Human transcription factor protein interaction networks

Helka Göös
Institute of Biotechnology, HiLIFE, University of Helsinki
Matias Kinnunen
Institute of Biotechnology, HiLIFE, University of Helsinki
Leena Yadav
University of Eastern Finland
Zenglai Tan
Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland
Kari Salokas
https://orcid.org/0000-0002-4471-6698
Qin Zhang
University of Oulu
Gong-hong Wei
University of Oulu
Markku Varjosalo (markku.varjosalo@helsinki.fi)
University of Helsinki https://orcid.org/0000-0002-1340-9732

Article

Keywords:

Posted Date: February 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-228624/v1

Version of Record: A version of this preprint was published at Nature Communications on February 9th, 2022. See the published version at https://doi.org/10.1038/s41467-022-28341-5.

Human transcription factor protein interaction networks

Helka Göös[1], Matias Kinnunen[1], Leena Yadav[1], Zenglai Tan[2#], Kari Salokas[1#], Qin Zhang[2], Gong-Hong Wei[2,3], Markku Varjosalo[1*]

[1]Institute of Biotechnology, HiLIFE, University of Helsinki, Helsinki, Finland
[2]Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland
[3]Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Fudan University Shanghai Cancer Center, Shanghai Medical College of Fudan University, Shanghai, China

# equal contribution

* Correspondence, lead contact: Markku Varjosalo, markku.varjosalo@helsinki.fi

Abbreviations: ACTB, Beta-actin; ACTBL, Beta-actin-like protein 2; AP-MS, affinity purification mass spectrometry; BioID, proximity-dependent biotinylation; DBD, DNA binding domain; GO-BP, Gene Ontology Biological Process; GTF, general transcription factor; HAT, histone acetyl transferase; KDM2B, Lysine-specific demethylase 2B; NFI, Nuclear Factor 1 family TFs; NLS, nuclear localisation signal; NM1, Nuclear myosin 1; NR, nuclear receptor; NSL, non-specific lethal; PIC, pre-initiation complex; Pol-II, RNA polymerase II; PPI, protein-protein interaction; TBP, TATA-binding protein; TF, transcription factor

Abstract

In participation of transcriptional regulation, transcription factors (TFs) interact with several other proteins. Here, we identified 7233 and 2176 protein-protein interactions for 110 different human TFs through proximity-dependent biotinylation (BioID) and affinity purification mass spectrometry (AP-MS), respectively. The BioID analysis resulted more high-confident interactions, highlighting the transient and dynamic nature of many of the TF interactions.

Using clustering and correlation analyses, we identified subgroups of TFs associated with specific biological functions, such as RNA-splicing, actin signalling or chromatin remodeling. We also observed 203 TF-TF interactions, of which 175 were interactions with Nuclear Factor 1 (NFI) -family members, indicating uncharacterized cross-talk between NFI signalling and numerous other TF signalling. Moreover, TF interactions with basal transcription machinery were mainly observed through TFIID and SAGA complexes.

This study, not only, provides a rich resource of human TF interactions, but also act as starting point directing future studies aimed at understanding TF mediated transcription.

Keywords: Transcription factors, Protein-protein interactions, transcriptional regulation, NFIA, splicing, interaction proteomics, mass spectrometry

Introduction

'The central dogma' states that genetic sequence information from DNA is transcribed to RNA and subsequently translated into proteins. These processes are tightly regulated and employ a plethora of proteins. Transcription, the first step, is regulated by transcription factors (TFs), which represent one of the largest families of the human genes. In human, 6–9% (~1,400-1,900) of proteins are predicted to regulate gene expression through DNA binding ([1-3], https://www.proteinatlas.org), and the most recent manual curation identified 1639 likely human TFs[4].

Complex and multilayer regulation of transcription involves not only direct binding of TFs to a target gene's regulatory element(s) but there exists a complicated interplay between TFs and TF binding proteins. These include several cofactors, the Mediator complex, basal transcription machinery, TF-activity modulating enzymes (such as phosphatases and kinases), dimerization partners, subunits and inhibitory proteins[5-8]. Moreover, as the chromosomal DNA is packed into chromatin to prevent uncontrolled transcription, TFs also interact with several chromatin remodeling proteins. The formed complexes are necessary to regulate the accessibility of DNA to allow chromatin opening and thereby gene transcription.

TFs play crucial roles in regulating numerous cellular mechanisms and are key regulators of tissue growth and embryonic development - processes which may cause cancer and other disorders when aberrantly controlled. Therefore, understanding the TF network at systems level would form an important crucial foundation for future studies as well as for therapeutic applications[7]. While the binding of TFs to DNA is extensively studied, for the most part we are still lacking a global understanding of the TF protein-protein interactions (PPIs) and their roles in the regulation of transcription. Therefore, we sought to fill this knowledge gap by using the recently developed state-of-the-art PPI identification methods, which allow unprecedented sensitivity and depth. In this study, we systemically characterized the PPIs of a selected set of 110 human TFs using affinity purification mass spectrometry (AP-MS) and proximity-dependent biotinylation (BioID) mass spectrometry. We

identified 7233 PPIs in the BioID analysis and 2176 PPIs in the AP-MS analysis. Most of the detected interactions were nuclear and linked to transcription and transcriptional regulation. These interactions paint a picture on how transcription factors are activated or repressed, and also add experimental evidence on the potential relevance of transient interactions in the advent of transcription related nuclear condensates and phase separation.

This large interactome network of TFs allowed us to recognise several interactome subgroups of TFs, such as TFs linked to actin and myosin signalling, TFs linked to mRNA splicing and TFs linked to chromatin remodeling. In addition, we observed that most of the studied TFs interacted with nuclear factor 1 (NFI) TFs, which are essential for several developmental and oncogenic processes. In sum, this work represents a rich resource to direct future studies aimed at understanding TF mediated transcription and how the TF formed interactions regulate important cellular phenomena in both health and disease.

Results

Identification of TF protein-protein interactions

To systematically investigate the protein-protein interactions formed by the human TFs, we selected a representative set of 110 TF genes from different TF families (Table S1A), which were then subjected to two independent mass spectrometry based interactome analysis methods. (Figure 1A). First, the stable TF complexes were purified using single-step Strep-tag affinity purification (AP-MS). Secondly, a proximity-dependent labelling approach (BioID) utilizing a minimal biotin ligase tag BirA* was used to detected also transient and proximal interactions of the TFs[9,10]. The expression of studied TFs were adjusted on the corresponding transgenic cell lines to close-to-physiological level by the tetracycline inducible and adjustable Tet-On expression system[11].

In total, we identified 7,233 high-confidence PPIs using BioID analysis and 2,176 PPIs using AP-MS method (Table S1 B-C; Figure 1B, C). For initial sanity check for the obtained TF interactomes, we mapped the interactors to their known subcellular localisations from the Cell Atlas[12]. This analysis revealed that > 80% of the TF interactors were detected to have nuclear localisation (yellow nodes; Figure 1B; Table S2), confirming the expected nuclear compartmentalisation of the studied TFs and their interactors. Remarkably, the majority (>75%) of the interactions within the TF interactome were previously unreported (Figure 1C; Table S1). On average, we identified 65 PPIs/TF in the BioID data and 21 PPIs/TF in the AP-MS (Figure 1D). Higher number of identified interactions by BioID compared to AP-MS agrees with many recently reported medium- and large-scale interatomic studies[13-16], but interestingly, our results suggest that TFs do prefer to form more transient or proximal interactions than stable protein complexes. This finding is consistent with the phase separation model for TF interactions, where interactions incorporated in TF condensates are dynamic[17-20].

The BioID method has been suggested to be efficient for studying transient interactions[9], and this is supported by our results, strongly suggesting that BioID is the method-of-choice for studying TF

interactions (Figure 2A). Most of the TFs show more detected high-confidence interactions with the BioID-method, with only few exceptions with Krüppel-like factor (KLF) family of transcription factors (Figures 2A and 1C). There were prominent differences in number of detected PPIs between different TF -families; for example, KLFs, SPs, TLXs, HNFs, SOXs and PAXs had over 100 PPIs on average, whereas NFACs, IRFs, STATs, GLIs, ETVs and TEADs had fewer than 50 PPIs on average (Figure 2B; Table S3). Some families, such as KLFs, PAXs, FOXs and NFIs, had similar number of PPIs between the members, but few families, including SPs, TEADs, ELKs and ETVs, had high variability in number of PPIs per bait (Figure 1D).

The most common TF interactor observed in our study was lysine-specific demethylase 2B (KDM2B), which interacted with 62 TFs (Table S1B). In addition, two lysine methyltransferases were among the top five of the most frequent TF preys (KMT2D: 58 PPIs and KDM6A: 53 PPIs), which highlights the importance of the histone modification homeostasis in the regulation of transcription. The detected interactions of lysine methyltransferases with TFs are highly specific and very rarely detected in large-scale studies with other key cellular signalling molecules[16,21,22], and hardly ever detected as contaminants[23]. Other common TF interactors were NFIA (54 PPIs), TLE1 (53 PPIs), CIC (52 PPIs) and several zinc finger proteins (50–52 PPIs). In addition, the well-established corepressors BCOR (48 PPIs) and NCOR2 (48 PPIs) were high on the list. Not surprisingly, the most frequently observed TF interactors were transcriptional activators and repressors.

To obtain a glimpse to the biological nature of TF interactions, we performed Gene Ontology biological process (GO-BP) enrichment analysis for all BioID interactions (Figure 3A, Table S4). As expected, BP terms linked to transcription and its regulation were significantly enriched. The most significantly enriched term was ‘transcription, DNA-templated’ with a p-value of 6.09×10^{-104}. This was followed by biological processes linked to positive and negative regulation of transcription with p-values of less than 1.34×10^{-62}.

Comparison to other studies

As BioID can capture transient and proximal interactions, most experimental validation methods, such as coimmunoprecipitation, are not sufficiently sensitive to validate the results. We, therefore, compared the identified PPIs with previously published interactions. In a medium scale analysis, Li et al. screened the PPIs of 59 TFs by tagging them with a C-terminal SFB-tag (S protein-tag, Flag-tag and Streptavidin binding peptide) and identified the interacting proteins using tandem affinity purification coupled to MS[6]. This analysis identified 2,156 PPIs. Fourteen of the TFs analysed in their study were included in our set (CREB1, ETS1, FOS, FOXI1, FOXL1, FOXQ1, IRF3, MEF2A, MYC, NFKB1, PPARG, STAT3, TEAD2 and TP53). Their approach is close to our AP-MS analysis, which identifies more stable protein-protein interactions and protein complexes than transient interactions. Therefore, it is not surprising that the overlap between our BioID PPIs and their PPIs was low; only 6% of our PPIs were covered by their study (Table S1B). A comparison with our AP-MS results revealed more common interactions; 26% of our AP-MS PPIs were detected by their approach (Table S1C). These differences detected between Li et al.'s PPIs and those we identified are most likely due to the transient nature of TF interactions and the different tagging strategies used.

Next, we compared our PPIs to public interaction databases such as PINA2, STRING, IntAct and Biogrid (Tables S1B-C). Overall, 21% (1,525/ 7,233) of our BioID PPIs and 16% (345/ 2,176) of our AP-MS interactions were also found in public databases or from the affinity-based TF interatomic study conducted by Li et al. (Figure 1B;[6]) The PPIs of several TFs, such as SOX2, MYC, TYY1, PAX6, HNF4a and GATA2, overlapped with more than 50 known interactions in the databases, whereas the PPIs of other TFs, such as ZIC3, ELK4, EN1 and NHLH1, did not overlap with any known PPIs from the databases (Table S1B).

Clustering of transcription factor protein-protein interactions

TFs are often classified according to their DNA-binding domains (DBDs). The DBD distribution of studied TFs compared to all human TFs is shown in Figure 3B. The majority of the studied TFs had C2H2 zing finger (ZF) or Homeodomain DBD, which are the most common DBDs among the human TFs[4]. To study whether the identified PPIs of the various TFs correlated with their DBD- families, we performed a hierarchical clustering of baits by their prey intensities and compared that to bait DBDs. Only a modest correlation was seen between the PPIs and DBDs TLX and LHX homeodomain TFs and KLFs and TYY1 -C2H2 ZF TFs clustered together, but no other correlations with DBDs were observed (Figure 3C).

Next, we wanted to see if PPI clustering correlated with TF amino acid sequence. For that we aligned the full amino acid sequences and compared these to the hierarchical PPI clustering (Figure S1A). The sequence alignment comparison to PPI clustering revealed multiple clusters with similarities in PPIs and sequences (Figure S1A), including the clusters of ELFs, NFIs, LHXs and KLFs.

In addition, the DNA-binding motifs of the studied TFs were aligned using the matrix-clustering tool RSAT (Figure S1B)[24] and compared to PPI clustering (Figure S1C). Several TF clusters also exhibited similarity between DNA-binding motifs and protein interactomes (Figure S1C), including the clusters of MYB, TBR and PAXs and the clusters of IRFs, TEADs and STATs.

TF interactions with basal transcription machinery and Mediator complex

Eukaryotic gene transcription is mostly executed by RNA polymerase II (Pol-II), which binds in conserved core promoters. In addition to Pol-II, the core promoters also bind the SAGA complex and the basal transcription machinery (known also as pre-initiation complex, PIC), which is composed of Pol-II, Mediator complex and general TFs (GTFs). The GTFs are TATA-binding protein (TBP), TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Table S5;[25,26]. To assess how the studied TFs interacted with PIC components, we retrieved the GTFs and Mediator complex members from the

CORUM protein complex database and compared them to the identified PPIs (Table S5). We observed multiple interactions with both TFIID and SAGA complex components, but only few interactions with Mediator complex members (Table S5) and we did not detect interactions between the studied TFs and TFIIA, TFIIB, TFIIF, TFIIH or Pol-II complex components. This indicated that under the given conditions, TF activity from enhancers to core promoter and PIC is mainly mediated by TFIID, SAGA and Mediator complexes.

TF interactions with Nuclear factors

Interestingly, we found 203 TF-TF (bait-bait) interactions in our TF interactome (Figure 4A). Most of these interactions (175) were TF interactions with NFI family members (NFIA, NFIB, NFIC and NFIX; Figure 4B). A total of 54 TFs interacted with one or more NFIs (Figure 4B): 52 TFs interacted with NFIA, 33 with NFIB, 12 with NFIC and 6 with NFIX (Figure 4B). In addition, all four NFIs formed bidirectional interactions (when used reciprocally as bait proteins) to each other in both the AP-MS and BioID analyses (Figure 4B). NFIs had also multiple other shared interactions than the above mentioned interactions with TFs (Figure S2A).

From all eight studied SOX family members of TFs, only SOX4 had no interactions with NFI proteins (Figure 4A). This suggested a previously unknown crosstalk between SOX transcription factors and NFI signalling. Moreover, we found that all five studied PAXs interacted with NFIs (Figure 4B and S2B). In fact, PAX9 was the only TF in set that interacted with all four NFIs. To our knowledge, no link between PAX9 and NFIs has been reported before. In addition to SOXs and PAXs, four (out of six) LHXs, six (out of ten) KLFs and all three studied TLXs interacted with NFIs (Figure 4B). These and the other detected TF-NFI interactions (Figure 4B) indicated that NFIs take part in multiple cellular processes with other TFs.

GO-BP enrichment analysis of NFIs' interactomes using the total BioID interactome as a background showed transcription-related BP terms to be significantly enriched, indicating NFIs importance in transcription regulation in general (Table S4).

Given the fact that it interacted in our analyses with 52 TFs, it is possible that NFIA could regulate the transcriptional activity of other TFs. To test this, we generated luciferase-based reporter (DNA binding domains extracted from JASPAR[27]) assays for selected TFs interacting with NFIA. The reporter assays which displayed induction after introduction of the corresponding TF were chosen for NFIA via siRNA-mediated knockdown experiments. Knockdown efficiency was first confirmed by western blotting using a specific antibody against NFIA (Figure 4C). Next, the effect of NFIA depletion on selected reporter activity was tested in the presence and absence of the corresponding TF (Figure 4C). Interestingly, both KLF4 assays, detecting the repressive and activating response of KLF4, showed altered luciferase activity after NFIA silencing: both the repressive and activating responses after the KLF4 induction were reduced (Figure 4C). In addition, SOX2 and PAX6 showed reduced activity after the NFIA silencing, while EN1 activity was increased after the depletion of NFIA (Figure 4C).

To further examine whether NFIA regulates genome-wide chromatin binding of its interacting TFs, we performed SOX2 chromatin-immunoprecipitation and sequencing (ChIP-seq) upon depletion of NFIA as an example. This analysis revealed the substantial loss of 6,921 SOX2 binding sites, while only 362 sites remained and 1,341 were gained (Figure 4D; Table S6). Consistent with this global shift in the genomic binding profile, depletion of NFIA led to a drastic difference in pathway enrichment for SOX2 target genes (Figure S3), suggesting that NFIA might contribute to the pathophysiological function of its interacting TFs on a genome-wide scale.

In addition, we tested the activity changes with some of the TFs that we did not detect to interact with NFIA. The activity of some of the TFs were affected by NFIA silencing (Figure S4A), while others

were not (Figure S4B). This might indicate that, in addition to a direct regulation of other TFs' activity, NFIA might regulate some other TFs' activity indirectly, without a physical PPI.

Besides NFI-TF interactions, 20 TFs interacted with TYY1 and 16 TFs interacted with ELF1 and/or ELF2 (Figure S5). Twelve of the baits that interacted with TYY1 and ELF1 or ELF2 were the same, indicating similarities in their interactomes.

Prey-prey correlation analysis reveals several biological clusters

The TFs prey-prey correlation analysis, using ProHits-Viz[28], revealed 15 biological clusters (Figure 5A, Table S7). This analysis revealed clusters of preys that were often seen together between baits, suggesting that they might be part of the same complex or colocalize or both. Baits driving the same cluster had a similarity in interactomes, indicating possible shared or similar biological roles. The preys belonging to different clusters and the baits driving the clusters are shown in Table S7. Next, we describe some of the interesting clusters found in this correlation analysis.

Actin and myosin linked protein cluster

The second cluster was mainly formed of proteins linked to actin and myosin signalling. This cluster consisted of 54 proteins, 36 of which were linked to actin and myosin signalling (Figure 5B, Table S7). In addition, 10 proteins were linked to ATP synthesis and mitochondrial respiration (Table S7). Cluster 2 was mainly driven by FOS interactions as FOS interacted with all 54 preys (Fig 5B). Besides FOS, the cluster was also driven by STAT1 (18 interactions), FOXL1 (17 interactions) and 29 other baits with less than 10 interactions (Table S7). Interestingly, FOS interacted with all 36 actin- and myosin-linked proteins in the cluster; STAT1 interacted with 17 and FOXL1 with 6 actin and myosin linked proteins (Table S7; Figure 5B). Other baits did not have actin- and myosin-linked interactions.

FOS and STAT1 uniquely interacted for example with beta-actin (ACTB) and beta-actin-like protein 2 (ACTBL2) (Table S7; Figure 5B). In addition, FOS and STAT1 interacted with ARP3, FOS with ARPC3 and STAT1 with ARPC4, all members of the Arp2/3 complex which mediates the actin

polymerisation in the nucleus and cytoplasm[29,30]. Interestingly, no other TFs interacted with Arp2/3 proteins (Table S1B). Furthermore, FOS and STAT1 were found to interact with MYO1C, isoform 3 of which is also known as nuclear myosin 1 (NM1, Figure 5B). FOS and STAT1 also interacted with various other myosins and myosin-linked proteins, such as the myosin chains MYL6B and MYH14, as well as several unconventional myosins (Tables S1 and S5; Figure 5B).

To explore FOS and STAT1 interactomes further, we performed GO-BP enrichment analysis on all TF BioID interactomes. This resulted in high enrichment of terms linked to actin filaments and cell-cell adhesion (p-values <0.001; Table S5), which further highlights the link from FOS and STAT1 to actin and myosin signalling among the TFs studied.

Preys linked to RNA splicing and processing in Clusters 10 and 11

Next, we found significant enrichment of proteins linked to mRNA splicing and processing in Clusters 10 and 11 (Figure 5A). Cluster 10 consisted of 29 preys and Cluster 11 of 34 preys, of which 23 and 20, respectively, were linked to RNA splicing and processing (Figure 6A, Table S7). GO-BP analysis showed a significant enrichment of proteins linked to 'mRNA splicing, via spliceosome' (p-value 2.70×10^{-11} in Cluster 10 and 6.7×10^{-13} in Cluster 11). Cluster 10 was driven mainly by GATA1 (28 interactions), GATA3 (20 interactions) and SP7 (20 interactions) (Figure 6A). Cluster 11 consisted almost totally of SP7 interactions; SP7 interacted with all 34 proteins (Figure 6A, Table S7).

In our dataset SP7 was the only protein to interact with core spliceosomal components SNRPA (SnRNPa, U1A), RU17 (SnRP70, U1-70), CD2B2 (U5-52K) and RSMB (snRNP-B). These newly identified splicing related interactions indicated that GATA1, GATA3 and especially SP7 are related to splicing and RNA processing. This was also evident in GO-BP enrichment analysis of GATA1 and SP7 interactions using all the TF interactions from our study as a background; GO-BP term "mRNA splicing, via spliceosome" was again significantly enriched (p-values 2.52×10^{-4} for GATA1 and 1.25

x 10^{-6} for SP7; Table S4). These new splicing related interactions for GATAs and SP7, and possible roles in regulation of splicing are highly intriguing and require further studies.

TF interactions with chromatin modulating complexes in Cluster 14

Preys in Cluster 14 (Figure 5A, Figure 6B, Table S7) had clear biological roles in chromatin modulation, especially in histone H4 and H3 modifications. Of 57 preys, 36 were directly linked to histone and chromatin signalling (Figure 6B, Table S7). These included ten members and one putative regulator of the INO80 chromatin remodeling complex (Figure 6B), seven members of the non-specific lethal (NSL) complex and nine members of the MLL1-WDR5 histone-3-lysine-4-(H3K4) methyltransferase complex.

Cluster 14 was mainly formed of TYY1 interactions; TYY1 interacted with all 57 preys in the cluster (Figure 6B, Table S7), while ELF1 interacted with 24, ELF4 with 24, ELF2 with 20 and HNF4A with 15 preys. Other baits driving the cluster are listed in Table S7.

TYY1 is known to be part of the INO80 complex[31]. Predictably, we found TYY1 to interact with 10 subunits of the INO80 complex (Figure 6B; Table S7). TYY1 interactions with INO80 complex members appear to be very stable as many of the interactions were also detected in the AP-MS data (Table S1C). Besides TYY1, we found that ELF4 interacted with eight INO80 complex members and with its putative regulator UCHL (Figure 6B).

Cluster 14 also contained seven out of nine members of the NSL histone acetyltransferase complex. In addition, WDR5 was identified outside Cluster 14 (Figure 6B). In total, we found that TYY1 and MYC interacted with WDR5 and all seven identified subunits of the NSL complex, whereas ELF1, ELF2, ELF4 and HNF4A interacted with five subunits (Figure 6B, Table S7). Histone acetyltransferase KAT8 (main unit of the NSL complex, also known as MYST1), was found to interact with TYY1, ELF1, ELF2, ELF4, HNF4a and MYC.

The top four baits in cluster 14 (TYY1, ELF1, ELF4 and ELF2) were also clustered together in bait-prey clustering, indicating similarities in protein interactomes (Figure 3C). In addition, 12 other TFs studied formed bait-bait interactions with TYY1, ELF1 and ELF2 (Figure S6). These results suggest that TYY1, ELF1 and ELF2 may have shared or similar biological functions. Of all five ELFs (1-5) studied, ELF1, ELF2 and ELF4 were the most connected and shared the greatest number of interactions, including interactions with TYY1 (Figure S6).

TF interactions with histone acetyltransferase complexes in Cluster 15

Closer analysis of Cluster 15 (Figure 5A) revealed accessory chromatin modulating complexes, especially histone acetylation complexes. BP-GO analysis resulted in the significant enrichment of terms linked to histone H2A, histone H4 and histone H3 acetylation (p-values of less than 7.19×10^{-22}).

In total, Cluster 15 consisted of 50 preys of which 40 were directly linked to histone acetylation (Figure 6C, Table S7). These included 19 members of the SAGA complex and 14 members of the NuA4/Tip60 HAT complex A (Figure 6C). The cluster was mainly driven by MYC, which had interactions with all 50 preys. KLF6 interacted with 29, HNF4A with 21, KLF8 with 20, ELF4 with 17, TYY1 with 13 and ELF1 with 11 preys. Other baits with less than 10 interactions are listed in Table S7.

MYC interacted with all 19 subunits of the SAGA complex identified in our data (Table S7). Furthermore, KLF6 was found to interact with 16, and KLF8 and HNF4a with eight SAGA complex subunits (Table S7). In addition, 14 of 15 NuA4/Tip60 HAT complex A subunits were identified in Cluster 15. MYC was found to interact with all 14 identified subunits, while HNF4a and KLF6 interacted with nine subunits, and KLF8 and ELF4 with eight subunits of this complex (Figure 6C, Table S7).

Discussion

Chromatin opening, transcription, RNA splicing, RNA processing and their regulation are often studied as separate processes. However, our understanding of the simultaneous and co-transcriptional nature of these processes has exploded in recent years[32-36]. In our analyses, TFs were found to interact with proteins involved in chromatin remodeling, transcription, mRNA splicing and RNA processing, highlighting the cooperative nature and close proximity of these processes. This also showed that TFs are central in regulating these interconnected processes. The most common interaction partners for the TFs studied were histone-modifying enzymes, signifying that histone modification and chromatin accessibility regulation are central to all these transcriptional subprocesses.

Studied TFs interact with basal transcription machinery mainly via TFIID, SAGA and Mediator complexes

Mediator complex, SAGA complex and most of the GTFs are multimeric protein complexes, which are needed for Pol-II promoter recognition and transcription initiation[37-39]. In most studied models, the PIC assembly starts with the binding of TBP to TATA- or TATA-like core promoters[40]. TBP belongs to two GTF complexes: TFIID and SAGA. It has been indicated that both TFIID and SAGA participate in the transcription of various genes simultaneously[41], but that the regulation of expression might be dominated by either one of them[42,43]. Different promoters are alleged to prefer either TFIID or SAGA, and it has been indicated that the activity of SAGA/TATA-like promoters might be more dependent on the presence of transcriptional activators (regulated genes) than the activity of TFIID/TATA-promoters (housekeeping genes)[44]. However, this is still controversial, as the depletion of either SAGA or TFIID complex members decreased the transcription of both regulated and housekeeping genes[25,41].

We observed multiple interactions with both TFIID and SAGA components, supporting the theory that both are needed for the transcription of regulated genes. In phase separation condensates

enhancer-bound TFs are physically separated from PIC with multiple cofactor complexes[17-20]. Our data indicated that mainly TFIID and SAGA serve as these cofactors (Table S5).

Interestingly, we detected only few interactions with Mediator complex (Table S5), even though Mediator is generally thought to mediate the regulatory signals between TFs to Pol-II[45,46]. The mediator complex is reported to interact with multiple TFs, and it is thought to form phase separation condensates with many TFs[20,46]. However, it is not comprehensively known how directly TFs interact with Mediator complex members. Multiple TFs are seen to colocalize with Mediator complex members in phase separation complexes *in vitro*[20]. However, our data indicated that the interactions between TFs and Mediator complex members might be mediated through other proteins, such as histone modifiers. We suggest that for these studied TFs and under the given conditions, the signal is primarily transferred to the Mediator complex and to PIC via other cofactors, such as SAGA, TFIID or other chromatin remodeling complexes.

TFs' interactions with the Nuclear Factor I family members

Interestingly, we detected a total of 203 bait-bait interactions within the studied TFs (Figure 4A), most of which (175) were interactions with NFI family members. NFIs are CCAAT-box-binding TFs that have similar DBDs and bind as hetero- or homodimers to the same common consensus sequence[47-49]. There are four NFI family members (NFIA, NFIB, NFIC and NFIX) in humans and most vertebrates[50,51]. Originally, the NFIs were identified as essential for adenovirus replication[52], but over the years they have been found to control a variety of genes in cancer and in development[53-58]. For example, NFIs are found to have multiple translocations leading to oncogenic gene fusion proteins in several cancer types[58], and knockout studies of NFIA, NFIB, NFIC, NFIX have revealed their necessity in lung, central nervous system, brain, tooth, skeletal and muscle development[53-55,57,59-61].

In our data, especially SOXs, PAXs, LHXs, KLFs and TLXs had multiple interactions with NFIs (Table S1B). These interactions indicated that NFIs take part in many cellular processes with other TFs. NFI family members have been found to interact with some individual TFs, such as with few SOX family proteins[62,63]. However, TF-NFI interactions in this scale have not been reported before. Given the important role of NFIs in regulation of developmental processes and their impact on cancer development, the high number of TF-NFI interactions might indicate that the activity of NFIs is generally regulated by other TFs, or vice versa. To test this, we generated the reporter gene assays for selected NFIA interacting TFs and discovered that RNAi silencing resulted in altered TF activity (Figure 4C). This strengthens the theory that NFIs have extensive and not a well characterised role in the regulation of other TFs' activity in gene expression regulation.

Moreover, the widely altered DNA binding of SOX2 after the NFIA silencing (Figure 4D) indicated that NFIA-TF PPI might be essential for genome-wide biding of certain TFs, and thus, be crucial for the normal regulatory functions of NFIA-interacting TFs.

FOS and STAT1 interact with nuclear actin and myosin related proteins

Nuclear actin is associated to chromatin remodeling complexes and is part of the Pol-II transcription machinery[64]. Actin dynamics have also been directly linked to gene transcription regulation[65,66].

In our analysis, the interactomes of FOS and STAT1 were found to be enriched with proteins linked to nuclear actin and myosin signalling. FOS and STAT1 interacted uniquely, e.g., with beta-actin (ACTB), ACTBL, and MYO1C, isoform 3 of which is also known as NM1 (Table S7, Figure 5B). NM1 and ACTB are suggested to be associated with each other and to play important roles in Pol-II transcription[64,67]. ACTB is also part of several chromatin remodeling complexes, such as BAF, Tip60 and INO80[68]. Indeed, we also detected FOS interacting with 13 BAF complex members (Table S1B, Figure 5B), suggesting that FOS activity is regulated trough BAF complex and actin dynamics. FOS

has been reported to be linked to BAF signalling before[69,70], but how actin is involved in this process remains largely unknown.

STAT1 activity is known to be regulated by actin cytoskeleton and extracellular matrix proteins[71]. Interestingly, STAT1 did not show any interactions with BAF complex members, indicating a different mechanism or different mediating proteins between STAT1 and actin proteins.

SP7 interacts with RNA-splicing proteins

In our analysis, proteins involved in RNA-splicing were enriched in interactomes of SP7, GATA1 and GATA3. In addition, we found that SP7 and GATA3 interacted with EP300 (p300) along with other members of the p300-CBP-p270-SWI/SNF HAT complex (Figure 6A, Table S1B).

TFs are known to affect RNA splicing in three ways; they can bind to RNA to recruit coregulators that also take part in splicing, block the associations of splicing factors with mRNA, and/or influence the transcription elongation rates, which are known to impact on splicing by skipping the weak 3' splice sites at a high rate[72]. One way to alter the elongation rate is through TF-mediated recruitment of EP300 which induces the histone acetylation of nearby promoters, increases the elongation rate and promotes exon skipping[73]. Therefore, SP7 and GATA3 interaction to EP300 suggest that they may be connected to the p300 chromatin remodeling complex and, thus, to the regulation of elongation rate.

Some TFs, such as steroid hormone receptors, nuclear receptors (NRs) and certain non-NR TFs, are known to regulate mRNA splicing by recruiting splicing-linked coregulators[72,74]. One of these coregulators is RBM14 (also known as CoAA), which is an NCOA6 (also known as TRBP) binding protein[75]. We found that, along with other splicing-related proteins, both RBM14 and NCOA6 interacted with SP7 (Figure 6A, Table S1B). This might indicate that SP7 recruits similar splicing-regulation-related coregulators.

SP7 interactions with four core spliceosomal components, and all other splicing related components suggested that SP7 has a largely unstudied role in recruiting the splicing machinery to the nascent pre-mRNA – a role that needs to be studied more. Like some other C2H2 zinc finger TFs, such as CTCF, VEZF1, MAZ and WT1, that are known to regulate mRNA splicing[72], SP7 might participate in pre-mRNA splicing.

TF interactions with chromatin modulating complexes

As expected, multiple TFs had interactions to chromatin modulating proteins. We found several TFs to interact e.g. with INO80, NSL, SAGA and NuA4/Tip60 HAT complexes (Table S1B). TF interactions to these complexes will be next discussed in more detail.

INO80 ATP-dependent chromatin remodeling complex activates transcription[31], regulates genomic stability through DNA repair[76], contributes to DNA replication[77] and, by shifting the nucleosomes, remodels the chromatin[78,79]. Predictably, TYY1, that is known to be part of INO80 complex, had multiple interactions with INO80 complex members in our analyses. More interestingly, we found that ELF4 had also multiple interactions with INO80 members. To our knowledge, ELFs have not been previously linked to INO80 signalling or TYY1 and this should be studied further.

Next, we found that various TFs, such as TYY1, ELF1, ELF2, ELF4, HNF4a and MYC, interacted with members of NSL histone acetyltransferase complex. The interactions between TYY1 and three NSL complex subunits (MCRS1, HCFC1, WDR5) have been reported (PINA2;[80]), but the role of TYY1 in NSL regulation is still unknown. Moreover, the roles of ELF1, ELF2 and ELF4 in the NSL complex (as in the INO80 complex) and their interactions with WDR5 are not well understood and require further study.

As mentioned earlier, the SAGA complex is a multi-module complex that has an important role in Pol-II recruitment for all expressed genes[81]. In addition, SAGA takes part in mRNA synthesis and export, maintenance of DNA integrity and histone modifications, such as histone acetylation,

succinylation and ubiquitylation[81-87]. In our data, multiple interactions were detected between SAGA complex members and the studied TFs (e.g. MYC, KLF6, KL8, HNF4a; Tables S1B and S5). MYC connections to SAGA are known[88] and KLF6 interaction with one of the subunits, TAF9, is reported in PINA2 database. In addition, KLF6 is known to interact with HDAC3 in preadipocyte differentiation[89]. However, we found no other connections between KLF6 and SAGA complex or histone modification. Our data indicated that KLF6 connections to SAGA are *bona fide* and should be studied further.

Finally, we found MYC, HNF4a, KLF6, KLF8 and ELF4 to interact with the NuA4/Tip60 HAT complex (Tables S1B and S7) that plays essential roles in cell cycle control, transcription and DNA repair, and act in the N-terminal acetylation of histones H4 and H2A[90]. NuA4/Tip60 HAT complex, along with other HAT complexes, is known to participate in MYC-signaling[91]. Accordingly, we found 14 interactions between MYC and NuA4/Tip60 HAT complex (Tables S1B and S7). Like that of MYC, HNF4a's association with the NuA4/Tip60 HAT complex is reported in a previous study[92]. Interestingly, we found also strong connection (eight to nine interactions) between KLF6, KLF8 or ELF4 and NuA4/Tip60 HAT complex. However, as mentioned earlier with regard to the SAGA complex, not much is known about the link between KLF6 link and histone modification. Therefore, KLF6's role in HAT complexes remains largely unstudied and requires further investigations.

Taken together, TYY1, ELF4, ELF1, ELF2 (Cluster14) and MYC, KLF6, KLF8 and HNF4a (Cluster 15) had several interactions with chromatin remodeling complexes. Some research has been conducted on the contributions of TYY1, MYC and HNF4a to histone modification and chromatin remodeling, but the roles of ELFs and KLFs in chromatin remodeling remain largely unexplored.

Interestingly, even though chromatin remodeling and histone modifications are known to be important for almost all TF signaling and most of the studied TFs interact with proteins that mediate these processes, only a fraction of TFs seemed to have interactions with almost-complete histone-modifying or chromatin remodeling complexes. These interactions include TYY1 and ELF1

interactions with the INO80 complex; TYY1 and MYC interactions with the NSL complex; MYC and KLF6 interactions with the SAGA complex; and MYC, HNF4a, KLF6, KLF8 and ELF4 interactions with the NuA4/Tip60 HAT complex. This observation suggests that these TFs act in close relation to these complexes and take part in them at least in certain conditions.

Conclusions

While TF binding to DNA is much studied, there is still a lack of comprehensive systems levels understanding of human TF protein interactions. The protein interactions of other human large proteins families, such as kinases and phosphatases, have been studied on the systems level[16,21,22,93], but the TFs protein interactions remain globally unstudied. This study provides so far, the most comprehensive systems-level analysis of human TFs identifying the largest reported cohort of TF PPIs and serving as a rich resource for further research and development of pharmaceutical treatment for TF-related diseases. It also allows to profile TFs protein interactomes in the context of more than 100 TF interactomes. Moreover, this is the first large-scale study to identify the dynamic PPIs of TFs using transient and proximal interactions catching BioID method. Finally, as defects in TF signalling often lead to severe pathological conditions[94-96], and TFs function as downstream players of multiple signalling cascades[3], identifying TF PPIs make a crucial contribution to pharmacological targeting of TF-related diseases.

METHODS

Used cell lines

Cell lines stably expressing selected TFs were generated from Flp-In™ T-REx™ 293 cells (Life Sciences). siRNA silencing and lusiferase experiments were performed in HEK 293 cell line (Atmerican Type Culture Collection, Manassas, VA). All the cells were cultured in low glucose tetracycline-free DMEM (Sigma Aldrich) complemented with 10% FBS and 100 µg/ml Penisillin/streptomysin (Life Technologies) at 37 °C with 5 % CO_2.

Generation of TF expression constructs and stable, inducible Flp-In™ 293 T-REx cell lines

110 TFs from different TF families were selected for this study. Using Gateway® cloning, the TF coding sequences without stop codons were obtained from ORF libraries and commercially cloned into pDONR221 entry vectors (GenScript). To generate tetracycline-inducible stable cell lines, constructs were cloned into N-terminal pTO_HA_StrepIII_BirA-N_GW_FRT, pTO_HA_StrepIII-N_GW_FRT or MAC-tag vectors and introduced into Flp-In™ T-REx™ 293 cells (Life Technologies, Carlsbad, CA) to generate stable, isogenic and inducible cell lines as described by Liu et al[15].

Affinity purification and mass spectrometry analysis

Approximately $1x10^8$ Flp-In™ T-REx™ 293 cells stably expressing human TFs were induced with 2 µg/ml of tetracycline (AP-MS and BioID) and 50 µM biotin (BioID) for 24 hours. The cells were pelleted using centrifugation, snap frozen in liquid nitrogen and stored at -80°C. The samples were then suspended in 3 ml of lysis buffer (50 mM HEPES pH 8.0, 5 mM EDTA, 150 mM NaCl, 50 mM NaF, 0.5% NP40, 1.5 mM Na_3VO_4, 1 mM PMSF, 1 x protease inhibitors cocktail, Sigma) on ice.

508 The BioID lysis buffer was completed with 0.1% SDS and 80 U/ml Benzonase Nuclease (Santa Cruz
509 Biotechnology, Dallas, TX), and the lysis was followed by incubation on ice for 15 min and three
510 cycles of sonication (3 min) and incubation (5 min) on ice.

511 All samples were then cleaned up by centrifugation and the supernatants were poured into microspin
512 columns (Bio-Rad, USA) that were pre-loaded with 200 μl of Strep-Tactin beads (IBA GmbH) and
513 allowed to drain under gravity. The beads were washed with 3 x 1 ml lysis buffer and then 4 x 1ml
514 lysis buffer without the detergents and inhibitors (wash buffer). The purified proteins were eluted
515 from the beads with 600 μl of wash buffer containing 0.5 mM biotin. To reduce and alkylate the
516 cysteine bonds, the proteins were treated to a final concentration of 5 mM TCEP (Tris(2-
517 carboxyethyl) phosphine) and 10 mM iodoacetamide, respectively. Finally, the proteins were
518 digested into tryptic peptides by incubating them with 1 μg sequencing grade trypsin (Promega)
519 overnight at 37°C. The digested peptides were purified using C-18 microspin columns (The Nest
520 Group Inc.) as instructed by the manufacturer. For the mass spectrometry analysis, the vacuum dried
521 samples were dissolved in buffer A (1% acetonitrile and 0.1% trifluoroacetic acid in MS grade water).

522 The peptides were analysed on EASY-nLC II system connected to an Orbitrap Elite ETD hybrid mass
523 spectrometer (Thermo Fisher Scientific, Waltham, MA). The digested peptides were first guided into
524 a precolumn (C18-packing; EASY-Column™ 2cm x 100 µm, 5 µm, 120 Å, Thermo Fisher Scientific)
525 and then into analytical column (C18-packing; EASY-Column™ 10 cm x 75 µm, 3 µm, 120 Å,
526 Thermo Fisher Scientific). The separation was completed with a 60-min linear gradient from 5 to
527 35% of buffer B (98% acetonitrile, 0.1% formic acid and 0.01% trifluoroacetic acid in MS grade
528 water) at a stable flow rate of 300 nl/min. Data-dependent acquisition analysis was performed as
529 follows: after one high-resolution (60,000) FTMS full scan (m/z 300-1700), top20 CID- MS2 scans
530 in ion trap were performed (energy 35). The highest fill time was 200 ms for FTMS (Full AGC target
531 1,000,000) and 200 ms for the ion trap (MSn AGC target of 50,000). Only precursor ions with higher

than 500 ion counts were chosen for MSn. The preview mode was applied for the FTMS scan to achieve a high resolution.

Protein identification

The proteins were identified using SEQUEST search engine in Proteome Discoverer™ software (version 1.4, Thermo Scientific). The raw data were analysed against the reviewed human proteins from the UniProt-database (release 2018_01; 20,192 entries). FASTA library was complemented with BSA, tag sequences, trypsin, biotin and GFP. Biotinylation (+226.078 Da) of lysine residues and oxidation (+15.994491 Da) of methionine or N-terminus were used as dynamic modification. In addition, cysteine residues' carbamidomethylation (+57.021464 Da) was used as static modification. A maximum of two missed cleavages and 15 ppm monoisotopic mass error were allowed. The peptide false discovery rate (FDR) was set to <0.05.

The identified proteins were filtered using SAINT software tools[97] with a SAINT score cut-off of 0.74. All the TFs were analysed in two or four replicates. TFs are known to have variable expression levels and patterns, and some of them are present in cells at extremely low copy numbers[98]. To efficiently filter the real interactions, we used 44 and 75 similarly tagged and analysed GFP control runs for the BioID analysis and AP-MS analysis, respectively. We also included GFPs with a nuclear localisation signal (NLS) to efficiently filter out unspecific nuclear interactions. The large nuclear dataset further facilitated the frequency-based deletion of contaminating proteins. The Cytoscape software platform was used to visualize the high-confidence TF PPIs[99].

Data analysis

The subcellular localisations of interacting proteins were obtained from the Cell Atlas[12]. Enriched biological process Gene Ontology terms for all PPIs were obtained from DAVID Bioinformatics Resources[100]. We also used DAVID to study the enrichment of separate TF interactomes against all

556 the PPIs identified in our study. All the terms with the corresponding p-values and FDR are reported
557 in Table S4.

558 A hierarchical clustering of baits (studied TFs) by their preys (interacting proteins) was performed
559 using ProHits-viz with default settings[28]. Comparison of two cluster dendrograms were done using
560 *dendexted* R package (https://www.datanovia.com/en/lessons/comparing-cluster-dendrograms-in-r/).
561 The full amino acid sequences of the studied TFs were downloaded from UniProt[101]. The DNA-
562 binding motifs of the studied TFs were mainly extracted from the JASPAR database[27]. Motifs not
563 found in JASPAR were extracted from the HT-SELEX and ENCODE databases[102,103]. All extracted
564 DNA-binding motifs were aligned using the matrix-clustering tool RSAT[24]. Finally, the prey-prey
565 correlation analysis of the BioId data was performed using ProHits-viz's correlation tool
566 (https://prohits-viz.lunenfeld.ca/Correlation/), where Pearson correlation and hierarchical clustering
567 with Euclidean distance metric was used[28]. Filtered SAINT-interactions were used as input. Apart
568 from default settings, score column was set to SaintScore and cut-off values for filtering were
569 removed as already filtered interaction data was used as input.

570 TF activity was accessed by luciferase assays in three replicates (Figure 4C and S4). Firefly luciferase
571 signals were normalized to renilla luciferase signals and the Student t-test was used to detect the
572 significance of the changes. Stars in figures (Figure 4C and S4) indicate following cut-offs: ***:
573 $p<0.001$, **: $p<0.01$, *:$p<0.05$.

574

575 ***NFIA silencing and reporter gene assays***

576 HEK293 cells were cultured in 96-well plates (7000 cell/well) for 24 hours. This was followed by
577 NFIA siRNA (Dharmacon J-008661-06) transfection in final concentration of 100 nM using
578 Dharmafect transfection reagent (0.35 µl/well). After 24 hours siRNA silencing, the culturing media
579 was replaced with fresh one, and cells were transfected with 50 ng of selected TF or empty vector

(pTO-SH-GW-FRT) along with 47.5 ng reporter construct. The reporter constructs contained 6-8x TF binding sites (TFBSs;[27], minimal promoter and firefly-luciferase reporter. Only the constructs displaying induction after introduction of the corresponding TF were chosen for further analysis. These include reporters for KLF4 (both activating: [TFBS: 6x GGGTGTGG] and repressive: [TFBS: 8x TAAAGGAAGG]), SOX2 (TFBS: 6x CTTTGTT), PAX6 (TFBS: 6x TTCACGCTTGAGTT) and EN1 (TFBS: 8x AAGTAGTGCCC).

In addition, cells were transfected with 2.5 ng of renilla-luciferase construct. After 24 hours cells were collected and the firefly-luciferase and renilla-luciferase signals were detected using Dual-GLO® luciferase Assay System (Promega). Firefly luciferase signals were normalized to renilla luciferase signals and the analysis was performed in three replicates. NFIA silencing was confirmed after 48 hours from siRNA transfection by western blotting using the specific antibody against NFIA (Abcam, ab228897).

NFIA silencing and SOX2 Chromatin Immunoprecipitation Sequencing (ChIP-seq)

HEK293 cells were seeded in 10 cm dishes; 24 hours later, tetracycline with a final concentration of 2 μg/ml was added to induce the expression of SOX2 protein. siRNAs against NFIA or control were transfected with a final concentration of 100 nM by the lipofectamine RNAiMAX reagent (Thermo Fisher Scientific). HEK293 cells with depletion of NFIA were applied in ChIP-seq assays as previously described[104]. In brief, HEK293 cells were cross-linked with 1% formaldehyde for 10 min at room temperature after 48 hours of siRNA transfection. The reaction was quenched with 125 mM glycine. Cells were collected after washing twice by pre-cold PBS and resuspended in hypotonic lysis buffer (20 mM Tris-Cl, PH 8.0, with 10% glycerol, 10 mM KCl, 2 mM DTT, and complete protease inhibitor cocktail (Roche)) to isolate nuclei. The nuclei pellets were washed with pre-cold PBS and resuspended in 1:1 ratio of SDS lysis buffer (50 mM Tris-HCl, pH 8.1, with 1% SDS, 10 mM EDTA, and complete Protease Inhibitor) and ChIP dilution buffer (16.7 mM Tris-HCl, pH 8.1, with 0.01%

SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl and Complete Protease Inhibitor). A Q800R sonicator (QSonica) was used to generate an average size of 300 bp chromatin fragments at 4°C. Dynabead protein G (Invitrogen) was washed with blocking buffer ((0.5% BSA in IP buffer (20 mM Tris-HCl, pH8.0, with 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, and Protease inhibitor cocktail)) and incubated with antibody against HA (ab18181, Abcam). Chromatin lysate was precipitated with Dynabead protein G for 12 hours, then wash the beads 4 times with washing buffer (50 mM HEPES, PH 7.6, 1 mM EDTA, 0.7% sodium deoxycholate, 1% NP-40, 0.5 M LiCl) and 2 times 100 mM ammonium hydrogen carbonate (AMBIC) solution. The DNA-protein complexes extracted from the beads by eluting in extraction buffer (10 mM Tris-HCl, pH8.0, with 1 mM EDTA, and 1% SDS), Proteinase K and RNase A were added to reverse the cross-links. A Mini-Elute PCR purification kit (Qiagen) was used to purify the DNA. The purified DNA was subjected to ChIP-seq library preparation through using the TruSeq ChIP Sample Preparation kit (Illumina). Briefly, DNA was blunted by using an End Repair Mix, then A-Tailing Mix was used to add a nucleotide to the 3' Ends of the DNA fragments. RNA Adaptor Indices were ligated to the DNA fragments and fragment size of 200-500 bp were selected on a 2% agarose gel. The DNA were enrichment by PCR amplification and quantified by using KAPA Library Quantification Kit (Roche). A NextSeq550 sequencing system (Illumina) was used to sequence the DNA library.

Bioinformatics analysis of ChIP-Seq data

The ChIP-seq library was sequenced to generate 76 bp single-end reads. FastQC[105] was applied to assess the quality of raw data and followed by Trimmomatic[106] for quality control. The cleaned reads were aligned to the human genome hg38 assembly using Bowtie2[107]. The ChIP-seq peaks were identified by applying findPeaks.pl from Hypergeometric Optimization of Motif Enrichment (HOMER v4.10)[108] with parameter "-mapq 20", while all other parameters were kept as default. Motif enrichment analysis was performed using HOMER findMotifsGenome.pl and peaks were annotated by "annotatePeaks.pl". Bioconductor package ChIPseeker (1.18.0)[109] was applied to perform pathway

enrichment analysis. Bam files were first converted to bigWig files by using bamCoverage from deepTools2. Heatmaps of aligned reads and average signal plots were generated by Samtools (v1.9)[110] and deepTools2 (v3.3.2)[111].

DATA AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Markku Varjosalo (markku.varjosalo@helsinki.fi)

Material Availability

Plasmids generated in this study will be deposited in Addgene. No other unique reagents were generated in this study.

MAC-tag-N destination vector (Addgene, plasmid no. 108078; RRID: Addgene_108078)

Data and Code Availability

The MS peptide raw data from the MS runs have been deposited to the Peptide Atlas (http://www.peptideatlas.org) under accession number XXXXX and the identified high-confidence protein-protein interactions are downloaded to IntAct -protein interaction database (https://www.ebi.ac.uk/intact). Filtered protein-protein interactions are also available as table S1 and ChIP-seq peak lists as table S6.

References

1 Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. Structure and evolution of transcriptional regulatory networks. *Current opinion in structural biology* **14**, 283-291, doi:10.1016/j.sbi.2004.05.004 (2004).

2 Fulton, D. L. *et al.* TFCat: the curated catalog of mouse and human transcription factors. *Genome biology* **10**, R29, doi:10.1186/gb-2009-10-3-r29 (2009).

3 Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nature reviews. Genetics* **10**, 252-263, doi:10.1038/nrg2538 (2009).

4 Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665, doi:10.1016/j.cell.2018.01.029 (2018).

5 Brivanlou, A. H. & Darnell, J. E., Jr. Signal transduction and the control of gene expression. *Science (New York, N.Y.)* **295**, 813-818, doi:10.1126/science.1066355 (2002).

6 Li, X. *et al.* Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Molecular systems biology* **11**, 775, doi:10.15252/msb.20145504 (2015).

7 Fontaine, F., Overman, J. & Francois, M. Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and obstacles. *Cell regeneration (London, England)* **4**, 2, doi:10.1186/s13619-015-0015-x (2015).

8 Rivera-Reyes, R., Kleppa, M. J. & Kispert, A. Proteomic analysis identifies transcriptional cofactors and homeobox transcription factors as TBX18 binding proteins. *PloS one* **13**, e0200964, doi:10.1371/journal.pone.0200964 (2018).

9 Varnaite, R. & MacNeill, S. A. Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. *Proteomics* **16**, 2503-2518, doi:10.1002/pmic.201600123 (2016).

10 Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of cell biology* **196**, 801-810, doi:10.1083/jcb.201112098 (2012).

11 Glatter, T., Wepf, A., Aebersold, R. & Gstaiger, M. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Molecular systems biology* **5**, 237, doi:10.1038/msb.2008.75 (2009).

12 Thul, P. J. *et al.* A subcellular map of the human proteome. *Science (New York, N.Y.)* **356**, doi:10.1126/science.aal3321 (2017).

13. Abbasi, S. & Schild-Poulter, C. Mapping the Ku Interactome Using Proximity-Dependent Biotin Identification in Human Cells. *Journal of proteome research* **18**, 1064-1077, doi:10.1021/acs.jproteome.8b00771 (2019).

14 Lambert, J. P., Tucholska, M., Go, C., Knight, J. D. & Gingras, A. C. Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *Journal of proteomics* **118**, 81-94, doi:10.1016/j.jprot.2014.09.011 (2015).

15 Liu, X. *et al.* An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nature communications* **9**, 1188, doi:10.1038/s41467-018-03523-2 (2018).

16 Yadav, L. *et al.* Systematic Analysis of Human Protein Phosphatase Interactions and Dynamics. *Cell systems* **4**, 430-444.e435, doi:10.1016/j.cels.2017.02.011 (2017).

17 Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (New York, N.Y.)* **361**, 412-415, doi:10.1126/science.aar4199 (2018).

18 Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.)* **361**, doi:10.1126/science.aar3958 (2018).

19 Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science (New York, N.Y.)* **361**, doi:10.1126/science.aar2555 (2018).

20 Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e1816, doi:10.1016/j.cell.2018.10.042 (2018).

21 Varjosalo, M. *et al.* The protein interaction landscape of the human CMGC kinase group. *Cell reports* **3**, 1306-1320, doi:10.1016/j.celrep.2013.03.027 (2013).

22 Varjosalo, M. *et al.* Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. *Nature methods* **10**, 307-314, doi:10.1038/nmeth.2400 (2013).

23 Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature methods* **10**, 730-736, doi:10.1038/nmeth.2557 (2013).

24 Nguyen, N. T. T. *et al.* RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic acids research* **46**, W209-w214, doi:10.1093/nar/gky317 (2018).

25 Baptista, T. *et al.* SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Molecular cell* **68**, 130-143.e135, doi:10.1016/j.molcel.2017.08.016 (2017).

26 Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295-301, doi:10.1038/nature10799 (2012).

27 Mathelier, A. *et al.* JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic acids research* **44**, D110-115, doi:10.1093/nar/gkv1176 (2016).

28 Knight, J. D. R. *et al.* ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nature methods* **14**, 645-646, doi:10.1038/nmeth.4330 (2017).

29 Yoo, Y., Wu, X. & Guan, J. L. A novel role of the actin-nucleating Arp2/3 complex in the regulation of RNA polymerase II-dependent transcription. *The Journal of biological chemistry* **282**, 7616-7623, doi:10.1074/jbc.M607596200 (2007).

30 Welch, M. D., Iwamatsu, A. & Mitchison, T. J. Actin polymerization is induced by Arp2/3 protein complex at the surface of Listeria monocytogenes. *Nature* **385**, 265-269, doi:10.1038/385265a0 (1997).

31 Cai, Y. *et al.* YY1 functions with INO80 to activate transcription. *Nature structural & molecular biology* **14**, 872-874, doi:10.1038/nsmb1276 (2007).

32 Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annual review of biochemistry* **84**, 165-198, doi:10.1146/annurev-biochem-060614-034242 (2015).

33 Dahan, N. & Choder, M. The eukaryotic transcriptional machinery regulates mRNA translation and decay in the cytoplasm. *Biochimica et biophysica acta* **1829**, 169-173, doi:10.1016/j.bbagrm.2012.08.004 (2013).

34 Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688-700, doi:10.1016/j.cell.2009.02.001 (2009).

35 Komili, S. & Silver, P. A. Coupling and coordination in gene expression processes: a systems biology view. *Nature reviews. Genetics* **9**, 38-48, doi:10.1038/nrg2223 (2008).
36 Reed, R. Coupling transcription, splicing and mRNA export. *Current opinion in cell biology* **15**, 326-331 (2003).
37 Meyer, K. D., Lin, S. C., Bernecky, C., Gao, Y. & Taatjes, D. J. p53 activates transcription by directing structural shifts in Mediator. *Nature structural & molecular biology* **17**, 753-760, doi:10.1038/nsmb.1816 (2010).
38 Poss, Z. C., Ebmeier, C. C. & Taatjes, D. J. The Mediator complex and transcription regulation. *Critical reviews in biochemistry and molecular biology* **48**, 575-608, doi:10.3109/10409238.2013.840259 (2013).
39 Joo, Y. J. *et al.* Downstream promoter interactions of TFIID TAFs facilitate transcription reinitiation. *Genes & development* **31**, 2162-2174, doi:10.1101/gad.306324.117 (2017).
40 Luse, D. S. The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled? *Transcription* **5**, e27050, doi:10.4161/trns.27050 (2014).
41 Fischer, V., Schumacher, K., Tora, L. & Devys, D. Global role for coactivator complexes in RNA polymerase II transcription. *Transcription* **10**, 29-36, doi:10.1080/21541264.2018.1521214 (2019).
42 Lee, T. I. *et al.* Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* **405**, 701-704, doi:10.1038/35015104 (2000).
43 Huisinga, K. L. & Pugh, B. F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Molecular cell* **13**, 573-585 (2004).
44 de Jonge, W. J. *et al.* Molecular mechanisms that distinguish TFIID housekeeping from regulatable SAGA promoters. *The EMBO journal* **36**, 274-290, doi:10.15252/embj.201695621 (2017).
45 Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nature reviews. Molecular cell biology* **16**, 155-166, doi:10.1038/nrm3951 (2015).
46 Borggrefe, T. & Yue, X. Interactions between subunits of the Mediator complex with gene-specific transcription factors. *Seminars in cell & developmental biology* **22**, 759-768, doi:10.1016/j.semcdb.2011.07.022 (2011).
47 Gronostajski, R. M. Analysis of nuclear factor I binding to DNA using degenerate oligonucleotides. *Nucleic acids research* **14**, 9117-9132, doi:10.1093/nar/14.22.9117 (1986).
48 Gronostajski, R. M., Adhya, S., Nagata, K., Guggenheimer, R. A. & Hurwitz, J. Site-specific DNA binding of nuclear factor I: analyses of cellular binding sites. *Molecular and cellular biology* **5**, 964-971, doi:10.1128/mcb.5.5.964 (1985).
49 Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327-339, doi:10.1016/j.cell.2012.12.009 (2013).
50 Gronostajski, R. M. Roles of the NFI/CTF gene family in transcription and development. *Gene* **249**, 31-45, doi:10.1016/s0378-1119(00)00140-2 (2000).
51 Fletcher, C. F., Jenkins, N. A., Copeland, N. G., Chaudhry, A. Z. & Gronostajski, R. M. Exon structure of the nuclear factor I DNA-binding domain from C. elegans to mammals. *Mammalian genome : official journal of the International Mammalian Genome Society* **10**, 390-396 (1999).
52 Nagata, K., Guggenheimer, R. A., Enomoto, T., Lichy, J. H. & Hurwitz, J. Adenovirus DNA replication in vitro: identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 6438-6442, doi:10.1073/pnas.79.21.6438 (1982).
53 Steele-Perkins, G. *et al.* Essential role for NFI-C/CTF transcription-replication factor in tooth root development. *Molecular and cellular biology* **23**, 1075-1084, doi:10.1128/mcb.23.3.1075-1084.2003 (2003).
54 Steele-Perkins, G. *et al.* The transcription factor gene Nfib is essential for both lung maturation and brain development. *Molecular and cellular biology* **25**, 685-698, doi:10.1128/mcb.25.2.685-698.2005 (2005).
55 Campbell, C. E. *et al.* The transcription factor Nfix is essential for normal brain development. *BMC developmental biology* **8**, 52, doi:10.1186/1471-213x-8-52 (2008).
56 Fane, M., Harris, L., Smith, A. G. & Piper, M. Nuclear factor one transcription factors as epigenetic regulators in cancer. *International journal of cancer* **140**, 2634-2641, doi:10.1002/ijc.30603 (2017).
57 Mason, S., Piper, M., Gronostajski, R. M. & Richards, L. J. Nuclear factor one transcription factors in CNS development. *Molecular neurobiology* **39**, 10-23, doi:10.1007/s12035-008-8048-6 (2009).
58 Chen, K. S., Lim, J. W. C., Richards, L. J. & Bunt, J. The convergent roles of the nuclear factor I transcription factors in development and cancer. *Cancer letters* **410**, 124-138, doi:10.1016/j.canlet.2017.09.015 (2017).
59 Piper, M., Gronostajski, R. & Messina, G. Nuclear Factor One X in Development and Disease. *Trends in cell biology* **29**, 20-30, doi:10.1016/j.tcb.2018.09.003 (2019).
60 Driller, K. *et al.* Nuclear factor I X deficiency causes brain malformation and severe skeletal defects. *Molecular and cellular biology* **27**, 3855-3867, doi:10.1128/mcb.02293-06 (2007).
61 Shu, T., Butz, K. G., Plachez, C., Gronostajski, R. M. & Richards, L. J. Abnormal development of forebrain midline glia and commissural projections in Nfia knock-out mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **23**, 203-212 (2003).
62 Glasgow, S. M. *et al.* Mutual antagonism between Sox10 and NFIA regulates diversification of glial lineages and glioma subtypes. *Nature neuroscience* **17**, 1322-1329, doi:10.1038/nn.3790 (2014).
63 Kang, P. *et al.* Sox9 and NFIA coordinate a transcriptional regulatory cascade during the initiation of gliogenesis. *Neuron* **74**, 79-94, doi:10.1016/j.neuron.2012.01.024 (2012).
64 Grummt, I. Actin and myosin as transcription factors. *Current opinion in genetics & development* **16**, 191-196, doi:10.1016/j.gde.2006.02.001 (2006).
65 Miralles, F., Posern, G., Zaromytidou, A. I. & Treisman, R. Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell* **113**, 329-342, doi:10.1016/s0092-8674(03)00278-2 (2003).
66 Olson, E. N. & Nordheim, A. Linking actin dynamics and gene transcription to drive cellular motile functions. *Nature reviews. Molecular cell biology* **11**, 353-365, doi:10.1038/nrm2890 (2010).
67 de Lanerolle, P. Nuclear actin and myosins at a glance. *Journal of cell science* **125**, 4945-4949, doi:10.1242/jcs.099754 (2012).
68 Xie, X. *et al.* beta-Actin-dependent global chromatin organization and gene expression programs control cellular identity. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **32**, 1296-1314, doi:10.1096/fj.201700753R (2018).
69 Ito, T. *et al.* Identification of SWI.SNF complex subunit BAF60a as a determinant of the transactivation potential of Fos/Jun dimers. *The Journal of biological chemistry* **276**, 2852-2857, doi:10.1074/jbc.M009633200 (2001).
70 Vierbuchen, T. *et al.* AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Molecular cell* **68**, 1067-1082.e1012, doi:10.1016/j.molcel.2017.11.026 (2017).
71 Chen, Z. *et al.* Negative regulation of interferon-gamma/STAT1 signaling through cell adhesion and cell density-dependent STAT1 dephosphorylation. *Cellular signalling* **23**, 1404-1412, doi:10.1016/j.cellsig.2011.04.003 (2011).

72 Rambout, X., Dequiedt, F. & Maquat, L. E. Beyond Transcription: Roles of Transcription Factors in Pre-mRNA Splicing. *Chemical reviews* **118**, 4339-4364, doi:10.1021/acs.chemrev.7b00470 (2018).
73 Duskova, E., Hnilicova, J. & Stanek, D. CRE promoter sites modulate alternative splicing via p300-mediated histone acetylation. *RNA biology* **11**, 865-874, doi:10.4161/rna.29441 (2014).
74 Auboeuf, D. *et al.* Differential recruitment of nuclear receptor coactivators may determine alternative RNA splice site choice in target genes. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2270-2274, doi:10.1073/pnas.0308133100 (2004).
75 Auboeuf, D. *et al.* CoAA, a nuclear receptor coactivator protein at the interface of transcriptional coactivation and RNA splicing. *Molecular and cellular biology* **24**, 442-453, doi:10.1128/mcb.24.1.442-453.2004 (2004).
76 Wu, S. *et al.* A YY1-INO80 complex regulates genomic stability through homologous recombination-based repair. *Nature structural & molecular biology* **14**, 1165-1172, doi:10.1038/nsmb1332 (2007).
77 Hur, S. K. *et al.* Roles of human INO80 chromatin remodeling enzyme in DNA replication and chromosome segregation suppress genome instability. *Cellular and molecular life sciences : CMLS* **67**, 2283-2296, doi:10.1007/s00018-010-0337-3 (2010).
78 Chen, L. *et al.* Subunit organization of the human INO80 chromatin remodeling complex: an evolutionarily conserved core complex catalyzes ATP-dependent nucleosome remodeling. *The Journal of biological chemistry* **286**, 11283-11289, doi:10.1074/jbc.M111.222505 (2011).
79 Jin, J. *et al.* A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. *The Journal of biological chemistry* **280**, 41207-41212, doi:10.1074/jbc.M509128200 (2005).
80 Wang, J. *et al.* YY1 Positively Regulates Transcription by Targeting Promoters and Super-Enhancers through the BAF Complex in Embryonic Stem Cells. *Stem cell reports* **10**, 1324-1339, doi:10.1016/j.stemcr.2018.02.004 (2018).
81 Bonnet, J. *et al.* The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes & development* **28**, 1999-2012, doi:10.1101/gad.250225.114 (2014).
82 Atanassov, B. S. *et al.* Gcn5 and SAGA regulate shelterin protein turnover and telomere maintenance. *Molecular cell* **35**, 352-364, doi:10.1016/j.molcel.2009.06.015 (2009).
83 Evangelista, F. M. *et al.* Transcription and mRNA export machineries SAGA and TREX-2 maintain monoubiquitinated H2B balance required for DNA repair. *The Journal of cell biology* **217**, 3382-3397, doi:10.1083/jcb.201803074 (2018).
84 Riss, A. *et al.* Subunits of ADA-two-A-containing (ATAC) or Spt-Ada-Gcn5-acetyltrasferase (SAGA) Coactivator Complexes Enhance the Acetyltransferase Activity of GCN5. *The Journal of biological chemistry* **290**, 28997-29009, doi:10.1074/jbc.M115.668533 (2015).
85 Rodriguez-Navarro, S. *et al.* Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell* **116**, 75-86, doi:10.1016/s0092-8674(03)01025-0 (2004).
86 Spedale, G., Timmers, H. T. & Pijnappel, W. W. ATAC-king the complexity of SAGA during evolution. *Genes & development* **26**, 527-541, doi:10.1101/gad.184705.111 (2012).
87 Wang, Y. *et al.* KAT2A coupled with the alpha-KGDH complex acts as a histone H3 succinyltransferase. *Nature* **552**, 273-277, doi:10.1038/nature25003 (2017).
88 Liu, X., Tesfai, J., Evrard, Y. A., Dent, S. Y. & Martinez, E. c-Myc transformation domain recruits the human STAGA complex and requires TRRAP and GCN5 acetylase activity for transcription activation. *The Journal of biological chemistry* **278**, 20405-20412, doi:10.1074/jbc.M211795200 (2003).
89 Li, D. *et al.* Kruppel-like factor-6 promotes preadipocyte differentiation through histone deacetylase 3-dependent repression of DLK1. *The Journal of biological chemistry* **280**, 26941-26952, doi:10.1074/jbc.M500463200 (2005).
90 Doyon, Y., Selleck, W., Lane, W. S., Tan, S. & Cote, J. Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Molecular and cellular biology* **24**, 1884-1896, doi:10.1128/mcb.24.5.1884-1896.2004 (2004).
91 Frank, S. R. *et al.* MYC recruits the TIP60 histone acetyltransferase complex to chromatin. *EMBO reports* **4**, 575-580, doi:10.1038/sj.embor.embor861 (2003).
92 Daigo, K. *et al.* Proteomic analysis of native hepatocyte nuclear factor-4alpha (HNF4alpha) isoforms, phosphorylation status, and interactive cofactors. *The Journal of biological chemistry* **286**, 674-686, doi:10.1074/jbc.M110.154732 (2011).
93 St-Denis, N. *et al.* Phenotypic and Interaction Profiling of the Human Phosphatases Identifies Diverse Mitotic Regulators. *Cell reports* **17**, 2488-2501, doi:10.1016/j.celrep.2016.10.078 (2016).
94 Goos, H. *et al.* Gain-of-function CEBPE mutation causes noncanonical autoinflammatory inflammasomopathy. *The Journal of allergy and clinical immunology*, doi:10.1016/j.jaci.2019.06.003 (2019).
95 Kaustio, M. *et al.* Damaging heterozygous mutations in NFKB1 lead to diverse immunologic phenotypes. *The Journal of allergy and clinical immunology* **140**, 782-796, doi:10.1016/j.jaci.2016.10.054 (2017).
96 Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251, doi:10.1016/j.cell.2013.02.014 (2013).
97 Choi, H. *et al.* SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature methods* **8**, 70-73, doi:10.1038/nmeth.1541 (2011).
98 Simicevic, J. & Deplancke, B. Transcription factor proteomics-Tools, applications, and challenges. *Proteomics* **17**, doi:10.1002/pmic.201600317 (2017).
99 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
100 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
101 Pundir, S., Martin, M. J. & O'Donovan, C. UniProt Protein Knowledgebase. *Methods in molecular biology (Clifton, N.J.)* **1558**, 41-55, doi:10.1007/978-1-4939-6783-4_2 (2017).
102 A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS biology* **9**, e1001046, doi:10.1371/journal.pbio.1001046 (2011).
103 Ayala, R. *et al.* Structure and regulation of the human INO80-nucleosome complex. *Nature* **556**, 391-395, doi:10.1038/s41586-018-0021-6 (2018).
104 Gao, P. *et al.* Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. *Cell* **174**, 576-589.e518, doi:10.1016/j.cell.2018.06.003 (2018).
105 Andrews, S. (2010).
106 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
107 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
108 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

109 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).
110 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
111 Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

Author Contributions

HG, MK and MV designed the study. HG and MK with help of LY generated the cell lines and HG and MK performed the mass spectrometry analyses. HG did the data filtering, data-analyses, NFIA silencing experiments, prepared the figures and wrote the manuscript. KS participated in data analysis. GW, ZT and QZ designed and performed the ChIP-Seq analyses. All the authors reviewed the manuscript carefully.

Acknowledgements

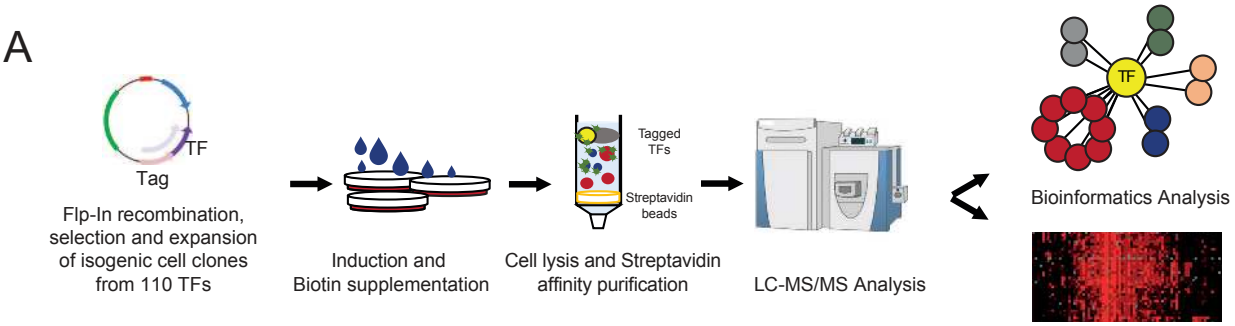
For the founding, we would like to thank Orion Research Foundation sr, Finnish Cultural Foundation (HG), Academy of Finland (288475 and 294173), Sigrid Jusélius Foundation, Biocentrum Helsinki, University of Helsinki Three-year Research Grant, Biocentrum Finland, HiLIFE, Instrumentarium Research Foundation (MV), Jane and Aatos Erkko Foundation and Finnish Cancer Foundation (GW).

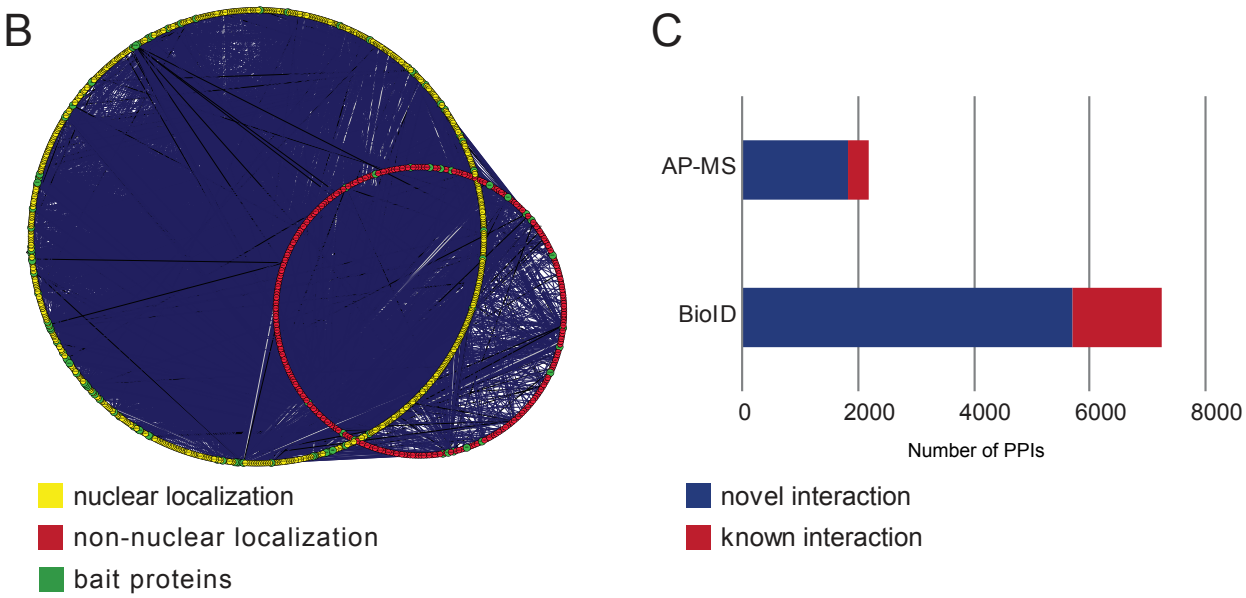
We would also like to thank Petri Auvinen and Juha Kere for the critical reading of the manuscript.

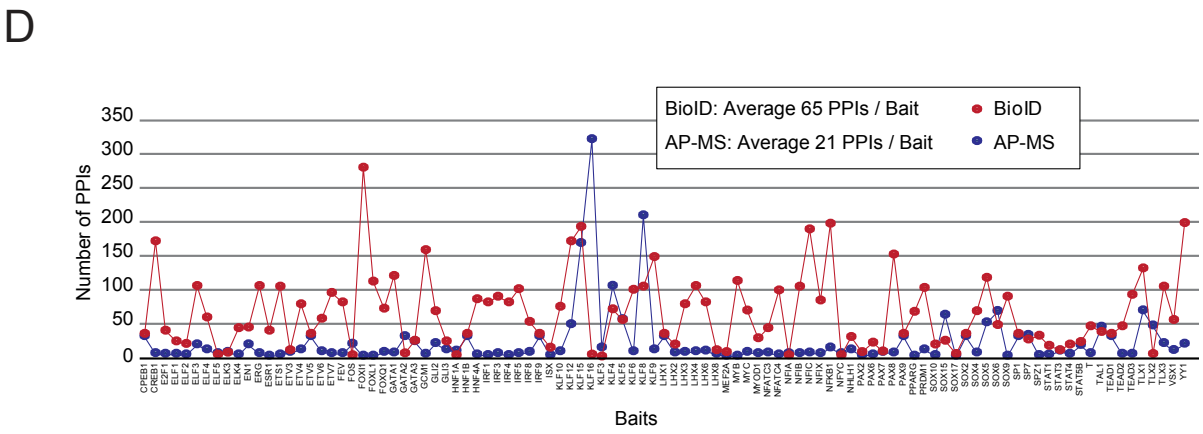
Competing interests

Authors report no competing interests.

887 **FIGURES**







888

889 **Figure 1. TF protein interactome identified using the BioID and AP-MS methods**

890 **A**. Schematic illustration of used methods. TFs were tagged N-terminally with a MAC, StrepIII-HA
891 or BirA -tags (Table S1A) and co-transfected with Flp-In recombinase to generate stable isogenic and
892 inducible cell lines. Cells were induced by tetracycline addition and, for the BioID analysis,
893 supplemented with biotin for 24 hours. This was followed by harvesting, lysis and affinity purification
894 with Streptavidin beads. Purified proteins were further digested into peptides and analysed by LC-
895 MS/MS. Proteins were later identified and analysed using different bioinformatic methods.

896 **B.** Localisation of interacting prey-proteins according the annotated localisations of Cell Atlas [12].
897 Yellow nodes indicate nuclear localisation and red non-nuclear. From mapped proteins, more than
898 80% had nuclear localisation.

899 **C**. Protein-protein interactions identified using the AP-MS (2176) and BioID (7232) methods.
900 Interactions were compared to interactions from the PINA2, Intact, Biogrid and String experimental
901 protein interaction databases and to interactions from a study by Li et al., resulting in 345 and 1525
902 previously reported interactions in the AP-MS and BioID data, respectively. The proportions of
903 known interactions are shown in red.

904 **D**. Number of high-confidence protein-protein interactions of different TF baits detected by AP-MS
905 (blue) or BioID (red) affinity purification combined to mass spectrometry.

906

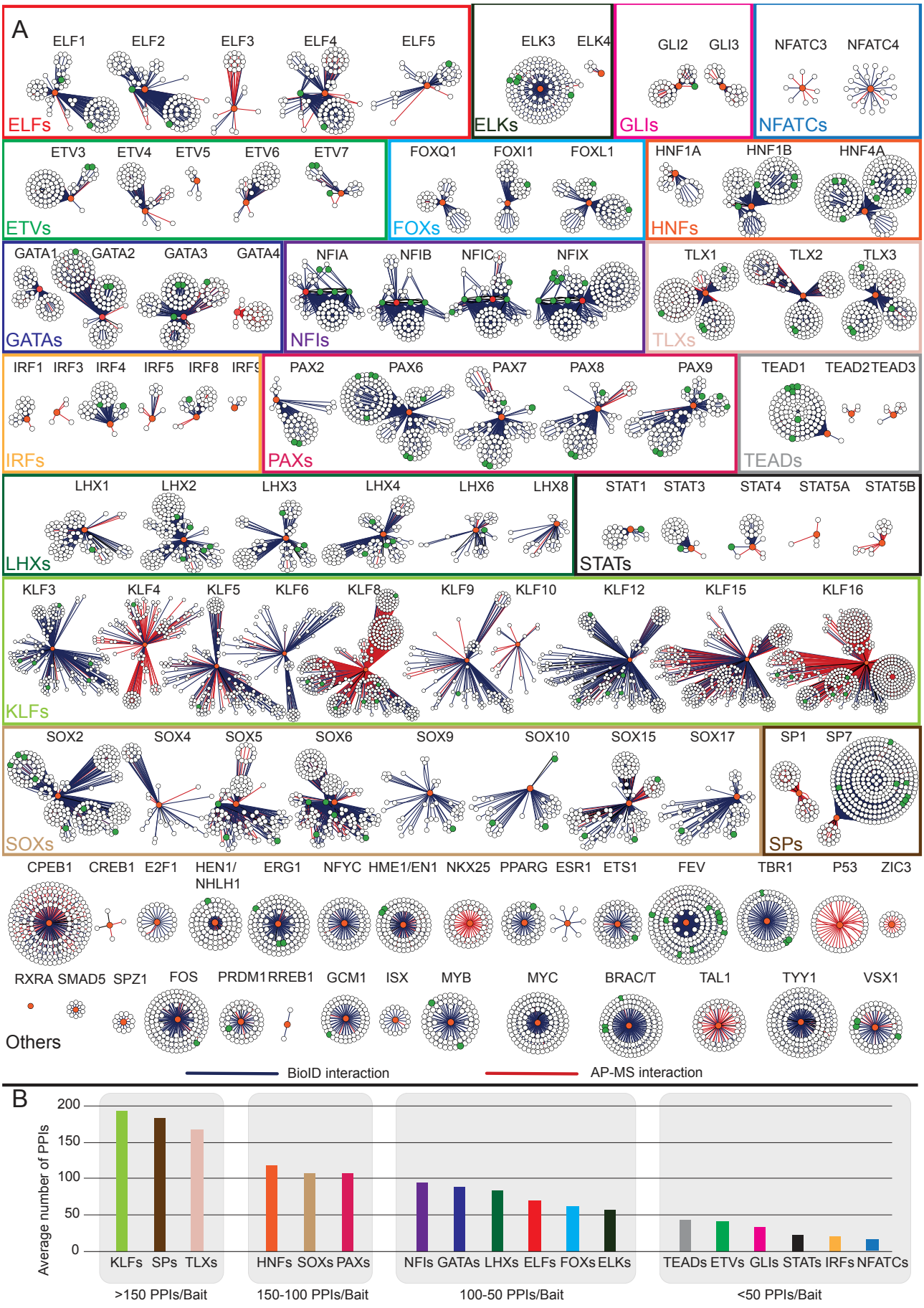

A
ELF1
ELF2
ELF3
ELF4
ELF5
ELFs
ELK3
ELK4
ELKs
GLI2
GLI3
GLIs
NFATC3
NFATC4
NFATCs
ETV3
ETV4
ETV5
ETV6
ETV7
ETVs
FOXQ1
FOXI1
FOXL1
FOXs
HNF1A
HNF1B
HNF4A
HNFs
GATA1
GATA2
GATA3
GATA4
GATAs
NFIA
NFIB
NFIC
NFIX
NFIs
TLX1
TLX2
TLX3
TLXs
IRF1
IRF3
IRF4
IRF5
IRF8
IRFs
PAX2
PAX6
PAX7
PAX8
PAX9
PAXs
TEAD1
TEAD2
TEAD3
TEADs
LHX1
LHX2
LHX3
LHX4
LHX6
LHX8
LHXs
STAT1
STAT3
STAT4
STAT5A
STAT5B
STATs
KLF3
KLF4
KLF5
KLF6
KLF8
KLF9
KLF10
KLF12
KLF15
KLF16
KLFs
SOX2
SOX4
SOX5
SOX6
SOX9
SOX10
SOX15
SOX17
SOXs
SP1
SP7
SPs
CPEB1
CREB1
E2F1
HEN1/
NHLH1
ERG1
NFYC
HME1/EN1
NKX25
PPARG
ESR1
ETS1
FEV
TBR1
P53
ZIC3
RXRA
SMAD5
SPZ1
FOS
PRDM1
RREB1
GCM1
ISX
MYB
MYC
BRAC/T
TAL1
TYY1
VSX1
Others
BioID interaction
AP-MS interaction
B
Average number of PPIs
200
150
100
50
0
KLFs SPs TLXs
HNFs SOXs PAXs
NFIs GATAs LHXs ELFs FOXs ELKs
TEADs ETVs GLIs STATs IRFs NFATCs
>150 PPIs/Bait
150-100 PPIs/Bait
100-50 PPIs/Bait
<50 PPIs/Bait

Figure 2. Comprehensive protein interactomes of the studied TF families

A. TFs belonging to a given family are indicated in orange nodes, other interacting TFs in green and the rest of the interacting proteins in white. Blue edges indicate interactions from the BioID analysis, red from the AP-MS analysis and black from both.

B. The average number of PPIs of different TF families are shown under the interaction maps.

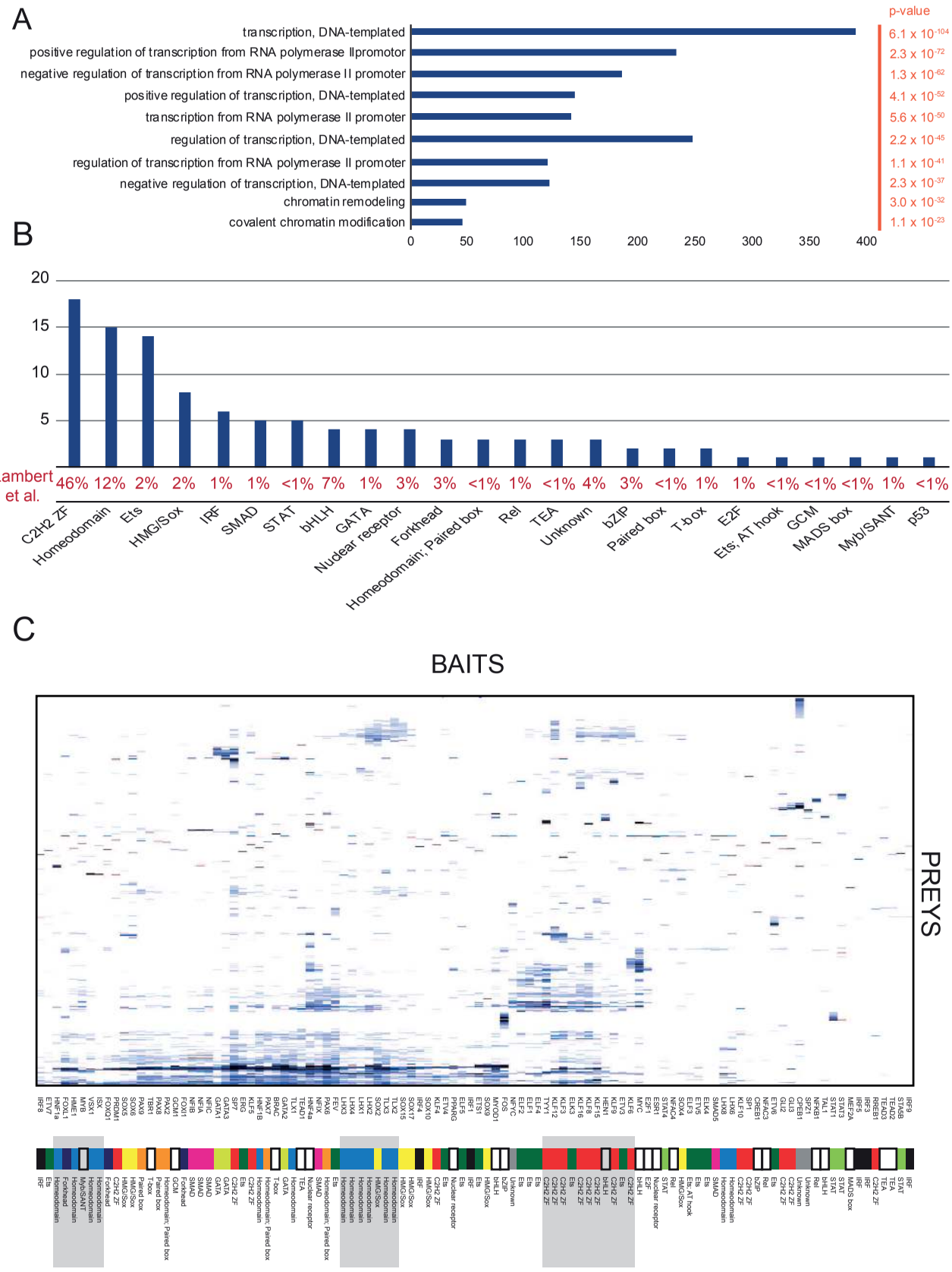

A
p-value
transcription, DNA-templated
positive regulation of transcription from RNA polymerase IIpromotor
negative regulation of transcription from RNA polymerase II promoter
positive regulation of transcription, DNA-templated
transcription from RNA polymerase II promoter
regulation of transcription, DNA-templated
regulation of transcription from RNA polymerase II promoter
negative regulation of transcription, DNA-templated
chromatin remodeling
covalent chromatin modification
6.1 x 10-104
2.3 x 10-72
1.3 x 10-62
4.1 x 10-52
5.6 x 10-50
2.2 x 10-45
1.1 x 10-41
2.3 x 10-37
3.0 x 10-32
1.1 x 10-23
0
50
100
150
200
250
300
350
400
B
20
15
10
5
Lambert et al.
46% 12% 2% 2% 1% 1% <1% 7% 1% 3% 3% <1% 1% <1% 4% 3% <1% 1% 1% <1% <1% <1% 1% <1%
C2H2 ZF
Homeodomain
Ets
HMG/Sox
IRF
SMAD
STAT
bHLH
GATA
Nuclear receptor
Forkhead
Homeodomain; Paired box
Rel
TEA
Unknown
bZIP
Paired box
T-box
E2F
Ets; AT hook
GCM
MADS box
Myb/SANT
p53
C
BAITS
PREYS

Figure 3. Hierarchical clustering of baits by preys and its correlations to DNA-binding domains and protein sequences.

A. Enriched GO-BP terms of interacting proteins from the BioID analysis.

B. The distribution of the DNA-binding domains of the studied TFs. The corresponding proportion of each DNA-binding domain from 1,639 TFs in the study of Lambert et al. is shown as a percentage value below the graph.

C. TFs (baits, named below the heatmap) and their interacting proteins (preys) were hierarchically clustered (Biohit-viz). Corresponding colour coded DNA -binding domains are shown below the baits.

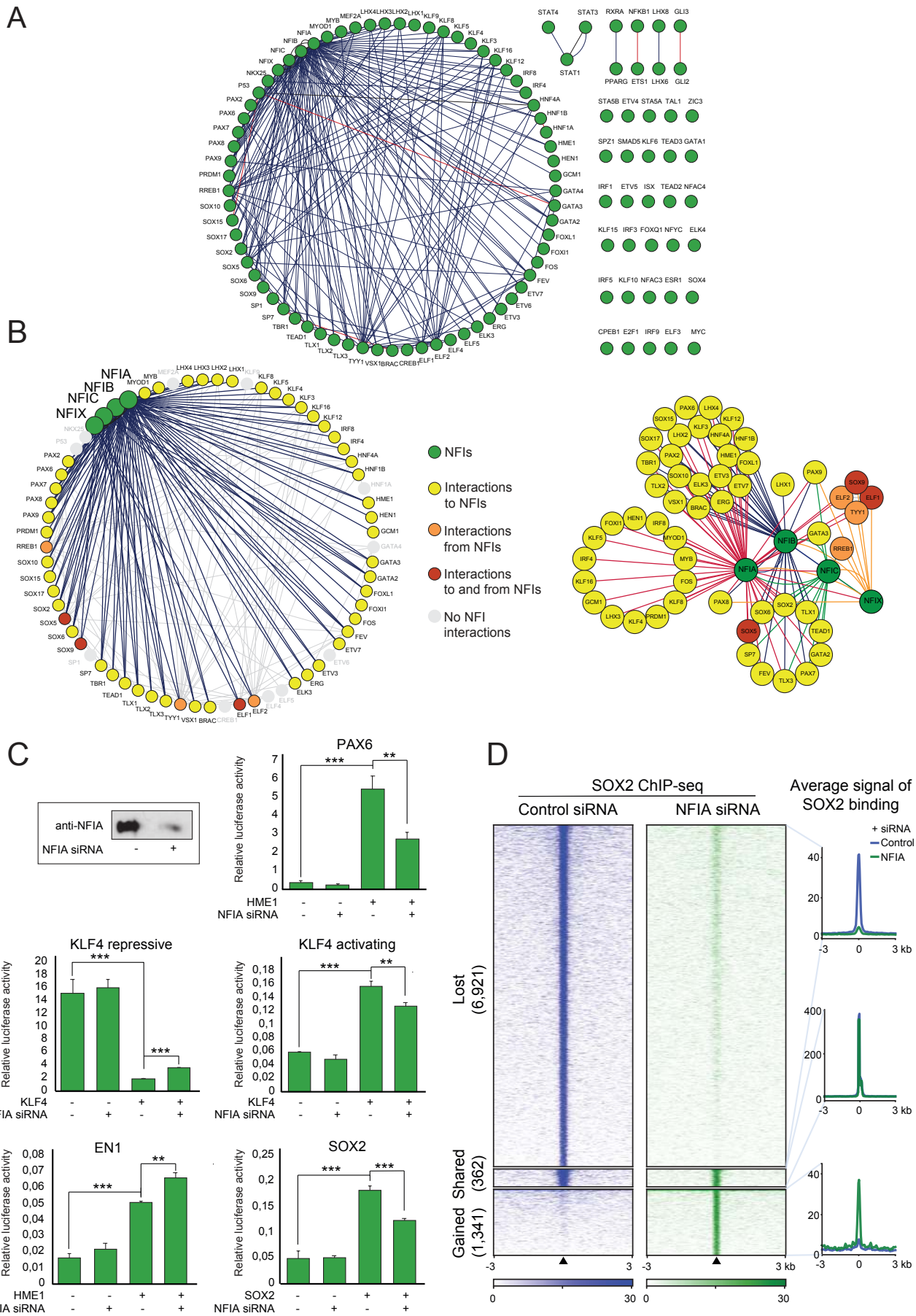

A
B
NFIA
NFIB
NFIC
NFIX
NFIs
Interactions to NFIs
Interactions from NFIs
Interactions to and from NFIs
No NFI interactions
C
anti-NFIA
NFIA siRNA
PAX6
Relative luciferase activity
HME1
KLF4 repressive
KLF4 activating
KLF4
EN1
SOX2
D
SOX2 ChIP-seq
Control siRNA
NFIA siRNA
Average signal of SOX2 binding
+ siRNA
Control
NFIA
Lost (6,921)
Gained Shared (1,341) (362)

Figure 4. TF-TF (bait-bait) interactions of 110 TFs studied

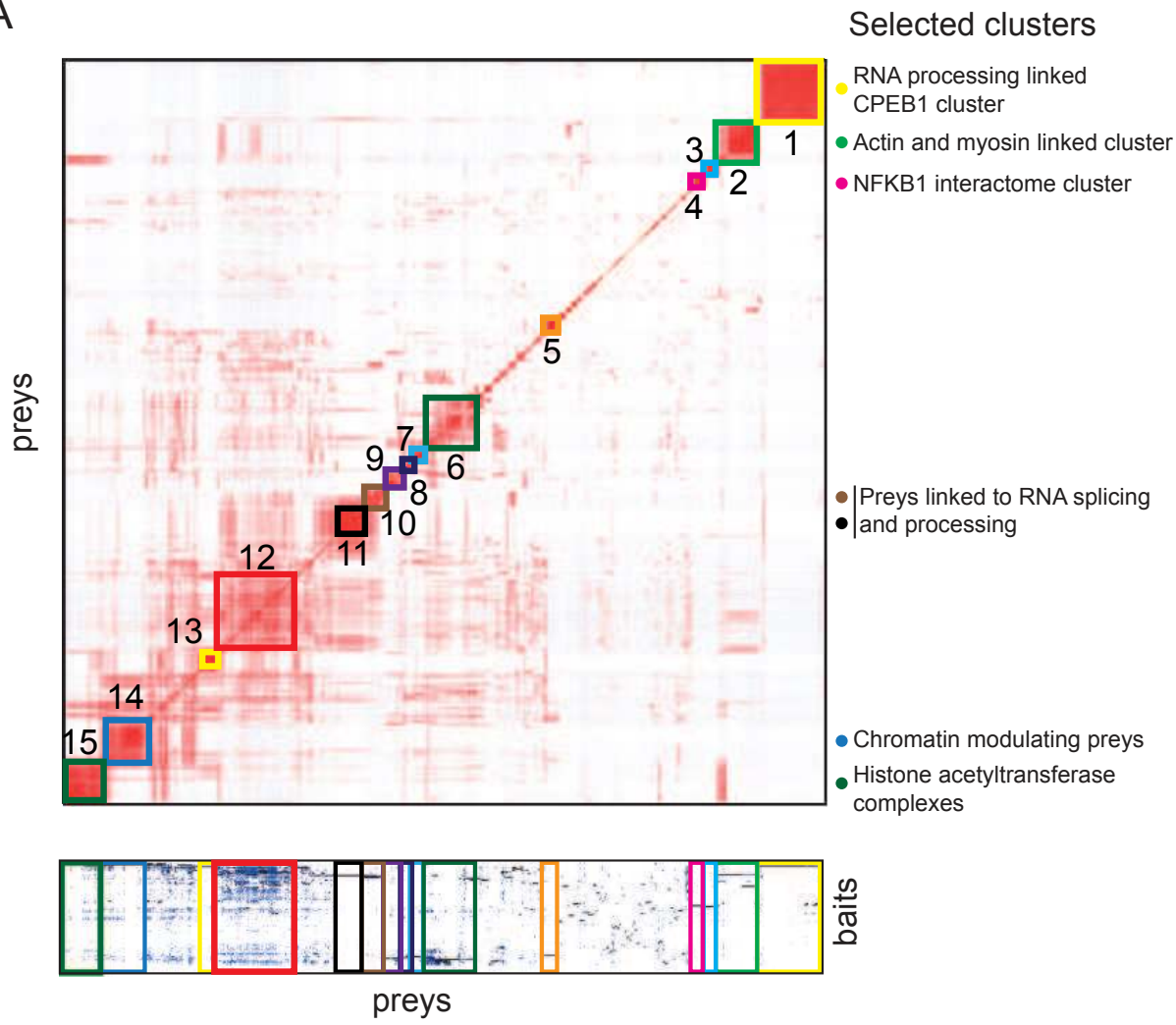
A. Of 110 studied TFs, 80 had 203 interactions with other studied TFs. Blue edges indicate interactions from the BioID analysis, red from the AP-MS analysis and black from both.

B. Most of these TF-TF (175) interactions were TF interactions with NFIs (left panel). The right panel shows the separate groups shared by one or multiple NFIs. Colour code: Green nodes = NFIs, yellow nodes = interactions to NFIs, orange nodes = interactions from NFIs, red nodes = interactions to and from NFIs and grey nodes = no interactions to or from NFIs. Colour coding of the nodes is shown in the right side of the figure.

C. NFIA was silenced using siRNA transfection and NFIA levels were detected 48 hours after transfection by western blotting using specific antibody against NFIA. TFs' activity was investigated after NFIA silencing using both repressive and activating reporter gene analysis. Both repressing and activating functions of KLF4 were reduced upon the NFIA silencing. In addition, SOX2 and PAX6 activity was reduced, while EN1 activity was increased upon NFIA silencing. N=3, ***: $p<0.001$, **: $p<0.01$, *:$p<0.05$

D. Heatmap representation of SOX2 binding intensity based on ChIP-seq signals in 293T cells while treated with siControl and siNFIA, respectively. Signals within 3 kb around the center of binding peaks are displayed in a descending order for each SOX2 binding event (lost, shared, and gained upon siNFIA). (Right) Plots of average signal of SOX2 binding at each corresponding region.

A


Selected clusters
RNA processing linked CPEB1 cluster
Actin and myosin linked cluster
NFKB1 interactome cluster
Preys linked to RNA splicing and processing
Chromatin modulating preys
Histone acetyltransferase complexes
preys
baits
preys
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

B

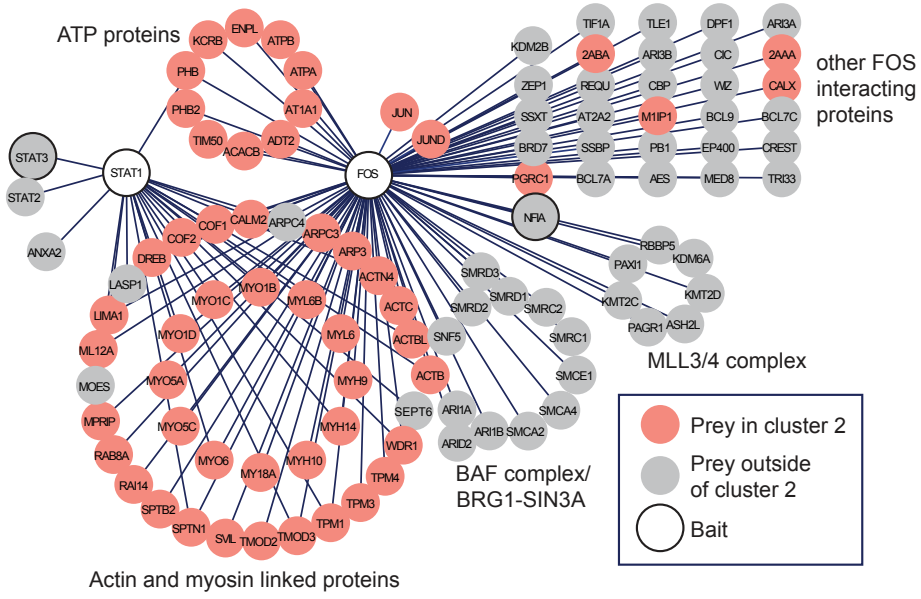

Cluster 2: STAT1 and FOS interactomes
ATP proteins
other FOS interacting proteins
MLL3/4 complex
BAF complex/ BRG1-SIN3A
Actin and myosin linked proteins
Prey in cluster 2
Prey outside of cluster 2
Bait

Figure 5. Biological clusters from prey-prey correlation analysis

A. Prey-prey correlation analysis (ProHits-viz) on preys identified in the BioID experiments. The results are shown in the heatmap with preys in both the x and y axis. A corresponding bait-prey heat map is shown below to assist in determining which and how many baits are driving the prey clusters. Preys in clusters (1-15) and baits driving the clusters are shown in Table S7. Color scale indicates Pearson correlation of prey PSMs that was done by Prohits-viz prior to hierarchical clustering.

B. FOS and STAT1 interactomes. Many proteins interacting with FOS and STAT1 were clustered in Cluster 2 of the prey-prey correlation analysis (highlighted in red). Actin and myosin, ATP signalling, BAF complex and MLL3/4 complex linked proteins are highlighted in different groups.

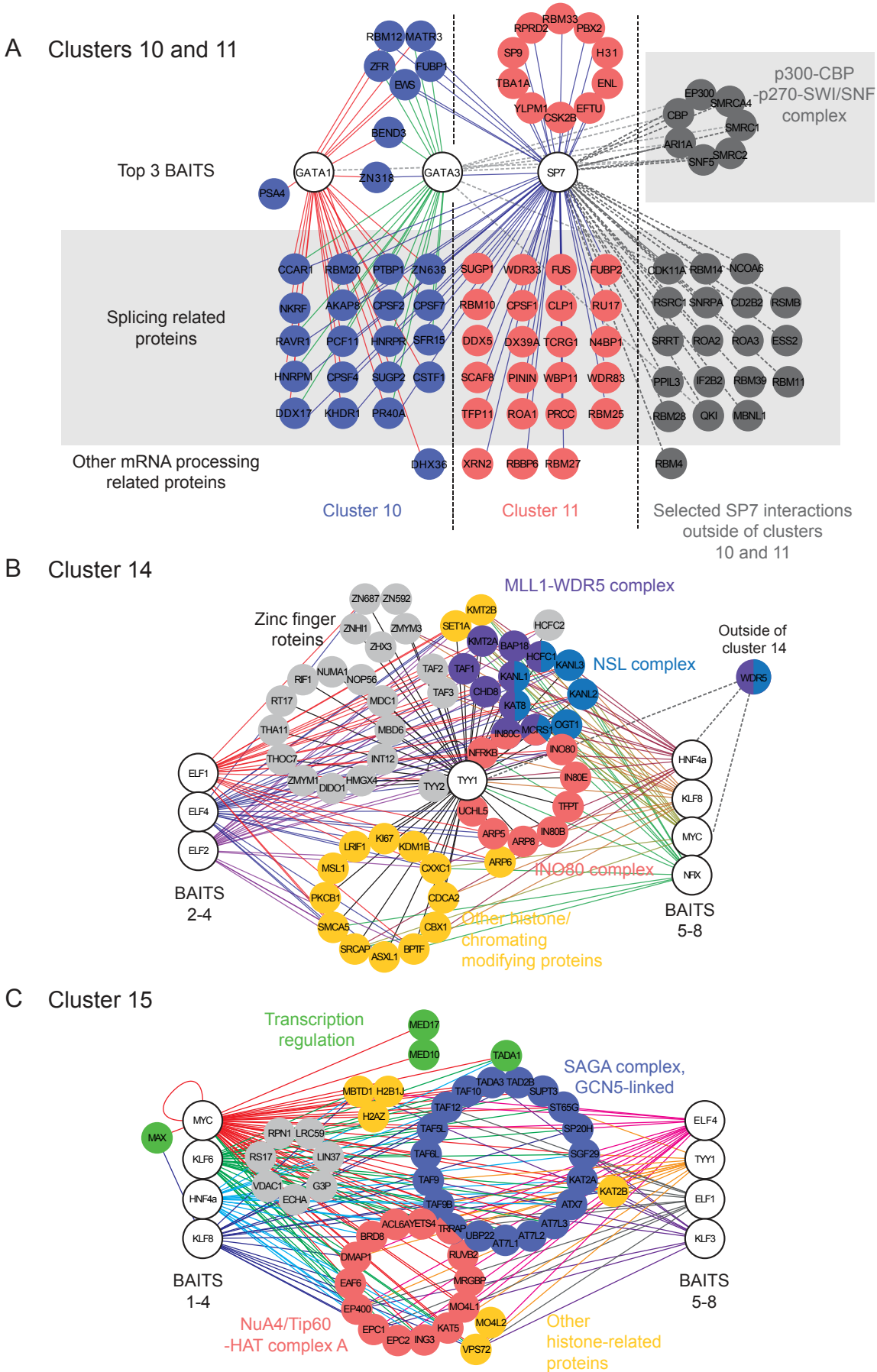

A Clusters 10 and 11
Top 3 BAITS
GATA1
GATA3
SP7
ZN318
PSA4
BEND3
RBM12
MATR3
ZFR
FUBP1
EWS
RBM33
RPRD2
PBX2
SP9
H31
TBA1A
ENL
YLPM1
CSK2B
EFTU
p300-CBP
-p270-SWI/SNF
complex
EP300
SMRCA4
CBP
SMRC1
ARI1A
SMRC2
SNF5
Splicing related
proteins
CCAR1
RBM20
PTBP1
ZN638
SUGP1
WDR33
FUS
FUBP2
CDK11A
RBM14
NCOA6
NKRF
AKAP8
CPSF2
CPSF7
RBM10
CPSF1
CLP1
RU17
RSRC1
SNRPA
CD2B2
RSMB
RAVR1
PCF11
HNRPR
SFR15
DDX5
DX39A
TCRG1
N4BP1
SRRT
ROA2
ROA3
ESS2
HNRPM
CPSF4
SUGP2
CSTF1
SCAF8
PININ
WBP11
WDR83
PPIL3
IF2B2
RBM39
RBM11
DDX17
KHDR1
PR40A
TFP11
ROA1
PRCC
RBM25
RBM28
QKI
MBNL1
Other mRNA processing
related proteins
DHX36
XRN2
RBBP6
RBM27
RBM4
Cluster 10
Cluster 11
Selected SP7 interactions
outside of clusters
10 and 11
B Cluster 14
MLL1-WDR5 complex
Zinc finger
roteins
ZN687
ZN592
ZNHI1
ZMYM3
ZHX3
KMT2B
SET1A
KMT2A
HCFC2
BAP18
HCFC1
Outside of
cluster 14
NSL complex
KANL3
KANL1
KANL2
WDR5
RIF1
NUMA1
NOP56
TAF2
TAF1
TAF3
CHD8
KAT8
RT17
MDC1
THA11
MBD6
OGT1
MCRS1
IN80C
INO80
NFRKB
THOC7
INT12
ZMYM1
DIDO1
HMGX4
TYY2
TYY1
IN80E
TFPT
UCHL5
ARP5
ARP8
IN80B
ARP6
LRIF1
KI67
KDM1B
MSL1
CXXC1
INO80 complex
PKCB1
CDCA2
SMCA5
CBX1
SRCAP
ASXL1
BPTF
Other histone/
chromating
modifying proteins
ELF1
ELF4
ELF2
BAITS
2-4
HNF4a
KLF8
MYC
NFX
BAITS
5-8
C Cluster 15
Transcription
regulation
MED17
MED10
TADA1
SAGA complex,
GCN5-linked
TADA3
TAD2B
SUPT3
TAF10
TAF12
ST65G
SP20H
TAF5L
TAF6L
SGF29
TAF9
KAT2A
KAT2B
TAF9B
ATX7
AT7L3
AT7L2
AT7L1
UBP22
TRRAP
ACL6A
YETS4
BRD8
DMAP1
RUVB2
EAF6
MRGBP
EP400
MO4L1
MO4L2
EPC1
EPC2
ING3
KAT5
VPS72
NuA4/Tip60
-HAT complex A
Other
histone-related
proteins
MYC
MAX
KLF6
HNF4a
KLF8
BAITS
1-4
MBTD1
H2B1J
H2AZ
RPN1
LRC59
RS17
LIN37
VDAC1
G3P
ECHA
ELF4
TYY1
ELF1
KLF3
BAITS
5-8

Figure 6. The interactions in Clusters 10, 11, 14 and 15 of the prey-prey correlation analysis

A. The three baits (GATA1, GATA3, SP7) with the most interactions in Clusters 10 and 11 are shown in the centre of the figure with white nodes. Most of the preys in Clusters 10 (blue nodes) and 11 (red nodes) were linked to mRNA splicing (highlighted in the grey box). Moreover, SP7 interactions linked to mRNA splicing, mRNA processing and the p300-CBP-p270-SWI/SNF complex found outside of Clusters 10 and 11 are presented on the right side of the figure (grey nodes).

B. Protein-protein interactions in Cluster 14. The eight baits with the most interactions are highlighted in white and the interacting protein complexes are colour coded. WDR5's (outside of Cluster 14) interactions are shown with dashed lines in the upper right corner of the figure.

C. Protein-protein interactions in Cluster 15. The eight baits with the most interactions are highlighted in white and the interacting protein complexes are colour coded.

Figures

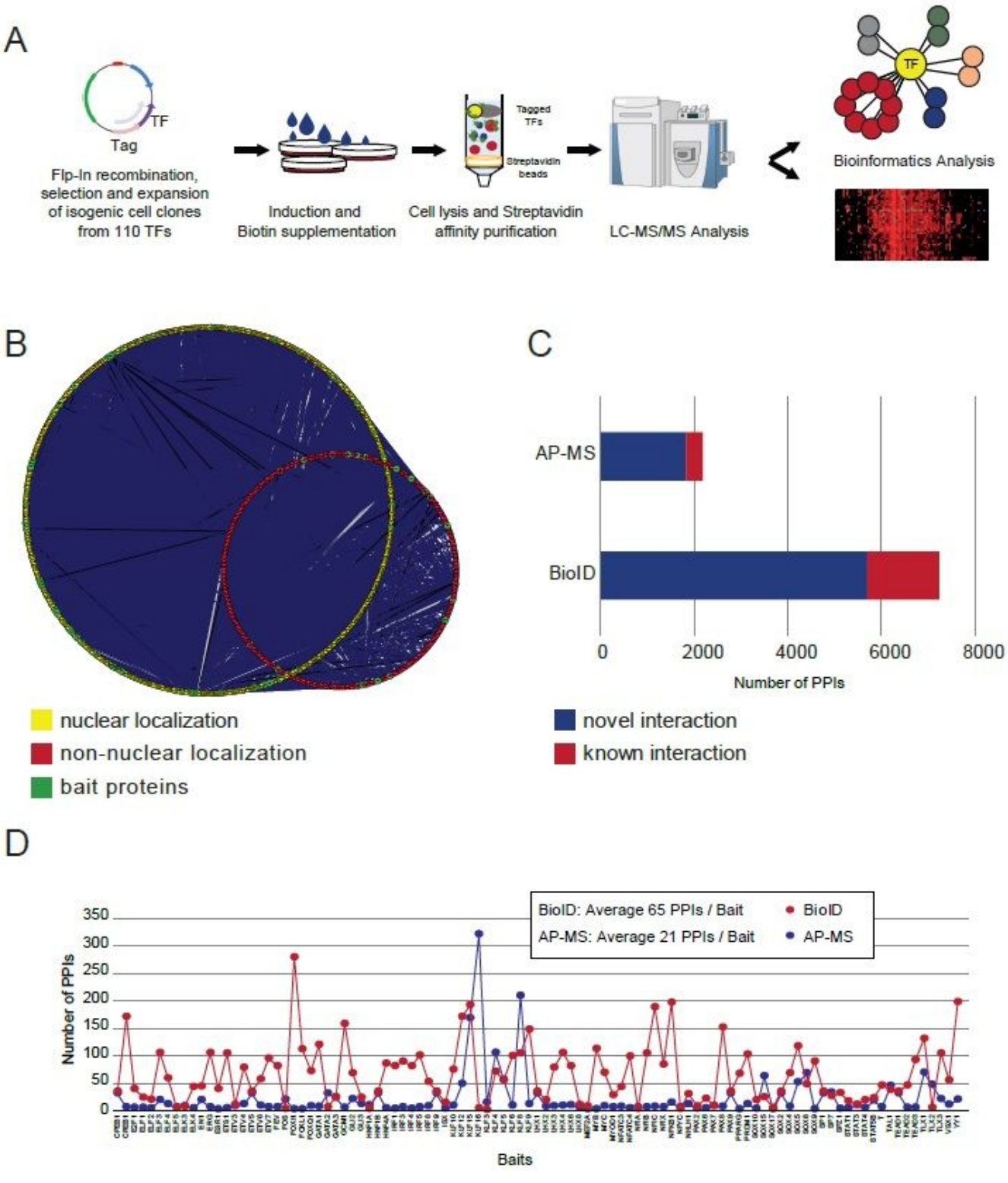


Figure 1

TF protein interactome identified using th 889 e BioID and AP-MS methods A. Schematic illustration of used methods. TFs were tagged N-terminally with a MAC, StrepIII-HA or BirA -tags (Table S1A) and co-transfected with Flp-In recombinase to generate stable isogenic and inducible cell lines. Cells were

induced by tetracycline addition and, for the BioID analysis, supplemented with biotin for 24 hours. This was followed by harvesting, lysis and affinity purification with Streptavidin beads. Purified proteins were further digested into peptides and analysed by LC MS/MS. Proteins were later identified and analysed using different bioinformatic methods. B. Localisation of interacting prey-proteins according the annotated localisations of Cell Atlas 12. Yellow nodes indicate nuclear localisation and red non-nuclear. From mapped proteins, more than 80% had nuclear localisation. C. Protein-protein interactions identified using the AP-MS (2176) and BioID (7232) methods. Interactions were compared to interactions from the PINA2, Intact, Biogrid and String experimentaln protein interaction databases and to interactions from a study by Li et al., resulting in 345 and 1525 previously reported interactions in the AP-MS and BioID data, respectively. The proportions of known interactions are shown in red. D. Number of high-confidence protein-protein interactions of different TF baits detected by AP-MS (blue) or BioID (red) affinity purification combined to mass spectrometry.

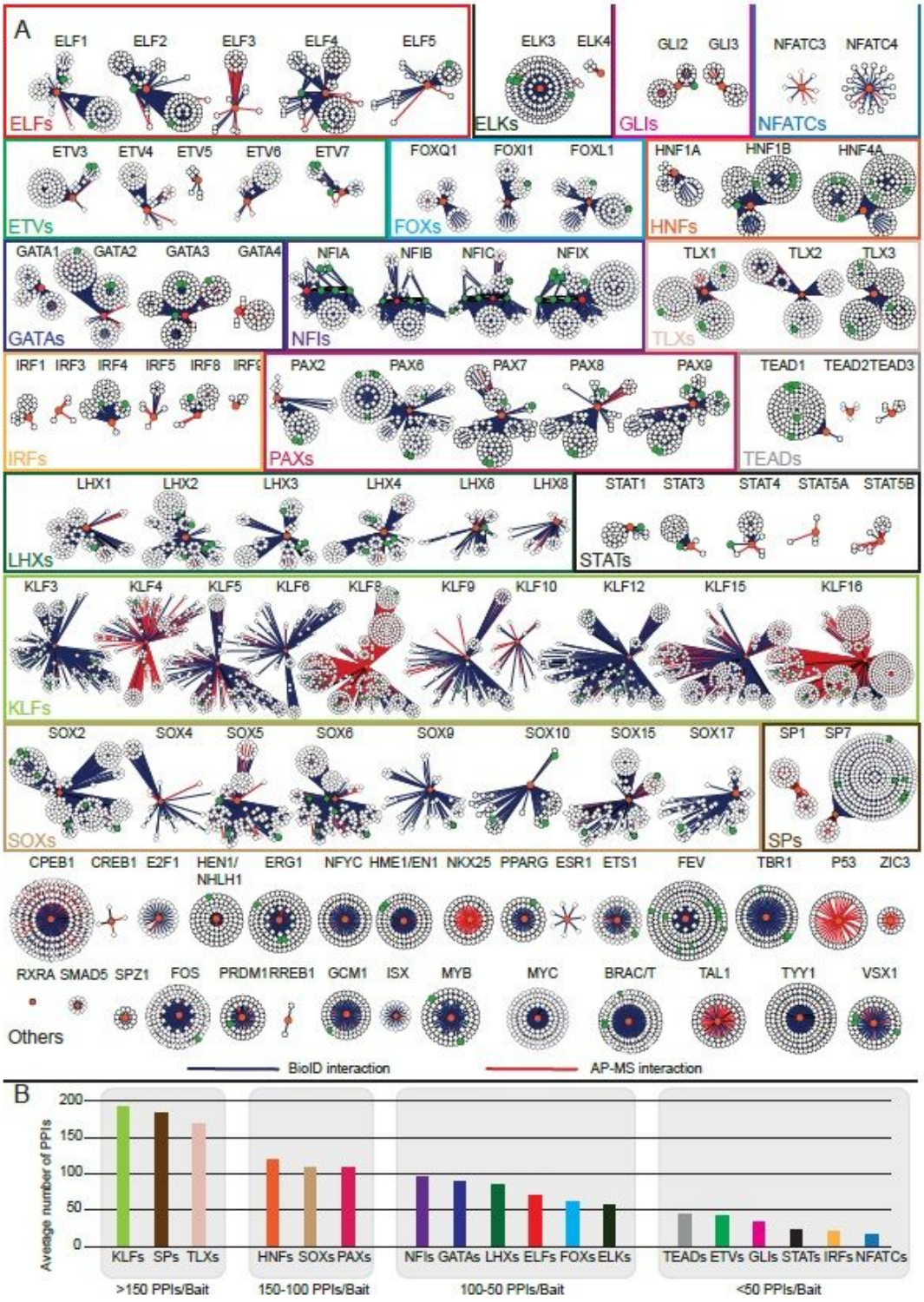


Figure 2

Comprehensive protein interactomes 908 of the studied TF families A. TFs belonging to a given family are indicated in orange nodes, other interacting TFs in green and the rest of the interacting proteins in white. Blue edges indicate interactions from the BioID analysis, red from the AP-MS analysis and black from both. B. The average number of PPIs of different TF families are shown under the interaction maps.

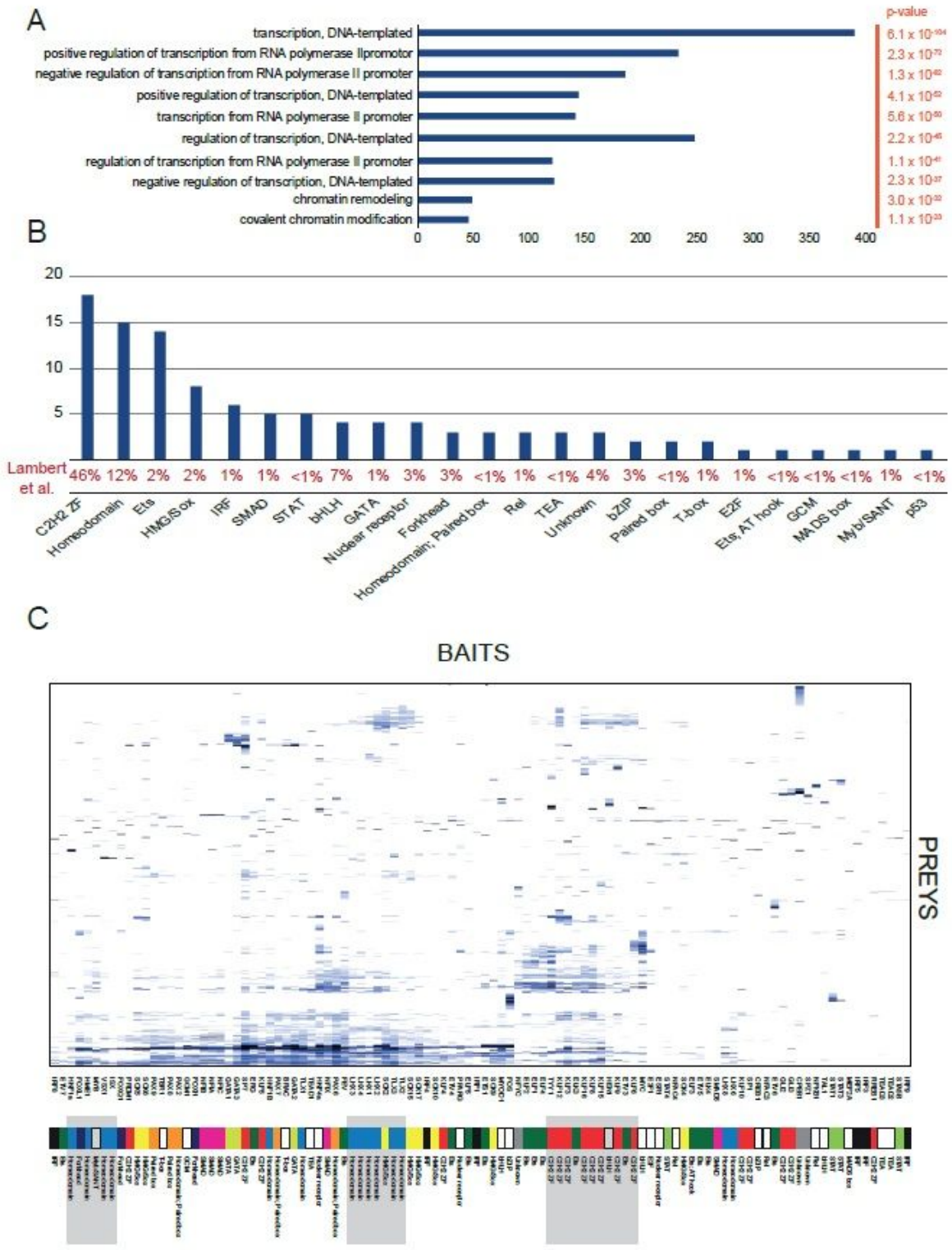


Figure 3

Hierarchical clustering of baits by preys and its correlations 914 to DNA-binding domains and protein sequences. A. Enriched GO-BP terms of interacting proteins from the BioID analysis. B. The distribution of the DNA-binding domains of the studied TFs. The corresponding proportion of each DNA-binding domain from 1,639 TFs in the study of Lambert et al. is shown as a percentage value below the graph. C. TFs

(baits, named below the heatmap) and their interacting proteins (preys) were hierarchically clustered (Biohit-viz). Corresponding colour coded DNA -binding domains are shown below the baits.

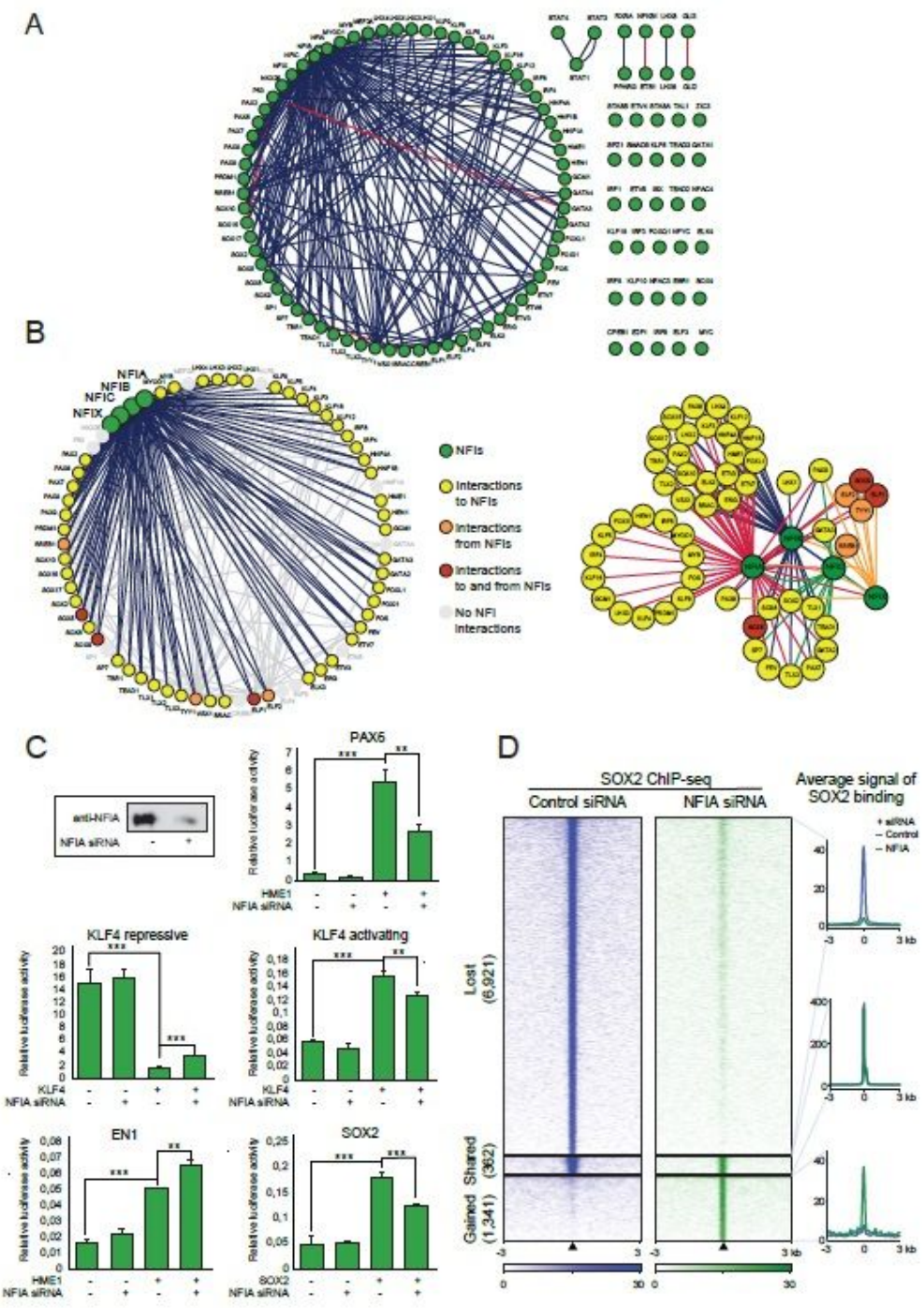


Figure 4

TF-TF (bait-924 bait) interactions of 110 TFs studied A. Of 110 studied TFs, 80 had 203 interactions with other studied TFs. Blue edges indicate interactions from the BioID analysis, red from the AP-MS analysis and black from both. B. Most of these TF-TF (175) interactions were TF interactions with NFIs (left

panel). The right panel shows the separate groups shared by one or multiple NFIs. Colour code: Green nodes = NFIs, yellow nodes = interactions to NFIs, orange nodes = interactions from NFIs, red nodes = interactions to and from NFIs and grey nodes = no interactions to or from NFIs. Colour coding of the nodes is shown in the right side of the figure. C. NFIA was silenced using siRNA transfection and NFIA levels were detected 48 hours after transfection by western blotting using specific antibody against NFIA. TFs' activity was investigated after NFIA silencing using both repressive and activating reporter gene analysis. Both repressing and activating functions of KLF4 were reduced upon the NFIA silencing. In addition, SOX2 and PAX6 activity was reduced, while EN1 activity was increased upon NFIA silencing. N=3, ***: $p<0.001$, **: $p<0.01$, *:$p<0.05$ D. Heatmap representation of SOX2 binding intensity based on ChIP-seq signals in 293T cells whil treated with siControl and siNFIA, respectively. Signals within 3 kb around the center of binding peaks are displayed in a descending order for each SOX2 binding event (lost, shared, and gained upon siNFIA). (Right) Plots of average signal of SOX2 binding at each corresponding region.

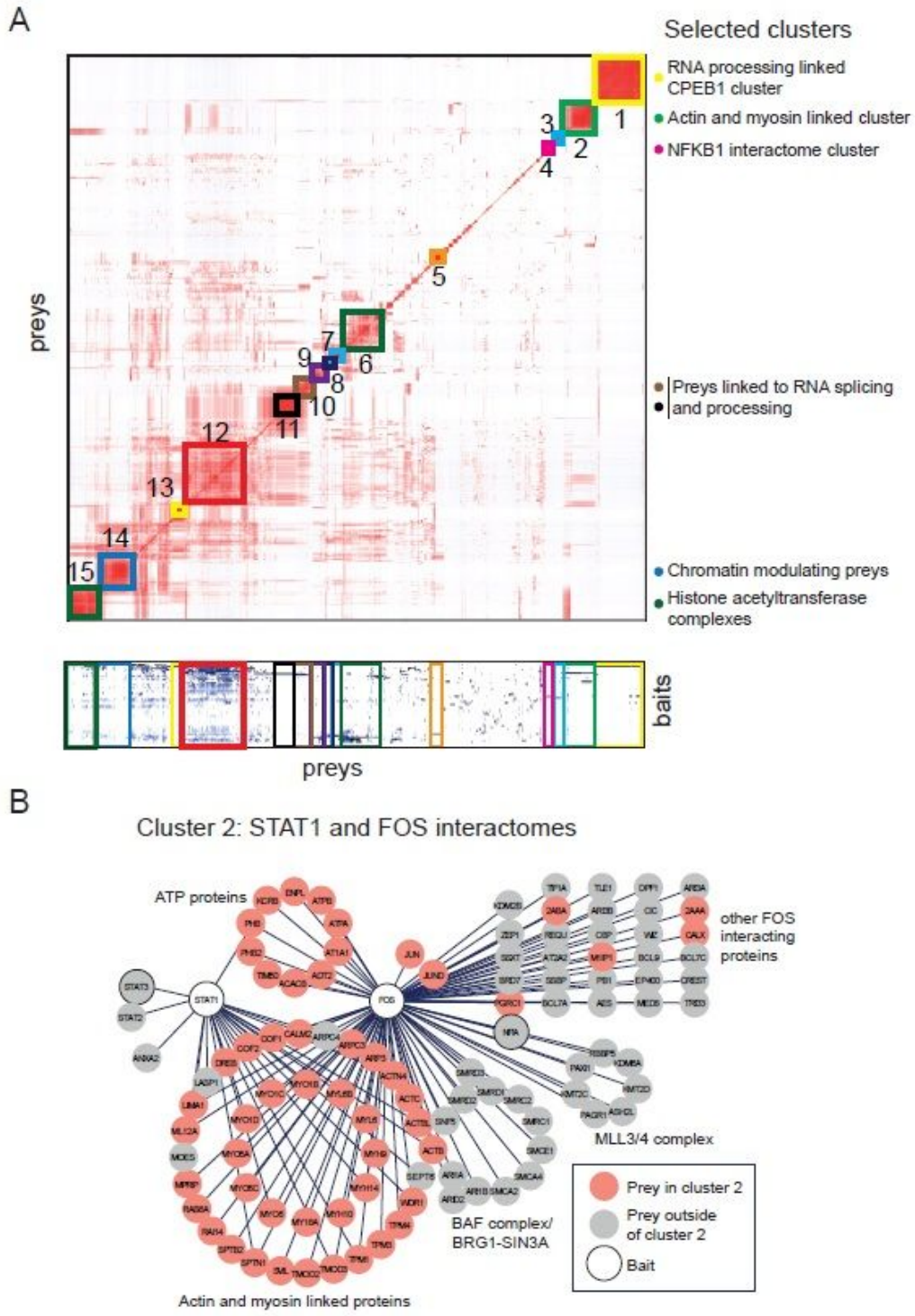


Figure 5

Biological clusters from prey-943 prey correlation analysisA. Prey-prey correlation analysis (ProHits-viz) on preys identified in the BioID experiments. The results are shown in the heatmap with preys in both the x and y axis. A corresponding bait-prey heat map is shown below to assist in determining which and how many baits are driving the prey clusters. Preys in clusters (1-15) and baits driving the clusters are shown in Table S7. Color scale indicates Pearson correlation of prey PSMs that was done by Prohits-viz prior to

hierarchical clustering. B. FOS and STAT1 interactomes. Many proteins interacting with FOS and STAT1 were clustered in Cluster 2 of the prey-prey correlation analysis (highlighted in red). Actin and myosin, ATP signalling, BAF complex and MLL3/4 complex linked proteins are highlighted in different groups.

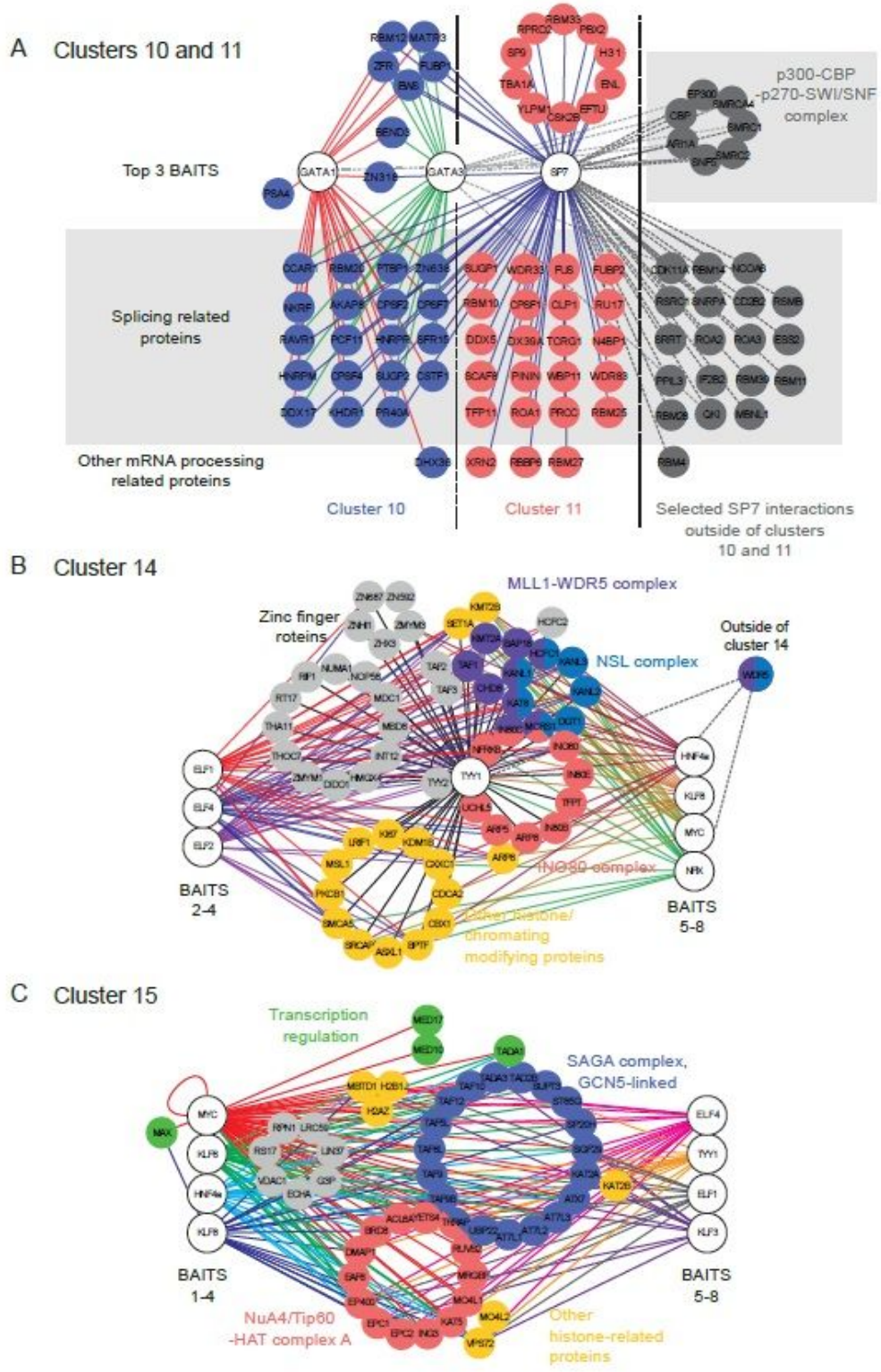


Figure 6

The interactions in Clusters 10, 11, 14 and 15 of the prey-953 prey correlation analysis A. The three baits (GATA1, GATA3, SP7) with the most interactions in Clusters 10 and 11 are shown in the centre of the

figure with white nodes. Most of the preys in Clusters 10 (blue nodes) and 11 (red nodes) were linked to mRNA splicing (highlighted in the grey box). Moreover, SP7 interactions linked to mRNA splicing, mRNA processing and the p300-CBP-p270-SWI/SNF complex found outside of Clusters 10 and 11 are presented on the right side of the figure (grey nodes). B. Protein-protein interactions in Cluster 14. The eight baits with the most interactions are highlighted in white and the interacting protein complexes are colour coded. WDR5's (outside of Cluster 14) interactions are shown with dashed lines in the upper right corner of the figure. C. Protein-protein interactions in Cluster 15. The eight baits with the most interactions are highlighted in white and the interacting protein complexes are colour coded.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx
- TableS7.xlsx
- SupplementaryFigures.pdf