



Two complementary proteomic analyses uncover new regulators of TBX18 transcriptional function

A thesis submitted for the degree of

Doctor of Philosophy (PhD)

in the subject of Molecular Biology

by

Reginaldo Rivera Reyes, M.Sc. Biochemistry

September 2019

Hannover Medical School

International MD/PhD program “Molecular Medicine”

at Hannover Biomedical Research School (HBRS)

Institute for Molecular Biology

Acknowledged by the PhD committee and the head of Hannover Medical School

President: Prof. Dr. med. Michael P. Manns

Supervisor: Prof. Dr. rer. nat. Andreas Kispert

Cosupervisors: Prof. Dr. Dr. med. Thomas Thum; Prof. Dr. rer. nat. Ina Gruh

External expert: Prof. Dr. Vincent Christoffels

Internal expert: Prof. Dr. Halyna Shcherbata

Day of final exam / public defense: 8th of November, 2019

Das Denken ist mit einem Nimbus umgeben. — Sein Wesen, die Logik, stellt eine Ordnung dar, und zwar die Ordnung a priori der Welt, d.i. die Ordnung der Möglichkeiten, die Welt und Denken gemeinsam sein muß. Diese Ordnung aber, so scheint es, muß höchst einfach sein. Sie ist vor aller Erfahrung; muß sich durch die ganze Erfahrung hindurchziehen; ihr selbst darf keine erfahrungsmäßige Trube oder Unsicherheit anhaften. — Sie muß vielmehr vom reinsten Kristall sein. Dieser Kristall aber erscheint nicht als eine Abstraktion; sondern als etwas Konkretes, ja als das Konkreteste, gleichsam Härteste.

Ludwig Wittgenstein, Philosophische Untersuchungen (1953)

The man who embraces a new paradigm at an early stage must often do so in defiance of the evidence provided by the problem- solving. He must, that is, have faith that the new paradigm will succeed with the many large problems that confront it, knowing only that the old paradigm has failed with a few. A decision of that kind can only be made on faith.

Thomas Kuhn, The Structure of Scientific Revolutions (1962)

Summary

The TBX18 protein is a DNA binding transcription factor that regulates cell patterning and differentiation programs during embryonic development in the primordia of the posterior pole of the heart, the vertebral column, the limb, the inner ear, the prostate and the ureter. Although the functional relevance of TBX18 for the development of these organs has been described, the molecular mechanisms by which TBX18 exerts its transcription regulation effect is still poorly understood. In this work, we set out to identify and characterize protein interaction partners of TBX18 in a cell culture context to circumvent the lack of suitable anti-TBX18 antibodies and the restricted expression of TBX18. We made use of two proteomic screens carried out with LC-MS to identify candidate proteins that interact with TBX18. First through overexpression of dual-tagged TBX18 in 293 cells and later with endogenous tagged TBX18 from 293 and A549 cells modified with the CRISPR/Cas9 system. Given the completely different strategies we used, we found and validated distinct TBX18 interaction partners from each approach in biochemical and cellular assays. Despite this finding, the validated candidates from both screenings belong to 3 groups of proteins related to transcription; transcription cofactors, homeobox and zinc-finger transcription factors. Based on the newly validated partners, we propose two molecular mechanism through which TBX18 can exert its transcription repression function. Most notably, we validated the interaction between TBX18 and ZMYM2 that could help explain the molecular basis for predisposition to congenital nephropathies, as mutations in either of these genes were recently genetically associated with this affliction in human patients. The relevance of the proposed mechanisms will be the subject of testing in future genetic interaction analyses between TBX18 and its interaction partners.

Table of contents

Summary	3
Table of contents	4
Introduction.....	5
Gene regulation shapes morphogenesis	5
Regulation of transcription: activation vs repression	5
The T-box family of transcription factors regulates gene expression during development	7
The T-box 18 transcription factor prepatterns tissues for differentiation	8
Mutations in <i>TBX18</i> underlie congenital defects of the urinary tract	8
Structure and function of the TBX18 protein	10
Aim of this thesis	13
Part 1 – An unbiased proteomic approach to identify TBX18 interacting proteins	15
" Proteomic analysis identifies transcriptional cofactors and homeobox transcription factors as TBX18 binding proteins".....	15
Part 2 - Characterizing endogenous TBX18 protein interaction partners.....	55
"Proteomic analysis identifies ZMYM2 as endogenous binding partner of TBX18" ..	16
Concluding remarks and outlook	97
The molecular mechanisms of TBX18 repression.....	17
TBX18 interacts with Homeobox TFs and corepressors	18
The zinc finger transcription factor ZMYM2 is an endogenous interaction partner of TBX18.....	20
Future work	22
References	104
Acknowledgements	108
Appendix.....	109
Curriculum vitae.....	110
List of publications	114
Declaration	115

Introduction

Gene regulation shapes morphogenesis

Metazoan organisms undergo complex morphological and cellular changes during embryonic development to transform early and simple precursor structures into organ systems with a highly organized architecture (Gilmour, Rembold, & Leptin, 2017). These changes occur in a strict spatial and temporal manner under the control of conserved developmental mechanisms which are underlaid by alterations of gene expression (Ettensohn, 2013).

Gene expression is controlled at different levels through the activity of macromolecular complexes including that for transcription, splicing, translation and posttranslational modifications (Maniatis & Reed, 2002). Transcription regulation is the starting point for genes to be expressed or repressed and is not only essential for development but also for the homeostasis in adult life (Vaquerizas, Kummerfeld, Teichmann, & Luscombe, 2009).

Regulation of transcription: activation vs repression

Transcription of a gene comprises the production of an immature messenger RNA (pre-RNA) from the genomic DNA (gDNA) template that is found associated with histone proteins in chromatin (Caramori, Ruggeri, Mumby, Atzeni, & Adcock, 2019). There is a complex interplay between the chromatin state and transcription machinery, normally only “open” or “relaxed” chromatin allows for transcription to take place (Li et al., 2011).

If the chromatin is in this relaxed permissible state, it is possible for the basal transcription machinery, including the RNA polymerase holozyyme to transcribe the pre-RNA from the DNA. However, the basal transcription machinery is normally not sufficient to initiate transcription (Goldman, Ebright, & Nickels, 2009). At the center of the transcriptional control mechanisms are tissue-specific transcription factors (TFs), proteins that bind to specific regions of the genomic DNA and regulate the expression of adjacent transcription units in a spatial and temporal manner by interacting with the basal transcription machinery or through cofactor proteins. The regions of gDNA that grant this specificity are called DNA regulatory elements and they can be recognized by TFs based on their DNA nucleotide sequence (Spitz & Furlong, 2012) (Fig. 1).

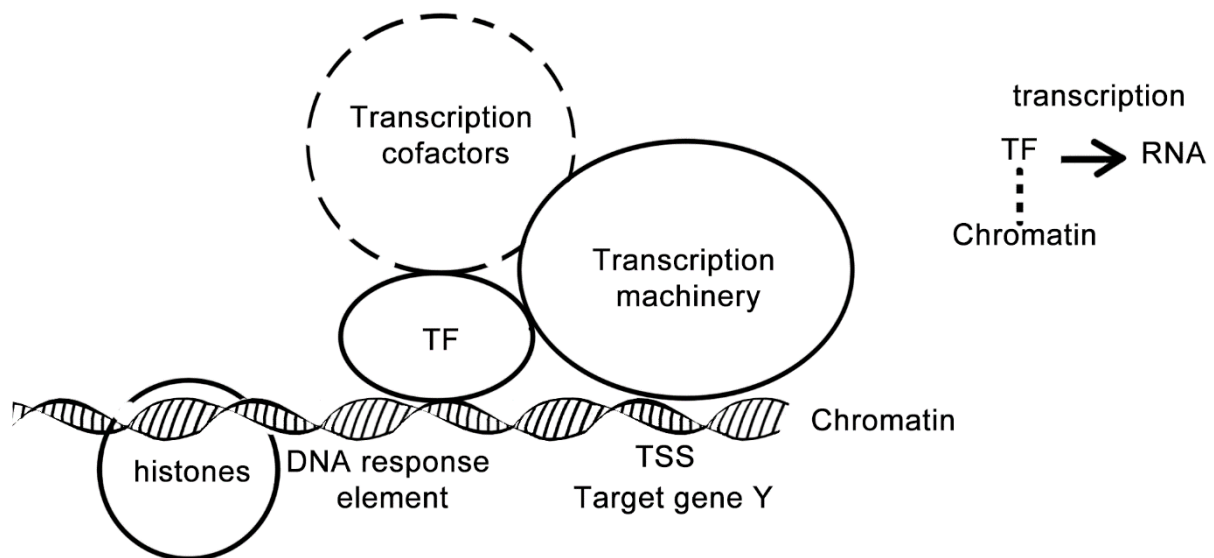


Figure 1. Genes are transcribed from the gDNA in permissive chromatin into pre-RNA with the help of different transcription machinery components. A certain gene Y is transcribed with the help of the basal transcription machinery after transcription factors and possibly transcription cofactors interact with it. The basal transcription machinery binds to DNA elements near the transcription start site (TSS) while the TFs bind to specific regulatory elements of gene Y.

Transcription factors can be functionally divided into activators or repressors regarding their effect on transcription outcome. If they favor transcription by binding to response elements like promoters or enhancers to recruit the RNA polymerase, they are called activators, but if they bind to regulatory elements to prevent transcription, they are called repressors. Transcriptional repression or activation may be achieved by the association with other transcription factors as well as transcriptional cofactors that together modify the chromatin to attract or avoid the activity of RNA polymerases (Caramori et al., 2019). (Soutourina, 2018; Spitz & Furlong, 2012).

In the case of transcriptional repression, the repressor TFs may act in different ways to prevent transcription. They may modify the chromatin state to prevent the transcription machinery to reach the regulatory elements and initiate transcription or they may antagonize the transcription machinery proteins of the basal transcriptional machinery or activating transcription factors (Li et al., 2011).

In both cases, the repressor TFs need to interact physically directly with other proteins to prevent transcription. If they stop the transcription machinery they need to interact with proteins that stop the machinery, like the corepressor GROUCHO, that is thought to pause the RNA polymerase (Chambers et al., 2017) or they may antagonize the corresponding activating TF activity by physically competing for the regulatory

elements like in the case of the TBX18-TBX5/NKX2.5/GATA4 TF axis (Farin et al., 2007). On the other hand, repressor TFs are thought to achieve long term repression by recruiting enzymatically active protein complexes that chemically modify the histones or DNA from the chromatin (Perissi, Jepsen, Glass, & Rosenfeld, 2010). These two mechanisms are not exclusive, and some corepressor proteins such as GROUCHO may later recruit chromatin modifying complexes in time (Kaul, Schuster, & Jennings, 2014)

The T-box family of transcription factors regulates gene expression during development

The T-box (TBX) family is one of different families of TFs that regulate the molecular programs that drive embryonic development. TBX proteins are characterized by a conserved T-box domain that is able to bind specifically to conserved DNA sequences called T-box binding elements (TBE). The TBEs are short DNA sequences found single or as double repeats in the genome arranged either in a parallel or antiparallel fashion across both DNA strands (and therefore may also be called T-half sites). Studies have shown that T-box proteins bind T-half sites as dimers, and therefore orientation of the T-half sites and dimerization capabilities of the T-box region may play an important role in T-box protein function (Muller & Herrmann, 1997). T-box proteins interact with each other but also with other transcription factors and cofactors to eventually regulate transcription positively or negatively (Kaltenbrun et al., 2013; Naiche, Harrelson, Kelly, & Papaioannou, 2005; Waldron et al., 2016).

The T-box family of TFs has been subdivided into 5 subfamilies based on amino acid sequence similarity (Naiche et al., 2005). Most of these subfamilies have already been established at the onset of metazoan development (Papaioannou, 2014) but the number of family members steadily increased in evolution by gene duplication events. T-box proteins are known to regulate gene expression in many different contexts of early embryonic development in almost all vertebrates. Correspondingly, loss of function of some of the family members have been associated with dramatic phenotypes in humans: *TBX1* mutations lead to DiGeorge syndrome, which comprises heart and vascular abnormalities; *TBX3* mutations cause Ulnar-mammary syndrome, that includes limb defects, hypoplastic mammary glands and defects in cardiac conduction; *TBX5* mutations lead to Holt-Oram syndrome, which encompasses heart

and hand abnormalities (Ghosh, Brook, & Wilsdon, 2017; Naiche et al., 2005; Papaioannou, 2014) to name only some important ones.

The T-box 18 transcription factor prepatterns tissues for differentiation

T-box 18 (*Tbx18*) is a member of the ancient *Tbx1* subfamily, which is present since the emergence of the bilaterian lineage. *Tbx18* is a more recent addition to this family in chordates that may have diversified from *Tbx20* (Sebe-Pedros & Ruiz-Trillo, 2017). TBX18 is 94.4% identical between mice and humans when assessed by an amino acid pairwise alignment (Sievers et al., 2011). The functional relevance of *Tbx18* in development has been mainly addressed in the mouse. It turned out that *Tbx18* is important during mouse embryonic development for correct formation of the axial skeleton, inner ear, cardiac pacemaker, epicardium, and urogenital system. *Tbx18* maintains cell identity of the anterior somite halves as an antiapoptotic factor. When *Tbx18* is deleted there is an abnormal pedicle and rib expansion, causing perinatal death due to respiratory complications. In inner ear development, *Tbx18* compartmentalizes the otic mesenchyme to allow adequate fibrocyte differentiation. Loss of *Tbx18* in the inner ear causes deafness in adult mice. During heart development, it was shown that *Tbx18* maintains the proliferation of mesenchymal cells in the inflow tract that contribute to the sinoatrial node head. Upon deletion of *Tbx18*, the sinoatrial node is greatly reduced. *Tbx18* is also expressed in the proepicardial organ and epicardium and is thought to control epicardial development through cell signaling. Loss of *Tbx18* in this context leads to defective coronary vessel structure and function (Bussen et al., 2004; Farin, Mansouri, Petry, & Kispert, 2008; Trowe, Maier, Schweizer, & Kispert, 2008; Wiese et al., 2009; Wu, Dong, Regan, Su, & Majesky, 2013)

Mutations in *TBX18* underlie congenital defects of the urinary tract

The ureter is the tubular conduct that transports the urine produced in the kidneys to the bladder for disposal. This is achieved by peristaltic activity of a smooth muscle layer that resides in the wall of this tube. *Tbx18* plays a crucial role in establishing this contractile tube. *Tbx18* is first expressed at embryonic stage (E)11.5 in a stripe of mesenchymal cells in between those of the metanephros and the Wolffian duct. At E12.5, this *Tbx18*-positive ureteric mesenchyme segregates into an inner (adluminal) and outer (abluminal) mesenchyme. *Tbx18* expression is maintained in the inner

mesenchymal layer until E14.5 when it becomes downregulated prior to smooth muscle differentiation. The deletion of *Tbx18* in the ureteric mesenchyme leads to reduced proliferation in the mesenchymal progenitor pool, and a failure of smooth muscle differentiation, which ultimately results in functional ureter obstruction, hydroureter and hydronephrosis (Fig. 2)

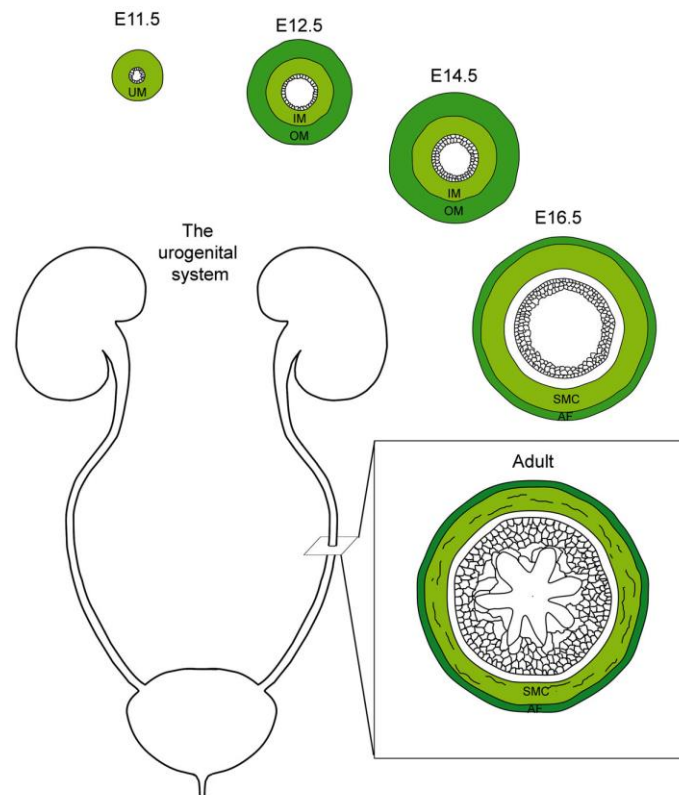


Figure 2. *Tbx18* is expressed early in the ureteric mesenchyme. *Tbx18* is expressed from E11.5 until E14.5 in the inner ureteric mesenchyme. The mesenchymal components and derivatives are shown in green, *Tbx18* is expressed in the light green domains. UM: ureteric mesenchyme, IM: inner mesenchyme, OM: outer mesenchyme, SMC: smooth muscle cells, AF: tunica adventitia.

More detailed molecular analyses showed that TBX18 in this context prepatterns the ureteric mesenchyme to respond to signals emitted from the ureteric epithelium. By doing this, TBX18 suppresses a renal stromal program in favor of a ureteric smooth muscle differentiation program (Airik, Bussen, Singh, Petry, & Kispert, 2006; Bohnenpoll et al., 2013).

In agreement with the relevance of *Tbx18* for ureter development in the mouse, mutations in *TBX18* were found to cause congenital anomalies of the kidney and urinary tract (CAKUT) in human patients (Vivante et al., 2015). The identified mutations

reduced the repressive function of TBX18 by ablating the binding of TBX18 to TBEs or by generating dominant negative loss-of-function versions of the protein.

Structure and function of the TBX18 protein

Based on the sequence conservation of the T-box, TBX18 protein can be divided in three regions. In the mouse, a region of 154 residues at the N-terminus precedes 182 highly conserved residues which make up the T-box domain, and a C-terminal domain consisting of 277 residues. The N-terminal domain of TBX18 contains an eh1-motif that interacts with TLE3/GROUCHO corepressors to partly regulate transcription, and a nuclear localization signal needed for TBX18 to shuttle to the nucleus. The T-box domain mediates binding to classical TBEs but also interacts with T-boxes of other family members. Despite its length, no additional motifs have been identified in the C-terminal domain of the protein (Fig. 3) (Farin et al., 2007).

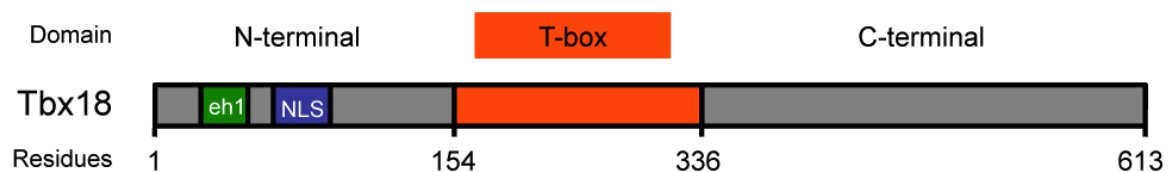


Figure 3. Biochemical structure of mouse TBX18. Identified features of the protein are highlighted in color. eh1: engrailed homology 1 motif, NLS: nuclear localization signal.

In vitro DNA binding and activation assays uncovered that TBX18 represses transcription in a GAL4-upstream activating sequence luciferase reporter system in a concentration dependent manner (Farin et al., 2007). Using the same system, it was also shown that the different domains of TBX18 contribute to its repressive activity differently, with the N-terminal plus T-Box domain being the dominant ones.

Identification of a classic eh1-motif in the N-terminal domain revealed that TBX18 represses transcription at least partially by binding to the cofactor TLE3/GROUCHO. This was proven by finding that deleting this eh1-motif causes a mild derepression. This mild change in repression activity mediated by the TLE3 corepressor could theoretically be explained by the fact that it can interact with HDAC2 and together they may be linked to the NUCleosome Remodeling and Deacetylase (NuRD) repressor complex (Kaltenbrun et al., 2013; Lai & Wade, 2011). Alternatively, TLE3 may act with

transcription factors to condensate chromatin (Sekiya & Zaret, 2007) or repress genes by binding to nearby T-cell factor/lymphoid enhancer factor (TCF/LEF) transcription factors (Arce, Pate, & Waterman, 2009; Brantjes, Roose, van De Wetering, & Clevers, 2001; Cadigan & Waterman, 2012).

A number of additional protein interaction partners have been found in recent years. TBX18 binds to GATA4 and NKX2.5 through its N-terminal and T-box domains *in vitro*. When assessing the importance of this interaction in a transactivation assay in cells, the combination of these factors allowed for the repression of transcription of a luciferase reporter of the *Nppa* promoter that is activated by TBX5, showing that TBX18 competes with TBX5 for binding of NKX2.5 and GATA4 (Farin et al., 2007).

TBX18 was found to interact with PAX3 in a yeast two-hybrid assay. Subsequently, PAX3, PAX1, PAX7 and PAX9 were confirmed to interact with the T-box of TBX18 through their paired domain *in vitro*. The importance of the interaction with PAX3 was confirmed *in vivo* since *Pax3* and *Tbx18* gene functions were combinatorially required for maintaining the anterior somite half by suppressing posterior somite identity (Farin et al., 2008).

The SIX1 protein was shown to interact with TBX18 genetically and physically to direct smooth muscle cell differentiation. A synergistic effect on smooth muscle precursor proliferation and apoptosis was observed upon *Six1* and *Tbx18* heterozygous loss which led to hydronephrosis and hydroureter formation. SIX1 is a homeobox TF that binds DNA through its homeobox domain and also possesses a Six domain to interact with other proteins, but no exact mechanism for the molecular functions of TBX18/SIX1 is yet known. Interestingly, point mutations in SIX1 found in branchio-oto-renal (BOR) syndrome patients prevented or decreased the TBX18-SIX1 interaction *in vitro*. Considering that 6% of BOR patients present renal defects, including hydroureter and hydronephrosis, this interaction may explain cooperativity of SIX1 and TBX18 to regulate the smooth muscle cell differentiation gene program (Nie, Sun, Gordon, Cai, & Xu, 2010).

An *in vitro* proteomic analysis found TBX18 interaction with TBX20 in 293 cells. TBX20 was linked to the NuRD repressor complex but no evidence for TBX18 interactions with repressor complexes was found (Kaltenbrun et al., 2013)

All of the summarized interacting proteins are DNA-binding zinc-finger or homeobox transcription factors that are selectively expressed in discrete times and domains but not necessarily together with TBX18. This makes it unlikely that they fully account for

TBX18 function *in vivo*. Additionally, all these proteins by themselves lack protein domains or enzymatic activities required to influence gene expression directly and do not provide a direct link to complexes that can regulate gene expression.

Moreover, the finding of mutations in *TBX18* in human CAKUT patients unveiled that the C-terminal domain of TBX18 plays an important role in transcription regulation when evaluated with a luciferase reporter under the control of two palindromic T-half sites (Farin et al., 2007; Vivante et al., 2015). To which proteins the C-terminal domain and the entire TBX18 protein bind to exert transcriptional repression *in vivo* has, therefore, remained largely elusive. Further, not a single target gene of TBX18 has been identified. Identifying protein interaction partners as well as target genes of TBX18 is of outmost importance to understand how this transcription factor regulates such a variety of developmental programs including ureter, heart, inner ear and skeletal development.

Aim of this thesis

The diverse morphological functions of the transcription factor TBX18 have been characterized by gene targeting in mice, where it was shown to prepattern and program diverse tissues for differentiation. Despite being responsible for well-known effects on cell identity in some of its expression domains during embryonic development, the exact molecular mechanisms through which TBX18 functions are still unknown. In order to fully understand the molecular mechanisms responsible for TBX18 transcriptional regulation function, we seek to identify and validate new interaction partner proteins that can functionally relate TBX18 to transcription regulation function. In order to identify the candidate proteins, we aim to use unbiased proteomic approaches through affinity purification of TBX18 containing complexes. Importantly, due to the lack of suitable anti-TBX18 antibodies and the restricted TBX18 protein expression during development, we seek to use an alternative source of TBX18 protein complexes.

The candidate partner proteins identified must be linked to transcriptional regulation in order to provide a feasible relation between TBX18 and its transcriptional regulation activity.

The candidate proteins must be validated in independent cellular and biochemical assays. The functional interaction between TBX18 and the partner proteins must take place in a context that allows transcription regulation, the nucleus of the cell, while direct or indirect biochemical interaction between the candidate proteins and TBX18 needs to be examined.

In order to explain the function of TBX18, the validated proteins need to be put into the context of known transcription regulation, a link between TBX18 and known transcription regulation processes (e.g. chromatin modification) needs to be proposed. The possible molecular mechanism for TBX18 must be outlined by interactions between TBX18 and proteins known to carry transcriptional regulation processes.

Functional transcription regulation must be assessed in order to evaluate the potential of the interactions between TBX18 and validated partner proteins.

The identification and characterization of new TBX18 interaction partners will be the first step towards the proposal of a mechanism that explains TBX18 molecular function. A future aim will be to demonstrate the relevance of the proposed mechanisms in the actual biological contexts of TBX18.

Part 1 – An unbiased proteomic approach to identify TBX18 interacting proteins
" Proteomic analysis identifies transcriptional cofactors and homeobox transcription factors as TBX18 binding proteins"

In this paper I performed all experiments except the initial proteomic screening. I also performed all data analysis and wrote the manuscript.

RESEARCH ARTICLE

Proteomic analysis identifies transcriptional cofactors and homeobox transcription factors as TBX18 binding proteins

Reginaldo Rivera-Reyes[☯], Marc-Jens Kleppa[☯], Andreas Kispert^{*}

Institut für Molekularbiologie, Medizinische Hochschule Hannover, Hannover, Germany

☯ These authors contributed equally to this work.

* kispert.andreas@mh-hannover.de



OPEN ACCESS

Citation: Rivera-Reyes R, Kleppa M-J, Kispert A (2018) Proteomic analysis identifies transcriptional cofactors and homeobox transcription factors as TBX18 binding proteins. PLoS ONE 13(8): e0200964. <https://doi.org/10.1371/journal.pone.0200964>

Editor: Claude Prigent, Institut de Genetique et Developpement de Rennes, FRANCE

Received: April 17, 2018

Accepted: May 30, 2018

Published: August 2, 2018

Copyright: © 2018 Rivera-Reyes et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Cluster of Excellence REBIRTH (From Regenerative Biology to Reconstructive Therapy) at Hannover Medical School, and by the individual grant DFG KI728/7-2 (AK). The funders had no role in study design, data

Abstract

The TBX18 transcription factor is a crucial developmental regulator of several organ systems in mice, and loss of its transcriptional repression activity causes dilative nephropathies in humans. The molecular complexes with which TBX18 regulates transcription are poorly understood prompting us to use an unbiased proteomic approach to search for protein interaction partners. Using overexpressed dual tagged TBX18 as bait, we identified by tandem purification and subsequent LC-MS analysis TBX18 binding proteins in 293 cells. Clustering of functional annotations of the identified proteins revealed a highly significant enrichment of transcriptional cofactors and homeobox transcription factors. Using nuclear recruitment assays as well as GST pull-downs, we validated CBFB, GAR1, IKZF2, NCOA5, SBNO2 and CHD7 binding to the T-box of TBX18 *in vitro*. From these transcriptional cofactors, CBFB, CHD7 and IKZF2 enhanced the transcriptional repression of TBX18, while NCOA5 and SBNO2 dose-dependently relieved it. All tested homeobox transcription factors interacted with the T-box of TBX18 in pull-down assays, with members of the *Pbx* and *Prrx* subfamilies showing coexpression with *Tbx18* in the developing ureter of the mouse. In summary, we identified and characterized new TBX18 binding partners that may influence the transcriptional activity of TBX18 *in vivo*.

Introduction

T-box (*Tbx*) genes encode a family of proteins that share a highly conserved domain for DNA-binding, the T-box [1, 2]. They act as transcription factors that activate or repress target gene expression upon binding to a conserved short variably spaced and oriented DNA-binding site, the T-box binding element (TBE) [3, 4]. In mammals, genome mining identified 17 members of this gene family that are grouped in five subfamilies depending on sequence conservation. Gene targeting experiments in mice revealed critical functions of some of these genes in the formation and differentiation of the germ layers, and in the development of various organ systems. Mutational analysis in man characterized mutations in *TBX* genes as causes for congenital diseases, demonstrating the importance of the gene family as regulators of developmental programs in mammals (for reviews see [5–8]).

collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Tbx18 is a member of a vertebrate specific subgroup within the *Tbx1*-subfamily [9]. Similar to the closely related *Tbx15* and *Tbx22* genes [10–12], *Tbx18* possesses a highly complex and dynamic array of developmental expression sites that in the mouse embryo include the anterior somite halves in the paraxial mesoderm, the developing heart, the mesenchyme surrounding the otic vesicle, the urogenital ridge including the ureteric and prostate mesenchyme, the epicardium, the mandibular/maxillary region, the proximal mesenchyme of the limb buds, and the skin [9, 13–16]. Mice homozygous for an engineered null allele of *Tbx18* died shortly after birth due to severe malformations of the axial skeleton [17]. The capsule and fibrocytes of the inner ear were disrupted, and the sinoatrial node was reduced in size [15, 18]. The epicardium and the coronary vasculature appeared altered [19, 20]. The periductal smooth muscle stromal cells in the prostate were reduced [16]. The ureter was shortened and lacked the smooth muscle coating resulting in hydroureter and hydronephrosis [13]. These defects were traced to independent functions of *Tbx18* in patterning and differentiation of the primordia of the affected organs, the somites, the otic mesenchyme, the sinus venosus, and the prostate and ureteric mesenchyme [13–17].

While these studies testified the multitude of developmental functions of the *Tbx18* gene, we still lack insight to what other transcription factors and cofactors TBX18 binds to form transcription regulation complexes, and what target genes are controlled by these complexes at the diverse sites of expression in the embryo. A couple of years ago, our lab started the analysis of the biochemical properties of TBX18 as a transcription factor [21]. We found that the murine protein constitutively localizes to the nucleus and that a classical nuclear localization signal at the N-terminus of the protein accounts for this behavior. PCR based cyclic enrichment of DNA-fragments confirmed that the T-box of TBX18 binds to a combination of conserved TBEs *in vitro*. The T-box does not only mediate DNA-binding, but also constitutes a protein interaction region [21]. Interaction in *in vitro* binding assays was shown with other T-box proteins, including TBX18 itself and TBX15 but also with members of other classes of tissue-specific transcription factors like NKX2.5, GATA4, PAX1, PAX3, PAX9 and SIX1 [21–23].

We further showed that fusion proteins of either TBX18, or its N-terminal, C-terminal and T-box regions with the DNA binding domain of the yeast GAL4 protein repressed transcription of a reporter gene under the control of GAL4 binding sites suggesting that TBX18 acts as a transcriptional repressor and that multiple regions of the protein are able to exert this effect. Removal of a conserved eh1-motif, which acts as a binding site for members of the Groucho (TLE) protein family of transcriptional corepressors abolished the transcriptional repression activity of TBX18 by 50%, stressing the notion that interaction with other cofactors at different sites of the protein confer additional repressive activity to TBX18 [21]. Indications that such interactions exist and are highly relevant at least for the function of TBX18 in the developing ureter *in vivo* were recently provided by the identification of mutations in *TBX18* in patients with congenital anomalies of the kidney and the urinary tract (CAKUT), including ureter-pelvic and vesico-ureteric junction obstruction and hydronephrosis. While one disease-causing mutation led to an exchange of a conserved residue in the T-box, another one led to deletion of the large C-terminal protein domain, and a third resulted in an amino acid exchange in this region. All of these mutant proteins showed reduced transcriptional repression and in the case of the C-terminal deletion mutant even a complete derepression compared to the wildtype protein [24].

Here, we set out to identify and characterize novel protein interaction partners of TBX18 to improve our comprehension of its transcriptional properties. Using an unbiased proteomic screen in 293 cells, we identified transcriptional corepressors and tissue specific transcription factors as binding candidates of TBX18. We validated their binding and present data on their significance for TBX18 transcriptional function.

Material and methods

Ethics statement

All animal work conducted for this study was performed according to European and German legislation. Breeding, handling and sacrifice of mice for isolation of embryos was approved by the Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit (Permit Number: AZ33.12-42502-04-13/1356, AZ33.12-42502-04-13/1875).

Mice

Embryonic day (E) 12.5 embryos for expression analysis were derived from NMRI wild-type mice which were purchased in house from the animal facility of the Medizinische Hochschule Hannover. Three to six mice per cage were housed with *ad libitum* access to food and water under conditions of regulated temperature (22°C) and humidity (50%) and a 12 h light/dark cycle. For timing of the pregnancies, vaginal plugs were checked in the morning after mating and noon was designated as embryonic day (E) 0.5. Female mice were sacrificed by cervical dislocation. Embryos and organs were harvested in PBS, decapitated, fixed in 4% paraformaldehyde overnight and stored in 100% methanol at -20°C before further use.

Semi-quantitative reverse transcription PCR

Cells were grown on 10 cm dishes until they reached confluence. Total RNA was extracted with peqGOLD RNAPure (PeqLab), and first-strand cDNA synthesis was performed with RevertAid reverse transcriptase (Fermentas). For quantitative PCR amplification of *Tbx18*, the forward primer 5' -GGTGGCAGGTAATGCTGACT and the reverse primer 5' -ACTTGCATTGCCTTGCTTGG were annealed at 56°C for 30 min, before elongation occurred at 72°C for 30 sec. The number of cycles was adjusted to the mid-logarithmic phase. For normalization, *Gapdh* was used (forward primer: 5' - ACCACAGTCCATGCCATCAC, reverse primer: 5' - TCCACCACCCTGTTGCTGTA, annealing at 56°C for 30 sec, elongation time 20 sec [25]. Quantification was performed with ImageJ 1.47 [26]. Assays were repeated at least three times in duplicate, and statistical analysis was done as previously described [27].

Identification of TBX18 interacting proteins in 293 cells

A cDNA fragment encoding a triple FLAG tag was generated by assembly PCR from several overlapping oligonucleotides. *NcoI* and *SmaI* sites were introduced at the ends to replace the single FLAG tag in *p.EF1α.FLAG.BioTag* (kind gift of A.P. McMahon) [28] with this fragment. A cDNA fragment with the open reading frame of *Tbx18* was amplified from a full-length *Tbx18* cDNA [9], and cloned into the *BamHI/XbaI* site of the resulting *p.EF1α.3xFLAG.BioTag* plasmid generating the *p.EF1α.3xFLAG.BioTag.Tbx18* plasmid. A *IRES.NLS.EGFP.BirA.BGHpA* fragment was released from *p.EF1α.Six2DE.IRES.NLS.EGFP.BirA.BGHpA* (kind gift of A.P. McMahon) [28] and cloned into the *NheI* site of *p.EF1α.3xFLAG.BioTag.Tbx18* yielding the plasmid *p.EF1α.3xFLAG.BioTag.Tbx18.IRES.NLS.EGFP.BirA.BGHpA*. This plasmid encodes a TBX18 protein fused to a N-terminal triple FLAG tag and a biotinylation signal peptide while also expressing the bacterial BirA enzyme required for biotinylation. 15 µg of the *p.EF1α.3xFLAG.BioTag.Tbx18.IRES.NLS.EGFP.BirA.BGHpA* and of the mock plasmid *p.EF1α.IRES.NLS.EGFP.BirA.BGHpA*, respectively, were transfected into 293 cells grown in a 15-cm dish at 70% confluence at the day of transfection using the calcium phosphate method [29]. Transfection efficiency was verified by fluorescence microscopy to be 80%. The next day the cells were expanded on 10 x 15-cm cell culture dishes and cultured for two more days until they reached 90–95% confluence. Nuclear extracts were prepared from these cells and 10 mg

protein content subjected to tandem affinity purification using first anti-FLAG M2 antibody (Sigma, #F3165) coupled to Protein A agarose beads (Santa Cruz, #sc-2001), with subsequent elution with a triple FLAG peptide (Sigma, #F4799), and second, a Strep-Tactin Superflow (IBA, #2-1206-002) binding column as previously published [30]. The purified complexes were resolved by SDS-PAGE, protein bands were excised and sent to the MHH Proteomics Facility where they were analyzed by liquid chromatography-mass spectrometry (LC-MS).

Sample preparation for MS analysis

Proteins were mixed, alkylated by acrylamide and further processed as described [31]. Peptide samples were analysed with a shot-gun approach and data dependent analysis in a LC-MS system (RSLC, LTQ Orbitrap Velos, both Thermo Fisher) as recently described [31]. Raw MS data were processed using Proteom discoverer 1.4 (Thermo Scientific) and Max Quant software (version 1.5) [32] and a data base containing human and viral proteins and common contaminants. Proteins were stated identified by a false discovery rate of 0.01 on protein and peptide level.

Bioinformatic analysis of MS datasets

Gene Ontology (GO) analysis was performed on the 143 proteins identified from TBX18 immunoprecipitation experiments to determine the enrichment of annotated molecular functions (MF), biological processes (BP) and cellular components (CC) using the Database for Annotation, Visualization and Integrated Discovery (DAVID) web program [33].

Expression constructs

Vectors for expression of proteins in 293 cells, for *in vitro* translation of proteins and for transactivation assays were acquired from researchers, companies or generated in house as detailed in [S1 Table](#). Cloning was done by PCR amplification of ORFs from donor plasmids and insertion of the restricted fragments into suitable sites in plasmids for *in vitro* expression (*pSP64.G.Myc/HA*) [21], and further shuttling into the eukaryotic expression plasmid *pcDNA3* (Invitrogen).

In vitro translation of radioactively labeled proteins

We used a coupled *in vitro* transcription/translation system from reticulocyte lysates (TNT SP6 or T7 Quick Coupled Transcription/Translation System, Promega) to generate proteins *in vitro*. 2 µg of the plasmid with a SP6 or T7 promoter were used for 50 µl reaction set-ups. ³⁵S labeled Methionine/Cysteine (Hartmann Analytic, Braunschweig, Germany) was added to a final concentration of 0.2–0.4 µCi/µl and reactions were incubated for 1.5 hr or 3 hr for longer proteins at 30°C.

GST-TBX18 fusion protein production

GST-TBX18 fusion proteins were produced as previously shown [21]. Briefly, the fragments encoding the N-terminal (aa 1–157), N-terminal plus T-box (aa 1–345), T-box (aa 148–345) and C-terminal region (aa 336–613) of TBX18 were inserted into the *pGEX-4T3* plasmid and transformed into *E.coli* BL-21 bacteria. After culture and IPTG induction, the bacteria were lysed and proteins were prepared by batch-purification using Glutathion Sepharose 4B beads (GE Healthcare, Chicago, Illinois). The protein content of the recovered beads was resolved by SDS-PAGE and visualized by Coomassie staining and the volumes of beads to achieve a standardized content of all of the GST-TBX18 fusion proteins were calculated with the help of ImageJ [26].

GST pull-down assays

The GST pull-down assays were performed as previously reported [21]. One or more 10-cm culture plates of 293 cells were transfected with 20 µg of eukaryotic expression plasmid using the calcium phosphate method. 2 days after transfection, proteins were extracted from each plate with 250 µl of NP-40 buffer with Phostop and cOmplete inhibitors (Roche) and a 1/10 aliquot was checked for protein expression by Western Blot. The rest of the protein extract was incubated with the standardized bead volumes of the GST-TBX18 domain fusion proteins for two hours before washing 3 times with pull-down buffer (20 mM HEPES, pH 7.9, 100 mM NaCl, 10 mM KCl, 5 mM MgCl₂, 0.5 mM EDTA, 5% glycerol, 0.05% Triton X-100, 1 mM dithiothreitol). The beads were boiled in 1x Laemmli buffer and the proteins were separated by SDS-PAGE. After blotting onto PVDF membranes, interactions were detected by Western Blot with anti-HA, anti-MYC and anti-FLAG antibodies (#ab1265 and #ab62928, Abcam, #F3165, Sigma).

For the GST pull-down with *in vitro* translated proteins, 20 to 40 µl of transcription/translation reaction were incubated with the standardized bead volume of each of the GST-TBX18 domain fusion proteins for two hours before washing 3 times with pull-down buffer. Analysis of bound protein was performed by SDS-PAGE followed by drying of the gels onto cellulose filters. Autoradiographic detection was performed by exposing the filters for 24 hr to Fujifilm BAS-IP MS 2025 phosphorescent imaging plates and scanning them in a Fujifilm FLA7000 Laser scanner.

Cell culture and transient transfections

293 cells were purchased (ACC 305, DSMZ, Braunschweig, Germany), all other cell lines were kindly provided by T. von Hahn (Medizinische Hochschule Hannover, Germany). Cell cultures were handled under strict sterile conditions. 293 cells were cultured in standard DMEM medium (GIBCO) containing 10% fetal bovine serum (Biochrom) and kept in an incubator at 37°C with 5% CO₂. The transient transfections were performed with the calcium phosphate method as previously described [29]. The *pd2E.GFP-N1* plasmid was independently transfected to check for efficiency of transfection, which was verified by epifluorescence microscopy.

Immunofluorescence analysis

293 cells were cultured in 12 well plates and transfected with *pcDNA3.1.TBX18ΔNLS* (500 ng), *pcDNA3.1.TBX18* (500 ng) or *pcDNA3.1* expression plasmids of the candidate proteins (500 ng). For recruitment assays, co-transfections were performed with either *pcDNA3.1.TBX18ΔNLS* (250 ng) or *pcDNA3.1.TBX18* (250 ng) plasmids, and *pcDNA3.1* expression plasmids of the candidate proteins (250 ng). After 2 days, immunofluorescent detection of TBX18 and the candidate proteins using the corresponding tags as previously reported [21] and imaged with a DM6000 microscope (Leica).

Reporter constructs and transactivation assays

To generate a reporter for TBX18-dependent transcriptional activity, we annealed the single-stranded DNA fragments 5′-GATCCGGTGCAGTAGGTGTGAAATCGCACCTGGGGA-3′ and 5′-GATCTCCCCAGGTGCGATTTCACACCTACTGCACCG-3′ and ligated the resulting double-stranded fragment into the *Bgl*II site of the *pGL3.Promoter* vector (Promega) to obtain *pGL3.Prom.Tbx18BS2*. Dual luciferase assays were performed in 293 cells in duplicates. Cells were seeded to 70% confluence in 6-well plates and 1 day later transfected using the calcium

phosphate method with 2.5 μ g of plasmid mix. 100 ng of *pRL-TK* were used for normalization, 250 ng of *pGL3.Prom.Tbx18BS2* were used to measure transcriptional activity, 250 ng of *pcDNA3.Tbx18* plasmid were used to evaluate the effect of TBX18, and in separate wells, additional 25, 250 and 500 ng of vectors encoding the interaction candidates were added to analyze the effect on TBX18 regulation of transcription. *pcDNA3* plasmid was used to fill up to 2.5 μ g. After 12 h, medium was refreshed and 1 day later the cells were extracted and analyzed according to the Dual-Luciferase Reporter Assay System kit (Promega #E1910) in two technical replicates.

RNA *in situ* hybridization

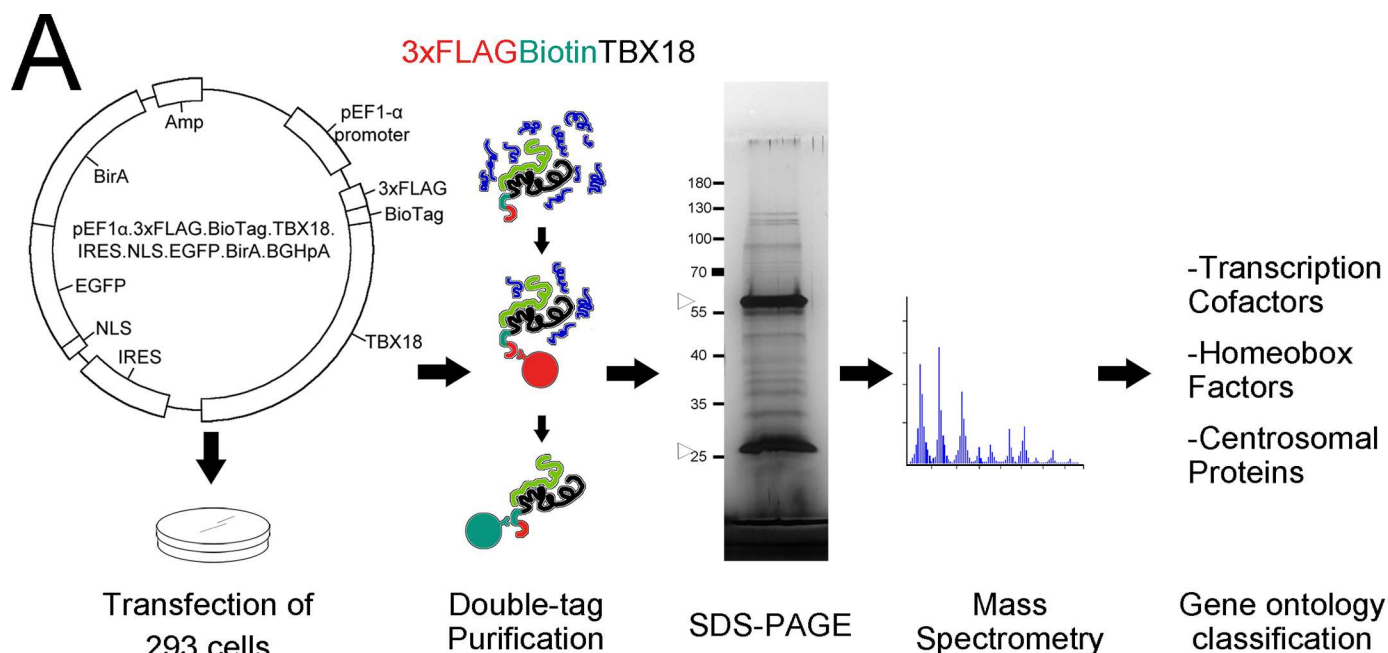
10- μ m paraffin sections of the posterior trunk region of embryonic day (E) 12.5 wildtype NMR mice containing the proximal ureter region were subjected to RNA hybridization with digoxigenin-labeled antisense riboprobes as previously described [34].

Results

Mass spectrometry analysis identifies novel TBX18 interacting proteins in 293 cells

Expression of *Tbx18* occurs at low levels in small progenitor pools of mammalian organ primordia hampering biochemical approaches to isolate endogenous complexes of TBX18 and interacting proteins. Since transcription factors exert transcriptional modulation functions upon expression in most cultured cells, we deemed that cell lines represent a more easily accessible source for the identification of at least some of the proteins that bind to TBX18 and affect its transcriptional activity. We chose 293 cells as a source for our proteomic screen since these cells are easy to grow and transfect, and are derived from an embryonic organ, the human kidney. Moreover, these cells express *TBX18* arguing that binding partners of TBX18 naturally occur in these cells (S1 Fig). Since available antibodies proved unsuitable for precipitation of endogenous protein complexes of TBX18 in our hands, we decided to transiently overexpress a version of TBX18 tailored for tandem affinity precipitation of native complexes [30]. For this, we transfected an expression construct encoding the mouse TBX18 protein fused to a N-terminal triple FLAG tag followed by a biotinylation signal peptide and the bacterial BirA enzyme required for biotinylation (Fig 1A). After cell expansion and nuclear extraction, the protein complexes containing TBX18 were purified by a two-step affinity purification strategy using Anti-FLAG antibody coupled to Protein A agarose beads and subsequent streptavidin affinity chromatography [28, 30].

The purified protein complexes were recovered by heat treatment and resolved by SDS-PAGE. The polyacrylamide gels of three independent experiments and controls were sent to the Hannover Medical School Proteomics Facility for protein extraction and subsequent LC-MS analysis (Fig 1A). Fragments of 143 human proteins were identified in any of the three independent IP-MS analyses we did but not in the control which lacked exogenous TBX18 expression (S2 Table). Ontological classification of annotated functions, biological processes and functional domains revealed a list of 84 enriched terms, in which nucleosome, homeobox, DNA-binding, nuclear chromatin were amongst the most highly significant ones (S3 Table). Furthermore, DAVID functional clustering with high stringency settings grouped these terms into 5 clusters, with nucleosome/chromatin and homeobox/DNA-binding/transcription related proteins having the highest enrichment scores (S4 Table). Histone H2A was by far the most strongly enriched protein indicating together that overexpressed TBX18 associates with other transcriptional regulatory proteins on chromatin in 293 cells. We manually mined the



B

Symbol	Name	UNIPROT	Annotations	Source
Transcription Cofactors				
BASP1	Brain acid soluble protein 1	P80723	Transcription regulatory region DNA binding	UP
CBFB	Core-binding factor subunit beta	Q13951	Transcription factor activity, sequence-specific DNA binding	GO
CHD7	Chromodomain-helicase-DNA-binding protein 7	Q9P2D1	Transcription, transcription regulation	UP, GO
GAR1	H/ACA ribonucleoprotein complex subunit 1	Q9NY12	RNA binding, telomerase holoenzyme complex	UP
IKZF2	Zinc finger protein Helios	Q9UKS7	Transcription, transcription regulation	UP, GO
NCOA5	Nuclear receptor coactivator 5	Q9HCD5	Transcription, transcription regulation	UP, GO
RCOR3	REST corepressor 3	Q9P2K3	Transcription, transcription regulation	UP, GO
SBNO2	Protein strawberry notch homolog 2	Q9Y2G9	Transcription, transcription regulation	UP, GO
SSXB2	Synovial sarcoma, X member B2	Q80ZT4	Regulation of transcription, DNA-templated	UP
SUV39H2	Histone-lysine N-methyltransferase SUV39H2	Q9H5I1	Transcription, transcription regulation	UP, GO
Transcription Factors				
DUX1	Double homeobox protein 1	O43812	Homeodomain, DNA binding	UP, GO, IP, SM
EVX1	Homeobox even-skipped homolog protein 1	P49640	Homeodomain, DNA binding	UP, GO, IP, SM
EVX2	Homeobox even-skipped homolog protein 2	Q03828	Homeodomain, DNA binding	UP, GO, IP, SM
GSC	Homeobox protein gooseoid	P56915	Homeodomain, DNA binding	UP, GO, IP, SM
PAX1	Paired box protein Pax-1	P15863	Homeodomain, DNA binding	IP
PBX4	Pre-B-cell leukemia transcription factor 4	Q9BYU1	Homeodomain, DNA binding	UP, GO, IP, SM
PRRX2	Paired mesoderm homeobox protein 2	Q99811	Homeodomain, DNA binding	UP, GO, IP, SM
Centrosomal proteins				
B9D2	B9 domain-containing protein 2	Q9BPU9	Centrosome, cilium biogenesis	UP, GO
CCDC39	Coiled-coil domain-containing protein 39	Q9UFE4	Cilium, cytoskeleton	UP
CENPE	Centromere-associated protein E	Q02224	Microtubule	UP, GO
CEP170	Centrosomal protein of 170 kDa	Q5SW79	Centrosome, microtubule	GO
CEP89	Centrosomal protein of 89 kDa	Q96ST8	Centrosome, cilium biogenesis	UP, GO
HAUS8	HAUS augmin-like complex subunit 8	Q9BT25	Centrosome	GO
MACF1	Microtubule-actin cross-linking factor 1	Q9UPN3	Microtubule	UP, GO
MAP9	Microtubule-associated protein 9	Q49MG5	Microtubule	UP
MICAL3	Protein-methionine sulfoxide oxidase MICAL3	Q7RTP6	cytoskeleton organization	UP
PARP3	Poly [ADP-ribose] polymerase 3	Q9Y6F1	Centriole	UP
PCM1	Pericentriolar material 1 protein	Q15154	Centrosome, cilium biogenesis	UP, GO
SYBU	Syntabulin	Q9NX95	Microtubule	GO
TPX2	Targeting protein for Xklp2	Q9ULW0	Microtubule	UP, GO

Fig 1. Mass spectrometry analysis identifies novel TBX18 interacting proteins in 293 cells. (A) Diagram of the identification strategy. A construct encoding mouse TBX18 protein fused to a N-terminal triple FLAG tag followed by a biotinylation signal peptide while also encoding the bacterial BirA enzyme required for biotinylation was transfected into 293 cells. Protein complexes containing TBX18 were purified from nuclear cell extracts by a two-step affinity purification strategy

using Anti-FLAG (red) and anti-Biotin chromatography (green). The purified protein complexes were resolved by SDS-PAGE. Silver staining of a sample aliquot is displayed, arrowheads show IgG heavy and light chains. After extraction, they were subjected to LC-MS analysis. Proteins were functionally classified and clustered. (B) Gene annotation enrichment analysis uncovered clusters of transcriptional cofactors, transcription factors of the homeobox family, and centrosomal proteins. Gene ontology analysis is provided. UNIPROT: Uniprot accession number. Database annotations are UP: Uniprot, GO: Gene Ontology, IP: Interpro, SM: Smart.

<https://doi.org/10.1371/journal.pone.0200964.g001>

available literature for each candidate and curated a list of 10 proteins (BASP1, CBF, CHD7, GAR1, IKZF2, NCOA5, RCOR3, SBNO2, SSXB2 and SUV39H2) with a functional assignment as transcriptional cofactors and a second list of 13 proteins (DUX1, DUX3, DUX4, DUX4L2, DUX4L4, DUX4L9, DUX5, EVX1, EVX2, GSC, PAX1, PRRX2, PBX4) characterized as tissue-specific transcription factors of the homeobox superfamily. Functional clustering and individual literature cross-referencing unexpectedly identified 13 proteins (B9D2, CCDC39, CENPE, CEP170, CEP89, HAUS8, MACF1, MAP9/ASAP, MICAL3, PARP3, PCM1, SYBU, TPX2) with a centrosomal and microtubule association (Fig 1B). Based on the recent report that the transcription factor ATF5 has an independent centrosomal function [35], we wished to determine whether these proteins might relate to an additional unexpected cellular function of TBX18. We rejected intermediate filament, keratin, actin filament, iron-sulfur, translation elongation, kinase, cell adhesion and ankyrin containing proteins for further validation as the established cytosolic or extracellular localization of these proteins point to an unspecific nature of the interaction.

Transcriptional cofactors bind to TBX18 in the nucleus of 293 cells

We validated the interactions of TBX18 with the candidate transcriptional cofactors using a number of independent assays. We started with interrogating whether in fact TBX18 colocalizes with and binds to these proteins in 293 cells. For this, we first transfected expression constructs for full length MYC- or HA-tagged candidate proteins, confirmed protein integrity by Western blot and characterized their localization by anti-MYC or anti-HA immunofluorescence analysis (S5A Table, S2A Fig). In addition to TBX18, BASP1, CHD7, GAR1, IKZF2, NCOA5, SBNO2, SSXB2 and SUV39H2 localized to the nucleus, while CBF and RCOR3 were found in the cytoplasm (S2B Fig). To test, whether TBX18 productively interacts with the candidate cofactors in 293 cells, we made use of a nuclear recruitment assay that we previously developed [21]. It is based on the identification of a classical nuclear localization signal (NLS) at the N-terminus of the TBX18 protein. When this NLS is deleted, the resulting protein (TBX18ΔNLS) is excluded from the nucleus but can be shuttled back to this compartment by binding to a protein which carries such a signal. Upon coexpression of TBX18ΔNLS and those candidate cofactors that showed nuclear localization in 293 cells, we observed a nuclear translocation of TBX18ΔNLS in the presence of BASP1, CHD7, GAR1, IKZF2, NCOA5, SBNO2 and SSXB2 proteins while coexpression of SUV39H2 left the cytoplasmic localization of TBX18ΔNLS unchanged (Fig 2A). To study TBX18 interaction with CBF and RCOR3 that reside in the cytoplasm, we modified this assay by cotransfecting expression constructs for full-length TBX18, and CBF or RCOR3, respectively. We observed that coexpression of TBX18 sufficed to shuttle the two proteins into the nucleus (Fig 2B). Together, this shows that TBX18 interacts with all transcriptional cofactor candidates except SUV39H2 in 293 cells.

Several transcriptional cofactors bind to TBX18 in a physically direct manner

To independently assess the physical interaction of TBX18 with the candidate transcriptional cofactors, we performed pull-down assays using bacterially expressed GST-TBX18 proteins that were incubated with lysates from 293 cells transfected with expression constructs for

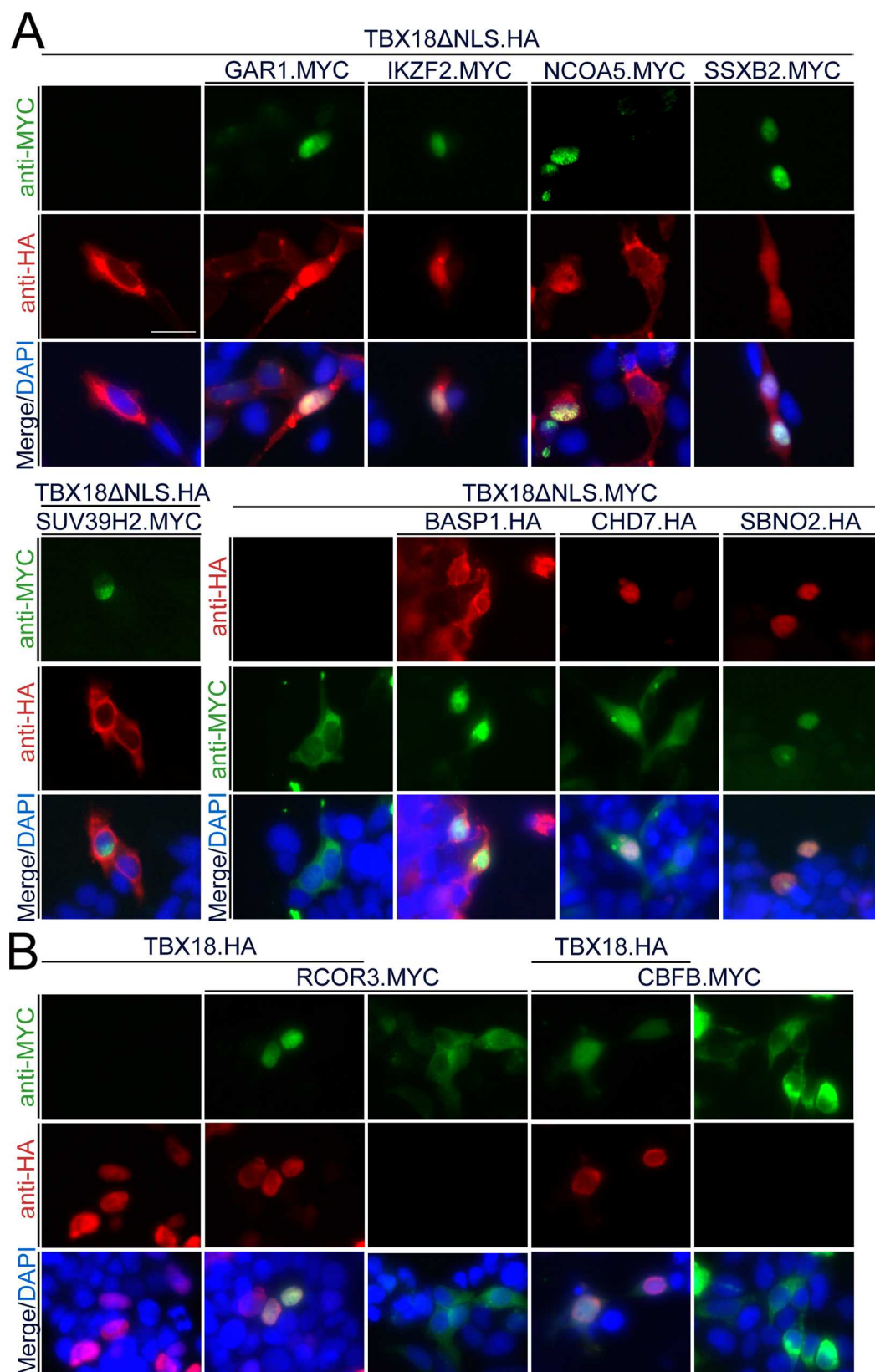


Fig 2. TBX18 colocalizes with most of the candidate transcriptional cofactors in the nucleus of 293 cells. (A) 293 cells were transfected with expression constructs for MYC-tagged GAR1, IKZF2, NCOA5, SSXB2 or SUV39H2 (green) in the presence of HA-tagged TBX18 lacking the nuclear localization signal (TBX18ΔNLS, red), or with expression constructs for HA-tagged BASP1, CHD7 or SBNO2 (red) in the presence of MYC-tagged TBX18ΔNLS (green). Immunofluorescence analysis shows that NLS-deficient TBX18 protein is efficiently shuttled from the cytoplasm to the nucleus by all candidate proteins except SUV39H2. (B) 293 cells were transfected with expression constructs for MYC-tagged RCOR3 or CBFB (green) in the presence of HA-tagged full-length TBX18 protein (TBX18ΔNLS, red). Immunofluorescence analysis shows that TBX18 protein recruits RCOR3 and CBFB from the cytoplasm into the nucleus. DAPI, 4,6-diamidino-2-phenylindole nuclear counterstain. Scale bar length is 25 μm.

<https://doi.org/10.1371/journal.pone.0200964.g002>

MYC- or HA tagged candidate proteins (S5A Table). Since GST-TBX18 full-length protein could not be expressed in bacteria, we used a previously described series of bacterially expressed fusion proteins of GST with the N- and C-terminal region and the T-box of TBX18 (S3 Fig) which simultaneously allowed to delimit the interaction domain with TBX18 [21]. Under the chosen experimental conditions, NKX2.5 bound well to the T-box of TBX18 as previously reported [21]. From the candidate cofactors, BASP1, RCOR3, SSXB2 and SUV39H2 did not bind to TBX18. CBFB bound to the N-, and C-terminal regions and the T-box, GAR1 to the T-box and weakly to the N-terminal region, IKZF2 and SBNO2 bound to the T-box, NCOA5 to the N-terminal region and more weakly to the T-box, the large CHD7 protein bound to the N- and T-box regions of TBX18 (Fig 3A).

Since 293 cell lysates represent a highly complex protein mix, it is conceivable that binding of the cofactors to TBX18 is mediated by additional proteins from the lysates. Furthermore, proteins from the lysate may preferentially bind to TBX18 thereby masking binding sites for our candidate cofactors. To address these concerns, we used reticulocyte lysates as an alternative and less complex source for the synthesis of the candidate cofactors (S5B Table). We labeled the proteins by addition of ³⁵S-Methionin to the *in vitro* translation reaction, and visualized their interaction with GST-TBX18 fusion proteins by autoradiographic imaging after SDS-PAGE of pulled-down complexes. Again, we did not find binding of BASP1, RCOR3, SSXB2 and SUV39H2 to TBX18 protein fragments. CBFB bound to the T-box and weakly to the C-terminal region, GAR1, IKZF2, and SBNO2 bound to the T-box of TBX18, NCOA5 and a subfragment of CHD7 ranging from 1533–2380 [36] bound to the T-box and weakly to the N-terminal region (Fig 3B). Together, these pull-down assays identify the T-box as the TBX18 subregion with which GAR1, IKZF2, SBNO2, NCOA5 and CHD7 strongly and most likely directly interact.

Some of transcriptional cofactors influence TBX18 transcriptional activity *in vitro*

We next performed luciferase reporter assays in 293 cells to analyze whether and how the candidate cofactors modulate TBX18 transcriptional activity. Cotransfection of an expression plasmid of TBX18 and a reporter plasmid in which we cloned a palindromic repeat of two TBEs which we previously described [21] in front of an SV40 minimal promoter and a firefly luciferase gene, resulted in 50% repression of luciferase activity compared to an empty expression vector. Cotransfection of increasing amounts of expression plasmids for BASP1 and GAR1 did not affect repression by TBX18. CBFB and CHD7 led to a weakly enhanced repression at least at one of the chosen concentrations. IKZF2 addition resulted in a dose-dependent increase of repression. In contrast, RCOR3 slightly relieved repression, while NCOA5, SBNO2 and SSXB2 dose-dependently counteracted the repressive activity of TBX18 (Fig 4A, S6 Table). Together, these assays identify CBFB, CHD7 and IKZF2 as transcriptional cofactors that bind the T-box of TBX18 directly and enhance (at least moderately) transcriptional repression of TBX18, while NCOA5 and SBNO2 bind but relieve repression (Fig 4B).

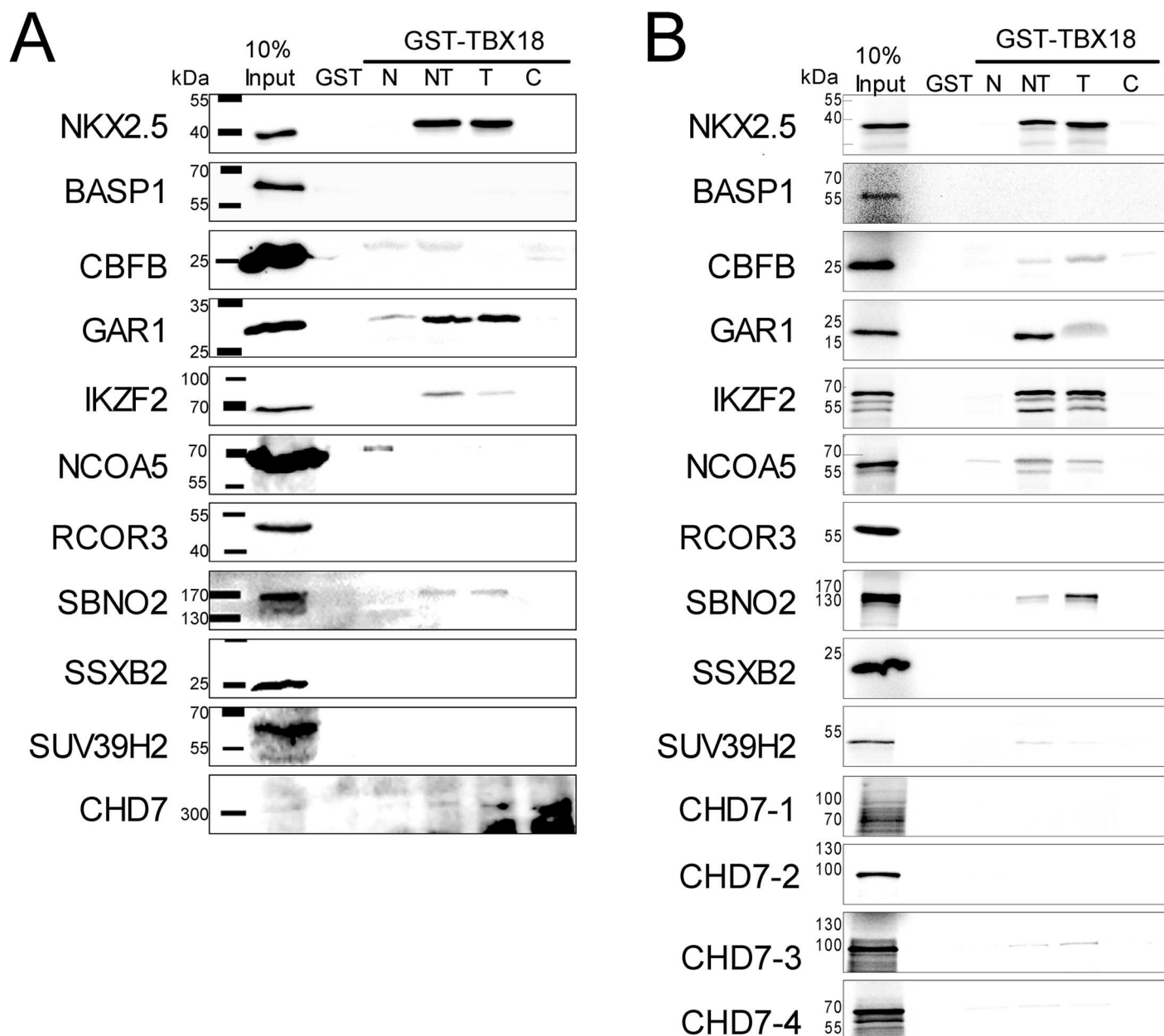


Fig 3. TBX18 physically interacts with candidate transcriptional cofactors preferentially via its T-box. (A) Western blot analysis of pull-down assays performed with GST and fusion proteins of GST with N-, N+T (NT), T-, and C-domains of TBX18 obtained from *E. coli* extracts, and protein extracts from 293 cells transfected with MYC- or HA-tagged full-length expression constructs of candidate transcriptional cofactors. Detection was performed with anti-MYC immunohistochemistry for CBFB, GAR1, IKZF2, NCOA5, RCOR3, SSXB2 and SUV39H2, and anti-HA immunohistochemistry for NKX2.5, BASP1, SBNO2 and CHD7. (B) Autoradiographic analysis of pull-down assays performed with the same GST-TBX18 fusion proteins and reticulocyte lysates programmed for *in vitro* translation of ³⁵S-labelled full-length candidate transcriptional cofactors. CBFB, GAR1, IKZF2, SBNO2, NCOA5 and CHD7 interact with the T-box of TBX18 as does the control protein NKX2.5. Due to the large size of CHD7, subfragments were expressed by *in vitro* translation and used in the pull-down assay. CHD7-1, amino acid residues 1–799; CHD7-2, 732–1567; CHD7-3, 1533–2380; CHD7-4, 2325–2997.

<https://doi.org/10.1371/journal.pone.0200964.g003>

Homeobox transcription factors bind to the T-box of TBX18 and are coexpressed during ureter development

Our proteomic screen identified 13 tissue-specific transcription factors that all featured a homeobox as DNA-binding region (Fig 1B). To validate the binding of these proteins to

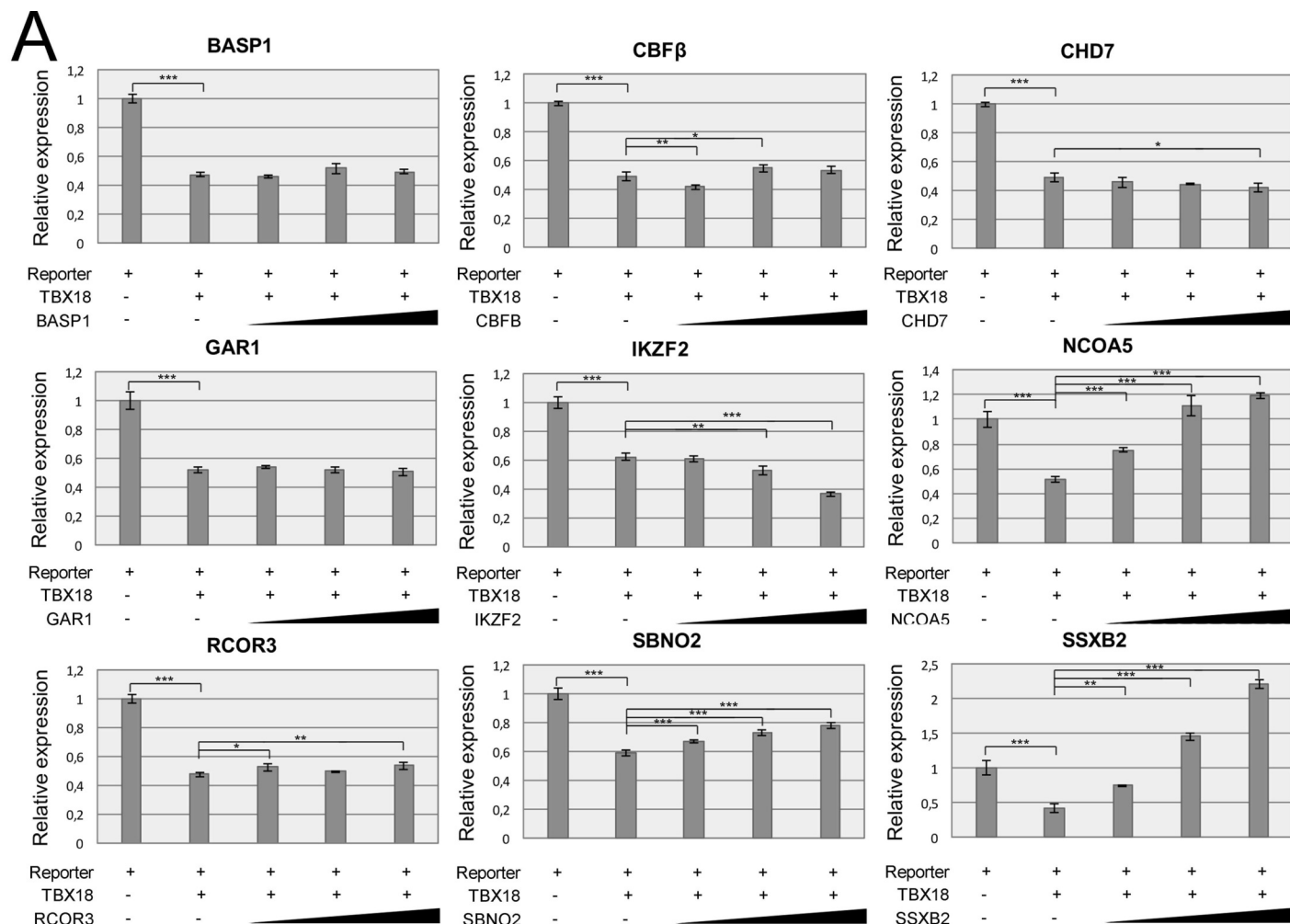


Fig 4. Candidate transcriptional cofactors modify the transcriptional repression activity of TBX18. (A) Luciferase assays from extracts of 293 cells cotransfected with 250 ng of the *pGL3.Prom.Tbx18* reporter plasmid, 250 ng of the TBX18 expression plasmid *pcDNA3.Tbx18* and 25, 250 or 500 ng of plasmids for expression of the transcriptional cofactors. Luciferase activity is normalized to a cotransfection of the reporter plasmid with an empty *pcDNA3* expression vector. TBX18 represses luciferase activity to about 50% of the control value. CBFB, CHD7, IKZF2, RCOR3 further augment repression by TBX18; NCOA5, SBNO2 and SSXB2 relieve TBX18 repression activity. Values are displayed as mean $\pm$ sd. * $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$; two-tailed Student's t-test. (For statistical values see S6 Table). (B) Summary of TBX18 binding and activation assays for transcriptional cofactor candidates. Three cofactors (CBFB, CHD7, IKZF2) reliably interact with the T-box of TBX18 and repress transcription. Two cofactors interact with the TBX18 T-box but relieve repression.

<https://doi.org/10.1371/journal.pone.0200964.g004>

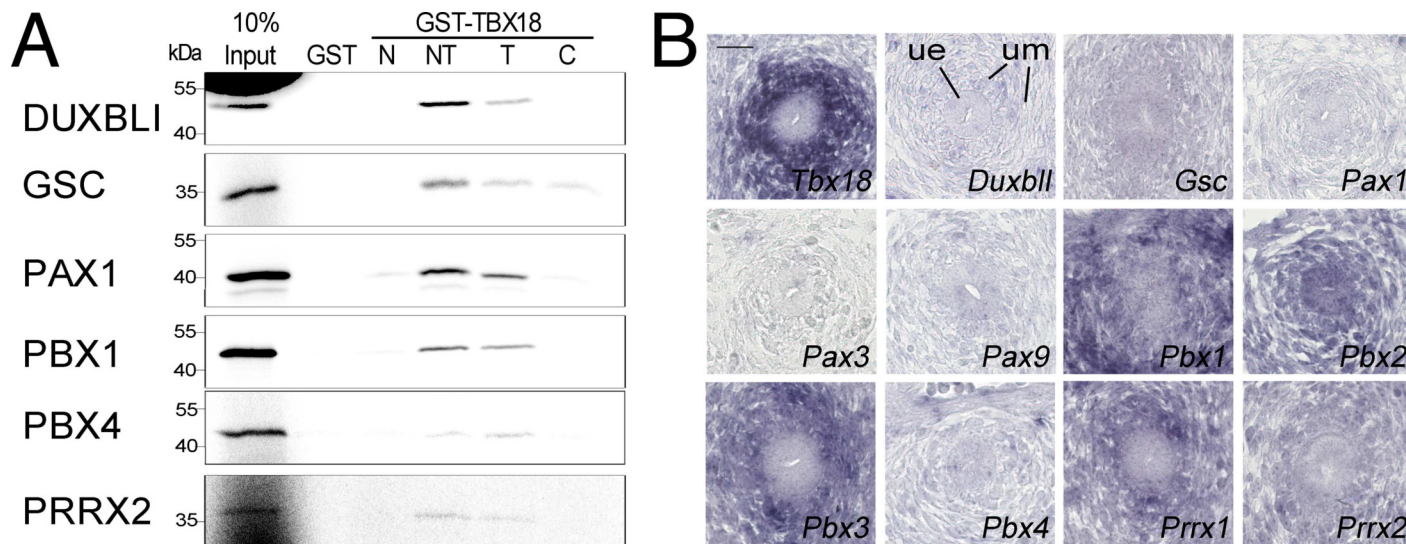


Fig 5. Homeobox transcription factors directly bind to the T-box of TBX18 and are coexpressed with *Tbx18* during ureter development. (A) Autoradiographic analysis of pull-down assays of GST and of fusion proteins of GST with N-, N+T (NT)-, T-, and C-domains of TBX18 from *E. coli* extracts, and reticulocyte lysates programmed for *in vitro* translation of ³⁵S-labelled full-length candidate transcriptional cofactors. DUXBL1, GSC, PAX1, PBX1, PBX4 and PRRX2 interact with the T-box of TBX18. (B) Comparative RNA *in situ* hybridization analysis on transverse sections of the proximal ureter of E12.5 embryos of *Tbx18* and of genes encoding homeobox transcription factor candidates and related subfamily members. ue, ureteric epithelium, um, ureteric mesenchyme. Scale bar length is 25 μ m.

<https://doi.org/10.1371/journal.pone.0200964.g005>

TBX18, we used the pull-down assay with GST-TBX18 fusion proteins and reticulocyte lysates programmed for the translation of radioactively labeled full-length proteins (S5C Table). Strikingly, all tested homeobox transcription factors (DUXBL1, the unique homologous mouse gene for the human DUX family, GSC, PAX1, PBX1, PBX4, PRRX2) bound to the T-box of TBX18. GSC additionally weakly interacted with the C-terminal region (Fig 5A). To determine a possible *in vivo* relevance of this finding, we analyzed whether the encoding genes as well as related members of the specific subfamily are coexpressed with *Tbx18* in the developing ureter. *In situ* hybridization on sections of E12.5 mouse ureters showed that *Pbx1*, *Pbx2*, *Pbx3*, *Prrx1* and *Prrx2* are coexpressed with *Tbx18* in the mesenchymal compartment of the organ, holding the promise of a functional interaction in ureter development (Fig 5B).

TBX18 does not bind candidate proteins at the centrosome

To evaluate the possibility of direct interactions between TBX18 and the identified centrosomal proteins (Fig 1B), we performed GST pull-down assays again with *in vitro* translated proteins as described (S5D Table). Interaction was found between TBX18's T-box and MAP9, PARP3, SYBU and TPX2. HAUS8 bound to the N-terminal domain, while B9D2 and (weakly) SYBU additionally bound to the C-terminal domain (Fig 6A). We next checked whether in fact TBX18 colocalizes with and binds to some of these proteins in 293 cells. For this, we first transfected expression constructs for full-length epitope-tagged candidate proteins, confirmed protein integrity by Western blot and characterized their localization by immunofluorescence analysis against their tags (S5E Table, S4 Fig). To our surprise, we did not detect any centrosomal accumulations but found that B9D2 and TPX2 localized to the nucleus, HAUS8 and ASAP to the cytoplasm, and PARP3 to cellular protrusions (Fig 6B). Upon coexpression of TBX18 Δ NLS with B9D2 and TPX2, we observed a nuclear translocation of TBX18 Δ NLS. In contrast, full-length TBX18 was not able to recruit HAUS8, ASAP and PARP3 to the nucleus (Fig 6C). Finally, we wished to ascertain that TBX18 does not localize to the centrosome. For

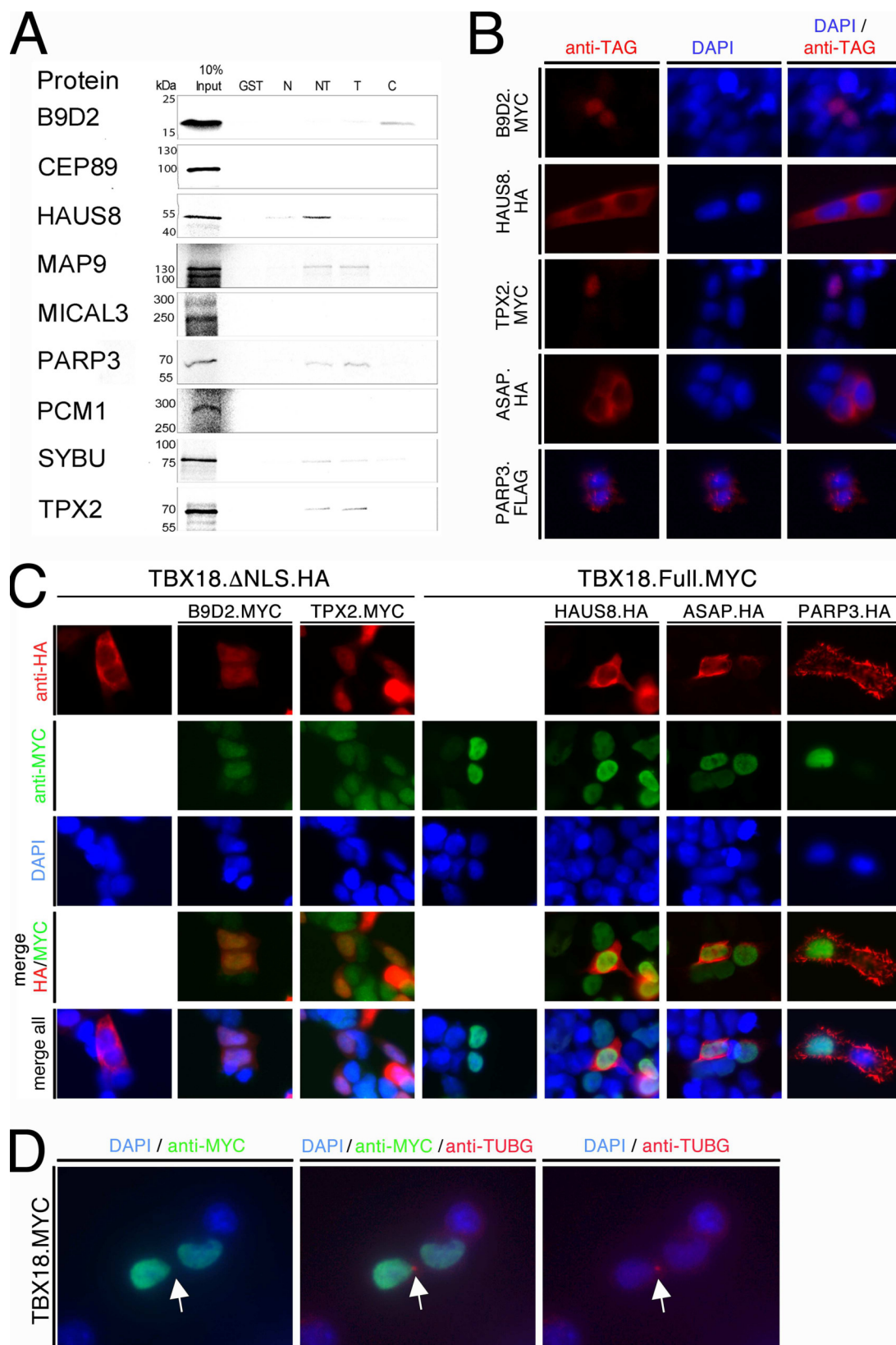


Fig 6. TBX18 does not bind to candidate interaction partners at the centrosome. (A) Autoradiographic analysis of pull-down assays of GST and fusion proteins of GST with N-, N+T- (NT), T-, and C-domains of TBX18 from *E. coli* extracts, and reticulocyte lysates programmed for *in vitro* translation of ³⁵S-labelled full-length candidate centrosomal proteins. B9D2, HAUS8, MAP9, PARP3, SYBU and TPX2 interact with the T-box of TBX18; B9D2 and SYBU additionally bind to the C-, HAUS8 to the N-terminal domain. (B) Immunofluorescent staining against epitope tags of candidate centrosomal proteins expressed in 293 cells reveals nuclear localization of B9D2 and TPX2, cytoplasmic localization of HAUS8 and ASAP, and confinement of PARP3 to cellular protrusions. (C) 293 cells were transfected with expression constructs for MYC-tagged B9D2 and TPX2 (*green*) in the presence of HA-tagged TBX18 lacking the nuclear localization signal (TBX18ΔNLS, *red*), or with expression constructs for HA-tagged HAUS8, ASAP and PARP3 (*red*) in the presence of MYC-tagged full-length TBX18 (*green*). Immunofluorescence analysis shows that NLS-deficient TBX18 protein is efficiently shuttled from the cytoplasm to the nucleus by B9D2 and TPX2 while the extranuclear localization of HAUS8, ASAP and PARP3 is unaffected by coexpression of full-length TBX18. (D) 293 cells were transfected with an expression construct for MYC-tagged TBX18 (*green*). Co-immunofluorescence analysis shows that TBX18 protein does not localize to the centrosome marked by TUBG expression (*red*) but to the nucleus. Scale bar length is 25 μm.

<https://doi.org/10.1371/journal.pone.0200964.g006>

this, we transfected an expression construct for full-length MYC-tagged TBX18 in 293 cells and performed co-immunofluorescence analysis of the MYC tag and the centrosomal marker protein TUBG (gamma-Tubulin) [37]. TBX18 protein was not observed at the centrosome but exclusively localized to the nucleus (Fig 6D). We conclude that TBX18 is able to interact with some of the proteins annotated as “centrosomal” in 293 cells (B9D2, TPX2) but that this interaction occurs in the nucleus.

Discussion

The tandem affinity purification (TAP)-Tag method has been successfully used to identify interactome members of different proteins in 293 cells [38, 39]. Here, we used this approach with dual tagged overexpressed TBX18 and identified protein interaction partners of TBX18 in 293 cells. Our results show that TBX18 preferentially binds to chromatin and interacts with transcriptional cofactors (preferentially corepressors) and homeobox transcription factors. Binding to proteins annotated as “centrosomal” may reflect an unknown nuclear function of these proteins rather than a novel centrosomal role of TBX18.

We have previously shown that TBX18 interacts with members of the TLE (Groucho) family via a conserved eh1-motif and that this interaction accounts for half of the repressive activity of TBX18 in transactivation assays *in vitro* [21]. TBX18 protein with loss of the C-terminal region and unchanged eh1-motif from a CAKUT patient exhibited reduced repressive activity *in vitro* [24] indicating that additional corepressors exist both in 293 cells as well as *in vivo* and that they may bind to the large C-terminal region of TBX18 to augment its repressive activity. Amongst 10 transcriptional cofactors identified in our proteomic screen, we validated robust and probably direct binding of 6 proteins (CBFB, CHD7, GAR1, IKZF2, NCOA5, SBNO2) to TBX18. Only CBFB interacted weakly with the C-terminal region, while all proteins bound to the T-box. This result was somehow disappointing to us since we had hoped to identify a C-terminally binding repressor that may explain the human disease mechanism. It is important to note that our proteomic approach did not reach saturation since we only detected a single protein in more than one of the three independent IP experiments. The list of cofactors and interactors, respectively, is therefore unlikely to be complete but may represent cofactors that bind relatively strongly to TBX18. Alternatively, the cofactor that operates in the embryonic mesenchyme of the ureter may simply be not present in 293 cells. Given the finding that both our candidate cofactors and the identified homeobox transcription factors interacted with the T-box, the argument may arise that the identified cofactors may bind indirectly e.g. via the homeobox factors to TBX18. While this may be well the case in cells, we assume that the interaction between GTS-TBX18 fusion proteins from bacteria and *in vitro* translated proteins from reticulocytes lacks such mediators, and therefore, is direct.

We identified three cofactors that bound to TBX18 and variably enhanced the repressive activity of TBX18: CBFB, CHD7 and IKZF2. CBFB represents the beta subunit of the core-binding factor protein. The alpha subunit is encoded by one of three RUNX proteins. The complex acts as a transcription factor with CBFB increasing the affinity of the alpha subunit for DNA and enhancing transcriptional activation by RUNX proteins [40]. Compatible with this function CBFB only poorly affected the transcriptional repression activity of TBX18.

CHD7 is a member of the Chromodomain Helicase DNA-binding (CHD) protein family that acts as ATP-dependent chromatin remodeler [41]. CHD7 was shown to bind to TBX20, a protein closely related to TBX18 [42], and to genetically interact with *TBX1*, another related family member in the development of the great vessels [43]. Interestingly, a mutation in amino acid residue 1684 of CHD7, was identified as a penetrant cause for CAKUT [44], and CHD7 was found to bind to CBFB through RUNX1 to repress RUNX1-induced genes [45]. Additionally, CHD4, another CHD family member that is a part of the Mi-2/NuRD (Nucleosome Remodeling Deacetylase) complex [46, 47] interacts with IKZF2 extensively during blood cell development [48–50]. IKZF2 is a member of the Ikaros Zn-finger containing DNA-binding protein family (thus not a classical cofactor) that additionally interacts with HDACs [51]. Together with the enhanced repressive activity of TBX18 by IKZF2, this suggests that TBX18 forms a complex with members of the CHD family (CHD4, CHD7 or others), the Ikaros family, CBFB and HDACs to compact chromatin and deacetylate histones for repression.

SBNO2 forms together with SBNO1 the small family of DExD/H helicase Strawberry notch-like proteins. SBNO2 can repress transcription of genes [52] but also activate [53] indicating a context-dependent role as transcriptional cofactor. Similarly, NCOA5, a factor that regulates nuclear receptor transcriptional activity, was shown to negatively and positively impact on gene expression [54, 55]. In our transcription assays both proteins relieved the transcriptional repression by TBX18 indicating that they act as transcriptional activators. This does not exclude a role as a repressor *in vivo*, given the fact that differentially spaced and oriented TBEs can translate into different transcriptional responses [4].

GAR1 is a component of box H/ACA ribonucleoproteins (RNPs), protein-RNA complexes responsible for post-transcriptional modification of cellular RNAs, splicing and ribosome synthesis [56] explaining the lack of this protein to affect transcription. The possible significance of TBX18 interaction remains unclear.

Previous work showed that *in vitro* TBX18 interacts with at least three different classes of tissue specific transcription factors, namely Zn-finger proteins (GATA4), T-box proteins (TBX15, TBX18) and homeobox proteins (NKX2.5, PAX3 and SIX1) [21–23]. We were surprised to identify in our proteomic screen in 293 only homeobox containing transcription factors. All of them interacted with the T-box of TBX18 in pull-down assays. Although this may indicate a general promiscuity of T-box interaction with this class of proteins, it seems likely that certain subfamilies of homeobox proteins are preferred interaction partners *in vivo* as well. Our screen identified as a binding partner, PAX1, a member of the paired-type homeobox transcription factors [57]. Interestingly, we have previously shown that another member of this family, PAX3 binds to TBX18, and that loss-of-function alleles of *Pax3* and *Tbx18* genetically interact in limb girdle development [22]. Although *Pax1* is not coexpressed with *Tbx18* in the ureteric mesenchyme, *Pax1*, *Pax9* or other family members may bind to *TBX18* in other embryological contexts including the somitic mesoderm in which they are partially coexpressed [17].

We also identified two of the four members of the PBX subfamily (PBX1, PBX4) as TBX18 binding partners *in vitro* [58]. Three members of this subfamily (*Pbx1*, *Pbx2*, *Pbx3*) are coexpressed with *Tbx18* in the ureteric mesenchyme. Importantly, a set of mutations recently identified *Pbx1* as monogenic cause of CAKUT making a functional interaction with TBX18 in

ureter development likely [44]. Other candidates for functional interaction are the two *Prrx* genes that are coexpressed with *Tbx18* in the ureteric mesenchyme but also in the developing limb [59].

Since a recent report ascribed an additional centrosomal function to the transcription factor ATF [35], we were excited that functional annotation and literature mining identified a cluster of centrosomal proteins as binding candidates of TBX18 in our proteomic screen in 293 cells. We validated direct binding of 6 of these candidates to the T-box of TBX18, and confirmed interaction of two of them, B9D2 and TPX2, with TBX18 in 293 cells. To our surprise, B9D2, which was reported as a ciliary transition zone protein [60], was mainly located in the cytoplasm of 293 cells, while TPX2 a microtubule binding protein that is necessary for spindle assembly was found in the nucleus. Interestingly, it was recently reported that TPX2 resides preferentially in the nucleus during interphase where it plays a role in amplification of the DNA damage response [61, 62]. Since TBX18 does not localize to the centrosome even when overexpressed in 293 cells, we deem a centrosomal function of TBX18 unlikely, but suggest that ciliary or microtubule-binding proteins may be shuttled to the nucleus to affect the function of transcription factors such as TBX18.

Together, our proteomic screen strongly supports an exclusive nuclear function of TBX18 as a repressor of chromatin accessibility. At least in 293 cells, this function is likely mediated by recruiting chromatin remodeling complexes of the CHD family and histone modifying complexes including HDACs in concert with other transcription factors, most likely of the homeobox family. Whether these complexes are conserved at the different sites of TBX18 expression in the embryo, will be a demanding yet exciting task to find out for genetic interaction screens and proteomic approaches with endogenously tagged TBX18 in the future.

Supporting information

S1 Fig. *Tbx18* is expressed in 293 cells but not in other tested cell lines. Semi-quantitative RT-PCR analysis of *TBX18* expression in 293, SK-N-MC, SW-13, Huh-7.5, SH-SY5Y and 786-O cells. *GAPDH* was used as a housekeeping control and all values were calculated as relative gene per *GAPDH* ratios. 293 cells exhibit a strong expression of *TBX18* with a mean $\pm$ SD ratio of 1.1408 ± 0.2394 . In all other tested cell lines *TBX18* expression was not above background level.

(TIF)

S2 Fig. Expression and localization of identified transcriptional cofactor candidates in 293 cells. (A) Western Blot analysis of over-expressed co-factors in 293 cells. Expression constructs (as listed in S5A Table) were transfected into 293 cells and exogenous proteins were detected using antibodies against the corresponding tags. All detected proteins were of the expected size except BASP1, which as previously appeared larger. (B) Immunofluorescence analysis of localization of candidate transcriptional cofactors after transfection of 293 cells with the expression plasmids listed in (A). BASP1, CHD7, GAR1, IKZF2, NCOA5, SBNO2, SSXB2, SUV39H2 localized to the nucleus, CBFB and RCOR3 to the cytoplasm. Scale bar length is 25 μ m.

(TIF)

S3 Fig. GST-TBX18 fusion proteins for pull-down assays. (A) Schematic representation of the primary structure of TBX18, and of the subfragments used to express GST fusion proteins. The T-box (T) is shaded in orange, and the N- and C-terminal domains (N and C) are shown in grey. The numbers refer to the amino acid position in the full-length TBX18 protein. The localization of the Groucho binding region (eh1), the nuclear localization signal (NLS) are highlighted in the N-terminal domain. (B) GST and fusion proteins of GST and N-, N+T-, T-

and C-domains of TBX18 were purified from *E. coli* extracts and analyzed for integrity and quantity by Coomassie Brilliant Blue staining of SDS-polyacrylamide gels. Asterisks mark the full-length proteins.

(TIF)

S4 Fig. Western Blot analysis of over-expressed “centrosomal” candidate proteins in 293 cells. Expression constructs (as listed in [S4E Table](#)) were transfected into 293 cells and exogenous proteins were detected using antibodies against the corresponding tags.

(TIF)

S1 Table. Summary of cloning strategies for expression plasmids.

(XLSX)

S2 Table. Mass spectrometry analysis identifies proteins copurified with TBX18 from 293 cells. List of 143 proteins detected by mass spectrometry in extracts of proteins copurified with TBX18 in 293 cells. Shown are Uniprot accession code (first row), the protein symbol (second row) and the protein name (third row). Color code identifies the three major protein clusters identified: transcription factors of the homeobox family (grey), transcriptional cofactors (green) and centrosomal proteins (yellow).

(XLSX)

S3 Table. DAVID analysis identifies 83 enriched functional annotations in the set of 143 proteins identified as possible TBX18 interaction partners in 293 cells. From the 143 identified proteins we excluded immunoglobulins, histone proteins and cytoskeletal proteins which leads to 83 possible interaction partners.

(XLSX)

S4 Table. DAVID analysis identifies 5 enriched clusters of TBX18 interacting proteins. All 143 identified proteins were subjected to a Functional Annotation Clustering (DAVID Bioinformatics Resources 6.8) using the following settings: Similarity Term Overlap 7; Similarity Threshold 0.35; Initial Group Membership 2; Final Group Membership 2; Multiple Linkage Threshold 0.15; EASE 0.3; UP_KEYWORDS and GOTHERM_CC_DIRECT. 5 clusters were identified with enrichment scores for cluster 1 of 6.07, for cluster 2 of 3.48, for cluster 3 of 1.85, for cluster 4 of 1.35 and cluster 5 of 1.19.

(XLSX)

S5 Table. Summary of expression of candidate proteins for TBX18 interaction in various assays. (A) Cloning and expression of candidate transcriptional cofactors in 293 cells. Shown are the name of the proteins, the species they are derived from (m, mouse; h, human), the expected molecular weight (in kDa), the plasmids used for expression of the proteins in 293 cells, the antibodies used to detect the tag on the candidate proteins, and the reference for the used plasmids. (B-D) Cloning and *in vitro* translation of candidate transcriptional cofactors (B), homeobox transcription factors (C) and centrosomal proteins (D). Shown are the name of the proteins, the species they are derived from (m, mouse; h, human), the expected molecular weight (in kDa), the plasmids used for expression of the proteins in reticulocyte lysate, the polymerase used for the *in vitro* transcription reaction, and the reference for the used plasmids. (E) Cloning and expression of candidate “centrosomal” proteins in 293 cells. Shown are the name of the proteins, the species they are derived from (m, mouse; h, human), the expected molecular weight (in kDa), the plasmids used for expression of the proteins in 293 cells, the antibodies used to detect the tag on the candidate proteins, and the reference for the used plasmids.

(XLSX)

S6 Table. Statistical analysis of transactivation assays of candidate transcriptional cofactors with TBX18 in 293 cells. Shown are the experimental setup for each sample of the transactivation assays and the statistical analysis performed to the transactivation assay results. Raw luminescence data was normalized to the average of the reporter alone measurements. Further, average and standard deviation were quantified from the two technical and two experimental replicates. The significance of luciferase expression between the TBX18 induced repression and the reporter and the effects of the cofactors on TBX18 repression was evaluated by a student t-test (TTEST) carried out according to an analysis of variance equality (FTEST). p-values<0.05 (*) were considered significant, while p-values<0.01 (**) and p-values<0.0001 (***) were considered very significant. (XLSX)

Acknowledgments

For plasmid gifts, we would like to thank Prof. Pauli Silke from Universitätsmedizin Göttingen, Dr. Peter Murray from St. Jude's Children Research Hospital in Memphis, Dr. Kai Jiao from the University of Alabama in Birmingham, Dr. Shireen Saleque from the City College of New York, Dr. Stefan Roberts from the University of Bristol, Dr. Kenta Maruyama from Osaka University, Dr. Anna Akhmanova from Utrecht University, Dr. Sylvie Rouquier from the Centre National de la Recherche Scientifique in Montpellier, Dr. Stuart Rulten from the University of Sussex, Dr. James Sillibourne from the Hopital Saint Louis Paris, Laurence Pelletier from Mount Sinai Hospital in Toronto and Dr. Andrew P. McMahon from the University of Southern California. We thank Thomas v. Hahn from Hannover Medical School for the donation of cell lines. We would like to acknowledge the assistance of the MS Core Facility Proteomics at the Medizinische Hochschule Hannover.

Author Contributions

Conceptualization: Marc-Jens Kleppa, Andreas Kispert.

Data curation: Marc-Jens Kleppa, Andreas Kispert.

Formal analysis: Reginaldo Rivera-Reyes, Marc-Jens Kleppa.

Funding acquisition: Andreas Kispert.

Investigation: Reginaldo Rivera-Reyes, Marc-Jens Kleppa.

Methodology: Reginaldo Rivera-Reyes, Marc-Jens Kleppa.

Project administration: Andreas Kispert.

Supervision: Marc-Jens Kleppa, Andreas Kispert.

Validation: Reginaldo Rivera-Reyes, Marc-Jens Kleppa.

Visualization: Reginaldo Rivera-Reyes, Marc-Jens Kleppa.

Writing – original draft: Reginaldo Rivera-Reyes, Andreas Kispert.

Writing – review & editing: Reginaldo Rivera-Reyes, Marc-Jens Kleppa, Andreas Kispert.

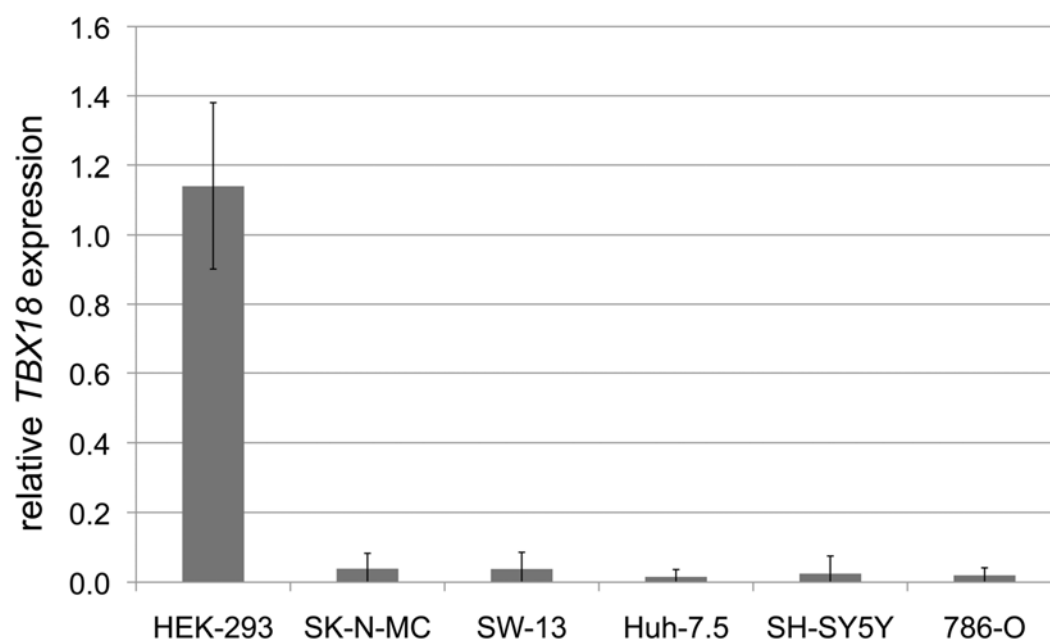
References

1. Bollag RJ, Siegfried Z, Cebra-Thomas JA, Garvey N, Davison EM, Silver LM. An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T locus. *Nat Genet.* 1994; 7: 383–9. <https://doi.org/10.1038/ng0794-383> PMID: 7920656

2. Agulnik SI, Bollag RJ, Silver LM. Conservation of the T-box gene family from *Mus musculus* to *Caenorhabditis elegans*. *Genomics*. 1995; 25: 214–9. PMID: [7774921](#)
3. Kispert A, Herrmann BG. The Brachyury gene encodes a novel DNA binding protein. *EMBO J*. 1993; 12: 3211–20. PMID: [8344258](#)
4. Kispert A, Koschorz B, Herrmann BG. The T protein encoded by Brachyury is a tissue-specific transcription factor. *EMBO J*. 1995; 14: 4763–72. PMID: [7588606](#)
5. Naiche LA, Harrelson Z, Kelly RG, Papaioannou VE. T-box genes in vertebrate development. *Annu Rev Genet*. 2005; 39: 219–39. <https://doi.org/10.1146/annurev.genet.39.073003.105925> PMID: [16285859](#)
6. Papaioannou VE. The T-box gene family: emerging roles in development, stem cells and cancer. *Development*. 2014; 141: 3819–33. <https://doi.org/10.1242/dev.104471> PMID: [25294936](#)
7. Ghosh TK, Brook JD, Wilsdon A. T-Box Genes in Human Development and Disease. *Curr Top Dev Biol*. 2017; 122: 383–415. <https://doi.org/10.1016/bs.ctdb.2016.08.006> PMID: [28057271](#)
8. Sebe-Pedros A, Ruiz-Trillo I. Evolution and Classification of the T-Box Transcription Factor Family. *Curr Top Dev Biol*. 2017; 122: 1–26. <https://doi.org/10.1016/bs.ctdb.2016.06.004> PMID: [28057261](#)
9. Kraus F, Haenig B, Kispert A. Cloning and expression analysis of the mouse T-box gene *Tbx18*. *Mech Dev*. 2001; 100: 83–6. PMID: [11118889](#)
10. Bush JO, Lan Y, Maltby KM, Jiang R. Isolation and developmental expression analysis of *Tbx22*, the mouse homolog of the human X-linked cleft palate gene. *Dev Dyn*. 2002; 225: 322–6. <https://doi.org/10.1002/dvdy.10154> PMID: [12412015](#)
11. Herr A, Meunier D, Muller I, Rump A, Fundele R, Ropers HH, et al. Expression of mouse *Tbx22* supports its role in palatogenesis and glossogenesis. *Dev Dyn*. 2003; 226: 579–86. <https://doi.org/10.1002/dvdy.10260> PMID: [12666195](#)
12. Singh MK, Petry M, Haenig B, Lescher B, Leitges M, Kispert A. The T-box transcription factor *Tbx15* is required for skeletal development. *Mech Dev*. 2005; 122: 131–44. <https://doi.org/10.1016/j.mod.2004.10.011> PMID: [15652702](#)
13. Airik R, Bussen M, Singh MK, Petry M, Kispert A. *Tbx18* regulates the development of the ureteral mesenchyme. *J Clin Invest*. 2006; 116: 663–74. <https://doi.org/10.1172/JCI26027> PMID: [16511601](#)
14. Christoffels VM, Mommersteeg MT, Trowe MO, Prall OW, de Gier-de Vries C, Soufan AT, et al. Formation of the venous pole of the heart from an *Nkx2-5*-negative precursor population requires *Tbx18*. *Circ Res*. 2006; 98: 1555–63. <https://doi.org/10.1161/01.RES.0000227571.84189.65> PMID: [16709898](#)
15. Trowe MO, Maier H, Schweizer M, Kispert A. Deafness in mice lacking the T-box transcription factor *Tbx18* in otic fibrocytes. *Development*. 2008; 135: 1725–34. <https://doi.org/10.1242/dev.014043> PMID: [18353863](#)
16. Bolt CC, Negi S, Guimaraes-Camboa N, Zhang H, Troy JM, Lu X, et al. *Tbx18* Regulates the Differentiation of Periductal Smooth Muscle Stroma and the Maintenance of Epithelial Integrity in the Prostate. *PLoS One*. 2016; 11: e0154413. <https://doi.org/10.1371/journal.pone.0154413> PMID: [27120339](#)
17. Bussen M, Petry M, Schuster-Gossler K, Leitges M, Gossler A, Kispert A. The T-box transcription factor *Tbx18* maintains the separation of anterior and posterior somite compartments. *Genes Dev*. 2004; 18: 1209–21. <https://doi.org/10.1101/gad.300104> PMID: [15155583](#)
18. Wiese C, Grieskamp T, Airik R, Mommersteeg MT, Gardiwal A, de Gier-de Vries C, et al. Formation of the sinus node head and differentiation of sinus node myocardium are independently regulated by *Tbx18* and *Tbx3*. *Circ Res*. 2009; 104: 388–97. <https://doi.org/10.1161/CIRCRESAHA.108.187062> PMID: [19096026](#)
19. Greulich F, Farin HF, Schuster-Gossler K, Kispert A. *Tbx18* function in epicardial development. *Cardiovasc Res*. 2012; 96: 476–83. <https://doi.org/10.1093/cvr/cvs277> PMID: [22926762](#)
20. Wu SP, Dong XR, Regan JN, Su C, Majesky MW. *Tbx18* regulates development of the epicardium and coronary vessels. *Dev Biol*. 2013; 383: 307–20. <https://doi.org/10.1016/j.ydbio.2013.08.019> PMID: [24016759](#)
21. Farin HF, Bussen M, Schmidt MK, Singh MK, Schuster-Gossler K, Kispert A. Transcriptional repression by the T-box proteins *Tbx18* and *Tbx15* depends on Groucho corepressors. *J Biol Chem*. 2007; 282: 25748–59. <https://doi.org/10.1074/jbc.M703724200> PMID: [17584735](#)
22. Farin HF, Mansouri A, Petry M, Kispert A. T-box protein *Tbx18* interacts with the paired box protein *Pax3* in the development of the paraxial mesoderm. *J Biol Chem*. 2008; 283: 25372–80. <https://doi.org/10.1074/jbc.M802723200> PMID: [18644785](#)
23. Nie X, Sun J, Gordon RE, Cai CL, Xu PX. *SIX1* acts synergistically with *TBX18* in mediating ureteral smooth muscle formation. *Development*. 2010; 137: 755–65. <https://doi.org/10.1242/dev.045757> PMID: [20110314](#)

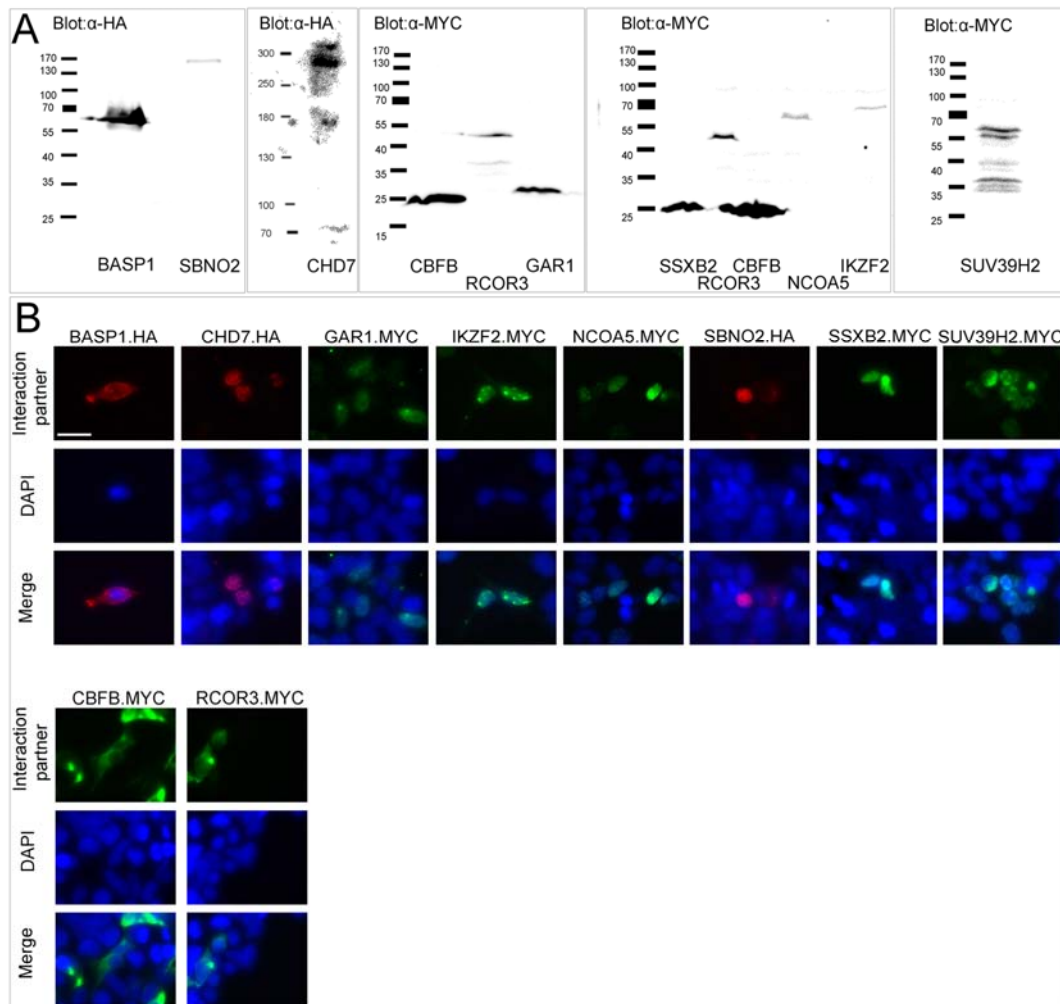
24. Vivante A, Kleppa MJ, Schulz J, Kohl S, Sharma A, Chen J, et al. Mutations in TBX18 Cause Dominant Urinary Tract Malformations via Transcriptional Dysregulation of Ureter Development. *Am J Hum Genet.* 2015; 97: 291–301. <https://doi.org/10.1016/j.ajhg.2015.07.001> PMID: 26235987
25. Tanimizu N, Tsujimura T, Takahide K, Kodama T, Nakamura K, Miyajima A. Expression of Dlk/Pref-1 defines a subpopulation in the oval cell compartment of rat liver. *Gene Expr Patterns.* 2004; 5: 209–18. <https://doi.org/10.1016/j.modgep.2004.08.003> PMID: 15567716
26. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012; 9: 671–5. PMID: 22930834
27. Zirzow S, Ludtke TH, Brons JF, Petry M, Christoffels VM, Kispert A. Expression and requirement of T-box transcription factors Tbx2 and Tbx3 during secondary palate development in the mouse. *Dev Biol.* 2009; 336: 145–55. <https://doi.org/10.1016/j.ydbio.2009.09.020> PMID: 19769959
28. O'Brien LL, Guo Q, Bahrami-Samani E, Park JS, Hasso SM, Lee YJ, et al. Transcriptional regulatory control of mammalian nephron progenitors revealed by multi-factor cistronic analysis and genetic studies. *PLoS Genet.* 2018; 14: e1007181. <https://doi.org/10.1371/journal.pgen.1007181> PMID: 29377931
29. Pear WS, Nolan GP, Scott ML, Baltimore D. Production of high-titer helper-free retroviruses by transient transfection. *Proc Natl Acad Sci U S A.* 1993; 90: 8392–6. PMID: 7690960
30. Wang J, Cantor AB, Orkin SH. Tandem affinity purification of protein complexes in mouse embryonic stem cells using in vivo biotinylation. *Curr Protoc Stem Cell Biol.* 2009;Chapter 1: Unit1B 5.
31. Jochim N, Gerhard R, Just I, Pich A. Impact of clostridial glucosylating toxins on the proteome of colonic cells determined by isotope-coded protein labeling and LC-MALDI. *Proteome Sci.* 2011; 9: 48. <https://doi.org/10.1186/1477-5956-9-48> PMID: 21849038
32. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* 2008; 26: 1367–72. <https://doi.org/10.1038/nbt.1511> PMID: 19029910
33. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4: 44–57. <https://doi.org/10.1038/nprot.2008.211> PMID: 19131956
34. Moorman AF, Houweling AC, de Boer PA, Christoffels VM. Sensitive nonradioactive detection of mRNA in tissue sections: novel application of the whole-mount in situ hybridization protocol. *J Histochem Cytochem.* 2001; 49: 1–8. <https://doi.org/10.1177/002215540104900101> PMID: 11118473
35. Madarampalli B, Yuan Y, Liu D, Lengel K, Xu Y, Li G, et al. ATF5 Connects the Pericentriolar Materials to the Proximal End of the Mother Centriole. *Cell.* 2015; 162: 580–92. <https://doi.org/10.1016/j.cell.2015.06.055> PMID: 26213385
36. Schulz Y, Freese L, Manz J, Zoll B, Volter C, Brockmann K, et al. CHARGE and Kabuki syndromes: a phenotypic and molecular link. *Hum Mol Genet.* 2014; 23: 4396–405. <https://doi.org/10.1093/hmg/ddu156> PMID: 24705355
37. Graf R, Euteneuer U, Ueda M, Schliwa M. Isolation of nucleation-competent centrosomes from Dictyostelium discoideum. *Eur J Cell Biol.* 1998; 76: 167–75. [https://doi.org/10.1016/S0171-9335\(98\)80031-9](https://doi.org/10.1016/S0171-9335(98)80031-9) PMID: 9716263
38. Gloeckner CJ, Kinkl N, Schumacher A, Braun RJ, O'Neill E, Meitinger T, et al. The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity. *Hum Mol Genet.* 2006; 15: 223–32. <https://doi.org/10.1093/hmg/ddi439> PMID: 16321986
39. den Hollander AI, Koenekoop RK, Mohamed MD, Arts HH, Boldt K, Towns KV, et al. Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis. *Nat Genet.* 2007; 39: 889–95. <https://doi.org/10.1038/ng2066> PMID: 17546029
40. Yoshida CA, Furuichi T, Fujita T, Fukuyama R, Kanatani N, Kobayashi S, et al. Core-binding factor beta interacts with Runx2 and is required for skeletal development. *Nat Genet.* 2002; 32: 633–8. <https://doi.org/10.1038/ng1015> PMID: 12434152
41. Bouazoune K, Kingston RE. Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders. *Proc Natl Acad Sci U S A.* 2012; 109: 19238–43. <https://doi.org/10.1073/pnas.1213825109> PMID: 23134727
42. Kennedy L, Kaltenbrun E, Greco TM, Temple B, Herring LE, Cristea IM, et al. Formation of a TBX20-CASZ1 protein complex is protective against dilated cardiomyopathy and critical for cardiac homeostasis. *PLoS Genet.* 2017; 13: e1007011. <https://doi.org/10.1371/journal.pgen.1007011> PMID: 28945738
43. Randall V, McCue K, Roberts C, Kyriakopoulou V, Beddow S, Barrett AN, et al. Great vessel development requires biallelic expression of Chd7 and Tbx1 in pharyngeal ectoderm in mice. *J Clin Invest.* 2009; 119: 3301–10. <https://doi.org/10.1172/JCI37561> PMID: 19855134
44. Heidet L, Moriniere V, Henry C, De Tomasi L, Reilly ML, Humbert C, et al. Targeted Exome Sequencing Identifies PBX1 as Involved in Monogenic Congenital Anomalies of the Kidney and Urinary Tract. *J Am Soc Nephrol.* 2017; 28: 2901–14. <https://doi.org/10.1681/ASN.2017010043> PMID: 28566479

45. Zhen T, Kwon EM, Zhao L, Hsu J, Hyde RK, Lu Y, et al. Chd7 deficiency delays leukemogenesis in mice induced by Cbfb-MYH11. *Blood*. 2017; 130: 2431–42. <https://doi.org/10.1182/blood-2017-04-780106> PMID: [29018080](#)
46. Low JK, Webb SR, Silva AP, Saathoff H, Ryan DP, Torrado M, et al. CHD4 Is a Peripheral Component of the Nucleosome Remodeling and Deacetylase Complex. *J Biol Chem*. 2016; 291: 15853–66. <https://doi.org/10.1074/jbc.M115.707018> PMID: [27235397](#)
47. Kolla V, Naraparaju K, Zhuang T, Higashi M, Kolla S, Blobel GA, et al. The tumour suppressor CHD5 forms a NuRD-type chromatin remodelling complex. *Biochem J*. 2015; 468: 345–52. <https://doi.org/10.1042/BJ20150030> PMID: [25825869](#)
48. Bottardi S, Mavoungou L, Pak H, Daou S, Bourgoin V, Lakehal YA, et al. The IKAROS interaction with a complex including chromatin remodeling and transcription elongation activities is required for hematopoiesis. *PLoS Genet*. 2014; 10: e1004827. <https://doi.org/10.1371/journal.pgen.1004827> PMID: [25474253](#)
49. O'Neill DW, Schoetz SS, Lopez RA, Castle M, Rabinowitz L, Shor E, et al. An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells. *Mol Cell Biol*. 2000; 20: 7572–82. PMID: [11003653](#)
50. Kim J, Sif S, Jones B, Jackson A, Koipally J, Heller E, et al. Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity*. 1999; 10: 345–55. PMID: [10204490](#)
51. Fan Y, Lu D. The Ikaros family of zinc-finger proteins. *Acta Pharm Sin B*. 2016; 6: 513–21. <https://doi.org/10.1016/j.apsb.2016.06.002> PMID: [27818917](#)
52. El Kasmi KC, Smith AM, Williams L, Neale G, Panopoulos AD, Watowich SS, et al. Cutting edge: A transcriptional repressor and corepressor induced by the STAT3-regulated anti-inflammatory signaling pathway. *J Immunol*. 2007; 179: 7215–9. PMID: [18025162](#)
53. Maruyama K, Uematsu S, Kondo T, Takeuchi O, Martino MM, Kawasaki T, et al. Strawberry notch homologue 2 regulates osteoclast fusion by enhancing the expression of DC-STAMP. *J Exp Med*. 2013; 210: 1947–60. <https://doi.org/10.1084/jem.20130512> PMID: [23980096](#)
54. Sauve F, McBroom LD, Gallant J, Moraitis AN, Labrie F, Giguere V. CIA, a novel estrogen receptor coactivator with a bifunctional nuclear receptor interacting determinant. *Mol Cell Biol*. 2001; 21: 343–53. <https://doi.org/10.1128/MCB.21.1.343-353.2001> PMID: [11113208](#)
55. Gillespie MA, Gold ES, Ramsey SA, Podolsky I, Aderem A, Ranish JA. An LXR-NCOA5 gene regulatory complex directs inflammatory crosstalk-dependent repression of macrophage cholesterol efflux. *EMBO J*. 2015; 34: 1244–58. <https://doi.org/10.15252/embj.201489819> PMID: [25755249](#)
56. Leulliot N, Godin KS, Hoareau-Aveilla C, Quevillon-Cheruel S, Varani G, Henry Y, et al. The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p. *J Mol Biol*. 2007; 371: 1338–53. <https://doi.org/10.1016/j.jmb.2007.06.031> PMID: [17612558](#)
57. Blake JA, Ziman MR. Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development*. 2014; 141: 737–51. <https://doi.org/10.1242/dev.091785> PMID: [24496612](#)
58. Capellini TD, Zappavigna V, Selleri L. Pbx homeodomain proteins: TALEnted regulators of limb patterning and outgrowth. *Dev Dyn*. 2011; 240: 1063–86. <https://doi.org/10.1002/dvdy.22605> PMID: [21416555](#)
59. Lu MF, Cheng HT, Lacy AR, Kern MJ, Argao EA, Potter SS, et al. Paired-related homeobox genes cooperate in handplate and hindlimb zeugopod morphogenesis. *Dev Biol*. 1999; 205: 145–57. <https://doi.org/10.1006/dbio.1998.9116> PMID: [9882503](#)
60. Zhao C, Malicki J. Nephrocystins and MKS proteins interact with IFT particle and facilitate transport of selected ciliary cargos. *EMBO J*. 2011; 30: 2532–44. <https://doi.org/10.1038/emboj.2011.165> PMID: [21602787](#)
61. Neumayer G, Belzil C, Gruss OJ, Nguyen MD. TPX2: of spindle assembly, DNA damage response, and cancer. *Cell Mol Life Sci*. 2014; 71: 3027–47. <https://doi.org/10.1007/s00018-014-1582-7> PMID: [24556998](#)
62. Neumayer G, Nguyen MD. TPX2 impacts acetylation of histone H4 at lysine 16: implications for DNA damage response. *PLoS One*. 2014; 9: e110994. <https://doi.org/10.1371/journal.pone.0110994> PMID: [25365214](#)



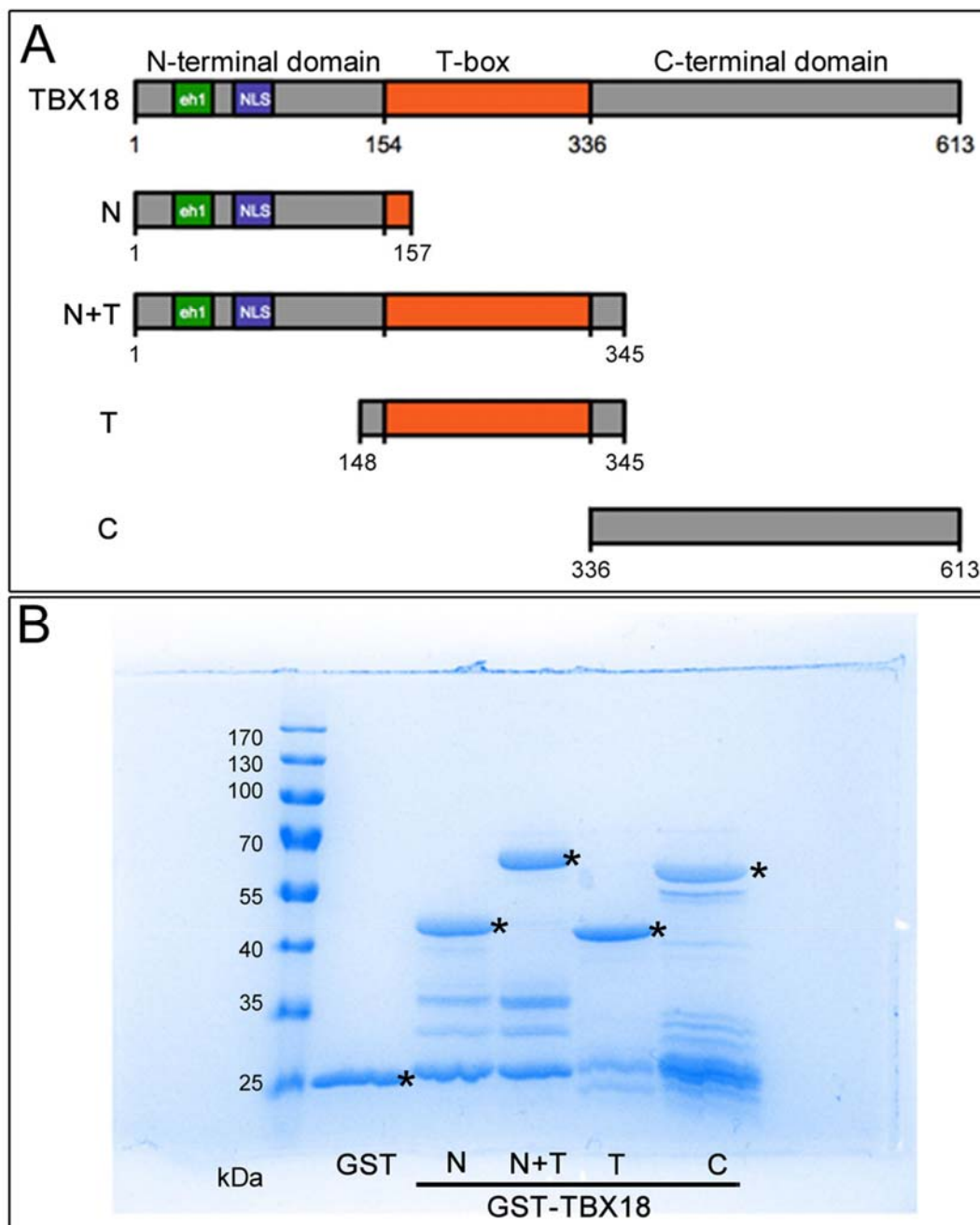
S1 Fig. *Tbx18* is expressed in 293 cells but not in other tested cell lines.

Semi-quantitative RT-PCR analysis of *TBX18* expression in 293, SK-N-MC, SW-13, Huh-7.5, SH-SY5Y and 786-O cells. *GAPDH* was used as a housekeeping control and all values were calculated as relative gene per *GAPDH* ratios. 293 cells exhibit a strong expression of *TBX18* with a mean $\pm$ SD ratio of 1.1408 ± 0.2394 . In all other tested cell lines *TBX18* expression was not above background level.



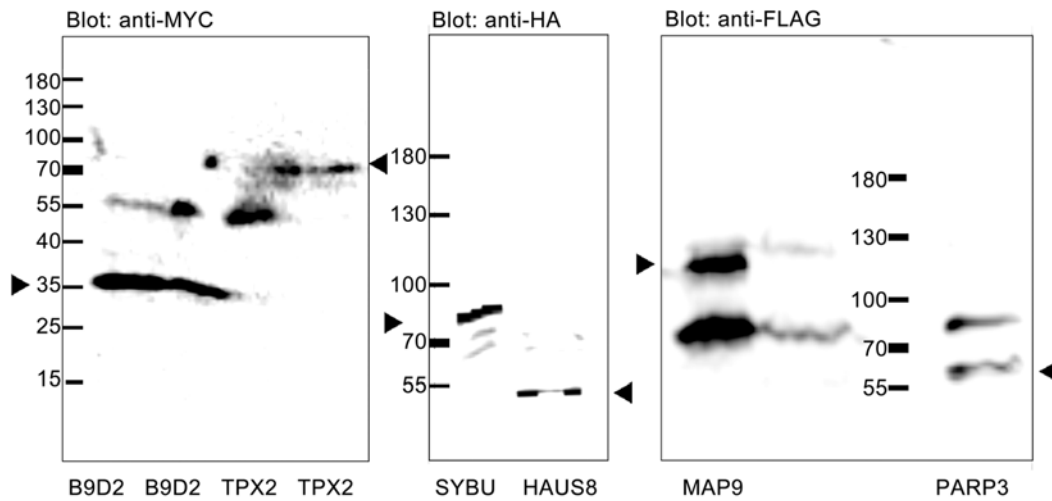
S2 Fig. Expression and localization of identified transcriptional cofactor candidates in 293 cells.

(A) Western Blot analysis of over-expressed co-factors in 293 cells. Expression constructs (as listed in [S5A Table](#)) were transfected into 293 cells and exogenous proteins were detected using antibodies against the corresponding tags. All detected proteins were of the expected size except BASP1, which as previously appeared larger. (B) Immunofluorescence analysis of localization of candidate transcriptional cofactors after transfection of 293 cells with the expression plasmids listed in (A). BASP1, CHD7, GAR1, IKZF2, NCOA5, SBNO2, SSXB2, SUV39H2 localized to the nucleus, CBFB and RCOR3 to the cytoplasm. Scale bar length is 25 μ m.



S3 Fig. GST-TBX18 fusion proteins for pull-down assays.

(A) Schematic representation of the primary structure of TBX18, and of the subfragments used to express GST fusion proteins. The T-box (T) is shaded in orange, and the N- and C-terminal domains (N and C) are shown in grey. The numbers refer to the amino acid position in the full-length TBX18 protein. The localization of the Groucho binding region (eh1), the nuclear localization signal (NLS) are highlighted in the N-terminal domain. (B) GST and fusion proteins of GST and N-, N+T-, T- and C-domains of TBX18 were purified from *E. coli* extracts and analyzed for integrity and quantity by Coomassie Brilliant Blue staining of SDS-polyacrylamide gels. Asterisks mark the full-length proteins.



S4 Fig. Western Blot analysis of over-expressed “centrosomal” candidate proteins in 293 cells. Expression constructs (as listed in [S4E Table](#)) were transfected into 293 cells and exogenous proteins were detected using antibodies against the corresponding tags.

S1 Table. Summary of cloning strategies for expression plasmids.

Plasmid	Fragment (or ORF)	Template (parental plasmid)	Template reference	Cloning Strategy	Cloning sites	Primer1 fwd 5' to 3'	Primer2 fwd 5' to 3'	Accepting vector	Accepting vector reference	Notes
TBX18 Protein partner Identification										
<i>p.EF1a.3xFLAG.BioTag</i>	103 bp	-	-	Assembly PCR	SmaI-NcoI	CCCCCGGGCTGCAGGAATTCCA CAGATGCTCTTCTTGCCCTTGTC	AGATGCTCTTCTTGCCCTTGTCATCG TCATCCTTGTAATCGATATCATGAT CTTTAT ACTGCCATGGACTACAAGACCA TGACGGTGATTATAAAGATCATGA TATCGATT	<i>pEF1a.Flag.BioTag</i>	O'Brien et al., 2018	SmaI-NcoI-digested PCR fragment cloned into NcoI partially digested <i>pEF1a.Flag.BioTag</i>
<i>p.EF1a.3xFLAG.BioTag.Tbx18</i>	1.8 kb	<i>pSP64.G.Tbx18.myc</i>	Farin et al., 2007	PCR, digestion and ligation	BamHI-XbaI	TTGTTCCGATCCCTGGCGGAGAA GCGGAGGGGCTC	GTGCTATCTAGATCAGACCATATG TGCAGACAC	<i>pEF1a.3xFLAG.BioTag</i>	this work	In two steps. First a 700bp BamHI-XbaI fragment was cloned directionally, followed by a 1.1kb fragment insertion into the BamHI site checked for direction
<i>p.EF1a.3xFLAG.BioTag.Tbx18.IRES.NLS.EGFP.BirA.BGHpA</i>	ca 2.8 kb	<i>p.EF1a.Six2DE.IRES.NLS.EGFP.BirA.BGHpA</i>	O'Brien et al., 2018	Shuttling by digestion and ligation	NheI			<i>pEF1a.3xFLAG.BioTag.Tbx18</i>	this work	
Cell culture-expressed cofactor localization and pulldown assays										
<i>pcDNA3.Nkx2.5.HA</i>	972 bp	<i>pcDNA3.cxx</i>	gift from Vincent Christoffels, Hartfaakcentrum, Amsterdam, Netherlands	PCR, digestion and ligation	BamHI/EcoRI	TGGATCCTCGCCACCATGTTCCC CAG	GGAATTCGGCTCGGATGCCGTGCA G	<i>pcDNA3.HA</i>	Invitrogen, Carlsbad, California	
<i>pcDNA3.Cbfb.myc</i>	903 bp	<i>pSP64.Cbfb.myc</i>	this work	Shuttling by digestion and ligation	HindIII/EcoRI	-	-	<i>pcDNA3</i>	Invitrogen, Carlsbad, California	
<i>pcDNA3.Gar1.myc</i>	1020 bp	<i>pSP64.Gar1.myc</i>	this work	Shuttling by digestion and ligation	HindIII-XbaI	-	-	<i>pcDNA3</i>	Invitrogen, Carlsbad, California	
Transactivation assays										
<i>pGL3.Prom.Tbx18BS2</i>	36 bp	-	Farin et al., 2007	Assembly PCR	BglII	GATCCGGTGCAGTAGGTGTGAAA TCGCACCTGGGGA	GATCTCCCCAGGTGCGATTTCACA CCTACTGCACCG	<i>pGL3.Promoter</i>	Promega, Madison, Wisconsin	
In vitro translated-cofactor pulldown assays										
<i>pSP64.Nkx2.5.myc</i>	972 bp	<i>pcDNA3.Nkx2.5.HA</i>	this work	PCR, digestion and ligation	NcoI-NdeI	AACACCATGGTCCCCAGCCCTGC G	ACCTCATATGTCCCCTCCAGGCTC GGATGCC	<i>pSP64.G.myc</i>	Farin et al., 2007	
<i>pSP64.Basp1.myc</i>	681 bp	<i>pcDNA3.Basp1.FLAG.HA</i>	Goodfellow et al., 2011	PCR, digestion and ligation	NcoI-NdeI	ATCGCCATGGGAGGCAAGCTCAG CAAG	CGATCATATGTCTTTCACGGTTA CGGTTTGG	<i>pSP64.G.myc</i>	Farin et al., 2007	
<i>pSP64.Cbfb.myc</i>	561 bp	E12.5 whole mouse cDNA/-	-	PCR, digestion and ligation	NcoI/NdeI	ATCGCCATGGGACCGCGCTGCT CCCGGACC	CGATCATATGACGAAGTTGAGAT ATCACC GC	<i>pSP64.G.myc</i>	Farin et al., 2007	
<i>pSP64.Gar1.myc</i>	693 bp	E12.5 whole mouse cDNA/-	-	PCR, digestion and ligation	NcoI-NdeI	ATCGCCATGGGATCTTTCGAGG CGGAGGTCGC	CGATCATATGATGTCCCTTCCTC GGAATCCG	<i>pSP64.G.myc</i>	Farin et al., 2007	
<i>pSP64.Ncoa5.myc</i>	1.7 kb	<i>pCMV6.Ncoa5.myc.FLA</i>	Origene #MR209077 Technologies, Rockville, USA	PCR, digestion and ligation	NcoI-NdeI	ATCGCCATGGGAAATACGGCTCC ATCAAGACCCAG	CGATCATATGGTAATGCCTCTGGT AAGATCCC	<i>pSP64.G.myc</i>	Farin et al., 2007	In two steps. First a 400 bp NcoI-NdeI fragment directionally, followed by a 1300 bp fragment insertion into the NdeI site checked for direction
<i>pEF4.Sbno2.myc.His</i>	4.1 kb	<i>pLenti.GIII.CMV.Sbno2</i>	Maruyama et al., 2013	Shuttling by digestion and ligation	EcoRV	-	-	<i>pEF4.myc.His</i>	Invitrogen, Carlsbad, California	

In vitro translated-transcription factor pulldown assays									
<i>pSP64.Goosecoid.myc</i>	768 bp	<i>pWZL.Blast.HA.Goosecoid</i>	Addgene #14096, Cambridge, USA	PCR and ligation	NcoI-NdeI	ATCGCCATGGGGATGCCGCCAGCATG	CGATCATATGGCTGTCCGAGTCCA	<i>pSP64.G.myc</i>	Farin et al., 2007
<i>pSP64.Pax1.myc</i>	1.3 kb	<i>pKS.Pax1</i>	gift from Ahmed Mansouri	PCR and ligation	Sall	GATCGTCGACATGGAGCAGACGTACGGCG	GCATGTCGACCTAGGAGGTCGAGGCTGGG	<i>pSP64.G.Sall.myc</i>	Farin et al., 2007
<i>pSP64.Prrx2.myc</i>	740 bp	<i>pSG5-Prrx2</i>	Addgene #21009, Cambridge, USA	PCR and ligation	NcoI-NdeI	ATCGCCATGGACAGCGCGGCCG	CGATCATATGGTTCACTGTGGGCACTTGGC	<i>pSP64.G.myc</i>	Farin et al., 2007
In vitro translated-centrosomal pulldown assays									
<i>pSP64.B9d2.myc</i>	525 bp	<i>mEmerald.B9D2-N-18</i>	Addgene #54012, Cambridge, USA	PCR, digestion and ligation	NcoI-NdeI	ATCGCCATGGCTGAAGTGCACGTG	ACGATCATATGACACTCCACTCCA	<i>pSP64.G.myc</i>	Farin et al., 2007
<i>pSP64.Cep89.myc</i>	2.3 kb	<i>pEGFP.C1.Cep123</i>	Sillibourne et al., 2011	PCR, digestion and ligation	NdeI	ATCGCATATGCTCTGGGATTTCGAGAG	CGATCATATGGCAGGTGGGGCATGAGAC	<i>pSP64.G.myc</i>	Farin et al., 2007
<i>pSP64.Haus8.HA</i>	1.2 kb	<i>pEGFP.C1.HAUS8</i>	Lawo et al., 2009	PCR, digestion and ligation	NcoI-NdeI	ATCGCCATGGCGGATTCCTCGGGCGG	CGATCATATGTGACAAGTCCCTCCCTGAACG	<i>pSP64.G.HA</i>	Farin et al., 2007
<i>pSP64.Haus8.myc</i>	1.2 kb	<i>pEGFP.C1.HAUS8</i>	Lawo et al., 2009	PCR, digestion and ligation	NcoI-NdeI	ATCGCCATGGCGGATTCCTCGGGCGG	CGATCATATGTGACAAGTCCCTCCCTGAACG	<i>pSP64.G.myc</i>	Farin et al., 2007
<i>pSP64.Mical3.HA</i>	6 kb	<i>GFP.Mical3.FL.WT</i>	Grigoriev et al., 2011	PCR, digestion and ligation	NdeI	ATCGCATATGGAGAGAGGAAGCATGAG	CGATCATATGGGACCAGTTAAGGCTGAAGCC	<i>pSP64.G.HA</i>	Farin et al., 2007
<i>pSP64.Parp3.myc</i>	1.6 kb	293 cell cDNA	-	PCR, digestion and ligation	NdeI	ATCGCATATGTCCCTGCTTTTCTTGCC	CGATCATATGGAGGTGGACCTCCAGCAGG	<i>pSP64.G.myc</i>	Farin et al., 2007
<i>pSP64.Pcm1.HA</i>	6 kb	293 cell cDNA	-	PCR, digestion and ligation	NdeI	ATCGCATATGGCCACAGGAGGAGGTCC	CGATCATATGTACTCTGGGCTCCACCG	<i>pSP64.G.HA</i>	Farin et al., 2007
<i>pSP64.Sybu.myc</i>	2 kb	293 cell cDNA	-	PCR, digestion and ligation	NcoI	ATCGCCATGGGGCCCTCCCGGAG	CGATCCATGGGGTTTIGATACGGAGGCGGTG	<i>pSP64.G.myc</i>	Farin et al., 2007
<i>pSP64.Tpx2.myc</i>	2.2 kb	<i>pmCherry.TPX2</i>	Addgene #31227, Cambridge, USA	PCR, digestion and ligation	NcoI	ATCGCCATGGGGATGTCAACAAGTTAAAGCTCTTATC	CGATCCATGGGCAGTGAATCGAGTGAGAATT	<i>pSP64.G.myc</i>	Farin et al., 2007
Centrosomal localization assays									
<i>pcDNA3.b9d2.myc</i>	864 bp	<i>pSP64.B9d2.myc</i>	this work	Shuttling by digestion and ligation	HindIII-EcoRI			<i>pcDNA3</i>	Invitrogen, Carlsbad, California
<i>pcDNA3.Haus8.HA</i>	1.6 kb	<i>pSP64.Haus8.HA</i>	this work	Shuttling by digestion and ligation	HindIII-EcoRI			<i>pcDNA3</i>	Invitrogen, Carlsbad, California
<i>pcDNA3.Sybu.HA</i>	2.3 kb	<i>pSP64.Sybu.myc</i>	this work	Shuttling by digestion and ligation	HindIII-XbaI			<i>pcDNA3</i>	Invitrogen, Carlsbad, California
<i>pcDNA3.Tpx2.myc</i>	2.6 kb	<i>pSP64.Tpx2.myc</i>	this work	Shuttling by digestion and ligation	HindIII-XbaI			<i>pcDNA3</i>	Invitrogen, Carlsbad, California

S2 Table. Mass spectrometry analysis identifies proteins copurified with TBX18 from 293 cells.

List of 143 proteins detected by mass spectrometry in extracts of proteins copurified with TBX18 in 293 cells. Shown are Uniprot accession code (first row), the protein symbol (second row) and the protein name (third row). Color code identifies the three major protein clusters identified: transcription factors of the homeobox family (grey), transcriptional cofactors (green) and centrosomal proteins (yellow).

UniProtKB	Protein symbol	Protein name
O95870	ABHD16A	Abhydrolase domain-containing protein 16A
Q8NCV1	ADAD2	Adenosine deaminase domain-containing protein 2
E9PIL2	ANKRD42	Ankyrin repeat domain-containing protein 42
O15085	ARHGEF11	Isoform 2 of Rho guanine nucleotide exchange factor 11
O43861	ATP9B	Isoform 2 of Probable phospholipid-transporting ATPase IIB
Q9BP19	B9D2	B9 domain-containing protein 2
O14514	BAI1	Brain-specific angiogenesis inhibitor 1
P80723	BASP1	Isoform 2 of Brain acid soluble protein
Q96P09	BIRC8	Baculoviral IAP repeat-containing protein 8
P53004	BLVRA	Biliverdin reductase A
P31944	CASP14	Caspase-14
J3KTD8	CBFB	Core-binding factor subunit beta (Fragment)
E7EUS2	CCDC110	Coiled-coil domain-containing protein 110
F8WBW2	CCDC39	Coiled-coil domain-containing protein 39
A6NKG9	CCDC85	Coiled-coil domain-containing protein 85C
C9J8J8	CDH2	Cadherin-2 (Fragment)
Q00535	CDK5	Isoform 2 of Cyclin-dependent kinase 5
B1AMJ5	CDK5RAP2	CDK5 regulatory subunit-associated protein 2 (Fragment)
D6R8W0	CENPE	Centromere-associated protein E (Fragment)
ESRG47	CEP170	Centrosomal protein of 170 kDa (Fragment)
Q96L14	CEP170P1	Cep170-like protein
K7EQ12	CEP89	Centrosomal protein of 89 kDa (Fragment)
Q9NX63	CHCHD3	Coiled-coil-helix-coiled-coil-helix domain-containing protein 3, mitochondrial
Q9P2D1	CHD7	Chromodomain-helicase-DNA-binding protein 7
Q5TAQ5	WDR42A	DDB1- and CUL4-associated factor 8 (Fragment)
P81605	DCD	Dermcidin
Q02413	DSG1	Desmoglein-1
O43812	DUX1	Double homeobox protein 1
Q96PT4	DUX3	Isoform 2 of Putative double homeobox protein 3
Q9UBX2	DUX4	Double homeobox protein 4
P0CJ85	DUX4L2	Double homeobox protein 4-like protein 2
P0CJ87	DUX4L4	Double homeobox protein 4-like protein 4
Q6RFH8	DUX4L9	Double homeobox protein 4C
Q96PT3	DUX5	Double homeobox protein 5
P68104	EEF1A1	Elongation factor 1-alpha 1
Q5VTE0	EEF1A1P5	Putative elongation factor 1-alpha-like 3
Q05639	EEF1A2	Elongation factor 1-alpha 2
F8W9E7	ELMO3	Engulfment and cell motility protein 3
Q16134	ETFDH	Electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial
HOYH69	ETNK1	Ethanolamine kinase 1 (Fragment)
P49640	EVX1	Homeobox even-skipped homolog protein 1
Q03828	EVX2	Homeobox even-skipped homolog protein 2
Q9ULE4	FAM184B	Protein FAM184B OS=Homo sapiens
Q9UK61	FAM208A	Protein FAM208A OS=Homo sapiens
Q8IY13	FBLN7	FBLN7 protein (Fragment)
Q5D862	FLG2	Filaggrin-2 OS=Homo sapiens
Q9NY12	GAR1	H/ACA ribonucleoprotein complex subunit 1
P14136	GFAP	Glial fibrillary acidic protein
D6RF75	GFM2	Ribosome-releasing factor 2, mitochondrial (Fragment)
O60547	GMDS	GDP-mannose 4,6 dehydratase
P56915	GSC	Homeobox protein goosecoid
Q9BTM1	H2AFJ	Histone H2A.J
Q71U19	H2AFV	Histone H2A
P16104	H2AFX	Histone H2A.x
P0C0S5	H2AFZ	Histone H2A.Z
Q9BT25	HAUS8	Isoform 2 of HAUS augmin-like complex subunit 8
Q9BXC0	HCAR1	Hydroxycarboxylic acid receptor 1
P04908	HIST1H2AB	Histone H2A type 1-B/E
P20671	HIST1H2AB	Histone H2A type 1-D
Q99878	HIST1H2AC	Histone H2A type 1-J
P0C0S8	HIST1H2AG	Histone H2A type 1
Q96KK5	HIST1H2AH	Histone H2A type 1-H
Q96QV6	HIST1H2AK	Histone H2A type 1-A
Q93077	HIST1H2AO	Histone H2A type 1-C
Q6F113	HIST2H2AA2	Histone H2A type 2-A
Q8IU66	HIST2H2AB	Histone H2A type 2-B
Q16777	HIST2H2AC	Histone H2A type 2-C
Q7L7L0	HIST3H2A	Histone H2A type 3
F5H4X2	HMBS	Porphobilinogen deaminase
Q86YZ3	HRNR	Hornerin
P54652	HSPA2	Heat shock-related 70 kDa protein 2
P06310	IGKV2-30	Ig kappa chain V-II region RPM1 6410
P01615	IGKV2D-28	Ig kappa chain V-II region GM607 (Fragment)
P01614	IGKV2D-40	Ig kappa chain V-II region Cum
Q9UKS7	IKZF2	Zinc finger protein Helios
HOY8Z1	INTS3	Integrator complex subunit 3 (Fragment)
Q5T749	KPRP	Keratinocyte proline-rich protein
P12035	KRT3	Keratin, type II cytoskeletal 3
Q14CN4	KRT72	Isoform 2 of Keratin, type II cytoskeletal 72
HOYIC5	KRT73	Keratin, type II cytoskeletal 73 (Fragment)
F8W1S1	KRT74	Keratin, type II cytoskeletal 74
P55268	LAMB2	Laminin subunit beta-2

Q8TC57	C2orf65	Isoform 2 of Uncharacterized protein C2orf65
Q9UPN3	MACF1	Isoform 2 of Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5
O95382	MAP3K6	Isoform 3 of Mitogen-activated protein kinase kinase kinase 6
Q49MG5	MAP9	Microtubule-associated protein 9
Q9NU22	MDN1	Midasin
Q5JXC5	MFN2	Mitofusin-2 (Fragment)
Q7RTP6	MICAL3	Isoform 3 of Protein-methionine sulfoxide oxidase MICAL3
Q9UKX3	MYH13	Myosin-13
Q9HCD5	NCOA5	Nuclear receptor coactivator 5
E9PN51	NDUFS8	NADH dehydrogenase [ubiquinone] iron-sulfur protein 8, mitochondrial (Fragment)
P48681	NES	Nestin OS=Homo sapiens
Q8NEJ9	NGDN	Isoform 2 of Neuroguidin
Q96RS6	NUDCD1	Isoform 2 of NudC domain-containing protein 1
Q5VST9	OBSCN	Isoform 2 of Obscurin
M0QZ32	OCEL1	Ocludin/ELL domain-containing protein 1 (Fragment)
Q9H074	PAIP1	Isoform 2 of Polyadenylate-binding protein-interacting protein 1
Q9Y6F1	PARP3	Isoform 2 of Poly [ADP-ribose] polymerase 3
P15863	PAX1	Isoform 2 of Paired box protein Pax-1
Q9BYU1	PBX4	Pre-B-cell leukemia transcription factor 4
Q15154	PCMI	Isoform 2 of Pericentriolar material 1 protein
F8W9F3	PCMTD2	Protein-L-isaspartate O-methyltransferase domain-containing protein 2
M0R0U1	PIK3C3	Phosphatidylinositol 3-kinase catalytic subunit type 3 (Fragment)
E9PDN7	PIK3CG	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform (Fragment)
Q5VTI5	PLEKHA6	Pleckstrin homology domain containing, family A member 6
Q86V59	PNMAL1	Isoform 2 of PNMA-like protein 1
Q6S8J3	POTEE	POTE ankyrin domain family member E
ASA3E0	POTEF	POTE ankyrin domain family member F
Q9BYX7	POTEKP	Putative beta-actin-like protein 3
Q99811	PRRX2	Paired mesoderm homeobox protein 2
K7EJR3	PSMD8	26S proteasome non-ATPase regulatory subunit 8 (Fragment)
H0YF22	RCOR3	REST corepressor 3 (Fragment)
Q9Y2G9	SBN02	Isoform 2 of Protein strawberry notch homolog 2
Q6UWP8	SBSN	Suprabasin
Q8WXD2	SCG3	Isoform 2 of Secretogranin-3
Q4G0G5	SCGB2B2	Secretoglobulin family 2B member 2
P21912	SDHB	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial
P31947	SFN	14-3-3 protein sigma
Q9BYB0	SHANK3	Isoform 2 of SH3 and multiple ankyrin repeat domains protein 3
Q8IWU4	SLC30A8	Zinc transporter 8
P49223	SPINT3	Kunitz-type protease inhibitor 3
O60224	SSXB2	Protein SSX4
Q96FJ0	STAMBPL1	Isoform 2 of AMSH-like protease
Q6J9G0	STYK1	Tyrosine-protein kinase STYK1
M0R216	SUGP2	SURP and G-patch domain-containing protein 2 (Fragment)
G5E9A8	SUV39H2	Histone-lysine N-methyltransferase SUV39H2
Q9NX95	SYBU	Isoform 2 of Syntabulin
P37837	TALDO1	Transaldolase OS=Homo sapiens
Q86SX3	C14orf80	Isoform 2 of Uncharacterized protein C14orf80
Q53EQ6	TIGD5	Isoform 2 of Tigger transposable element-derived protein 5
Q9NXH8	TOR4A	Torsin-4A OS=Homo sapiens
Q9ULW0	TPX2	Targeting protein for Xklp2
Q5T4S7	UBR4	Isoform 2 of E3 ubiquitin-protein ligase UBR4
F8VZH8	UNC13A	Protein unc-13 homolog A
Q9Y484	WDR45	Isoform 2 of WD repeat domain phosphoinositide-interacting protein 4
P98170	XIAP	E3 ubiquitin-protein ligase XIAP
P63104	YWHAZ	14-3-3 protein zeta/delta
Q9NXT0	ZNF586	Isoform 2 of Zinc finger protein 586
Q8N2I2	ZNF619	Isoform 2 of Zinc finger protein 619
A4D1E1	ZNF804B	Zinc finger protein 804B
Q6ZNA1	ZNF836	Zinc finger protein 836
A6NDW0	ZRSR2	U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit-related protein 2

S3 Table. DAVID analysis identifies 83 enriched functional annotations in the set of 143 proteins identified as possible TBX18 interaction partners in 293 cells.
From the 143 identified proteins we excluded immunoglobulins, histone proteins and cytoskeletal proteins which leads to 83 possible interaction partners.

Category	Term	Count	%	Pvalue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KEYWORDS GOTERM _CC_DIRECT	Nucleosome core	12	8,8235	1,39E-11	HIST1H2AB, HIST1H2AC, HIST2H2AB, H2AFV, HIST1H2AG, HIST2H2AC, H2AFZ, HIST1H2AH, H2AFX, HIST3H2A, HIST1H2AK, H2AFJ	135	89	20581	20,5553059	2,86E-09	2,86E-09	1,75E-08
	GO:0000786 ~nucleosome	11	8,0882	1,04E-09	HIST1H2AB, HIST1H2AC, HIST2H2AB, H2AFV, HIST1H2AG, HIST2H2AC, H2AFZ, HIST1H2AH, H2AFX, HIST1H2AK, H2AFJ	128	94	18224	16,6609043	1,97E-07	1,97E-07	1,29E-06
	Chromosome	16	11,765	5,48E-08	HIST1H2AB, HIST1H2AC, HIST1H2AG, CENPE, H2AFJ, FAM208A, SUV39H2, HIST2H2AB, H2AFV, HIST2H2AC, H2AFZ, NGDN, HIST1H2AH, HIST1H2AK, HIST3H2A, H2AFX	135	400	20581	6,09807407	1,12E-05	5,62E-06	6,90E-05
UP_KEYWORDS	GO:0070062 ~extracellular exosome	45	33,088	6,63E-08	HIST1H2AB, DCD, HIST1H2AC, HRNR, POTEKP, YWHAZ, TALDO1, HIST1H2AG, KPRP, CHCHD3, FLG2, CDH2, SFN, HIST2H2AB, LAMB2, HSPA2, H2AFV, SBSN, HIST2H2AC, H2AFZ, IGKV2-30, IGKV2D-28, H2AFX, KRT3, HIST3H2A, CDK5RAP2, POTEF, POTE, PSMD8, EEF1A1, GMD5, H2AFJ, BASP1, KRT73, BLVRA, KRT72, SDHB, KRT74, CASP14, EEF1A1P5, DSG1, FBLN7, HIST1H2AH, MYH13, HIST1H2AK	128	2811	18224	2,27921558	1,25E-05	6,27E-06	8,23E-05
GOTERM _CC_DIRECT	Methylation	24	17,647	1,24E-07	HIST1H2AB, EEF1A1, HIST1H2AC, GFAP, HRNR, POTEKP, RCOR3, HIST1H2AG, EEF1A2, PAIP1, CENPE, H2AFJ, SHANK3, HIST2H2AB, CHD7, HSPA2, DCAF8, EEF1A1P5, HIST2H2AC, H2AFZ, HIST1H2AH, MYH13, HIST1H2AK, HIST3H2A	135	1001	20581	3,65518926	2,54E-05	8,46E-06	1,56E-04
UP_KEYWORDS	Citrullination	9	6,6176	2,32E-07	HIST1H2AB, HIST1H2AC, HIST2H2AB, GFAP, HIST1H2AG, HIST2H2AC, HIST1H2AH, HIST3H2A, HIST1H2AK	135	97	20581	14,1450172	4,77E-05	1,19E-05	2,93E-04
UP_KEYWORDS	Homeobox	12	8,8235	1,19E-06	DUX4L9, GSC, EVX1, EVX2, DUX4L2, DUX1, PRRX2, DUX4, DUX4L4, DUX3, PBX4, DUX5	135	262	20581	6,98252757	2,44E-04	4,89E-05	0,001501
UP_KEYWORDS	Coiled coil	41	30,147	5,87E-06	YWHAZ, GFAP, CHCHD3, CCDC39, MAP3K6, CHD7, LAMB2, MACF1, CEP170, SYBU, KRT3, ZRSR2, CDK5RAP2, MDN1, POTEF, POTE, OBSN, NES, RCOR3, CEP89, MICAL3, TPX2, CENPE, PCMI, ANKRD42, SHANK3, ARHGEF11, KRT73, MFN2, KRT72, PLEKHA6, CCDC110, KRT74, DCAF8, FAM184B, FBLN7, NGDN, MYH13, HAUS8, MAP9, UNC13A	135	3036	20581	2,05880301	0,0012035	2,01E-04	0,007392
UP_KEYWORDS GOTERM _CC_DIRECT	GO:0000790 ~nuclear chromatin	10	7,3529	8,41E-06	HIST1H2AB, HIST1H2AC, HIST2H2AB, H2AFV, HIST1H2AG, HIST2H2AC, HIST1H2AH, H2AFX, HIST1H2AK, H2AFJ	128	193	18224	7,37694301	0,0015878	5,30E-04	0,01044
	DNA-binding	31	22,794	1,72E-05	HIST1H2AB, HIST1H2AC, DUX4L9, EVX1, HIST1H2AG, EVX2, ZNF619, DUX4L2, PRRX2, DUX4L4, PAX1, HIST2H2AB, CHD7, H2AFV, HIST2H2AC, H2AFZ, H2AFX, HIST3H2A, TIGD5, DUX1, DUX4, DUX3, DUX5, IKZF2, GSC, H2AFJ, ZNF836, HIST1H2AH, HIST1H2AK, ZNF586, PBX4	135	2050	20581	2,30536947	0,0035266	5,05E-04	0,021685
UP_KEYWORDS	GO:0005634 ~nucleus	61	44,853	2,23E-05	HRNR, GAR1, INTS3, PRRX2, PAX1, CBF, HIST2H2AB, NUDCD1, H2AFV, HIST2H2AC, H2AFZ, H2AFX, HIST3H2A, ZRSR2, TIGD5, MDN1, PSMD8, GSC, RCOR3, MICAL3, TPX2, H2AFJ, BASP1, CDK5, DCAF8, NCOA5, EEF1A1P5, ZNF586, HIST1H2AB, HIST1H2AC, YWHAZ, DUX4L9, TALDO1, XIAP, HIST1H2AG, B9D2, CHCHD3, FLG2, ZNF619, SFN, DUX4L2, DUX4L4, CHD7, HSPA2, DUX1, DUX4, DUX3, DUX5, EEF1A1, IKZF2, EEF1A2, CENPE, ZNF836, FAM208A, SUV39H2, KRT73, CCDC110, CASP14, HIST1H2AH, HIST1H2AK, PARP3, PBX4	128	5415	18224	1,60385503	0,0042038	0,0010526	0,027675
GOTERM _CC_DIRECT	Nucleus	57	41,912	2,61E-05	EVX1, EVX2, GAR1, INTS3, PRRX2, PAX1, CBF, HIST2H2AB, NUDCD1, H2AFV, HIST2H2AC, H2AFZ, H2AFX, HIST3H2A, ZRSR2, TIGD5, MDN1, GSC, RCOR3, MICAL3, UBR4, TPX2, H2AFJ, CDK5, DCAF8, NCOA5, ZNF586, HIST1H2AB, HIST1H2AC, DUX4L9, XIAP, HIST1H2AG, B9D2, CHCHD3, ZNF619, SFN, DUX4L2, DUX4L4, CHD7, DUX1, DUX4, DUX3, DUX5, EEF1A1, IKZF2, EEF1A2, ZNF836, FAM208A, SUV39H2, SUGP2, CCDC110, CASP14, HIST1H2AH, NGDN, HIST1H2AK, PARP3, PBX4	135	5244	20581	1,65708535	0,005332	6,68E-04	0,032815
UP_KEYWORDS	Isopeptide bond	21	15,441	4,40E-05	HIST1H2AB, HIST1H2AC, NES, IKZF2, XIAP, HIST1H2AG, TPX2, CENPE, H2AFJ, BASP1, HIST2H2AB, HIST2H2AC, H2AFZ, BIRC8, HIST1H2AH, MYH13, HIST1H2AK, HIST3H2A, H2AFX, POTEF, POTE	135	1132	20581	2,8281704	0,0089862	0,0010025	0,055399
UP_KEYWORDS	Ubl conjugation	26	19,118	9,68E-05	HIST1H2AB, EEF1A1, HIST1H2AC, NES, IKZF2, XIAP, HIST1H2AG, TPX2, CENPE, H2AFJ, BASP1, SFN, PCMI, ARHGEF11, MFN2, HIST2H2AB, H2AFV, HIST2H2AC, H2AFZ, HIST1H2AH, BIRC8, MYH13, H2AFX, HIST3H2A, HIST1H2AK, POTEF	135	1705	20581	2,32477897	0,0196519	0,0019828	0,121767
UP_KEYWORDS	Intermediate filament	6	4,4118	1,47E-04	KRT73, KRT72, NES, GFAP, KRT74, KRT3	135	77	20581	11,8793651	0,0297488	0,0027417	0,185223
UP_KEYWORDS	Phosphoprotein	75	55,147	2,66E-04	GFAP, HRNR, KPRP, INTS3, CCDC39, CBF, HIST2H2AB, MAP3K6, NUDCD1, HIST2H2AC, PIK3C2, H2AFX, HIST3H2A, ZRSR2, CDK5RAP2, MDN1, PSMD8, PIK3CG, RCOR3, CEP89, HMB5, MICAL3, TPX2, UBR4, H2AFJ, BASP1, PCMI, CDK5, ARHGEF11, MFN2, DCAF8, NCOA5, HAUS8, UNC13A, HIST1H2AB, HIST1H2AC, YWHAZ, TALDO1, XIAP, HIST1H2AG, CHCHD3, TOR4A, FLG2, SFN, CDH2, CHD7,	135	8246	20581	1,38659822	0,0530782	0,0045346	0,334245
UP_KEYWORDS												

GOTERM _CC_DIRECT	GO:0098574 ~cytoplasmic side of lysosomal membrane Acetylation	3	2,2059	2,86E-04	LAMB2, HSPA2, MACF1, CEP170, SYBU, ETFDH, STAMBPL1, SCG3, EEF1A1, OBSCN, NES, IKZF2, GMD5, EEF1A2, CENPE, SHANK3, FAM208A, SUV39H2, BLVRA, PLEKHA6, SUGP2, CCDC110, KRT74, DSG1, NGDN, HIST1H2AH, MYH13, HIST1H2AK, MAP9	128	4	18224	106,78125	0,0527108	0,0107717	0,355153
		39	28,676	4,58E-04	EEF1A1, GFAP, EEF1A2	135	3424	20581	1,73645509	0,0897053	0,0072037	0,575309
UP_KEYWORDS UP_KEYWORDS	Elongation factor Cytoplasm	4	2,9412	9,63E-04	HIST1H2AB, DCD, HIST1H2AC, YWHAZ, TALDO1, HIST1H2AG, CHCHD3, INTS3, HIST2H2AB, MACF1, H2AFV, HIST2H2AC, ETFDH, STAMBPL1, H2AFZ, H2AFX, HIST3H2A, MDN1, EEF1A1, NES, IKZF2, GMD5, EEF1A2, HMBS, PAIP1, UBR4, TPX2, H2AFJ, PCM1, CDK5, FAM208A, BLVRA, SDHB, NCOA5, HIST1H2AH, NGDN, HIST1H2AK, HAUS8, MAP9	135	30	20581	20,3269136	0,1792962	0,0140147	1,205666
		47	34,559	0,002644	GFM2, EEF1A1, EEF1A1P5, EEF1A2	135	4816	20581	1,48779839	0,4188777	0,0355394	3,277224
UP_KEYWORDS UP_KEYWORDS	Cytoskeleton	17	12,5	0,003002	POTEKP, YWHAZ, GFAP, TALDO1, XIAP, B9D2, CHCHD3, KPRP, FLG2, SFN, CCDC39, HSPA2, NUDCD1, MACF1, CEP170, SYBU, ETNK1, BIRC8, CDK5RAP2, POTEF, PIK3CG, M1AP, EEF1A1, OBSCN, CEP89, PAIP1, HMBS, MICAL3, UBR4, TPX2, CENPE, PCM1, CDK5, ELMO3, SHANK3, ARHGEF11, BLVRA, DCAF8, CASP14, EEF1A1P5, PCMTD2, NGDN, MYH13, PARP3, HAUS8, MAP9, UNC13A	135	1138	20581	2,27740025	0,4600418	0,0377841	3,71248
		6	4,4118	0,004252	POTEKP, CEP89, B9D2, MICAL3, UBR4, TPX2, CENPE, CCDC39, PCM1, HSPA2, MACF1, CEP170, SYBU, PARP3, CDK5RAP2, HAUS8, MAP9	128	152	18224	5,62006579	0,5530771	0,1256095	5,153922
GOTERM _CC_DIRECT	GO:0072562 ~blood microparticle	5	3,6765	0,005269	YWHAZ, HSPA2, IGV2-30, IGV2D-28, POTEF, POTE	128	100	18224	7,11875	0,6315934	0,1329447	6,350258
		5	3,6765	0,008076	KRT73, KRT72, KRT74, CASP14, KRT3	128	113	18224	6,29977876	0,7840131	0,1743359	9,578826
GOTERM _CC_DIRECT	GO:0005882 ~intermediate filament	7	5,1471	0,010138	KRT73, KRT72, NES, GFAP, KRT3	135	280	20581	3,8112963	0,8761868	0,1156313	12,03572
		7	5,1471	0,021845	MACF1, CEP170, TPX2, CENPE, HAUS8, CDK5RAP2	128	311	18224	3,20458199	0,9846165	0,371129	23,98755
UP_KEYWORDS UP_KEYWORDS	Nucleotide-binding	20	14,706	0,022854	MACF1, CEP170, SYBU, TPX2, CENPE, HAUS8, CDK5RAP2	135	1788	20581	1,70527798	0,9912555	0,2314859	25,24398
		3	2,2059	0,024151	PIK3CG, OBSCN, EEF1A1, POTEKP, EEF1A2, TOR4A, CENPE, CDK5, MFN2, MAP3K6, GFM2, STYK1, CHD7, HSPA2, ATP9B, EEF1A1P5, PIK3C3, ETNK1, MYH13, MDN1	135	37	20581	12,360961	0,9933411	0,2318566	26,48396
GOTERM _CC_DIRECT	GO:0005853 ~eukaryotic translation elongation factor 1 complex	2	1,4706	0,027588	SDHB, NDUFS8, ETFDH	128	4	18224	71,1875	0,9949447	0,410648	29,34706
		4	2,9412	0,040704	EEF1A1, EEF1A2	128	109	18224	5,22477064	0,9996118	0,5103198	40,31146
GOTERM _CC_DIRECT	GO:0000922 ~spindle pole	4	2,9412	0,044509	CEP89, TPX2, HAUS8, CDK5RAP2	128	113	18224	5,03982301	0,9998168	0,511829	43,18534
		7	5,1471	0,045573	CEP89, CEP170, PARP3, PCM1	128	371	18224	2,68632075	0,9998516	0,4924353	43,96618
GOTERM _CC_DIRECT	GO:0005856 ~cytoskeleton	3	2,2059	0,050962	POTEKP, MACF1, MICAL3, SYBU, CDK5RAP2, BASP1, CDK5	128	52	18224	8,21394231	0,9999491	0,5064539	47,77101
		6	4,4118	0,056327	NES, GFAP, MDN1	128	296	18224	2,88597973	0,9999826	0,5183298	51,32146
GOTERM _CC_DIRECT	GO:0045111 ~intermediate filament cytoskeleton	46	33,824	0,05723	DCD, NES, LAMB2, CASP14, SBSN, DSG1	128	5222	18224	1,25416507	0,9999855	0,5014977	51,89667
		3	2,2059	0,058297	HRNR, YWHAZ, GFAP, TALDO1, XIAP, CHCHD3, KPRP, FLG2, SFN, CDH2, CCDC39, NUDCD1, MACF1, CEP170, ETNK1, BIRC8, CDK5RAP2, MDN1, PIK3CG, M1AP, EEF1A1, NES, GMD5, CEP89, PAIP1, EEF1A2, HMBS, MICAL3, UBR4, CENPE, BASP1, PCM1, CDK5, ELMO3, SHANK3, ARHGEF11, BLVRA, CCDC110, KRT74, DCAF8, CASP14, EEF1A1P5, PCMTD2, NGDN, PARP3, UNC13A	135	60	20581	7,62259259	0,9999955	0,4597219	53,04131
GOTERM _CC_DIRECT	GO:0000775 ~chromosome, centromeric region	3	2,2059	0,059989	SDHB, NDUFS8, ETFDH	128	57	18224	7,49342105	0,9999916	0,497306	53,6155
		2	1,4706	0,061012	NGDN, CENPE, SUV39H2	128	9	18224	31,6388889	0,9999932	0,4836682	54,23892
GOTERM _CC_DIRECT	GO:1990023 ~mitotic spindle midzone	15	11,029	0,069013	CENPE, MAP9	135	1391	20581	1,64398115	0,9999996	0,502456	59,33994
		4	2,9412	0,077076	PIK3CG, MAP3K6, OBSCN, POTEKP, STYK1, CHD7, HSPA2, ATP9B, PIK3C3, ETNK1, TOR4A, MYH13, CENPE, CDK5, MDN1	135	152	20581	4,01189084	0,9999999	0,5264021	63,556
UP_KEYWORDS UP_KEYWORDS	Protein biosynthesis	4	2,9412	0,077076	GFM2, EEF1A1, EEF1A1P5, EEF1A2	135	152	20581	4,01189084	0,9999999	0,5264021	63,556
		7	5,1471	0,077721	OBSCN, UBR4, MYH13, CDK5RAP2	128	426	18224	2,33949531	0,9999998	0,5528285	63,38411
UP_KEYWORDS UP_KEYWORDS	Calmodulin-binding	4	2,9412	0,077076	CEP89, CEP170, B9D2, UBR4, HAUS8, CDK5RAP2, PCM1	135	153	20581	3,98566933	0,9999999	0,516548	64,16333
		2	1,4706	0,093297	PIK3CG, PIK3C3	128	14	18224	20,3392857	1	0,6036808	70,36463
GOTERM _CC_DIRECT	GO:0005942 ~phosphatidylinositol 3-kinase complex	4	2,9412	0,078308	KRT73, KRT72, KRT74, KRT3	135	153	20581	3,98566933	0,9999999	0,516548	64,16333
		2	1,4706	0,093297	PIK3CG, PIK3C3	128	14	18224	20,3392857	1	0,6036808	70,36463

GOTERM	GO:0001741 ~XY body	2	1,4706	0,093297	H2AFX, PBX4	128	14	18224	20,3392857	1	0,6036808	70,36463
_CC_DIRECT												
GOTERM	GO:0001673 ~male germ cell nucleus	2	1,4706	0,1059	HSPA2, H2AFX	128	16	18224	17,796875	1	0,6348478	75,093
_CC_DIRECT												
UP_KEYWORDS	Cell division	6	4,4118	0,109779	PIK3C3, TPX2, CENPE, HAUS8, MAP9, CDK5	135	388	20581	2,35750286	1	0,6296389	76,8556
UP_KEYWORDS	Primary microcephaly	2	1,4706	0,122522	CENPE, CDK5RAP2	135	20	20581	15,2451852	1	0,6575976	80,69595
GOTERM	GO:0000242 ~pericentriolar material	2	1,4706	0,12448	CDK5RAP2, PCMI	128	19	18224	14,9868421	1	0,6808365	80,81011
_CC_DIRECT												
UP_KEYWORDS	Cell cycle	8	5,8824	0,132234	PIK3C3, TPX2, H2AFX, CENPE, HAUS8, MAP9, CDK5, SUV39H2	135	650	20581	1,87633048	1	0,6731657	83,21898
UP_KEYWORDS	Hypotrichosis	2	1,4706	0,139567	KRT74, DSG1	135	23	20581	13,2566828	1	0,6806024	84,91881
GOTERM	GO:0035869 ~ciliary transition zone	2	1,4706	0,154599	B9D2, PCMI	128	24	18224	11,8645833	1	0,7484366	87,5754
_CC_DIRECT												
GOTERM	GO:0005829 ~cytosol	29	21,324	0,154941	YWHAZ, GFAP, POTEKP, TALDO1, XIAP, B9D2, SFN, HSPA2, PIK3C3, ETNK1, CDK5RAP2, PSMD8, PIK3CG, EEF1A1, OBSCN, GMD5, CEP89, HMBS, PAIP1, TPX2, CENPE, PCMI, CDK5, ARHGEF11, MFN2, BLVRA, DSG1, HAUS8, WDR45	128	3315	18224	1,24551282	1	0,7343954	87,63771
_CC_DIRECT												
UP_KEYWORDS	Electron transport	3	2,2059	0,156289	SDHB, NDUFS8, ETFDH	135	108	20581	4,23477366	1	0,7118413	88,21923
GOTERM	GO:0030016 ~myofibril	2	1,4706	0,177952	OBSCN, MYH13	128	28	18224	10,1696429	1	0,7726872	91,22569
_CC_DIRECT												
UP_KEYWORDS	GTP-binding	5	3,6765	0,185415	MFN2, GFM2, EEF1A1, EEF1A1P5, EEF1A2	135	343	20581	2,2223302	1	0,7653534	92,4287
UP_KEYWORDS	Repressor	7	5,1471	0,189433	SBNO2, RCOR3, NCOA5, CHCHD3, NGDN, FAM208A, SUV39H2	135	592	20581	1,80264014	1	0,7619186	92,8855
UP_KEYWORDS	Cell projection	8	5,8824	0,190819	MACF1, B9D2, NGDN, BASP1, PCMI, CCDC39, CDK5, SHANK3	135	721	20581	1,69156008	1	0,7534447	93,03714
GOTERM	GO:0035861 ~site of double-strand break	2	1,4706	0,195045	H2AFX, PARP3	128	31	18224	9,18548387	1	0,7934474	93,2409
_CC_DIRECT												
UP_KEYWORDS	Kinase	8	5,8824	0,203478	PIK3CG, MAP3K6, OBSCN, YWHAZ, STYK1, PIK3C3, ETNK1, CDK5	135	735	20581	1,65933988	1	0,7671652	94,29035
UP_KEYWORDS	Ubiquinone	2	1,4706	0,204524	NDUFS8, ETFDH	135	35	20581	8,71153439	1	0,7586303	94,384
UP_KEYWORDS	Autophagy	3	2,2059	0,207614	MFN2, PIK3C3, WDR45	135	130	20581	3,51811966	1	0,7541616	94,65239
GOTERM	GO:0031965 ~nuclear membrane	4	2,9412	0,214527	EVX1, CEP170, DUX4, PCMI	128	229	18224	2,48689956	1	0,8155333	95,01391
_CC_DIRECT												
UP_KEYWORDS	Centromere	3	2,2059	0,224318	NGDN, SHANK, SUV39H2	135	137	20581	3,33836172	1	0,7741275	95,91009
UP_KEYWORDS	Autism spectrum disorder	2	1,4706	0,230131	EEF1A2, SHANK3	135	40	20581	7,62259259	1	0,7744682	96,27952
GOTERM	GO:0030658 ~transport vesicle membrane	2	1,4706	0,233571	SCG3, SLC30A8	128	38	18224	7,49342105	1	0,833971	96,32385
_CC_DIRECT												
UP_KEYWORDS	Mitosis	4	2,9412	0,243549	TPX2, CENPE, HAUS8, MAP9	135	262	20581	2,32750919	1	0,7869999	97,01805
UP_KEYWORDS	ANK repeat	4	2,9412	0,247022	POTEF, ANKRD42, SHANK3, POTEE	135	264	20581	2,30987654	1	0,783592	97,18582
GOTERM	GO:0015630 ~microtubule cytoskeleton	3	2,2059	0,247459	MFN2, TPX2, CENPE	128	137	18224	3,11770073	1	0,8432098	97,07059
_CC_DIRECT												
UP_KEYWORDS	Cilium biogenesis /degradation	3	2,2059	0,250745	CEP89, B9D2, PCMI	135	148	20581	3,09024024	1	0,7807183	97,356
UP_KEYWORDS	Transcription regulation	19	13,971	0,256465	EEF1A1, SBNO2, DUX4L9, IKZF2, RCOR3, CHCHD3, ZNF619, DUX4L2, ZNF836, DUX4L4, PAX1, FAM208A, SUV39H2, CHD7, NCOA5, DUX1, DUX4, PBX4, ZNF586	135	2332	20581	1,24210342	1	0,7810126	97,5991
UP_KEYWORDS	GO:0030864 ~cortical actin cytoskeleton	2	1,4706	0,259964	EEF1A1, CDH2	128	43	18224	6,62209302	1	0,8499309	97,62094
_CC_DIRECT												
UP_KEYWORDS	Actin-binding	4	2,9412	0,264496	MACF1, MICAL3, MYH13, SHANK3	135	274	20581	2,22557448	1	0,7847593	97,90579
GOTERM	GO:0005876 ~spindle microtubule	2	1,4706	0,265134	XIAP, BIRC8	128	44	18224	6,47159091	1	0,8471372	97,81927
_CC_DIRECT												
UP_KEYWORDS	Alternative splicing	74	54,412	0,268206	GFAP, EVX1, GARI, INTS3, CCDC39, PAX1, CBFB, MAP3K6, H2AFV, NUDCD1, SBSN, TIGD5, CDK5RAP2, M1AP, RCOR3, CEP89, HMBS, MICAL3, TPX2, UBR4, H2AF1, BASP1, PCMI, ELMO3, CDK5, ARHGEF11, MFN2, DCAF8, ATP9B, PCMTD2, HAUS8, WDR45, ZNF586, ADAD2, DCD, YWHAZ, SBNO2, ZNF619, SFN, CDH2, GFM2, CHD7, ABHD16A, MACF1, CEP170, SYBU, ETFDH, STAMBPL1, SCG3, ETNK1, PNMA1, DUX1, SLC30A8, DUX3, POTEE, DUX5, EEF1A1, OBSCN, IKZF2, GMD5, PAIP1, CENPE, ANKRD42, FAM208A, SUV39H2, KRT73, KRT72, SUGP2, CCDC110, DSG1, FBLN7, NGDN, PARP3, MAP9	135	10587	20581	1,06559337	1	0,7821851	98,03492
UP_KEYWORDS	Differentiation	7	5,1471	0,26824	M1AP, OBSCN, SBNO2, HSPA2, CASP14, UNC13A, SUV39H2	135	667	20581	1,59994447	1	0,7743764	98,03607
UP_KEYWORDS	Immunoglobulin V region	2	1,4706	0,26941	IGKV2-30, IGKV2D-28	135	48	20581	6,35216049	1	0,768344	98,07521
UP_KEYWORDS	Apoptosis	6	4,4118	0,27101	MFN2, XIAP, TPX2, BIRC8, CDK5, ELMO3	135	536	20581	1,70655058	1	0,7630694	98,12758
UP_KEYWORDS	Developmental protein	9	6,6176	0,277772	OBSCN, NES, HRNR, GSC, EVX1, EVX2, CENPE, PRRX2, PAX1	135	949	20581	1,4458026	1	0,765481	98,3348
GOTERM	GO:0043209 ~myelin sheath	3	2,2059	0,286184	GFAP, HSPA2, EEF1A2	128	152	18224	2,81003289	1	0,8634649	98,47992
_CC_DIRECT												
UP_KEYWORDS	Transcription	19	13,971	0,296407	EEF1A1, SBNO2, DUX4L9, IKZF2, RCOR3, CHCHD3, ZNF619, DUX4L2, ZNF836, DUX4L4, PAX1, FAM208A, SUV39H2, CHD7, NCOA5, DUX1, DUX4, PBX4, ZNF586	135	2398	20581	1,20791709	1	0,7841951	98,80162
UP_KEYWORDS	S-nitrosylation	2	1,4706	0,297558	XIAP, BIRC8	135	54	20581	5,64636488	1	0,778742	98,82605
UP_KEYWORDS	Neurodegeneration	4	2,9412	0,298074	MFN2, DCAF8, CDK5, WDR45	135	293	20581	2,08125395	1	0,7725252	98,83686

S4 Table. DAVID analysis identifies 5 enriched clusters of TBX18 interacting proteins.

All 143 identified proteins were subjected to a Functional Annotation Clustering (DAVID Bioinformatics Resources 6.8) using the following settings: Similarity Term Overlap 7; Similarity Threshold 0.35; Initial Group Membership 2; Final Group Membership 2; Multiple Linkage Threshold 0.15; EASE 0.3; UP_KEYWORDS and GOTHERM_CC_DIRECT. 5 clusters were identified with enrichment scores for cluster 1 of 6.07, for cluster 2 of 3.48, for cluster 3 of 1.85, for cluster 4 of 1.35 and cluster 5 of 1.19.

Annotation Cluster 1		Enrichment Score: 6.069435122354389														
Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR				
UP_KEYWORDS	Nucleosome core	12	8,89	1,28E-11	HIST1H2AB, HIST1H2AC, HIST2H2AB, H2AFV, HIST1H2AG, HIST2H2AC, H2AFZ, HIST1H2AH, H2AFX, HIST3H2A, HIST1H2AK, H2AFJ	134	89	20581	20,7087	2,63E-09	2,63E-09	1,62E-08				
GOTERM_CC_DIRECT	GO:0000786 ~nucleosome	11	8,15	8,90E-10	HIST1H2AB, HIST1H2AC, HIST2H2AB, H2AFV, HIST1H2AG, HIST2H2AC, H2AFZ, HIST1H2AH, H2AFX, HIST1H2AK, H2AFJ	126	94	18224	16,9254	1,67E-07	1,67E-07	1,10E-06				
UP_KEYWORDS	Chromosome	16	11,9	4,96E-08	HIST1H2AB, HIST1H2AC, HIST1H2AG, CENPE, H2AFJ, FAM208A, SUV39H2, HIST2H2AB, H2AFV, HIST2H2AC, H2AFZ, NGDN, HIST1H2AH, HIST1H2AK, HIST3H2A, H2AFX	134	400	20581	6,14358	1,02E-05	5,08E-06	6,24E-05				
UP_KEYWORDS	Citullination	9	6,67	2,19E-07	HIST1H2AB, HIST1H2AC, HIST2H2AB, GFAP, HIST1H2AG, HIST2H2AC, HIST1H2AH, HIST3H2A, HIST1H2AK	134	97	20581	14,2506	4,50E-05	1,50E-05	2,76E-04				
UP_KEYWORDS	Methylation	23	17	4,49E-07	HIST1H2AB, EEF1A1, HIST1H2AC, GFAP, HRNR, POTEKP, RCOR3, HIST1H2AG, EEF1A2, PAIP1, CENPE, H2AFJ, SHANK3, HIST2H2AB, CHD7, HSPA2, EEF1A1P5, HIST2H2AC, H2AFZ, HIST1H2AH, MYH13, HIST1H2AK, HIST3H2A	134	1001	20581	3,52903	9,21E-05	2,30E-05	5,66E-04				
GOTERM_CC_DIRECT	GO:0000790 ~nuclear chromatin	10	7,41	7,39E-06	HIST1H2AB, HIST1H2AC, HIST2H2AB, H2AFV, HIST1H2AG, HIST2H2AC, HIST1H2AH, H2AFX, HIST1H2AK, H2AFJ	126	193	18224	7,49404	0,0013878	4,63E-04	0,009164				
UP_KEYWORDS	Isopeptide bond	21	15,6	3,95E-05	HIST1H2AB, HIST1H2AC, NES, IKZF2, XIAP, HIST1H2AG, TPX2, CENPE, H2AFJ, BASP1, HIST2H2AB, HIST2H2AC, H2AFZ, BIRC8, HIST1H2AH, MYH13, HIST1H2AK, HIST3H2A, H2AFX, POTEF, POTE	134	1132	20581	2,84928	0,008058	0,0010108	0,049655				
UP_KEYWORDS	Ubl conjugation	26	19,3	8,54E-05	HIST1H2AB, EEF1A1, HIST1H2AC, NES, IKZF2, XIAP, HIST1H2AG, TPX2, CENPE, H2AFJ, BASP1, SFN, PCM1, ARHGEF11, MFN2, HIST2H2AB, H2AFV, HIST2H2AC, H2AFZ, HIST1H2AH, BIRC8, MYH13, H2AFX, HIST3H2A, HIST1H2AK, POTEF	134	1705	20581	2,34213	0,0173473	0,0017484	0,107369				
UP_KEYWORDS	Phospho protein	74	54,8	3,73E-04	GFAP, HRNR, KPRP, INTS3, CCDC39, CBF, HIST2H2AB, MAP3K6, NUDCD1, HIST2H2AC, PIK3C3, H2AFX, HIST3H2A, ZRSR2, CDK5RAP2, MDN1, PSMD8, PIK3CG, RCOR3, CEP89, HMBS, MICAL3, TPX2, UBR4, H2AFJ, BASP1, PCM1, CDK5, ARHGEF11, MFN2, NCOA5, HAUS8, UNC13A, HIST1H2AB, HIST1H2AC, YWHAZ, TALDO1, XIAP, HIST1H2AG, CHCHD3, TOR4A, FLG2, SFN, CDH2, CHD7, LAMB2, HSPA2, MACF1, CEP170, SYBU, ETFDH, STAMBPL1, SCG3, EEF1A1, OBSCN, NES, IKZF2, GMD5, EEF1A2, CENPE, SHANK3, FAM208A, SUV39H2, BLVRA, PLEKHA6, SUGP2, CCDC110, KRT74, DSG1, NGDN, HIST1H2AH, MYH13, HIST1H2AK, MAP9	134	8246	20581	1,37832	0,0735521	0,0069212	0,467894				
UP_KEYWORDS	Acetylation	39	28,9	3,90E-04	HIST1H2AB, DCD, HIST1H2AC, YWHAZ, TALDO1, HIST1H2AG, CHCHD3, INTS3, HIST2H2AB, MACF1, H2AFV, HIST2H2AC, ETFDH, STAMBPL1, H2AFZ, H2AFX, HIST3H2A, MDN1, EEF1A1, NES, IKZF2, GMD5, EEF1A2, HMBS, PAIP1, UBR4, TPX2, H2AFJ, PCM1, CDK5, FAM208A, BLVRA, SDHB, NCOA5, HIST1H2AH, NGDN, HIST1H2AK, HAUS8, MAP9	134	3424	20581	1,74941	0,0768553	0,0066419	0,489716				
Annotation	Cluster 2	Enrichment Score: 3.4785277413497244														
Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR				
UP_KEYWORDS	Homeobox	12	8,89	1,11E-06	DUX4L9, GSC, EVX1, EVX2, DUX4L2, DUX1, PRRX2, DUX4, DUX4L4, DUX3, PBX4, DUX5	134	262	20581	7,03464	2,27E-04	4,54E-05	0,001393				
UP_KEYWORDS	DNA-binding	31	23	1,48E-05	HIST1H2AB, HIST1H2AC, DUX4L9, EVX1, HIST1H2AG, EVX2, ZNF619, DUX4L2, PRRX2, DUX4L4, PAX1, HIST2H2AB, CHD7, H2AFV, HIST2H2AC, H2AFZ, H2AFX, HIST3H2A, TIGD5, DUX1, DUX4, DUX3, DUX5, IKZF2, GSC, H2AFJ, ZNF836, HIST1H2AH, HIST1H2AK, ZNF586, PBX4	134	2050	20581	2,32257	0,0030219	4,32E-04	0,018577				
GOTERM_CC_DIRECT	GO:0005634 ~nucleus	60	44,4	2,69E-05	HRNR, GAR1, INTS3, PRRX2, PAX1, CBF, HIST2H2AB, NUDCD1, H2AFV, HIST2H2AC, H2AFZ, BIRC8, H2AFX, TIGD5, MDN1, PSMD8, GSC, RCOR3, MICAL3, TPX2, H2AFJ, BASP1, CDK5, NCOA5, EEF1A1P5, ZNF586, HIST1H2AB, HIST1H2AC, YWHAZ, DUX4L9, TALDO1, XIAP, HIST1H2AG, B9D2, CHCHD3, FLG2, ZNF619, SFN, DUX4L2, DUX4L4, CHD7, HSPA2, DUX1, DUX4, DUX3, DUX5, EEF1A1, IKZF2, EEF1A2, CENPE, ZNF836, FAM208A, SUV39H2, KRT73, CCDC110, CASP14, HIST1H2AH, HIST1H2AK, PARP3, PBX4	126	5415	18224	1,6026	0,0050444	0,0012635	0,033369				
UP_KEYWORDS	Nucleus	56	41,5	4,36E-05	EVX1, EVX2, GAR1, INTS3, PRRX2, PAX1, CBF, HIST2H2AB, NUDCD1, H2AFV, HIST2H2AC, H2AFZ, H2AFX, HIST3H2A, ZRSR2, TIGD5, MDN1, GSC, RCOR3, MICAL3, UBR4, TPX2, H2AFJ, CDK5, NCOA5, ZNF586, HIST1H2AB, HIST1H2AC, DUX4L9, XIAP, HIST1H2AG, B9D2, CHCHD3, ZNF619, SFN, DUX4L2, DUX4L4, CHD7, DUX1, DUX4, DUX3, DUX5, EEF1A1, IKZF2, EEF1A2, ZNF836, FAM208A, SUV39H2, SUGP2, CCDC110, CASP14, HIST1H2AH, NGDN, HIST1H2AK, PARP3, PBX4	134	5244	20581	1,64016	0,008892	9,92E-04	0,054815				
UP_KEYWORDS	Transcription regulation	19	14,1	0,24624	EEF1A1, SBNO2, DUX4L9, IKZF2, RCOR3, CHCHD3, ZNF619, DUX4L2, ZNF836, DUX4L4, PAX1, FAM208A, SUV39H2, CHD7, NCOA5, DUX1, DUX4, PBX4, ZNF586	134	2332	20581	1,25137	1	0,9282105	97,1486				
UP_KEYWORDS	Transcription	19	14,1	0,2853	EEF1A1, SBNO2, DUX4L9, IKZF2, RCOR3, CHCHD3, ZNF619, DUX4L2, ZNF836, DUX4L4, PAX1, FAM208A, SUV39H2, CHD7, NCOA5, DUX1, DUX4, PBX4, ZNF586	134	2398	20581	1,21693	1	0,9432469	98,54043				
Annotation	Cluster 3	Enrichment Score: 1.8477005499067418														
Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR				

UP_KEYWORDS	Cytoskeleton	17	12,6	0,00278	POTEKP, CEP89, B9D2, MICAL3, UBR4, TPX2, CENPE, CCDC39, PCM1, HSPA2, MACF1, CEP170, SYBU, PARP3, CDK5RAP2, HAUS8, MAP9	134	1138	20581	2,2944	0,435142	0,0429857	3,445629
UP_KEYWORDS	Microtubule	7	5,19	0,00979	MACF1, CEP170, TPX2, CENPE, HAUS8, CDK5RAP2, MAP9	134	280	20581	3,83974	0,8670034	0,1258432	11,6485
GOTERM_CC_DIRECT	GO:0005874	7	5,19	0,02038	MACF1, CEP170, SYBU, TPX2, CENPE, HAUS8, CDK5RAP2	126	311	18224	3,25545	0,9791666	0,5389458	22,54525
GOTERM_CC_DIRECT	~microtubule GO:0005813~centrosome	7	5,19	0,07322	CEP89, CEP170, B9D2, UBR4, HAUS8, CDK5RAP2, PCM1	126	426	18224	2,37663	0,9999994	0,8702683	61,07183
Annotation	Cluster 4	Enrichment Score: 1.3529513190210538										
Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KEYWORDS	Cytoplasm	45	33,3	0,0069	POTEKP, YWHAZ, GFAP, TALDO1, XIAP, B9D2, CHCHD3, KPRP, FLG2, SFN, CCDC39, HSPA2, NUDCD1, MACF1, CEP170, SYBU, ETKN1, BIRC8, CDK5RAP2, POTEF, PIK3CG, EEF1A1, OBSCN, CEP89, PAIP1, HMBS, MICAL3, UBR4, TPX2, CENPE, PCM1, CDK5, ELMO3, SHANK3, ARHGEF11, BLVRA, CASP14, EEF1A1P5, PCMTD2, NGDN, MYH13, PARP3, HAUS8, MAP9, UNC13A	134	4816	20581	1,43512	0,7580688	0,0963963	8,342966
GOTERM_CC_DIRECT	GO:0005737~cytoplasm	44	32,6	0,09389	HRNR, YWHAZ, GFAP, TALDO1, XIAP, CHCHD3, KPRP, FLG2, SFN, CDH2, CCDC39, NUDCD1, MACF1, CEP170, ETKN1, BIRC8, CDK5RAP2, MDN1, PIK3CG, EEF1A1, NES, GMD5, CEP89, PAIP1, EEF1A2, HMBS, MICAL3, UBR4, CENPE, BASP1, PCM1, CDK5, ELMO3, SHANK3, ARHGEF11, BLVRA, CCDC110, KRT74, CASP14, EEF1A1P5, PCMTD2, NGDN, PARP3, UNC13A	126	5222	18224	1,21868	1	0,9014213	70,57171
GOTERM_CC_DIRECT	GO:0005829~cytosol	29	21,5	0,13483	YWHAZ, GFAP, POTEKP, TALDO1, XIAP, B9D2, SFN, HSPA2, PIK3C3, ETKN1, CDK5RAP2, PSMD8, PIK3CG, EEF1A1, OBSCN, GMD5, CEP89, HMBS, PAIP1, TPX2, CENPE, PCM1, CDK5, ARHGEF11, MFN2, BLVRA, DSG1, HAUS8, WDR45	126	3315	18224	1,26528	1	0,9514563	83,41815
Annotation	Cluster 5	Enrichment Score: 1.1855762592464065										
Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KEYWORDS	Nucleotide - binding	20	14,8	0,02129	PIK3CG, OBSCN, EEF1A1, POTEKP, EEF1A2, TOR4A, CENPE, CDK5, MFN2, MAP3K6, GFM2, STYK1, CHD7, HSPA2, ATP9B, EEF1A1P5, PIK3C3, ETKN1, MYH13, MDN1	134	1788	20581	1,718	0,9878566	0,2409484	23,72176
UP_KEYWORDS	ATP-binding	15	11,1	0,06569	PIK3CG, MAP3K6, OBSCN, POTEKP, STYK1, CHD7, HSPA2, ATP9B, PIK3C3, ETKN1, TOR4A, MYH13, CENPE, CDK5, MDN1	134	1391	20581	1,65625	0,9999991	0,5592737	57,47411
UP_KEYWORDS	Kinase	8	5,93	0,19846	PIK3CG, MAP3K6, OBSCN, YWHAZ, STYK1, PIK3C3, ETKN1, CDK5	134	735	20581	1,67172	1	0,8846197	93,82069

S5 Table. Summary of expression of candidate proteins for TBX18 interaction in various assays.

(A) Cloning and expression of candidate transcriptional cofactors in 293 cells. Shown are the name of the proteins, the species they are derived from (m, mouse; h, human), the expected molecular weight (in kDa), the plasmids used for expression of the proteins in 293 cells, the antibodies used to detect the tag on the candidate proteins, and the reference for the used plasmids. (B-D) Cloning and *in vitro* translation of candidate transcriptional cofactors (B), homeobox transcription factors (C) and centrosomal proteins (D). Shown are the name of the proteins, the species they are derived from (m, mouse; h, human), the expected molecular weight (in kDa), the plasmids used for expression of the proteins in reticulocyte lysate, the polymerase used for the *in vitro* transcription reaction, and the reference for the used plasmids. (E) Cloning and expression of candidate “centrosomal” proteins in 293 cells. Shown are the name of the proteins, the species they are derived from (m, mouse; h, human), the expected molecular weight (in kDa), the plasmids used for expression of the proteins in 293 cells, the antibodies used to detect the tag on the candidate proteins, and the reference for the used plasmids.

A Western Blot, GST pull-downs, localization and recruitment assays with transcriptional cofactor candidates expressed in 293 cells						
Protein	Species	Molecular weight (kDa)	Plasmid	Antibody	Plasmid reference	
NKX2.5	h	34.1	<i>pcDNA3.Nkx2.5.HA</i>	Anti-MYC	this work	
BASP1	h	22.7	<i>pcDNA3.Basp1.FLAG.HA</i>	Anti-HA	Goodfellow et al., 2011	
CBFB	m	22	<i>pcDNA3.Cbfb.myc</i>	Anti-MYC	this work	
CHD7	h	336	<i>pReceiver.3xHA.Chd7</i>	Anti-HA	Liu et al., 2014	
GAR1	m	23.5	<i>pcDNA3.Gar1.myc</i>	Anti-MYC	this work	
IKZF2	m	59.4	<i>pCMV6.Ikzf2.myc.FLAG</i>	Anti-MYC	Origene Technologies, Rockville, USA	
NCOA5	m	65.3	<i>pCMV6.Ncoa5.myc.FLAG</i>	Anti-MYC	Origene Technologies, Rockville, USA	
RCOR3	m	49.8	<i>pEF4.Rcor3.myc.His</i>	Anti-MYC	Upadhyay et al., 2014	
SBNO2	h	150.3	<i>pLenti.GIII.CMV.Sbno2.HA</i>	Anti-HA	Maruyama et al., 2013	
SUV39H2	m	54.1	<i>pCMV6.Suv39h2.myc.FLAG</i>	Anti-MYC	Origene Technologies, Rockville, USA	
SSXB2	m	15.3	<i>pCMV6.Ssxb2.myc.FLAG</i>	Anti-MYC	Origene Technologies, Rockville, USA	

B GST pull-downs with radioactively labeled in vitro translated transcriptional cofactor candidates						
Protein	Species	Molecular weight (kDa)	Plasmid	In vitro transcription	Reference	
NKX2.5	h	34.1	<i>pSP64.Nkx2.5.myc</i>	SP6	this work	
BASP1	h	22.7	<i>pSP64.Basp1.myc</i>	SP6	this work	
CBFB	m	22	<i>pSP64.Cbfb.myc</i>	SP6	this work	
CHD7-1	h	87.4	<i>pGBKT7.Chd7-1</i>	T7	Schulz et al., 2014	
CHD7-2	h	98.9	<i>pGBKT7.Chd7-2</i>	T7	Schulz et al., 2014	
CHD7-3	h	98.8	<i>pGBKT7.Chd7-3</i>	T7	Schulz et al., 2014	
CHD7-4	h	74	<i>pGBKT7.Chd7-4</i>	T7	Schulz et al., 2014	
GAR1	m	23.5	<i>pSP64.Gar1.myc</i>	SP6	this work	
IKZF2	m	59.4	<i>pCMV6.Ikzf2.myc.FLAG</i>	T7	Origene Technologies, Rockville, USA	
NCOA5	m	65.3	<i>pSP64.Ncoa5.myc</i>	SP6	this work	
RCOR3	m	49.8	<i>pEF4.Rcor3.myc.His</i>	T7	Upadhyay et al., 2014	
SBNO2	h	150.3	<i>pEF4.Sbno2.myc.His</i>	T7	this work	
SUV39H2	m	54.1	<i>pCMV6.Suv39h2.myc.FLAG</i>	T7	Origene Technologies, Rockville, USA	
SSXB2	m	15.3	<i>pCMV6.Ssxb2.myc.FLAG</i>	SP6	Origene Technologies, Rockville, USA	

C	GST pull-downs with radioactively labeled in vitro translated homeobox transcription factors					
	Protein	Species	Molecular weight (kDa)	Plasmid	In vitro transcription	Reference
	DUXBLI	m	41	<i>Topo.V5.His.Duxbl-L</i>	T7	Wu et al., 2010
	GSC	m	28	<i>pSP64.Goosecoid.myc</i>	SP6	this work
	PAX1	m	46.3	<i>pSP64.Pax1.myc</i>	SP6	this work
	PBX1	m	46.6	<i>pCS2.Pbx1a</i>	SP6	Shanmugam et al., 1999
	PBX4	h	40.8	<i>pCMV6-Entry-Pbx4</i>	T7	Origene Technologies, Rockville, USA
	PRRX2	m	26.4	<i>pSP64.Prrx2.myc</i>	SP6	this work
D	GST pull-downs with radioactively labeled in vitro translated "centrosomal" proteins					
	Protein	Species	Molecular weight (kDa)	Plasmid	In vitro transcription	Reference
	B9D2	m	19.2	<i>pSP64.B9d2.myc</i>	SP6	this work
	CEP89	h	89.6	<i>pSP64.Cep89.myc</i>	SP6	this work
	HAUS8	h	44.8	<i>pSP64.Haus8.myc</i>	SP6	this work
	MAP9	h	74.2	<i>pcDNA3.Flag.HA.ASAP</i>	T7	Saffin et al., 2005
	MICAL3	h	224.3	<i>pSP64.Mical3.HA</i>	SP6	this work
	PARP3	h	60	<i>pSP64.Parp3.myc</i>	SP6	this work
	PCM1	h	228.5	<i>pSP64.Pcm1.HA</i>	SP6	this work
	SYBU	h	72.4	<i>pSP64.Sybu.myc</i>	SP6	this work
	TPX2	h	85.6	<i>pSP64.Tpx2.myc</i>	SP6	this work
E	Western Blot, localization and recruitment assays with "centrosomal" candidates expressed in 293 cells					
	Protein	Species	Molecular weight (kDa)	Plasmid	Antibody	Reference
	B9D2	m	19. Feb	<i>pCDNA3.B9d2.myc</i>	Anti-MYC	this work
	HAUS8	h	44.8	<i>pCDNA3.Haus8.HA</i>	Anti-HA	this work
	MAP9	h	74.2	<i>pcDNA3.Flag.HA.Asap</i>	Anti-HA	Saffin et al., 2005
	PARP3	h	60	<i>pCD2E.Parp3.Flag</i>	Anti-FLAG	Dai et al., 2015
	SYBU	h	72.4	<i>pCDNA3.Sybu.HA</i>	Anti-HA	this work
	TPX2	h	85.6	<i>pCDNA3.Tpx2.myc</i>	Anti-MYC	this work

References

- Dai, X., Rulten, S. L., You, C., Caldecott, K. W., and Wang, Y. (2015) Identification and Functional Characterizations of N-Terminal alpha-N-Methylation and Phosphorylation of Serine 461 in Human Poly(ADP-ribose) Polymerase 3. *J Proteome Res* 14, 2575-2582
- Farin, H.F., Bussen, M., Schmidt, M. K., Singh, M. K., Schuster-Gossler, K. and Kispert, A. (2007) Transcriptional Repression by the T-box Proteins Tbx18 and Tbx15 Depends on Groucho Corepressors. *J. Biol. Chem.* 282, 25748–25759
- Goodfellow, S. J., Rebello, M. R., Toska, E., Zecf, L. A., Rudd, S. G., Medler, K. F., and Roberts, S. G. (2011) WT1 and its transcriptional cofactor BASP1 redirect the differentiation pathway of an established blood cell line. *Biochem J* 435, 113-125
- Grigoriev, I., Yu, K. L., Martinez-Sanchez, E., Serra-Marques, A., Smal, I., Meijering, E., Demmers, J., Peranen, J., Pasterkamp, R. J., van der Sluijs, P., Hoogenraad, C. C., and Akhmanova, A. (2011) Rab6, Rab8, and MICAL3 cooperate in controlling docking and fusion of exocytotic carriers. *Curr Biol* 21, 967-974
- Lawo, S., Bashkurov, M., Mullin, M., Ferreria, M. G., Kittler, R., Habermann, B., Pinchev, D., Buchholz, F., Peters, J.M., Hyman, A.A., Gingras, A.C., Pelletier, L. (2009). HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity. *Curr. Biol.*, 19, 816-826.
- Liu, Y., Harmelink, C., Peng, Y., Chen, Y., Wang, Q., and Jiao, K. (2014) CHD7 interacts with BMP R-SMADs to epigenetically regulate cardiogenesis in mice. *Hum Mol Genet* 23, 2145-2156
- Maruyama, K., Uematsu, S., Kondo, T., Takeuchi, O., Martino, M. M., Kawasaki, T., & Akira, S. (2013). Strawberry notch homologue 2 regulates osteoclast fusion by enhancing the expression of DC-STAMP. *J. Exp. Med.*, 210, 1947-1960.
- O'Brien, L. L., Guo, Q., Bahrami-Samani, E., Park, J. S., Hasso, S. M., Lee, Y. J., Fang, A., Kim, A. D., Guo, J., Hong, T. M., Peterson, K. A., Lozanoff, S., Raviram, R., Ren, B., Fogelgren, B., Smith, A. D., Valouev, A., and McMahon, A. P. (2018) Transcriptional regulatory control of mammalian nephron progenitors revealed by multi-factor cistronic analysis and genetic studies. *Plos Genet* 14, e1007181
- O'Neill, D. W., Schoetz, S. S., Lopez, R. A., Castle, M., Rabinowitz, L., Shor, E., Krawchuk, D., Goll, M.G., Renz, M., Seelig, H.P., Han, S., Seong, R.H., Park, S.D., Agaloti, T., Munshi, N., Thanos, D. Erdjument-Bromage, H., Tempst, P., Bank, A. (2000). An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells. *Mol. Cell. Biol.*, 20, 7572-7582.
- Saffin, J. M., Venoux, M., Prigent, C., Espeut, J., Poulat, F., Giorgi, D., Abrieu, A., and Rouquier, S. (2005) ASAP, a human microtubule-associated protein required for bipolar spindle assembly and cytokinesis. *Proc Natl Acad Sci U S A* 102, 11302-11307
- Schulz, Y., Freese, L., Mänz, J., Zoll, B., Völter, C., Brockmann, K., Bögershausen, N., Becker, J., Wollnik, B., Pauli, S. (2014) CHARGE and Kabuki syndromes: a phenotypic and molecular link. *Hum. Mol. Gen.* 23, 4396-4405.
- Shanmugam, K., Green, N. C., Rambaldi, I., Saragovi, H. U., and Featherstone, M. S. (1999) PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins. *Mol Cell Biol* 19, 7577-7588
- Sillibourne, J. E., Specht, C. G., Izeddin, I., Hurbain, I., Tran, P., Triller, A., Darzacq, X., Dahan, M., and Bormens, M. (2011) Assessing the localization of centrosomal proteins by PALM/STORM nanoscopy. *Cytoskeleton (Hoboken)* 68, 619-627
- Upadhyay, G., Chowdhury, A. H., Vaidyanathan, B., Kim, D., and Saleque, S. (2014) Antagonistic actions of Rcor proteins regulate LSD1 activity and cellular differentiation. *Proc Natl Acad Sci U S A* 111, 8071-8076
- Wu, S. L., Tsai, M. S., Wong, S. H., Hsieh-Li, H. M., Tsai, T. S., Chang, W. T., Huang, S. L., Chiu, C. C., and Wang, S. H. (2010) Characterization of Genomic Structures and Expression Profiles of Three Tandem Repeats of a Mouse Double Homeobox Gene: Duxbl. *Dev Dynam* 239, 927-940

S6 Table. Statistical analysis of transactivation assays of candidate transcriptional cofactors with TBX18 in 293 cells.

Shown are the experimental setup for each sample of the transactivation assays and the statistical analysis performed to the transactivation assay results. Raw luminescence data was normalized to the average of the reporter alone measurements. Further, average and standard deviation were quantified from the two technical and two experimental replicates. The significance of luciferase expression between the TBX18 induced repression and the reporter and the effects of the cofactors on TBX18 repression was evaluated by a student t-test (TTEST) carried out according to an analysis of variance equality (FTEST). p-values<0.05 (*) were considered significant, while p-values<0.01 (**) and p-values<0.0001 (***) were considered very significant.

BASP1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
-------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Part 2 - Characterizing endogenous TBX18 protein interaction partners

"Proteomic analysis identifies ZMYM2 as endogenous binding partner of TBX18"

In this paper I performed all experiments except the initial genetic engineering of the cells and the proteomic screening. I also performed data analysis and writing.

Proteomic analysis identifies ZMYM2 as endogenous binding partner of TBX18

Reginaldo Rivera-Reyes^{1*}, Marc-Jens Kleppa^{1*}, Timo H.-W. Lüdtke¹, Friedhelm Hildebrandt² and Andreas Kispert¹

* These authors contributed equally to this work.

¹Institut für Molekularbiologie, Medizinische Hochschule Hannover, 30625 Hannover, Germany. ²Harvard Medical School and Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115, USA. Correspondence and requests for materials should be addressed to A.K. (email: kispert.andreas@mh-hannover.de)

Abstract

The TBX18 transcription factor regulates patterning and differentiation programs in the primordia of many organs yet the molecular complexes in which TBX18 resides to exert its crucial transcriptional functions in these embryonic contexts have remained elusive. Here, we used 293 and A549 cells as an accessible cell source to search for endogenous protein interaction partners of TBX18 by an unbiased proteomic approach. We tagged endogenous TBX18 by CRISPR/Cas9 targeted genome editing with a triple FLAG peptide, and identified by anti-FLAG affinity purification and subsequent LC-MS analysis the ZMYM2 protein to be statistically enriched together with TBX18 in both 293 and A549 nuclear extracts. Using a variety of assays, we confirmed binding of TBX18 to ZMYM2, a component of the CoREST transcriptional corepressor complex. *Tbx18* is coexpressed with *Zmym2* in the developing ureter of the mouse, and mutations both in *TBX18* as well as in *ZMYM2* were recently linked to congenital anomalies in the kidney and urinary tract (CAKUT) in line with a possible *in vivo* relevance of these protein interaction. TBX18 binding partners characterized in this assay are distinct from those identified with overexpressed dually tagged TBX18 in 293 cells, indicating that the design of the assay dramatically affects the recovery of interacting proteins.

Introduction

T-box (*Tbx*) genes encode a large family of proteins that are characterized by a conserved DNA-binding domain, the T-box [1]. This 180 amino acid long region recognizes a short conserved stretch of DNA, the T-box binding element (TBE), which can occur in repeats of variable spacing and orientation [2, 3]. Binding to these elements results in transcriptional changes of adjacent genes, the nature of which, i.e. activation or repression, depends on both the composition of the DNA binding site as well as on protein regions outside the T-box domain to which transcriptional cofactors can bind [3-6]. Besides DNA-binding, the T-box domain mediates interaction with other transcription factors, which is likely to increase target gene specificity. Functional analysis characterized T-box transcription factors as essential regulators of various cellular processes in diverse tissue contexts during metazoan development [4, 7]. Not surprisingly, mutations in T-box genes have been identified to cause a large set of human congenital diseases [8].

TBX18 is a member of a vertebrate-specific subgroup of T-box genes. Mutations in *TBX18* were recently recognized to underlie congenital forms of urinary tract malformations in human patients [9]. These phenotypic changes relate to a function of *TBX18* in the early development of the ureteric mesenchyme as shown by analysis of mice carrying a null allele of the gene [10, 11]. Besides dilatations of the ureter and the renal pelvis (hydroureteronephrosis) [10], mutant mice feature defects in the otic capsule and the otic fibrocyte compartment [12], in smooth muscles of the prostate [13], and in the pericardium and epicardium [14-16]. They also have a reduction of the sinoatrial node and the caval vein myocardium [17, 18], and malformation of the vertebral column and the rib cage, the latter of which causes postnatal lethality [19]. Defects in these organ systems were traced to changes in patterning and

differentiation of the respective tissue primordia during early organogenesis [10-12, 17, 19]. To date, both the molecular complexes in which TBX18 acts as well as the transcriptional targets of its activity have remained unexplored in all of its expression domains.

Murine *Tbx18* encodes a protein of 613 amino acid residues (aa) that can be divided in an N-terminal region (154 aa), a middle T-box domain (182 aa) and a large C-terminal region (277 aa) [20]. A couple of years ago, we started the biochemical characterization of the TBX18 protein. We found that the T-box domain of TBX18 preferentially binds to a palindromic array of two TBEs *in vitro*, and that the protein represses transcription upon binding to these elements in cellular transactivation assays. Transcriptional repression depends on a conserved EH1-motif in the N-terminal region that acts as a binding site for GROUCHO/TLE3 corepressors [20]. GROUCHO proteins attenuate transcription by facilitating pausing of RNA polymerase II and/or by recruiting histone deacetylases to condensate chromatin [21-24].

The N-terminal region of TBX18 also harbours a short stretch of basic amino acid residues that is necessary and sufficient to mediate nuclear localization in cells [20]. No conserved protein motifs have been identified in the large C-terminal region. However, TBX18 proteins lacking the C-terminal region lose part of their repression function indicating that additional cofactors for transcription bind to this domain [9].

In vitro, TBX18 forms homodimers and heterodimerizes with related T-box proteins. TBX18 also binds to members of the PAX, SIX and NKX homeobox families and to the Zn-finger transcription factor GATA4 [6, 20, 25, 26]. *In vivo* relevance has been shown for the interaction with PAX3 in somite patterning and with SIX1 in ureter development [25, 26].

More recently, we embarked on a proteomic approach to search for TBX18 protein interaction partners in an unbiased fashion. Using overexpressed dual tagged TBX18 as bait, we identified by tandem purification and subsequent liquid chromatography-mass spectrometry (LC-MS) analysis several transcriptional cofactors and homeobox transcription factors as preferred binding partners in 293 cells. The transcriptional cofactors CBFB, GAR1, IKZF2, NCOA5, SBNO2 and CHD7 and all identified homeobox proteins (members of the DUX, GSC, PAX, PBX1 and PRRX families) bound to the T-box of TBX18 *in vitro* and interacted with TBX18 upon coexpression in 293 cell nuclei [27].

However, overexpression in cells may force unnatural protein interactions, particularly when a relatively promiscuous protein interaction motif such as the T-box is involved. Therefore, we aimed to identify and characterize proteins with which TBX18 forms functionally relevant endogenous complexes in cells. Here, we present a proteomic screen for interaction partners of endogenously tagged TBX18 protein in 293 and A549 cells obtained through CRISPR/Cas9 gene editing. We identified the zinc finger transcription factor ZMYM2 as a binding partner of TBX18 in both cell lines, validated its binding, present data on its significance for TBX18 transcriptional function and discuss the relevance of TBX18 interaction with ZMYM2 in congenital anomalies of the kidney and the urinary tract (CAKUT).

Material and Methods

Mice

NMRI wildtype mice were housed with *ad libitum* access to food and water under conditions of regulated temperature (22°C) and humidity (50%) and a 12 h light/dark cycle. For timing of pregnancies, vaginal plugs were checked in the morning after mating and noon was designated as embryonic day (E) 0.5. Pregnant mice were sacrificed by cervical dislocation, embryos were dissected in PBS, decapitated, fixed in 4% paraformaldehyde (PFA) overnight and stored in 100% methanol at -20°C before further use. All experiments were in accordance with the German Animal Welfare Legislation (§4, TierSchG) and approved by the local Institutional Animal Care and Research Advisory Committee and permitted by the Lower Saxony State Office for Consumer Protection and Food Safety (reference number 42500/1H).

Cell culture and transient transfections

293 cells were purchased (ACC 305, DSMZ, Braunschweig, Germany), all other cell lines were kindly provided by Thomas von Hahn (Medizinische Hochschule Hannover, Germany). Cells were cultured in standard DMEM medium (#61965026, GIBCO, Carlsbad, California, USA) containing 10% fetal bovine serum (#S0115, Biochrom, Berlin, Germany) and kept in a Sanyo MCO-19AIC incubator (Sanyo, Gunma, Japan) at 37°C with 5% CO₂ under strict sterile conditions. For 293 cells, transient transfections were performed using the calcium phosphate method as previously described [28]; for A549 cells, we used Lipofectamine 3000 (#L3000008, ThermoFisher, Waltham, Massachusetts, USA). We plated 6x10⁵ cells in one well of 6-well plates and transfected them with 2.5 µg of DNA and 10 µl of Lipofectamine diluted in 1000 µl of Opti-MEM (#31985062, GIBCO). The *pd2EGFP-N1* plasmid

(Clontech, Mountain View, CA, USA) was independently transfected to check for efficiency of transfection, which was verified by epifluorescence microscopy.

Semi-quantitative reverse transcription PCR

Analysis of *TBX18* mRNA expression in human cell lines was done by semi-quantitative RT-PCR with normalization against *GAPDH* exactly as described before [27].

Generation of A549 and 293 cell lines with FLAG-tagged TBX18

We used a CRISPR/Cas9-mediated knock-in approach based on a protocol from the Medenhall and Myers laboratory [29] to tag endogenous TBX18 with a triple FLAG peptide. Briefly, the homology arms of the *TBX18* locus were PCR-amplified from genomic DNA of A549 cells and cloned into the *pFETCh-Donor* (EMM0021) plasmid (deposited by Eric Medenhall & Richard M. Myers, plasmid #63934, Addgene, Watertown, Massachusetts, USA) with Gibson Assembly Master Mix (#E2611S, NEB, Ipswich, Massachusetts, USA). Two variants were generated. For Homology Directed Repair (donor) plasmid 1, the 5'-arm was amplified using the primer pair 5'-*TCCCCGACCTGCAGCCCAGCTACAGGGTTAACTACCCTCTC*-3' and 5'-*CAAGTATCTGCACATATGGTCGGGAGCGGAGGAGGTTCCGG*-3' (homology arm length of 994 bp, **complementary to *TBX18*, complementary to *pFETCh-Donor***); the 3'-arm was amplified with primers 5'-*AGTTCTTCTGATTGGAACATCGTCTAACATATTTTCTTTTTC*-3 and 5'-*CTTATTTATATCTTGTAATTCCCTTGGAAGTCCTCTCCA*-3' (homology arm length of 1023 bp, **complementary to *TBX18*, complementary to *pFETCh-Donor***). For donor plasmid 2, the 3'-arm was amplified with 5'-

AGTTCTTCTGATTCTGAACATCA**AAAGAACTCTCTGTGGTTTA**-3' and 5'-**CTTATTTATATCTTGTAATTCCCTTGGAAGTCCTCTCCA**-3' which generates a 19 bp deletion compared to donor plasmid 1 in the 3'-UTR of *TBX18*. In all donor plasmids the endogenous *TBX18* stop codons (two consecutive stop codons) were deleted to generate a TBX18 triple FLAG (TBX18.3xFLAG) fusion protein. For details see Supplementary Table S1.

Three guide RNAs, located around the stop codon of *TBX18* were designed using the CRISPR design platform CrispRGold [30]. Guide1: 5'-CCAAATGTCATTCTAACTTAAAGG-3', location chr6:84736657-84736679, guide2: 5'-CCTTTAAGTTAAATGACATT-3', location chr6:84736657-84736679, guide3: 5'-CCACATAGCTTTTAAAAAGAAA-3', location chr6:84736613-84736635. Guide RNAs were ligated into the CRISPR/Cas9 expression plasmid *pX330 (pX330-U6-Chimeric_BB-CBh-hSpCas9)* (deposited by Feng Zhang at Addgene, plasmid #42230) following the published protocol of the Orkin laboratory [31]. For TBX18 tagging, 293 or A549 cells were transfected with one or a combination of two guide RNAs (guide1, 2, 3, 1+3 or 2+3), and selected with G418 (#A291-25, Biochrom) (300 µg/ml for 293 cells, 450 µg/ml for A549 cells) as described [29]. Surviving colonies were expanded and subsequently screened for correct integration by RFLP analysis.

Southern Blot

Genomic DNA from single cell derived cultures was screened for correct integration of the 3'-homology arm, the 5'-integration and neomycin by Southern blot analysis of *Pst*I-digested genomic DNA. The 5'-probe was amplified from the genomic region adjacent but outside the targeting vector using the primer pair 5'-GCATGGATCCGCAGAGGTCCAACAAATCAACAG-3' and 5'-

CGATGAATTCCCTAGAACTCAGGCACTAAAAG-3' and subcloned as 828 bp *Bam*HI/*Eco*RI fragment. This probe recognizes a 5.5 kbp *Pst*I fragment in the wildtype and a 2.9 kbp *Pst*I fragment in the mutant owing to the insertion of a new *Pst*I site inside the FLAG-neomycin cassette. The 3'-probe was amplified using the primer pair 5'-GCATGGATCCGGGGCCACCAAGGATTCCTG-3' and 5'-CGATGAATTCGGTGTATAATTCAAGAACAGAGAG-3' and subcloned as an 876 bp *Bam*HI/*Eco*RI fragment. This probe detects a 5.5 kbp *Pst*I fragment in the wildtype and a 3.7 kbp *Pst*I fragment in the targeted allele. The *neomycin* (*neo*) probe is a 613 bp fragment, PCR-amplified out of the donor vector using the primers 5'-GGACGAGGCAGCGCGGCTAT-3' and 5'-TCAGAAGAACTCGTCAAGAAGGC-3'. It detects a 3.7 kbp *Pst*I fragment in the targeted allele. Probe synthesis was performed with the Rediprime II DNA Labeling System (#RPN1633, Amersham, Little Chalfont, UK) from 50 ng DNA fragment, and labelled with [α -³²P]dCTP (#SCP-205, Hartmann Analytic, Braunschweig, Germany). Probe synthesis reactions were column purified with NucleoSpin gel and PCR Clean-up kit (#740609.250, Macherey-Nagel, Düren, Germany). Probe hybridization was performed in Church buffer according to standard protocols[32] before exposing the membranes to Fujifilm imaging plates (#BAS-IP MS 2025, Fujifilm, Tokio, Japan) for 3 days. Detection of the DNA fragments was done on a FLA-7000 Laser photodocumentation system (Fujifilm).

Identification of TBX18 interacting proteins in 293 and A549 cells by LC-MS

To immunoprecipitate TBX18.3xFLAG containing complexes, we followed a modified batch purification strategy [33]. 293- and A549-TBX18.3xFLAG knock-in cells along with parental control cells were used for nuclear extraction. Twenty dishes of 293 cells at 90% confluency were used to obtain 3x10E9 cells and yielded 90 mg per

sample. Ten 15 cm dishes of A549 cells at 90% confluency were used to obtain $3\text{--}4 \times 10^8$ cells and 40–45 mg of protein per sample. The resulting extracts were used to immunoprecipitate TBX18.3xFLAG containing complexes by batch purification with anti-FLAG M2 affinity gel (#A2220, Sigma Aldrich, St. Louis, Missouri, USA) according to the manufacturer's instructions. Three independent repetitions for each cell line along with parental cell controls were run in one-well 10% SDS polyacrylamide gels, and the proteins were silver stained. The gels were delivered to the MHH Proteomics facility for fractionation and liquid chromatography-mass spectrometry (LC-MS) analysis. There, proteins were mixed, alkylated by acrylamide and further processed as described [34]. Peptide samples were analyzed with a shotgun approach and data dependent analysis in a LC-MS system (RSLC, LTQ Orbitrap Velos, both Thermo Fisher) as described [34]. Raw MS data were processed using Proteome discoverer 1.4 (Thermo Scientific, Waltham, Massachusetts, USA) and Max Quant software (version 1.5) [35], and a database containing human and viral proteins and common contaminants. Proteins were stated identified by a false discovery rate of 0.01 on protein and peptide level. A two tailed Student's t-test was performed to evaluate the enrichment of proteins identified in the immunoprecipitates of 3xFLAG 293 and A549 knock-in cell lines versus the parental controls. Protein classification was done with the Uniprot database; ontological classification was performed with the database for Annotation, Visualization and Integrated Discovery (DAVID) web program but no functional clustering from the annotations was found [36].

Expression constructs

The expression plasmid for ZMYM2 was obtained from Origene (#RC205350), the KDM1A plasmid was a kind gift from Prof. Yang Shi from Harvard Medical School. Additional seamless cloning of the donor plasmid was performed with the Gibson assembly master mix from NEB (#E2611S, NEB), while ligations of the guide RNA plasmid were performed with T4 DNA ligase (#M0202, NEB). For details see Supplementary Table S1.

Antibodies

For Western blot analysis, we used mouse anti-FLAG (1:10000, #F3165, Sigma-Aldrich), rabbit anti-ZMYM2 (1:5000, #PA031765, Sigma-Aldrich), goat anti-HA-HRP (1:10000, #ab1265, Abcam), sheep anti-mouse HRP (1:10000, #515-035-003, Dianova, Hamburg, Germany) and goat anti-rabbit-HRP (1:10000, #111-035-045, Dianova). For Immunofluorescent stainings, we used mouse anti-FLAG (1:200, #F3165, Sigma-Aldrich), rat anti-HA (1:300, #11867423001, Roche), rabbit anti-ZMYM2 (1:125, #PA031765, Sigma-Aldrich), goat anti-rat-Alexa Fluor 555 (1:300, #A21434, #A21202, Invitrogen, Carlsbad, California, USA) and donkey anti-mouse-Alexa Fluor 488 (1:500, #A21202, Invitrogen). For *in situ* hybridization experiments we used sheep anti-Digoxigenin-AP Fab fragments (#11093274910, Roche).

Western blot

Cell extracts from 90% confluent cells were obtained with RIPA buffer (25 mM Tris-HCl, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, pH 7.6) containing 1x complete protease inhibitor (#4693132001, Roche). After homogenization and centrifugation, supernatants were kept at -20°C until further analysis. 20 to 40 µg of protein extract were denatured with 4x Laemmli buffer and

loaded onto 5-10% SDS polyacrylamide gels. After running the gels, the proteins were transferred to PVDF membranes (#T830.1, Roth, Karlsruhe, Germany) in semidry blot transfer chambers (#1703848, Bio-Rad, Hercules, California, USA). The membranes were then blocked with 5% non-fat dry milk (#A0830, Applichem, Darmstadt, Germany) in PBS-T (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 0.1% Tween-20) for 1 h and then incubated an additional hour with manufacturers' recommended antibody dilutions in blocking solution. After 3 PBS-T washes for 15 min, the membranes were incubated with secondary antibodies coupled to HRP in blocking solution for 1 h. After 3 additional PBS-T washes, the membranes were developed with the CheLuminate-HRP FemtoDetect kit (#A7807, Applichem) and a LAS-4000 photodocumentation system (Fujifilm) with increment exposure and super resolution detection modes unless stated otherwise.

Immunofluorescent stainings

Immunofluorescent staining was performed on 50-60% confluent cells in 24-well plates. Cells were washed once with PBS and then fixed in either 4% PFA for 10 min or cold methanol for 15 min at 4°C and later permeabilized with 0.5% Triton X-100 for 10 min. Epitopes were blocked with TNB blocking solution (#FP1012, Perkin-Elmer, Waltham, Massachusetts, USA) for 1 h and later stained with the antibody dilutions recommended by the manufacturers in TNB blocking solution for an hour or overnight at 4°C. After washing 3 times with PBS-T for 15 min they were incubated with secondary antibody dilutions in TNB blocking solution for an hour. After counterstaining with DAPI (4',6-Diamidino-2-phenylindole dihydrochloride, #6843.2, Roth) and 3 last washing steps, the stainings were documented in a DM6000 inverted microscope (Leica, Wetzlar, Germany).

Nuclear recruitment assays

To check for interactions inside the nucleus, we cotransfected 293 cells with plasmids encoding the tagged identified candidate proteins with either TBX18, or TBX18 with a deleted nuclear localization signal (TBX18 Δ NLS) as previously reported [20]. One day after transfection, cells were fixed with PFA and immunofluorescent stainings against the different protein tags were carried out and documented with a DM6000 microscope.

GST pull-down assays

To check for physical interaction, we performed Glutathione S-Transferase (GST) fusion protein pull-downs as described previously [20]. GST-fusion proteins of the different TBX18 domains were produced in the BL21 *E. coli* strain after IPTG induction. The GST fusion proteins were batch purified from the bacterial lysates using glutathione sepharose 4B beads (#17075601, GE, Chicago, Illinois, USA). To produce radioactively tagged candidate proteins we used a reticulocyte lysate-based TNT Quick Coupled Transcription/Translation system (#L2080, #L1170, Promega) with ³⁵S-labeled Methionine/Cysteine 75/25% (#SRIS-103, Hartmann Analytic) on plasmids with a suitable promoter (see Supplementary Table S1 for details on plasmids used). To perform the pull-down assays, the bead solutions containing the GST-TBX18 fusion proteins were incubated with the *in vitro* produced radioactively-labelled candidate proteins and washed several times with HEPES buffer (20 mM HEPES, 10 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA, 5% Glycerol, 0.5% Triton X-100, 80 mM NaCl, pH 7.5). Analysis of the pulled-down proteins was achieved by separating the proteins by denaturing SDS-PAGE, heat-drying the resulting gels on

filter paper and then exposing them to phosphorescent imaging plates for 1 day. Detection of the proteins was done on a FLA-7000 Laser photodocumentation system (Fujifilm) with a sensitivity of S10000 and a pixel size of 50 μm .

Immunoprecipitation assays

To perform immunoprecipitations of endogenous proteins, we used whole cell extracts from 293 and 293-TBX18.3xFLAG knock-in cells. Cells were cultured to 90% confluency on 10 cm dishes and then washed with PBS before harvesting them twice with PBS and a plastic scraper. The cells were lysed with 1 ml RIPA buffer, and the lysate homogenized and centrifuged. 100 or 200 μl of the supernatants were used to detect the interaction in either direction either with 10 μl of direct anti-FLAG M2 affinity gel batch immunoprecipitation (#A2220, Sigma Aldrich) or immunoprecipitation with 2 μl of the primary antibodies against ZMYM2 together with 20 μl of protein A-agarose beads (#sc-2001, Santa Cruz, Santa Cruz, California, USA) according to the manufacturer's protocol. The FLAG beads were washed four times with TBS-T buffer, while the protein A beads with or without antibody were washed thrice with HEPES Buffer 2 (25 mM HEPES, 1 mM MgCl_2 , 10% Glycerol, 0.1% NP40, 75 mM NaCl, 1 mM DTT, 10 $\mu\text{g/ml}$ BSA, pH 7.4) before boiling the beads with Laemmli buffer to release the pulled proteins. Western Blot analysis was used to detect the presence of the proteins in either direction of interaction.

Transactivation assays

Transactivation assays were done on lysates of 293 cells with the Dual-Luciferase Reporter Assay System (#E1910, Promega) as previously reported [27]. We seeded 50-60% confluent 293 cells on 6-well plates and transfected a total of 2.5 mg of the

plasmid combinations, including *pCMV-EGFP* and *pRL-tk* as transfection controls, into the cells with the calcium phosphate method. One day after transfection, we lysed the cells and followed the manufacturer's protocol to automatically measure the firefly and renilla luciferase activities in duplicates with a Glomax luminometer (#9100-002, Turner Biosystems, Sunnyvale, California, USA). Relative luciferase units were calculated by dividing the firefly to renilla luciferase activity ratio from each well and later normalizing this ratio to the firefly luciferase reporter transfection (*pGL3.Prom.Tbx18BS2*) [27] control. Two technical replicates each from two biological replicates were used for statistical analysis. A two-tailed Student's t-test was carried between all normalized measurements between pairs of samples adjusting for equal or unequal variance according to an F-test. Significance was defined as $P \leq 0.05$ (*) significant, $P \leq 0.01$ (**) very significant or $P \leq 0.001$ (***) extremely significant.

RNA *in situ* hybridization

10- μ m paraffin sections of the posterior trunk region of embryonic day (E) 12.5 wildtype NMRI mice containing the proximal ureter region were subjected to RNA hybridization with digoxigenin-labelled antisense riboprobes as previously described [37].

Results

CRISPR/Cas9 technology confers a 3xFLAG tag to endogenous TBX18 in 293 and A549 cell lines

Identification of TBX18 interaction partners in embryonic contexts is hampered by the scarcity of the material, whereas overexpression approaches *in vitro* may be compromised by unnaturally high protein levels that force unnatural interactions. We reasoned that cell lines with endogenous *TBX18* expression provide a useful alternative for identification of relevant binding partners of TBX18. We recently reported that 293 embryonic kidney cells show increased expression of *TBX18* in a set of 6 human cell lines tested [27]. An expanded search identified elevated levels of *TBX18* in A549 lung carcinomas and HPMEC pulmonary microcapillary endothelial cells as well (Supplementary Fig. S1).

Since available antibodies proved unsuitable for precipitation of endogenous protein complexes of TBX18 in these cell lines in our hands, we decided to exploit the CRISPR/Cas9 technology to introduce a triple FLAG (3xFLAG) affinity tag fragment [38, 39] into the *TBX18* locus of both 293 and A549 cells [29, 40]. Human *TBX18* is localized on the long arm of chromosome 6 (6q14.3) and comprises 8 exons. To correctly integrate the 3xFLAG at the C-terminus of the encoded TBX18 protein, we designed guide RNA plasmids to direct the Cas9 nuclease to a protospacer adjacent motif (PAM) close to the *TBX18* stop codon, and homology directed recombination (donor) plasmids containing a modified region of the 8th exon to erase the stop codon, followed by a 3xFLAG motif, a P2A cleavage sequence and a *neomycin* cassette for antibiotic resistance, flanked by 5'- and 3'-regions homologous to the *TBX18* locus (Fig. 1A, for details of these and all other constructs see Supplementary Table S1).

Cells transfected with these two constructs were grown under neomycin selection using G418, and individual surviving clones expanded. We validated the targeting event by RFLP analysis on genomic DNA extracted from the clones. For A549 cells, 4 out of 39 tested clones carried a heterozygous knock-in, 1 was homozygous; all 5 clones were devoid of additional integrations of the *neo* cassette (Supplementary Fig. S2). For 293 cells, we identified 5 clones with correct heterozygous targeting of *TBX18* and lack of additional off-target integrations of the selection cassette (n=45) (Supplementary Fig. S3). We choose one 293 heterozygous clone (293-P2C2), one heterozygous (A549-12C3) and one homozygous A549 clone (A549-12C4) for further characterization. Western blot analysis using an anti-FLAG antibody detected a band at approximately 90 kDa slightly bigger than the calculated theoretical mass of 68.7 kDa in the targeted but not in the parental cell lines (Fig. 1B, Supplementary Fig. S4). Immunofluorescent stainings revealed a speckle-like pattern of the FLAG tag in the nucleus of the TBX18.3xFLAG cell lines but not in the parental cell lines (Fig. 1C) in agreement with an exclusive nuclear localization of TBX18 in 293 and A549 cells.

LC-MS identifies ZMYM2 as a novel TBX18 interacting protein in 293 and A549 cells

To identify TBX18 interacting proteins in 293 and A549 cells, we used our newly generated TBX18.3xFLAG cell lines (heterozygous 293-P2C2, homozygous A549-12C4) for affinity chromatography of native complexes. After cell expansion and nuclear extraction, protein complexes harbouring TBX18.3xFLAG were purified with an anti-FLAG batch affinity strategy [41, 42]. The purified protein complexes were recovered by heat treatment and resolved by SDS-PAGE. The polyacrylamide gels of three independent experiments and controls for each cell line were sent to the

Hannover Medical School Proteomics Facility for protein extraction and subsequent LC-MS analysis. A total of 115 human proteins were enriched in anti-FLAG immunoprecipitates from 293-P2C2 compared to parental 293 cells, from which 10 passed a stringent t-test statistical analysis of the intensity scores of the three samples of the 293-TBX18.3xFLAG cells compared to the three samples of the 293 parental cells. Four of the 10 candidates were Zn-finger proteins (PATZ2, ZNF143, ZMYM2, ZMYM3) (Fig. 2, Supplementary Table S2). In anti-FLAG immunoprecipitates from A549-12C4 cells, 270 human proteins were more strongly recovered compared to the parental A549 cells. Here, 36 proteins were statistically significantly enriched in the three IP samples of the A549-TBX18.3xFLAG cells compared to A549 control cells (Fig. 2, Supplementary Table S3). These included the transcription factor RFX1, nuclear corepressors like NCOR2 and EMSY, a cluster of proteins from the 60S ribosomal subunit (RPL4, RPL14, RPL15, RPL18) and the zinc finger transcription factors ZNF609, ZMYM2 and ZMYM3. Importantly, TBX18 was the most enriched protein in immunoprecipitates from both cell lines confirming correct FLAG-tagging of the protein by our approach.

To recover cell context-independent binding partners of TBX18, we overlapped the lists of statistically significantly enriched candidates obtained from 293-P2C2 and A549-12C4 cell precipitates and identified besides TBX18 a single protein, Zinc finger, MYM-type 2 (ZMYM2), a DNA-binding transcription factor known to be part of the CoREST/BHC repressive complex [43-46]. Importantly, ZMYM2 was only found once as background in 136 datasets of FLAG precipitations in 293 cells as listed in the crapome database for common affinity purification contaminants [47] making it very unlikely that ZMYM2 is merely a contaminant in our IP experiment.

ZMYM2 interacts with endogenous TBX18 in the nucleus of 293 cells

To validate the finding of our proteomic approach, we independently assayed for endogenous interaction of ZMYM2 and TBX18 in 293 cells. Using a commercial antibody against ZMYM2 we detected a band of the expected size (154.9 kDa) in lysates of both 293 parental and TBX18.3xFLAG knock-in cells (Fig. 3A). In immunofluorescent analysis, ZMYM2 antibodies detected nuclear antigens in 293-TBX18.3xFLAG cells (Fig. 3B). In co-immunoprecipitation experiments in 293-TBX18.3xFLAG cells, we detected the presence of ZMYM2 protein after immunoprecipitation with the anti-FLAG antibody. In the reverse reaction, we did not detect FLAG-tagged TBX18 in precipitates with the ZMYM2 antibody possibly due to the low amounts of proteins present in these cells (Fig. 3C). Thus, we identified ZMYM2 as a nuclear protein that interacts with TBX18 in 293-TBX18.3xFLAG cells.

Overexpressed ZMYM2 binds to TBX18 in the nucleus of 293 cells

To further validate ZMYM2-TBX18 interaction, we analyzed whether TBX18 and ZMYM2 co-localize in the nucleus when overexpressed in 293 cells. For this, we transfected 293 cells with an expression construct for full-length MYC-ZMYM2, confirmed protein integrity by Western blot (Fig. 4A), and characterized the localization of the protein by immunofluorescence analysis against the MYC-tag. Similar to TBX18, ZMYM2 localized to the nucleus, accumulating in speckle-like structures (Fig. 4B).

To test, whether overexpressed TBX18 productively interacts with ZMYM2 in 293 cells, we made use of a nuclear recruitment assay that we previously described [20]. It is based on the identification of a classical nuclear localization signal (NLS) at the N-terminus of the TBX18 protein. When this NLS is deleted, the resulting protein

(TBX18 Δ NLS) is excluded from the nucleus but can be shuttled back to this compartment by binding to a protein, which carries such a signal. Upon coexpression of TBX18 Δ NLS and ZMYM2, we observed a nuclear translocation of TBX18 Δ NLS (Fig. 4C). Together, this shows that TBX18 interacts with ZMYM2 in the nucleus when overexpressed in 293 cells.

ZMYM2 interacts with the T-box region of TBX18 in a physically direct manner

Interaction between TBX18 and ZMYM2 can occur directly or is mediated by auxiliary proteins. To more stringently test for direct physical interaction, we performed pull-down assays using bacterially expressed GST-TBX18 fragment proteins and radioactively labelled ZMYM2 produced with an *in vitro* expression system. Since GST-TBX18 full-length protein could not be expressed in bacteria, we used a previously described series of bacterially expressed fusion proteins of GST with the N- and C-terminal region and the T-box of TBX18 which simultaneously allows to delimit the interaction domain with TBX18 [20]. ZMYM2 interacted with the T-box region and weakly with the N-terminal plus T-box region of TBX18 (Fig. 5).

ZMYM2 does not modulate TBX18 transcriptional repression

To analyse whether ZMYM2 modulates TBX18 transcriptional activity, we performed dual luciferase reporter assays in 293 cells. We used a recently described reporter plasmid in which a palindromic repeat of two TBEs is cloned in front of an SV40 minimal promoter and a firefly luciferase gene, expression constructs of TBX18 and candidate proteins in various combinations, and a Renilla luciferase control plasmid [27] (Fig. 6A, Supplementary Table S4). Under the chosen conditions, TBX18 reduced the TBX18 reporter activity by approximately a third in comparison to an

empty expression vector. TLE3/GROUCHO further increased this repression as previously shown [20]. Cotransfection of an expression plasmid for ZMYM2 had no significant effect on TBX18 repression alone in line with the lack of a transcription modulation domain in this protein [48, 49]. However, ZMYM2 was shown to be a member of the CoREST corepressor complex, that harbors additional proteins such as KDM1A, that was also recovered in our proteomic screens [43-46]. In cotransfection/coactivation experiments KDM1A repressed in a concentration-dependent manner transcription together with TBX18. Notably, the addition of ZMYM2 to KDM1A relieved the corepression exerted by KDM1A.

***Zmym2* is coexpressed with *Tbx18* in the ureteric mesenchyme at E12.5**

To unveil whether ZMYM2 could potentially mediate TBX18 function *in vivo*, we investigated the expression pattern of the mouse *Zmym2* orthologue in the developing ureter at E12.5, an embryonic stage when *Tbx18* has a critical function in the mesenchymal compartment of this organ [10, 11]. RNA *in situ* hybridization detected *Zmym2* in the epithelium and overlapping with *Tbx18* in the mesenchyme of the developing ureter indicating a possible *in vivo* relevance of TBX18-ZMYM2 interaction (Fig. 6B).

Discussion

Here, we used affinity purification and mass spectrometry to characterize the interactome of a tagged version of endogenous TBX18 in 293 and A549 cells. We found ZMYM2 as the only statistically enriched protein in anti-FLAG immunoprecipitates from both 293 and A549 cells with endogenously tagged TBX18, and confirmed direct binding of ZMYM2 to the T-box of TBX18. Our work implicates the CoREST repressor complex in TBX18 function, and suggests that TBX18 and ZMYM2 interact in the ureteric mesenchyme *in vivo*.

Immunoprecipitations of an endogenous FLAG tagged TBX18 protein identifies only one protein significantly enriched in both A549 and 293 cells

Our affinity mediated immunoprecipitation approach against an engineered FLAG tag at the C-terminus of TBX18 protein identified both in 293 and A549 cells a large number of proteins coprecipitating with TBX18. These extensive lists (115 and 270 proteins, respectively) comprised proteins localized outside the nucleus (e.g. cytoskeletal proteins) indicating that the nuclear extracts used for the immunoprecipitations were not purified to homogeneity. Furthermore, we detected many proteins (e.g. the ribonucleoproteins RALY and HNRNPN) that have been frequently detected in similar anti-FLAG immunoprecipitation experiments as documented in the crapome database [47]. They may represent proteins that are unspecifically recognized by the anti-FLAG antibody, that adhere strongly to the binding matrix or that may preferentially denature to form adhesive aggregates. In any case, our findings enforce the notion that an affinity purification against a single tag provides more false positive candidates compared to a tandem purification strategy available for dual tagged proteins.

Given these considerations, we decided to increase the stringency of our approach by selecting proteins that were statistically significantly enriched in all six individual immunoprecipitations both in 293- and A549-TBX18.3xFLAG cells. Such a strategy is not without caveats since it ignores the fact that 293 and A549 cells derive from two different organs, thus, express a large number of different cell-type specific proteins. However, and surprisingly, we did not find proteins that bear a “renal” (293 cells) or a “pulmonary” signature (A549 cells) in the individual candidate lists of either cell line, making it less likely that we missed out on cell-type specific interesting candidates in the final list of proteins. Unexpectedly, this approach identified only one protein, ZMYM2, as binding partner of TBX18 in both cell lines. Importantly, ZMYM2 is a nuclear protein related to gene regulation and transcriptional repression function as is TBX18, arguing that we indeed increased the specificity within the overlap.

We recently used overexpressed dual tagged TBX18 as bait to identify by tandem affinity purification and subsequent LC-MS analysis TBX18 interacting proteins in 293 cells. We identified several transcriptional cofactors and a large number of homeobox transcription factors as preferred binding partners [27]. To our surprise, the interactome revealed for endogenous TBX18 in our current study was composed differently. Most notably, we found only few tissue-specific transcription factors which belonged to the Zn-finger family and not to the homeobox protein family. It seems possible that in an overexpression approach the endogenous binding partners of TBX18 are limited and that weak or even unnatural interactions are overrepresented or newly established. Such a behavior is eased by the fact that the T-box seems to act rather promiscuously by binding to many very different types of proteins or motifs. Alternatively, TBX18 expression in A549 and 293 cells may be extremely low to interact with very few other proteins. Irrespective of the precise mechanism, it shows

that the expression level (e.g. low endogenous vs highly overexpressed) of a protein and the mode of purification (single tag vs dual tag) can dramatically affect the identification of relevant binding partners in cellular systems.

ZMYM2 may present a biologically meaningful protein interaction partner of TBX18

ZMYM2 is a Zn-finger protein that binds to chromatin and recruits a repressive complex consisting of the histone demethylase KDM1/LSD1, the cofactor CoREST (RCOR1 or RCOR2), and a histone deacetylase HDAC1 [46, 50]. ZMYM2 was the protein with the highest significance score found in both 293 and A549 precipitates. It interacted with endogenous and overexpressed TBX18 in 293 cells, and bound to the T-box of TBX18 *in vitro*. KDM1 was not found in the nucleus and did not physically interact with TBX18 (data not shown). However, KDM1A increased the repression by TBX18. ZMYM2 was not only bound to TBX18 directly inside the nucleus but it seemed to restore the level of repression from the level exerted by KDM1A and TBX18 together in a luciferase reporter assay. This effect may be explained by competition with TBX18 since ZMYM2 is able to compete with the transcription factor REST for binding to RCOR1 [46]. It is possible that large amounts of ZMYM2 impair the functional interaction of TBX18 with the CoREST complex.

We recently showed that RCOR3, an occasional member of the CoREST complex [44] is able to interact with TBX18 in a nuclear recruitment assay, but this interaction was indirect and only altered transcription repression by TBX18 in a concentration-independent manner [27]. The protein ZMYM3 was also statistically enriched in the 293 cells, and KDM1A was increased, though not statistically significantly. Both of these proteins have been described as members of the CoREST complex as well.

Hence, enrichment of 4 proteins involved in the CoREST complex in TBX18 precipitates (RCOR3, KDM1A, ZMYM2 and ZMYM3) indicates the potential significance in mediating transcriptional repression by TBX18.

A recent genomic association study identified *ZMYM2* as a gene underlying CAKUT in human patients. Importantly, the phenotypic changes overlapped with that recently reported for patients with inactivating mutations in *TBX18* (Friedhelm Hildebrandt, Harvard Medical School, personal communication). Moreover, our expression analysis showed that *Tbx18* and *Zmym2* are coexpressed in the ureteric mesenchyme of mouse embryos at E12.5, the stage at which *Tbx18* is required for specification of this tissue. It is tempting to speculate that ZMYM2 and TBX18 physically interact in the ureteric mesenchyme and corepress a set of genes mediating this function. Further genetic interaction and molecular studies *in vivo* will address this hypothesis in the future.

References

1. Sebe-Pedros A, Ruiz-Trillo I. Evolution and Classification of the T-Box Transcription Factor Family. *Curr Top Dev Biol.* 2017;122:1-26.
2. Kispert A, Herrmann BG. The Brachyury gene encodes a novel DNA binding protein. *EMBO J.* 1993;12(8):3211-20.
3. Kispert A, Koschorz B, Herrmann BG. The T protein encoded by Brachyury is a tissue-specific transcription factor. *EMBO J.* 1995;14(19):4763-72.
4. Papaioannou VE. The T-box gene family: emerging roles in development, stem cells and cancer. *Development.* 2014;141(20):3819-33..
5. Waldron L, Steimle JD, Greco TM, Gomez NC, Dorr KM, Kweon J, et al. The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. *Dev Cell.* 2016;36(3):262-75. d
6. Kaltenbrun E, Greco TM, Slagle CE, Kennedy LM, Li T, Cristea IM, et al. A Gro/TLE-NuRD corepressor complex facilitates Tbx20-dependent transcriptional repression. *J Proteome Res.* 2013;12(12):5395-409.
7. Naiche LA, Harrelson Z, Kelly RG, Papaioannou VE. T-box genes in vertebrate development. *Annu Rev Genet.* 2005;39:219-39.
8. Ghosh TK, Brook JD, Wilsdon A. T-Box Genes in Human Development and Disease. *Curr Top Dev Biol.* 2017;122:383-415.
9. Vivante A, Kleppa MJ, Schulz J, Kohl S, Sharma A, Chen J, et al. Mutations in TBX18 Cause Dominant Urinary Tract Malformations via Transcriptional Dysregulation of Ureter Development. *Am J Hum Genet.* 2015;97(2):291-301.
10. Airik R, Bussen M, Singh MK, Petry M, Kispert A. Tbx18 regulates the development of the ureteral mesenchyme. *J Clin Invest.* 2006;116(3):663-74.
11. Bohnenpoll T, Bettenhausen E, Weiss AC, Foik AB, Trowe MO, Blank P, et al. Tbx18 expression demarcates multipotent precursor populations in the developing urogenital system but is exclusively required within the ureteric mesenchymal lineage to suppress a renal stromal fate. *Dev Biol.* 2013;380(1):25-36.
12. Trowe MO, Maier H, Schweizer M, Kispert A. Deafness in mice lacking the T-box transcription factor Tbx18 in otic fibrocytes. *Development.* 2008;135(9):1725-34. doi: 10.1242/dev.014043. PubMed PMID: 18353863.
13. Bolt CC, Negi S, Guimaraes-Camboa N, Zhang H, Troy JM, Lu X, et al. Tbx18 Regulates the Differentiation of Periductal Smooth Muscle Stroma and the Maintenance of Epithelial Integrity in the Prostate. *PLoS One.* 2016;11(4):e0154413.
14. Norden J, Grieskamp T, Christoffels VM, Moorman AF, Kispert A. Partial absence of pleuropericardial membranes in Tbx18- and Wt1-deficient mice. *PLoS One.* 2012;7(9):e45100.
15. Greulich F, Farin HF, Schuster-Gossler K, Kispert A. Tbx18 function in epicardial development. *Cardiovasc Res.* 2012;96(3):476-83.
16. Wu SP, Dong XR, Regan JN, Su C, Majesky MW. Tbx18 regulates development of the epicardium and coronary vessels. *Dev Biol.* 2013;383(2):307-20.
17. Christoffels VM, Mommersteeg MT, Trowe MO, Prall OW, de Gier-de Vries C, Soufan AT, et al. Formation of the venous pole of the heart from an Nkx2-5-negative precursor population requires Tbx18. *Circ Res.* 2006;98(12):1555-63.
18. Wiese C, Grieskamp T, Airik R, Mommersteeg MT, Gardiwal A, de Gier-de Vries C, et al. Formation of the sinus node head and differentiation of sinus node myocardium are independently regulated by Tbx18 and Tbx3. *Circ Res.* 2009;104(3):388-97.
19. Bussen M, Petry M, Schuster-Gossler K, Leitges M, Gossler A, Kispert A. The T-box transcription factor Tbx18 maintains the separation of anterior and posterior somite compartments. *Genes Dev.* 2004;18(10):1209-21.
20. Farin HF, Bussen M, Schmidt MK, Singh MK, Schuster-Gossler K, Kispert A. Transcriptional repression by the T-box proteins Tbx18 and Tbx15 depends on Groucho corepressors. *J Biol Chem.* 2007;282(35):25748-59.

21. Chen G, Fernandez J, Mische S, Courey AJ. A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in *Drosophila* development. *Genes Dev.* 1999;13(17):2218-30.
22. Kaul A, Schuster E, Jennings BH. The Groucho co-repressor is primarily recruited to local target sites in active chromatin to attenuate transcription. *PLoS Genet.* 2014;10(8):e1004595.
23. Kaul AK, Schuster EF, Jennings BH. Recent insights into Groucho co-repressor recruitment and function. *Transcription.* 2015;6(1):7-11..
24. Sekiya T, Zaret KS. Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo. *Mol Cell.* 2007;28(2):291-303.
25. Farin HF, Mansouri A, Petry M, Kispert A. T-box protein Tbx18 interacts with the paired box protein Pax3 in the development of the paraxial mesoderm. *J Biol Chem.* 2008;283(37):25372-80. d
26. Nie X, Sun J, Gordon RE, Cai CL, Xu PX. SIX1 acts synergistically with TBX18 in mediating ureteral smooth muscle formation. *Development.* 2010;137(5):755-65.
27. Rivera-Reyes R, Kleppa MJ, Kispert A. Proteomic analysis identifies transcriptional cofactors and homeobox transcription factors as TBX18 binding proteins. *PLoS One.* 2018;13(8):e0200964.
28. Pear WS, Nolan GP, Scott ML, Baltimore D. Production of high-titer helper-free retroviruses by transient transfection. *Proc Natl Acad Sci U S A.* 1993;90(18):8392-6.
29. Savic D, Partridge EC, Newberry KM, Smith SB, Meadows SK, Roberts BS, et al. CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Res.* 2015;25(10):1581-9.
30. Chu VT, Graf R, Wirtz T, Weber T, Favret J, Li X, et al. Efficient CRISPR-mediated mutagenesis in primary immune cells using CrispRGold and a C57BL/6 Cas9 transgenic mouse line. *Proc Natl Acad Sci U S A.* 2016;113(44):12514-9.
31. Bauer DE, Canver MC, Orkin SH. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J Vis Exp.* 2015;(95):e52118.
32. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*: Cold spring harbor laboratory press; 1989.
33. Wang Z. Epitope tagging of endogenous proteins for genome-wide chromatin immunoprecipitation analysis. *Methods Mol Biol.* 2009;567:87-98.
34. Jochim N, Gerhard R, Just I, Pich A. Impact of clostridial glucosylating toxins on the proteome of colonic cells determined by isotope-coded protein labeling and LC-MALDI. *Proteome Sci.* 2011;9:48.
35. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* 2008;26(12):1367-72.
36. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44-57.
37. Moorman AF, Houweling AC, de Boer PA, Christoffels VM. Sensitive nonradioactive detection of mRNA in tissue sections: novel application of the whole-mount in situ hybridization protocol. *J Histochem Cytochem.* 2001;49(1):1-8.
38. Knappik A, Pluckthun A. An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments. *Biotechniques.* 1994;17(4):754-61.
39. Tsai A, Carstens RP. An optimized protocol for protein purification in cultured mammalian cells using a tandem affinity purification approach. *Nat Protoc.* 2006;1(6):2820-7.
40. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science.* 2013;339(6121):819-23.
41. Wang Y, Zhang H, Chen Y, Sun Y, Yang F, Yu W, et al. LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell.* 2009;138(4):660-72.

42. O'Brien LL, Guo Q, Bahrami-Samani E, Park JS, Hasso SM, Lee YJ, et al. Transcriptional regulatory control of mammalian nephron progenitors revealed by multi-factor cistromic analysis and genetic studies. *PLoS Genet.* 2018;14(1):e1007181.
43. Hakimi MA, Dong Y, Lane WS, Speicher DW, Shiekhatair R. A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. *J Biol Chem.* 2003;278(9):7234-9.
44. Bantscheff M, Hopf C, Savitski MM, Dittmann A, Grandi P, Michon AM, et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotechnol.* 2011;29(3):255-65.
45. Laurent B, Ruitu L, Murn J, Hempel K, Ferrao R, Xiang Y, et al. A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. *Mol Cell.* 2015;57(6):957-70.
46. Gocke CB, Yu H. ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS One.* 2008;3(9):e3255.
47. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods.* 2013;10(8):730-6.
48. Popovici C, Adelaide J, Ollendorff V, Chaffanet M, Guasch G, Jacrot M, et al. Fibroblast growth factor receptor 1 is fused to FIM in stem-cell myeloproliferative disorder with t(8;13). *Proc Natl Acad Sci U S A.* 1998;95(10):5712-7.
49. Reiter A, Sohal J, Kulkarni S, Chase A, Macdonald DH, Aguiar RC, et al. Consistent fusion of ZNF198 to the fibroblast growth factor receptor-1 in the t(8;13)(p11;q12) myeloproliferative syndrome. *Blood.* 1998;92(5):1735-42.
50. Maiques-Diaz A, Somerville TC. LSD1: biologic roles and therapeutic targeting. *Epigenomics.* 2016;8(8):1103-16.
51. Grossman CE, Qian Y, Banki K, Perl A. ZNF143 mediates basal and tissue-specific expression of human transaldolase. *J Biol Chem.* 2004;279(13):12190-205.
52. Thompson JJ, Kaur R, Sosa CP, Lee JH, Kashiwagi K, Zhou D, et al. ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. *Nucleic Acids Res.* 2018;46(19):10034-51.
53. Peng L, Chen G, Zhu Z, Shen Z, Du C, Zang R, et al. Circular RNA ZNF609 functions as a competitive endogenous RNA to regulate AKT3 expression by sponging miR-150-5p in Hirschsprung's disease. *Oncotarget.* 2017;8(1):808-18.
54. Armache JP, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, et al. Localization of eukaryote-specific ribosomal proteins in a 5.5-A cryo-EM map of the 80S eukaryotic ribosome. *Proc Natl Acad Sci U S A.* 2010;107(46):19754-9.
55. Cheng X, Kao HY. G protein pathway suppressor 2 (GPS2) is a transcriptional corepressor important for estrogen receptor alpha-mediated transcriptional regulation. *J Biol Chem.* 2009;284(52):36395-404.
56. Xu Y, Sengupta PK, Seto E, Smith BD. Regulatory factor for X-box family proteins differentially interact with histone deacetylases to repress collagen alpha2(I) gene (COL1A2) expression. *J Biol Chem.* 2006;281(14):9260-70.

Acknowledgements

We thank Christopher Binz for technical assistance, and Prof. Yang Shi (Harvard Medical School) for plasmids. This work was supported by funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Cluster of Excellence REBIRTH (From Regenerative Biology to Reconstructive Therapy) at Hannover Medical School, and by the individual grant DFG KI728/7-2 (A.K.).

Author Contributions

Study concept and design, R.R.-R., M-J.K and A.K.. Acquisitions of data, R.R.-R., M-J.K., T.H.L.. Drafting of the manuscript R.R.-R. and A.K.. Critical revision of manuscript, all authors. Statistical analysis, R.R.-R., M-J.K.

Data Availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

Additional Information

Supplementary information accompanies this paper at...

Competing Interests: The authors declare that they have no competing interests.

Figures

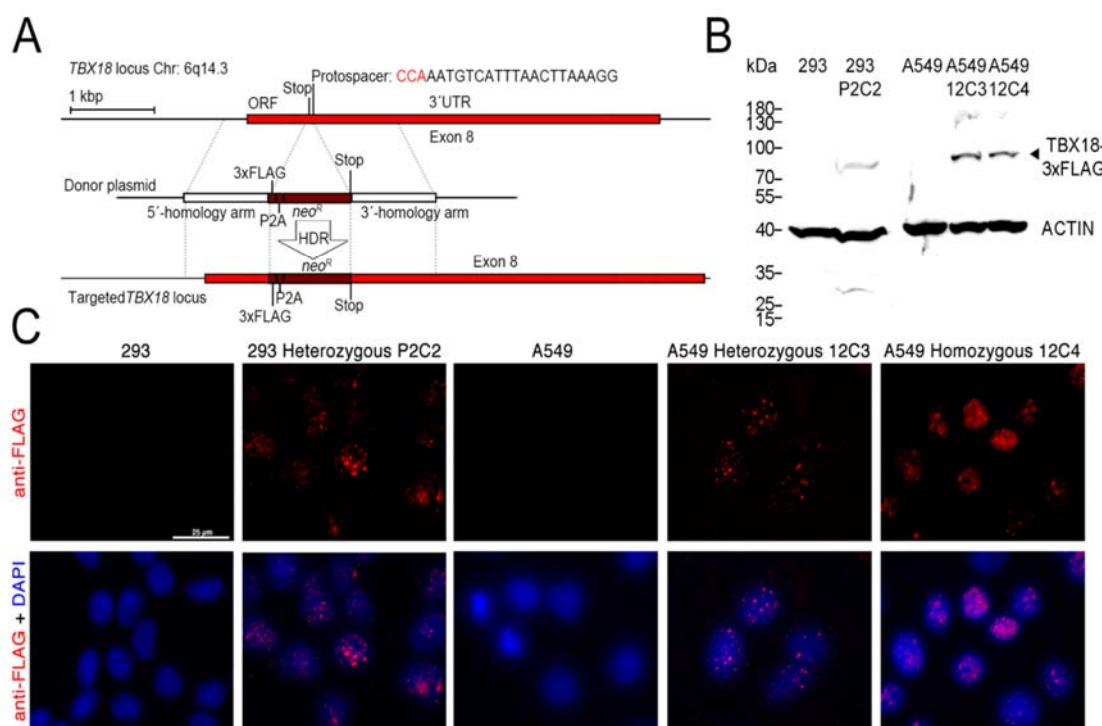


Figure 1. CRISPR/Cas9-mediated FLAG-tagging allows detection and localization of endogenous TBX18 protein in 293 and A549 cells. **(A)** Scheme of the CRISPR/Cas9 and homology mediated repair strategy to insert a triple FLAG peptide at the C-terminal end of TBX18. Red boxes indicate exon 8 of *TBX18* on human chromosome 6 in which the stop codon is located. Thin dashed lines mark the homology region between the donor plasmid (white boxes) and the *TBX18* locus before and after targeting to insert a cassette containing a triple FLAG tag, a P2A protein cleavage site, the *neomycin* resistance gene and the stop codon. **(B)** Western blot analysis of 293 parental and of 293-TBX18.3xFLAG heterozygous (293-P2C2) cells, and of A549 parental, A549-TBX18.3xFLAG heterozygous (A549-12C3) and homozygous (A549-12C4) cells with anti-FLAG antibodies detects a band slightly heavier (80-90kDa) than the calculated 68.7 kDa in the knock-in cells only (white arrowhead). ACTIN immunodetection serves as a loading control (black arrowhead). **(C)** Immunofluorescent staining of A549 and 293 parental and TBX18.3xFLAG knock-in cells with the anti-FLAG antibody detects a speckle-like pattern in the nucleus of TBX18.3xFLAG knock-in cells. Counterstaining of the nucleus was made with DAPI, 4,6-diamidino-2-phenylindole nuclear counterstain.

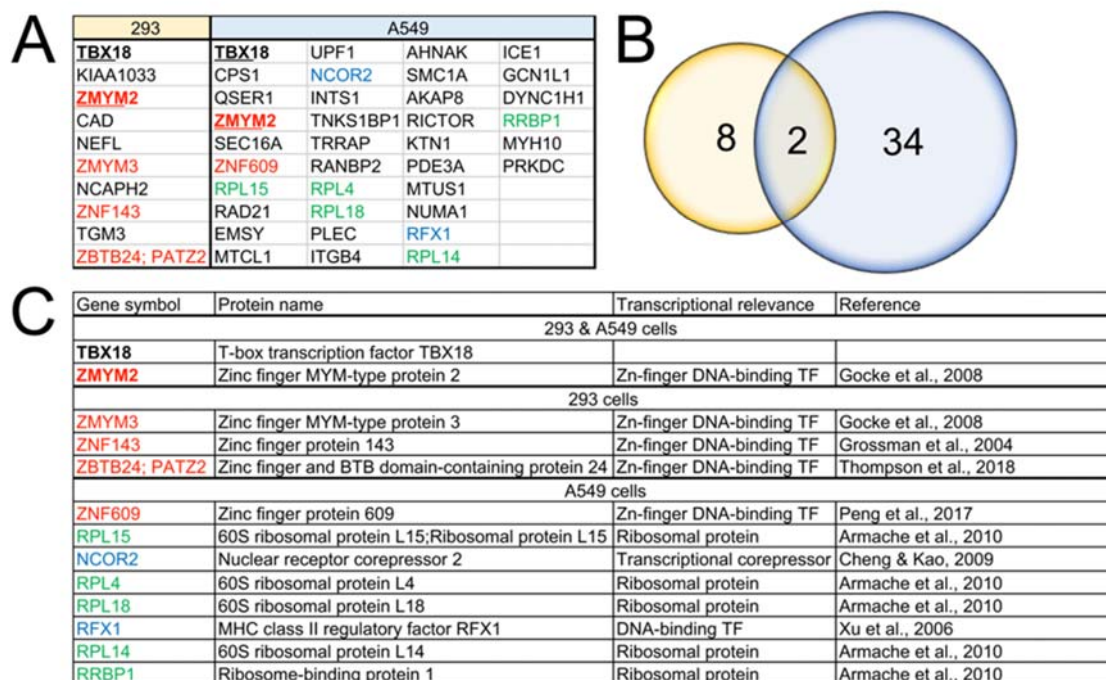


Figure 2. LC-MS identifies novel protein binding partners of endogenous TBX18 protein in 293 and A549 cells. **(A)** Two lists showing 36 peptides statistically enriched in FLAG precipitates from A549-TBX18.3xFLAG cells (A549-12C4) and 10 proteins from 293-TBX18.3xFLAG cells (293-P2C2) compared to the unmodified parental lines. **(B)** Venn diagram depicting the overlap between the candidate protein lists for 293 (yellow) and A549 (blue) cells. **(C)** Manual literature mining identified proteins with transcriptional relevance; Zn finger proteins in red, ribosomal proteins in green, corepressors in blue. References: Gocke et al., 2008 [46]; Grossman et al., 2004 [51]; Thompson et al., 2018 [52]; Peng et al., 2017 [53]; Armache et al., 2010 [54]; Cheng and Kao, 2009 [55]; Xu et al., 2006 [56].

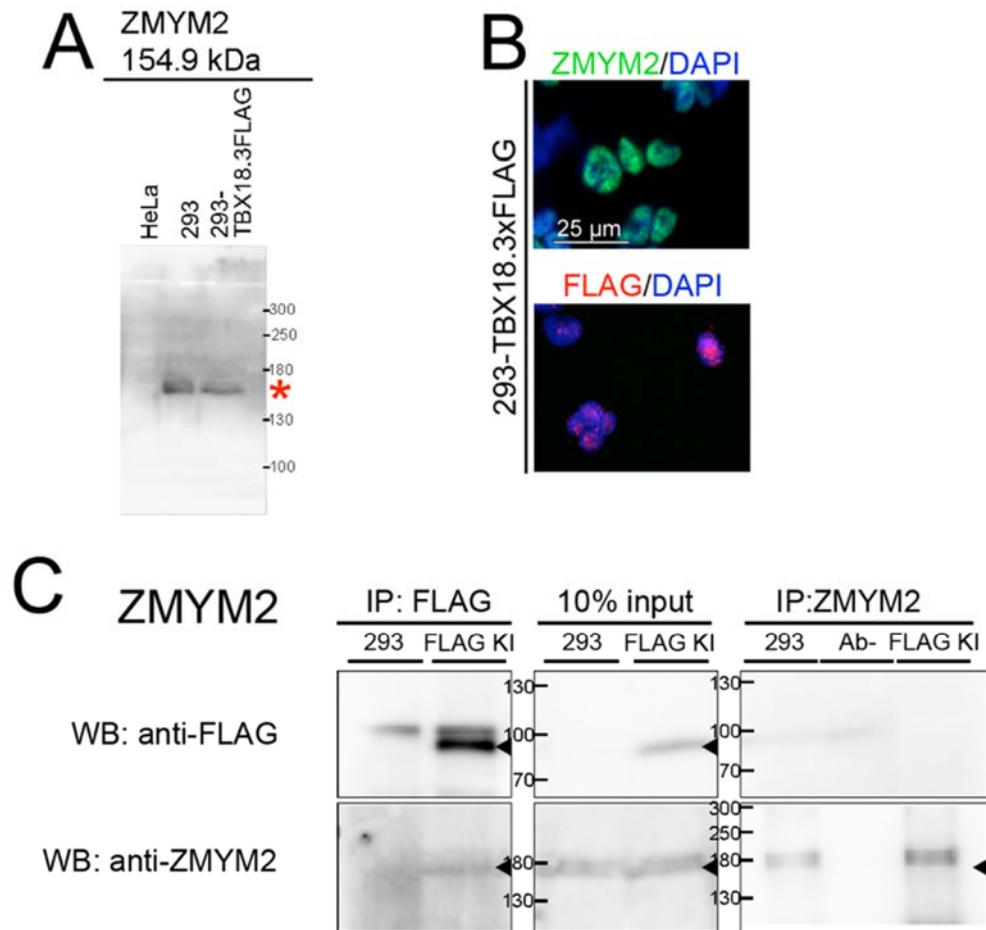


Figure 3. Endogenous TBX18 interacts with ZMYM2 in 293 cells. **(A)** Western blot analysis of endogenous ZMYM2 expression in different cell extracts, showing the antibody against ZMYM2 detects two bands of approximately 160-170 kDa size corresponding to two isoforms of ZMYM2 in 293-TBX18.3xFLAG and 293 parental control cells but not in HeLa cells. **(B)** Immunofluorescent staining of endogenous TBX18.3xFLAG and ZMYM2. Each staining was carried with the indicated primary antibody, with both stainings found in the nucleus. Counterstaining of nuclei was performed with DAPI. **(C)** Representative Western blot analysis of immunoprecipitation assays for TBX18 and ZMYM2. Immunoprecipitation assays were performed with anti-FLAG antibody or the ZMYM2 antibody in lysates of 293-TBX18.3xFLAG or 293 parental control cells. The same antibodies were used to detect the co-precipitated proteins. Left, FLAG immunoprecipitation from both cells. Middle, a ten percent input control is shown. Right, ZMYM2 immunoprecipitation; for the ZMYM2 primary antibody, a no-antibody control immunoprecipitation sample was also analyzed. $n \geq 2$. For each experiment, a single membrane was used to detect both FLAG and candidate proteins in the control and immunoprecipitation samples. However, higher exposure settings are sometimes shown to detect the coimmunoprecipitated proteins. All blots and conditions are detailed in Supplementary Fig. S5. Black arrowheads mark TBX18.3xFLAG protein (upper panel: WB: anti-FLAG), and the candidate protein ZMYM2 (lower panel: WB: anti-ZMYM2).

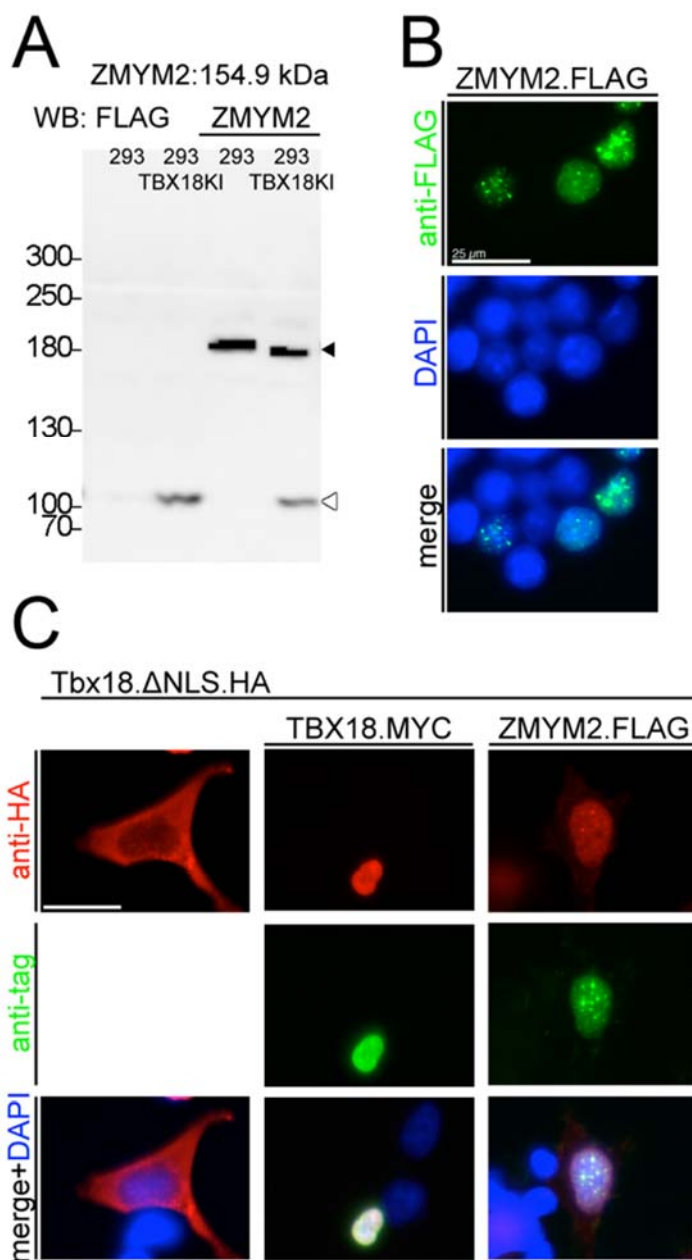


Figure 4. Overexpressed ZMYM2 colocalizes with TBX18 in the nucleus of 293 cells. **(A)** 293 cells were transfected with an expression construct for FLAG-tagged ZMYM2 and whole cell extracts were used to detect the overexpressed protein in both 293-TBX18.3xFLAG or 293 parental control cells. A black arrowhead marks the ZMYM2 protein, while a white arrowhead points to the detected TBX18.3xFLAG protein from the 293 knock-in cells. **(B)** Immunofluorescent analysis detects nuclear localization of overexpressed ZMYM2.FLAG protein in 293 cells. **(C)** Nuclear recruitment assay in 293 cells. Cells were transfected with an expression construct for FLAG-tagged ZMYM2 or Myc-tagged TBX18 in the presence of HA-tagged TBX18 lacking the nuclear localization signal (TBX18ΔNLS.HA). Immunofluorescence analysis shows that NLS-deficient TBX18 protein (*red*) is efficiently shuttled from the cytoplasm to the nucleus by overexpressed TBX18 or ZMYM2 (*green*).

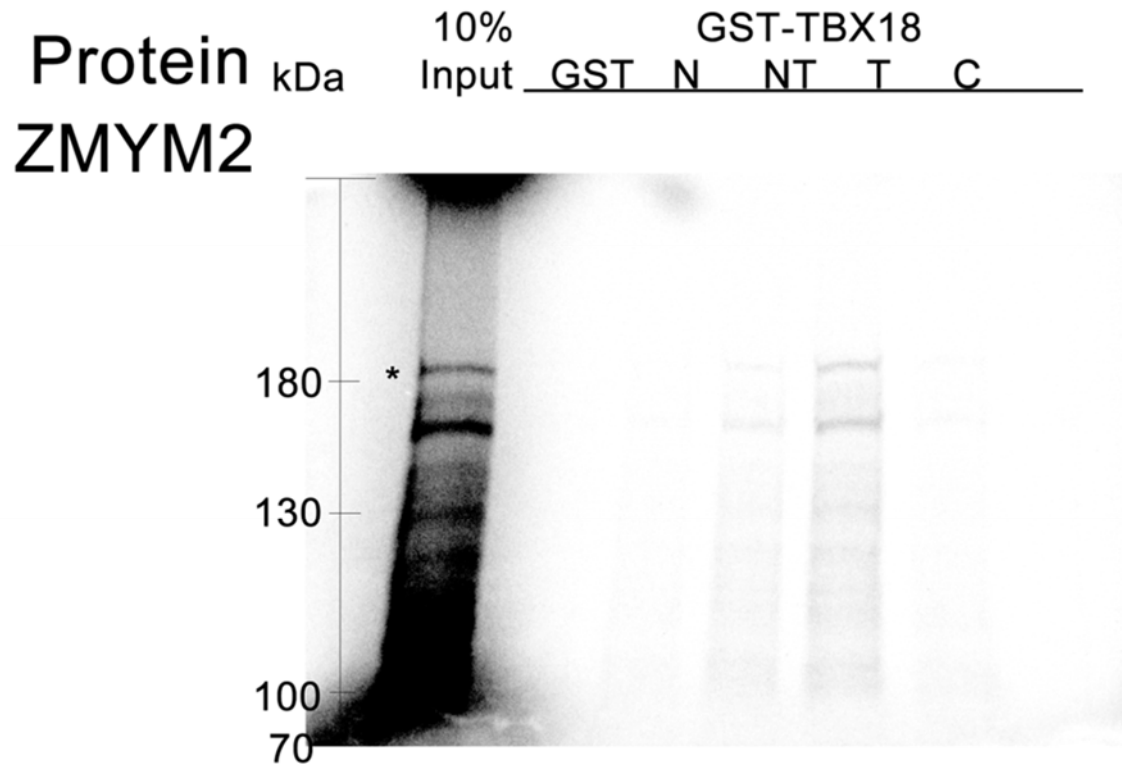


Figure 5. ZMYM2 interacts physically with the T-box region of TBX18. Autoradiographic analysis of pull-down assays performed with GST and fusion proteins of GST with N-, N+T (NT), T-, and C-regions of TBX18 obtained from *E. coli* extracts, and reticulocyte lysates programmed for *in vitro* translation of ^{35}S -labelled tagged full-length ZMYM2. ZMYM2 preferentially interacts with the T-box region of TBX18. Ten percent assay input and GST mock negative controls are shown. $n \geq 2$. Both of the produced isoforms that are also endogenously expressed can be distinguished by an asterisk around 170 kDa.

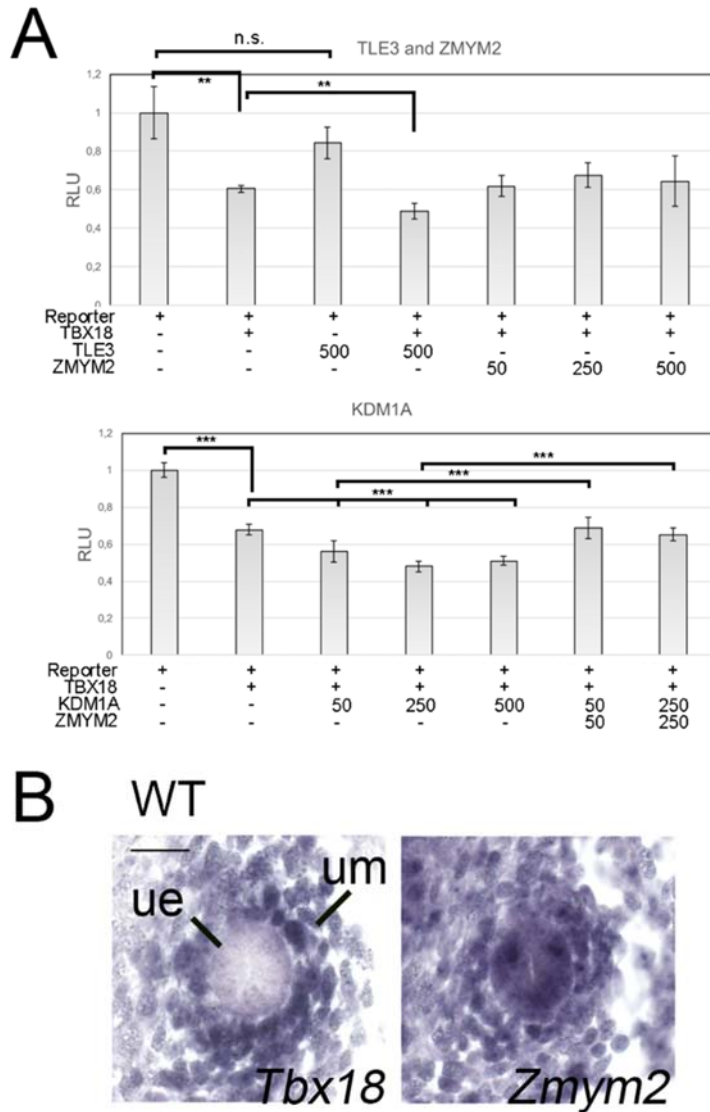
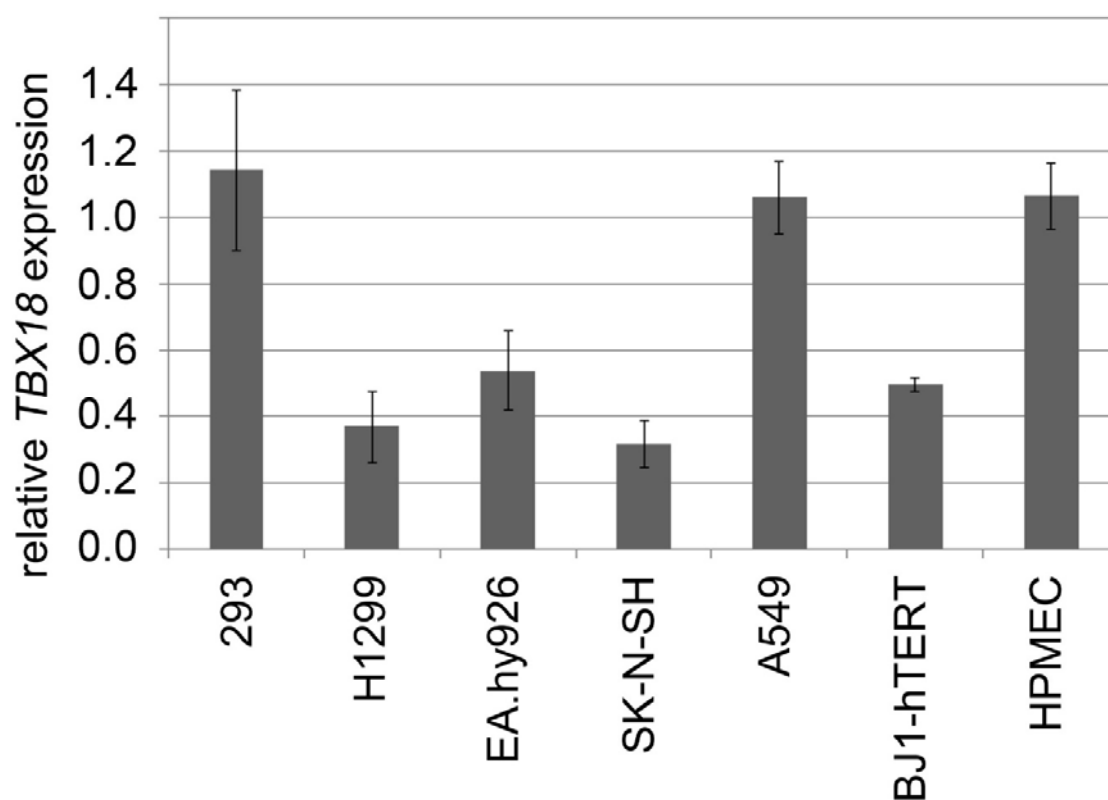
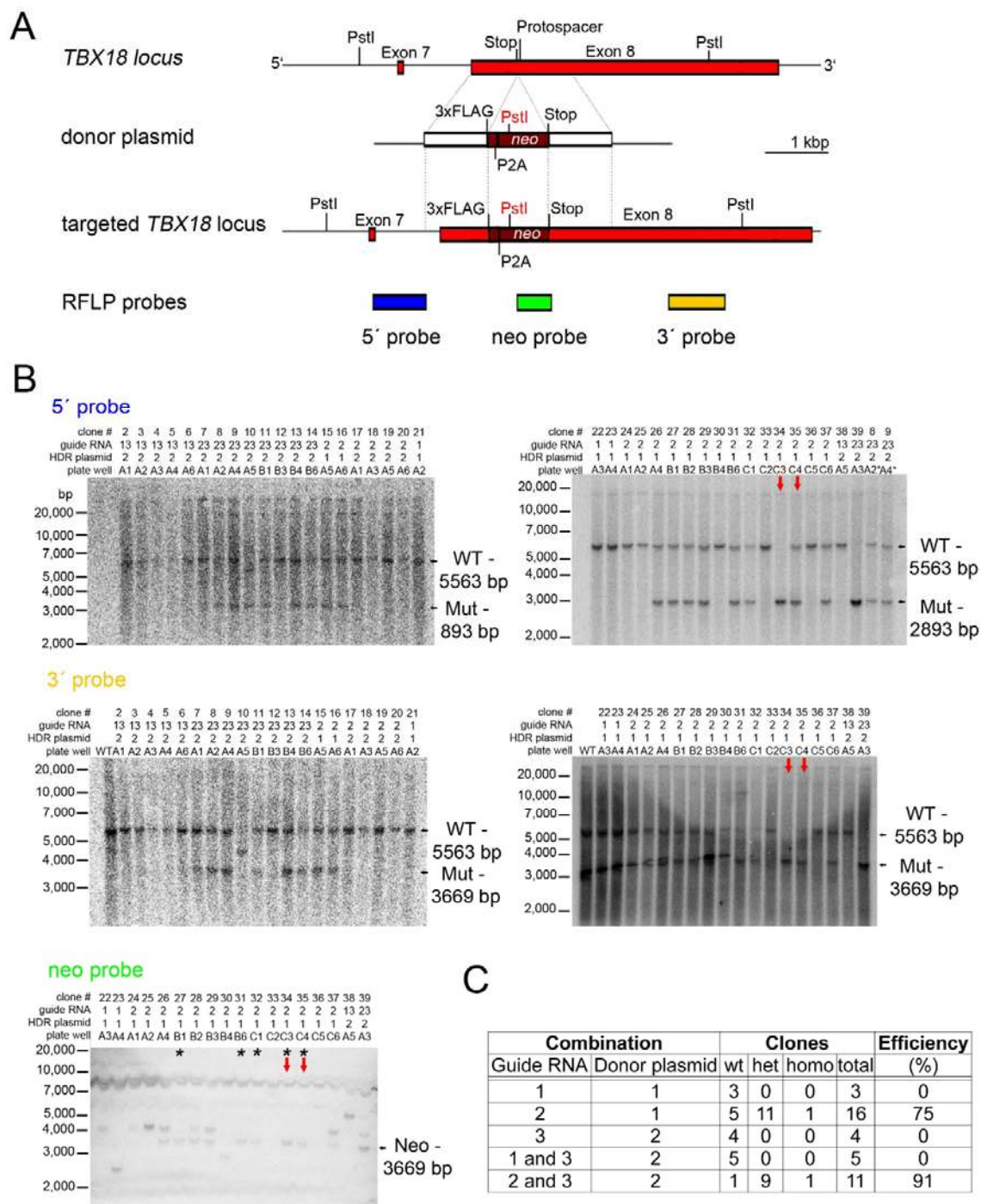


Figure 6. (A) Transactivation assays demonstrate functional regulation of a TBX18 reporter gene construct. Dual luciferase assays were performed with extracts of 293 cells cotransfected with constructs for the TBX18 luciferase reporter and expression constructs for full-length TBX18 in combination with TLE3 (positive control) ZMYM2 and KDM1A plus ZMYM2 proteins (amounts in ng). Upper panel, cotransfection of TBX18 and the reporter reduces luciferase activity significantly by 30-40%. The TLE3 corepressor further represses reporter activity only in presence of TBX18. Increasing concentrations of ZMYM2 do not affect transcriptional repression by TBX18. Lower panel, Lysine Demethylase 1A (KDM1A) represses luciferase reporter activity in presence of TBX18 in a concentration dependent manner. The addition of ZMYM2 reversed this effect. An experiment comprising results of two biological replicates each with two technical replicates are shown in relative light units (RLU) and normalized to the RLU of the reporter alone. One representative experiment is shown ($n \geq 2$). The statistical analysis and setup description are found in Supplementary Table S4 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(B)** Murine *Zmym2* is coexpressed with *Tbx18* in the ureteric mesenchyme. Comparative RNA *in situ* hybridization analysis of expression of *Tbx18* and *Zmym2* on transverse sections of the proximal ureter of E12.5 wildtype embryos. Size bar is 25 μ m. ue, ureteric epithelium; um, ureteric mesenchyme.

Supplementary Figures

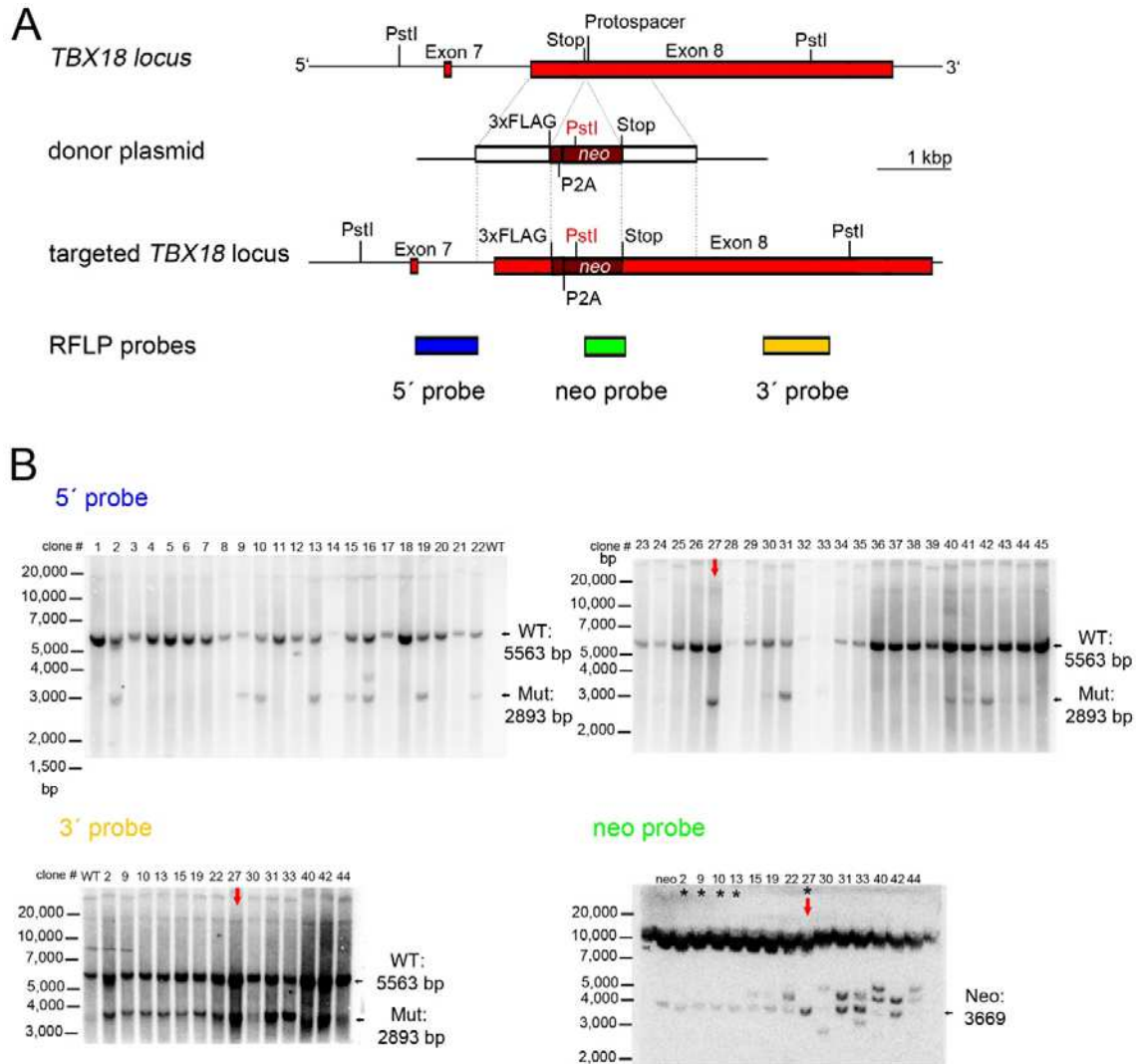


Supplementary Figure S1. *TBX18* is expressed in various human cell lines. Semi-quantitative RT-PCR analysis of *TBX18* mRNA expression in 293 human embryonic kidney cells, H1299 non-small cell cancer cells, EA.hy926 endothelial cells, SK-N-MC neuroepithelioma cells, SK-N-SH neuroblastoma cells, A549 lung carcinoma cells, BJ1-hTERT telomerase immortalized fibroblasts, and HPMEC primary human pulmonary microvascular endothelial cells. *GAPDH* was used as a housekeeping control and all values were calculated as relative gene per *GAPDH* ratios. A549 (1.06 ± 0.11) and HPMEC (1.06 ± 0.101) cells exhibit a similarly strong expression of *TBX18* as 293 cells; in all other tested cell lines *TBX18* expression was considerably lower.

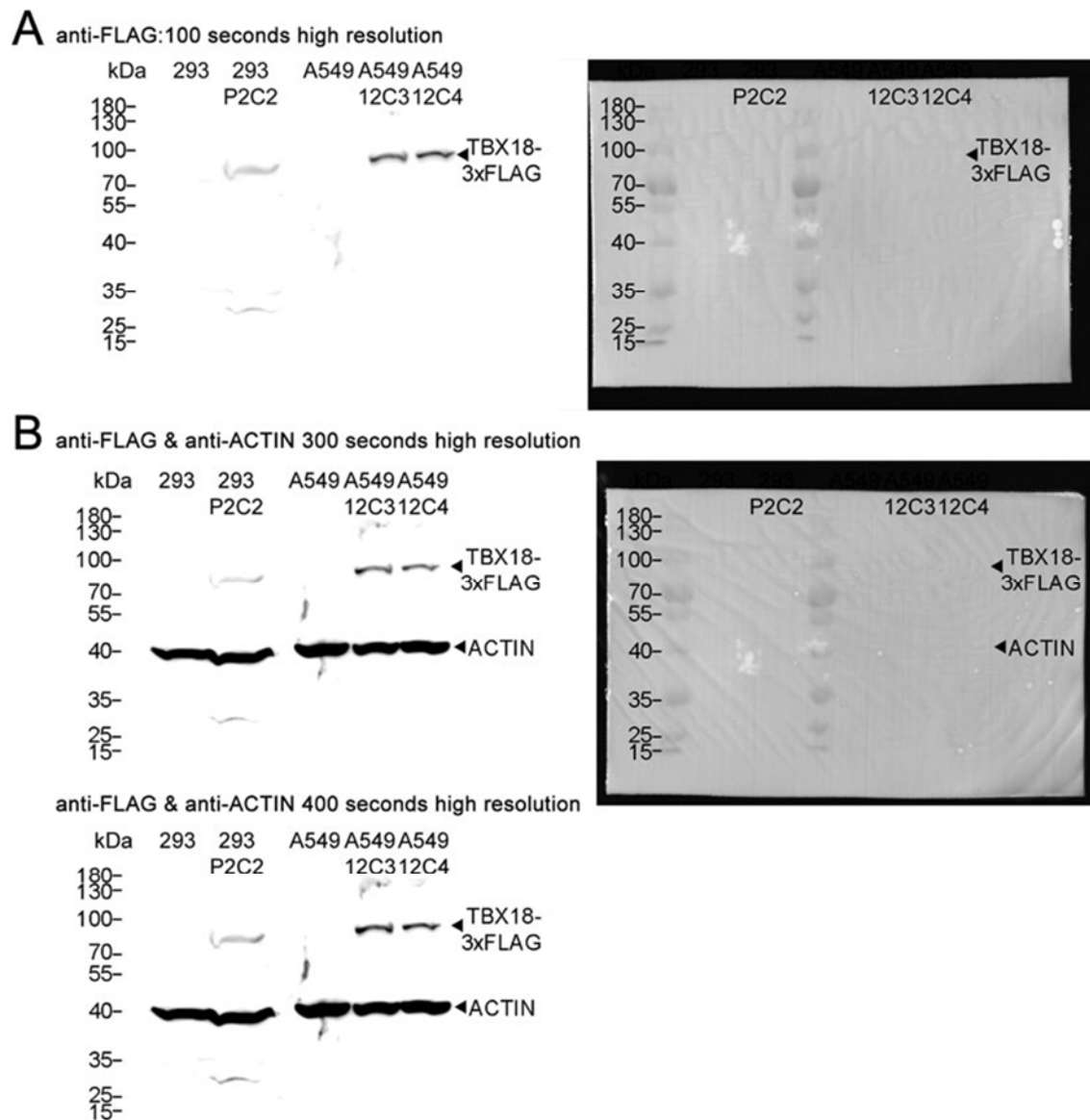


Supplementary Figure S2.

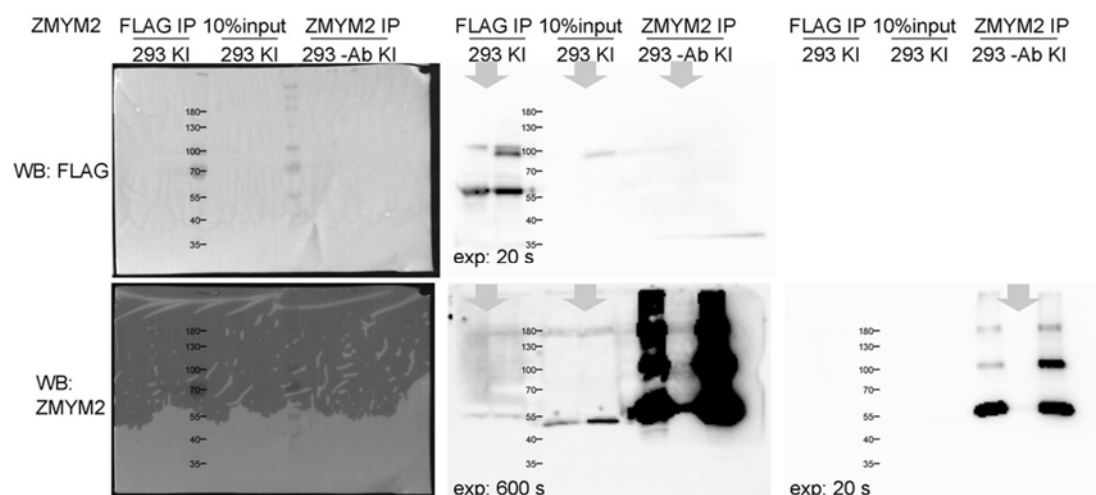
Supplementary Figure S2. RFLP analysis identifies A549 clones with CRISPR/Cas9-mediated FLAG-tagging of the *TBX18* locus. (A) Scheme of the CRISPR/Cas9 and homology mediated repair strategy to insert a triple FLAG peptide at the C-terminal end of *TBX18*. Red boxes indicate exon 7 and 8 of the *TBX18* locus on human chromosome 6. Thin dashed lines mark the homology region between the donor plasmid (white boxes) and the *TBX18* locus before and after targeting to insert a cassette containing a triple FLAG tag, a P2A protein cleavage site, a *neomycin* resistance gene (*neo*) and a stop codon. *Pst*I restriction sites and probes used for RFLP analysis of the integration event are shown. (B) Southern blot analysis of *Pst*I digested genomic DNA of individual A549 cell clones. All clones were screened with the 5'- and 3'-probes, some additionally with a *neo* probe. The sizes of the expected wildtype and mutant fragments for the different RFLPs are indicated. Note the labelling system: the upper number is the unique running number for each clone. The second number (1 and/or 2 and/or 3) shows which guide RNA(s) was/were used. The third number (1 or 2) shows the number of the donor plasmid used. The fourth number refers to an internal positioning system. (C) Analysis of the efficiency of homologous recombination using different combinations of guide RNA and donor plasmids. Note that only guide RNA 2 mediated *TBX18* targeting, whereas both donor plasmids worked. 20 of the targeted clones were heterozygous (het), 2 were homozygous (homo). Only five of these clones had a unique *neo* insertion (*). One heterozygous clone (A549-12C3) and one homozygous clone (A549-12C4), marked with red arrows, were used for subsequent experiments.



Supplementary Figure S3. RFLP analysis identifies 293 clones with CRISPR/Cas9-mediated FLAG-tagging of the *TBX18* locus. (A) Scheme of the CRISPR/Cas9 and homology mediated repair strategy to insert a triple FLAG peptide at the C-terminal end of *TBX18*. Red boxes indicate exon 7 and 8 of the *TBX18* locus on human chromosome 6. Thin dashed lines mark the homology region between the donor plasmid (white boxes) and the *TBX18* locus before and after targeting to insert a cassette containing a triple FLAG tag, a P2A protein cleavage site, the *neomycin* resistance gene (*neo*) and the stop codon. *Pst*I restriction sites and probes used for RFLP analysis of the integration event are shown. Note that for 293 cells only the combination of guide 2 RNA plasmid and donor plasmid 1 was used for electroporation. (B) Southern blot analysis of *Pst*I digested genomic DNA for homologous recombination in the *TBX18* locus. 5'-probe (blue): all clones were screened with the 5'-probe. 3'-probe (yellow): DNA of the 15 clones with the correct 5'-RFLP were subjected to hybridization with the 3'-probe. All of them exhibited a mutant band of the expected size. Neo probe (green): hybridization with the *neo* probe revealed 5 clones with correct unique *neo* integration in the *TBX18* locus (*). Red arrows mark DNA of the clone (#27, P2C2) that was finally used for further analysis.



Supplementary Figure S4. FLAG-tagging allows detection of endogenous TBX18 protein in 293 and A549 cells through Western blot. (A) Western blot analysis of cell lysates from parental and TBX18-FLAG targeted clones with an anti-FLAG antibody with low exposure settings. (B) Western blot analysis the same membrane as in (A) with an anti-FLAG antibody together with anti-ACTIN antibody with lower (up) and higher (down) exposure settings. Molecular weight markers are shown in the membrane.



Supplementary Figure S5. Endogenous TBX18 interacts with ZMYM2 in 293 cells.

Coimmunoprecipitation of endogenous ZMYM2 with TBX18 was analysed by Western blot. Cell lysates were preincubated with anti-FLAG beads or anti-ZMYM2 antibody plus protein A beads, the beads were washed and analysed for coimmunoprecipitation of TBX18 or ZMYM2. Each experiment was analysed in a single blot as shown in matching pictures, and the antibodies against FLAG (up) or ZMYM2 (down) were developed sequentially. To detect the coimmunoprecipitated proteins higher exposure settings were sometimes required, for this reason, we show the pictures of both exposure settings included in Figure 3, the cropped correct weight bands are marked with a grey arrow.

Concluding remarks and outlook

Characterizing the molecular mechanisms of TBX18 function is prerequisite to understand both its wide roles during embryonic development as well as the mechanisms behind CAKUT caused by TBX18 loss-of-transcriptional function in humans. In order to elucidate these mechanisms, I performed proteomic approaches to identify TBX18 interaction partners relevant for transcriptional regulation. Using two independent approaches in 293 and A549 cells, I identified and characterized candidate proteins that shed light on the nature of transcriptional repression by TBX18.

The molecular mechanisms of TBX18 repression

TBX18 has been characterized as a repressor of transcription by independent gene reporter assays, both on its own and in heterotypic TF complexes, acting upon diverse regulatory elements. The full-length protein and sub-domains were shown to repress transcription of a GAL/UAS (yeast derived trans-acting elements) reporter inside HeLa cells. In the same cells, the full-length protein also acted as a competitor to the TBX6-TCF1E/B-CAT proteins and repressed transcription of an enhancer-promoter construct derived from the mouse *Dll1* regulatory elements. Additionally, TBX18 was found to repress CArG-box containing constructs related to smooth muscle differentiation. TBX18 repressed *SM22 α* and *c-fos* regulatory elements on its own and also prevented the activation of *SM γ A* regulatory elements by MYOCD and SRF (Farin et al., 2007; Farin et al., 2008; Vivante et al., 2015; Wu et al., 2013). However, unlike other T-Box proteins, no transrepressor domain has been identified in the TBX18 protein and the link between its DNA-recognition capabilities and transcription repression activity is poorly understood.

Conceptually, to repress transcription of such diverse molecular activities, TBX18 needs to either interact directly with the transcription machinery or indirectly through proteins that negatively regulate this event. Moreover, TBX18 and its interactors may regulate transcription via the transcription machinery and/or by changing the accessibility of the chromatin (Fig 4).

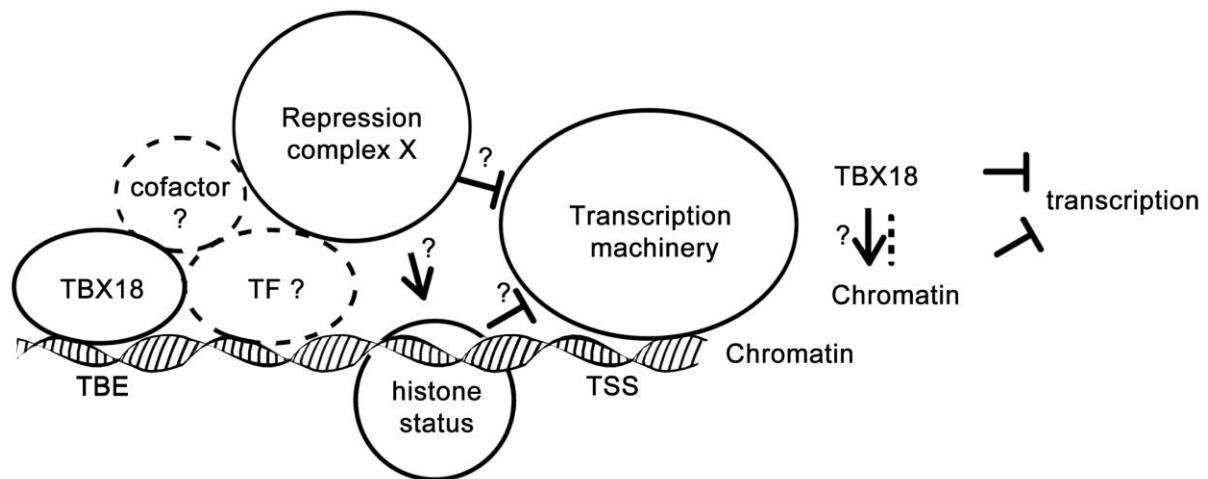


Fig 4. Possible molecular interactions responsible for TBX18 transcriptional repression.

TBX18 has been extensively described as a repressor of transcription in various cellular contexts. Independent of the precise mode, protein interactions are needed for TBX18 to interact with the transcription machinery directly or to actively modify the chromatin status. These proteins may be represented by specific context-dependent TFs or common transcriptional cofactors.

TBX18 interacts with Homeobox TFs and corepressors

To identify and validate the protein interactions needed for TBX18 function, we embarked on the analysis of the first unbiased interactome of TBX18 using a TBX18 overexpression approach in 293 cells. The identification was performed with nuclear extracts from cells overexpressing a double-tagged TBX18 version versus controls, which greatly decreases the detection of false positives. With this approach, we identified and validated two different classes of molecules as TBX18 interactors.

First, homeobox transcription factors (DUXBLI, GSC, PAX1, PBX1, PBX4 and PRRX2) bound to TBX18 in a direct *in vitro* assay. Except PAX1, that doesn't have a homeodomain but can bind TBX18 through its paired-box domain (Farin et al., 2008), all these proteins have a homeodomain as a common feature but belong to very different homeobox subclasses (Holland, Booth, & Bruford, 2007). Given that the T-box domain has previously been shown to interact with the homeobox domain (Hiroi et al., 2001) and that DUXBLI is mostly comprised of a homeobox domain, the assumption is valid that the homeobox domain of all candidate proteins binds directly to the T-box domain of TBX18. From these TFs, we showed TBX18 coexpression with members of the PBX and PRRX TF families in the ureteric mesenchyme, but only PRRX family proteins count with an OAR trans nhibition domain to influence transcription directly and therefore, additional proteins would be needed to exert

transcriptional regulation (Lu et al., 2011). Thus, TBX18 interaction with homeobox transcription factors may need accessory proteins for transcriptional relevance.

Secondly, we identified transcriptional cofactors as a second group of TBX18 interacting proteins. We validated interaction with 9 proteins (BASP1, CBFB, CHD7, GAR1, IKZF2, NCOA5, RCOR3, SBNO2 and SSXB2) in the nucleus of 293 cells, of which 6 directly physically interacted with TBX18 (CBFB, CHD7, GAR1, IKZF2, NCOA5 and SBNO2). Importantly, three of these candidates clearly regulated transcription together with TBX18 in a luciferase reporter assay (IKZF2 acted as corepressor, NCOA5 and SBNO2 as coactivators). The transcription reporter assay employed in this study makes use of a palindromic arrangement of 2 TBE repeats in front of a minimal promoter to drive luciferase expression in a 293 cellular context, in this assay, however no effects on the chromatin status should be observed, since it makes use of an episomal plasmid with no histone content. Therefore, the joint transcriptional repression effect of TBX18 and IKZF2 should relate directly to the transcription machinery activity. We can also highlight the fact that two of the aforementioned interacting proteins, IKZF2 and CHD7 are part of the Nucleosome Remodeling Deacetylase complex (NuRD). IKZF2 is a member of this complex and CHD units of this complex are interchangeable (at least between CHD3, 4 and 5) (Georgopoulos, 2017) (Low et al., 2016). This points to the possibility that the repressive function of TBX18 is mediated by deacetylation of histones by this complex. The CHD7 protein itself contains a helicase/ATPase domain that can enzymatically rearrange the chromatin state (Hall & Georgel, 2007). These interactions support a proposed NuRD-mediated TBX18 repression mechanism through the modification of the chromatin state. Another relevant protein of the NuRD complex for repression, the HDAC family proteins 1 or 2, that modify chromatin state and target gene accessibility (Low et al., 2016) were not found in our screen, but will be interesting candidates for further evaluation.

This first approach was designed to use an accessible source of proteins through bait overexpression and a high stringency for candidate protein identification by a 2 step purification process, as shown by the high number of physically directly bound proteins we examined (19/25), but was not without side effects. The clustering employed to restrict transcriptionally interesting candidates pointed us to a group of “centrosomal” proteins that localized to the cytoplasm and did not interact inside the cell nucleus with TBX18, for this reason they do not constitute relevant partners to explain TBX18

function, and we deemed the purification of these proteins an artefact of overexpressing TBX18.

The zinc factor transcription factor ZMYM2 is an endogenous interaction partner of TBX18

As a complementary and independent approach, we sought to screen the interactome of TBX18 in cell lines exhibiting endogenous TBX18 expression. This approach employed the CRISPR/Cas9 system to genetically modify the TBX18 locus to integrate a non-invasive tag for purification of TBX18 and its bound proteins.

We chose to modify the genome of 2 cell lines with the highest TBX18 expression in our hands, 293 and A549, which derive from very different biological sources (293, embryonic kidney cells; A549, lung carcinoma cells). After validating the integration of the 3xFLAG tag to the C-terminus of TBX18 in these cells, we identified candidate interacting proteins with a single step purification process. Our identification results were screened with caution owing to the less stringent purification approach. The statistical analysis revealed that 10 proteins were statistically enriched in 293 cells, 36 were statistically enriched in A549 cells versus the control parental cells.

To our surprise, we identified much different and less candidate proteins for TBX18 interaction compared to our first study. In 293 cells, several zinc finger proteins relevant for transcription stood out of the short list, while in A549 cells, the enriched proteins contained not only zinc-finger transcription factors but also 2 corepressors and a group of ribosomal proteins. To search if TBX18 had a common molecular mechanism between these different contexts, we chose to overlap the enriched proteins from the two cell lines and found out a single common protein partner, ZMYM2 which is a member of the CoREST repressor complex.

We characterized that TBX18 and ZMYM2 interact in a physically direct manner inside the nucleus of cells. In contrast to the IKZF2 zinc-finger protein, we initially validated, ZMYM2 did not regulate TBX18-dependent repression of transcription in the same TBX18 reporter assay. The IKZF2 structure comprises 4 N-terminal classic C2H2 zinc-finger domains important for NuRD and DNA interaction and 2 more C-terminal zinc finger domains important for dimerization with other IKZF family members, with additional PEST motifs (Georgopoulos, 2017). In comparison, ZMYM2 contains 9-10 tandem MYM type zinc finger domains, which are still unmapped to a particular function, a proline-valine rich region, a NLS and a domain of unknown function

(DUF3504) (Reiter et al., 1998). Although IKZF2 and ZMYM2 are both classified as zinc-finger proteins, their zinc/finger domains are differently composed (C2-H2, C2-C2 frameworks) and arranged, and the way they interact with TBX18 and the transcription machinery is still uncharacterized.

To test if ZMYM2 was relevant for transcription repression inside a complex, we tested if another member of the CoREST complex relevant for its enzymatic repression activity, KDM1A could co-regulate TBX18-dependent transcriptional repression. KDM1A repressed transcription cooperatively with TBX18 and when ZMYM2 was also added to this model, we found ZMYM2 derepressed the transcription exerted by TBX18-KDM1A.

This approach to identify and validate the interaction partners of TBX18 has downsides compared to our first study. The one-step purification process is less stringent, as reflected by our finding of less statistically enriched proteins for each cell line used. However, we identified and validated a new TF interaction partner of TBX18 and related TBX18 function to another corepressor complex. Notably, ZMYM2 has just been described by the laboratory of Prof. Friedhelm Hildebrand (Harvard Medical School, personal communication to Prof. Andreas Kispert) to cause the same kind of uro- and nephropathies as TBX18 in a large human patient cohort study. We then found both TFs coexpressed in mice during ureter development, providing a strong hint that both factors may indeed interact in the development of the ureteric mesenchyme *in vivo*.

Although no consistent overlap was found between our two proteomic approaches, even within the 293 cell line data sets, we still found the same classes of interaction partners between both studies, namely transcriptional cofactors, homeobox and zinc-finger transcription factors. This triad of T-box, homeobox and Zn-finger TFs was previously recognized as crucial interactors in heart development (TBX5/GATA4/NKX2.5) (Luna-Zurita et al., 2016) (Fig.5)

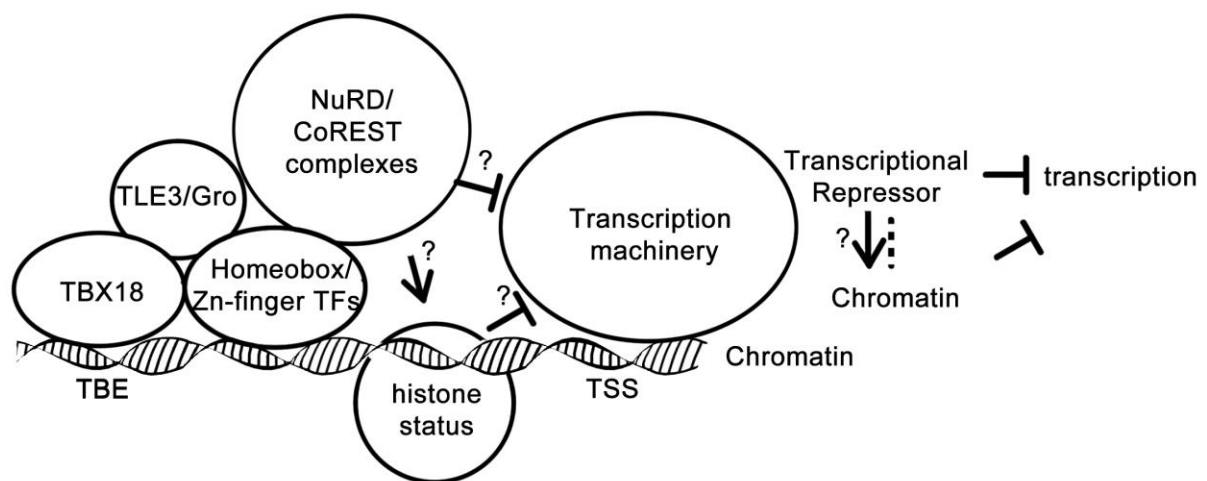


Fig. 5. Scheme of TBX18 and its interacting proteins as identified and validated with both overexpression and endogenous purification approaches in the regulation of transcription.

The TBX18 TF can bind to corepressors, homeobox and zinc-finger TFs. We found several members of the CoREST and NuRD repressor complexes represented in our datasets. The next task would be to find how they act on the transcription machinery and the chromatin. In the case of TLE3, it may have an aggregative local effect on chromatin binding and pause transcription.

Future work

We screened for interaction partners of TBX18 in two different approaches. First by overexpression in 293 cells. Second, by looking at endogenous TBX18 bound proteins from the 293 and A549 cell lines. From these screenings, we validated three classes of proteins that interact with TBX18, both in vitro and inside the nucleus of 293 cells; transcriptional cofactors, homeobox and zinc-finger transcription factors. Taking together, proteins that interact with TBX18 from both studies, we propose two new mechanisms through which TBX18 may regulate transcription.

First, the interaction with IKZF2 points to a direct repressor effect exerted on the transcriptional machinery while the interactions with IKZF2 and CHD7 suggest a possible involvement of the NuRD corepressor complex on transcriptional regulation. Second, the interaction of TBX18 with ZMYM2, RCOR3 and KDM1A points to the involvement of the CoREST repressor complex on transcriptional regulation. The proposed mechanisms of regulation of transcription may not only affect directly the transcription machinery but they are known to regulate the chromatin state of genes. In order to validate these mechanisms, it would be optimal to assess the

relevance of the interactions found on the natural molecular contexts in which TBX18 functions.

An important task for this matter is the validation of *bona fide* target genes, which would allow for the assessment of the joint transcription regulation by these proteins on actual regulatory elements from a target gene and define whether the effects rely on the transcription machinery or the chromatin state surrounding the regulatory elements.

To confirm endogenous involvement of TBX18 with these proteins *in vivo*, it would be useful to tag the endogenous TBX18 locus of an organism model, like the mouse, to circumvent the lack of TBX18 antibodies suitable to study protein interaction. This would also allow for the identification of endogenous TBX18 target genes through a chromatin immunoprecipitation and sequencing (CHIP-Seq) method. Initially, these mechanisms could be tested in the context of the ureter development, where TBX18 function has been carefully described.

To assess the biological relevance of the TBX18 interactions with NuRD complex proteins, additional efforts should be put first to study the interactions between the possible subunits involved with TBX18 interaction, importantly HDAC1/2, and whether CHD7 can be confirmed as the member of the NuRD complex. No obvious shared phenotype has been observed between *Tbx18*, *Chd7* or *Ikzf2* knockouts, and only targeted mutations for *Chd7* have been characterized in mouse.

On the other hand, the relevance of the CoREST complex proteins interactions can be examined *in vivo* through genetic tools. Targeted mutations of the loci of *Tbx18* and *Zmym2* in the mouse have been established and the mouse lines can be crossed to assess if heterozygosity of both genes produces a genetic interaction compared to the single knockout of both genes. The readout for this interaction would be the presence and severity of nephropathies, as in the individual complete loss-of-function situation. It remains an interesting question whether TBX18 operates in the same molecular complexes in different biological contexts. If this is the case, it would mean there is a common molecular mechanisms of transcription regulation by TBX18, otherwise, the molecular mechanisms would be context dependent and the specificity of TBX18 transcription regulation program would rely on specific interaction partners in each biological context.

References

- Airik, R., Bussen, M., Singh, M. K., Petry, M., & Kispert, A. (2006). Tbx18 regulates the development of the ureteral mesenchyme. *J Clin Invest*, 116(3), 663-674. doi:10.1172/JCI26027
- Arce, L., Pate, K. T., & Waterman, M. L. (2009). Groucho binds two conserved regions of LEF-1 for HDAC-dependent repression. *BMC Cancer*, 9, 159. doi:10.1186/1471-2407-9-159
- Bohnenpoll, T., Bettenhausen, E., Weiss, A. C., Foik, A. B., Trowe, M. O., Blank, P., . . . Kispert, A. (2013). Tbx18 expression demarcates multipotent precursor populations in the developing urogenital system but is exclusively required within the ureteric mesenchymal lineage to suppress a renal stromal fate. *Dev Biol*, 380(1), 25-36. doi:10.1016/j.ydbio.2013.04.036
- Brantjes, H., Roose, J., van De Wetering, M., & Clevers, H. (2001). All Tcf HMG box transcription factors interact with Groucho-related co-repressors. *Nucleic Acids Res*, 29(7), 1410-1419. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11266540>
- Bussen, M., Petry, M., Schuster-Gossler, K., Leitges, M., Gossler, A., & Kispert, A. (2004). The T-box transcription factor Tbx18 maintains the separation of anterior and posterior somite compartments. *Genes Dev*, 18(10), 1209-1221. doi:10.1101/gad.300104
- Cadigan, K. M., & Waterman, M. L. (2012). TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harb Perspect Biol*, 4(11). doi:10.1101/cshperspect.a007906
- Caramori, G., Ruggeri, P., Mumby, S., Atzeni, F., & Adcock, I. M. (2019). Transcription Factors. *eLS*, 1-8.
- Chambers, M., Turki-Judeh, W., Kim, M. W., Chen, K., Gallaher, S. D., & Courey, A. J. (2017). Mechanisms of Groucho-mediated repression revealed by genome-wide analysis of Groucho binding and activity. *BMC Genomics*, 18(1), 215. doi:10.1186/s12864-017-3589-6
- Ettensohn, C. A. (2013). Encoding anatomy: developmental gene regulatory networks and morphogenesis. *Genesis*, 51(6), 383-409. doi:10.1002/dvg.22380
- Farin, H. F., Bussen, M., Schmidt, M. K., Singh, M. K., Schuster-Gossler, K., & Kispert, A. (2007). Transcriptional repression by the T-box proteins Tbx18 and Tbx15 depends on Groucho corepressors. *J Biol Chem*, 282(35), 25748-25759. doi:10.1074/jbc.M703724200
- Farin, H. F., Mansouri, A., Petry, M., & Kispert, A. (2008). T-box protein Tbx18 interacts with the paired box protein Pax3 in the development of the paraxial mesoderm. *J Biol Chem*, 283(37), 25372-25380. doi:10.1074/jbc.M802723200

Georgopoulos, K. (2017). The making of a lymphocyte: the choice among disparate cell fates and the IKAROS enigma. *Genes Dev*, 31(5), 439-450. doi:10.1101/gad.297002.117

Ghosh, T. K., Brook, J. D., & Wilsdon, A. (2017). T-Box Genes in Human Development and Disease. *Curr Top Dev Biol*, 122, 383-415. doi:10.1016/bs.ctdb.2016.08.006

Gilmour, D., Rembold, M., & Leptin, M. (2017). From morphogen to morphogenesis and back. *Nature*, 541(7637), 311-320. doi:10.1038/nature21348

Goldman, S. R., Ebright, R. H., & Nickels, B. E. (2009). Direct detection of abortive RNA transcripts in vivo. *Science*, 324(5929), 927-928. doi:10.1126/science.1169237

Hall, J. A., & Georgel, P. T. (2007). CHD proteins: a diverse family with strong ties. *Biochem Cell Biol*, 85(4), 463-476. doi:10.1139/O07-063

Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R., & Komuro, I. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat Genet*, 28(3), 276-280. doi:10.1038/90123

Holland, P. W., Booth, H. A., & Bruford, E. A. (2007). Classification and nomenclature of all human homeobox genes. *BMC Biol*, 5, 47. doi:10.1186/1741-7007-5-47

Kaltenbrun, E., Greco, T. M., Slagle, C. E., Kennedy, L. M., Li, T., Cristea, I. M., & Conlon, F. L. (2013). A Gro/TLE-NuRD corepressor complex facilitates Tbx20-dependent transcriptional repression. *J Proteome Res*, 12(12), 5395-5409. doi:10.1021/pr400818c

Kaul, A., Schuster, E., & Jennings, B. H. (2014). The Groucho co-repressor is primarily recruited to local target sites in active chromatin to attenuate transcription. *PLoS Genet*, 10(8), e1004595. doi:10.1371/journal.pgen.1004595

Lai, A. Y., & Wade, P. A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer*, 11(8), 588-596. doi:10.1038/nrc3091

Li, X. Y., Thomas, S., Sabo, P. J., Eisen, M. B., Stamatoyannopoulos, J. A., & Biggin, M. D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biol*, 12(4), R34. doi:10.1186/gb-2011-12-4-r34

Low, J. K., Webb, S. R., Silva, A. P., Saathoff, H., Ryan, D. P., Torrado, M., . . . Mackay, J. P. (2016). CHD4 Is a Peripheral Component of the Nucleosome Remodeling and Deacetylase Complex. *J Biol Chem*, 291(30), 15853-15866. doi:10.1074/jbc.M115.707018

Lu, X., Beck, G. R., Jr., Gilbert, L. C., Camalier, C. E., Bateman, N. W., Hood, B. L., . . . Nanes, M. S. (2011). Identification of the homeobox protein Prx1 (MHox, Prrx-1) as a regulator of osterix

expression and mediator of tumor necrosis factor alpha action in osteoblast differentiation. *J Bone Miner Res*, 26(1), 209-219. doi:10.1002/jbmr.203

Luna-Zurita, L., Stirnimann, C. U., Glatt, S., Kaynak, B. L., Thomas, S., Baudin, F., . . . Mileikovsky, M. (2016). Complex interdependence regulates heterotypic transcription factor distribution and coordinates cardiogenesis. *Cell*, 164(5), 999-1014.

Maniatis, T., & Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature*, 416(6880), 499-506. doi:10.1038/416499a

Muller, C. W., & Herrmann, B. G. (1997). Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature*, 389(6653), 884-888. doi:10.1038/39929

Naiche, L. A., Harrelson, Z., Kelly, R. G., & Papaioannou, V. E. (2005). T-box genes in vertebrate development. *Annu Rev Genet*, 39, 219-239. doi:10.1146/annurev.genet.39.073003.105925

Nie, X., Sun, J., Gordon, R. E., Cai, C. L., & Xu, P. X. (2010). SIX1 acts synergistically with TBX18 in mediating ureteral smooth muscle formation. *Development*, 137(5), 755-765. doi:10.1242/dev.045757

Papaioannou, V. E. (2014). The T-box gene family: emerging roles in development, stem cells and cancer. *Development*, 141(20), 3819-3833. doi:10.1242/dev.104471

Perissi, V., Jepsen, K., Glass, C. K., & Rosenfeld, M. G. (2010). Deconstructing repression: evolving models of co-repressor action. *Nat Rev Genet*, 11(2), 109-123. doi:10.1038/nrg2736

Reiter, A., Sohal, J., Kulkarni, S., Chase, A., Macdonald, D. H., Aguiar, R. C., . . . Cross, N. C. (1998). Consistent fusion of ZNF198 to the fibroblast growth factor receptor-1 in the t(8;13)(p11;q12) myeloproliferative syndrome. *Blood*, 92(5), 1735-1742. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9716603>

Sebe-Pedros, A., & Ruiz-Trillo, I. (2017). Evolution and Classification of the T-Box Transcription Factor Family. *Curr Top Dev Biol*, 122, 1-26. doi:10.1016/bs.ctdb.2016.06.004

Sekiya, T., & Zaret, K. S. (2007). Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo. *Mol Cell*, 28(2), 291-303. doi:10.1016/j.molcel.2007.10.002

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, 7, 539. doi:10.1038/msb.2011.75

- Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol*, 19(4), 262-274. doi:10.1038/nrm.2017.115
- Spitz, F., & Furlong, E. E. (2012). Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*, 13(9), 613-626. doi:10.1038/nrg3207
- Trowe, M. O., Maier, H., Schweizer, M., & Kispert, A. (2008). Deafness in mice lacking the T-box transcription factor Tbx18 in otic fibrocytes. *Development*, 135(9), 1725-1734. doi:10.1242/dev.014043
- Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*, 10(4), 252-263. doi:10.1038/nrg2538
- Vivante, A., Kleppa, M. J., Schulz, J., Kohl, S., Sharma, A., Chen, J., . . . Hildebrandt, F. (2015). Mutations in TBX18 Cause Dominant Urinary Tract Malformations via Transcriptional Dysregulation of Ureter Development. *Am J Hum Genet*, 97(2), 291-301. doi:10.1016/j.ajhg.2015.07.001
- Waldron, L., Steimle, J. D., Greco, T. M., Gomez, N. C., Dorr, K. M., Kweon, J., . . . Conlon, F. L. (2016). The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. *Dev Cell*, 36(3), 262-275. doi:10.1016/j.devcel.2016.01.009
- Wiese, C., Grieskamp, T., Airik, R., Mommersteeg, M. T., Gardiwal, A., de Gier-de Vries, C., . . . Christoffels, V. M. (2009). Formation of the sinus node head and differentiation of sinus node myocardium are independently regulated by Tbx18 and Tbx3. *Circ Res*, 104(3), 388-397. doi:10.1161/CIRCRESAHA.108.187062
- Wu, S. P., Dong, X. R., Regan, J. N., Su, C., & Majesky, M. W. (2013). Tbx18 regulates development of the epicardium and coronary vessels. *Dev Biol*, 383(2), 307-320. doi:10.1016/j.ydbio.2013.08.019

Acknowledgements

I would like to thank Prof. Dr. Kispert for the unique opportunity I was given to participate in several scientific projects. I will forever acknowledge the trust he placed upon me to work under his guidance.

I thank my examiners who kindly accepted to evaluate this work.

I thank Dr. Susanne Kruse, Prof. Dr. Reinhold E. Schmidt and the Hannover Biomedical Research School program for all their incredible support.

I thank my cosupervisors and teachers, who showed me the way into the intricate and astonishing world of life sciences. Special thanks to Marc Kleppa and Carsten Rudat for all the fruitful discussions.

I thank my labmates and colleagues who were always open to help and talk during many work hours.

I thank my family for all their unconditional and continuous love along the way. Father, mother and sister, you will be for ever in my heart and thoughts.

I thank my friends and flatmates for their never-ending support in most aspects of living abroad. Cheers my bendehos!

Appendix

Curriculum vitae

List of publications

Declarations

List of publications

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

Morales-González, D., Flores-Martínez, E., Zamora-Bustillos, R., Rivera-Reyes, R., Michel-Morfín, J. E., Landa-Jaime, V., ... & Aguilar, M. B. (2015). Diversity of A-conotoxins of three worm-hunting cone snails (*Conus brunneus*, *Conus nux*, and *Conus princeps*) from the Mexican Pacific coast. *Peptides*, 68, 25-32.

Zamora-Bustillos, R., Rivera-Reyes, R., Aguilar, M. B., Michel-Morfín, E., Landa-Jaime, V., Falcón, A., & Heimer, E. P. (2014). Identification, by RT-PCR, of eight novel I2-conotoxins from the worm-hunting cone snails *Conus brunneus*, *Conus nux*, and *Conus princeps* from the eastern Pacific (Mexico). *Peptides*, 53, 22-29.

Unpublished manuscripts:

Rivera-Reyes, R., Kleppa, M. J., Hildebrandt, F. & Kispert, A.
“Proteomic analysis identifies ZMYM2 as endogenous binding partner of TBX18”

Weiss, A. C., Rivera-Reyes, R. & Kispert, A.
“Expansion of the renal capsular stroma, ureteric bud branching defects and cryptorchidism in mice with Wilms tumor 1 gene deletion in the stromal compartment of the developing kidney”

Declaration

Herewith, I confirm that I have written the present PhD thesis myself and independently, and that I have not submitted it at any other university worldwide.

Herewith, I agree that MHH can check my thesis by plagiarism detection software as well as randomly check the primary data. I am aware that in case of suspicion, ombudsman proceedings according to § 9 of MHH 'Guidelines of Hannover Medical School to guarantee good scientific practice and dealing with scientific fraud' will be initiated. During such proceedings, the PhD process is paused.

Hannover, (January, 2020)
