SPACE on a biologically challenging system, using <2 million cells. VCP or p97 is a hexameric protein that is conserved across all eukaryotes. VCP is a member of the type II AAA+ ATPase (ATPases Associated with diverse cellular Activities) family of proteins, involved in multiple cellular processes, including protein degradation, intracellular trafficking, DNA repair and replication, and the regulation of cell cycle[17]. Moreover, multiple proteomic methods have reported it to bind RNA[9, 10] and chromatin, with multiple chromatin-related functions[18]. Mutations in VCP cause ALS (amyotrophic lateral sclerosis)[19, 20]; however the disease-causing mechanism is unknown. We compared chromatin compositions in neural differentiated (day 14) human induced pluripotent stem cells (hiPSCs) between three cell lines (M1, M2 and M3) with VCP mutations (one R191Q and two R155C; hereafter VCPmut) and control lines from three healthy donors (C1, C2 and C3; Fig. 5A). hiPSCs were differentiated into ventral spinal motor neuron precursors (NPs[21]). We identified 1,639 proteins in total, with 1,540 proteins quantified in at least four samples (Supplementary Table 4); of these, 89 proteins are already annotated for involvement in neurodegenerative diseases including TDP-43, MATR3, SFPQ, FUS, C9orf72, SOD1 and VCP (Supplementary Table 4 and Supplementary Fig. 5A). Among the quantified proteins, 40 proteins were statistically significantly more abundant in control lines, and 44 were higher in VCPmut samples (moderated t-test adj. p-value <0.1, fold-change > 1.5; Fig. 5B). The control and VCPmut cells could be clearly separated by PCA analysis (Supplementary Fig. 5B). Notably, VCP is higher in controls, indicating that VCP mutations decrease the capacity of VCP to bind chromatin. Conversely, Plastin-3 was higher in VCPmut cells: previous studies[22, 23] proposed plausible neuroprotective roles of Plastin-3 against SMA (Spinal Muscular Atrophy), suggesting that VCPmut cells might upregulate Plastin-3 in order to withstand neurodegenerative effects.

We also observe decreased chromatin-association of many proteins related to DNA damage response in VCPmut cells, including TP53BP1. TP53BP1 promotes non-homologous end joining repair of damage sites[24], and its decrease in VCPmut cells agrees with the finding that VCP directly promotes the recruitment of TP53BP1 to the DNA double-strand breaks[25]. SMC1A (cohesion complex) and MED1 (RNA polymerase II mediator complex) are also decreased in VCPmut cells, in agreement with their direct interactions with TP53BP1 and roles in DNA repair[26, 27](Fig. 5C). Thus, our data suggest that the decreased capacity of mutant VCP to bind chromatin decreases the chromatin binding of TP53BP1 and associated proteins involved in DNA repair.

Among the up- or down-regulated chromatin-associated proteins are several RBPs involved in RNA-processing and splicing, nonsense-mediated decay (NMD), RNA transport, transcriptional control and neuronal processes (Fig. 5C and Supplementary Fig. 5D). Remarkably, UPF1 is significantly higher in VCPmut cells. UPF1 is an RNA helicase that remodels RNPs[28], contributes to release of mRNA from DNA, shuttles between nucleus and cytoplasm[29], and plays key roles in NMD[30]. Interestingly, UPF1 also has neuroprotective effects against the accumulation of mutant RBPs, such as TDP43 and FUS, and improves survival in a neuronal model of ALS[31, 32]. Thus, the observed upregulation of RNA transport and NMD components on chromatin may be part of a rescue mechanism in VCPmut cells. Altogether, chromatin composition analysis sheds light on how mutations in VCP might make the cells vulnerable to DNA damage, and reveals changes in recruitment of many RBPs. This opens the opportunity to understand if these RBPs contribute to the vulnerability of cells to degeneration, or if they might, along with UPF1, help to protect them.

Discussion

This study presents SPACE, a robust method for purification of chromatin-associated proteins by silica magnetic beads for proteomic analysis. Recent ChIP-seq studies revealed dozens of RBPs that can bind to chromatin[33], and strikingly, our study now detects 63% of the nuclear RBPs known to be associated with chromatin. Besides, by combining SPACE with SICAP, we developed a high-throughput approach to identify direct DNA-binders, we showed that many RBPs, specifically mRNA splicing factors, are capable of directly binding to DNA. Our results are in line with co-transcriptional RNA processing concept[34], and indicate that proteins binding both to RNA and DNA may play a role in physically integrating transcription and RNA processing. The straightforward and cost-effective nature of SPACE makes it well-suited to study chromatin remodelling, which we evaluated in three experiments: First, we compared the global chromatin composition in 2iL and serum conditions of mESCs, which showed that the differential chromatin binding detected by SPACE upon altered pluripotency states was generally twice larger than the changes detected at the total protein level. For instance, transcription factors involved in pluripotency such as b-Catenin, Nanog, Cbfa2t2, Zfp42, Znf281, Zscan10 (12/15 proteins, Fig 3B) showed differential changes using SPACE (adj. p-value < 0.1 and fold-change >2), whereas full proteome changes were not significant (fold-change <2). Thus, SPACE provides a very sensitive method to detect chromatin composition changes in response to the pluripotency state.

One of the unexpected differential proteins was the RBP Dazl, which is highly abundant on chromatin specifically 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, which is necessary for germ cell survival[35, 36]. Moreover, cytoplasmic Dazl was reported to regulate Tet1 translation and hence DNA demethylation in 2i condition in mESCs[37]. Dazl

was previously reported to also localise in the nucleus[38], but its nuclear function has not yet been explored. Therefore we used ChIP-seq, which found that Dazl associates with the same chromatin sites as PRC2, including TSS of many developmental genes such as Hox genes, Wnt ligands, Wnt receptors and some pluripotency genes (Esrrb and Sox2). This is in line with the finding that RBPs often interact with enhancers, promoters and transcriptionally active regions on chromatin (31251911). Considering that we apply RNase treatment during the SPACE procedure, the co-location of Dazl and PRC2 on chromatin is not indirect via RNA. Nevertheless, it is likely that Dazl binds to and regulates the non-coding RNAs that can associate with the PRC2-bound chromatin sites, and we speculate that Dazl might influence the reported capacity of RNA to displace PRC2 from the chromatin[39].

In the second experiment, we use SPACE to assess local chromatin dynamics on the Nanog-binding CRMs during the transition from serum to 2iL condition. ChIP-seq analyses showed that roughly 50% of the binding sites of the many well-known TFs do not contain the cognate motifs[40]. Our results are in line with the cooperative model of TF binding to the hotspots or clusters that are built through a complex network of protein-protein, in addition to protein-DNA, interactions[41, 42]. Specifically, our data suggest that shortly after induction of naive pluripotency, proteins with both activating and inhibiting functions bind to the Nanog-regulated CRMs, and then the activators take over the CRMs 48h after the induction, indicating how the CRMs are remodelled during this transition.

Finally, to explore the capacity of SPACE to provide insight into disease-related chromatin remodelling, we examined the impact of VCP mutations on NPs. We observed a significant decrease in chromatin abundance of mutant VCP, as well as several proteins involved in DNA repair, such as TP53BP1, which relates to the well-recognised role of DNA damage in neurodegeneration[43-45]. Moreover, VCP contributes to genome stability, and mutant VCP neurons are highly sensitive to DNA damage-induced transcriptional stress[46, 47]. Interestingly,

VCPmut cells have increased chromatin abundance of RBPs involved in RNA clearance from chromatin, such as NMD and RNA export from the nucleus (Fig. 5D). It will be interesting to test whether the presence of these RBPs affects the role that nascent RNA plays as a competitor for many chromatin-binding proteins[39], and whether their effect decreases or increases the vulnerabilities of mutant cells. All in all, our study demonstrated the capability of SPACE for unravelling the chromatin remodelling that underlies the dynamics of gene regulation and genome maintenance in development and disease.

Acknowledgement

We would like to thank Jeroen Krijgsveld for revising the manuscript, and the proteomics platform at the Francis Crick Institute, Bram Snyder, Steve Howel and Helen Flynn for sharing their technical expertise.

M.R. 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), the UK Medical Research Council (FC010110), and the Wellcome Trust (FC010110). NML is a Winton Group Leader in recognition of the Winton Charitable Foundation's support towards the establishment of the Francis Crick Institute. NML is 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

M.R., JU and NML designed the research. M.R, JAZ and GT performed the experiments. M.R. analyzed the data. M.R. and NML wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests

Figure legends

Fig. 1: Overview of the SPACE and SPACE-SICAP procedures.
1: Cells are crosslinked by 1% formaldehyde, resuspended in the lysis buffer containing guanidinium, and iso-propanol and silica magnetic beads are added to the lysate. 2: Chromatin binds to the magnetic beads, and is separated from the lysate. 3: The beads are washed with lysis buffer and ethanol, and are treated with RNase A. 4a: In SPACE, the beads are washed again with ethanol and Acetonitrile, and Trypsin/LysC are added to digest the chromatin-associated proteins on the beads. 4b: In SPACE-SICAP, chromatin is eluted from the silica magnetic beads DNA is end-labelled by TdT and biotin-dCTP, andt chromatin is captured by protease-resistant streptavidin beads. 5: Chromatin is stringently washed, and chromatin-associated proteins are digested by LysC. 6: The supernatant is collected, and digested with Trypsin to generate peptides for chromatin and DNA-binders (SPACE-SICAP I). The streptavidin beads are washed and heated to reverse the crosslinking, and the released peptides are digested by Trypsin to generate peptides of DNA-binders (SPACE-SICAP II,). The peptides are cleaned and injected to the mass spectrometer.

Fig. 2: Chromatin composition in mESC.

(A) Proteins identified by SPACE procedure in comparison to the non-crosslinked control. The red line shows the threshold for the enriched proteins. CL: crosslinked by formaldehyde, nCL: the non-crosslinked control. The proteins were categorized into 4 groups: 1-'DNA-binding proteins' (DB, dark green), 2: 'Chromatin- or chromosome-binding proteins' (CB) that are not DB (green). 3: Remaining proteins 'present in nucleus' (PN) are those that are not DB or CB (pale green). 4: Proteins that do not fall into the previous categories are so-called 'unexpected' (UE, yellow). (B) Comparing the enriched proteins by SPACE with the full proteome of the cells. The percentiles in the left three bars were calculated based on the number of the proteins. The percentiles in the right three bars were calculated based on the protein abundances. The PN sections of the SPACE results (marked by a star) were shown separately to depict the frequency of RBPs and proteins involved in RNA splicing (purple sections). (C) Proteins identified by SPACE-SICAP I procedure. (D) The bars show the percentiles of the enriched proteins by SPACE-SICAP based on the number and the abundance of the enriched proteins. (E) Overlap of the enriched proteins among SPACE, SPACE-SICAP and the DNase-treated negative control (The Venn-diagram). The PN section of the intersect (marked by a star) was shown as a bar (bottom) to depict the frequency of the RBPs and proteins involved in RNA splicing(F) Proteins identified by SPACE-SICAP II procedure to obtain DNA-binders (please refer to Fig. 1). (G) The enriched proteins from that are not known as DNA-binding proteins were depicted by green rectangles (CB), pale green diamonds (PN). RBPs were marked by purple borders. The edges show physical protein-protein interactions.

Fig. 3: Chromatin composition in 2iL and serum states of mESC.

(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 fold-change>2). The rest of the proteins were depicted by grey. Proteins involved in pluripotency, mESC self-renewal or differentiation were marked by black dots. (B) Comparing 2iL/serum ratios in full proteome analysis with SPACE results. (C) Physical interaction network of the proteins involved in pluripotency or blastocyst development. RBPs were marked by purple borders. (D) The proteins in panel C were clustered based on their 2iL/serum ratios and abundances (iBAQ). (E) Annotation of Dazl ChIP-seq peaks (idr < 0.01). (F) Profile of Dazl peaks on the genome in comparison with Suz12, Aebp2 and H3K27me3 peaks in mESC. The last ChIP profiles of Suz12, Aebp2 and H3K27me3 were obtained from [48].

Fig. 4: Dynamics of proteins associated with Nanog-binding CREs.

(A) Overview of the experiment. (B) Frequency of the enriched proteins based on their abundances (left) and their numbers (middle). The PN section (marked by a star) was separated to show the frequency of the RBPs and RNA splicing factors (purple sections, right). (C) The clusters show the profiles of protein abundances throughout the time points. (D) Proposed model based on the pattern of binding events to the Nanog CREs. In serum condition, the CREs are bound by constitutive binders such as Esrrb and Nanog (Cluster 1). After changing the medium to 2iL, the CREs gain transient states (12-24h) by recruiting many proteins to the CREs. Some of them are activators such as Oct4, Sox2, Klf4 and Smarca4 (Brg1) (Cluster 2), and some others are inhibitors such as Rnf2 (PRC1), or competitors such as Lin28a and Znf281 (Cluster 3). Some activators fluctuate such as Prdm14 and Cbfa2t2 (Cluster 4). The inhibitors and other competitors abandon the CREs by 48 h, and activators remain on the CREs.

Fig. 5: Chromatin composition in VCPmut and control NPs.

(A) Overview of the experiment. M1 and M2 contain R191Q, and M3 contains R155C mutations of VCP. (B) Differential proteins in control and VCPmut NPs were shown by green and red, respectively (moderated t-test adj. p-value < 0.1 and log2(fold-change) > 0.5). Proteins with orange cross and circle are involved in neurodegenerative diseases. (C) Differential proteins involved in neuronal process, DNA damage response, RNA processing, Nonsense-mediated decay, RNA transport and transcriptional control were shown. RBPs were marked by purple borders. (D) Proposed model based on the differentially bound proteins to chromatin. In VCPmut cells, VCP shows a decrease in binding to chromatin. As a result, other members of DNA damage response such as TP53BP1 are not recruited to the DNA damage sites efficiently. Over time mutations are accumulated in the NPs which might lead to transcription stress. As a result, UPF1 is recruited to chromatin more abundantly to trigger RNA clearance mechanisms such as NMD and RNA export.

Supplementary Figure legends

Supplementary Fig. 1: Identification of direct DNA binding proteins by SPACE-SICAP procedure.

After capturing chromatin by protease-resistant streptavidin, chromatin-associated proteins are digested by LysC. Majority of the proteins are detached (red fragments), however, the crosslinked parts of the direct DNA-binders remain crosslinked to DNA (orange fragments). After reversing the crosslinking, DNA-binding fragments are digested by Trypsin to be identified by mass spectrometry (left route). Digesting supernatant of the initial digestion provides both DNA and chromatin-binders (right route).

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

The bars show the frequencies of the RBPs based on the protein numbers (A) and protein abundances (B) in proteins enriched by SPACE. The enriched proteins were categorized into 4 groups: 1:DB = DNA-binding proteins, 2:CB = chromosome- and chromatin-binding proteins that are not DB. 3:PN= proteins present in nucleus that are not DB or CB. 4: UE = the so-called unexpected proteins that do not fall into the previous categories. (C) Top enriched GO terms in each category. (D) Peptides identified using SPACE-SICAP II were mapped to the protein domains. The plot shows the number of peptides that fall into the protein domains or they are < 50 a.a. far from the domain. (E) Map of the protein domains in RBPs enriched by SPACE-SICAP II and position of the identified peptides.

Supplementary Fig. 3: Comparing 2iL and serum states of mESC using SPACE.

(A) Comparing proteins identified by SPACE with the full proteome of the cells. The proteins were categorized into 4 groups: 1:DB = DNA-binding proteins (dark green), 2:CB =

chromosome- and chromatin-binding proteins that are not DB, (green). 3:PN= proteins present in the nucleus or that are not DB or CB (pale green). 4: Unexpected = proteins that do not fall into the previous categories (yellow). (B) Proteins involved in pluripotency that were only identified by SPACE. (C) The boxplot shows the distribution of 2iL/serum ratios obtained by full proteome analysis and SPACE. 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. (D) Frequency of the identified proteins by comparative SPACE, calculated based on their numbers (left) and their abundances (right). (E) The volcano plot shows RBPs with differential binding ratios to chromatin in 2i and serum conditions. (F) Proteins identified in a Dazl IP-MS with formaldehyde crosslinking. Proteins were sorted based on their intensities. Dazl is the most abundant protein as the target of the IP. (G) KEGG pathway analysis of the genes with a Dazl ChIP-seq peak. (H) Distribution of Dazl iCLIP peaks across RNA regions. (I) Intersect of Dazl ChIP-seq and iCLIP-seq peaks at the gene level.

Supplementary Fig. 4: local view of chromatin-associated proteins.
(A) Overview of ChIP-SPACE procedure. (B) Comparing potential false positives (PFP) and potential true positives (PTP) in Nanog ChIP-MS, ChIP-SICAP and SPACE using relative abundance of the enriched proteins. (C) The enriched proteins involved in pluripotency, differentiation or blastocyst development were clustered based on their abundance profiles during the timepoints.

Supplementary Fig. 5: Comparing chromatin composition in VCPmut and control NPs.
(A) Disease Ontology (DO) analysis of the proteins identified by SPACE. The network shows enriched diseases (brown nodes) based on the identified proteins. The colour of the proteins shows their VCPmut over control log2 ratios. (B) Dimensionality reduction of the

variations in the SPACE results by PCA analysis. (C) Comparing differential chromatin-associated proteins in VCPmut and control cells. (D) The box plots compare the intensity of the proteins in VCPmut and control cells. The box covers the inter-quantile region (IQR), the whiskers show 1.5x IQR, the middle line is the median, and the outliers are shown by black points. The adj. p-values were calculated by moderated t-tests

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. After filtering out Reverse, potential contaminants and proteins only identified by site, protein groups were quantified by their SILAC or iBAQ intensities. In the experiments related to mESC chromatin composition (Fig 2), if a SILAC ratio of CL/nCL was NA then iBAQ intensities were used to calculate the ratios. Differential proteins were determined by the limma R package. Proteins with log2-fold change > 1 and adj. p-value < 0.1 were considered as “Enriched ++”, and proteins with log2-fold change > 1 were considered as “Enriched +”. Uniprot and DAVID were used as the source of Gene Ontology and protein domain information. To categorize the proteins, the DNA-binders were labeled as DB, then chromatin and chromosome binders were labeled as CB. This means if a protein is a DNA-binder and chromatin-binder it is labeled as DB. Then proteins present in the nucleus were labeled as PN, then the rest of the proteins were labeled as “unexpected” or UE. In the specific case of SPACE-SICAP to determine direct DNA-binders we filtered proteins > 20 kD to remove small proteins. The Clusterprofiler R package was used for Gene enrichment analysis. In the experiments related to chromatin composition in 2iL and serum states of mESC (Fig 3), physical protein-protein interaction data was imported from STRING database (11.0), and visualized by Cytoscape (3.8.0). The proteins were clustered by k-means clustering algorithm. The protein clusters were visualized by the Factoextra R package. In the experiments related to the Dynamics of proteins associated with Nanog-binding CREs (Fig 4), the iBAQ intensities were quantile normalized by the preprocessCore package, and the mean of two replicates per time-point was used as the average protein intensity to cluster the proteins by k-means clustering algorithm. In the experiments related to the chromatin

composition in VCPmut and control NPs (Fig 5), the iBAQ intensities were quantile normalized by the preprocessCore package, and differential proteins were determined by limma package. Protein information was downloaded from Uniprot and DAVID Gene Ontology database. The DOSE R package was used for disease ontology.

Dazl ChIP-seq experiment and data analysis

The ChIP procedure and analysis were carried out essentially as described previously[11]. 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[49]. 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.

ChIP-SICAP

Nanog IP was carried out using CST antibody #8822. ChIP-SICAP was carried out essentially as described previously[11, 12]. The step by step protocol of ChIP-SICAP is available in the following database:
dx.doi.org/10.17504/protocols.io.bcrriv56

Cell culture and generation of precursor motor neurons from hiPSC

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.

Human iPSCs were maintained using Geltrex (Life Technologies) with Essential 8 Medium media (Life Technologies) and passaged using EDTA (Life Technologies, 0.5mM). Three of the control lines used are commercially available and were purchased from Coriell (cat. number ND41866*C), ThermoFisher Scientific (cat. number A18945) and Cedars Sinai (Cat.number CS02iCTR-NTn4) respectively. Two of the iPSC lines with VCP mutation contained R155C mutation, and one iPS line contained R191Q mutation. Motor neuron differentiation was carried out as described previously[21]. Briefly, human iPSCs were first differentiated to neuroepithelium by plating to 100% confluency in chemically defined medium (DMEM/F12 Glutamax, Neurobasal, L-Glutamine, N2 supplement, non-essential amino acids, B27 supplement, β-mercaptoethanol (all from Life Technologies) and insulin (Sigma)). Treatment with small molecules from day 0-7 was as follows: 1µM Dorsomorphin (Millipore), 2µM SB431542 (Tocris Bioscience), and 3.3µM CHIR99021 (Miltenyi Biotec). At day 8, cells patterned for 7 days with 0.5µM retinoic acid and 1µM Purmorphamine. On day 14 spinal cord MN precursors were harvested for experiments. Throughout the neural conversion and patterning phase (D0-18) the neuroepithelial layer was enzymatically dissociated twice (at D4-5 and D10-12) using dispase (GIBCO, 1 mg ml^{-1}).

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD020037. The accession numbers for the Dazl ChIP-seq and iCLIP reported in this paper are ArrayExpress: E-MTAB-9302 and E-MTAB-9332, respectively.

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, G3272-500G)
- 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 should 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 inputs.
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.

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[12]

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.

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: if you wish to keep protein-RNA complexes on chromatin skip this step.
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.

Reference

1. Boukouris, A.E., Zervopoulos, S.D. & Michelakis, E.D. Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription. *Trends Biochem Sci* **41**, 712-730 (2016).
2. Ginno, P.A., Burger, L., Seebacher, J., Iesmantavicius, V. & Schubeler, D. Cell cycle-resolved chromatin proteomics reveals the extent of mitotic preservation of the genomic regulatory landscape. *Nat Commun* **9**, 4048 (2018).
3. van Mierlo, G. et al. Integrative Proteomic Profiling Reveals PRC2-Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency. *Cell Stem Cell* **24**, 123-137 e128 (2019).
4. Aranda, S. et al. Chromatin capture links the metabolic enzyme AHCY to stem cell proliferation. *Sci Adv* **5**, eaav2448 (2019).
5. Kliszczak, A.E., Rainey, M.D., Harhen, B., Boisvert, F.M. & Santocanale, C. DNA mediated chromatin pull-down for the study of chromatin replication. *Sci Rep* **1**, 95 (2011).
6. Sirbu, B.M. et al. Identification of proteins at active, stalled, and collapsed replication forks using isolation of proteins on nascent DNA (iPOND) coupled with mass spectrometry. *J Biol Chem* **288**, 31458-31467 (2013).
7. Alabert, C. et al. Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nat Cell Biol* **16**, 281-293 (2014).
8. Kohlmeier, F., Maya-Mendoza, A. & Jackson, D.A. EdU induces DNA damage response and cell death in mESC in culture. *Chromosome Res* **21**, 87-100 (2013).
9. Baltz, A.G. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46**, 674-690 (2012).
10. Queiroz, R.M.L. et al. Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat Biotechnol* **37**, 169-178 (2019).
11. Rafiee, M.R., Girardot, C., Sigismondo, G. & Krijgsveld, J. Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins. *Mol Cell* **64**, 624-635 (2016).
12. Rafiee, M.R. et al. Protease-resistant streptavidin for interaction proteomics. *Mol Syst Biol* **16**, e9370 (2020).
13. Castello, A. et al. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* **63**, 696-710 (2016).
14. Kinoshita, M. & Smith, A. Pluripotency Deconstructed. *Dev Growth Differ* **60**, 44-52 (2018).
15. Konig, J. et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**, 909-915 (2010).
16. Creyghton, M.P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936 (2010).
17. Meyer, H. & Weihl, C.C. The VCP/p97 system at a glance: connecting cellular function to disease pathogenesis. *J Cell Sci* **127**, 3877-3883 (2014).
18. Dantuma, N.P., Acs, K. & Luijsterburg, M.S. Should I stay or should I go: VCP/p97-mediated chromatin extraction in the DNA damage response. *Exp Cell Res* **329**, 9-17 (2014).
19. Johnson, J.O. et al. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68**, 857-864 (2010).
20. Taylor, J.P. Multisystem proteinopathy: intersecting genetics in muscle, bone, and brain degeneration. *Neurology* **85**, 658-660 (2015).

21. Hall, C.E. et al. Progressive Motor Neuron Pathology and the Role of Astrocytes in a Human Stem Cell Model of VCP-Related ALS. *Cell Rep* **19**, 1739-1749 (2017).
22. Alrafiah, A. et al. Plastin 3 Promotes Motor Neuron Axonal Growth and Extends Survival in a Mouse Model of Spinal Muscular Atrophy. *Mol Ther Methods Clin Dev* **9**, 81-89 (2018).
23. Hosseinibarkooie, S. et al. The Power of Human Protective Modifiers: PLS3 and CORO1C Unravel Impaired Endocytosis in Spinal Muscular Atrophy and Rescue SMA Phenotype. *Am J Hum Genet* **99**, 647-665 (2016).
24. Mirman, Z. & de Lange, T. 53BP1: a DSB escort. *Genes Dev* **34**, 7-23 (2020).
25. Acs, K. et al. The AAA-ATPase VCP/p97 promotes 53BP1 recruitment by removing L3MBTL1 from DNA double-strand breaks. *Nat Struct Mol Biol* **18**, 1345-1350 (2011).
26. Litwin, I., Pilarczyk, E. & Wysocki, R. The Emerging Role of Cohesin in the DNA Damage Response. *Genes (Basel)* **9** (2018).
27. Parsons, B.L. MED1: a central molecule for maintenance of genome integrity and response to DNA damage. *Proc Natl Acad Sci U S A* **100**, 14601-14602 (2003).
28. Fiorini, F., Bagchi, D., Le Hir, H. & Croquette, V. Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. *Nat Commun* **6**, 7581 (2015).
29. Singh, A.K. et al. The RNA helicase UPF1 associates with mRNAs co-transcriptionally and is required for the release of mRNAs from gene loci. *Elife* **8** (2019).
30. Dehecq, M. et al. Nonsense-mediated mRNA decay involves two distinct Upf1-bound complexes. *EMBO J* **37** (2018).
31. Barmada, S.J. et al. Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proc Natl Acad Sci U S A* **112**, 7821-7826 (2015).
32. Ortega, J.A. et al. Nucleocytoplasmic Proteomic Analysis Uncovers eRF1 and Nonsense-Mediated Decay as Modifiers of ALS/FTD C9orf72 Toxicity. *Neuron* **106**, 90-107 e113 (2020).
33. Xiao, R. et al. Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell* **178**, 107-121 e118 (2019).
34. Herzel, L., Ottoz, D.S.M., Alpert, T. & Neugebauer, K.M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**, 637-650 (2017).
35. Zagore, L.L. et al. DAZL Regulates Germ Cell Survival through a Network of PolyA-Proximal mRNA Interactions. *Cell Rep* **25**, 1225-1240 e1226 (2018).
36. Li, H. et al. DAZL is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* **6**, 455-468 (2019).
37. Welling, M. et al. DAZL regulates Tet1 translation in murine embryonic stem cells. *EMBO Rep* **16**, 791-802 (2015).
38. Moore, F.L., Jaruzelska, J., Dorfman, D.M. & Reijo-Pera, R.A. Identification of a novel gene, DZIP (DAZ-interacting protein), that encodes a protein that interacts with DAZ (deleted in azoospermia) and is expressed in embryonic stem cells and germ cells. *Genomics* **83**, 834-843 (2004).
39. Skalska, L., Beltran-Nebot, M., Ule, J. & Jenner, R.G. Regulatory feedback from nascent RNA to chromatin and transcription. *Nat Rev Mol Cell Biol* **18**, 331-337 (2017).
40. Liu, Z. et al. Enhancer activation requires trans-recruitment of a mega transcription factor complex. *Cell* **159**, 358-373 (2014).

41. Siersbaek, R. et al. Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. *Cell Rep* **7**, 1443-1455 (2014).
42. Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol* **47**, 1-8 (2017).
43. Mitra, J. et al. Motor neuron disease-associated loss of nuclear TDP-43 is linked to DNA double-strand break repair defects. *Proc Natl Acad Sci U S A* **116**, 4696-4705 (2019).
44. Wang, H. et al. Mutant FUS causes DNA ligation defects to inhibit oxidative damage repair in Amyotrophic Lateral Sclerosis. *Nat Commun* **9**, 3683 (2018).
45. Stein, D. & Toiber, D. DNA damage and neurodegeneration: the unusual suspect. *Neural Regen Res* **12**, 1441-1442 (2017).
46. Vaz, B., Halder, S. & Ramadan, K. Role of p97/VCP (Cdc48) in genome stability. *Front Genet* **4**, 60 (2013).
47. Lans, H., Hoeijmakers, J.H.J., Vermeulen, W. & Marteijn, J.A. The DNA damage response to transcription stress. *Nat Rev Mol Cell Biol* **20**, 766-784 (2019).
48. Healy, E. et al. PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol Cell* **76**, 437-452 e436 (2019).
49. Blazquez, L. et al. Exon Junction Complex Shapes the Transcriptome by Repressing Recursive Splicing. *Mol Cell* **72**, 496-509 e499 (2018).

Figure 1

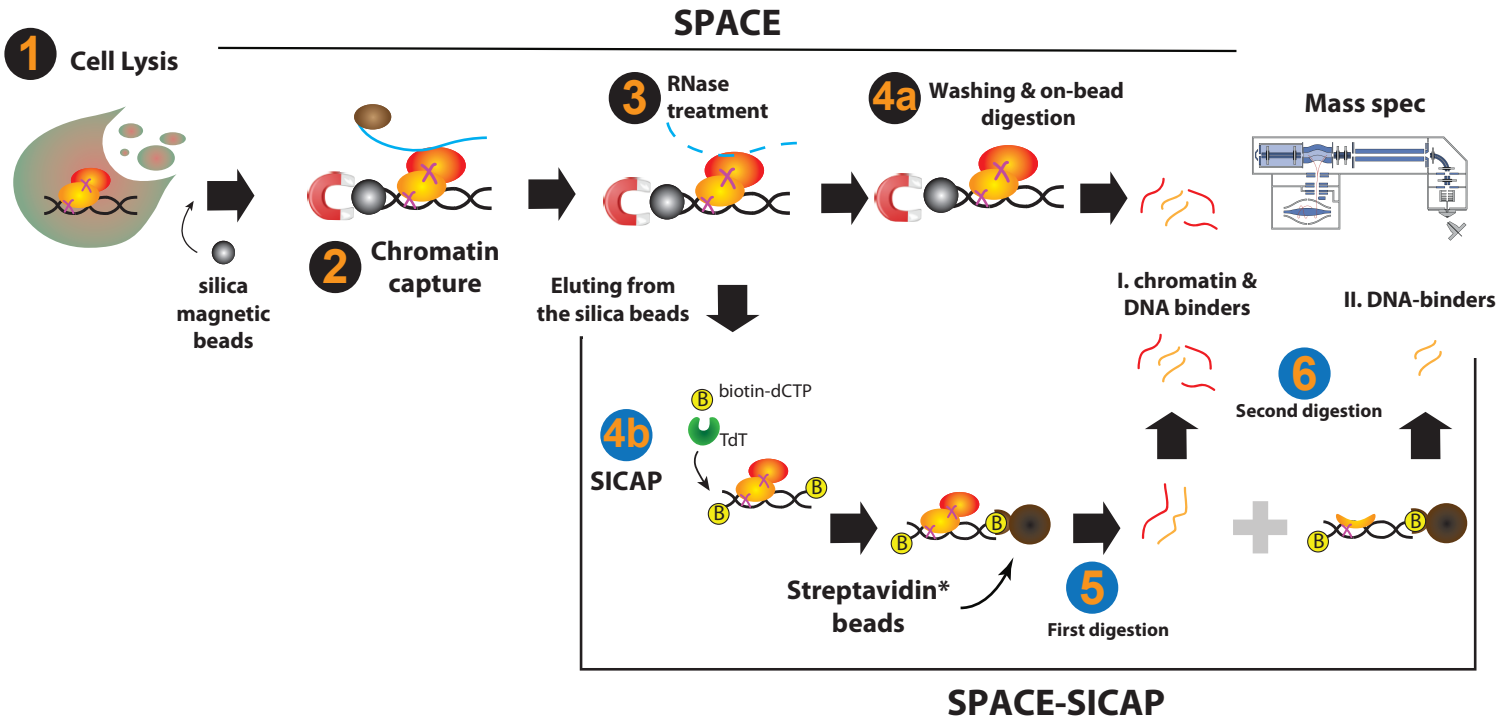

SPACE
1 Cell Lysis
silica magnetic beads
2 Chromatin capture
3 RNase treatment
4a Washing & on-bead digestion
Mass spec
Eluting from the silica beads
I. chromatin & DNA binders
II. DNA-binders
4b SICAP
biotin-dCTP
TdT
Streptavidin* beads
5 First digestion
6 Second digestion
SPACE-SICAP

Figure 2

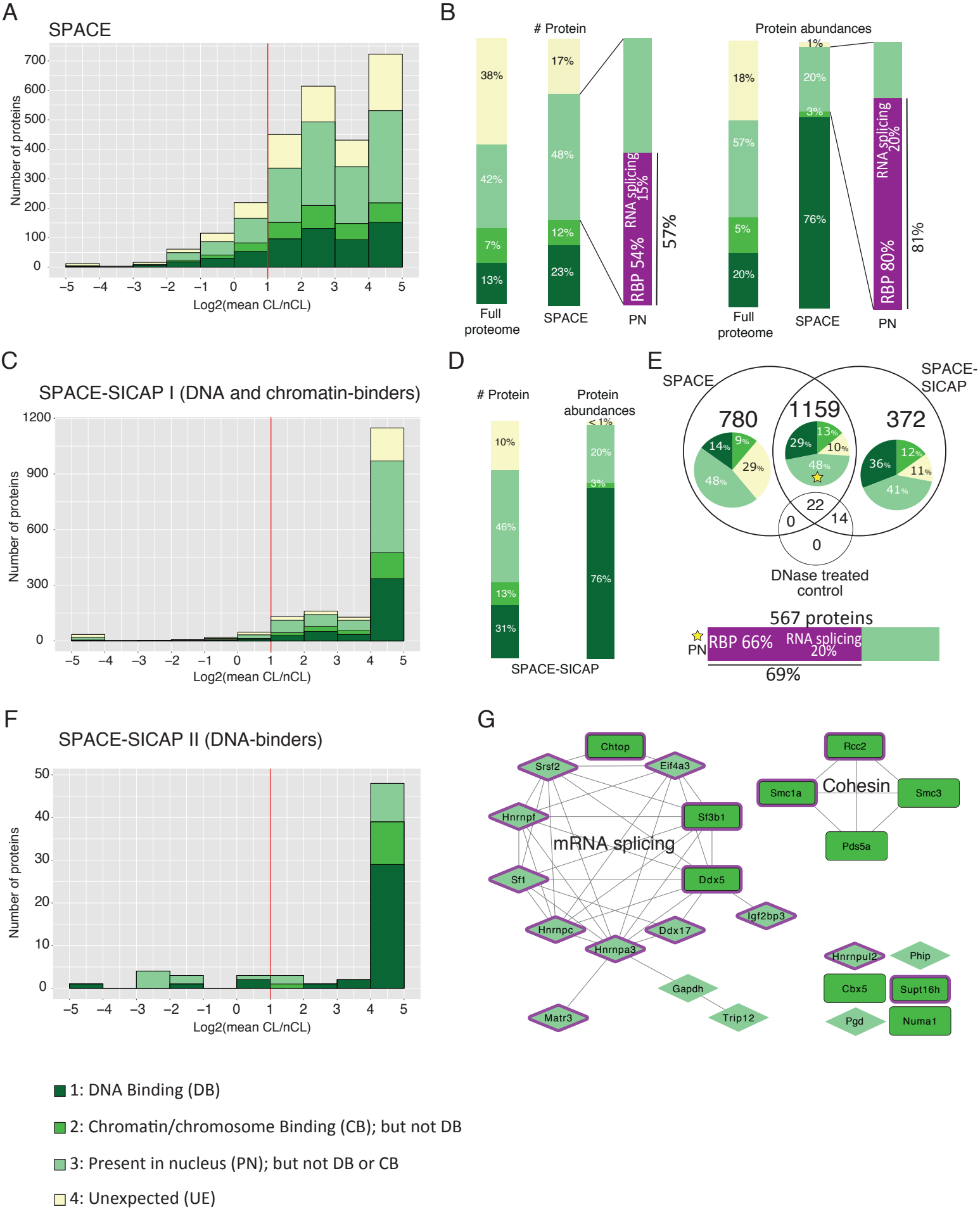

A
SPACE
Number of proteins
Log2(mean CL/nCL)
B
Protein
Protein abundances
38%
42%
7%
13%
17%
48%
12%
23%
RNA splicing 15%
RBP 54%
57%
18%
57%
5%
20%
1%
20%
3%
76%
RNA splicing 20%
RBP 80%
81%
Full proteome
SPACE
PN
C
SPACE-SICAP I (DNA and chromatin-binders)
Number of proteins
Log2(mean CL/nCL)
D
Protein
Protein abundances
10%
46%
13%
31%
< 1%
20%
3%
76%
SPACE-SICAP
E
SPACE
SPACE-SICAP
780
1159
372
22
0
14
0
DNase treated control
567 proteins
PN
RBP 66%
RNA splicing 20%
69%
F
SPACE-SICAP II (DNA-binders)
Number of proteins
Log2(mean CL/nCL)
G
Chtop
Srsf2
Eif4a3
Hnrnpf
Sf3b1
mRNA splicing
Sf1
Ddx5
Hnrnpc
Ddx17
Hnrnpa3
Igf2bp3
Gapdh
Matr3
Trip12
Rcc2
Smc1a
Cohesin
Smc3
Pds5a
Hnrnpul2
Phip
Cbx5
Supt16h
Pgd
Numa1
1: DNA Binding (DB)
2: Chromatin/chromosome Binding (CB); but not DB
3: Present in nucleus (PN); but not DB or CB
4: Unexpected (UE)

Figure 3

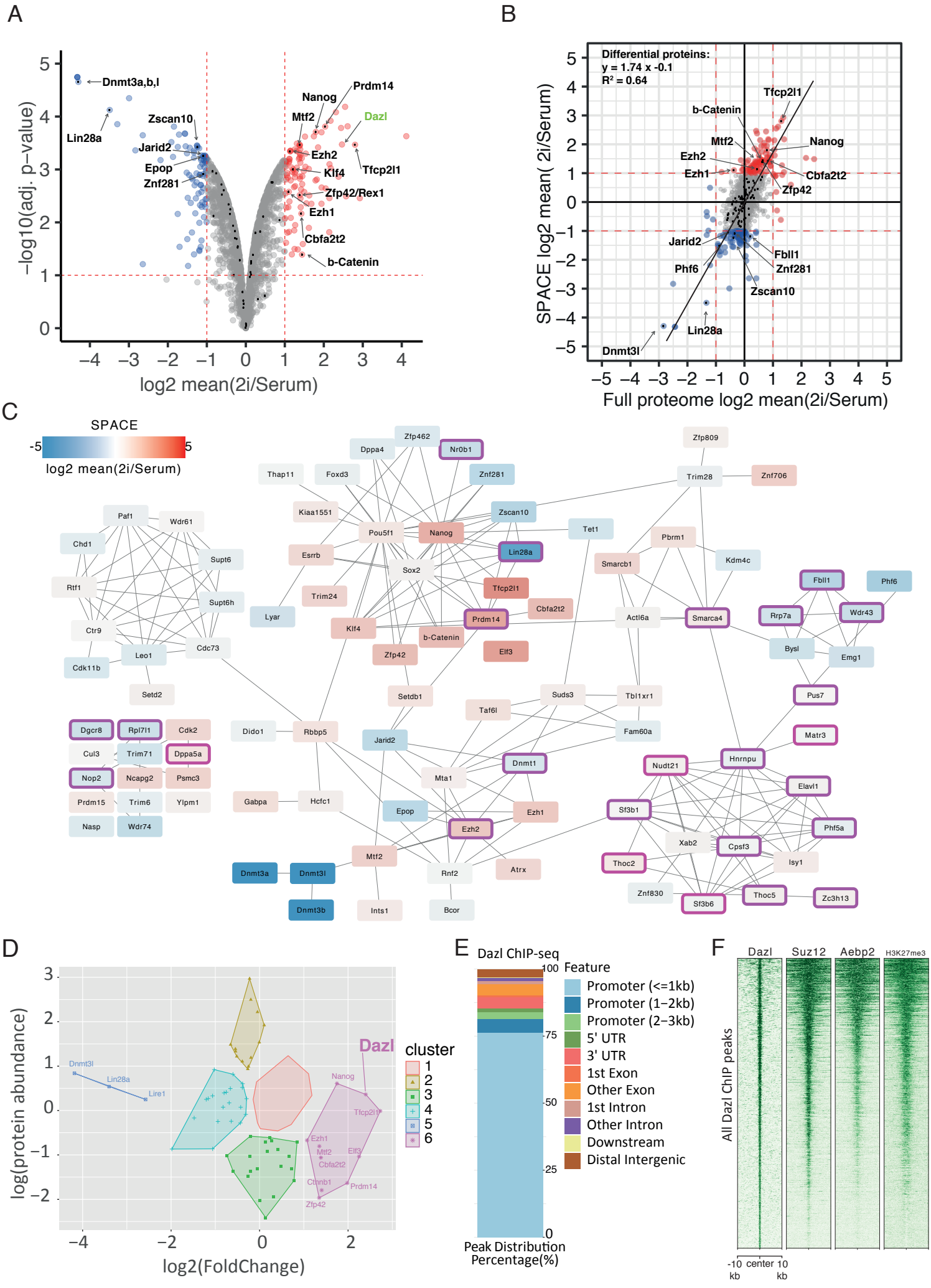

A
−log10(adj. p−value)
log2 mean(2i/Serum)
Dnmt3a,b,l
Lin28a
Zscan10
Jarid2
Epop
Znf281
Nanog
Prdm14
Mtf2
Dazl
Ezh2
Klf4
Tfcp2l1
Zfp42/Rex1
Ezh1
Cbfa2t2
b-Catenin
B
Differential proteins:
y = 1.74 x -0.1
R² = 0.64
SPACE log2 mean(2i/Serum)
Full proteome log2 mean(2i/Serum)
Tfcp2l1
b-Catenin
Mtf2
Nanog
Ezh2
Ezh1
Cbfa2t2
Zfp42
Jarid2
Fbll1
Znf281
Phf6
Zscan10
Lin28a
Dnmt3l
C
SPACE
log2 mean(2i/Serum)
D
log(protein abundance)
log2(FoldChange)
Dnmt3l
Lin28a
Lire1
Dazl
Nanog
Tfcp2l1
Ezh1
Mtf2
Elf3
Cbfa2t2
Ctnnb1
Prdm14
Zfp42
cluster
1
2
3
4
5
6
E
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
Peak Distribution Percentage(%)
F
Dazl
Suz12
Aebp2
H3K27me3
All Dazl ChIP peaks
-10 kb
center
10 kb

Figure 4

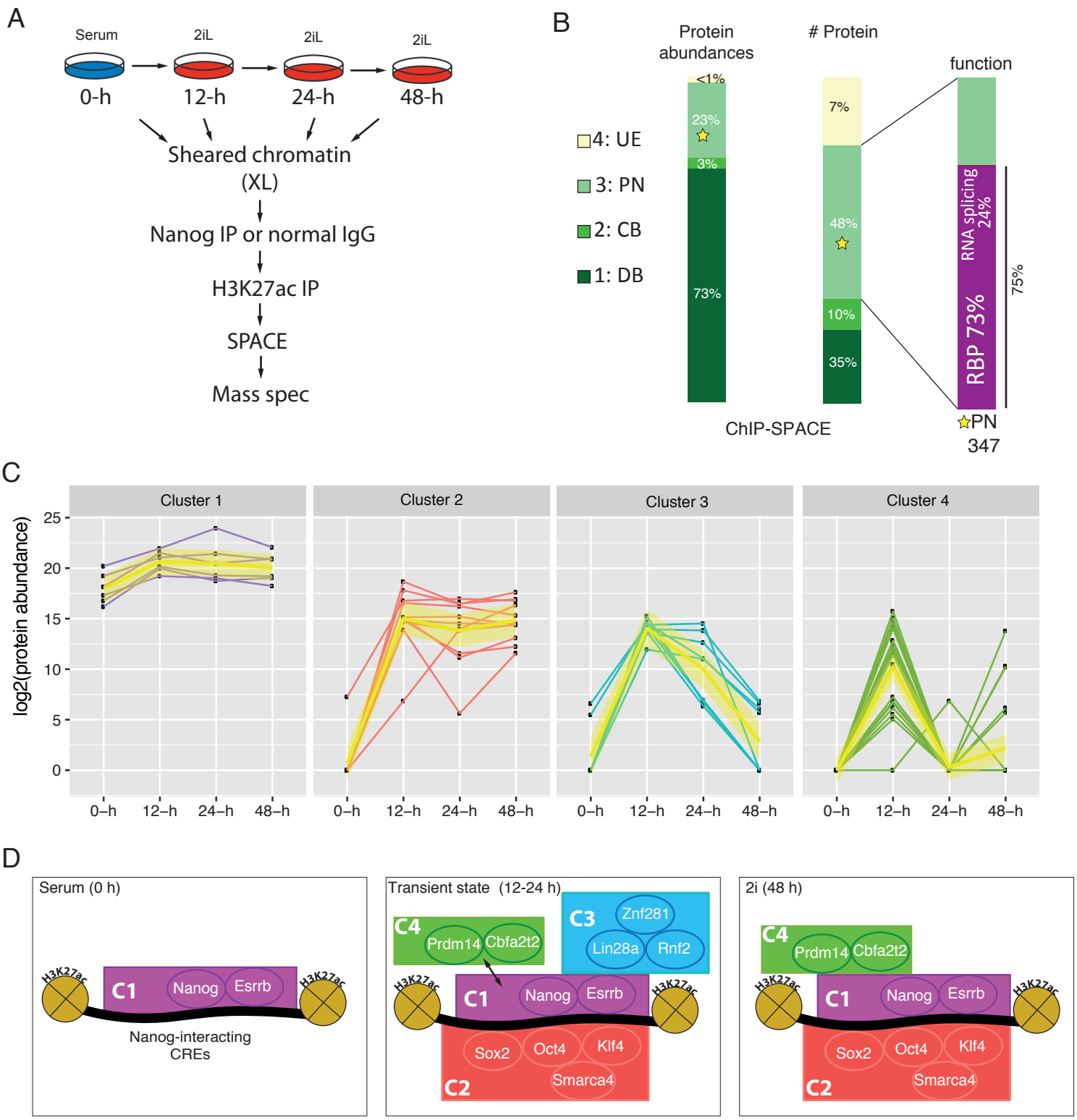
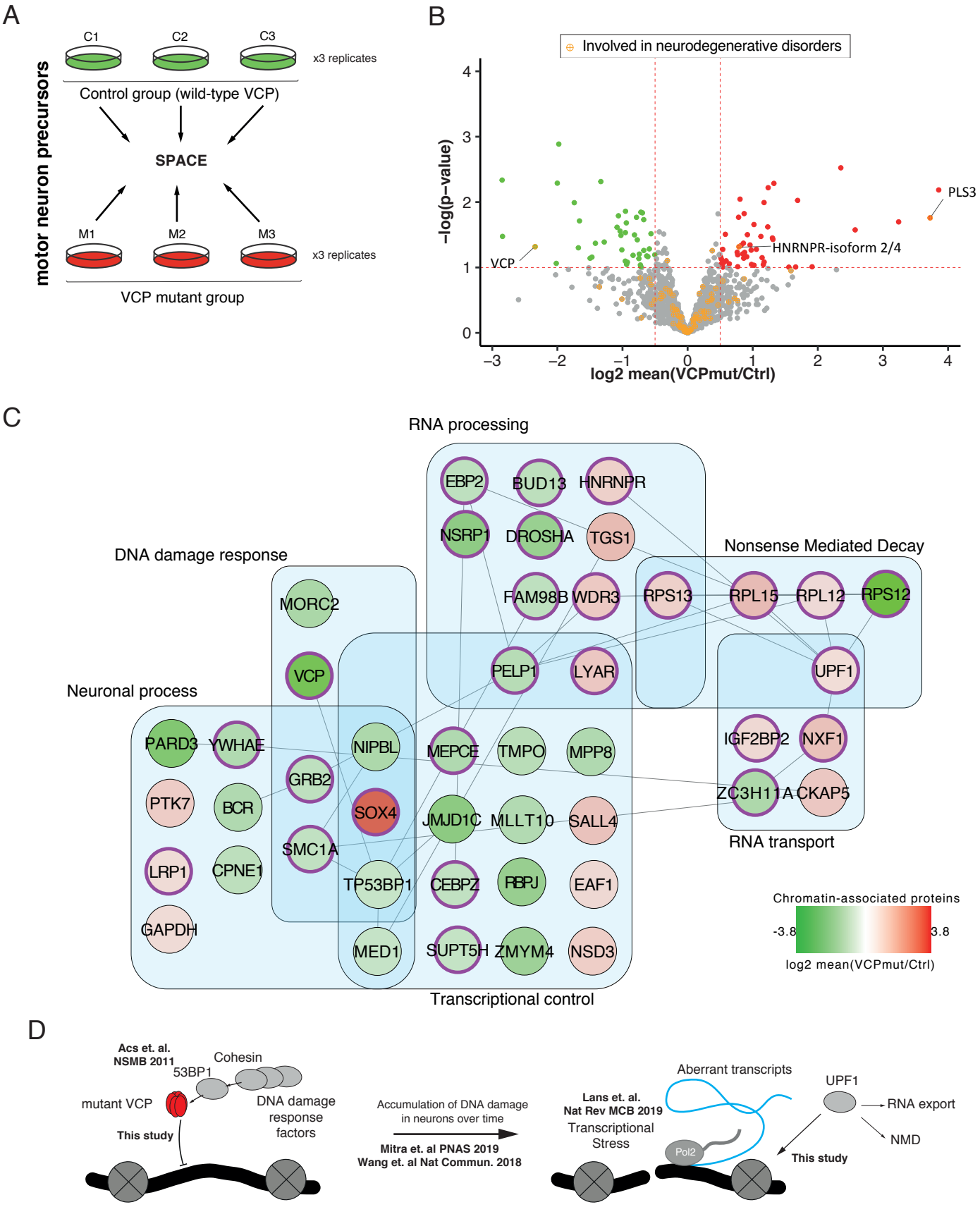

A
Serum
2iL
2iL
2iL
0-h
12-h
24-h
48-h
Sheared chromatin (XL)
Nanog IP or normal IgG
H3K27ac IP
SPACE
Mass spec
B
Protein abundances
Protein
function
4: UE
3: PN
2: CB
1: DB
<1%
23%
3%
73%
7%
48%
10%
35%
RNA splicing 24%
RBP 73%
75%
ChIP-SPACE
PN 347
C
Cluster 1
Cluster 2
Cluster 3
Cluster 4
log2(protein abundance)
0–h
12–h
24–h
48–h
D
Serum (0 h)
H3K27ac
C1
Nanog
Esrrb
Nanog-interacting CREs
Transient state (12-24 h)
C4
Prdm14
Cbfa2t2
C3
Znf281
Lin28a
Rnf2
Sox2
Oct4
Klf4
Smarca4
C2
2i (48 h)

Figure 5


A
motor neuron precursors
C1
C2
C3
x3 replicates
Control group (wild-type VCP)
SPACE
M1
M2
M3
x3 replicates
VCP mutant group
B
Involved in neurodegenerative disorders
−log(p−value)
log2 mean(VCPmut/Ctrl)
VCP
PLS3
HNRNPR-isoform 2/4
C
RNA processing
EBP2
BUD13
HNRNPR
NSRP1
DROSHA
TGS1
FAM98B
WDR3
PELP1
LYAR
DNA damage response
MORC2
VCP
Nonsense Mediated Decay
RPS13
RPL15
RPL12
RPS12
UPF1
Neuronal process
PARD3
YWHAE
PTK7
BCR
LRP1
CPNE1
GAPDH
GRB2
SMC1A
NIPBL
SOX4
TP53BP1
MED1
MEPCE
TMPO
MPP8
JMJD1C
MLLT10
SALL4
CEBPZ
RBPJ
EAF1
SUPT5H
ZMYM4
NSD3
Transcriptional control
IGF2BP2
NXF1
ZC3H11A
CKAP5
RNA transport
Chromatin-associated proteins
-3.8
3.8
log2 mean(VCPmut/Ctrl)
D
Acs et. al.
NSMB 2011
Cohesin
53BP1
mutant VCP
DNA damage
response
factors
This study
Accumulation of DNA damage
in neurons over time
Mitra et. al PNAS 2019
Wang et. al Nat Commun. 2018
Lans et. al.
Nat Rev MCB 2019
Transcriptional
Stress
Aberrant transcripts
Pol2
UPF1
RNA export
NMD
This study