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>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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

2

3 Helka Göös¹, Matias Kinnunen¹, Leena Yadav¹, Zenglai Tan^{2#}, Kari Salokas^{1#}, Qin Zhang², Gong-
4 Hong Wei^{2,3}, Markku Varjosalo^{1*}

5
6 ¹Institute of Biotechnology, HiLIFE, University of Helsinki, Helsinki, Finland

7 ²Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu,
8 Finland

9 ³Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Fudan
10 University Shanghai Cancer Center, Shanghai Medical College of Fudan University, Shanghai,
11 China

12

13

14 # equal contribution

15

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

17

18

19

20 Abbreviations: ACTB, Beta-actin; ACTBL, Beta-actin-like protein 2; AP-MS, affinity purification

21 mass spectrometry; BioID, proximity-dependent biotinylation; DBD, DNA binding domain; GO-BP,

22 Gene Ontology Biological Process; GTF, general transcription factor; HAT, histone acetyl

23 transferase; KDM2B, Lysine-specific demethylase 2B; NFI, Nuclear Factor 1 family TFs; NLS,

24 nuclear localisation signal; NM1, Nuclear myosin 1; NR, nuclear receptor; NSL, non-specific lethal;

25 PIC, pre-initiation complex; Pol-II, RNA polymerase II; PPI, protein-protein interaction; TBP,

26 TATA-binding protein; TF, transcription factor

27

28 **Abstract**

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

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

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

42

43

44

45

46

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

49

50 **Introduction**

51 ‘The central dogma’ states that genetic sequence information from DNA is transcribed to RNA and
52 subsequently translated into proteins. These processes are tightly regulated and employ a plethora of
53 proteins. Transcription, the first step, is regulated by transcription factors (TFs), which represent one
54 of the largest families of the human genes. In human, 6–9% (~1,400-1,900) of proteins are predicted
55 to regulate gene expression through DNA binding (¹⁻³, <https://www.proteinatlas.org>), and the most
56 recent manual curation identified 1639 likely human TFs⁴.

57 Complex and multilayer regulation of transcription involves not only direct binding of TFs to a target
58 gene’s regulatory element(s) but there exists a complicated interplay between TFs and TF binding
59 proteins. These include several cofactors, the Mediator complex, basal transcription machinery, TF-
60 activity modulating enzymes (such as phosphatases and kinases), dimerization partners, subunits and
61 inhibitory proteins⁵⁻⁸. Moreover, as the chromosomal DNA is packed into chromatin to prevent
62 uncontrolled transcription, TFs also interact with several chromatin remodeling proteins. The formed
63 complexes are necessary to regulate the accessibility of DNA to allow chromatin opening and thereby
64 gene transcription.

65 TFs play crucial roles in regulating numerous cellular mechanisms and are key regulators of tissue
66 growth and embryonic development - processes which may cause cancer and other disorders when
67 aberrantly controlled. Therefore, understanding the TF network at systems level would form an
68 important crucial foundation for future studies as well as for therapeutic applications⁷. While the
69 binding of TFs to DNA is extensively studied, for the most part we are still lacking a global
70 understanding of the TF protein-protein interactions (PPIs) and their roles in the regulation of
71 transcription. Therefore, we sought to fill this knowledge gap by using the recently developed state-
72 of-the-art PPI identification methods, which allow unprecedented sensitivity and depth. In this study,
73 we systemically characterized the PPIs of a selected set of 110 human TFs using affinity purification
74 mass spectrometry (AP-MS) and proximity-dependent biotinylation (BioID) mass spectrometry. We

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

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

87

88

89 **Results**

90 **Identification of TF protein-protein interactions**

91 To systematically investigate the protein-protein interactions formed by the human TFs, we selected
92 a representative set of 110 TF genes from different TF families (Table S1A), which were then
93 subjected to two independent mass spectrometry based interactome analysis methods. (Figure 1A).
94 First, the stable TF complexes were purified using single-step Strep-tag affinity purification (AP-
95 MS). Secondly, a proximity-dependent labelling approach (BioID) utilizing a minimal biotin ligase
96 tag BirA* was used to detected also transient and proximal interactions of the TFs^{9,10}. The expression
97 of studied TFs were adjusted on the corresponding transgenic cell lines to close-to-physiological level
98 by the tetracycline inducible and adjustable Tet-On expression system¹¹.

99 In total, we identified 7,233 high-confidence PPIs using BioID analysis and 2,176 PPIs using AP-MS
100 method (Table S1 B-C; Figure 1B, C). For initial sanity check for the obtained TF interactomes, we
101 mapped the interactors to their known subcellular localisations from the Cell Atlas¹². This analysis
102 revealed that > 80% of the TF interactors were detected to have nuclear localisation (yellow nodes;
103 Figure 1B; Table S2), confirming the expected nuclear compartmentalisation of the studied TFs and
104 their interactors. Remarkably, the majority (>75%) of the interactions within the TF interactome were
105 previously unreported (Figure 1C; Table S1). On average, we identified 65 PPIs/TF in the BioID data
106 and 21 PPIs/TF in the AP-MS (Figure 1D). Higher number of identified interactions by BioID
107 compared to AP-MS agrees with many recently reported medium- and large-scale interatomic
108 studies¹³⁻¹⁶, but interestingly, our results suggest that TFs do prefer to form more transient or proximal
109 interactions than stable protein complexes. This finding is consistent with the phase separation model
110 for TF interactions, where interactions incorporated in TF condensates are dynamic¹⁷⁻²⁰.

111 The BioID method has been suggested to be efficient for studying transient interactions⁹, and this is
112 supported by our results, strongly suggesting that BioID is the method-of-choice for studying TF

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

121 The most common TF interactor observed in our study was lysine-specific demethylase 2B
122 (KDM2B), which interacted with 62 TFs (Table S1B). In addition, two lysine methyltransferases
123 were among the top five of the most frequent TF preys (KMT2D: 58 PPIs and KDM6A: 53 PPIs),
124 which highlights the importance of the histone modification homeostasis in the regulation of
125 transcription. The detected interactions of lysine methyltransferases with TFs are highly specific and
126 very rarely detected in large-scale studies with other key cellular signalling molecules^{16,21,22}, and
127 hardly ever detected as contaminants²³. Other common TF interactors were NFIA (54 PPIs), TLE1
128 (53 PPIs), CIC (52 PPIs) and several zinc finger proteins (50–52 PPIs). In addition, the well-
129 established corepressors BCOR (48 PPIs) and NCOR2 (48 PPIs) were high on the list. Not
130 surprisingly, the most frequently observed TF interactors were transcriptional activators and
131 repressors.

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

138 **Comparison to other studies**

139 As BioID can capture transient and proximal interactions, most experimental validation methods,
140 such as coimmunoprecipitation, are not sufficiently sensitive to validate the results. We, therefore,
141 compared the identified PPIs with previously published interactions. In a medium scale analysis, Li
142 et al. screened the PPIs of 59 TFs by tagging them with a C-terminal SFB-tag (S protein-tag, Flag-
143 tag and Streptavidin binding peptide) and identified the interacting proteins using tandem affinity
144 purification coupled to MS⁶. This analysis identified 2,156 PPIs. Fourteen of the TFs analysed in their
145 study were included in our set (CREB1, ETS1, FOS, FOXI1, FOXL1, FOXQ1, IRF3, MEF2A, MYC,
146 NFKB1, PPARG, STAT3, TEAD2 and TP53). Their approach is close to our AP-MS analysis, which
147 identifies more stable protein-protein interactions and protein complexes than transient interactions.
148 Therefore, it is not surprising that the overlap between our BioID PPIs and their PPIs was low; only
149 6% of our PPIs were covered by their study (Table S1B). A comparison with our AP-MS results
150 revealed more common interactions; 26% of our AP-MS PPIs were detected by their approach (Table
151 S1C). These differences detected between Li et al.'s PPIs and those we identified are most likely due
152 to the transient nature of TF interactions and the different tagging strategies used.

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

160

161 **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 zinc finger (ZF) or Homeodomain DBD, which are the most common DBDs among the human TFs⁴. 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.

178

179 **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, TFIIIE, 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

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

192 **TF interactions with Nuclear factors**

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

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

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

211 Given the fact that it interacted in our analyses with 52 TFs, it is possible that NFIA could regulate
212 the transcriptional activity of other TFs. To test this, we generated luciferase-based reporter (DNA
213 binding domains extracted from JASPAR²⁷) assays for selected TFs interacting with NFIA. The
214 reporter assays which displayed induction after introduction of the corresponding TF were chosen for
215 NFIA via siRNA-mediated knockdown experiments. Knockdown efficiency was first confirmed by
216 western blotting using a specific antibody against NFIA (Figure 4C). Next, the effect of NFIA
217 depletion on selected reporter activity was tested in the presence and absence of the corresponding
218 TF (Figure 4C). Interestingly, both KLF4 assays, detecting the repressive and activating response of
219 KLF4, showed altered luciferase activity after NFIA silencing: both the repressive and activating
220 responses after the KLF4 induction were reduced (Figure 4C). In addition, SOX2 and PAX6 showed
221 reduced activity after the NFIA silencing, while EN1 activity was increased after the depletion of
222 NFIA (Figure 4C).

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

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

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

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

237 **Prey-prey correlation analysis reveals several biological clusters**

238 The TFs prey-prey correlation analysis, using ProHits-Viz²⁸, revealed 15 biological clusters (Figure
239 5A, Table S7). This analysis revealed clusters of preys that were often seen together between baits,
240 suggesting that they might be part of the same complex or colocalize or both. Baits driving the same
241 cluster had a similarity in interactomes, indicating possible shared or similar biological roles. The
242 preys belonging to different clusters and the baits driving the clusters are shown in Table S7. Next,
243 we describe some of the interesting clusters found in this correlation analysis.

244 *Actin and myosin linked protein cluster*

245 The second cluster was mainly formed of proteins linked to actin and myosin signalling. This cluster
246 consisted of 54 proteins, 36 of which were linked to actin and myosin signalling (Figure 5B, Table
247 S7). In addition, 10 proteins were linked to ATP synthesis and mitochondrial respiration (Table S7).
248 Cluster 2 was mainly driven by FOS interactions as FOS interacted with all 54 preys (Fig 5B). Besides
249 FOS, the cluster was also driven by STAT1 (18 interactions), FOXL1 (17 interactions) and 29 other
250 baits with less than 10 interactions (Table S7). Interestingly, FOS interacted with all 36 actin- and
251 myosin-linked proteins in the cluster; STAT1 interacted with 17 and FOXL1 with 6 actin and myosin
252 linked proteins (Table S7; Figure 5B). Other baits did not have actin- and myosin-linked interactions.
253 FOS and STAT1 uniquely interacted for example with beta-actin (ACTB) and beta-actin-like protein
254 2 (ACTBL2) (Table S7; Figure 5B). In addition, FOS and STAT1 interacted with ARP3, FOS with
255 ARPC3 and STAT1 with ARPC4, all members of the Arp2/3 complex which mediates the actin

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

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

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

266 Next, we found significant enrichment of proteins linked to mRNA splicing and processing in
267 Clusters 10 and 11 (Figure 5A). Cluster 10 consisted of 29 preys and Cluster 11 of 34 preys, of which
268 23 and 20, respectively, were linked to RNA splicing and processing (Figure 6A, Table S7). GO-BP
269 analysis showed a significant enrichment of proteins linked to ‘mRNA splicing, via spliceosome’ (p-
270 value 2.70×10^{-11} in Cluster 10 and 6.7×10^{-13} in Cluster 11). Cluster 10 was driven mainly by GATA1
271 (28 interactions), GATA3 (20 interactions) and SP7 (20 interactions) (Figure 6A). Cluster 11
272 consisted almost totally of SP7 interactions; SP7 interacted with all 34 proteins (Figure 6A, Table
273 S7).

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

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

282 *TF interactions with chromatin modulating complexes in Cluster 14*

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

289 Cluster 14 was mainly formed of TYY1 interactions; TYY1 interacted with all 57 preys in the cluster
290 (Figure 6B, Table S7), while ELF1 interacted with 24, ELF4 with 24, ELF2 with 20 and HNF4A with
291 15 preys. Other baits driving the cluster are listed in Table S7.

292 TYY1 is known to be part of the INO80 complex³¹. Predictably, we found TYY1 to interact with 10
293 subunits of the INO80 complex (Figure 6B; Table S7). TYY1 interactions with INO80 complex
294 members appear to be very stable as many of the interactions were also detected in the AP-MS data
295 (Table S1C). Besides TYY1, we found that ELF4 interacted with eight INO80 complex members and
296 with its putative regulator UCHL (Figure 6B).

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

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

309 *TF interactions with histone acetyltransferase complexes in Cluster 15*

310 Closer analysis of Cluster 15 (Figure 5A) revealed accessory chromatin modulating complexes,
311 especially histone acetylation complexes. BP-GO analysis resulted in the significant enrichment of
312 terms linked to histone H2A, histone H4 and histone H3 acetylation (p-values of less than 7.19×10^{-22}).
313

314 In total, Cluster 15 consisted of 50 preys of which 40 were directly linked to histone acetylation
315 (Figure 6C, Table S7). These included 19 members of the SAGA complex and 14 members of the
316 NuA4/Tip60 HAT complex A (Figure 6C). The cluster was mainly driven by MYC, which had
317 interactions with all 50 preys. KLF6 interacted with 29, HNF4A with 21, KLF8 with 20, ELF4 with
318 17, TYY1 with 13 and ELF1 with 11 preys. Other baits with less than 10 interactions are listed in
319 Table S7.

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

326 Discussion

327 Chromatin opening, transcription, RNA splicing, RNA processing and their regulation are often
328 studied as separate processes. However, our understanding of the simultaneous and co-transcriptional
329 nature of these processes has exploded in recent years³²⁻³⁶. In our analyses, TFs were found to interact
330 with proteins involved in chromatin remodeling, transcription, mRNA splicing and RNA processing,
331 highlighting the cooperative nature and close proximity of these processes. This also showed that TFs
332 are central in regulating these interconnected processes. The most common interaction partners for
333 the TFs studied were histone-modifying enzymes, signifying that histone modification and chromatin
334 accessibility regulation are central to all these transcriptional subprocesses.

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

337 Mediator complex, SAGA complex and most of the GTFs are multimeric protein complexes, which
338 are needed for Pol-II promoter recognition and transcription initiation³⁷⁻³⁹. In most studied models,
339 the PIC assembly starts with the binding of TBP to TATA- or TATA-like core promoters⁴⁰. TBP
340 belongs to two GTF complexes: TFIID and SAGA. It has been indicated that both TFIID and SAGA
341 participate in the transcription of various genes simultaneously⁴¹, but that the regulation of expression
342 might be dominated by either one of them^{42,43}. Different promoters are alleged to prefer either TFIID
343 or SAGA, and it has been indicated that the activity of SAGA/TATA-like promoters might be more
344 dependent on the presence of transcriptional activators (regulated genes) than the activity of
345 TFIID/TATA-promoters (housekeeping genes)⁴⁴. However, this is still controversial, as the depletion
346 of either SAGA or TFIID complex members decreased the transcription of both regulated and
347 housekeeping genes^{25,41}.

348 We observed multiple interactions with both TFIID and SAGA components, supporting the theory
349 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 complexes¹⁷⁻²⁰. 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 sequence⁴⁷⁻⁴⁹. 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 replication⁵², but over the years they have been found to control a variety of genes in cancer and in development⁵³⁻⁵⁸. For example, NFIs are found to have multiple translocations leading to oncogenic gene fusion proteins in several cancer types⁵⁸, 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}.

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

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

386 *FOS and STAT1 interact with nuclear actin and myosin related proteins*

387 Nuclear actin is associated to chromatin remodeling complexes and is part of the Pol-II transcription
388 machinery⁶⁴. Actin dynamics have also been directly linked to gene transcription regulation^{65,66}.

389 In our analysis, the interactomes of FOS and STAT1 were found to be enriched with proteins linked
390 to nuclear actin and myosin signalling. FOS and STAT1 interacted uniquely, e.g., with beta-actin
391 (ACTB), ACTBL, and MYO1C, isoform 3 of which is also known as NM1 (Table S7, Figure 5B).
392 NM1 and ACTB are suggested to be associated with each other and to play important roles in Pol-II
393 transcription^{64,67}. ACTB is also part of several chromatin remodeling complexes, such as BAF, Tip60
394 and INO80⁶⁸. Indeed, we also detected FOS interacting with 13 BAF complex members (Table S1B,
395 Figure 5B), suggesting that FOS activity is regulated through BAF complex and actin dynamics. FOS

396 has been reported to be linked to BAF signalling before^{69,70}, but how actin is involved in this process
397 remains largely unknown.

398 STAT1 activity is known to be regulated by actin cytoskeleton and extracellular matrix proteins⁷¹.
399 Interestingly, STAT1 did not show any interactions with BAF complex members, indicating a
400 different mechanism or different mediating proteins between STAT1 and actin proteins.

401 *SP7 interacts with RNA-splicing proteins*

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

405 TFs are known to affect RNA splicing in three ways; they can bind to RNA to recruit coregulators
406 that also take part in splicing, block the associations of splicing factors with mRNA, and/or influence
407 the transcription elongation rates, which are known to impact on splicing by skipping the weak 3'
408 splice sites at a high rate⁷². One way to alter the elongation rate is through TF-mediated recruitment
409 of EP300 which induces the histone acetylation of nearby promoters, increases the elongation rate
410 and promotes exon skipping⁷³. Therefore, SP7 and GATA3 interaction to EP300 suggest that they
411 may be connected to the p300 chromatin remodeling complex and, thus, to the regulation of
412 elongation rate.

413 Some TFs, such as steroid hormone receptors, nuclear receptors (NRs) and certain non-NR TFs, are
414 known to regulate mRNA splicing by recruiting splicing-linked coregulators^{72,74}. One of these
415 coregulators is RBM14 (also known as CoAA), which is an NCOA6 (also known as TRBP) binding
416 protein⁷⁵. We found that, along with other splicing-related proteins, both RBM14 and NCOA6
417 interacted with SP7 (Figure 6A, Table S1B). This might indicate that SP7 recruits similar splicing-
418 regulation-related coregulators.

419 SP7 interactions with four core spliceosomal components, and all other splicing related components
420 suggested that SP7 has a largely unstudied role in recruiting the splicing machinery to the nascent
421 pre-mRNA – a role that needs to be studied more. Like some other C2H2 zinc finger TFs, such as
422 CTCF, VEZF1, MAZ and WT1, that are known to regulate mRNA splicing⁷², SP7 might participate
423 in pre-mRNA splicing.

424 *TF interactions with chromatin modulating complexes*

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

428 INO80 ATP-dependent chromatin remodeling complex activates transcription³¹, regulates genomic
429 stability through DNA repair⁷⁶, contributes to DNA replication⁷⁷ and, by shifting the nucleosomes,
430 remodels the chromatin^{78,79}. Predictably, TYY1, that is known to be part of INO80 complex, had
431 multiple interactions with INO80 complex members in our analyses. More interestingly, we found
432 that ELF4 had also multiple interactions with INO80 members. To our knowledge, ELFs have not
433 been previously linked to INO80 signalling or TYY1 and this should be studied further.

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

440 As mentioned earlier, the SAGA complex is a multi-module complex that has an important role in
441 Pol-II recruitment for all expressed genes⁸¹. In addition, SAGA takes part in mRNA synthesis and
442 export, maintenance of DNA integrity and histone modifications, such as histone acetylation,

443 succinylation and ubiquitylation⁸¹⁻⁸⁷. In our data, multiple interactions were detected between SAGA
444 complex members and the studied TFs (e.g. MYC, KLF6, KL8, HNF4a; Tables S1B and S5). MYC
445 connections to SAGA are known⁸⁸ and KLF6 interaction with one of the subunits, TAF9, is reported
446 in PINA2 database. In addition, KLF6 is known to interact with HDAC3 in preadipocyte
447 differentiation⁸⁹. However, we found no other connections between KLF6 and SAGA complex or
448 histone modification. Our data indicated that KLF6 connections to SAGA are *bona fide* and should
449 be studied further.

450 Finally, we found MYC, HNF4a, KLF6, KLF8 and ELF4 to interact with the NuA4/Tip60 HAT
451 complex (Tables S1B and S7) that plays essential roles in cell cycle control, transcription and DNA
452 repair, and act in the N-terminal acetylation of histones H4 and H2A⁹⁰. NuA4/Tip60 HAT complex,
453 along with other HAT complexes, is known to participate in MYC-signaling⁹¹. Accordingly, we found
454 14 interactions between MYC and NuA4/Tip60 HAT complex (Tables S1B and S7). Like that of
455 MYC, HNF4a's association with the NuA4/Tip60 HAT complex is reported in a previous study⁹².
456 Interestingly, we found also strong connection (eight to nine interactions) between KLF6, KLF8 or
457 ELF4 and NuA4/Tip60 HAT complex. However, as mentioned earlier with regard to the SAGA
458 complex, not much is known about the link between KLF6 link and histone modification. Therefore,
459 KLF6's role in HAT complexes remains largely unstudied and requires further investigations.

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

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

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

472 *Conclusions*

473 While TF binding to DNA is much studied, there is still a lack of comprehensive systems levels
474 understanding of human TF protein interactions. The protein interactions of other human large
475 proteins families, such as kinases and phosphatases, have been studied on the systems level^{16,21,22,93},
476 but the TFs protein interactions remain globally unstudied. This study provides so far, the most
477 comprehensive systems-level analysis of human TFs identifying the largest reported cohort of TF
478 PPIs and serving as a rich resource for further research and development of pharmaceutical treatment
479 for TF-related diseases. It also allows to profile TFs protein interactomes in the context of more than
480 100 TF interactomes. Moreover, this is the first large-scale study to identify the dynamic PPIs of TFs
481 using transient and proximal interactions catching BioID method. Finally, as defects in TF signalling
482 often lead to severe pathological conditions⁹⁴⁻⁹⁶, and TFs function as downstream players of multiple
483 signalling cascades³, identifying TF PPIs make a crucial contribution to pharmacological targeting of
484 TF-related diseases.

485

486 **METHODS**

487

488 *Used cell lines*

489 Cell lines stably expressing selected TFs were generated from Flp-In™ T-REx™ 293 cells (Life
490 Sciences). siRNA silencing and luciferase experiments were performed in HEK 293 cell line
491 (American Type Culture Collection, Manassas, VA). All the cells were cultured in low glucose
492 tetracycline-free DMEM (Sigma Aldrich) complemented with 10% FBS and 100 µg/ml
493 Penicillin/streptomycin (Life Technologies) at 37 °C with 5 % CO₂.

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

495 110 TFs from different TF families were selected for this study. Using Gateway® cloning, the TF
496 coding sequences without stop codons were obtained from ORF libraries and commercially cloned
497 into pDONR221 entry vectors (GenScript). To generate tetracycline-inducible stable cell lines,
498 constructs were cloned into N-terminal pTO_HA_StrepIII_BirA-N_GW_FRT, pTO_HA_StrepIII-
499 N_GW_FRT or MAC-tag vectors and introduced into Flp-In™ T-REx™ 293 cells (Life
500 Technologies, Carlsbad, CA) to generate stable, isogenic and inducible cell lines as described by Liu
501 et al¹⁵.

502 *Affinity purification and mass spectrometry analysis*

503 Approximately 1x10⁸ Flp-In™ T-REx™ 293 cells stably expressing human TFs were induced with
504 2 µg/ml of tetracycline (AP-MS and BioID) and 50 µM biotin (BioID) for 24 hours. The cells were
505 pelleted using centrifugation, snap frozen in liquid nitrogen and stored at -80°C. The samples were
506 then suspended in 3 ml of lysis buffer (50 mM HEPES pH 8.0, 5 mM EDTA, 150 mM NaCl, 50 mM
507 NaF, 0.5% NP40, 1.5 mM Na₃VO₄, 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 μ l 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 μ l 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 μ g 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 μ m, 5 μ m, 120 Å, Thermo Fisher Scientific)
525 and then into analytical column (C18-packing; EASY-Column™ 10 cm x 75 μ m, 3 μ m, 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

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

534 ***Protein identification***

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

543 The identified proteins were filtered using SAINT software tools⁹⁷ with a SAINT score cut-off of
544 0.74. All the TFs were analysed in two or four replicates. TFs are known to have variable expression
545 levels and patterns, and some of them are present in cells at extremely low copy numbers⁹⁸. To
546 efficiently filter the real interactions, we used 44 and 75 similarly tagged and analysed GFP control
547 runs for the BioID analysis and AP-MS analysis, respectively. We also included GFPs with a nuclear
548 localisation signal (NLS) to efficiently filter out unspecific nuclear interactions. The large nuclear
549 dataset further facilitated the frequency-based deletion of contaminating proteins. The Cytoscape
550 software platform was used to visualize the high-confidence TF PPIs⁹⁹.

551

552 ***Data analysis***

553 The subcellular localisations of interacting proteins were obtained from the Cell Atlas¹². Enriched
554 biological process Gene Ontology terms for all PPIs were obtained from DAVID Bioinformatics
555 Resources¹⁰⁰. 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 settings²⁸. Comparison of two cluster dendrograms were done using
560 *dendextend* 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 UniProt¹⁰¹. The DNA-
562 binding motifs of the studied TFs were mainly extracted from the JASPAR database²⁷. 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 RSAT²⁴. 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 used²⁸. 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 μ l/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

580 (pTO-SH-GW-FRT) along with 47.5 ng reporter construct. The reporter constructs contained 6-8x
581 TF binding sites (TFBSs;²⁷, minimal promoter and firefly-luciferase reporter. Only the constructs
582 displaying induction after introduction of the corresponding TF were chosen for further analysis.
583 These include reporters for KLF4 (both activating: [TFBS: 6x GGGTGTGG] and repressive: [TFBS:
584 8x TAAAGGAAGG]), SOX2 (TFBS: 6x CTTTGTT), PAX6 (TFBS: 6x TTCACGCTTGAGTT) and
585 EN1 (TFBS: 8x AAGTAGTGCCC).

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

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

593 HEK293 cells were seeded in 10 cm dishes; 24 hours later, tetracycline with a final concentration of
594 2 µg/ml was added to induce the expression of SOX2 protein. siRNAs against NFIA or control were
595 transfected with a final concentration of 100 nM by the lipofectamine RNAiMAX reagent (Thermo
596 Fisher Scientific). HEK293 cells with depletion of NFIA were applied in ChIP-seq assays as
597 previously described¹⁰⁴. In brief, HEK293 cells were cross-linked with 1% formaldehyde for 10 min
598 at room temperature after 48 hours of siRNA transfection. The reaction was quenched with 125 mM
599 glycine. Cells were collected after washing twice by pre-cold PBS and resuspended in hypotonic lysis
600 buffer (20 mM Tris-Cl, PH 8.0, with 10% glycerol, 10 mM KCl, 2 mM DTT, and complete protease
601 inhibitor cocktail (Roche)) to isolate nuclei. The nuclei pellets were washed with pre-cold PBS and
602 resuspended in 1:1 ratio of SDS lysis buffer (50 mM Tris-HCl, pH 8.1, with 1% SDS, 10 mM EDTA,
603 and complete Protease Inhibitor) and ChIP dilution buffer (16.7 mM Tris-HCl, pH 8.1, with 0.01%

604 SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl and Complete Protease Inhibitor). A Q800R
605 sonicator (QSonica) was used to generate an average size of 300 bp chromatin fragments at 4°C.
606 Dynabead protein G (Invitrogen) was washed with blocking buffer ((0.5% BSA in IP buffer (20 mM
607 Tris-HCl, pH8.0, with 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, and Protease inhibitor
608 cocktail)) and incubated with antibody against HA (ab18181, Abcam). Chromatin lysate was
609 precipitated with Dynabead protein G for 12 hours, then wash the beads 4 times with washing buffer
610 (50 mM HEPES, PH 7.6, 1 mM EDTA, 0.7% sodium deoxycholate, 1% NP-40, 0.5 M LiCl) and 2
611 times 100 mM ammonium hydrogen carbonate (AMBIC) solution. The DNA-protein complexes
612 extracted from the beads by eluting in extraction buffer (10 mM Tris-HCl, pH8.0, with 1 mM EDTA,
613 and 1% SDS), Proteinase K and RNase A were added to reverse the cross-links. A Mini-Elute PCR
614 purification kit (Qiagen) was used to purify the DNA. The purified DNA was subjected to ChIP-seq
615 library preparation through using the TruSeq ChIP Sample Preparation kit (Illumina). Briefly, DNA
616 was blunted by using an End Repair Mix, then A-Tailing Mix was used to add a nucleotide to the 3'
617 Ends of the DNA fragments. RNA Adaptor Indices were ligated to the DNA fragments and fragment
618 size of 200-500 bp were selected on a 2% agarose gel. The DNA were enrichment by PCR
619 amplification and quantified by using KAPA Library Quantification Kit (Roche). A NextSeq550
620 sequencing system (Illumina) was used to sequence the DNA library.

621 ***Bioinformatics analysis of ChIP-Seq data***

622 The ChIP-seq library was sequenced to generate 76 bp single-end reads. FastQC¹⁰⁵ was applied to
623 assess the quality of raw data and followed by Trimmomatic¹⁰⁶ for quality control. The cleaned reads
624 were aligned to the human genome hg38 assembly using Bowtie2¹⁰⁷. The ChIP-seq peaks were
625 identified by applying findPeaks.pl from Hypergeometric Optimization of Motif Enrichment
626 (HOMER v4.10)¹⁰⁸ with parameter “-mapq 20”, while all other parameters were kept as default. Motif
627 enrichment analysis was performed using HOMER findMotifsGenome.pl and peaks were annotated
628 by “annotatePeaks.pl”. Bioconductor package ChIPseeker (1.18.0)¹⁰⁹ was applied to perform pathway

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

632

633 **DATA AVAILABILITY**

634 *Lead contact*

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

637 *Material Availability*

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

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

641 *Data and Code Availability*

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

647

648

649

650

652 1 Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. Structure and evolution of transcriptional regulatory networks.
653 *Current opinion in structural biology* **14**, 283-291, doi:10.1016/j.sbi.2004.05.004 (2004).
654 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
655 (2009).
656 3 Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and
657 evolution. *Nature reviews. Genetics* **10**, 252-263, doi:10.1038/nrg2538 (2009).
658 4 Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665, doi:10.1016/j.cell.2018.01.029 (2018).
659 5 Brivanlou, A. H. & Darnell, J. E., Jr. Signal transduction and the control of gene expression. *Science (New York, N.Y.)* **295**, 813-818,
660 doi:10.1126/science.1066355 (2002).
661 6 Li, X. *et al.* Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Molecular systems biology* **11**,
662 775, doi:10.15252/msb.20145504 (2015).
663 7 Fontaine, F., Overman, J. & Francois, M. Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and
664 obstacles. *Cell regeneration (London, England)* **4**, 2, doi:10.1186/s13619-015-0015-x (2015).
665 8 Rivera-Reyes, R., Kleppa, M. J. & Kispert, A. Proteomic analysis identifies transcriptional cofactors and homeobox transcription factors as TBX18
666 binding proteins. *PLoS one* **13**, e0200964, doi:10.1371/journal.pone.0200964 (2018).
667 9 Varnaite, R. & MacNeill, S. A. Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. *Proteomics*
668 **16**, 2503-2518, doi:10.1002/pmic.201600123 (2016).
669 10 Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian
670 cells. *The Journal of cell biology* **196**, 801-810, doi:10.1083/jcb.201112098 (2012).
671 11 Glatter, T., Wepf, A., Aebersold, R. & Gstaiger, M. An integrated workflow for charting the human interaction proteome: insights into the PP2A
672 system. *Molecular systems biology* **5**, 237, doi:10.1038/msb.2008.75 (2009).
673 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).
674 13 Abbasi, S. & Schild-Poulter, C. Mapping the Ku Interactome Using Proximity-Dependent Biotin Identification in Human Cells. *Journal of proteome*
675 *research* **18**, 1064-1077, doi:10.1021/acs.jproteome.8b00771 (2019).
676 14 Lambert, J. P., Tucholska, M., Go, C., Knight, J. D. & Gingras, A. C. Proximity biotinylation and affinity purification are complementary approaches
677 for the interactome mapping of chromatin-associated protein complexes. *Journal of proteomics* **118**, 81-94,
678 doi:10.1016/j.jprot.2014.09.011 (2015).
679 15 Liu, X. *et al.* An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations.
680 *Nature communications* **9**, 1188, doi:10.1038/s41467-018-03523-2 (2018).
681 16 Yadav, L. *et al.* Systematic Analysis of Human Protein Phosphatase Interactions and Dynamics. *Cell systems* **4**, 430-444.e435,
682 doi:10.1016/j.cels.2017.02.011 (2017).
683 17 Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (New York, N.Y.)* **361**, 412-
684 415, doi:10.1126/science.aar4199 (2018).
685 18 Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.)* **361**,
686 doi:10.1126/science.aar3958 (2018).
687 19 Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science (New York, N.Y.)* **361**,
688 doi:10.1126/science.aar2555 (2018).
689 20 Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-
690 1855.e1816, doi:10.1016/j.cell.2018.10.042 (2018).
691 21 Varjosalo, M. *et al.* The protein interaction landscape of the human CMGC kinase group. *Cell reports* **3**, 1306-1320,
692 doi:10.1016/j.celrep.2013.03.027 (2013).
693 22 Varjosalo, M. *et al.* Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. *Nature methods* **10**,
694 307-314, doi:10.1038/nmeth.2400 (2013).
695 23 Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature methods* **10**, 730-736,
696 doi:10.1038/nmeth.2557 (2013).
697 24 Nguyen, N. T. T. *et al.* RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic acids research* **46**, W209-w214,
698 doi:10.1093/nar/gky317 (2018).
699 25 Baptista, T. *et al.* SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Molecular cell* **68**, 130-143.e135,
700 doi:10.1016/j.molcel.2017.08.016 (2017).
701 26 Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295-301,
702 doi:10.1038/nature10799 (2012).
703 27 Mathelier, A. *et al.* JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic*
704 *acids research* **44**, D110-115, doi:10.1093/nar/gkv1176 (2016).
705 28 Knight, J. D. R. *et al.* ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nature methods* **14**, 645-646,
706 doi:10.1038/nmeth.4330 (2017).
707 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.
708 *The Journal of biological chemistry* **282**, 7616-7623, doi:10.1074/jbc.M607596200 (2007).
709 30 Welch, M. D., Iwamatsu, A. & Mitchison, T. J. Actin polymerization is induced by Arp2/3 protein complex at the surface of *Listeria*
710 *monocytogenes*. *Nature* **385**, 265-269, doi:10.1038/385265a0 (1997).
711 31 Cai, Y. *et al.* YY1 functions with INO80 to activate transcription. *Nature structural & molecular biology* **14**, 872-874, doi:10.1038/nsmb1276
712 (2007).
713 32 Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin
714 structure. *Annual review of biochemistry* **84**, 165-198, doi:10.1146/annurev-biochem-060614-034242 (2015).
715 33 Dahan, N. & Choder, M. The eukaryotic transcriptional machinery regulates mRNA translation and decay in the cytoplasm. *Biochimica et*
716 *biophysica acta* **1829**, 169-173, doi:10.1016/j.bbagr.2012.08.004 (2013).
717 34 Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688-700,
718 doi:10.1016/j.cell.2009.02.001 (2009).

719 35 Komili, S. & Silver, P. A. Coupling and coordination in gene expression processes: a systems biology view. *Nature reviews. Genetics* **9**, 38-48,
720 doi:10.1038/nrg2223 (2008).

721 36 Reed, R. Coupling transcription, splicing and mRNA export. *Current opinion in cell biology* **15**, 326-331 (2003).

722 37 Meyer, K. D., Lin, S. C., Bernecky, C., Gao, Y. & Taatjes, D. J. p53 activates transcription by directing structural shifts in Mediator. *Nature*
723 *structural & molecular biology* **17**, 753-760, doi:10.1038/nsmb.1816 (2010).

724 38 Poss, Z. C., Ebmeier, C. C. & Taatjes, D. J. The Mediator complex and transcription regulation. *Critical reviews in biochemistry and molecular*
725 *biology* **48**, 575-608, doi:10.3109/10409238.2013.840259 (2013).

726 39 Joo, Y. J. *et al.* Downstream promoter interactions of TFIID TAFs facilitate transcription reinitiation. *Genes & development* **31**, 2162-2174,
727 doi:10.1101/gad.306324.117 (2017).

728 40 Luse, D. S. The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled? *Transcription* **5**, e27050,
729 doi:10.4161/trns.27050 (2014).

730 41 Fischer, V., Schumacher, K., Tora, L. & Devys, D. Global role for coactivator complexes in RNA polymerase II transcription. *Transcription* **10**, 29-
731 36, doi:10.1080/21541264.2018.1521214 (2019).

732 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).

733 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*
734 *cerevisiae*. *Molecular cell* **13**, 573-585 (2004).

735 44 de Jonge, W. J. *et al.* Molecular mechanisms that distinguish TFIID housekeeping from regulatable SAGA promoters. *The EMBO journal* **36**, 274-
736 290, doi:10.15252/embj.201695621 (2017).

737 45 Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nature reviews. Molecular cell biology* **16**, 155-166,
738 doi:10.1038/nrm3951 (2015).

739 46 Borggreffe, T. & Yue, X. Interactions between subunits of the Mediator complex with gene-specific transcription factors. *Seminars in cell &*
740 *developmental biology* **22**, 759-768, doi:10.1016/j.semcdb.2011.07.022 (2011).

741 47 Gronostajski, R. M. Analysis of nuclear factor I binding to DNA using degenerate oligonucleotides. *Nucleic acids research* **14**, 9117-9132,
742 doi:10.1093/nar/14.22.9117 (1986).

743 48 Gronostajski, R. M., Adhya, S., Nagata, K., Guggenheimer, R. A. & Hurwitz, J. Site-specific DNA binding of nuclear factor I: analyses of cellular
744 binding sites. *Molecular and cellular biology* **5**, 964-971, doi:10.1128/mcb.5.5.964 (1985).

745 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).

746 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
747 (2000).

748 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
749 from *C. elegans* to mammals. *Mammalian genome : official journal of the International Mammalian Genome Society* **10**, 390-396 (1999).

750 52 Nagata, K., Guggenheimer, R. A., Enomoto, T., Lichy, J. H. & Hurwitz, J. Adenovirus DNA replication in vitro: identification of a host factor that
751 stimulates synthesis of the preterminal protein-dCMP complex. *Proceedings of the National Academy of Sciences of the United States of*
752 *America* **79**, 6438-6442, doi:10.1073/pnas.79.21.6438 (1982).

753 53 Steele-Perkins, G. *et al.* Essential role for NFI-C/CTF transcription-replication factor in tooth root development. *Molecular and cellular biology* **23**,
754 1075-1084, doi:10.1128/mcb.23.3.1075-1084.2003 (2003).

755 54 Steele-Perkins, G. *et al.* The transcription factor gene Nfib is essential for both lung maturation and brain development. *Molecular and cellular*
756 *biology* **25**, 685-698, doi:10.1128/mcb.25.2.685-698.2005 (2005).

757 55 Campbell, C. E. *et al.* The transcription factor Nfix is essential for normal brain development. *BMC developmental biology* **8**, 52,
758 doi:10.1186/1471-213x-8-52 (2008).

759 56 Fane, M., Harris, L., Smith, A. G. & Piper, M. Nuclear factor one transcription factors as epigenetic regulators in cancer. *International journal of*
760 *cancer* **140**, 2634-2641, doi:10.1002/ijc.30603 (2017).

761 57 Mason, S., Piper, M., Gronostajski, R. M. & Richards, L. J. Nuclear factor one transcription factors in CNS development. *Molecular neurobiology*
762 **39**, 10-23, doi:10.1007/s12035-008-8048-6 (2009).

763 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.
764 *Cancer letters* **410**, 124-138, doi:10.1016/j.canlet.2017.09.015 (2017).

765 59 Piper, M., Gronostajski, R. & Messina, G. Nuclear Factor One X in Development and Disease. *Trends in cell biology* **29**, 20-30,
766 doi:10.1016/j.tcb.2018.09.003 (2019).

767 60 Driller, K. *et al.* Nuclear factor I X deficiency causes brain malformation and severe skeletal defects. *Molecular and cellular biology* **27**, 3855-
768 3867, doi:10.1128/mcb.02293-06 (2007).

769 61 Shu, T., Butz, K. G., Plachez, C., Gronostajski, R. M. & Richards, L. J. Abnormal development of forebrain midline glia and commissural projections
770 in Nfia knock-out mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **23**, 203-212 (2003).

771 62 Glasgow, S. M. *et al.* Mutual antagonism between Sox10 and NFIA regulates diversification of glial lineages and glioma subtypes. *Nature*
772 *neuroscience* **17**, 1322-1329, doi:10.1038/nn.3790 (2014).

773 63 Kang, P. *et al.* Sox9 and NFIA coordinate a transcriptional regulatory cascade during the initiation of gliogenesis. *Neuron* **74**, 79-94,
774 doi:10.1016/j.neuron.2012.01.024 (2012).

775 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
776 (2006).

777 65 Miralles, F., Posern, G., Zaromytidou, A. I. & Treisman, R. Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell* **113**, 329-
778 342, doi:10.1016/s0092-8674(03)00278-2 (2003).

779 66 Olson, E. N. & Nordheim, A. Linking actin dynamics and gene transcription to drive cellular motile functions. *Nature reviews. Molecular cell*
780 *biology* **11**, 353-365, doi:10.1038/nrm2890 (2010).

781 67 de Lanerolle, P. Nuclear actin and myosins at a glance. *Journal of cell science* **125**, 4945-4949, doi:10.1242/jcs.099754 (2012).

782 68 Xie, X. *et al.* beta-Actin-dependent global chromatin organization and gene expression programs control cellular identity. *FASEB journal : official*
783 *publication of the Federation of American Societies for Experimental Biology* **32**, 1296-1314, doi:10.1096/fj.201700753R (2018).

784 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*
785 *of biological chemistry* **276**, 2852-2857, doi:10.1074/jbc.M009633200 (2001).

786 70 Vierbuchen, T. *et al.* AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Molecular cell* **68**, 1067-
787 1082.e1012, doi:10.1016/j.molcel.2017.11.026 (2017).

788 71 Chen, Z. *et al.* Negative regulation of interferon-gamma/STAT1 signaling through cell adhesion and cell density-dependent STAT1
789 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-acetyltransferase (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).

861 109 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*
862 **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).
863 110 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
864 111 Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-165,
865 doi:10.1093/nar/gkw257 (2016).

867

868

869 **Author Contributions**

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

875

876 **Acknowledgements**

877 For the founding, we would like to thank Orion Research Foundation sr, Finnish Cultural Foundation
878 (HG), Academy of Finland (288475 and 294173), Sigrid Jusélius Foundation, Biocentrum Helsinki,
879 University of Helsinki Three-year Research Grant, Biocentrum Finland, HiLIFE, Instrumentarium
880 Research Foundation (MV), Jane and Aatos Erkko Foundation and Finnish Cancer Foundation (GW).
881 We would also like to thank Petri Auvinen and Juha Kere for the critical reading of the manuscript.

882

883 **Competing interests**

884 Authors report no competing interests.

885

886

A

B

- nuclear localization
- non-nuclear localization
- bait proteins

C

■ novel interaction
■ known interaction

D

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

908 **Figure 2. Comprehensive protein interactomes of the studied TF families**

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

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

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

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

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

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

A

B

C

D

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

B

Cluster 2: STAT1 and FOS interactomes

943 **Figure 5. Biological clusters from prey-prey correlation analysis**

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

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

A Clusters 10 and 11

B Cluster 14

C Cluster 15

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

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

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

962 **C.** Protein-protein interactions in Cluster 15. The eight baits with the most interactions are highlighted
963 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 experimental 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 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.

Figure 5

Biological clusters from prey-943 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.

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](#)