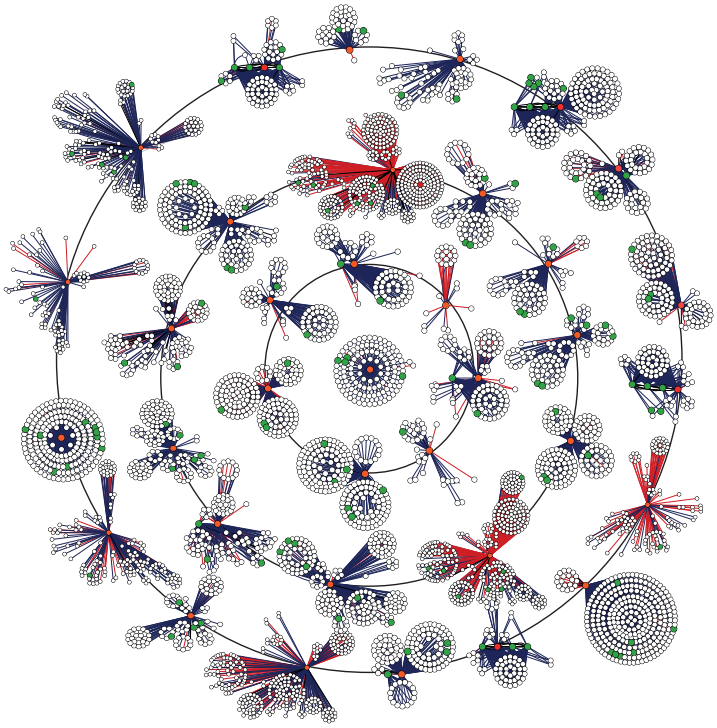
DISSERTATIONES SCHOLAE DOCTORALIS AD SANITATEM INVESTIGANDAM
UNIVERSITATIS HELSINKIENSIS

HELKA GÖÖS

HUMAN TRANSCRIPTION FACTOR PROTEIN-PROTEIN INTERACTIONS IN HEALTH AND DISEASE



INSTITUTE OF BIOTECHNOLOGY
HELSINKI INSTITUTE OF LIFE SCIENCE (HiLIFE) AND
DEPARTMENT OF BIOSCIENCES
FACULTY OF BIOLOGICAL AND ENVIROMENTAL SCIENCES
DOCTORAL PROGRAMME IN BIOMEDICINE
UNIVERSITY OF HELSINKI

Institute of Biotechnology
Helsinki Institute of Life Science (HiLIFE),
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

Doctoral Programme in Biomedicine (DPBM)
University of Helsinki
Helsinki, Finland

HUMAN TRANSCRIPTION FACTOR PROTEIN-PROTEIN INTERACTIONS IN HEALTH AND DISEASE

Helka Göös

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki,
for public examination in Auditorium 235 (Sali 2)
at Infokeskus Korona, Viikinkaari 11, Helsinki,
on 29th of November, 2019 at 12 o'clock.

Helsinki 2019

Supervision
Research Director Markku Varjosalo, Ph.D.
Institute of Biotechnology
Helsinki Institute of Life Science (HiLIFE)
University of Helsinki, Helsinki, Finland

Custodian,
Thesis Committee member
Professor
Juha Partanen, Ph.D.
Faculty of Biological and Environmental Sciences
University of Helsinki,
Helsinki, Finland

Thesis Committee member
Associate Professor
Ville Hietakangas, Ph.D.
Faculty of Biological and Environmental Sciences
University of Helsinki,
Helsinki, Finland

Thesis Reviewers
Professor Gonghong Wei, Ph.D
Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine,
University of Oulu,
Oulu, Finland

Associate Professor
Mikael Björklund, Ph.D.
ZJU-UoE INSTITUTE
Zhejiang University,
Haining, China

Opponent
Professor
Anna-Liisa Levonen, MD, Ph.D.
A.I.Virtanen Institute for Molecular Sciences
University of Eastern Finland
Kuopio, Finland

Faculty representative
Associate Professor
Susanna Fagerholm
Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki,
Helsinki, Finland

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis
ISBN 978-951-51-5480-4 (paperback.)
ISBN 978-951-51-5481-1 (PDF)
ISSN 2342-3161 (print)
ISSN 2342-317X (online)
http://ethesis.helsinki.fi
Cover layout: Anita Tienhaara

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on three original publications, which are referred using the following Roman numerals:

I. **Göös H**, Kinnunen M, Yadav L, Varjosalo M. Protein interaction landscape of human TFs. Manuscript 2019.

Together with MK, HG generated the cell lines stably expressing 110 human TFs and screened the protein-protein interactions using both AP-MS and BioID- methods. HG performed all of the data-analysis, wrote the manuscript and prepared the figures.

II. **Göös H*,** Fogarty CL*, Sahu B*, Plagnol V, Rajamäki K, Nurmi K, Liu X, Einarsdottir E, Jouppila A, Pettersson M, Vihinen H, Krjutskov K, Saavalainen P, Järvinen A, Muurinen M, Greco D, Scala D, Curtis J, Nordström D, Flaumenhaft R, Vaarala O, Kovanen P, Keskitalo S, Ranki A, Kere J, Lehto M, Notarangelo LD, Nejentsev S, MD, Eklund KK#, Varjosalo M#, Taipale J#, Seppänen M#. Gain-of-function *CEPBE* mutation causes non-canonical autoinflammatory inflammasopathy. J Allergy Clin Immunol. 2019 Jun 12. pii: S0091-6749(19)30762-6.

HG generated stable cell lines expressing WT or mutant C/EBPε. She designed and performed a BioID- protein-protein interaction analysis and a nanostring analysis. HG had major roles in study design, the coordination of tasks between the authors, data-analysis, manuscript writing, submission procedure and figure preparation.

III. Kaustio M*, Haapaniemi E*, **Göös H*,** Hautala T, Park G, Syrjänen J, Einarsdottir E, Sahu B, Kilpinen S, Rounioja S, Fogarty CL, Glumoff V, Kulmala P, Katayama S, Tamene F, Trotta L, Morgunova E, Krjutškov K, Nurmi K, Eklund K, Lagerstedt A, Helminen M, Martelius T, Mustjoki S, Taipale J, Saarela J#, Kere J#, Varjosalo M#, Seppänen M#. Damaging heterozygous mutations in NFKB1 lead to diverse immunologic phenotypes. J Allergy Clin Immunol. 140(3):782-796, 2017

HG generated stable cell lines expressing WT or mutant NFKB1s. She designed and performed an AP-MS and a BioID protein-protein interactions analyses, phosphorylation analysis, an expression analysis by WB and a proteasome inhibition assay using these cell lines. HG designed and performed a NFKB1 expression analysis of patient PBMC-cells by MS. HG had major contributions to the data interpretation, manuscript writing and figure preparation.

*/# Equal contribution

ABBREVIATIONS

AP	affinity purification
AP-MS	affinity purification coupled to mass spectrometry
APEX	ascorbate peroxidase
bHLH	basic helix-loop-helix
BIFC	bimolecular fluorescence complementation
BioID	proximity dependent biotin identification
BirA	modified biotin ligase
bZIP	basic leucine zipper
C2H2-ZF	C2H2-zinc finger
CAIN	C/EBPε-associated autoinflammation and immune impairment of neutrophils
ChIP-seq	ChIP-sequencing
CoIP	co-immunoprecipitation
CVID	common variable immunodeficiency
DBD	DNA-binding domain
DDR	DNA damage response
DSB	double-strand DNA breaks
GOF	gain-of-function
GTF	general transcription factor
Hh	Hedgehog
IDR	Intrinsically disordered region
IKB	NFKB inhibitor protein
IKK	inhibitor of nuclear factor kappa-B kinases
KISS	kinase substrate sensor
LOF	loss-of-function
MaMTH	mammalian membrane two hybrid
MAPPIT	mammalian protein-protein interaction trap
mRNA	messenger RNA
MS	mass spectrometry
NES	nuclear export signals
NLS	nuclear localisation signal
NM1	nuclear myosin I
NR	nuclear receptor
ORC	origin recognition complex
PBMC	patient peripheral blood mononuclear cell
PCA	protein-fragment complementation
PIC	pre-initiation complex
PID	primary immunodeficiency
Pol-II	RNA polymerase II
PPI	protein-protein interaction
pre-RC	pre-replicative complex
PTM	post-translational modification
RNA-seq	RNA-sequencing
SAGA	Spt-Ada-Gcn5-acetyltransferase
SAINT	Significance Analysis of INTeractome
SGD	specific granule deficiency
SRF	serum response factor
TAD	transactivation domain
TAF	TBP-associated protein
TBP	TATA-binding protein
TF	transcription factor
TFBS	TF-binding sites
WT	wild type

TIIVISTELMÄ

Transkriptiotekijät eli transkriptiofaktorit (TF:t) ovat tärkeitä proteiineja geenien luennan säätelyssä. Ne vaikuttavat kaikkien solujen toiminnan ylläpidossa ja erilaistumisessa, ja ovat siten välttämättömiä mm. sikiönkehityksessä. Virheet TF:ien signaloinnissa voivat aiheuttaa vakavia kehityshäiriöitä ja sairauksia. Näin ollen TF:ien toiminta soluissa on hyvä ymmärtää mahdollisimman kattavasti, jotta häiriöihin voidaan puuttua esimerkiksi kehittämällä lääkehoitoa.

TF:t säätelevät geenien ilmentymistä sitoutumalla perimäainekseen, DNA-juosteeseen, aktivoiden tai estäen kohdegeenien luennan ja valmistuksen aktiivisiksi proteiineiksi. DNA:han sitoutuminen ei kuitenkaan ole tarpeeksi geenien luennan säätelemiseksi, vaan TF:t vuorovaikuttavat useiden muiden proteiinien kanssa halutun vasteen aikaansaamiseksi. Tämän väitöskirjatyön tavoitteena on kartoittaa ihmisen TF:ien proteiinivuorovaikutuksia sekä normaalioloissa että tautitiloissa. Työ koostuu kolmesta osajulkaisusta.

Ensimmäisessä osajulkaisussa onnistuimme kartoittamaan solumalleissa yli 7000 proteiinivuorovaikutusta 110 TF:lle. Iso osa näistä vuorovaikutuksista liittyy geenien luennan säätelyyn. Osa TF:stä vuorovaikutti myös erityisten proteiiniryhmien, kuten RNA-silmukointiin liittyvien tai tuma-aktiiniin liittyvien proteiinien, kanssa. Kartoitimme myös TF:ien keskinäisiä vuorovaikutuksia tutkitussa aineistossa ja teimme yllättävän havainnon, että 54 TF:a 110:stä vuorovaikutti Nuclear Factor-perheen (NFI) TF:ien kanssa. Tämä oli mielenkiintoinen löydös, sillä NFI- TF:t ovat välttämättömiä mm. hermoston, hampaiden, aivojen, luuston ja lihasten kehittymisessä sekä ne on yhdistetty usean syövän kehittymiseen. Tulostemme mukaan on mahdollista, että NFI:ien toimintaa säädellään muiden TF:ien kautta.

Toisessa osajulkaisussa tutkittiin C/EBPε-TF:n mutaation vaikutuksia soluissa. Mutaatio löydettiin suomalaisesta suvusta, jonka jäsenet kärsivät määrittelemättömästä primääristä immuunipuutoksesta. Solutasolla mutaatio aiheutti laajoja häiriöitä C/EBPε:n toiminassa: virheellinen TF sitoutui enemmän DNA:han, se vuorovaikutti huomattavasti vähemmän TF:ien toimintaa estävien proteiinien kanssa sekä häiritsi yli 460 geenin luentaa. Nämä muutokset johtivat häiriintyneeseen immuunipuolustukseen, mm. yliaktiiviseen non-kanonisen inflammasomin aktitiivisuuteen ja autoimmuunioireisiin. Uusi tauti nimettiin CAIN:ksi (C/EBPε-associated autoinflammation and immune impairment of neutrophils).

Kolmannessa osajulkaisussa tutkittiin kolmen erillisen NFKB1 TF-mutaation vaikutuksia primäärissä immuunipuutoksessa kolmessa eri suomalaissuvussa. Mutaatiot eri kohdissa proteiinia vaikuttivat erilaisilla mekanismeilla, mutta jokainen niistä aiheutti virheitä immuunipuolustuksen toimintaan.

Kaiken kaikkiaan tämä väitöskirja tarjoaa tärkeän aineiston TF:ien proteiinivuorovaikutuksista, jota voidaan käyttää mm. uusien lääkkeiden ja hoitomuotojen kehittämiseen. Siinä myös kartoitetaan, miten yksittäinen virhe TF:ssa voi aiheuttaa ongelmia monella eri geenien luennan säätelyn tasolla ja miten samassa TF:ssa eri kohdissa olevat mutaatiot voivat aiheuttaa tauteja erilaisilla mekanismeilla.

ABSTRACT

Transcription factors (TFs) are one of the most important groups of proteins for the development and differentiation of cells. They control the gene expression of all cells in all stages of development. Defects in TF signalling may lead to severely altered development and diseases. However, while TF DNA binding has been widely studied, we are still lacking a systems-level understanding of human TF signalling. TFs' action in gene expression regulation is highly dependent on their interactions with multiple proteins, such as cofactors, dimerization partners, chromatin modulating proteins, enzymes, inhibitory proteins and general TFs. Therefore, the aim of this study is to shed light on TF protein-protein interactions and, more specifically, to examine the effect of TF mutations found in primary immunodeficiency patients.

A comprehensive interactome analysis of 110 TFs revealed over 7,000 TF protein-protein interactions, most of which are nuclear and play a role in transcriptional regulation **(I)**. The large number of TF interactions discovered in this study enabled us to conduct a systems-level analysis that revealed groups of TFs with specific biological functions, such as actin and myosin signalling and RNA splicing. Interestingly, 54 of the TFs studied interacted with the nuclear factor family of TFs. Nuclear factors are known to control a number of genes in development; for instance, they are essential for central nervous system, tooth, brain, skeletal, lung and muscle development. In addition, they are linked to several cancer types. Our data suggest that transcription control by NFIs may be regulated by nuclear factor interactions with other TFs.

A219H mutation in the C/EBPε TF was found in a Finnish family with immunodeficiency and autoinflammatory syndrome (II). A data-driven multiomics study of the mutation revealed a novel TF-related disease mechanism; mutation decreased association with transcriptional repressors, increased chromatin binding and widely dysregulated transcription. These changes resulted in disturbed non-canonical inflammasome activation due to the increased expression of NLRP3 and constitutively expressed CASP5.

Three different damaging mutations in NFKB1 resulted in diverse immunological phenotypes due to different mechanisms (III): H67R led to decreased nuclear entry, reduced association with RelB and decreased transcriptional activity; I553M led to decreased phosphorylation of S893 and p907 and enhanced p105 subunit degradation upon TNF treatment; and R157X led to an almost total loss of NFKB1 subunits due to proteasome-mediated dominant negative degradation.

This study provides valuable information on TF protein-protein interactions at systems level (I). In addition, this study provides examples of how single TF mutation may affect TF signalling on many levels, such as in protein interactions, DNA binding and transcription (II) and how different mutations in the same TF can have different outcomes (III). TFs are downstream players of many signalling cascades and targeting TF protein interactions can offer a high degree of specificity in future therapeutics applications.

I LITERATURE REVIEW

1. Transcription factors (TFs)

'The central dogma of biology,' is the translation of genetic information from DNA sequence to RNA and finally into active proteins. This process allows cells to respond to external and internal stimuli by changing the amount of RNA and active proteins using multiple subprocesses, such as transcription, translation and control of protein activity. These processes are regulated in chromatin, DNA, RNA and protein levels, of which the regulation of transcription in chromatin and DNA is the first and most important step.

The human genome consists of over three billion DNA base pairs, resulting in a chain over two meters long. To fit within the nucleus and prevent unwanted gene transcription, DNA is highly winded around histone proteins to form nucleosome complexes that are further packed into chromatin. In non-dividing cell, chromatin can be detected in lightly packed, transcriptional active euchromatin form and highly packed, transcriptionally inactive heterochromatin form. Protein-protein interaction (PPI) mediated post-translational modifications (PTMs) of histones and PPI-mediated recruitment of chromatin remodelling proteins control the unpacking of the DNA chain and thus regulate access of basal transcription machinery on gene promoters.

Transcriptional regulation is tightly controlled by different groups of proteins, such transcription factors (TFs), chromatin remodellers and histone modifying enzymes. In addition, small RNA molecules, such as miRNAs and siRNAs act as gene expression regulators. TFs are DNA-binding proteins that recognise and bind sequence-specific DNA motifs on gene promoters and enhancers through their DNA-binding domains (DBDs) to either activate or repress gene expression. TFs can regulate transcription either by recruiting chromatin modelling proteins to affect the opening chromatin state or by directly binding to promoters and enhancers to regulate the general transcription machinery's access to the transcription starting sites. In addition to regulating transcription, DNA-binding TFs have a role in DNA-modifying processes, such as in DNA replication, repair and rearrangement (Xie et al., 2011).

Out of all >20,000 human proteins, 6–9% (~1400–1900) are predicted to be TFs (https://www.proteinatlas.org; (Vaquerizas et al., 2009; Babu et al., 2004; Fulton et al., 2009). Previous manual curation of potential TFs resulted in 1639 known or likely human TFs (Lambert et al., 2018). Through their regulation of gene transcription, TFs are the key factors in many biological processes, including proliferation, apoptosis and differentiation. TFs are also central in developmental processes, since acting as pioneer factors, they elicit the initial cell reprogramming in embryonic development. Given their vitalness in many biological processes, TFs' impact on numerous diseases is unsurprising: TFs are overpresented in oncofusion proteins of soft tissue tumors (Mertens et al., 2016), and they take part in numerous pathological conditions such as inflammation, neurodegenerative diseases and cancer (Han et al., 2017; Martin-Martin et al., 2017; Wang et al., 2018).

1.2 TF structure and classification

The prototypical TF contains at least one DBD and one or more effector domains (Figure 1). Effector domains include (a) transactivation domains (TADs) that interact with components of the basal transcription machinery; (b) domains that mediate the interactions with other DNA-specific TFs (dimerization); (c) domains that mediate the PPIs with other cofactors, such as chromatin-modifying enzymes; (d) signal sensing domains (e.g., ligand-binding domains) and (e) domains with enzymatic activity (e.g., histone acetylase activity;(Frietze et al., 2011; Lambert et al., 2018). Effector domains, through their interactions with other proteins, may regulate gene expression by inducing changes in chromatin opening states, generating necessary platforms for cofactors by binding other factors or by changing the basal transcription machinery's conformation to either induce or repress the RNA polymerase II (Pol-II) mediated transcription. Effector domains are seldom specific to one co-regulatory protein; they may bind multiple different co-regulatory proteins, and the same co-regulatory protein can bind different effector domains.

TFs are mainly classified by their DBD structure (Lambert et al., 2018), but they can also be classified by other characteristics, such as functionality (Qian et al., 2006; Wingender et al., 2018; Yang et al., 2010). For instance, TFs can be classified as general transcription factors (GTFs), which are part of the basal transcription machinery, or upstream TFs that regulate the activity of the GTFs and RNA Pol-II. In this thesis, TFs most often refer to these upstream TFs. TFs can further be divided into two functional groups: TFs that are constitutively active and TFs that require activation. TFs are also classified based on their DBD structure. From all manually curated TFs (1639), the majority (747) have C2H2-zinc finger (C2H2-ZF) DBDs (Lambert et al., 2018) Figure 2). Of the remaining TFs, 196 have homeodomain DBDs, 108 have basic helix-loop-helix (bHLH) DBDs and 54 have basic leucine zipper (bZIP) DBDs. The remaining groups have less than 50 TFs (Figure 2). In total, Lambert et al. listed 65 different DBDs, of which 12 are actually a combination of two different DBDs. However, only 3% of TFs had these two different types of DBDs. Various TFs contain several copies of the same DBD, most of which are C2H2-ZF TFs that might have more than 30 copies of same DBD sequence.

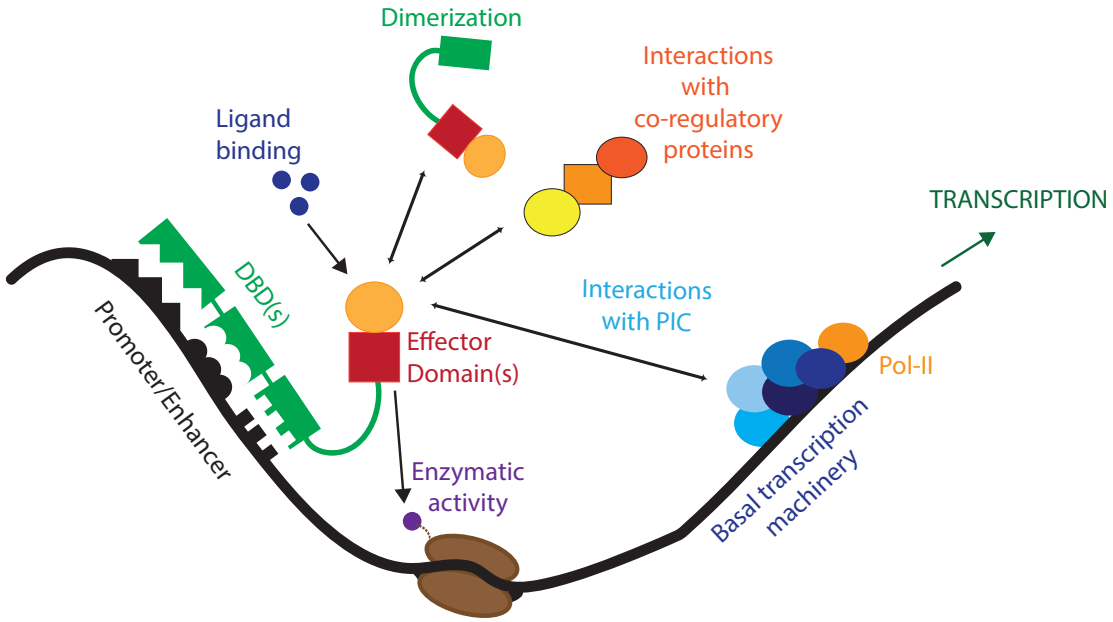


Figure 1: Schematic model of TF domain structure and effector domain functions. Similar schematic domain organization of TFs are used in figures 3, 4 and 5.

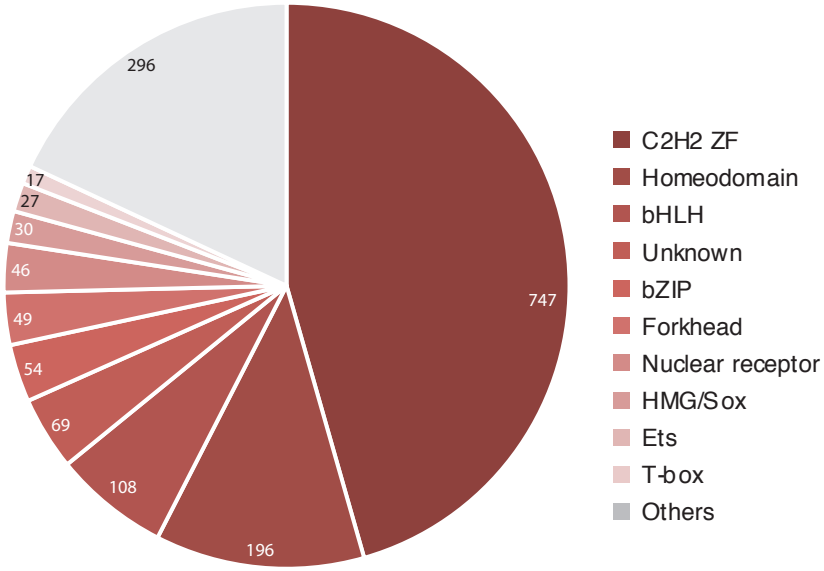


Figure 2: DNA-binding domains of 1639 TFs

1.3 TF DNA binding

TFs have a high affinity to specific DNA sequences, known as TF-binding sites (TFBSs). A TF's affinity to a specific TFBS can be more than 1000 times higher compared to its affinity to an unspecific sequence (Geertz et al., 2012). TFBSs are typically short (6–12 bp) DNA sequences, which are normally repeated several times within the target gene's cis-regulatory sequence. Cis-regulatory elements, such as enhancers and promoters, are the areas of non-coding DNA that serve as regulatory elements for different genes.

TFBS identifying methods, such as ChIP-seq, SELEX and protein-binding microarrays, have recently improved remarkably (reviewed in (Inukai et al., 2017), leading to an increased number of identified TFBS that are available in databases, such as JASPAR (Khan et al., 2018), TRANSFAC (Matys et al., 2006) and HT-SELEX (Jolma et al., 2013). TFBSs are not always straightforward: the most favourable TFBS for a particular TF might depend on specific conformations of DNA (Samee et al., 2019), DNA methylation statuses and protein interactions (Yin et al., 2017; Jolma et al., 2015). Generally, the consensus sequences (the TFBSs with the highest affinity) for each TFs are reported, but TFs may have other biologically relevant TFBSs that they bind to with different levels of affinity (Jolma et al., 2011; Jiang et al., 1993). The lowest affinity is to non-specific DNA, and it allows TFs to slide along the DNA. The higher affinity in specific binding sites allows the TF to bind the TFBS long enough to regulate the transcription (Jolma et al., 2011). TF DNA binding is also affected by genomic variations, mainly in non-coding DNA, which are extensively reviewed in (Deplancke et al., 2016).

TFs bind to their TFBS by DBDs with electrostatic interactions, such as hydrogen bonds, and Van der Waals forces. The specificity to certain TFBS may come from the specific amino acid organisation in the DBDs (Baker et al., 2007).

2. TF protein-protein interactions

Transcriptional regulation is the result of cross-talk between the TFs, basal transcription machinery and chromatin landscape (Li, Wang, et al., 2015) . However, the TFs are the only components in this network that are able to bind highly specific promoters or enhancers. Consequently, the complicated and multilayer transcriptional regulation system includes not only the direct binding of TFs in the target gene's binding sites, but also the complex network of interactions between the TFs and TF-binding proteins. This network includes interactions with cofactors; chromatin remodellers; proteins in the basal transcription machinery and Mediator complex; interactions with TF modulating proteins, such as phosphatases and kinases and interactions with dimerization partners, subunits and inhibitory proteins. Recently introduced phase separation model of TF PPIs indicates that many of TF PPIs are dynamic, structured and formed between intrinsically disordered region (IDRs) of TFs' effector domains (Boija et al., 2018).

2.1 TF activity regulation by protein-protein interactions

Cells communicate with external environments by changing the level of expressed genes and proteins. This happens through signalling cascades, which can be short (e.g., nuclear receptor [NR] signalling) or more complicated (e.g., Wnt or Hedgehog [Hh] signalling). However, the cascades control the activity of downstream TFs that regulate the target gene's expression. To regulate gene expression, the TFs must be activated or inhibited through a process such as cleavage, PTMs, TF binding (and releasing) to (and from) inhibitory proteins, dimerization, ligand binding, increased or decreased TF synthesis, localisation changes, or, most commonly, through a combination of these various mechanisms. The accessibility of TFBSs in DNA and the availability of co-regulatory proteins also affect TF activity. Besides external stimuli, TFs may also respond to intracellular signalling by changing the activity state. However, many of these activity control steps involve TF interactions with other proteins.

A direct way to change the activation status of a TF is through ligand binding to the TF. A classic example is NRs, which are directly activated by lipid-soluble hormones binding to their ligand-binding domains. NR ligands, such as Vitamin D3 for Vitamin D receptor and testosterone for Androgen receptor, are often steroid hormones, which can pass the cell membrane and directly bind the NRs inside the cell (Sever et al., 2013). TF-binding ligands may also be proteins. For example, Hippo signalling pathway activation leads to the activation of YAP and TAZ proteins that can serve as ligands for several TFs, such as TEADs that bind YAP and TAZ with a YAP-binding pocket (Li et al., 2018). However, some nomenclature unclarity exists regarding whether these protein ligands are referred to as ligands, cofactors or activators.

Some TFs are expressed in a long form that is cleaved depending on the TF activation status. For example, upon activation of the NFKB pathways, p105 and p100 forms of NFKB proteins are cleaved into p52 and p50 parts that can enter the nucleus and, depending on the dimerization partners, either activate or repress the target gene expression (Oeckinghaus et al., 2009). In contrast, GLI3 cleavage into a repressive form and GLI2 complete degradation are inhibited in response to the Hh pathway activation by the Hh ligand (Varjosalo et al., 2008). These cleavages and

degradations require TF protein interactions with proteases, such as NFKB1 interaction with 20S proteasome units (Moorthy et al., 2006).

Proteases also mediate the level of TFs in a cell by degradation. To control gene expression, the TF level in cells is highly regulated by synthesis and degradation. TF concentration is often controlled by a negative feedback loop in which the TF itself balances its synthesis and disposal (Pan et al., 2006; Bornstein et al., 2014; Harris et al., 2005). The synthesis and degradation of TFs, like any other protein, requires multiple interactions with other proteins; for example, TFs have been found to interact with ribosomal proteins, endoplasmic reticulum proteins, transport proteins and ubiquitin ligases (Li, Wang, et al., 2015).

The activity of some TFs, such as NFKB1 and bHLH TFs, is also controlled by binding them to an inhibitory protein that may block the nuclear localisation of the TF or its interactions with other genes or co-regulatory proteins. In an inactive state, NFKB1 is held in cytoplasm bound to NFKB inhibitor proteins (IKBs; IKBA IKBB, IKBE and IKBZ; (Totzke et al., 2006; Scherer et al., 1995; Oeckinghaus et al., 2009). The bHLH TFs are inhibited by the binding of inhibitor of DNA-binding proteins (IDs), helix-loop-helix proteins that bind bHLH TFs to form non-functional heterodimers (Ling et al., 2014). IDs have been shown to affect growth, differentiation and cancer (reviewed in (Ke et al., 2018; Ling et al., 2014).

2.1.1 Post-translational modifications in TF activity regulation

The TFs' activity state is often regulated by PTMs (Filtz et al., 2014). This process requires substrate-enzyme PPIs, for example with kinases, phosphatases, acetyltransferases, deacetylases, methyltransferases, demethylases, ubiquitin ligases, ubiquitin hydroxylases, carboxylases, (de)hydroxylases, glycosyl transferases and SUMO transferases.

TF PTMs may regulate the TF activity through various mechanisms (Figure 3; (Muratani et al., 2003; Tootle et al., 2005; Whitmarsh et al., 2000): First, the PTMs can affect the PPIs to other TFs, dimerization partners, co-regulatory proteins or the basal transcription machinery. Second, the PTMs may affect the TFs' DNA binding. Third, the TFs are often targeted to a specific cleavage or proteasomal degradation by the PTMs. Fourth, the PTMs may control the TFs' translocation to the nucleus, and their time spent there. Fifth, TF stability may depend on the PTMs, and, finally, the PTMs may regulate the binding of other PTMs to the same TF or nearby proteins. In addition to binding directly to the TFs, the PTMs play an important role in gene expression regulation by modifying other transcription-related proteins, such as histones, cofactors and inhibitory proteins. Histone modifications are crucial in changing the accessibility of DNA to the basal transcription machinery and TFs (as reviewed in (Zhao et al., 2018; An, 2007; Fan et al., 2015)). TF interactions with histone-modifying proteins will be discussed later.

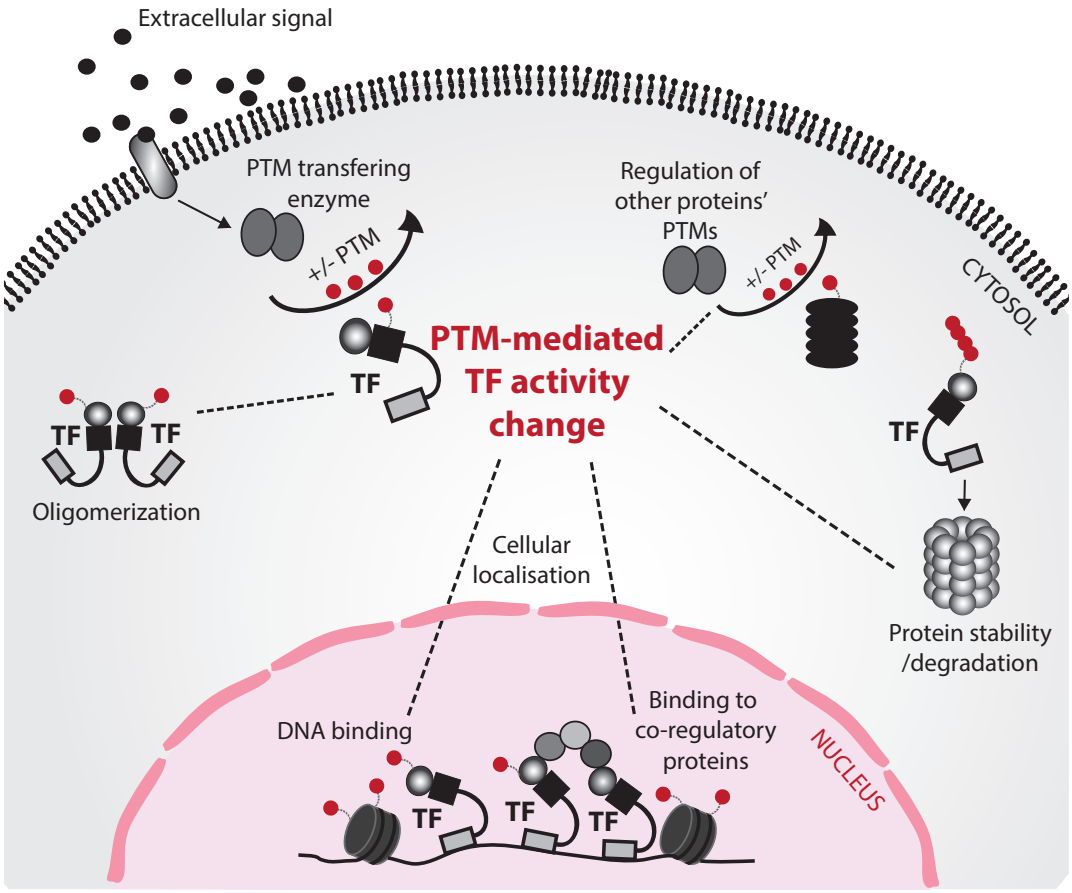


Figure 3. TF activity regulation by PTMs. PTMs in TF may affect its oligomerization, localisation, DNA binding, interaction to other proteins, stability or they can induce PTMs to other proteins. Red symbols indicate the PTMs, such as phosphorylation or ubiquitylation, that are transfered to TFs trough PTM transfering enzymes. Used schematic domain organization of TFs is described more in details in Figure 1.

A good example of TF activation by PTMs are STATs, which, upon activation of the JAK-STAT pathway, are phosphorylated in their C-terminal tyrosine by Janus kinases (JAKs;(Decker et al., 2000). This procedure allows for STAT dimerization and entrance into the nucleus.

While many TFs are constitutively nuclear, several, such as STATs, shuttle in and out from the nucleus. Nuclear imports and exports are often coded in the TF sequence as nuclear localisation signals (NLSs) and nuclear export signals (NESs). The phosphorylation of signal peptides can both induce or repress nuclear localisation (Nardozzi et al., 2010; Whitmarsh et al., 2000). For example, DYRK1A-mediated phosphorylation of GLI1 NLSs increases the nuclear import (Ehe et al., 2017), but DYRK-mediated heavy phosphorylation of NFAT NLSs blocks the nuclear import (Sharma et al., 2011). In addition, NFAT requires dephosphorylation for efficient nuclear transport. Similar to NLSs, the phosphorylation of NESs may induce an export, such as a p53 export from the nucleus in response to DNA damage (Zhang et al., 2001).

Besides the direct phosphorylation of TFs, nuclear localisation can be controlled by the phosphorylation of TF localisation controlling proteins, such as inhibitory proteins. For example, the phosphorylation of IKBs by inhibitor of nuclear factor kappa-B kinases (IKKs) releases p52 (NFKB1), allowing it to enter the nucleus (Oeckinghaus et al., 2009).

TF DNA binding can also be affected by phosphorylation. For example, FOXO1 phosphorylation in the DBD (S256) suppresses its binding to the DNA (Zhang et al., 2002). Indeed, FOXO proteins are extensively post-translationally modified (including phosphorylation, methylation, ubiquitylation and acetylation) and serve as a good example of how TFs are directly modified by kinases, phosphatases, ubiquitin ligases, acetyltransferases, deacetylases and methyltransferases altering the protein stability, DNA binding, localisation, interactions with other proteins and regulation of other PTMs. These mechanisms have been extensively reviewed by Brown and Webb (Brown et al., 2018).

Ubiquitylation is most common for marking a protein for degradation, but it may also regulate gene expression in a non-proteolytic way as direct ubiquitylation may affect the TF activity (Ndoja et al., 2014; Muratani et al., 2003). Inhibition of receptor activated SMADs by non-proteolytic ubiquitylation blocks the formation of active SMAD-dimers or binding to DNA without directing them to degradation (Tang et al., 2011; Inui et al., 2011). Similarly, the activity of PPARγ can be inhibited by non-proteolytic ubiquitination by Smurf2. Besides inhibition, non-proteolytic ubiquitylation may also enhance the transcriptional TF activity. For example, p53 is stably monoubiquitylated, resulting in nuclear localisation and increased DNA-binding affinity (Landre et al., 2017).

2.2 TF cooperativity and oligomerization

Under the physiological conditions, most TF DNA-binding sites are not occupied, so the identification of the binding sequence alone is not a reliable predictor of TF binding (Wasserman et al., 2004). In most cases, TFs need to cross-talk, or cooperate, with other TFs and cofactors to be able to bind specifically to DNA and mediate the signal further to the basal transcription machinery and Pol-II. Cooperativity allows the TFs enough flexibility and specificity to regulate the total transcription; for example, in developmental processes, cooperativity allows multiple TFs to regulate the generation of a large number of cell types (Spitz et al., 2012; Reiter et al., 2017).

TF cooperativity can occur in three overlapping levels: in DNA-binding, independent of DNA-binding and via PTMs. Cooperative TF DNA binding enhances the cell type specific binding as cooperative binding only occurs if all necessary TFs are expressed in appropriate concentrations (Barozzi et al., 2014; Heinz et al., 2010).

DNA-mediated cooperative TF binding appears when multiple TFs bind synchronously to binding sites that have a specific spacing and orientation relation (Figure 4; (Jolma et al., 2015). This occurs passively when several TFs bind to DNA without a physical PPI (Figure 4; (Reiter et al., 2017). Synergistic binding allows the TFs change the DNA accessibility while individual TFs are insufficient to complete the DNA binding with nucleosomes (Lickwar et al., 2012; Moyle-Heyrman et al., 2011). In active binding, TFs form protein interactions, such as homodimerization or heterodimerization, which provide increased specificity and an affinity to a regulatory element (Figure 4). Active binding can be DNA-mediated, in which case the binding sites guide the TFs together, or interaction-mediated, in which case contact between the TFs occurs before binding to the DNA (Morgunova et al., 2017). DNA binding may facilitate the multimer formation by increasing TF affinity to one other, such as by changing the TF conformation. Finally, cooperative TF binding may also occur in

sequential order: first the pioneer factor binds the DNA, initiating chromatin remodelling and allowing the other TFs to follow and recognise their binding sites (Figure 4; (Zaret et al., 2011; Iwafuchi-Doi et al., 2014).

Active TF cooperative DNA binding is not restricted to TF interactions with other TFs but includes interactions with co-regulatory proteins and even higher complexes (Spitz et al., 2012). A good example of cooperative binding is enhancesome, in which eight TFs are needed to bind the IFN-β enhancer region to reach the compulsory affinity and stability to recruit the KAT2A, CBP/p300 and switch (SWI)/SWF complexes to acetylate the nucleosomes, remodel the chromatin and enable the assembly of the basal transcription machinery (Panne, 2008).

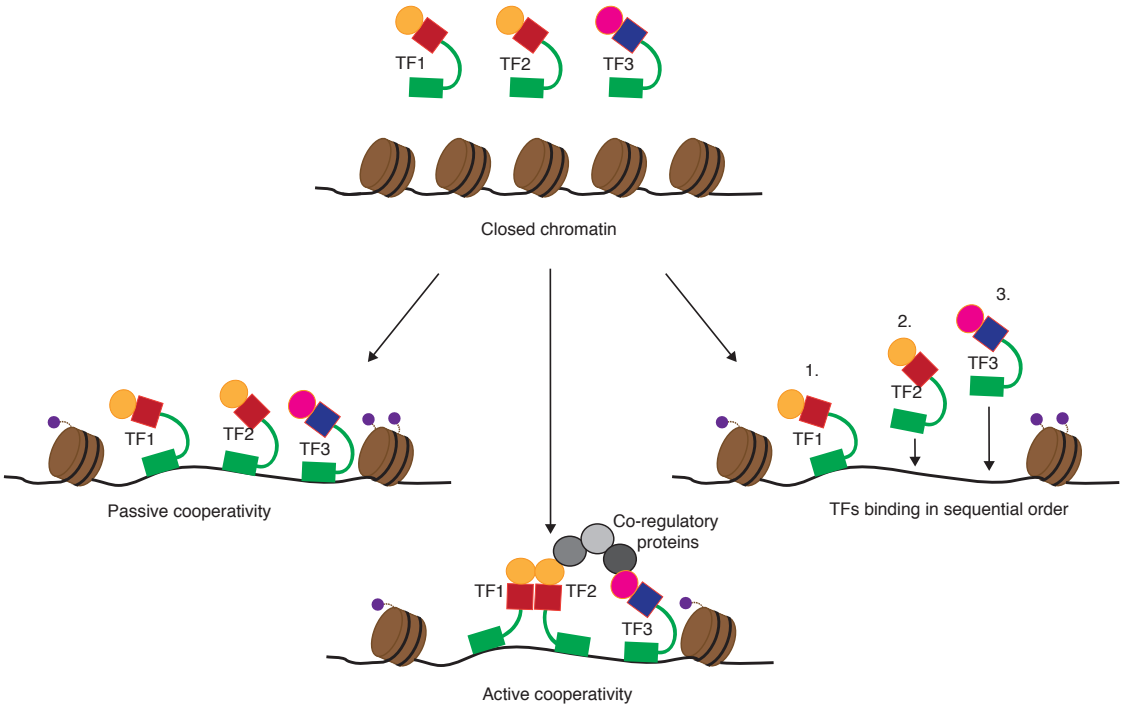


Figure 4. Cooperative DNA-binding models of TFs. Cooperative binding may be passive (TFs bind DNA without physical contact) or active (TFs bind each other directly or through co-regulatory proteins). Cooperative binding may occur in sequential order where binding of certain TF is needed for binding the next TF. Used schematic domain organization of TFs is described more in details in Figure 1.

Many TFs are known to form homodimers, heterodimers or even higher degrees of homomeric and heteromeric multimers (Amoutzias et al., 2008). Some TF families, such as HNFs, have specific dimerization domains for dimerization, whereas others, such as bHLH's and bZIP's dimerization domains, are not specific to the dimerization. Different multimer compositions might act as activators and repressors, and the same oligomer, depending on other interactions, can act as both. For example, p50 and p52, the activated forms of NFKB1 and NFKB2, together with other Rel-family proteins, form nine dimeric complexes (Oeckinghaus et al., 2009). Not all of these complexes are transcriptional activators: p50 and p52 homodimers often act as repressors. Regardless, binding Bcl-3 might change the regulation status to activator.

Cooperativity between the TFs, co or gene-regulatory proteins and basal transcription machinery may also occur via PTMs. Various cofactors act as enzymes that mediate the PTMs to their target proteins. Acetylation, methylation and other histone modifications, protein phosphorylation and other PTMs are part of the TF communication to the actual transcription machinery. This might occur through stable protein complexes or transient PPIs (Reiter et al., 2017).

2.3 TF protein-protein interactions with the basal transcription machinery

Eukaryotic gene transcription is performed by RNA Pol-II, which binds in highly conserved DNA sequences referred to as core promoters. Core promoters serve as binding sites for the basal transcription machinery (also referred to as a pre-initiation complex [PIC]), which is composed of Pol-II, Mediator complex and multiple GTFs (TBP, TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH; Table 1), of which most are actually multimeric protein complexes. GTFs are essential for the Pol-II recognition of the promoter and transcription initiation. After the complete assembly of the PIC and unwinding of the double-stranded DNA, Pol-II escapes the promoter and can perform the transcription elongation alone. The Pol-II activity is often regulated by TFs, but as the TFs do not directly bind the Pol-II components, the signal is transferred directly through PPIs with GTFs and/or Mediator complexes or indirectly through other cofactors, such as chromatin remodelling complexes.

2.3.1 TF protein-protein interactions with general transcription factors

To begin assembling the PIC, the TATA-binding protein (TBP) binds the TATA box (consensus TATAAWR) of the core promoter. Next, it recruits 13–20 TBP-associated proteins (TAFs) to form the TFIID complex (Table 1). The TFIID interacts with other activating TFs and also reads the genome's epigenetic marks with its subunits TAF1 and TAF3 (Vermeulen et al., 2007; Wassarman et al., 2001).

Besides in the TFIID complex, the TBP exists in at least one other transcriptionally active complex: in coactivator complex SAGA (Spt-Ada-Gcn5-acetyltransferase, Table 1; (Petrenko et al., 2019; Kuras et al., 2000). The SAGA complex binds the TBP with its two subunits, Stp3 and Stp8 (Sermwittayawong et al., 2006; Mohibullah et al., 2008) to help recruit TBP to TATA-like promoters (TATA-box with one or two mismatches; (Han et al., 2014; Sermwittayawong et al., 2006). The SAGA complex has recently been identified as a general cofactor for Pol-II- transcription (Baptista et al., 2017; Bonnet et al., 2014). It also shares multiple members, such as TAF9 and TAF10, with TFIID and other co-regulatory complexes (Table 1; (Helmlinger et al., 2017).

Several studies have suggested that both TFIID and SAGA contribute to the transcription of numerous, if not all, genes (Fischer et al., 2019), but the expression might be dominated by either of them (Lee, Causton, et al., 2000; Huisinga et al., 2004). Different promoters are indicated to favour either SAGA or TFIID, and it has been suggested that SAGA and TATA-like promoters might depend more on the presence of transcriptional activators (regulated genes) than TFIID and TATA-promoters (housekeeping genes; (de Jonge et al., 2017). However, this proposition is still controversial as a depletion of TAFs or SAGA components reduces the transcription significantly, and equally, in both cases (Fischer et al., 2019). It is reported that actually SAGA act as general cofactor in expression of all genes (Baptista et al., 2017).

Some TFs interact directly with TFIID and SAGA complex members. The TFs might form direct contacts with TAFs, TBP and SAGA complex members, leading to conformational PIC changes (Joo et al., 2017), or recruit other cofactors to mediate the signal. For example, TAF1 interacts with SP1 (Suzuki et al., 2003), p53 (Li et al., 2004), PAX3 (Boutet et al., 2010) and JUN (Lively et al., 2001). Some TFs also interact with multiple TFIID components; for instance, PAX6, TP53 and FOS interact with both TBP and TAF1 (Metz et al., 1994; Qadri et al., 2002; Cvekl et al., 1999). Besides TAF1 and TBP, other TFIID components have also been found to bind TFs.

Members of the SAGA complex have also been detected interacting with different TFs. KAT2A (also known as GCN5), the main catalytic subunit of the SAGA complex, as an example, interacts with C/EBPβ (Wiper-Bergeron et al., 2007), PBX1-E2A dimer (Holmlund et al., 2013) and various SMADs (SMAD6, SMAD3, SMAD2 and SMAD9; (Kahata et al., 2004). Some TFs interact with multiple SAGA subunits: TP53 interacts with TADA2B and KAT2A (Gamper et al., 2008), whereas MYC and E2F interact with KAT2A and TTRAP (Lang et al., 2001; Zhang et al., 2014; Liu et al., 2003). Out of all the GFTs, TFIID and SAGA appear to have the most direct TF contacts. TF binding to TFIID or SAGA components is a direct way of regulating the PIC conformation and Pol-II activity.

After the binding of TFIID and SAGA, TFIIA binds upstream of the TATA-box. This is followed by the binding of TFIIB. Binding of TFIIB changes the conformation of the TPB/TATA complex, enhances its stability (Hieb et al., 2007) and leads to the recruitment of TFIIF, which is bound to Pol-II (Thomas et al., 2006). Finally, after the binding of TFIIF, the PIC is completed by the binding of TFIIE and TFIIH. Binding the TFIIE assists the assembly and orientation of the final subunit, TFIIH (Schilbach et al., 2017). Both TFIIE and TFIIH are necessary to proceed from initiation to transcription elongation (Holstege et al., 1996). TFIIH has ATPase activity that is needed for promoter melting, transcription initiation and escaping from the promoter; in the absence of TFIIH and ATP, Pol-II might stall on the promoter (Dvir et al., 1997; Kugel et al., 1998; Kumar et al., 1998).

The direct TF PPIs with TFIIA, TFIIB, TFIIF and TFIIE subunits are not well known. However, GTF2B (TFIIB) might interact with FOXF2 (Hellqvist et al., 1998), and the subunits of TFIIF might interact with AR (McEwan et al., 1997) and MYC (McEwan et al., 1996). TFIIH components interactions with TFs have been studied more extensively: The CDK7 phosphorylates often NRs (Rochette-Egly, 2003), such as ESR1 (Chen et al., 2000), RARG (Bastien et al., 2000), AR (Lee, Duan, et al., 2000) and other TFs (e.g., SP1(Chuang et al., 2012). TP53 interacts with at least three TFIIH subunits (CDK7, CCNH and MNAT1 (Ko et al., 1997).

2.3.2 TF protein-protein interactions with the Mediator complex

The Mediator complex is a large, multisubunit protein complex whose basic function is to mediate regulatory signals from TFs to Pol-II (Table 1, (Allen et al., 2015; Borggrefe et al., 2011). It is often considered to be part of the PIC as it extensively interacts with Pol-II and broadly regulates the assembly of the PIC (Harper et al., 2018). The Mediator complex consists of 26 subunits in mammals, and the subunit composition may change according to the biological function (Harper et al., 2018). Its structure is highly dynamic, allowing it to flexibly change the conformation upon the binding of ligands, such as TFs and PIC components (Poss et al., 2013).

Numerous TFs have been identified to interact with one or more Mediator complex subunits (Currie et al., 2017; Yin et al., 2014; Poss et al., 2013). As TFs do not directly bind the Pol-II, they may affect Pol-II activity by changing the Mediator complex conformation either directly by binding its components or indirectly, for instance by interacting with chromatin remodelling complexes that mediate the signal to the Mediator complex (Poss et al., 2013; Harper et al., 2018). Different TFs may interact with different surfaces or subunits of Mediator and therefore induce diverse structural changes to the Mediator complex. These changes may affect the Mediator-Pol-II interactions and/or Mediator interactions to other Pol-II related cofactors, leading to changes in the transcription (Poss et al., 2013). In an example scenario, p53 is detected to activate the transcription by interacting with different Mediator subunits using its C-terminal activation domain, thus altering the Mediator conformation that affects the Pol-II C-terminal phosphorylation and finally, resulting in activated transcription (Meyer et al., 2010).

The TFs are often the mediating proteins between the cell signalling pathways and basal transcription machinery and/or Mediator complex. Subsequently, the basal transcription machinery and Mediator complex forward the signal to Pol-II. This requires either multiple PPIs between the TFs and PIC proteins or proteins that transfer the activation to the PIC. Some TFs interact directly with PIC components, such as with TAFs and Mediator complex members, but they may also recruit other cofactors, such as chromatin remodelling complexes, to mediate the signal to the PIC. The lack of TF interactions with TFIIA, TFIIB, TFIIF and TFIIE might indicate that the TFs communicate with the PIC mainly through the TFIID, SAGA, TFIIH and Mediator complexes.

Table 1. Protein composition of PIC subcomplexes obtained from Corum-database (http://mips.helmholtz-muenchen.de/corum/).

Protein description	Gene name	UniProt
TFIID complex (Corum complex 484)		
TATA-box-binding protein	TBP	P20226
Transcription initiation factor TFIID subunit 1	TAF1	P21675
Transcription initiation factor TFIID subunit 10	TAF10	Q12962
Transcription initiation factor TFIID subunit 11	TAF11	Q15544
Transcription initiation factor TFIID subunit 12	TAF12	Q16514
Transcription initiation factor TFIID subunit 13	TAF13	Q15543
Transcription initiation factor TFIID subunit 4	TAF4	O00268
Transcription initiation factor TFIID subunit 5	TAF5	Q15542
Transcription initiation factor TFIID subunit 6	TAF6	P49848
Transcription initiation factor TFIID subunit 7	TAF7	Q15545
Transcription initiation factor TFIID subunit 9	TAF9	Q16594
SAGA complex, GCN5-linked (Corum complex 6643)		
Adenosine deaminase	ADA	P00813
Ataxin-7-like protein 1	ATXN7L1	Q9ULK2
Ataxin-7-like protein 2	ATXN7L2	Q5T6C5
Ataxin-7-like protein 3	ATXN7L3	Q14CW9
Histone acetyltransferase KAT2A	KAT2A	Q92830
SAGA-associated factor 29	SGF29	Q96ES7
STAGA complex 65 subunit gamma	SUPT7L	O94864
TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5L	TAF5L	O75529
TAF6-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 6L	TAF6L	Q9Y6J9
Transcription factor SPT20 homolog	SUPT20H	Q8NEM7
Transcription initiation factor TFIID subunit 10	TAF10	Q12962
Transcription initiation factor TFIID subunit 12	TAF12	Q16514
Transcription initiation factor TFIID subunit 9	TAF9	Q16594
Transcription initiation factor TFIID subunit 9B	TAF9B	Q9HBM6
Transcription initiation protein SPT3 homolog	SUPT3H	O75486
Transcriptional adapter 2-beta	TADA2B	Q86TJ2
Transcriptional adapter 3	TADA3	O75528
Transformation/transcription domain-associated protein	TRRAP	Q9Y4A5
Ubiquitin carboxyl-terminal hydrolase 22	USP22	Q9UPT9
TFIIA complex (Corum complex 489)		
Transcription initiation factor IIA subunit 1	GTF2A1	P52655
Transcription initiation factor IIA subunit 2	GTF2A2	P52657
TFIIB		
Transcription initiation factor IIB	GTF2B	Q00403
TFIIF complex (Corum complex 153)		
General transcription factor IIF subunit 1	GTF2F1	P35269
General transcription factor IIF subunit 2	GTF2F2	P13984
TFIIE complex (Corum complex 152)		
General transcription factor IIE subunit 1	GTF2E1	P29083
Transcription initiation factor IIE subunit beta	GTF2E2	P29084
TFIIH transcription factor complex (Corum complex 1009)		
CDK-activating kinase assembly factor MAT1	MNAT1	P51948
Cyclin-dependent kinase 7	CDK7	P50613
Cyclin-H	CCNH	P51946
General transcription factor IIH subunit 1	GTF2H1	P32780
General transcription factor IIH subunit 2	GTF2H2	Q13888
General transcription factor IIH subunit 3	GTF2H3	Q13889
General transcription factor IIH subunit 4	GTF2H4	Q92759
General transcription factor IIH subunit 5	GTF2H5	Q6ZYL4
TFIIH basal transcription factor complex helicase XPB subunit	ERCC3	P19447
TFIIH basal transcription factor complex helicase XPD subunit	ERCC2	P18074

Protein description	Gene name	UniProt
Mediator complex (Corum complex 230)		
Cyclin-C	CCNC	P24863
Cyclin-dependent kinase 19	CDK19	Q9BWU1
Cyclin-dependent kinase 8	CDK8	P49336
Mediator of RNA polymerase II transcription subunit 1	MED1	Q15648
Mediator of RNA polymerase II transcription subunit 10	MED10	Q9BTT4
Mediator of RNA polymerase II transcription subunit 11	MED11	Q9P086
Mediator of RNA polymerase II transcription subunit 12	MED12	Q93074
Mediator of RNA polymerase II transcription subunit 13	MED13	Q9UHV7
Mediator of RNA polymerase II transcription subunit 13-like	MED13L	Q71F56
Mediator of RNA polymerase II transcription subunit 14	MED14	O60244
Mediator of RNA polymerase II transcription subunit 15	MED15	Q96RN5
Mediator of RNA polymerase II transcription subunit 16	MED16	Q9Y2X0
Mediator of RNA polymerase II transcription subunit 17	MED17	Q9NVC6
Mediator of RNA polymerase II transcription subunit 18	MED18	Q9BUE0
Mediator of RNA polymerase II transcription subunit 19	MED19	A0JLT2
Mediator of RNA polymerase II transcription subunit 20	MED20	Q9H944
Mediator of RNA polymerase II transcription subunit 21	MED21	Q13503
Mediator of RNA polymerase II transcription subunit 22	MED22	Q15528
Mediator of RNA polymerase II transcription subunit 23	MED23	Q9ULK4
Mediator of RNA polymerase II transcription subunit 24	MED24	O75448
Mediator of RNA polymerase II transcription subunit 25	MED25	Q71SY5
Mediator of RNA polymerase II transcription subunit 26	MED26	O95402
Mediator of RNA polymerase II transcription subunit 27	MED27	Q6P2C8
Mediator of RNA polymerase II transcription subunit 28	MED28	Q9H204
Mediator of RNA polymerase II transcription subunit 29	MED29	Q9NX70
Mediator of RNA polymerase II transcription subunit 30	MED30	Q96HR3
Mediator of RNA polymerase II transcription subunit 31	MED31	Q9Y3C7
Mediator of RNA polymerase II transcription subunit 4	MED4	Q9NPJ6
Mediator of RNA polymerase II transcription subunit 6	MED6	O75586
Mediator of RNA polymerase II transcription subunit 7	MED7	O43513
Mediator of RNA polymerase II transcription subunit 8	MED8	Q96G25
Mediator of RNA polymerase II transcription subunit 9	MED9	Q9NWA0
RNA polymerase II (RNAPII) (Corum complex 2685)		
DNA-directed RNA polymerase II subunit RPB1	POLR2A	P24928
DNA-directed RNA polymerase II subunit RPB11-a	POLR2J	P52435
DNA-directed RNA polymerase II subunit RPB2	POLR2B	P30876
DNA-directed RNA polymerase II subunit RPB3	POLR2C	P19387
DNA-directed RNA polymerase II subunit RPB4	POLR2D	O15514
DNA-directed RNA polymerase II subunit RPB7	POLR2G	P62487
DNA-directed RNA polymerase II subunit RPB9	POLR2I	P36954
DNA-directed RNA polymerases I, II, and III subunit RPABC1	POLR2E	P19388
DNA-directed RNA polymerases I, II, and III subunit RPABC2	POLR2F	P61218
DNA-directed RNA polymerases I, II, and III subunit RPABC3	POLR2H	P52434
DNA-directed RNA polymerases I, II, and III subunit RPABC4	POLR2K	P53803
DNA-directed RNA polymerases I, II, and III subunit RPABC5	POLR2L	P62875
General transcription factor IIF subunit 1	GTF2F1	P35269
General transcription factor IIF subunit 2	GTF2F2	P13984
RNA polymerase II subunit A C-terminal domain phosphatase	CTDP1	Q9Y5B0
RNA polymerase II-associated protein 1	RPAP1	Q9BWH6
Transcription initiation factor IIB	GTF2B	Q00403

2.4 TF protein-protein interactions with chromatin modulating proteins

Chromatin accessibility, controlled by DNA winding around the nucleosomes, is obligatory for the binding of the basal transcription machinery and TFs to allow the transcription initiation. In non-dividing cell, chromatin can be seen in two forms: highly packed heterochromatin and lightly packed euchromatin. Euchromatin shows higher transcriptional activity compared to heterochromatin. Chromatin accessibility is regulated mainly by two mechanisms (Figure 5): first, covalent histone PTMs in specific sites in histone tails affect the DNA-histone binding affinity and enable the recruitment of co-regulatory proteins. Second, the ATP-depended chromatin remodelling complex

replaces or translocases histones to release the DNA for the basal transcription machinery or other TFs (Langst et al., 2015). Besides these two mechanisms, DNA (de)methylation affect the accessibility (Figure 5; (Wang et al., 2007b, 2007a).

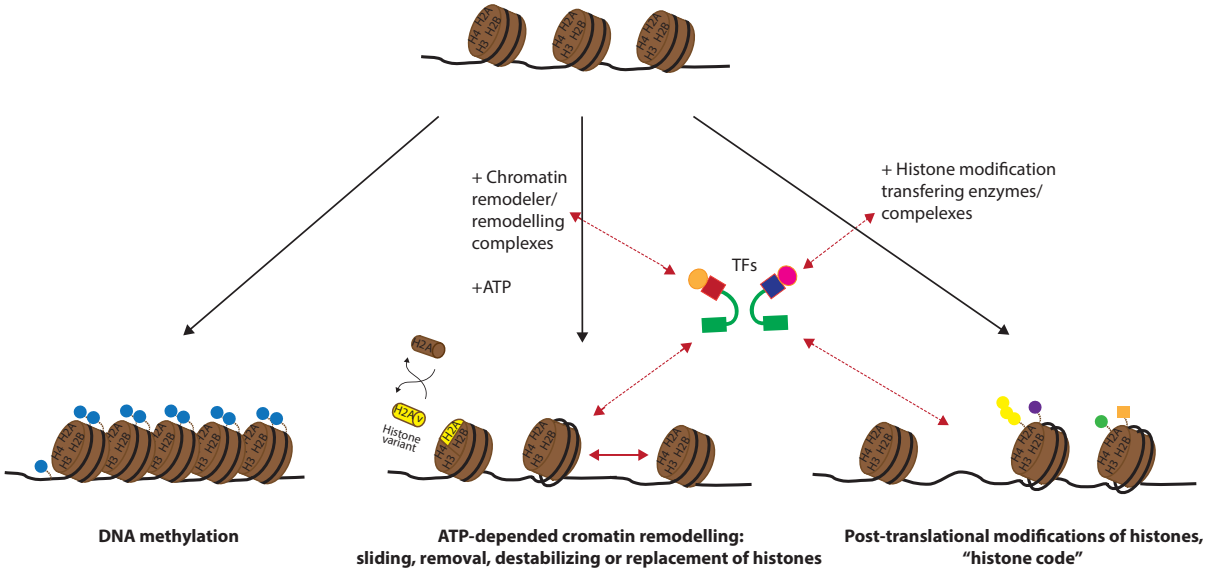


Figure 5. Main chromatin remodelling mechanisms. Chromatin accessibility can be regulated through DNA methylation, ATP-depended chromatin remodelling or histone modifications. TFs are known to interact with many ATP-dependent remodellers as well as with histone modifiers. Chromatin accessibility is required for TF DNA binding. Used schematic domain organization of TFs is described more in details in Figure 1.

Several specific enzymes and protein complexes, such as histone acetyltransferases and the SAGA complex, drive the histone PTMs (Zhang et al., 2016). Acetylation and methylation were the first identified histone modifications (Allfrey et al., 1964), but since then, more modifications have been identified, including phosphorylation, sumoylation, ubiquitylation, ADP-ribosylation and glycosylation (Zhao et al., 2018). These modifications, alone or together, form the histone code that may be recognised, or read, by other proteins, such as TFs. This recognition then determines the activity state of chromatin, the exchange between the transcriptionally active and inactive forms. Modifications may affect the histone-DNA binding affinity, allowing the histone to detach from the DNA.

TFs interact with several enzymes, such as phosphatases, kinases, methylases, ubiquitin ligases and ubiquitin hydroxylases not only for direct modification of themselves but also for activating and inactivating them to modify other co-regulatory proteins, such as histones. As described in section 2.3, multiple SAGA subunits have been identified interacting with different TFs. The SAGA complex is a multisubunit coactivator complex that, besides being part of the basal transcription machinery, has histone acetyltransferase and histone deubiquitynase activity (Helmlinger et al., 2017).

ATP-depended chromatin remodelling complexes contain an ATPase that produces the energy to break the DNA-histone contact, allowing the sliding, removal, destabilising or replacement of the histones (Clapier et al., 2009). Chromatin remodelling proteins, remodellers, fall into four subfamilies: INO80, CHD, SWI/SNF and ISWI (Clapier et al., 2017). Many of the complexes vary in

composition and may share subunits with other chromatin remodelling subcomplexes. This diversity allows them to interact specifically with particular TFs and co-regulatory proteins to respond to different signals with highly specific outcomes.

TFs may regulate the chromatin accessibility by interacting and recruiting both chromatin modifying proteins and ATP-depended chromatin remodelling complexes. TFs have been identified interacting with many histone-modifying proteins and chromatin remodelling complex members, including p300 histone acetyltransferase, histone deacetylases, histone deacetylase complex subunits, lysine-specific demethylases, Paired amphipathic helix proteins, INO80 complex proteins and numerous other proteins (Li, Wang, et al., 2015). TFs require remodelling complexes to make space both for their own DNA binding, for the binding of PIC components and for the binding of other site-specific TFs in a cooperative manner. Once a chromatin is opened, DNA-bound TFs can function as a barrier to prevent the nucleosome from repositioning (Li, Hada, et al., 2015).

2.5 TF protein-protein interactions with RNA splicing machinery

It has been suggested that some chromatin proteins interact with spliceosome components and play a role in RNA splicing (Allemand et al., 2016). These chromatin proteins include TFs that may regulate splicing in several mechanisms: TFs may bind to messenger RNA (mRNA) to recruit other splicing-related co-regulatory proteins; TFs may block the binding of splicing factors from mRNA or the transcription elongation rate of RNA splicing may be affected by the TF binding to promoters, gene bodies or pre-mRNA (Rambout et al., 2018). Increased elongation rates may enhance skipping the weak 3' splice sites, resulting in a different isoform of protein than the slow transcription. TFs can affect the splicing indirectly, as when changing the elongation rates, or directly by binding the RNA and splicing machinery, even independently of their DNA binding. Moreover, as with the expression of all genes, TFs control the splicing factors' transcription.

Some C2H2-ZF TFs, such as CTCF, VEZF1 and MAZ have been reported to bind the gene bodies and stall the Pol-II transitorily, enhancing the inclusion of alternative exon (slow transcription; (Rambout et al., 2018). Additionally, they may bind the RNA and alter the histone acetylation, which, in turn, takes part in RNA splicing. In addition to binding the gene bodies or promoters, TFs may regulate the splicing independently of DNA binding. For example, SOX6 and SOX9 have been reported to affect the splicing without contacting the chromatin but colocalizing with splicing factors (Girardot et al., 2018; Ohe et al., 2002). Other TFs, such as SPI1, may inhibit the splicing by binding specific splicing factors and blocking their interactions with mRNA (Hallier et al., 1996). SP1 binding, however, also functions conversely, as SPI1 binding to the splicing factors may affect the DNA binding of SPI1 (Hallier et al., 1998). Similarly to SPI1, another C2H2-ZF TF, FBI1, inhibits the binding of splicing factor SAM68 to mRNA, reducing the alternative splicing (Bielli et al., 2014).

These examples indicate that RNA splicing and its regulation is not independent from transcription, and TFs, with their DNA, RNA and protein interactions, participate in splicing regulation in multiple ways. However, many facets of the roles of different TFs and TF families remain unknown. In a recent review, Rambout et al. summarised the current knowledge of TFs in mRNA splicing (Rambout et al., 2018).

2.6 TF protein-protein interactions with nuclear acting and myosin signalling proteins

Nuclear actin, nuclear myosin and their interaction partners have been shown to take part in gene expression regulation. Nuclear actin is linked to all three RNA polymerase complexes: it serves as ATPase; is member of several SWI/SNF chromatin remodelling complexes, such as BAF complex, SWR1 and INO80; and is member of histone-modifying complexes, such as TIP60 and p400 (Grummt, 2006). In addition to chromatin remodelling, actin and actin-binding proteins, such as Arps, can regulate the activation of some TFs, often affecting the TF or its cofactor location.

The best-known example of actin-regulated TFs is the serum response factor (SRF), a TF that, in response to actin dynamics, regulates many genes linked to the cytoskeleton. The SRF's co-activator, MRTFA, transfers the signal from the actin cytoskeletal to the nucleus by binding the actin, which controls its nuclear import and activity (Miralles et al., 2003). In addition, YY1's shuttle between the cytoplasm and nucleus is reported to depend on the actin polymerization status in vascular smooth muscle development (Favot et al., 2005). Polymerized gamma-actins hold the YY1 in cytoplasm, while depolymerization releases it to enter the nucleus. Similarly, the subcellular localisation of PREP is controlled through cytoskeletal proteins binding to actin and tubulin, and the release from either of them enables the nuclear translocation (Haller et al., 2004).

Recent proteomics study of nuclear actin PPIs have further expanded the knowledge of actin's connection to transcription and mRNA splicing (Haller et al., 2004). In total, 338 interaction partners have been identified for nuclear actin, of which, according to Lambert et al.'s curation (Lambert et al., 2018), 22 were TFs. These include GTF2I, YBOX1, YBOX3 and multiple zinc finger (ZF) TFs. The affinity purification (AP) of FOXR1, FOXO6, FOXN1, MYC and TCF4 identified interactions with the actin-like protein 6A (Li, Wang, et al., 2015).

Together with nuclear actin, nuclear myosin I (NM1) takes part in gene expression regulation. Both actin and myosin are needed for the RNA polymerase I -mediated transcription of ribosomal RNA (Philimonenko et al., 2004). NM1 has other functions in the nucleus; for instance, it is crucial for the formation of the first phosphodiester bound in Pol-II-mediated transcription initiation. Besides NM1, other myosins (2, 5, 6, 10 and 16) have been described as located in nucleus (de Lanerolle, 2012), but their roles are not well understood. TF APs have identified multiple interactions between TFs and myosins, such as interactions with Myosin-10 and Myosin-7 (Li, Wang, et al., 2015).

However, the mechanisms relating to how TFs interact with actin- and myosin-signalling proteins remain poorly understood. There is evidence that TFs may interact directly with actin and myosin proteins, but the interactions are more often taught as occurring in the context of chromatin remodelling.

2.7 TF protein-protein interactions in DNA repair

In response to DNA damage, cells activate a large network of pathways, collectively termed as DNA damage response (DDR) (Ciccia et al., 2010). DDR pathway activation leads typically to cell cycle arrest, inhibition of transcription and translation, and, in case of considerable DNA damages, to apoptosis (Raschella et al., 2017). In addition, DNA repair machinery is activated.

DNA damage may only occur in one strand for which cells have three repair mechanisms: base excision repair, nucleotide excision repair and mismatch repair. However, most often, pathological conditions caused by DNA damage originate from double-strand DNA breaks (DSBs). DSBs are repaired by two major mechanisms: non-homologous end joining and homology-directed repair. These repair mechanisms require complex cross-talk between multiple DNA repairing proteins. The lesion repair requires the recognition by specific factors as well as changes in the chromatin structure (Lukas et al., 2011). Similar to chromatin modulation in transcriptional regulation, the chromatin structure is altered by histone and histone-modifying enzyme modifications, nucleosome translocations and replacements and the recruitment of an extensive number of proteins in the lesion area.

Even transcription is generally inhibited after DNA damage; numerous TFs take part in DNA repair. TFs facilitate DNA repair by increasing the transcription of DNA repair proteins (Christmann et al., 2013), but they also alter the repair in non-transcriptional mechanisms, such as altering the chromatin remodelling or acting directly in DNA repair complexes (Malewicz et al., 2014). Importantly, FOXO3a interacts and promotes the phosphorylation of AMT, which orchestrates the assembly and activation of downstream DNA repair mechanisms (Tsai et al., 2008). Silencing FOXO3a results in a lack of DNA repair. Similarly, ATF2 (Bhoumik et al., 2005), E2F1 (Biswas et al., 2012), NR4A (Malewicz et al., 2011) and SP1 (Beishline et al., 2012) have transcription independent roles in DNA repair. They all translocase in DNA repair loci in response to DNA damage and are most often phosphorylated in specific sites in response to DNA damage. A study that identified new DDR proteins using epitope-tagging coupled to DNA damage localisation screen revealed that more than 70% (35) randomly selected TFs (with no known role in DNA damage) localised in the DNA damage area (Izhar et al., 2015). Thirteen of these 35 TFs belong to ZF TFs.

TFs seem to have important role in DDR that includes both transcriptional and non-transcriptional actions. The non-transcriptional action in the DNA lesion site requires obvious TF protein interactions with DDR machinery. Li et al. supported this finding through their AP study (Li, Wang, et al., 2015), which observed several TFs interacting with DNA repair proteins, such as XRCC1, RAD50, MOV10 and MMS19. Furthermore, Ets1 has been reported to interact with multiple DNA repair proteins (PARP-1 and DNA-PK) with possible outcomes in cancer development (Legrand et al., 2013; Choul-Li et al., 2018).

2.8 TF protein-protein interactions in DNA replication

DNA replication in the S-phase of the cell cycle is performed by a multiprotein protein machine, replisome. Replisome includes, most importantly, helicases, primases, DNA polymerases, nucleases and ligases. In addition to replisome, many other accessory proteins are needed to perform the replication completely and accurately (Cortez, 2017).

The origin recognition complex (ORC) is an essential protein complex in initiating replication (Ding et al., 2011). To form the pre-replicative complex (pre-RC), the ORC recognises the replication origin and recruits the Cdt1, Cdc6, and MCM complex (Belsky et al., 2015).

TFs are known to control the expression of genes needed for the cell cycle transition different phages. For instance, E2F TFs are known to both activate and repress the transcription of

genes needed for G1/S transition (Bertoli et al., 2013). In early G1 phase, repressor E2F TFs are bound to promoters to prevent the transcription, whereas activator E2F TFs are bound to inhibitory proteins that are known as pocket proteins. Later in G1 phase, pocket proteins are phosphorylated by CDKs and activator E2F released to replace the repressor E2Fs and activate the gene expression of proteins needed for G1/S transition.

Besides transcription repression or activation, some TFs have been reported to interact with ORC and pre-RC components and regulate the replication in non-transcriptional mechanisms. For example, MYC interacts with multiple MCM complex members, such as MCM2-7, ORC2, Cdc6 and Cdt1 (Dominguez-Sola et al., 2007). MYC was reported to control the replication initiation in non-transcriptional mechanisms in a late G1 state (Dominguez-Sola et al., 2007), and the inactivation of MYC led to the inhibition of the G1/S transition (Prathapam et al., 2006). Similarly, in Drosophila, E2F1 (Bosco et al., 2001) and MYB (Beall et al., 2002) are reported to interact with ORC and pre-RC components.

3. TFs in development and diseases

3.1 TFs in development

TFs play important role in cell differentiation, tissue generation and embryonic development. Different cell types, maintained with specific gene expression patterns, are differentiated from a multipotent stem cell as a result of the synergistic cooperativity of a particular set of tissue-specific TFs. In early differentiation, specific TFs with the ability to respond to extracellular signals act as pioneer factors to initiate the chromatin opening and assembly of the PIC (Spitz et al., 2012).

Pioneer TFs, including e.g. OCT3/4, SOX2, FOXA, GATA4, KLF4, PAX7, ASCL1, p53 and Nanog (Iwafuchi-Doi, 2019), are able to target closed chromatin in the early stages of development. Therefore, they control the whole gene expression program of cell differentiation. Like many other TFs, pioneer TFs enhance the DNA-binding specificity by binding in oligomers and in a cooperative manner (see Section 2.2). A well-known example of this is the OCT3/4-SOX2 heterodimer that recognises the specific target sites of pluripotent stem cells (Rizzino et al., 2016). The binding of pioneer factors leads to the recruitment of chromatin remodellers, other TFs, PIC components, Mediator complex and, finally, to the transcription of cell type specific TFs. These changes in transcription allow cells to differentiate to certain morphology.

Particular TFs can change the phenotype when overexpressed ectopically. This ability is used in cell reprogramming. For example, an overexpressed MYOD1 can instruct change in the phenotype (Davis et al., 1987). However, often a specific combination of multiple TFs is needed to instruct cell fate (Niwa, 2018); for example, reprogramming somatic cells into pluripotent stem cells requires the introduction of four specific proteins: SOX2, MYC, OCT3/4 and KLF4 (Takahashi et al., 2006).

Like the transcription in any highly differentiated cells, transcription during the developmental processes requires various PPIs between TFs and TF-binding proteins. However, a disturbance in the activity of the TFs involved in differentiation is often highly damaging and can cause serious outcomes in human health. For example, a mutation in pioneer TF or its DNA-binding sequence may result in numerous cancer types (Iwafuchi-Doi, 2019; Jozwik et al., 2012).

3.2 TFs in diseases and disorders

Because of TFs' important roles in differentiating and maintaining multiple cellular processes, such as transcription, DNA replication and DNA repair, TF mutations might lead to serious pathological conditions. Lambert et al. (Lambert et al., 2018) estimated that 19.1% of human TFs are associated with at least one phenotype, whereas lower number, 16.2%, of all human genes are disease-associated.

Numerous TFs are involved in large number of diseases, including cancer, diabetes, developmental syndromes, obesity, autoimmunity, neurological disorders and immunodeficiencies (Lee et al., 2013). Genetic variation and mutations in a TF's coding genes may affect its ability to interact with DNA, RNA, ligands or other proteins or mutations may be truncating and lead to the degradation of the TF (Kaustio et al., 2017). These disturbances may lead for instance to altered

chromatin remodelling, cooperativity, cell cycle and DNA repair. Many of these mechanisms result from a disturbed ability of the TFs to bind to DNA or other proteins.

Although few mutations are reported to stabilise PPIs, most of the mutations are destabilising, mainly because of alterations in the binding energy due the changes in the electrostatic forces (Yue et al., 2005; Nishi et al., 2013). Mutations in the TFs' interfaces often results in more severe changes in interactome than mutations in amino acids packed inside the protein.

The pharmacological approach to target TF activity is usually to mimic the endogenous ligands or target upstream phosphokinases (Fontaine et al., 2015). However, the specificity is not always high, and there are limitations to identifying novel molecular targets. Therefore, targeting TF PPIs is an excellent addition to the pharmaceutical treatment of TF-related diseases. Targeting TFs that function as downstream players of signalling cascades serves more specificity compared to targeting upstream effectors, such as membrane receptors or kinases (Vaquerizas et al., 2009). TF protein interactions may be modulated, for example, with small compound PPI distributors. Usually they are small, rigid, hydrophobic and may have a peptide motif as part of the design (Wojcik et al., 2016; Fontaine et al., 2015).

3.2.1 TFs in cancer

A dysregulation of genes is one of the hallmarks of cancer. Mutations in TF coding genes have long been known to participate in tumorigenesis. Kandoth et al. analysed the mutation patterns of 3281 tumours across 12 major cancer types (Kandoth et al., 2013) and identified 127 significantly mutated genes (SMGs), of which, according the TF classification of Lambert et al. (Lambert et al., 2018), 26 were TFs. These included TP53, FOXA1, FOXA2, GATA3, NF2F2, NF2L3, SMAD4, SOX17, SOX9, TBX3 and WT1, but, interestingly, not MYC. TP53 had the highest frequency (42%) of all 127 SMGs. Vogeltein et al. obtained similar results to those of Kandoth et al. in a study that identified 27 of 138 cancer driver genes to be TFs (Vogelstein et al., 2013). The study by Vogeltein et al. identified similar genes to those identified by Kandoth et al., such as TP53, AR, WT1, GATA3, SOX9 and SMAD4, as well as other TFs, including MYC, MYCL1, MYCN, GATA1, GATA2, HNF4a, KLF4, NCOA3 and PAX5. Together these two studies identified 42 different TFs genes that are cancer driver genes and/or highly mutated in tumour progressions in various cancer types.

As an example, mutations in Mediator complex subunits are reported to impact various tumours having impact in protein-protein interactions, including uterine leiomyosarcomas (Makinen, Mehine, et al., 2011; Makinen, Heinonen, et al., 2011) and prostate cancer (Barbieri et al., 2012). MED12 exon2 was mutated in 70% of Finnish and 50% of South African uterine leiomyomas patients (Makinen, Mehine, et al., 2011; Makinen, Heinonen, et al., 2011). An investigation of the mechanism behind the disease performed using affinity purification coupled to mass spectrometry (AP-MS) revealed the disruptions in protein interactions: mutated MED12 had a specific decrease in interactions with cyclin C-CDK8/19, which led to lack of Mediator-associated cdk activity (Turunen et al., 2014). Moreover, in acute lymphoblastic leukemia, MED12 had a nonsense mutation in the 5' end (c.97G>T, p.E33X(Heikkinen et al., 2017). However, the C-terminal of MED12 was still expressed using an alternative translation initiation site. The lack of N-terminal NLSs resulted in the loss of interaction with importin-α and the nuclear import. This led to a

complete loss of nuclear protein-interactions, such as interactions with the rest of the Mediator complex or basal transcription machinery.

Moreover, TFs are often part of oncofusion proteins that are resulted from chromosomal translocations. Rearrangement may, for example, result in a fusion protein which have the DBD of TF fused to protein with interactions to transcriptional repressors ((Martens et al., 2010). Abnormal introduction of transcriptional repressors to transcriptional site might lead to several alterations in target genes' expression. As an example of TF-oncofusions, ETS are known to be part of various oncofusion proteins both in haematological cancers and solid tumours (Sizemore et al., 2017), and multiple TF-fusions, such as AML1 (RUNX1), RARa, HOXA9, EVI1, MYH11 and ERG fusions, are found in acute myeloid leukemia (Martens et al., 2010). However, TF-mediated transcription regulation may also be dysregulated by translocations with mechanism other than the TFs itself taking part in oncofusion proteins; the translocations may result in abnormally expressed coactivators that have an impact on TFs' function (Martens et al., 2010).

3.2.2 TFs in neurological diseases

Mutations in several TFs, especially in Mediator complex subunits, can induce neurological disorders (Lee et al., 2013). This is not a surprise as TFs play a key role in tissue development, including the development of the brain and nervous system. MED23 mutations disrupted interactions with the enhancer-bound TFs, leading to a dysregulation of transcription in early brain development (Hashimoto et al., 2011). Similarly, MED12 mutations have been linked to intellectual disabilities (Srivastava et al., 2019; Ding et al., 2008). In addition, TAF1 variants are associated with neurological manifestations, developmental delays and the development of X-linked ID syndrome (O'Rawe et al., 2015). Various TF mutations and polymorphisms are linked to Parkinson's disease, including variations in NR4A2 (Grimes et al., 2006; Le et al., 2003), PITX3 and EN1 (Haubenberger et al., 2011).

One of the mechanisms through which TF mutations impact neurological disorders is altered PPIs. For example, missense SIX3 mutations are linked to holoprosencephaly, which is a brain development disorder that causes the embryonic forebrain to fail to divide into right and left hemispheres (Laflamme et al., 2004). Wild type (WT) SIX3 has ability to bind and stimulate the transcriptional activity of another TF, NR NOR1 (NR4A3), but the patient mutation R257P disrupts the interaction and therefore changes the transcriptional activity of NR4A3.

3.2.3 TFs in diabetes

The gene expression patterns of pancreatic tissue in maturity-onset diabetes of the young (MODY) are mainly controlled by small set of master TFs that include HNFs (1α, 1β and 4α), PDX1 and NDF1 (NeuroD) (Malecki, 2005; Maestro et al., 2007). Therefore, any mutation in these master TFs can induce the development of MODY. However, in other forms of diabetes, various TFs play an important role. For example, mutations and polymorphisms in PAX4 are associated with Type 2 diabetes mellitus and MODY (Lorenzo et al., 2017), and FOXOs are linked to diabetes-related muscle atrophy (O'Neill et al., 2019) and diabetes-induced oxidative stress (Ponugoti et al., 2012).

In addition, STAT3 gain-of-function (GOF) mutations have been linked to Type 1 diabetes (Flanagan et al., 2014). Closer investigations of molecular mechanisms revealed an increased affinity

to transcriptional regulators and nuclear pore complex members, indicating an increased nuclear localisation of mutated STAT3 (Saarimaki-Vire et al., 2017).

3.2.4 TFs in cardiac diseases

TFs, together with chromatin remodellers, have been strongly linked to cardiac diseases (reviewed in (Bauer et al., 2017). A good example is the cardiogenic TF GATA4, in which the G296S mutation causes cardiac septal defects and cardiomyopathy (Ang et al., 2016). The mutation disrupts the recruitment of another cooperative TF, TBX5, resulting in a dysregulated transcription of the genes required for the normal phenotype, which, in turn, causes impaired contractility, calcium handling and metabolic activity. Mutations in NKX2-5 (Schott et al., 1998), GATA4, Tbx5, ETS1, GATA6 and ETS1 have been indicated to play a role in congenital heart disease development (McCulley et al., 2012). Similarly, cardiac hypertrophy is altered by multiple TFs, such as GATAs, MEF-2, SRF, TEAD, NFAT, FOXO, MIFT and YY1 (Kohli et al., 2011).

3.3 Primary immunodeficiencies caused by TF mutations

In addition to the disorders described in the previous subsections, several TF variations have been linked to the development of autoimmunity and inflammation-related diseases. A GWAS analysis by Lambert et. al. revealed an enrichment of GWAS signals for polygenic immunological diseases in proximal to TF coding genes, highlighting the importance of TFs in immunological processes (Lambert et al., 2018). Numerous TFs have been identified as contributing to immune responses and to the development of immune cells including NFKBs, STATs, IRFs, Bcl6, GATA3, PAX3, E2A, Ikaros, PU.1, CEBPs and EBF (Smale, 2014; Mercer et al., 2011). Logically, genetic variation and mutations in genes or regulatory areas of these TFs may lead to serious defects in immune cell development or immune responses. For example, dysregulations in NFKB signalling have been linked to cancer, inflammations, autoinflammation diseases and disrupted immune development. Mutations in NFKB1 are also the most general monogenic reason for primary immunodeficiencies (PIDs) in Europeans (Tuijnenburg et al., 2018). This example, along with others, such as the fact that mutations in AIRE result in Type I autoimmune polyendocrinopathy syndrome (Perniola, 2018), highlight the importance of TFs in immunological disorders. However, PIDs serve as an excellent example of monogenic TF mutations in disease development. These disorders will now be discussed in greater detail.

PIDs are rare and chronic diseases caused by inherent defects in immune system development and/or functions leading to increased susceptibility to infections (Ballow et al., 2009). Paradoxically, a significant subset of patients also develops autoimmunological symptoms (Todoric et al., 2013). Phenotypes of patients may include allergies and malignancies (Bousfiha et al., 2018). Because of the heterogenicity of diseases, no universal clinical care is available for the patients. However, treating PIDs often includes immunoglobulin replacements (Ballow et al., 2009) and hematopoietic stem cell transplantations (Thrasher et al., 2017).

PIDs are a diverse group of more than 350 diseases (Picard et al., 2018), most of which have a monogenic background (Bousfiha et al., 2018). An OMIM search (https://www.ncbi.nlm.nih.gov/omim) of the term ‘primary immunodeficiency’ resulted in 336

diseases that could be matched to 295 proteins (Uniprot). Thirty-eight of these proteins were TFs, including AIRE, CEBPs (a and b), BCs, ETV6, FOXO, FOXP3, GATA2, IRF8, MYC, KMTA2, STATs, TALs, RUNX1, NFKBs, RELs, NFACs, TP53 and PAX5. Due to significant improvements in next generation sequencing techniques as diagnostic tools, the number of identified PIDs causing mutations has been dramatically expanded during the last decade. The following subsections discuss some of the well-studied TFs in PIDs.

3.3.1 NFKBs

NFKBs are a family of proteins that are found in most human tissues where they control numerous cellular processes, such as transcription, cytokine production and cell survival (Sun, 2017). The mammalian NFKB family consists of five members: p105/p50 (NFKB1), p100/p52 (NFKB2), c-rel, RelB and p65 (RelA), all of which share the rel-homology domain. These family members associate with each other to form homo- and heterodimers, one binding partner often being either the full length NFKB1 or NFKB2 (p105 or p100) or processed forms (p50 or p52, respectively). However, as only c-rel, RelB and p65 contain a C-terminal activation domain, not all of the combinations, such as p50 or p52 homodimers, are transcriptionally active. Processing NFKB1 and NFKB2 into p50 and p52 subunits is necessary for nuclear entrance and transcriptional regulation.

Under the cell's resting state, most of the NFKBs are full length, cytosolic, inactive and bound to IKBs (Sun, 2017). An activation signal leads NFKBs to separate from the IKBs. This may happen through canonical activation pathway, where pathogen-associated molecular patterns (PAMPs) or cell surface receptors of pro-inflammatory cytokines bind their extracellular ligands and transfer the activation signal for intracellular IKK complex members. Activated IKKs catalyse the phosphorylation and polyubiquitylation of IKBs, leading to their complete degradation and dissociation from the NKFBs. Next, the NFKBs (mostly NFKB1 in the canonical pathway) are processed into shorter forms, allowing their dimerization (mostly p50-RelA dimers) and entrance to the nucleus.

The NFKB pathway can also be activated through a non-canonical process. Extracellular signals are transferred via LT-β or BAFF receptors, activating NIK, which, in turn, activates IKKa homodimers. This allows the IKKa to directly phosphorylate NFKB2 (p100), leading to NFKB2 processing to p52, dimerization (mostly with RelB) and nuclear entrance.

Recently, numerous NKFB1 mutations have been shown to cause PIDs. NFKB1 mutations in PIDs have been identified in patients with common variable immunodeficiency (CVID), which is the one of the largest subgroups of PIDs. In CVID, heterozygous mutations causing exon-skipping (p105ΔEx8, p50ΔEx9) or truncation (c.465dupA) lead to a CVID phenotype due the rapid degradation of affected protein leading to haploinsufficiency of the NFKB1 subunit p50 (Fliegauf et al., 2015). Moreover, c.491delG; p.G165A*31 frameshift mutation is reported to lead to haploinsufficiency and decreased p50 levels due to the reduced phosphorylation of p105 upon stimulation (Boztug et al., 2016). A recent a European cohort study indicated that loss-of-function (LOF) mutations are the most common monogenic reason for CVID (Tuijnenburg et al., 2018). In a cohort of 846 PID patients, Tuijnenjurg et al. identified 17 pathogenic variants in the NFKB1 gene, including heterozygous truncating, missense and gene deletion variants. These variants cause variable disease manifestations through decreased p50 levels.

Besides NFKB1, mutations in NFKB2 are linked to PID development. A recently published overview summarised 19 different NFKB2 mutations in 50 PID patients (Klemann et al., 2019). mutations include four missense mutations and 15 mutations that caused the truncation, either by a nonsense or frameshift mutation. Functional studies of mutants revealed various mechanisms. Most of the mutations (K855Sfs*7, A867Cfs*19, D865Vfs*17 and D865G) were in the C-terminal end of the protein and resulted in a decreased C-terminal phosphorylation of the p100 subunit, which inhibited proteasomal processing into the active subunit p52 (Lindsley et al., 2014; Lee et al., 2014; Chen et al., 2013). However, few mutations in the p52 coding sequence, E418X and R635X, were identified to be GOF mutations leading to constitute nuclear localisation and the activation of non-canonical and canonical NFKB pathways (Kuehn et al., 2017). Mutations also caused increased interaction with the dimerization partner RelB, leading to an enhanced nuclear localisation of RelB. As previously mentioned, interaction with RelB is obligatory for the transcriptional activation activity of p52 as RelB contains the C-terminal activation domain that is absent from p52. Homodimer p52-p52 functions as a transcriptional repressor.

3.3.2 STATs

STATs are TFs involved in the downstream signalling of Type I and II cytokine receptors. STAT1 and STAT2 are activated by the IFN-IFNR-JAK pathway, where extracellular IFNs bind to their receptors, allowing JAKs to activate, which in turn leads to the phosphorylation of STAT1 and STAT2. Activated STATs may form STAT1-STAT2 heterodimers and enter the nucleus. STATs are involved in multiple cellular processes, including apoptosis, proliferation, differentiation and immunity. STATs share structural similarities, including the SH2 domain involved in dimerization, a coiled coin domain needed for localisation to the nucleus, the DBD and the TAD. Mutations in different parts of the STATs may affect their localisation, expression levels and interactions with other proteins. STAT mutations are associated with several disorders, including STAT1, STAT2, STAT3 and STAT5B mutant PIDs.

The first identified mutation in STAT1, L706S, led to a susceptibility to mycobacterial infections (Dupuis et al., 2001). The total STAT1 deficiency due to a homozygous two amino acid deletion (1757.1757delAG) or homozygous missense mutation (L600P) in STAT1 resulted in an impaired response to the IFNs (IFNα and IFNβ) and early mortality due to viral infections (Dupuis et al., 2003). The partial STAT1 deficiency (heterozygous P696S) did not completely abolish the response to IFN signalling, but it impaired the response to several IFNs (Chapgier et al., 2009).

The first GOF mutations in STAT1 were described in chronic mucocutaneous candidiasis patients in 2011 (Liu et al., 2011; van de Veerdonk et al., 2011). The mutations caused increased responses to IFN signalling, resulting in, for example, autoinflammatory symptoms. Since then, numerous STAT1 LOF and GOF mutations in PIDs have been described (Mogensen, 2018; Martinez-Martinez et al., 2015; Lorenzini et al., 2017; Toubiana et al., 2016).

STAT2 mutations are a rare cause of PIDs, and only a few cases have been reported. STAT2 deficiency, caused by homozygous mutation in intron 4 and altered mRNA splicing, led to an abnormal failure in Type 1 IFN signalling and severe, even lethal, viral infections (Hambleton et al., 2013). An infectious phenotype with disseminated measles infection was also observed after an

MMR vaccination. Similarly, a G1576A mutation in the STAT2 gene caused altered splicing and led to a STAT2 deficiency, impaired IFN response and severe viral illness (Moens et al., 2017). Moreover, a STAT2 deficiency (STAT2: C612Y) in children with severe a neurological deterioration following a viral infection has been linked to a mitochondrial morphology disorder, indicating a link between the immunity and mitochondrial dynamics (Shahni et al., 2015).

The first STAT3 mutations causing PIDs were reported in 2007 when two groups reported 18 and five different STAT3 mutations that cause hyper-IgE syndrome (HIES) (Minegishi et al., 2007; Holland et al., 2007). The mutations were mainly located in the SH2 and DBDs of STAT3 and resulted in complex phenotypes, characterised by symptoms such as extremely high levels of IgE, increased innate immune responses and impaired IL6 signalling. Mutations in the DBDs did not change the expression, phosphorylation or location of STAT3, but mutations in the SH2 domain reduced STAT3 phosphorylation at tyrosine 705 (Renner et al., 2008).

In contrast to these STAT3 LOF mutations, several GOF mutations have been identified. The first GOF mutation was identified in patients suffering from early-onset multi-organ autoimmune diseases that includes Type 1 diabetes (Flanagan et al., 2014). Four different heterozygous de novo STAT3 mutations in the SH2 domain (K392R, N646K, T716M and K658N) caused increased STAT3 activity. Further studies of the K392R mutation in the development of diabetes have shown that, besides increased transcriptional activity, the mutant STAT3 increased nuclear localisation and upregulated affinity to nuclear pore complex members and proteins involved in transcriptional regulation (Saarimaki-Vire et al., 2017). STAT3 GOF mutations have also been reported to cause early onset lymphoproliferation and autoimmunity mechanisms, including defects in STAT5 and STAT1 phosphorylation (Milner et al., 2015).

STAT5B has an important role as a downstream component of the IL2 and growth hormone receptors. STAT5B deficiency is reported to cause PID, autoimmunity, IGF-1 deficiency and an insensitivity to growth hormones (Bernasconi et al., 2006; Acres et al., 2019; Hwa et al., 2005; Nadeau et al., 2011; Kofoed et al., 2003). These defects lead to a clinical phenotype of abnormal growth, severe infections, pulmonary disease and facial dysmorphism. Since the first reported patient in 2003, autosomal dominant and autosomal recessive forms of STAT5B deficiency have been reported in thirteen and ten patients, respectively (Acres et al., 2019). These mutations included missense mutations (e.g., A630P and L151P), frameshift mutations (e.g., N398EfsX16) and truncating mutations (e.g., R152X) (Bernasconi et al., 2006; Acres et al., 2019; Hwa et al., 2005; Nadeau et al., 2011; Kofoed et al., 2003). Recently, three heterozygous mutations were reported as not causing a complete loss of STAT5B, but rather acting dominant-negative through dimerization with WT STAT5B (Klammt et al., 2018). This result errors also in WT's nuclear localisation or DNA binding, leading to the total transcriptional inactivation of STAT5B.

3.3.3 CEBPE

C/EBPε, encoded by *CEBPE*, is a TF involved in cellular differentiation, especially the differentiation of myeloid lineages (Bedi et al., 2009; Lekstrom-Himes et al., 1998). Through RNA splicing and alternative translational start sites, four isoforms of C/EBPε exist, of which the longest (32 and 30

kDa) function as transcriptional activators. The 27 kDA isoform represses the GATA-1 transactivation in eosinophils promoters, and the 14 kDa isoform functions as a dominant-negative regulator.

The only known CEBPE disorder is a rare PID, specific granule deficiency (SGD). Patients with SGD have neutrophils that are differentiated but have atypical nuclear morphology, impaired neutrophil chemotaxis and abnormal granule number and/or morphology. These symptoms led to a clinical phenotype with recurrent infections and abscesses in the skin and airways.

Both homozygous and heterozygous CEBPE mutations are reported to cause SGD. To date, seven different SGDs causing CEBPE mutations have been identified (Khanna-Gupta et al., 2007; Stray-Pedersen et al., 2017; Shigemura et al., 2014; Wada et al., 2015; Lekstrom-Himes et al., 1999). As an example of the mechanism behind SGD, two amino acid deletions in the leucine zipper domain led to decreased transcriptional activity and weakened the association with two other TFs, GATA-1 and PU.1 (Wada et al., 2015). Moreover, five base pair deletions in the second exon led to the truncation and loss of DNA binding and dimerization domains (Lekstrom-Himes et al., 1999). This homozygous mutation led to a total loss of C/EBPε.

4. Methods for solving TF protein-protein interactions

PPIs of TFs and other proteins play a central role in almost every, if not all, cellular processes. Therefore, identifying and characterising PPIs is essential to understand the cellular processes and their disturbances. PPIs can be divided into stable and transient interactions, and there are various in vitro and in vivo methods to analyse both (Table 2). These include methods that require a library of potential interacting proteins (preys) fused to certain probes and discovery methods that can be used in mixtures of multiple intact proteins, such as endogenous proteins in cells or in cell lysates.

A common, longstanding method is yeast two hybrid (Y2H) (Fields et al., 1989) and its variations in the yeast and mammalian systems. In Y2H, the protein of interest (bait) is fused with the DBD of a split TF, whereas the prey is fused to an activation domain of the same TF. The interaction between the bait and prey brings the domains together, allowing for the transcriptional activity of the TF and expression of the reporter gene. Y2H requires a library, where the potential prey proteins are fused to the DBDs or activation domains of the TF. The Y2H method has limitations; for example, both interacting proteins must be in the nucleus, it can only identify stable PPIs and it has difficulties modelling PPIs from other organisms due the poor expression of interacting proteins or necessary cofactors in the yeast. Apart from these limitations, this method is affordable, simple, easy to set up in different laboratories and functions within in vivo environments.

There are several derivatives of Y2H, such as membrane yeast two hybrid (MYTH) (Stagljar et al., 1998), mammalian protein-protein interaction trap (MAPPIT) (Ulrichts et al., 2009), mammalian membrane two hybrid (MaMTH) (Petschnigg et al., 2014) and kinase substrate sensor (KISS) (Lievens et al., 2014). These methods overcome some of the limitations of Y2H, such as protein localisation and the lack of mammalian protein expression. MAPPIT, the Y2H derivative in mammalian cells, is based on the splitting of STAT3 (Ulrichts et al., 2009). Unlike in Y2H, the interaction might also occur in cytosol as the STAT3 is only located in the nucleus when activated. In KISS, the bait is fused to fused to a TYK2 kinase and the potential prey to TYK2's substrate motif containing gp130. The interaction leads to the phosphorylation of gp130 and finally to the activation of STAT3 and the reporter gene. The membrane protein assays (MYTH and MaMTH) are based on split ubiquitin, of which one part is fused to the membrane protein bait and the other to the potential prey (Petschnigg et al., 2014; Stagljar et al., 1998). The reunification of the ubiquitin via bait-prey interaction enables its recognition by cellular deubiquitinating enzymes that cleave and release the ubiquitin-bound artificial TF, which is then free to enter the nucleus and activate the transcription of the reporter gene.

Table 2. Summary of methods used for PPI detection. These include methods that require a library of preys fused to certain probes (library) and discovery methods that can be used in mixtures of multiple intact proteins, such as endogenous proteins in cells or in cell lysates (discovery).

Method	Short summary	Identified PPIs	Discovery/ Library	Approach	Reference
Affinity purification					
AP-MS: Affinity tag based	Bait protein is tagged with affinity tag (e.g. FLAG, HA, Strep), protein complexes are affinity purified and interacting proteins identified by MS.	Stable	Discovery	*In vitro*	(Chang, 2006; Dunham et al., 2012)
Antibody based	Protein of interest and it's interacting proteins are captured from the cell lysate by antibody and identified by MS.	Stable	Discovery	*In Vitro*	(Dunham et al., 2012)
Proximity labelling					
BioID: Biotinylation based	Bait protein is tagged with biotin ligase that biotinylates the proximal proteins. Biotinylated proteins are affinity purified and identified by MS.	Stable, transient	Discovery	*In Vitro*	(Kim et al., 2016; Roux et al., 2012)
APEX	Bait proteins are tagged with APEX, that generates biotin-phenol radicals that are fused to interacting proteins.				(Rhee et al., 2013)
MAC-MS: Combination of AP-MS and BioID-MS	Bait protein is tagged with affinity tag and biotin ligase allowing the purification of both transient and stable interactions.	Stable, transient	Discovery	*In Vitro*	(Liu et al., 2018)
Reporter gene based two hybrid approaches					
Y2H: Yeast two Hybrid	Tool for screening the bait protein's PPIs against potential preys in yeast system. Bait-Prey interaction results reporter gene activity.	Stable	Library	*In Vivo*	(Fields et al., 1989)
MYTH: Membrane yeast two hybrid	Design for detection of membrane protein's PPIs in yeast system. Bait-Prey interaction results reporter gene activity via reunification of split ubiquitin and release of artificial TF.	Stable	Library	*In Vivo*	(Stagljar et al., 1998)
MAPPIT: Mammalian protein–protein interaction trap	Design for detection of mammalian PPIs. Bait and prey are bound to STAT3 fragments and the interaction allows the reporter gene activity.	Stable	Library	*In Vitro*	(Ulrichts et al., 2009)
MaMTH: mammalian membrane two-hybrid	Design for detection of mammalian membrane protein's PPIs. Bait-Prey interaction results reporter gene activity via reunification of split ubiquitin and release of artificial TF.	Stable	Library	*In Vitro*	(Petschnigg et al., 2014)
KISS: kinase substrate sensor	Design for detection of intracellular PPIs. Bait is bound to kinase TYK2 and prey to its substrate motif containing gp130. Interaction leads to activation of STAT3 depended reporter gene.	Stable	Library	*In Vitro*	(Lievens et al., 2014)
Protein-fragment complementation (PCA)					
BiFC: bimolecular fluorescence complementation	Bait and prey are fused to parts of divided fluorescent protein and the interaction provides a fluorescent signal.	Stable	Library	*In Vitro*	(Kerppola, 2008)
NanoLuc: NanoLuc complementation assay	Bait and prey are fused to two fragments which interaction brings together. This results functional enzyme and luminescent signal.	Stable	Library	*In Vitro*	(Dixon et al., 2016)
DHFR-PCA: dihydrofolate reductase protein-fragment complementation assay	Survival-selection assay, where yeast cell cycle is arrested by inhibiting DHFR by methotrexate. Bait and prey are tagged with parts of divided methotrexate insensitive DHFR and the complementation restores the cell cycle allowing proliferation.	Stable	Library	*In Vivo*	(Tarassov et al., 2008)
Cross-linking protein interaction analysis					
Label transfer protein interaction analysis	Bait proteins are crosslinked with label transfer reagent and mixed with possible preys. UV exposure allows the label transfer to preys that can be identified e.g. by MS or WB.	Stable, transient	Discovery	*In Vitro*	(Liu et al., 2007)
Cross linking	Interacting proteins are crosslinke(Neiswinger et al., 2016)d using chemical or photoreactive crosslinkers, digested and identified by MS.	Stable, transient	Discovery	*In Vivo, In Vitro*	(Vasilescu et al., 2004)

Co-immunoprecipitation	Interacting proteins are captured using antibody against protein A or it's fusion tag. Protein B is detected by western blotting with specific antibody.	Stable	Library	*In Vitro*	Reviewed in (Lee, 2007)
Protein Microarrays	Full length proteins, protein domains, antibodies or cell lysates are arrayed into slide and the slides are then probed with proteins, antibodies or cell lysates to identify specific interactions.	Stable	Discovery	*In Vitro*	(Neiswinger et al., 2016)
FRET: fluorescence/Foerster resonance energy transfer	Bait and preys are fused to acceptor and donor fluorophores and the excitation signal upon interaction provides a detectable emission signal.	Stable, transient	Library	*In Vivo* *In Vitro*	Reviewed in (Kenworthy, 2001)
BRET: bioluminescence resonance energy transfer	Bait is fused to energy donor (Renilla l*uciferase*) and prey to fluorescent protein (GFP/YFP). Interaction provides a detectable fluorescent signal.	Stable, transient	Library	*In Vivo, In Vitro*	Reviewed in (Pfleger et al., 2006)
LUMIER: Luminescence-based mammalian interactome mapping	Co-immunoprecipitation based method where bait-prey interaction results a *luciferase* signal.	Stable	Library	*In Vitro*	(Barrios-Rodiles et al., 2005)
LRC-TriCEPS: Ligand–receptor capture – trifunctional chemoproteomics reagents	Involves three-part chemoproteomics reagent, of which one part is fused to possible ligand, one to biotin and last is allowed to bind to glycosylated receptor. Biotin allows the affinity capture and analyses by MS.	Stable, transient	Library	*In Vitro*	(Frei et al., 2013)
PLA: Proximity ligation assay	Design for detection of PPIs in fixed cells and tissues. Interaction of proximity probe tagged bait and prey leads to synthesis of DNA with elements for fluorophore binding allowing the visualization of interactions.	Stable, transient	Library	*In Vitro*	(Fredriksson et al., 2002)
LuTHy: combination of BRET and LuC (LUMINER)	Combines two methods: BRET and luminescence-based co-precipitation providing more insensitivity to tag location and complex stability.	Stable, transient	Library	*In Vitro*	(Trepte et al., 2018)

Another PPI analysis method is protein-fragment complementation (PCA), in which, similar to some previously mentioned two hybrid methods, a signalling protein is divided into two fragments. One of these fragments is fused to the bait and the other to the prey. Apart from in the two hybrid methods, unification does not result in the transcription of the reporter gene, but a direct signal, such as fluorescence, luminescence or cell survival. Bimolecular fluorescence complementation (BiFC) is probably the most used PCA method (Kerppola, 2008). In this method, the fluorescent protein (e.g., YFP or GFP) is divided into two non-fluorescent parts that are fused to the bait and prey. The interaction leads to the complementation and production of a fluorescent signal that can be detected by microscopy or flow cytometry techniques. The BiFC method allows the PPI to be detected in living cells and provides spatial information regarding the interaction. However, there is a delay in producing the fluorescent signal, and the reaction is irreversible. Consequently, the BiFC method cannot be used to detect PPI dynamics. This limitation is overcome using the NanoLuc PCA technique, in which a small luciferase, NanoLuc, is divided, and the fragments are fused to the interacting proteins (Dixon et al., 2016). Complementation enables interaction detection using a luminesce signal. The NanoLuc method also allows interaction kinetics to be detected.

PPIs are often validated using a co-immunoprecipitation coupled to western blot (CoIP) technique. In this approach, the interacting Proteins A and B are purified using an affinity tag or antibody for Protein A, and the interaction is verified from purified proteins by western blotting using an antibody against Protein B. CoIP is performed from whole cell lysates where the interaction proteins are present in native forms with the necessary modifications. This approach is simple, and it can be utilised in most laboratories. However, it has limitations as it depends on antibody specificity, has multiple ex vivo steps, requires stable interactions that are not lost during the

protocol and, like a normal western blot, is sensitive to errors in detection, such as too long of exposure times. An overexpression of both interacting proteins might also introduce artificial PPIs.

The interacting proteins can also be bound to each other by chemical or photoreactive cross-linkers that covalently link two non-covalently interacting proteins with their two reactive groups. Cross-linked proteins are the further analysed using a tool such as mass spectrometry (MS) (Vasilescu et al., 2004). Cross-linking provides spatial information of the proteins' interaction sites. Thus, it can be used to obtain structural information of interacting proteins and/or protein complexes. Cross-linking can be done both in vitro and in vivo, can identify transient interactions and is not very sensitive to sample handling after the cell lysis.

In functional in vitro protein microarrays, full-length proteins or protein domains are arrayed in a slide, and the slide is then probed with proteins, protein domains or cell lysates to identify specific PPIs with various biochemical reactions (Neiswinger et al., 2016). This method allows thousands of interactions to be screened simultaneously, more than is possible to access by a single MS run. The interactions are screened in an in vitro environment without a cellular context that might introduce artificial interactions.

4.1 Affinity purification coupled to mass spectrometry

In addition to the previously mentioned methods, AP-MS is a widely utilised technique in PPI analysis. In this method, baits, together with their interacting preys, are affinity purified using antibody or affinity tags, and the preys are identified by MS. Multiple variations of the method exist, which this study divides into three groups: antibody-based AP-MS, affinity tag-based AP-MS and proximity labelling.

Antibody-based purification can be done from primary cells or cell cultures without an additional transfection of bait or prey proteins. The bait is captured by a specific antibody and purified together with its interacting preys, for example by a secondary antibody coupled to Sepharose or magnetic beads. Proteins are then digested into peptides and identified using MS. This allows the purification of intact complexes with endogenous expression of interacting proteins. However, the expression of the interacting proteins in the primary cells might be significantly low, and many of the interactions are lost during the purification. Also, the availability of good antibodies may cause restrictions in the analysis.

These limitations have been overcome by introducing affinity tags to the bait prior to transfection. Widely used tags include FLAG, HA, STREP and GFP. Affinity tags allow for efficient purification, and the method can be utilised in high-throughput screening of PPIs of multiples baits. Tagging with multiple affinity tags also allows for double-step purification, resulting in a purer complex than in single-step purification.

Recently, complementary proximity labelling techniques coupled to AP and MS analyses have been introduced. These include proximity dependent biotin identification (BioID) (Lambert et al., 2015; Roux et al., 2012) and ascorbate peroxidase (APEX) techniques (Rhee et al., 2013). In a BioID analysis, the baits are tagged with modified biotin ligase (BirA), which is activated in cell cultures by adding biotin in the growth media. Activation of BirA allows it to biotinylate the interacting proteins with within close proximity (10 nm). Biotinylated proteins can be purified using

Streptavidin beads. The direct biotinylation of interacting proteins allows the use of harsh lysis conditions (e.g., sonication and benzonase) without losing the interacting proteins. This is important, especially in the analysis of nuclear or membrane proteins. In addition to the BioID analysis, a BioID2 method has been introduced (Kim et al., 2016). In the BioID2 method, the biotin ligase tag is smaller and requires less biotin in the culturing media; however, limitations include the fact that the endogenous biotin may activate the tag and cause more background.

Tagging with APEX is also used for labelling the proximal proteins (Rhee et al., 2013). APEX has the ability to oxidase phenol derivatives to phenol radicals. Adding biotin-phenol and H_2O_2 to culturing media leads to the generation of biotin-phenol radicals and their fusion to interacting proteins' Tyr, Trp, His and Cys amino acids within a labelling radius of 20 nm. Similar to the BioID analysis, biotin allows the interacting proteins to be purified using streptavidin beads.

APEX has a longer labelling radius than BioID analysis. This allows more interacting proteins to be purified, but it also introduces more noise to already unspecific data and makes further analysis more difficult. The labelling time of APEX is counted in minutes, while BioID analysis needs a hours -long labelling time (Trinkle-Mulcahy, 2019). This is benefit over the BioID method, but it should be kept in mind that adding H_2O_2, even for short time, might affect the oxidative status of cells and cause additional stress. With the shorter labelling time, APEX is more capable to catch "snapshot" -like interactomes, while BioID catches the interactions for a longer period of time. Both BioID and APEX methods are recently reviewed more in depth in (Trinkle-Mulcahy, 2019).

Proximity labelling techniques do not require physical associations between bait and preys, which may introduce false positives to results. Therefore, sufficient controls are needed to filter the data properly. Controls are also essential to filter out the unspecific endogenously biotinylated proteins, such as carboxylases. In addition, PPIs from proximity labelling lack the information of primary or secondary nature of PPIs. However, in some cases, such in studying PPIs in liquid-like membrane less organelles (known also as phase separation condensates), proximity-based identification of PPIs might be beneficial, as the PPIs may be weak, short, dynamic, have heterogenic conformation, and be formed between IDRs of interacting proteins (Tompa et al., 2008). For example, TFs are known to form condensates trough their effector domains, which are found to be enriched with IDRs (Boija et al., 2018).

II STUDY AIMS

While TF DNA binding has been well examined, a system-level understanding of TF PPIs on transcriptional regulation is still lacking. This information would extend the knowledge of how TF activities are regulated and how the signal is further transmitted to Pol-II and active transcription. The overall aim of this thesis project is to study human TF PPIs at the systems level and deepen the understanding of the role of mutated TFs and their PPIs in PIDs.

The specific aims of Studies I, II and II were as follows:

I. To perform a global and comprehensive protein interactomics analysis of 110 TFs.

II. To analyse the effect of biallelic missense mutation in the CEBPE gene on C/EBPε functions in PID patients.

III. To study the functional role of three NFKB1 patient mutations in PIDs.

III MATERIAL AND METHODS

1. DNA constructs (I–III)

The TF full-length coding sequences without stop codons were obtained in Gateway® compatible entry vectors, either from the ORF collection of the Genome Biology Unit of the University of Helsinki (https://www.helsinki.fi/en/researchgroups/genome-biology-unit) or commercially from GenScript. Using Gateway® cloning technology (Thermo Fisher Scientific), TF coding sequences were cloned into the destination vectors generated in the Varjosalo laboratory, which are listed in Table 3:

Table 3: Destination vectors used in different studies

Study	Vector	Tag	Tag terminus
I, III	pTO_HA_StrepIII_N_GW_FRT	HA, Strep x3	N-terminal
I, III	pTO_Myc_BirA_N_GW_FRT	Myc, BirA	N-terminal
I, II	pTO_HA_StrepIII_BirA-N_GW_FRT (MAC-tag)	MAC: HA, Strep x3, BirA	N-terminal

1.1 Mutagenesis

Patient mutations used in studies II and III were introduced to entry clones by PCR using the primers listed in Table 4:

Table 4: PCR primers to generate mutation constructs in Studies II and III

Study	Mutation construct	Primer
II	Human C/EBPε R219H	Forward: GCAACAACATCGCCGTGCACAAGAGCCGAGA
		Reverse: GCCTTGTCTCGGCTCTTGTGCACGGCGATGT
II	Mouse C/EBPε R219H	Forward: GTAACAACATCGCGGTGCACAAGAGCCGGGA
		Reverse: GCCTTGTCCCGGCTCTTGTGCACCGCGATGT
II	Human C/EBPε V219A	Forward: GCAACAACATCGCCGCGCGCAAGAGCCGAG
		Reverse: CTCGGCTCTTGCGCGCGGCGATGTTGTTGCG
III	NFKB1 H67R	Forward: TGTATGTACCTTTGTTGCCGT
		Reverse: TTCAGCTTAGGAGCGAAGGC
III	NFKB1 I553M	Forward: GAGAGCAGATTCCATTCTTGAGT
		Reverse: TAACATGCACCCAAACATAACAG
III	NFKB1 R157X	Forward: CTGGAAGCATGAATGACAGAGGCGTGTATAAGGG
		Reverse: CTCTGTCATTCATGCTTCCAGTGTTTCAAATACTTTTTTC
III	NFKB1 A156Sfs	Forward: AAGAAAAAAGTATTTGAAACACTGGAAAGCACGAATGACA
		Reverse: CGCCTCTGTCATTCGTGCTTTCCAGTGTTTCAAATA

2. Generation of cell lines (I–III)

Stable TF expressing cell lines were generated using an Flp-In™ transfection system (Thermo Fisher Scientific). Shortly, Flp-In™ 293 T-REx cells (Thermo Fisher Scientific) were transfected with expression vectors (Table 3) and a pO44 Flp recombinase vector using FUGENE 6 (Promega) as a transfection reagent. The Flp-In™ 293 T-Rex cell line contained a single FRT site in a transcriptionally active site. Co-transfection of the expression vector together with recombinase allowed for the stable integration of the gene of interest as a single copy into genome. This resulted in a monogenic cell line whose bait expression was both inducible and adjustable by tetracycline or doxycycline.

After transfection, cells were selected by Hygromysin B (Promega) for three weeks. Twenty-four hours prior to harvesting, the bait expression was induced by 2 μg/ml of tetracycline and, in case of the BioID analysis, the biotin ligation was activated by supplementing the media with 50 μM of biotin. Approximately 5 x 10^7 cells (5 x 15 cm plates) were harvested, centrifugated at 4° C, snap frozen in liquid nitrogen and stored at -70° C.

In Study II, approximately 5×10^6 Jurkat cells per sample were transfected by electroporation (Neon® Transfection system, Thermo Fisher Scientific) with 1350 V 10 ms pulse width and three cycles. Cells from ten tips (~5×10^7 cells) were combined for an AP of one construct.

3. Affinity purification (I–III)

For the AP, AP-MS samples were lysed in a HENN-lysis buffer (50 mM HEPES pH 8.0, 5 mM EDTA, 150 mM NaCl, 50 mM NaF, 1% N-dodecyl-b-D-maltoside [Sigma], 1.5 mM Na_3VO_4, 1 mM PMSF and 1× protease inhibitors cocktail). In the BioID analysis, a HENN lysis buffer was supplemented with 80 U/ml Benzonase Nuclease (Santa Cruz) and 0.1% SDS. In addition, harsher lysis conditions, including three cycles of three-minute sonication, were used.

After clearance by centrifugation, lysates from both the AP-MS and BioID analyses were loaded into spin columns containing Strep-Tactin beads (IBA), the unbound proteins were washed and proteins were eluted with 0.5 μM of biotin. Purified proteins were reduced using Tris (2-carboxyethyl) phosphine (TCEP), alkylated with iodoacetamide and digested into peptides with trypsin. Tryptic peptides were further purified using C-18 micro spin columns (The Nest Group, Inc.) in accordance with the manufacturer instructions. Finally, vacuum-dried peptides were dissolved in Buffer A (1% acetonitrile and 0.1% trifluoroacetic acid in MS grade water).

3.1 Affinity purification from patient peripheral blood mononuclear cells (III)

Approximately 2 x 10^6 patient peripheral blood mononuclear cells (PBMCs) were lysed in the HENN lysis buffer, and the lysate was precleared by incubating for one hour with rec-Protein-A-Sepharose 4B beads. This was followed by a four-hour incubation with anti-NFKB1 antibody (1:50, 3035: CST) prior to a one-hour incubation with rec-Protein-A-Sepharose 4B beads. After the washes, the proteins bound to the beads were digested and processed as previously described.

4. Mass spectrometry analysis (I–III)

Tryptic peptide mixtures were analysed using an EASY-nLC II system coupled to Orbitrap mass spectrometers (Elite: Orbitrap Elite™ Hybrid Ion Trap-Orbitrap or QE: Q Exactive™ Hybrid Quadrupole-Orbitrap™, Thermo Fisher Scientific). Peptides were run through a C-18 packed pre-column and analytical column using a 60-min linear gradient of increasing amounts (5–35%) of Buffer B (98% acetonitrile and 0.1% trifluoroacetic acid in MS grade water). The analysis was performed in a data-dependent mode: one high resolution MS1 scan was followed by CID fragmentation and MS2 scans of the 20 most intense precursors. A detailed description of the parameters used can be found in Kaustio et al.'s study (III).

5. Bioinformatics and data visualisation (I–III)

5.1 Protein identification and quantification (I–III)

Proteins were identified using Proteome Discoverer analysis software (I, II) coupled to a SEQUEST search engineer or MaxQuant-software coupled to an Andromeda search engine (III). The MS data were searched against the human component of a UniProt database (reviewed entries, SwissProt) complemented with tag sequences, trypsin, BSA and GFP. A maximum of two missed cleavages were allowed, and carbamidomethylation (C) and oxidation (M) were used as the fixed and variable modifications, respectively. False discovery rate (FDR) cut-offs of 0.01 and 0.05 were used in the AP-MS and BioID analyses, respectively. Proteins from AP-MS samples in Study III were MS1 quantified using MaxQuant. For the phosphorylation analysis (III), Ser, Thr and Tyr phosphorylation were set as the variable modifications in MaxQuant, followed by the quantification of phosphopeptides and their corresponding unmodified peptides and the calculation of the phosphorylation degree.

5.2 Filtering the specific protein-protein interactions (I–III)

Specific PPIs were filtered from an unspecific background using Significance Analysis of INTeractome (SAINT) express software (I) or by manual thresholds, such as frequency, abundance compared to control runs and comparison to the contaminant database (https://www.crapome.org; II, III). Similarly tagged and processed control GFP runs (with and without NLSs) were used as controls in both approaches.

5.3 Analysis of interaction data

The filtered PPIs were further analysed using the data analysis tools listed in Table 5:

Table 5: Tools used for PPI analysis

Tool	Used for	Web address	Study
Cytoscape	Visualization of PPI networks	https://cytoscape.org	I, II, III
Crapome	Filtering of unspecific PPIs	https://www.crapome.org	III
CORUM	Protein complex annotations	https://mips.helmholtz-muenchen.de/corum/	I, II
DAVID	Functional annotations of PPIs	https://david.ncifcrf.gov	I, II
PINA2	Known PPIs	http://omics.bjcancer.org/pina/	I, II, III
Uniprot	Functions of interacting proteins	https://www.uniprot.org	I, II, III
BioGrid	Known PPIs	https://thebiogrid.org	I
String	Known PPIs	https://string-db.org	I
Intact	Known PPIs	https://www.ebi.ac.uk/intact/	I
Human Cell Atlas	Subcellular locations of interacting proteins	https://www.humancellatlas.org	I
ProHits-viz	Clustering and correlation analysis of PPIs and preys	https://prohits-viz.lunenfeld.ca	I
Clustal Omega	TF sequence alingment	https://www.ebi.ac.uk/Tools/msa/clustalo	I
JASPAR	TFs DNA binding motifs	http://jaspar.genereg.net	I
RSAT	Alignmet of DNA binding motifs	http://rsat.sb-roscoff.fr	I
ProteinAtlas	TF tissue expression	https://www.proteinatlas.org	I
Interferome	IFN-related PPIs	http://www.interferome.org	II

6. Western blotting (III)

Western blots in Study II from Flp-In™T-REx™293 and PBMC cells were performed using antibodies listed in Table 6:

Table 6: Antibodies used for western blotting

Antibody	Vendor	Dilution
Anti-HA	HA.11, Covance	1:1500
Anti-p50/p105	#3035, CST	1:1000
Anti-tubulin	ab7291, Abcam	1:10000
Anti-mouse	NA931, GE	1:2000
Anti-rabbit	Po448, Dako	1:2000

6.1 Proteasome mediated degradation analysis (III)

Flp-In™T-REx™293 cells expressing WT or R157X NFKB1 were treated with 0, 5, 10 or 20 µM of the proteasome inhibitor MG132 (474790, Merck Millipore) for four hours. The p50 and p105 levels were detected by western blotting using an anti-p50/p105 antibody.

7. Nanostring (II)

The mRNA levels of patients and matched controls' PBMC cells were detected using a direct digital detection of mRNA by Nanostring technology (Nanostring Technologies) with a custom gene panel of 50 genes linked to immunological responses. Shortly, blood samples were collected and PBMCs were separated with Vacutainer® CPT™ cell preparation tubes with Sodium HeparinN (BD). Next, RNA was extracted using RNeasy® Mini Kit (Qiaqen) and introduced to the Nanostring protocol, which included mRNAs' fusion to the 5' reporter probes tagged with the fluorescent barcodes of the target genes. The fluorescent signals were scanned using an nCounter Digital Analyser.

The expression data were normalised and investigated in nSolver™ 4.0 analysis software (Nanostring Technologies). The ratio between the patients and the controls' gene expression was calculated, and a student t-test was used for statistics.

IV RESULTS AND DISCUSSION

1. Protein interaction landscape of human TFs

To map the PPIs of 110 TFs, we generated Flp-In™ T-REx™ 293 cells stably expressing selected TFs. Using these we utilised two AP methods for each TF: First, the TFs were tagged with a biotin ligase that allowed the biotinylation of nearby proteins (BioID). This was followed by an AP and MS-based identification of the biotinylated proteins. Second, stable TF complexes were affinity purified using a strep tag followed by an MS-based identification (AP-MS).

These methods allowed us to capture 7233 and 2176 PPIs with BioID and AP-MS, respectively. Of these, 1525 PPIs (BioID) and 345 PPIs (AP-MS) had been previously reported in String experimental, Biogrid, Intact or PINA2 databases, as well as in Li et al.'s TFs interactome study (Li, Wang, et al., 2015). The GO-BP enrichment analysis showed a significant enrichment of terms linked to transcription and transcriptional regulation, and, according to the Cell Atlas, 80% of the preys had nuclear localisation. The BioID method, which has the ability to catch transient and proximal interactions (Varnaite et al., 2016), yielded more interactions than the AP-MS method. This result indicates that the nature of TF PPIs is transient and weaker and that they do not generally form stable complexes. These transient PPIs are also seen in phase separation model of TFs', where PPIs might be weak, short, dynamic and have heterogenic conformation. Therefore, most of the following analyses described in study I were performed using BioID data.

Interestingly, we found 175 bait-bait interactions between 54 TFs and NFI-family members (NFIA, NFIB, NFIC and NFIX; Figure 6). NFIs have an important role in development, and they are often part of oncufusion proteins in several cancer types. Knock-out studies have identified them to be essential e.g. for central nervous system, lung, skeletal, muscle, brain and tooth development (Steele-Perkins et al., 2005; Mason et al., 2009; Piper et al., 2019; Driller et al., 2007; Campbell et al., 2008; Steele-Perkins et al., 2003; Shu et al., 2003). Some TF-NFI interactions had previously been reported, but NFIs are not known to interact with many other TFs. Our data indicates a cross-talk between NFI and other TF signalling: transcriptional activity of NFI may be mediated trough other TFs, of *vice versa*. This could explain their necessity for several developmental processes.

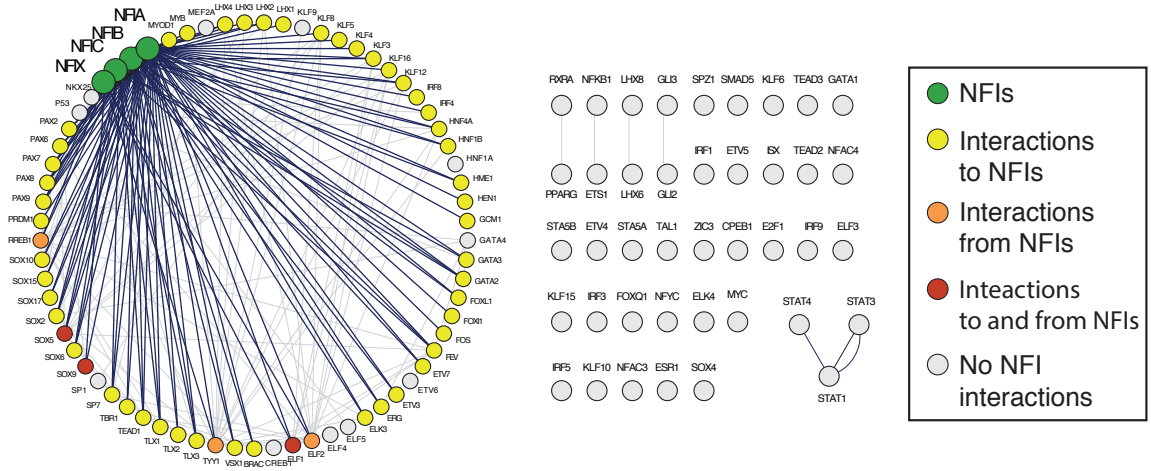


Figure 6: BioID bait-bait interactions within studied TFs. Blue edges indicate PPIs to/from NFIs. Nodes are colour coded as shown in right side of the figure.

Next, the identified TF PPIs were mapped to the basal transcription machinery, Pol-II complex and Mediator complex proteins. Several TFs interacted with SAGA and TFIID, the initiation complexes of the basal transcription machinery assembly. However, only six TFs interacted with the Mediator complex members, although the role of the Mediator complex is often described as mediating the signal from the TFs to the basal transcription machinery. This result indicates that under the studied conditions, in certain activity state, TFs tend to form PPIs rather with TFIID and SAGA than with mediator complex. None of the TFs interacted with the TFIIA, TFIIB, TFIIF, TFIIE, TFIIH or Pol-II complexes, indicating that the signal from the TFs to Pol-II was mostly transferred via SAGA, TFIID or the Mediator complex.

To further study the similarities and differences in TF PPIs, we performed a prey-prey correlation analysis (Figure 7). This analysis showed which preys are observed to act in a correlate manner between the baits, indicating co-localisation and/or participation in same complex. Baits that drive the same cluster share the interactions within the cluster and might have similar or shared biological functions.

The prey-prey analysis revealed 15 interesting biological clusters (Figure 7). For example, Cluster 2 consisted of proteins linked to actin and myosin signalling. This cluster was mainly formed from FOS PPIs, but STAT1 and FOXL1 had some similar actin and myosin linked interactions. FOS and STAT1 interacted with Beta-actin (ACTB) and nuclear myosin (MUO1C), both of which are linked to nuclear actin signalling.

Furthermore, Clusters 10 and 11 (Fig 7) were formed from preys linked to mRNA splicing and processing. Baits driving these clusters—baits having similar interactions—were mainly SP7, GATA1 and GATA3. In particular, SP7, which has not previously been linked to mRNA splicing, had multiple splicing related interactions and should be studied further in the context of RNA splicing.

Clusters 14 and 15 (Figure 7) included preys linked to ATP-depended chromatin remodelling complexes, such as INO80, and histone modifiers and modifying complexes, such as the SAGA,

NuA4/Tip60 HAT and NSL HAT complexes. Baits driving these clusters included TYY1, HNF4s, MYC, ELF4, ELF2, ELF1, KLF6 and KLF8, suggesting their importance in chromatin modulation.

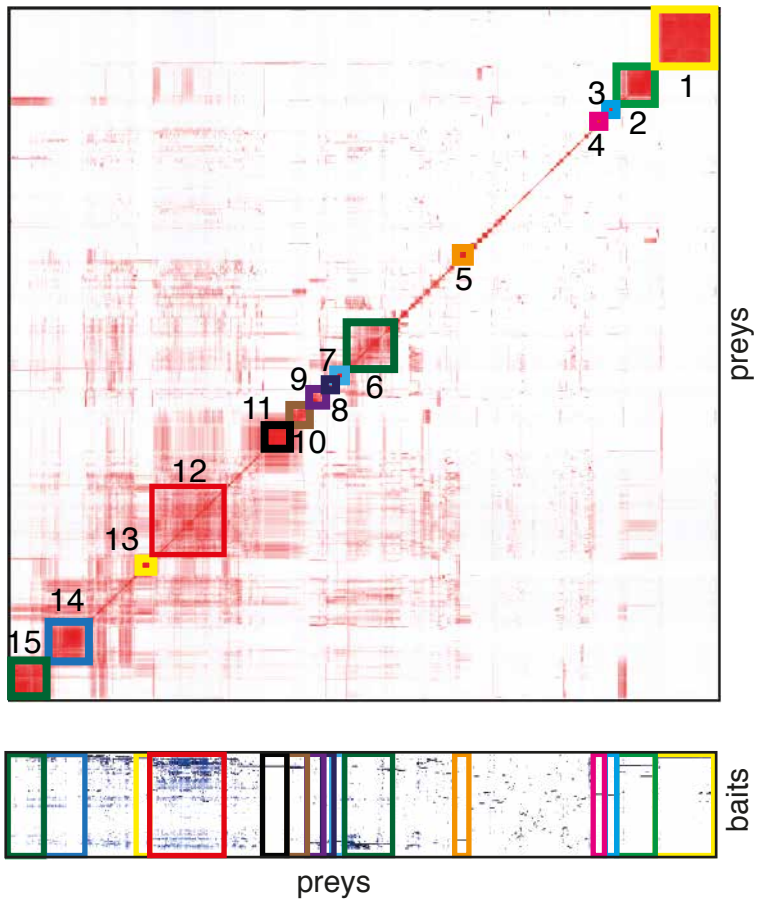


Figure 7. Prey-prey correlation analysis was performed on preys from BioID experiments using Prohits-viz-software tool. Results of correlation analysis indicate which preys are seen in a correlated manner between the baits. Correlation may suggest that preys belong to same complex and/or co-localize. Results are illustrated in heatmap with preys in both x- and y-axis. Corresponding bait-prey heat map below the heatmap shows which and how many baits are driving the prey clusters.

This study allowed us to extensively examine human TF PPIs. We were able to identify TFs' PPIs with many known TF interaction partner groups (presented in chapter 2), suggesting that most of the TFs were transcriptionally active. These interactions include PPIs with PTM transferring enzymes, PPIs with basal transcription machinery and Mediator complex, PPIs with chromatin modulation proteins, PPIs with RNA splicing and processing machinery, PPIs with nuclear actin and myosins and PPIs with DNA repair and replication machinery.

Proximity labelling with BioID method identified more PPIs than AP-MS, indicating that proximity labelling was more powerful tool to analyse TFs' PPIs. However, it would be also interesting to see, how results from APEX labelling would differ from results of BioID labelling, as the APEX labelling time is shorter and the labelling radius longer (see chapter 4.1 for comparison of BioID and APEX).

Large TFs' PPIs analysis allowed us to identify subgroups of TFs linked to specific biological functions. Using prey-prey correlation analysis, we could identify groups of TFs related to chromatin

remodelling, actin and myosin signalling, and RNA splicing. In addition, we found 54 TFs to interact with NFIs, suggesting previously unknown synergism between NFI and other TF signalling. These subgroups, on the other hand, highlight the cooperativity of human TFs in regulating cellular functions, but tells also that TFs may function in different cellular processes, even with DNA-binding independent mechanisms. Our data highlighted the simultaneous nature of nuclear gene expression processes, such as chromatin modulation, transcription, RNA processing and RNA splicing and regulation. The results of this study also showed that TFs play an important role in regulating these processes as they can interact with the proteins involved in each of them.

Furthermore, the large mapping of TF PPIs provides an opportunity to profile other TFs' interactomes against large set of identified TF interactomes. These profiling may produce valuable information not only about similarities in interactions, but also regarding specific roles of TFs in the context of more than 100 TFs.

Taken together, this study provides an excellent repository of human TF PPIs to be used in other studies.

2. CEBPE mutation causes non-canonical autoinflammatory inflammasomopathy

CEBPE encodes C/EBPε, which is a TF required for cell differentiation and functions in myeloid and lymphoid lineages cells (Bedi et al., 2009). Mutations in C/EBPε have previously been linked to SGDs, where patients' neutrophils display abnormal morphology and numbers of granules together with abnormalities in the nuclear shape.

This study described a novel C/EBPε associated disease, CAIN (C/EBPε-associated autoinflammation and immune impairment of neutrophils) in an index family that has been studied from the 1970s (Repo et al., 1979; Pasanen et al., 1987; Murros et al., 1974). Patients displayed normal neutrophils granules; however, the nucleus were hyposegmented, and the neutrophils had impaired chemotaxis. Phenotypes included both immunodeficiency and autoimmune symptoms. CAIN patients' symptoms included periodic attacks of fever, systemic inflammation and abdominal pain.

CAIN is caused by a previously unreported homozygous R219H C/EBPε mutation. Mutations were detected from affected family members by whole exome sequencing, while Sanger sequencing was used to validate the heterozygous carriers in the family.

The molecular mechanisms behind the disease were studied using multiomics and data-driven approach. This study included a PPI analysis using the BioID method, a ChIP-sequencing (ChIP-seq) analysis of WT and mutant C/EBPε DNA binding, RNA-sequencing (RNA-seq) of patients and controls' granulocytes with and without inductions, a Nanostring analysis of the patients and controls' PBMCs and inflammasome activation analyses of the patients' macrophages and monocytes.

The BioID method identified in total 141 PPIs for WT and mutant C/EBPε (Figure 8A). Of these, 108 were significantly changed between the WT and mutant C/EBPε: two had an increased and 106 had a decreased affinity to the mutant compared to WT (Figure 8B). Many proteins with decreased affinity to the mutant C/EBPε were transcription repressors, indicating changes in C/EBPε transcriptional activity. Therefore, transcriptional activity was accessed using a ChIP-seq analysis. This resulted in widely dysregulated DNA binding in a GOF manner: patients' granulocytes had 10,322 C/EBPε binding sites, while healthy controls had 3391.

As BioID data are not easy to validate with CoIP, we validated some of the PPI results in Jurkat cells and with mouse C/EBPε in Flp-In™ 293 T-REx cells. In addition, as the mutation in neighbouring amino acid (V218A) is known to cause SGD (Khanna-Gupta et al., 2007), we also analysed its interactome. However, the PPIs of V218A were not affected similarly than with R219H mutation indicating unique mechanism for the CAIN disease.

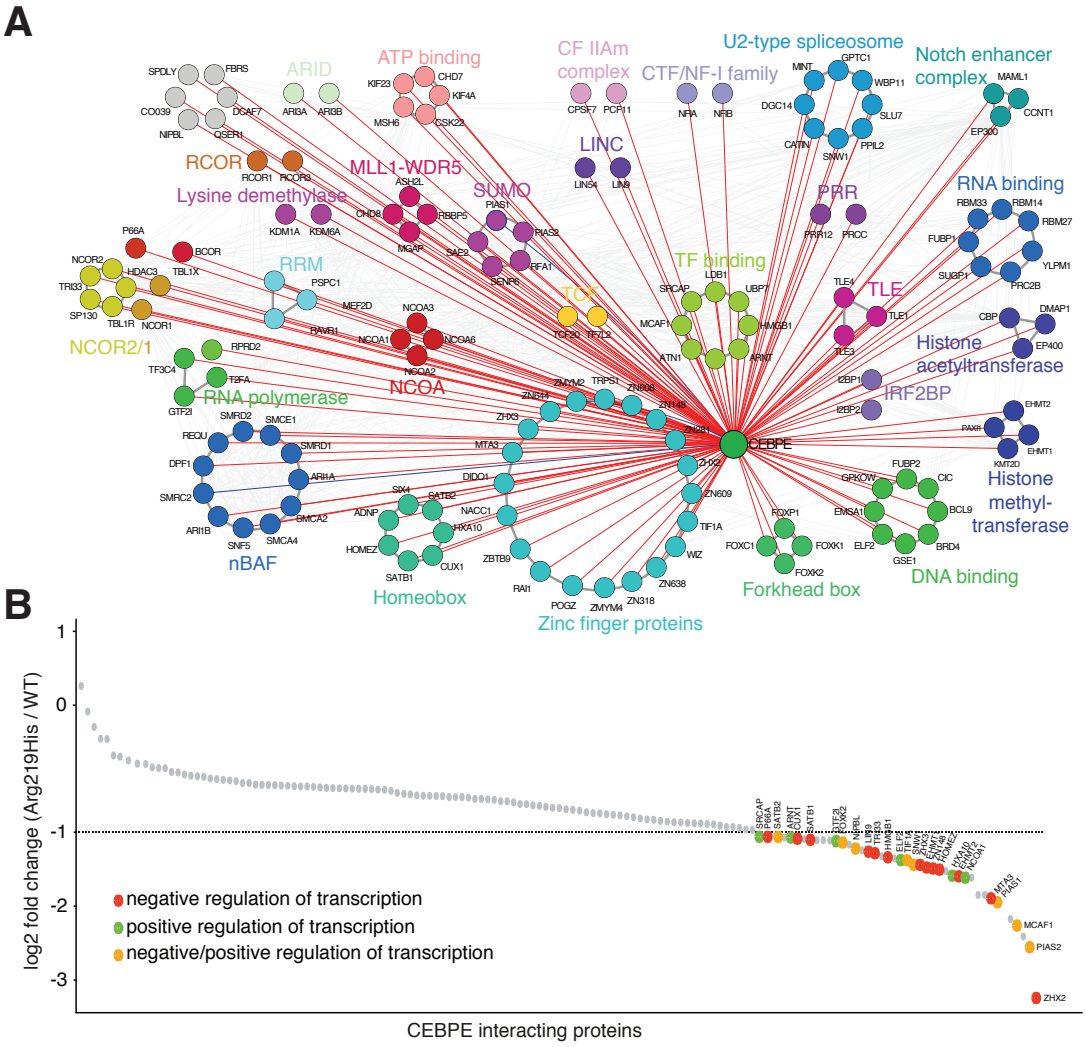


Figure 8. A) PPIs of C/EBPε. B) Log2 fold change of interacting proteins between WT and R219H C/EBPε

Changes in transcription were further investigated using an RNA-seq analysis of unstimulated granulocytes. This analysis revealed 464 differentially transcribed genes, of which 189 were related to IFN signalling. Furthermore, according to the ChIP-seq analysis, C/EBPε had binding sites in 271 of these differentially transcribed genes, of which 80 had patient-specific C/EBPε binding. This indicates these genes are under direct C/EBPε control.

RNA-seq analyses were also done after stimulations. As many dysregulated genes in unstimulated condition were related to IFN-signalling, granulocytes were first treated with IFNα2b and IFNγ. In addition, granulocytes were stimulated with LPS and bacterial DNA to model the clinical phenotype of bacterial infections. After IFNα2b, IFNγ and bacterial treatments, 534, 427 and 470 genes were differentially expressed, respectively. Of these, 266, 208 and 289 had C/EBPε binding sites in the ChIP-seq analysis.

A Nanostring analysis was performed using a custom gene panel of 50 immuno-related genes. mRNA levels of genes linked to the interferon signature did not differ between the patients and controls. However, the levels of genes involved with the JAK/STAT pathway and inflammasome

activation were increased in the patients. These included CASP5, NLRP3 and CASP8. Moreover, NLRP3 was upregulated in an RNA-seq analysis, and C/EBPε binding within its gene was upregulated.

Transcriptional changes and DNA-binding differences in inflammasome-related genes led us to investigate closer non-canonical and canonical inflammasome activation in patient macrophages and monocytes. Canonical inflammation activation did not differ between the patients and controls, but CASP5 mediated activation of non-canonical inflammasome by intracellular LPS resulted in increased activation, which was detected by elevated IL-1β and IL-18 levels. Elevated CASP5 mRNA levels were also detected in resting macrophages. Pro-CASP5 processing into an active CASP5 was investigated in greater detail by western blotting patients' PBMCs with a specific antibody. The results showed that the expression of pro-CASP5 was higher in patients, and, more importantly, significantly more of it was processed compared to healthy controls. Taken together, these results indicated that compared to healthy controls, the patients had increased non-canonical NLRP3 inflammasome activation due to the constrictively expressed CASP5.

The symptoms and the molecular mechanisms of CAIN differed from mechanisms of SGD (presented in Chapter 3.3.3) separating it clearly to own disease. Most importantly, the neutrophils displayed normal granules in neutrophils. SGD causing mutation in neighbouring amino acid (V218A) did not cause similar changes in PPIs and no similar proteomics changes were seen in CAIN than in V218A causing SGD.

Overall, this study provides valuable information regarding how a single TF mutation may cause diseases through a multi-level mechanism involving changes in TF PPIs, DNA binding and transcriptional regulation. The changes caused overactivation of non-canonical inflammasome and dysregulated IFN-signalling. To our knowledge, this study is the first instance to report that widely dysregulated transcription causes PIDs. The mechanism is most likely not exclusive to C/EBPε, and similar molecular mechanisms may be behind other TF-related immunodeficiencies or diseases.

3. Damaging heterozygous mutations in NFKB1 lead to diverse immunologic phenotypes

Affected members of three unrelated Finnish families (Figure 9) with diverse immunological phenotypes were sequenced by whole genome or whole exome sequencing. This uncovered three different mutations, H67R, I553M and R157X, in the *NFKB1* gene that encoded a TF important for multiple immunological processes. Using Sanger sequencing, the presence of these mutations was validated in other members of the families.

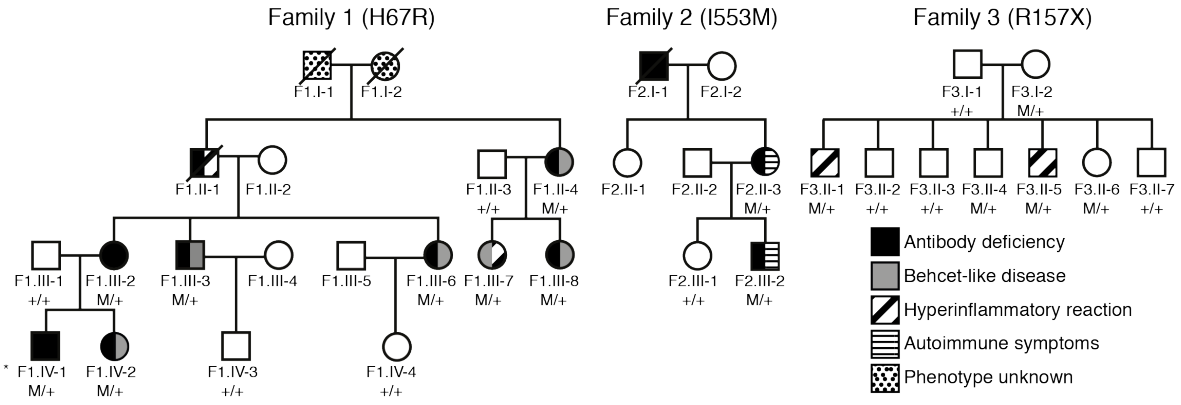


Figure 9: Pedigrees, clinical phenotypes and sequencing results of study subjects.

To study the molecular phenotype behind the symptoms, we generated Flp-In™ 293 T-REx cell lines stably expressing WT, H67R, I553M and R157X NFKB1. These were used for a PPI analysis with AP-MS and BioID, expression analyses, localisation analyses and luciferase assays. Moreover, patient PBMC cells were used to observe the NFKB1 expression using western blot and MS methods. MS was also used to check the phosphorylation status of NFKB1 in samples.

H67R mutated NFKB1 was detected to have decreased transcriptional activity due to various mechanisms: the H67R mutation decreased the nuclear entry of TNF-activated NFKB1, reduced NFKB1 interactions with its dimerization partner RelB, indicating decreased transcriptional activity (Figure 10), and resulted in increased interactions with NEMO, a cytosolic component of the NFKB1 complex.

A I553M mutation is located in C-terminus of NFKB1 that contains the ankyrin repeat areas. This area is responsible for interactions with inhibitory proteins that keep the NFKB1 in full length p105 form in cytosol. An AP-MS analysis showed that the mutant NFKB1 had increased affinity to one of the inhibitors, IKBε. In addition, the I553M mutation resulted in decreased phosphorylation of NFKB1 serines at positions S893 and S907 and caused p105 subunit degradation upon treatment with a TNF gradient.

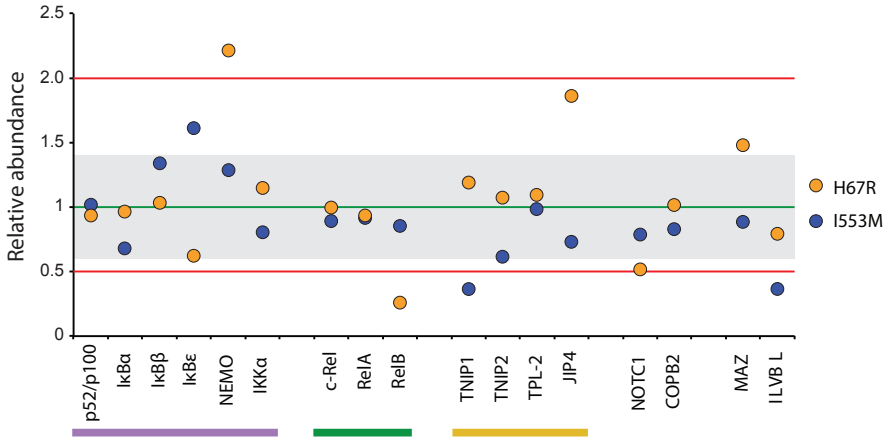


Figure 10: Relative abundance changes (mutant dived by WT) in NFKB1 protein interactions

R157X was truncating mutation for the NFKB1. Using a MS, only a few N-terminal peptides were detected. Similarly, when NFBK1 was purified from patient PBMC cells using an antibody, only a few peptides from the N-terminus were detected (Figure 11A). This finding suggests the existence of a dominant negative mechanism where both the R157X mutated NFKB1 and the NFKB1 from the healthy allele were degraded resulting in almost a complete loss of NFKB1 in the cell. The same result was observed using a western blot from HEK cell line models as well as from patient PBMCs (Figure 11B). Degradation could be rescued by treating the cells with an MG132 proteasome inhibitor, indicating that the degradation of NFKB1 from healthy allele is performed by proteasomes.

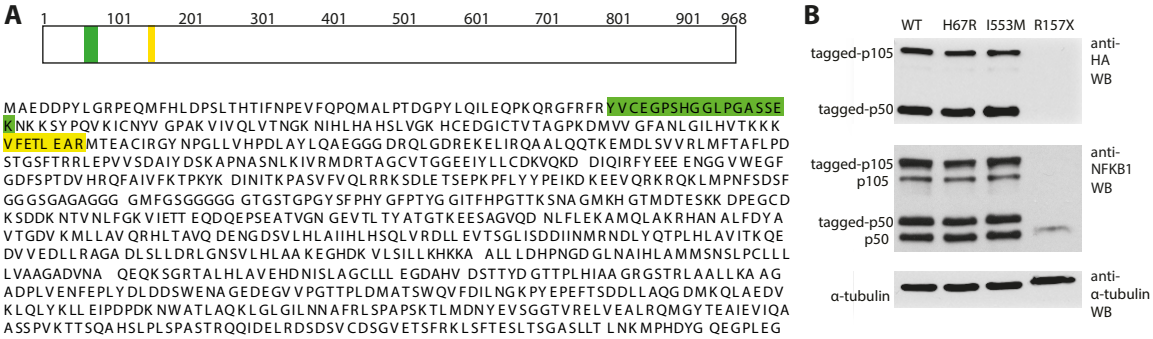


Figure 11. A) Peptides identified fromR157X mutation carrying patient PBMCs after the AP with NFKB1 antibody. Peptides were detected using MS. Green: FDR <0.01, yellow: FDR < 0.05. B) Expression of p50 and p105 forms of NFKB1 in Flp-In T-Rex- cell models were detected using WB.

As NFKB1 is involved in regulating NLRP3 inflammasome, inflammasome activation by LPS priming and ATP induction was tested in patients and healthy controls' macrophages. The R157X macrophages showed a significantly increased secretion of IL-1β (>25 times compared to the control magrophages). H67R showed a slightly reduced IL-1β secretion. Overactivated NLRP3 inflammasome activation might explain the life-threatening postoperative complications of R157X carriers.

This study illustrates how mutations in the same TF might have very different phenotypical outcomes. TF activity may be disturbed in its activation (I553M), nuclear localisation (H67R) or expression (R157X), all of which have different molecular mechanisms and cause different immune-related symptoms.

Before this study, not many NFKB1-related PIDs were known. NFKB1 was reported to cause CVIDs due to exon skipping or truncation (Fliegauf et al., 2015). However, the missense mutations covered in this study did not result in haploinsufficiency, and the truncating R157X resulted in an almost total loss of NFKB1. This study provides valuable information related to mechanisms other than haploinsufficiency. At present, NFKB1 mutations are reported to be one of the most common reason for PIDs, most of them being caused by haploinsufficiency (Tuijnenburg et al., 2018).

V CONCLUSION AND FUTURE PERSPECTIVES

Human TF interactions with other proteins are central in gene expression regulation. Any disturbance in these interactions may result in a pathological condition. The overall aim of this thesis was to shed light on TF proteomes and examine how certain TF mutations result in clinical phenotypes of immunodeficiency patients.

While TF binding to DNA is largely studied, there is still a lack of large systems levels analysis of TF protein interactions. Study I provides so far the most comprehensive systems-level analysis of human TF identifying a large cohort of PPIs, serving as excellent resource of TF PPIs for further studies and development of pharmaceutical treatment for TF-related diseases. It also provides an important overview of how TFs generally interact with other proteins. This overview correlated well with the previously reported TF PPI groups that are presented in Chapter 2. This indicates that the used assays and the filtering of unspecific interactions functioned properly.

As the most common PPIs interacting with TFs were histone modifying enzymes (lysine-specific demethylase and histone-lysine transferase), highlights the study I the importance of histone modification in TF-regulated gene expression. Besides this, TFs clearly take part in regulating other cellular processes, such as DNA repair, DNA replication and DNA recombination, both in a transcriptional and non-transcriptional manner. However, these processes appear not to be independent but rather to share multiple proteins and protein complexes. Two studies of yeast Mediator complex protein interactions (Uthe et al., 2017; Chereji et al., 2017) uncovered similar groups of interacting proteins than we identified to studied TFs, including proteins involved in actin assembly, RNA metabolism, chromatin remodelling, GTFs, SAGA complex and RNA splicing. These similar results highlight the the co-transcriptional nature of these processes.

Studies II and III provided more detailed information regarding how single-point mutations in TFs may cause diseases through different mechanisms. TF mutations may affect the DNA binding (II), PPIs (II and III), transcription (II), localisation (III), PTMs (III) or degradation (III) of TFs. Often the outcome of the mutation is a result of multiple interconnected mechanisms. However, these studies illustrate how mutations in the same protein may cause symptoms through different mechanisms (III) or how one mutation may effect multiple levels of transcriptional regulation (II).

Study I, especially, left uncovered topics that should be studied further, such as the possible general role of NFIs in TF-mediated transcriptional regulation, the role of SP7 in RNA splicing and the connections between STAT1 or FOS to actin and myosin signalling. Moreover, the identified large human TF PPI cohort may be used to profile other TFs in the context of more than 100 TFs.

Study II's data-driven multiomics approach uncovered a novel gain-of-function PID mechanism and previously unidentified disease, CAIN. Multiomics approach was also used in study III to investigate the mechanisms behind the three different diseases. Similar methods might be advantageous when studying other TF-related disorders in order to uncover the defects in multilayer transcription regulation by TFs.

This thesis provided a broad overview of human TF PPIs in studied conditions and attempted to fill the lack of a system-level understanding of TF protein interactomes. More systems level TF interactomics studies in different conditions are needed for further understand the complicated PPI

network of human transcription regulation. This knowledge is valuable in developing treatments for TF-related diseases or studying the changes in protein interactions in different disease states.

In addition, this thesis provides detailed information on how certain TF mutations affect different TF functions and lead to the development of PIDs. While study I allowed us to study the TF interactomes in systems level, studies II and III led us focus the TF PPIs and their changes more in specific way. These findings have already facilitated the treatment of patients participating in the studies, but they could also be used to diagnose and investigate diseases, as well as to develop treatments for patients with the same mutations or similar disease mechanisms

VI ACKNOWLEDGEMENTS

This work was carried out in the Systems Biology Research Group and Proteomics Unit at the Institute of Biotechnology, University of Finland. I would like to thank the Proteomics Unit and the Institute for providing me with excellent facilities and instrumentation to complete my work. I am very grateful for the Doctoral Programme in Biomedicine for both financial and educational support during the thesis. This work has also been financially supported by the Research Foundation of the University of Helsinki, the Instrumentarium Science Foundation, the Orion Research Foundation and the Finnish Cultural Foundation.

Thank you to my excellent supervisor Dr. Markku Varjosalo for giving me the opportunity to work in your group. You have introduced me the fascinating world of proteomics. Your support, encouragement, ideas, expertise and enthusiasm throughout this thesis process are highly valued. Thank you also for your kind introduction to the world of Kummeli and to the facial expression “tonnin seteli.” I am sure this expression will help me in my next steps in life.

I sincerely thank Professor Anna-Liisa Levonen for serving as the opponent for this thesis. I would also like to thank the pre-reviewers, Professor Gonghong Wei and Associate Professor Mikael Björklund for their careful reviews and valuable suggestions for improvement. Special thanks to you both for being flexible with the timetable so I could defend my thesis before my maternity leave.

I wish to thank my thesis committee members, Professor Juha Partanen and Associate Professor Ville Hietakangas. Thank you both for your support and ideas during these years. Juha, thanks for being an excellent responsible professor during the thesis process and thank you for serving as custodian at the defense.

The studies included in this thesis would not have been possible without a large network of collaborators. I would like to thank all the researchers and patients who participated in the studies. I especially want to thank Associate Professor Mikko Seppänen, the head of the Rare Disease Center at Helsinki University Hospital, for guiding me to the world of immunodeficiencies. Special thanks to Dr Christopher Fogarty, Dr Sahu Biswajyoti, Meri Kaustio, Dr Emma Haapaniemi, Professor Janna Saarela, Professor Juha Kere, Professor Jussi Taipale and Professor Kari Eklund for the fruitful collaborations. Without you all, the articles about *CEBPE* and *NFKB1* mutations in immunodeficiencies would not have been here.

My warmest thanks to all current and former members of Varjosalo Lab. Thank you Leena, Liu, Salla, Sini, Lisa, Tiina, Kari, Jaakko, Elina, Matias, Rigbae, Antti and Saara for all your support and the fun that we have had during the years we spent together. Leena and Liu, thank you that you for growing up with me in this proteomic world. Your support and help have been very precious to me throughout the completion of this thesis (especially your help in harvesting!). Salla, thank you for all your help, not only in scientific matters but also with kids and construction plans. I am sure that we would also have a future in the building industry! Sini, Tiina, Lisa and Elina, it has been a pleasure to work with you, have fun with you and eat lunch with you (thank you for your great sense of humor, and thanks to Kokkikartano for the food). Sini, thank you for keeping the lab running, even when I did not remember to write the product numbers in the ordering notes. Lisa, thank you for contributing with your interest in statistics and for commenting on the first version of the thesis. Jaakko, Kari, Matias, Antti and Rigbae, it has been great to work with you. Many thanks for all your

help and company during the years. Jaakko, you were really patient in our office, answering all my not always so wise questions. Saara, thank you for supporting me during this thesis, as a college friend, a friend and finally as a workmate.

I would like to thank the staff members of the Institute, especially Pasi, Petri and Tommi. Your work has been crucial in facilitating all the research I performed in the Institute. You were always willing to help me without any delay. Thank you!

I am lucky to have great people as my friends and relatives. I wish to thank all my friends for their friendship, support and encouragement during this thesis process and in different stages of my life. You all are really important to me. I wish to thank my lovely godmother and friend Sailaritta for everything over the years. Thank you also to all the members of the Göös family, especially Irene and Risto, for their support and interest in our lives and work. Special thanks to my own lovely family. I wish to thank you all for all your love and support! Many thanks to my beautiful sister Sara for sharing my life and supporting me during these Helsinki years. Thank you, Mom and Dad, Kirsti and Lauri, for the love, understanding and support you give me throughout my life. Your belief in me during this thesis process has been precious.

Finally, to my beloved husband Tommi and our great son Oiva, you are the base of my life. I love you. My biggest thanks to Tommi for your love, patience, understanding and belief in my thesis project. Oiva, even though you are only 2 years old, you have asked me multiple times when I will have the thesis ready and will I bring one for you. Here we go, this is for you. You are the best! I would also like to thank our dear unborn son, you really made me finish this. See you soon!

Helsinki, November 2019

Helka Göös

VII REFERENCES

Acres, M. J., et al. 2019. Signal transducer and activator of transcription 5B deficiency due to a novel missense mutation in the coiled-coil domain. *J Allergy Clin Immunol*, 143: 413-16.e4.

Allemand, E., et al. 2016. A Broad Set of Chromatin Factors Influences Splicing. *PLoS Genet*, 12: e1006318.

Allen, B. L., and D. J. Taatjes. 2015. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol*, 16: 155-66.

Allfrey, V. G., R. Faulkner, and A. E. Mirsky. 1964. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proc Natl Acad Sci U S A*, 51: 786-94.

Amoutzias, G. D., et al. 2008. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem Sci*, 33: 220-9.

An, W. 2007. Histone acetylation and methylation: combinatorial players for transcriptional regulation. *Subcell Biochem*, 41: 351-69.

Ang, Y. S., et al. 2016. Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell*, 167: 1734-49.e22.

Babu, M. M., et al. 2004. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol*, 14: 283-91.

Baker, C. M., and G. H. Grant. 2007. Role of aromatic amino acids in protein-nucleic acid recognition. *Biopolymers*, 85: 456-70.

Ballow, M., et al. 2009. Immunodeficiencies. *Clin Exp Immunol*, 158 Suppl 1: 14-22.

Baptista, T., et al. 2017. SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Mol Cell*, 68: 130-43.e5.

Barbieri, C. E., et al. 2012. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet*, 44: 685-9.

Barozzi, I., et al. 2014. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol Cell*, 54: 844-57.

Barrios-Rodiles, M., et al. 2005. High-throughput mapping of a dynamic signaling network in mammalian cells. *Science*, 307: 1621-5.

Bastien, J., et al. 2000. TFIIH interacts with the retinoic acid receptor gamma and phosphorylates its AF-1-activating domain through cdk7. *J Biol Chem*, 275: 21896-904.

Bauer, A. J., and K. A. Martin. 2017. Coordinating Regulation of Gene Expression in Cardiovascular Disease: Interactions between Chromatin Modifiers and Transcription Factors. *Front Cardiovasc Med*, 4: 19.

Beall, E. L., et al. 2002. Role for a Drosophila Myb-containing protein complex in site-specific DNA replication. *Nature*, 420: 833-7.

Bedi, R., et al. 2009. Human C/EBP-epsilon activator and repressor isoforms differentially reprogram myeloid lineage commitment and differentiation. *Blood*, 113: 317-27.

Beishline, K., et al. 2012. Sp1 facilitates DNA double-strand break repair through a nontranscriptional mechanism. *Mol Cell Biol*, 32: 3790-9.

Belsky, J. A., et al. 2015. Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. *Genes Dev*, 29: 212-24.

Bernasconi, A., et al. 2006. Characterization of immunodeficiency in a patient with growth hormone insensitivity secondary to a novel STAT5b gene mutation. *Pediatrics*, 118: e1584-92.

Bertoli, C., J. M. Skotheim, and R. A. de Bruin. 2013. Control of cell cycle transcription during G1 and S phases. *Nat Rev Mol Cell Biol*, 14: 518-28.

Bhoumik, A., et al. 2005. ATM-dependent phosphorylation of ATF2 is required for the DNA damage response. *Mol Cell*, 18: 577-87.

Bielli, P., et al. 2014. The transcription factor FBI-1 inhibits SAM68-mediated BCL-X alternative splicing and apoptosis. *EMBO Rep*, 15: 419-27.

Biswas, A. K., and D. G. Johnson. 2012. Transcriptional and nontranscriptional functions of E2F1 in response to DNA damage. *Cancer Res*, 72: 13-7.

Boija, A., et al. 2018. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*, 175: 1842-55.e16.

Bonnet, J., et al. 2014. The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes Dev*, 28: 1999-2012.

Borggrefe, T., and X. Yue. 2011. Interactions between subunits of the Mediator complex with gene-specific transcription factors. *Semin Cell Dev Biol*, 22: 759-68.

Bornstein, C., et al. 2014. A negative feedback loop of transcription factors specifies alternative dendritic cell chromatin States. *Mol Cell*, 56: 749-62.

Bosco, G., W. Du, and T. L. Orr-Weaver. 2001. DNA replication control through interaction of E2F-RB and the origin recognition complex. *Nat Cell Biol*, 3: 289-95.

Bousfiha, A., et al. 2018. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. *J Clin Immunol*, 38: 129-43.

Boutet, S. C., et al. 2010. Taf1 regulates Pax3 protein by monoubiquitination in skeletal muscle progenitors. *Mol Cell*, 40: 749-61.

Boztug, H., et al. 2016. NF-kappaB1 Haploinsufficiency Causing Immunodeficiency and EBV-Driven Lymphoproliferation. *J Clin Immunol*, 36: 533-40.

Brown, A. K., and A. E. Webb. 2018. Regulation of FOXO Factors in Mammalian Cells. *Curr Top Dev Biol*, 127: 165-92.

Campbell, C. E., et al. 2008. The transcription factor Nfix is essential for normal brain development. *BMC Dev Biol*, 8: 52.

Chang, I. F. 2006. Mass spectrometry-based proteomic analysis of the epitope-tag affinity purified protein complexes in eukaryotes. *Proteomics*, 6: 6158-66.

Chapgier, A., et al. 2009. A partial form of recessive STAT1 deficiency in humans. *J Clin Invest*, 119: 1502-14.

Chen, D., et al. 2000. Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIIH and participation of CDK7. *Mol Cell*, 6: 127-37.

Chen, K., et al. 2013. Germline mutations in NFKB2 implicate the noncanonical NF-kappaB pathway in the pathogenesis of common variable immunodeficiency. *Am J Hum Genet*, 93: 812-24.

Chereji, R. V., et al. 2017. Mediator binds to boundaries of chromosomal interaction domains and to proteins involved in DNA looping, RNA metabolism, chromatin remodeling, and actin assembly. *Nucleic Acids Res*, 45: 8806-21.

Choul-Li, S., et al. 2018. Ets-1 interacts through a similar binding interface with Ku70 and Poly (ADP-Ribose) Polymerase-1. *Biosci Biotechnol Biochem*, 82: 1753-59.

Christmann, M., and B. Kaina. 2013. Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation. *Nucleic Acids Res*, 41: 8403-20.

Chuang, J. Y., et al. 2012. Sp1 phosphorylation by cyclin-dependent kinase 1/cyclin B1 represses its DNA-binding activity during mitosis in cancer cells. *Oncogene*, 31: 4946-59.

Ciccia, A., and S. J. Elledge. 2010. The DNA damage response: making it safe to play with knives. *Mol Cell*, 40: 179-204.

Clapier, C. R., and B. R. Cairns. 2009. The biology of chromatin remodeling complexes. *Annu Rev Biochem*, 78: 273-304.

Clapier, C. R., et al. 2017. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol*, 18: 407-22.

Cortez, D. 2017. Proteomic Analyses of the Eukaryotic Replication Machinery. *Methods Enzymol*, 591: 33-53.

Currie, S. L., et al. 2017. ETV4 and AP1 Transcription Factors Form Multivalent Interactions with three Sites on the MED25 Activator-Interacting Domain. *J Mol Biol*, 429: 2975-95.

Cvekl, A., et al. 1999. Pax-6 interactions with TATA-box-binding protein and retinoblastoma protein. *Invest Ophthalmol Vis Sci*, 40: 1343-50.

Davis, R. L., H. Weintraub, and A. B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, 51: 987-1000.

de Jonge, W. J., et al. 2017. Molecular mechanisms that distinguish TFIID housekeeping from regulatable SAGA promoters. *Embo j*, 36: 274-90.

de Lanerolle, P. 2012. Nuclear actin and myosins at a glance. *J Cell Sci*, 125: 4945-9.

Decker, T., and P. Kovarik. 2000. Serine phosphorylation of STATs. *Oncogene*, 19: 2628-37.

Deplancke, B., D. Alpern, and V. Gardeux. 2016. The Genetics of Transcription Factor DNA Binding Variation. *Cell*, 166: 538-54.

Ding, N., et al. 2008. Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Mol Cell*, 31: 347-59.

Ding, Q., and D. M. MacAlpine. 2011. Defining the replication program through the chromatin landscape. *Crit Rev Biochem Mol Biol*, 46: 165-79.

Dixon, A. S., et al. 2016. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem Biol*, 11: 400-8.

Dominguez-Sola, D., et al. 2007. Non-transcriptional control of DNA replication by c-Myc. *Nature*, 448: 445-51.

Driller, K., et al. 2007. Nuclear factor I X deficiency causes brain malformation and severe skeletal defects. *Mol Cell Biol*, 27: 3855-67.

Dunham, W. H., M. Mullin, and A. C. Gingras. 2012. Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics*, 12: 1576-90.

Dupuis, S., et al. 2001. Impairment of mycobacterial but not viral immunity by a germline human STAT1 mutation. *Science*, 293: 300-3.

Dupuis, S., et al. 2003. Impaired response to interferon-alpha/beta and lethal viral disease in human STAT1 deficiency. *Nat Genet*, 33: 388-91.

Dvir, A., R. C. Conaway, and J. W. Conaway. 1997. A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. *Proc Natl Acad Sci U S A*, 94: 9006-10.

Ehe, B. K., et al. 2017. Identification of a DYRK1A-mediated phosphorylation site within the nuclear localization sequence of the hedgehog transcription factor GLI1. *Biochem Biophys Res Commun*, 491: 767-72.

Fan, J., et al. 2015. Metabolic regulation of histone post-translational modifications. *ACS Chem Biol*, 10: 95-108.

Favot, L., et al. 2005. Cytoplasmic YY1 is associated with increased smooth muscle-specific gene expression: implications for neonatal pulmonary hypertension. *Am J Pathol*, 167: 1497-509.

Fields, S., and O. Song. 1989. A novel genetic system to detect protein-protein interactions. *Nature*, 340: 245-6.

Filtz, T. M., W. K. Vogel, and M. Leid. 2014. Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharmacol Sci*, 35: 76-85.

Fischer, V., et al. 2019. Global role for coactivator complexes in RNA polymerase II transcription. *Transcription*, 10: 29-36.

Flanagan, S. E., et al. 2014. Activating germline mutations in STAT3 cause early-onset multi-organ autoimmune disease. *Nat Genet*, 46: 812-14.

Fliegauf, M., et al. 2015. Haploinsufficiency of the NF-kappaB1 Subunit p50 in Common Variable Immunodeficiency. *Am J Hum Genet*, 97: 389-403.

Fontaine, F., J. Overman, and M. Francois. 2015. Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and obstacles. *Cell Regen (Lond)*, 4: 2.

Fredriksson, S., et al. 2002. Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol*, 20: 473-7.

Frei, A. P., et al. 2013. Ligand-based receptor identification on living cells and tissues using TRICEPS. *Nat Protoc*, 8: 1321-36.

Frietze, S., and P. J. Farnham. 2011. Transcription factor effector domains. *Subcell Biochem*, 52: 261-77.

Fulton, D. L., et al. 2009. TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol*, 10: R29.

Gamper, A. M., and R. G. Roeder. 2008. Multivalent binding of p53 to the STAGA complex mediates coactivator recruitment after UV damage. *Mol Cell Biol*, 28: 2517-27.

Geertz, M., D. Shore, and S. J. Maerkl. 2012. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc Natl Acad Sci U S A*, 109: 16540-5.

Girardot, M., et al. 2018. SOX9 has distinct regulatory roles in alternative splicing and transcription. *Nucleic Acids Res*, 46: 9106-18.

Grimes, D. A., et al. 2006. Translated mutation in the Nurr1 gene as a cause for Parkinson's disease. *Mov Disord*, 21: 906-9.

Grummt, I. 2006. Actin and myosin as transcription factors. *Curr Opin Genet Dev*, 16: 191-6.

Haller, K., et al. 2004. Subcellular localization of multiple PREP2 isoforms is regulated by actin, tubulin, and nuclear export. *J Biol Chem*, 279: 49384-94.

Hallier, M., et al. 1998. The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. *J Biol Chem*, 273: 4838-42.

Hallier, M., A. Tavitian, and F. Moreau-Gachelin. 1996. The transcription factor Spi-1/PU.1 binds RNA and interferes with the RNA-binding protein p54nrb. *J Biol Chem*, 271: 11177-81.

Hambleton, S., et al. 2013. STAT2 deficiency and susceptibility to viral illness in humans. *Proc Natl Acad Sci U S A*, 110: 3053-8.

Han, J., and R. J. Kaufman. 2017. Physiological/pathological ramifications of transcription factors in the unfolded protein response. *Genes Dev*, 31: 1417-38.

Han, Y., et al. 2014. Architecture of the Saccharomyces cerevisiae SAGA transcription coactivator complex. *Embo j*, 33: 2534-46.

Harper, T. M., and D. J. Taatjes. 2018. The complex structure and function of Mediator. *J Biol Chem*, 293: 13778-85.

Harris, S. L., and A. J. Levine. 2005. The p53 pathway: positive and negative feedback loops. *Oncogene*, 24: 2899-908.

Hashimoto, S., et al. 2011. MED23 mutation links intellectual disability to dysregulation of immediate early gene expression. *Science*, 333: 1161-3.

Haubenberger, D., et al. 2011. Association of transcription factor polymorphisms PITX3 and EN1 with Parkinson's disease. *Neurobiol Aging*, 32: 302-7.

Heikkinen, T., et al. 2017. Somatic MED12 Nonsense Mutation Escapes mRNA Decay and Reveals a Motif Required for Nuclear Entry. *Hum Mutat*, 38: 269-74.

Heinz, S., et al. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, 38: 576-89.

Hellqvist, M., et al. 1998. The human forkhead protein FREAC-2 contains two functionally redundant activation domains and interacts with TBP and TFIIB. *J Biol Chem*, 273: 23335-43.

Helmlinger, D., and L. Tora. 2017. Sharing the SAGA. *Trends Biochem Sci*, 42: 850-61.

Hieb, A. R., et al. 2007. TFIIA changes the conformation of the DNA in TBP/TATA complexes and increases their kinetic stability. *J Mol Biol*, 372: 619-32.

Holland, S. M., et al. 2007. STAT3 mutations in the hyper-IgE syndrome. *N Engl J Med*, 357: 1608-19.

Holmlund, T., et al. 2013. GCN5 acetylates and regulates the stability of the oncoprotein E2A-PBX1 in acute lymphoblastic leukemia. *Leukemia*, 27: 578-85.

Holstege, F. C., P. C. van der Vliet, and H. T. Timmers. 1996. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *Embo j*, 15: 1666-77.

Huisinga, K. L., and B. F. Pugh. 2004. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Mol Cell*, 13: 573-85.

Hwa, V., et al. 2005. Severe growth hormone insensitivity resulting from total absence of signal transducer and activator of transcription 5b. *J Clin Endocrinol Metab*, 90: 4260-6.

Inui, M., et al. 2011. USP15 is a deubiquitylating enzyme for receptor-activated SMADs. *Nat Cell Biol*, 13: 1368-75.

Inukai, S., K. H. Kock, and M. L. Bulyk. 2017. Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev*, 43: 110-19.

Iwafuchi-Doi, M. 2019. The mechanistic basis for chromatin regulation by pioneer transcription factors. *Wiley Interdiscip Rev Syst Biol Med*, 11: e1427.

Iwafuchi-Doi, M., and K. S. Zaret. 2014. Pioneer transcription factors in cell reprogramming. *Genes Dev*, 28: 2679-92.

Izhar, L., et al. 2015. A Systematic Analysis of Factors Localized to Damaged Chromatin Reveals PARP-Dependent Recruitment of Transcription Factors. *Cell Rep*, 11: 1486-500.

Jiang, J., and M. Levine. 1993. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell*, 72: 741-52.

Jolma, A., and J. Taipale. 2011. Methods for Analysis of Transcription Factor DNA-Binding Specificity In Vitro. *Subcell Biochem*, 52: 155-73.

Jolma, A., et al. 2013. DNA-binding specificities of human transcription factors. *Cell*, 152: 327-39.

Jolma, A., et al. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, 527: 384-8.

Joo, Y. J., et al. 2017. Downstream promoter interactions of TFIID TAFs facilitate transcription reinitiation. *Genes Dev*, 31: 2162-74.

Jozwik, K. M., and J. S. Carroll. 2012. Pioneer factors in hormone-dependent cancers. *Nat Rev Cancer*, 12: 381-5.

Kahata, K., et al. 2004. Regulation of transforming growth factor-beta and bone morphogenetic protein signalling by transcriptional coactivator GCN5. *Genes Cells*, 9: 143-51.

Kandoth, C., et al. 2013. Mutational landscape and significance across 12 major cancer types. *Nature*, 502: 333-39.

Kaustio, M., et al. 2017. Damaging heterozygous mutations in NFKB1 lead to diverse immunologic phenotypes. *J Allergy Clin Immunol*, 140: 782-96.

Ke, J., et al. 2018. Inhibitor of DNA binding proteins: implications in human cancer progression and metastasis. *Am J Transl Res*, 10: 3887-910.

Kenworthy, A. K. 2001. Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. *Methods*, 24: 289-96.

Kerppola, T. K. 2008. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annu Rev Biophys*, 37: 465-87.

Khan, A., et al. 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res*, 46: D260-d66.

Khanna-Gupta, A., et al. 2007. Growth factor independence-1 (Gfi-1) plays a role in mediating specific granule deficiency (SGD) in a patient lacking a gene-inactivating mutation in the C/EBPepsilon gene. *Blood*, 109: 4181-90.

Kim, D. I., et al. 2016. An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell*, 27: 1188-96.

Klammt, J., et al. 2018. Dominant-negative STAT5B mutations cause growth hormone insensitivity with short stature and mild immune dysregulation. *Nat Commun*, 9: 2105.

Klemann, C., et al. 2019. Clinical and Immunological Phenotype of Patients With Primary Immunodeficiency Due to Damaging Mutations in NFKB2. *Front Immunol*, 10: 297.

Ko, L. J., et al. 1997. p53 is phosphorylated by CDK7-cyclin H in a p36MAT1-dependent manner. *Mol Cell Biol*, 17: 7220-9.

Kofoed, E. M., et al. 2003. Growth hormone insensitivity associated with a STAT5b mutation. *N Engl J Med*, 349: 1139-47.

Kohli, S., S. Ahuja, and V. Rani. 2011. Transcription factors in heart: promising therapeutic targets in cardiac hypertrophy. *Curr Cardiol Rev*, 7: 262-71.

Kuehn, H. S., et al. 2017. Novel nonsense gain-of-function NFKB2 mutations associated with a combined immunodeficiency phenotype. *Blood*, 130: 1553-64.

Kugel, J. F., and J. A. Goodrich. 1998. Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA. *Proc Natl Acad Sci U S A*, 95: 9232-7.

Kumar, K. P., S. Akoulitchev, and D. Reinberg. 1998. Promoter-proximal stalling results from the inability to recruit transcription factor IIH to the transcription complex and is a regulated event. *Proc Natl Acad Sci U S A*, 95: 9767-72.

Kuras, L., et al. 2000. TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo. *Science*, 288: 1244-8.

Laflamme, C., C. Filion, and Y. Labelle. 2004. Functional characterization of SIX3 homeodomain mutations in holoprosencephaly: interaction with the nuclear receptor NR4A3/NOR1. *Hum Mutat*, 24: 502-8.

Lambert, J. P., et al. 2015. Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J Proteomics*, 118: 81-94.

Lambert, Samuel A., et al. 2018. The Human Transcription Factors. *Cell*, 172: 650-65.

Landre, V., et al. 2017. Regulation of transcriptional activators by DNA-binding domain ubiquitination. *Cell Death Differ*, 24: 903-16.

Lang, S. E., et al. 2001. E2F transcriptional activation requires TRRAP and GCN5 cofactors. *J Biol Chem*, 276: 32627-34.

Langst, G., and L. Manelyte. 2015. Chromatin Remodelers: From Function to Dysfunction. *Genes (Basel)*, 6: 299-324.

Le, W. D., et al. 2003. Mutations in NR4A2 associated with familial Parkinson disease. *Nat Genet*, 33: 85-9.

Lee, C. 2007. Coimmunoprecipitation assay. *Methods Mol Biol*, 362: 401-6.

Lee, C. E., et al. 2014. Autosomal-dominant B-cell deficiency with alopecia due to a mutation in NFKB2 that results in nonprocessable p100. *Blood*, 124: 2964-72.

Lee, D. K., H. O. Duan, and C. Chang. 2000. From androgen receptor to the general transcription factor TFIIH. Identification of cdk activating kinase (CAK) as an androgen receptor NH(2)-terminal associated coactivator. *J Biol Chem*, 275: 9308-13.

Lee, T. I., et al. 2000. Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature*, 405: 701-4.

Lee, T. I., and R. A. Young. 2013. Transcriptional regulation and its misregulation in disease. *Cell*, 152: 1237-51.

Legrand, A. J., et al. 2013. The level of Ets-1 protein is regulated by poly(ADP-ribose) polymerase-1 (PARP-1) in cancer cells to prevent DNA damage. *PLoS One*, 8: e55883.

Lekstrom-Himes, J. A., et al. 1999. Neutrophil-specific granule deficiency results from a novel mutation with loss of function of the transcription factor CCAAT/enhancer binding protein epsilon. *J Exp Med*, 189: 1847-52.

Lekstrom-Himes, J., and K. G. Xanthopoulos. 1998. Biological role of the CCAAT/enhancer-binding protein family of transcription factors. *J Biol Chem*, 273: 28545-8.

Li, H. H., et al. 2004. Phosphorylation on Thr-55 by TAF1 mediates degradation of p53: a role for TAF1 in cell G1 progression. *Mol Cell*, 13: 867-78.

Li, M., et al. 2015. Dynamic regulation of transcription factors by nucleosome remodeling. *Elife*, 4.

Li, X., et al. 2015. Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Mol Syst Biol*, 11: 775.

Li, Y., et al. 2018. Structural and ligand-binding analysis of the YAP-binding domain of transcription factor TEAD4. *Biochem J*, 475: 2043-55.

Lickwar, C. R., et al. 2012. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature*, 484: 251-5.

Lievens, S., et al. 2014. Kinase Substrate Sensor (KISS), a mammalian in situ protein interaction sensor. *Mol Cell Proteomics*, 13: 3332-42.

Lindsley, A. W., et al. 2014. Combined immune deficiency in a patient with a novel NFKB2 mutation. *J Clin Immunol*, 34: 910-5.

Ling, F., B. Kang, and X. H. Sun. 2014. Id proteins: small molecules, mighty regulators. *Curr Top Dev Biol*, 110: 189-216.

Liu, B., et al. 2007. Label transfer chemistry for the characterization of protein-protein interactions. *J Am Chem Soc*, 129: 12348-9.

Liu, L., et al. 2011. Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. *J Exp Med*, 208: 1635-48.

Liu, X., et al. 2018. An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat Commun*, 9: 1188.

Liu, X., et al. 2003. c-Myc transformation domain recruits the human STAGA complex and requires TRRAP and GCN5 acetylase activity for transcription activation. *J Biol Chem*, 278: 20405-12.

Lively, T. N., et al. 2001. c-Jun binds the N terminus of human TAF(II)250 to derepress RNA polymerase II transcription in vitro. *J Biol Chem*, 276: 25582-8.

Lorenzini, T., et al. 2017. STAT mutations as program switchers: turning primary immunodeficiencies into autoimmune diseases. *J Leukoc Biol*, 101: 29-38.

Lorenzo, P. I., et al. 2017. The Diabetes-Linked Transcription Factor PAX4: From Gene to Functional Consequences. *Genes (Basel)*, 8.

Lukas, J., C. Lukas, and J. Bartek. 2011. More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nat Cell Biol*, 13: 1161-9.

Maestro, M. A., et al. 2007. Distinct roles of HNF1beta, HNF1alpha, and HNF4alpha in regulating pancreas development, beta-cell function and growth. *Endocr Dev*, 12: 33-45.

Makinen, N., et al. 2011. MED12 exon 2 mutations are common in uterine leiomyomas from South African patients. *Oncotarget*, 2: 966-9.

Makinen, N., et al. 2011. MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science*, 334: 252-5.

Malecki, M. T. 2005. Genetics of type 2 diabetes mellitus. *Diabetes Res Clin Pract*, 68 Suppl1: S10-21.

Malewicz, M., et al. 2011. Essential role for DNA-PK-mediated phosphorylation of NR4A nuclear orphan receptors in DNA double-strand break repair. *Genes Dev*, 25: 2031-40.

Malewicz, M., and T. Perlmann. 2014. Function of transcription factors at DNA lesions in DNA repair. *Exp Cell Res*, 329: 94-100.

Martens, J. H., and H. G. Stunnenberg. 2010. The molecular signature of oncofusion proteins in acute myeloid leukemia. *FEBS Lett*, 584: 2662-9.

Martin-Martin, N., A. Carracedo, and V. Torrano. 2017. Metabolism and Transcription in Cancer: Merging Two Classic Tales. *Front Cell Dev Biol*, 5: 119.

Martinez-Martinez, L., et al. 2015. A novel gain-of-function STAT1 mutation resulting in basal phosphorylation of STAT1 and increased distal IFN-gamma-mediated responses in chronic mucocutaneous candidiasis. *Mol Immunol*, 68: 597-605.

Mason, S., et al. 2009. Nuclear factor one transcription factors in CNS development. *Mol Neurobiol*, 39: 10-23.

Matys, V., et al. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*, 34: D108-10.

McCulley, D. J., and B. L. Black. 2012. Transcription factor pathways and congenital heart disease. *Curr Top Dev Biol*, 100: 253-77.

McEwan, I. J., et al. 1996. Functional interaction of the c-Myc transactivation domain with the TATA binding protein: evidence for an induced fit model of transactivation domain folding. *Biochemistry*, 35: 9584-93.

McEwan, I. J., and J. Gustafsson. 1997. Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF. *Proc Natl Acad Sci U S A*, 94: 8485-90.

Mercer, E. M., Y. C. Lin, and C. Murre. 2011. Factors and networks that underpin early hematopoiesis. *Semin Immunol*, 23: 317-25.

Mertens, F., C. R. Antonescu, and F. Mitelman. 2016. Gene fusions in soft tissue tumors: Recurrent and overlapping pathogenetic themes. *Genes Chromosomes Cancer*, 55: 291-310.

Metz, R., et al. 1994. c-Fos-induced activation of a TATA-box-containing promoter involves direct contact with TATA-box-binding protein. *Mol Cell Biol*, 14: 6021-9.

Meyer, K. D., et al. 2010. p53 activates transcription by directing structural shifts in Mediator. *Nat Struct Mol Biol*, 17: 753-60.

Milner, J. D., et al. 2015. Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations. *Blood*, 125: 591-9.

Minegishi, Y., et al. 2007. Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature*, 448: 1058-62.

Miralles, F., et al. 2003. Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell*, 113: 329-42.

Moens, L., et al. 2017. A novel kindred with inherited STAT2 deficiency and severe viral illness. *J Allergy Clin Immunol*, 139: 1995-97.e9.

Mogensen, T. H. 2018. IRF and STAT Transcription Factors - From Basic Biology to Roles in Infection, Protective Immunity, and Primary Immunodeficiencies. *Front Immunol*, 9: 3047.

Mohibullah, N., and S. Hahn. 2008. Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev*, 22: 2994-3006.

Moorthy, A. K., et al. 2006. The 20S proteasome processes NF-kappaB1 p105 into p50 in a translation-independent manner. *Embo j*, 25: 1945-56.

Morgunova, E., and J. Taipale. 2017. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol*, 47: 1-8.

Moyle-Heyrman, G., H. S. Tims, and J. Widom. 2011. Structural constraints in collaborative competition of transcription factors against the nucleosome. *J Mol Biol*, 412: 634-46.

Muratani, M., and W. P. Tansey. 2003. How the ubiquitin-proteasome system controls transcription. *Nat Rev Mol Cell Biol*, 4: 192-201.

Murros, J., and A. Konttinen. 1974. Recurrent attacks of abdominal pain and fever with familial segmentation arrest of granulocytes. *Blood*, 43: 871-4.

Nadeau, K., V. Hwa, and R. G. Rosenfeld. 2011. STAT5b deficiency: an unsuspected cause of growth failure, immunodeficiency, and severe pulmonary disease. *J Pediatr*, 158: 701-8.

Nardozzi, J. D., K. Lott, and G. Cingolani. 2010. Phosphorylation meets nuclear import: a review. *Cell Commun Signal*, 8: 32.

Ndoja, A., R. E. Cohen, and T. Yao. 2014. Ubiquitin signals proteolysis-independent stripping of transcription factors. *Mol Cell*, 53: 893-903.

Neiswinger, J., et al. 2016. Protein Microarrays: Flexible Tools for Scientific Innovation. *Cold Spring Harb Protoc*, 2016.

Nishi, H., et al. 2013. Cancer missense mutations alter binding properties of proteins and their interaction networks. *PLoS One*, 8: e66273.

Niwa, H. 2018. The principles that govern transcription factor network functions in stem cells. *Development*, 145.

O'Neill, B. T., et al. 2019. FoxO Transcription Factors Are Critical Regulators of Diabetes-Related Muscle Atrophy. *Diabetes*, 68: 556-70.

O'Rawe, J. A., et al. 2015. TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations. *Am J Hum Genet*, 97: 922-32.

Oeckinghaus, A., and S. Ghosh. 2009. The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harb Perspect Biol*, 1: a000034.

Ohe, K., E. Lalli, and P. Sassone-Corsi. 2002. A direct role of SRY and SOX proteins in pre-mRNA splicing. *Proc Natl Acad Sci U S A*, 99: 1146-51.

Pan, G., et al. 2006. A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *Faseb j*, 20: 1730-2.

Panne, D. 2008. The enhanceosome. *Curr Opin Struct Biol*, 18: 236-42.

Pasanen, A. V., et al. 1987. Impaired heme synthesis in a family with Pelger-Huet anomaly, recurrent abdominal pain attacks and impaired neutrophil motility in vitro. *Eur J Haematol*, 39: 274-7.

Perniola, R. 2018. Twenty Years of AIRE. *Front Immunol*, 9: 98.

Petrenko, N., et al. 2019. Requirements for RNA polymerase II preinitiation complex formation in vivo. *Elife*, 8.

Petschnigg, J., et al. 2014. The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. *Nat Methods*, 11: 585-92.

Pfleger, K. D., and K. A. Eidne. 2006. Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat Methods*, 3: 165-74.

Philimonenko, V. V., et al. 2004. Nuclear actin and myosin I are required for RNA polymerase I transcription. *Nat Cell Biol*, 6: 1165-72.

Picard, C., et al. 2018. International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity. *J Clin Immunol*, 38: 96-128.

Piper, M., R. Gronostajski, and G. Messina. 2019. Nuclear Factor One X in Development and Disease. *Trends Cell Biol*, 29: 20-30.

Ponugoti, B., G. Dong, and D. T. Graves. 2012. Role of forkhead transcription factors in diabetes-induced oxidative stress. *Exp Diabetes Res*, 2012: 939751.

Poss, Z. C., C. C. Ebmeier, and D. J. Taatjes. 2013. The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol*, 48: 575-608.

Prathapam, T., et al. 2006. Activated Src abrogates the Myc requirement for the G0/G1 transition but not for the G1/S transition. *Proc Natl Acad Sci U S A*, 103: 2695-700.

Qadri, I., M. Iwahashi, and F. Simon. 2002. Hepatitis C virus NS5A protein binds TBP and p53, inhibiting their DNA binding and p53 interactions with TBP and ERCC3. *Biochim Biophys Acta*, 1592: 193-204.

Qian, Z., Y. D. Cai, and Y. Li. 2006. Automatic transcription factor classifier based on functional domain composition. *Biochem Biophys Res Commun*, 347: 141-4.

Rambout, X., F. Dequiedt, and L. E. Maquat. 2018. Beyond Transcription: Roles of Transcription Factors in Pre-mRNA Splicing. *Chem Rev*, 118: 4339-64.

Raschella, G., G. Melino, and M. Malewicz. 2017. New factors in mammalian DNA repair-the chromatin connection. *Oncogene*, 36: 4673-81.

Reiter, F., S. Wienerroither, and A. Stark. 2017. Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev*, 43: 73-81.

Renner, E. D., et al. 2008. Novel signal transducer and activator of transcription 3 (STAT3) mutations, reduced T(H)17 cell numbers, and variably defective STAT3 phosphorylation in hyper-IgE syndrome. *J Allergy Clin Immunol*, 122: 181-7.

Repo, H., et al. 1979. Impaired neutrophil chemotaxis in Pelger-Huet anomaly. *Clin Exp Immunol*, 36: 326-33.

Rhee, H. W., et al. 2013. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, 339: 1328-31.

Rizzino, A., and E. L. Wuebben. 2016. Sox2/Oct4: A delicately balanced partnership in pluripotent stem cells and embryogenesis. *Biochim Biophys Acta*, 1859: 780-91.

Rochette-Egly, C. 2003. Nuclear receptors: integration of multiple signalling pathways through phosphorylation. *Cell Signal*, 15: 355-66.

Roux, K. J., et al. 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol*, 196: 801-10.

Saarimaki-Vire, J., et al. 2017. An Activating STAT3 Mutation Causes Neonatal Diabetes through Premature Induction of Pancreatic Differentiation. *Cell Rep*, 19: 281-94.

Samee, M. A. H., B. G. Bruneau, and K. S. Pollard. 2019. A De Novo Shape Motif Discovery Algorithm Reveals Preferences of Transcription Factors for DNA Shape Beyond Sequence Motifs. *Cell Syst*, 8: 27-42.e6.

Scherer, D. C., et al. 1995. Signal-induced degradation of I kappa B alpha requires site-specific ubiquitination. *Proc Natl Acad Sci U S A*, 92: 11259-63.

Schilbach, S., et al. 2017. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature*, 551: 204-09.
Schott, J. J., et al. 1998. Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science*, 281: 108-11.
Sermwittayawong, D., and S. Tan. 2006. SAGA binds TBP via its Spt8 subunit in competition with DNA: implications for TBP recruitment. *Embo j*, 25: 3791-800.
Sever, R., and C. K. Glass. 2013. Signaling by nuclear receptors. *Cold Spring Harb Perspect Biol*, 5: a016709.
Shahni, R., et al. 2015. Signal transducer and activator of transcription 2 deficiency is a novel disorder of mitochondrial fission. *Brain*, 138: 2834-46.
Sharma, S., et al. 2011. Dephosphorylation of the nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA-protein scaffold complex. *Proc Natl Acad Sci U S A*, 108: 11381-6.
Shigemura, T., et al. 2014. Clinical course in a patient with neutrophil-specific granule deficiency and rapid detection of neutrophil granules as a screening test. *J Clin Immunol*, 34: 780-3.
Shu, T., et al. 2003. Abnormal development of forebrain midline glia and commissural projections in Nfia knock-out mice. *J Neurosci*, 23: 203-12.
Sizemore, G. M., et al. 2017. The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer*, 17: 337-51.
Smale, S. T. 2014. Transcriptional regulation in the immune system: a status report. *Trends Immunol*, 35: 190-4.
Spitz, F., and E. E. Furlong. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*, 13: 613-26.
Srivastava, S., et al. 2019. Dysregulations of sonic hedgehog signaling in MED12-related X-linked intellectual disability disorders. *Mol Genet Genomic Med*, 7: e00569.
Stagljar, I., et al. 1998. A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. *Proc Natl Acad Sci U S A*, 95: 5187-92.
Steele-Perkins, G., et al. 2003. Essential role for NFI-C/CTF transcription-replication factor in tooth root development. *Mol Cell Biol*, 23: 1075-84.
Steele-Perkins, G., et al. 2005. The transcription factor gene Nfib is essential for both lung maturation and brain development. *Mol Cell Biol*, 25: 685-98.
Stray-Pedersen, A., et al. 2017. Primary immunodeficiency diseases: Genomic approaches delineate heterogeneous Mendelian disorders. *J Allergy Clin Immunol*, 139: 232-45.
Sun, S. C. 2017. The non-canonical NF-kappaB pathway in immunity and inflammation. *Nat Rev Immunol*, 17: 545-58.
Suzuki, T., et al. 2003. Functional interaction of the DNA-binding transcription factor Sp1 through its DNA-binding domain with the histone chaperone TAF-I. *J Biol Chem*, 278: 28758-64.
Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126: 663-76.
Tang, L. Y., et al. 2011. Ablation of Smurf2 reveals an inhibition in TGF-beta signalling through multiple mono-ubiquitination of Smad3. *Embo j*, 30: 4777-89.
Tarassov, K., et al. 2008. An in vivo map of the yeast protein interactome. *Science*, 320: 1465-70.
Thomas, M. C., and C. M. Chiang. 2006. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol*, 41: 105-78.
Thrasher, A. J., and D. A. Williams. 2017. Evolving Gene Therapy in Primary Immunodeficiency. *Mol Ther*, 25: 1132-41.
Todoric, K., et al. 2013. Autoimmunity in immunodeficiency. *Curr Allergy Asthma Rep*, 13: 361-70.
Tompa, P., and M. Fuxreiter. 2008. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci*, 33: 2-8.
Tootle, T. L., and I. Rebay. 2005. Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays*, 27: 285-98.
Totzke, G., et al. 2006. A novel member of the IkappaB family, human IkappaB-zeta, inhibits transactivation of p65 and its DNA binding. *J Biol Chem*, 281: 12645-54.
Toubiana, J., et al. 2016. Heterozygous STAT1 gain-of-function mutations underlie an unexpectedly broad clinical phenotype. *Blood*, 127: 3154-64.
Trepte, P., et al. 2018. LuTHy: a double-readout bioluminescence-based two-hybrid technology for quantitative mapping of protein-protein interactions in mammalian cells. *Mol Syst Biol*, 14: e8071.
Trinkle-Mulcahy, L. 2019. Recent advances in proximity-based labeling methods for interactome mapping. *F1000Res*, 8.
Tsai, W. B., et al. 2008. Functional interaction between FOXO3a and ATM regulates DNA damage response. *Nat Cell Biol*, 10: 460-7.
Tuijnenburg, P., et al. 2018. Loss-of-function nuclear factor kappaB subunit 1 (NFKB1) variants are the most common monogenic cause of common variable immunodeficiency in Europeans. *J Allergy Clin Immunol*, 142: 1285-96.
Turunen, M., et al. 2014. Uterine leiomyoma-linked MED12 mutations disrupt mediator-associated CDK activity. *Cell Rep*, 7: 654-60.
Ulrichts, P., et al. 2009. MAPPIT (mammalian protein-protein interaction trap) analysis of early steps in toll-like receptor signalling. *Methods Mol Biol*, 517: 133-44.
Uthe, H., J. T. Vanselow, and A. Schlosser. 2017. Proteomic Analysis of the Mediator Complex Interactome in Saccharomyces cerevisiae. *Sci Rep*, 7: 43584.
van de Veerdonk, F. L., et al. 2011. STAT1 mutations in autosomal dominant chronic mucocutaneous candidiasis. *N Engl J Med*, 365: 54-61.
Vaquerizas, J. M., et al. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*, 10: 252-63.
Varjosalo, M., and J. Taipale. 2008. Hedgehog: functions and mechanisms. *Genes Dev*, 22: 2454-72.

Varnaite, R., and S. A. MacNeill. 2016. Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. *Proteomics*, 16: 2503-18.

Vasilescu, J., X. Guo, and J. Kast. 2004. Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics*, 4: 3845-54.

Vermeulen, M., et al. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, 131: 58-69.

Vogelstein, B., et al. 2013. Cancer genome landscapes. *Science*, 339: 1546-58.

Wada, T., et al. 2015. A Novel In-Frame Deletion in the Leucine Zipper Domain of C/EBPepsilon Leads to Neutrophil-Specific Granule Deficiency. *J Immunol*, 195: 80-6.

Wang, G. G., C. D. Allis, and P. Chi. 2007a. Chromatin remodeling and cancer, Part I: Covalent histone modifications. *Trends Mol Med*, 13: 363-72.

———. 2007b. Chromatin remodeling and cancer, Part II: ATP-dependent chromatin remodeling. *Trends Mol Med*, 13: 373-80.

Wang, Y., et al. 2018. The role of STAT3 in leading the crosstalk between human cancers and the immune system. *Cancer Lett*, 415: 117-28.

Wassarman, D. A., and F. Sauer. 2001. TAF(II)250: a transcription toolbox. *J Cell Sci*, 114: 2895-902.

Wasserman, W. W., and A. Sandelin. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet*, 5: 276-87.

Whitmarsh, A. J., and R. J. Davis. 2000. Regulation of transcription factor function by phosphorylation. *Cell Mol Life Sci*, 57: 1172-83.

Wingender, E., et al. 2018. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res*, 46: D343-d47.

Wiper-Bergeron, N., et al. 2007. Glucocorticoid-stimulated preadipocyte differentiation is mediated through acetylation of C/EBPbeta by GCN5. *Proc Natl Acad Sci U S A*, 104: 2703-8.

Wojcik, P., and L. Berlicki. 2016. Peptide-based inhibitors of protein-protein interactions. *Bioorg Med Chem Lett*, 26: 707-13.

Xie, Z., et al. 2011. Systematic characterization of protein-DNA interactions. *Cell Mol Life Sci*, 68: 1657-68.

Yang, X. Y., et al. 2010. Classification of transcription factors using protein primary structure. *Protein Pept Lett*, 17: 899-908.

Yin, J. W., and G. Wang. 2014. The Mediator complex: a master coordinator of transcription and cell lineage development. *Development*, 141: 977-87.

Yin, Y., et al. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, 356.

Yue, P., Z. Li, and J. Moult. 2005. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol*, 353: 459-73.

Zaret, K. S., and J. S. Carroll. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*, 25: 2227-41.

Zhang, N., et al. 2014. MYC interacts with the human STAGA coactivator complex via multivalent contacts with the GCN5 and TRRAP subunits. *Biochim Biophys Acta*, 1839: 395-405.

Zhang, P., et al. 2016. An Overview of Chromatin-Regulating Proteins in Cells. *Curr Protein Pept Sci*, 17: 401-10.

Zhang, X., et al. 2002. Phosphorylation of serine 256 suppresses transactivation by FKHR (FOXO1) by multiple mechanisms. Direct and indirect effects on nuclear/cytoplasmic shuttling and DNA binding. *J Biol Chem*, 277: 45276-84.

Zhang, Y., and Y. Xiong. 2001. A p53 amino-terminal nuclear export signal inhibited by DNA damage-induced phosphorylation. *Science*, 292: 1910-5.

Zhao, S., X. Zhang, and H. Li. 2018. Beyond histone acetylation-writing and erasing histone acylations. *Curr Opin Struct Biol*, 53: 169-77

Recent Publications in this Series

45/2019 Mgbeahuruike Eunice Ego
Evaluation of the Medicinal Uses and Antimicrobial Activity of *Piper guineense* (Schumach & Thonn)
46/2019 Suvi Koskinen
Near-Occlusive Atherosclerotic Carotid Artery Disease: Study with Computed Tomography Angiography
47/2019 Flavia Fontana
Biohybrid Cloaked Nanovaccines for Cancer Immunotherapy
48/2019 Marie Mennesson
Kainate Receptor Auxiliary Subunits Neto1 and Neto2 in Anxiety and Fear-Related Behaviors
49/2019 Zehua Liu
Porous Silicon-Based On-Demand Nanohybrids for Biomedical Applications
50/2019 Veer Singh Marwah
Strategies to Improve Standardization and Robustness of Toxicogenomics Data Analysis
51/2019 Iryna Hlushchenko
Actin Regulation in Dendritic Spines: From Synaptic Plasticity to Animal Behavior and Human Neurodevelopmental Disorders
52/2019 Heini Liimatta
Effectiveness of Preventive Home Visits among Community-Dwelling Older People
53/2019 Helena Karppinen
Older People´s Views Related to Their End of Life: Will-to-Live, Wellbeing and Functioning
54/2019 Jenni Laitila
Elucidating Nebulin Expression and Function in Health and Disease
55/2019 Katarzyna Ciuba
Regulation of Contractile Actin Structures in Non-Muscle Cells
56/2019 Sami Blom
Spatial Characterisation of Prostate Cancer by Multiplex Immunohistochemistry and Quantitative Image Analysis
57/2019 Outi Lyytinen
Molecular Details of the Double-Stranded RNA Virus Replication and Assembly
58/2019 Markus Räsänen
Vascular Endothelial Growth Factor-B and the Bmx Tyrosine Kinase in Cardiac Hypertrophy and Revascularization
59/2019 Vuokko Nummi
Insights into Clinical and Laboratory Phenotypes of Von Willebrand Disease
60/2019 Shah Hasan
Challenges of Hyper-Prolificacy in the Pig: Colostrum and Gut Microbiota
61/2019 Sanna Matilainen
Pathomechanisms of Leigh Syndrome: Defects of Post-Transcriptional and Post-Translational Regulation of Mitochondrial Metabolism
62/2019 Kirsi Santti
Desmoid Tumor: Oncological Management and Prognostic Biomarkers
63/2019 Hesham E. Abdolhfid Mohamed
Evaluation of Prognostic Markers for Oropharyngeal Carcinoma Using Tissue Microarray
64/2019 Johanna Uhari-Väänänen
Contributions of μ- and κ-Opioidergic Systems to Ethanol Intake and Addiction
65/2019 Susanna Rapo-Pylkkö
Chronic Pain and Neuropathic Pain among Community-dwelling Older Adults in Primary Health Care Settings

Helsinki 2019 ISSN 2342-3161 ISBN 978-951-51-5480-4