

RESEARCH ARTICLE

Computational identification of tissue-specific transcription factor cooperation in ten cattle tissues

Lukas Steuernagel¹ , Cornelia Meckbach² , Felix Heinrich¹ , Sebastian Zeidler¹, Armin O. Schmitt^{1,3}, Mehmet Gültas^{1,3*} 

1 Breeding Informatics Group, Department of Animal Sciences, Georg-August University, Margarethe von Wrangell-Weg 7, 37075 Göttingen, Germany, **2** Institute of Medical Bioinformatics, Goldschmidtstraße 1, University Medical Center Göttingen, Georg-August-University, 37077 Göttingen, Germany, **3** Center for Integrated Breeding Research (CiBreed), Albrecht-Thaer-Weg 3, Georg-August University, 37075, Göttingen, Germany

 These authors contributed equally to this work.

* gueltas@cs.uni-goettingen.de



OPEN ACCESS

Citation: Steuernagel L, Meckbach C, Heinrich F, Zeidler S, Schmitt AO, Gültas M (2019) Computational identification of tissue-specific transcription factor cooperation in ten cattle tissues. *PLoS ONE* 14(5): e0216475. <https://doi.org/10.1371/journal.pone.0216475>

Editor: Raluca Mateescu, University of Florida, UNITED STATES

Received: October 18, 2018

Accepted: April 22, 2019

Published: May 16, 2019

Copyright: © 2019 Steuernagel 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 manuscript and its Supporting Information files.

Funding: CM was funded by ExiTox2 (grant: 031L0120B) of the BMBF (German Ministry of Education and Research). FH was partially funded by Deutsche Gesellschaft für Klinische Chemie und Laboratoriumsmedizin (DGKL). We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Göttingen University. The authors declare that the

Abstract

Transcription factors (TFs) are a special class of DNA-binding proteins that orchestrate gene transcription by recruiting other TFs, co-activators or co-repressors. Their combinatorial interplay in higher organisms maintains homeostasis and governs cell identity by finely controlling and regulating tissue-specific gene expression. Despite the rich literature on the importance of cooperative TFs for deciphering the mechanisms of individual regulatory programs that control tissue specificity in several organisms such as human, mouse, or *Drosophila melanogaster*, to date, there is still need for a comprehensive study to detect specific TF cooperations in regulatory processes of cattle tissues. To address the needs of knowledge about specific combinatorial gene regulation in cattle tissues, we made use of three publicly available RNA-seq datasets and obtained tissue-specific gene (TSG) sets for ten tissues (heart, lung, liver, kidney, duodenum, muscle tissue, adipose tissue, colon, spleen and testis). By analyzing these TSG-sets, tissue-specific TF cooperations of each tissue have been identified. The results reveal that similar to the combinatorial regulatory events of model organisms, TFs change their partners depending on their biological functions in different tissues. Particularly with regard to preferential partner choice of the transcription factors *STAT3* and *NR2C2*, this phenomenon has been highlighted with their five different specific cooperation partners in multiple tissues. The information about cooperative TFs could be promising: i) to understand the molecular mechanisms of regulating processes; and ii) to extend the existing knowledge on the importance of single TFs in cattle tissues.

Introduction

Regulation mechanisms of gene expression are of fundamental importance for different cellular processes, for instance, tissue development, differentiation or adaption to changing

research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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

environmental conditions [1–3]. Today, it is well known that the precise and effective regulation of the transcriptional machinery in higher organisms is often achieved by the cooperation of transcription factors (TFs) [1, 2, 4]. Such cooperative TFs frequently bind to the regulatory regions of the DNA (promoters as well as enhancers) in a cell-type specific manner to govern a large spectrum of biological processes, e.g. cell-cycle or homeostasis [1, 3, 5]. The identification of such cooperative TFs in higher organisms is important to distinguish common biological processes from individual regulatory programs that control tissue specificity [5] (for more details see the reviews [3, 6]).

In the last years, several groups have successfully studied tissue-specific combinatorial gene regulation based on the complex interplay between multiple TFs in different organisms [5, 7–15]. A small representative number of these studies is presented in Table 1. In addition, a variety of databases like TransCompel [16], BioGRID [17], STRING [18], or TRRUST [19] have been created to store both experimentally verified and computationally predicted cooperativity of TFs, as well as proteins in general.

Despite the rich literature on tissue-specific cooperations of TFs in different organisms, as mentioned in Table 1, their importance in the gene regulatory mechanisms of underlying biological processes in the cattle genome has not yet been extensively studied. Until now, only few research groups have investigated the crucial role of (single) TFs in the cattle genome. For this purpose, Lim et al. [23] analyzed the promoters of differentially expressed genes of the Korean cattle breed Hanwoo and determined significant tissue-specific TFs for fat-, muscle-, and liver-tissues. Moreover, Bickhart et al. [24] performed a large-scale genome-wide analysis to predict 379333 transcription factor binding sites (TFBSs) and their associations with known SNPs by considering the promoters of 7764 annotated genes in the cattle genome. Recently, Weber et al. [25] examined feed conversion in Angus by analyzing RNA-seq data of metabolism related tissues (pituitary, visceral adipose, duodenum, liver and skeletal muscle). Applying partial correlation and information theory (PCIT) based methods, they constructed coexpression

Table 1. Representative studies for the tissue-specific combinatorial gene regulation based on TF cooperations.

Authors	Synopsis of study	Type of data
Ament et al. [7]	Modeling of transcriptional network controlling mouse and human striatum as well as exploring the role of 48 TF-TF interactions in mouse models of Huntington’s disease	RNA-seq and microarray gene expression data
Sonawane et al. [5]	Investigation of cooperative TFs in regulatory networks for 38 human tissues	RNA-seq data from the Genotype-Tissue Expression project
Zeidler et al. [8]	Exploration of interacting TFs to understand the gene regulatory mechanisms during heart development	RNA-seq time series dataset including five time points
Song et al. [9]	Understanding and explanation of the role of 21 environmental stress related TF and their cooperativeness in the comprehensive regulatory network of <i>Arabidopsis thaliana</i>	Chip-seq and RNA-seq data
Rhee et al. [10]	Genome-wide analysis performed for <i>Drosophila melanogaster</i> in order to determine crucial biological functions of TF cooperations in tissue specification	RNA-seq data of 29 tissues and developmental time points from the modENCODE project
Nandi et al. [11]	Modeling of non-random functional dimers between the transcription factor MyoD and some muscle specific factors in the promoters of human genes	Human promoter sequences from the DBTSSs [20]
Laresgoit et al. [12]	Explanation of the essential role of the cooperation between transcription factors E2F2 and CREB for the regulation of transcriptional activity of cell cycle genes in mice	Data from CHIP-chip experiments
Myšíčková et al. [13]	Systematic large-scale analysis for the characterization of tissue-specific TF interactions of 22 human tissues	Expressed Sequence Tags (EST) data
Girgis et al. [21]	Prediction of cis-regulatory motifs in 72 human tissues and identification of related TFs	Expression data from GNF Atlas
Hu et al. [15]	Systematic large-scale analysis for the identification of tissue-specific TF interactions for 79 human tissues	Gene expression data from GNF Atlas2 gene expression database (gnfAtlas2) [22]
Yu et al. [14]	Systematic large-scale analysis for the characterization of tissue-specific combinatorial gene regulation based on TF interactions for 30 human tissues	Tissue-specific genes from NCBI EST database

<https://doi.org/10.1371/journal.pone.0216475.t001>

networks and determined the hub TFs, which act as important regulators in a tissue specific manner [25].

In order to address the limited knowledge available about crucial biological functions of tissue-specific TF cooperations in cattle, we analyzed in this study the promoter regions of tissue-specific genes (TSGs) of ten cattle tissues for the identification of their specific combinatorial gene regulation mechanisms. For this aim, using three publicly available RNA-seq datasets, we first identified a set of TSGs for each tissue according to their expression values and the significant TF cooperations for each tissue were determined using the PC-TraFF approach [1]. Subsequently, by applying the extension of the PC-TraFF approach (PC-TraFF⁺ [26]), the significant pairs have been assigned to two distinct groups as: i) TSG-set-specific TF cooperations; and ii) common (generally important) TF cooperations. As a result of our analysis, we obtained for each tissue a list of TSG-set-specific cooperative TF pairs which are likely to be fundamentally implicated in the regulation of transcriptional activity of a particular tissue. In the Result section of our study, we focused on these pairs and exemplarily explained their importance as well as potential roles in ten cattle tissues by providing further insight into the regulatory programs controlling specific biological processes such as tissue specificity or development.

Materials and methods

RNA-seq datasets

Datasets from three publicly available studies (described below) were selected for the identification of tissue-specific genes.

The first dataset has been published by Weber et al. [25] and consists of 16 samples from five cattle tissues (skeletal muscle, liver, visceral adipose tissue, pituitary and duodenum). The dataset contains of 17016 out of in total 24737 annotated cattle genes that have an expression value ≥ 0.2 (unit of expression: reads per kilobase of gene per million mapped reads, RPKM). In their study, Weber et al. described a gene g_i to be specific for the tissue t_j , ($j = 1, \dots, 5$) if the expression value of g_i in t_j is \geq two thirds of g_i 's accumulated expression values over all tissues. In total they identified 1026 genes as TSG: 285 genes for pituitary, 220 for skeletal muscle, 275 for liver, 33 for visceral adipose tissue, and 213 for duodenum.

In order to extend the number of tissues and increase the confidence in the TSGs, we included two further RNA-seq data sets available through the EBI Expression Atlas in our analysis [27]. The data from Merkin et al. (Accession: E-MTAB-2798) consists of 27 samples from nine tissues (brain, colon, heart, kidney, liver, lung, skeletal muscle, spleen and testis) and was part of a study on tissue-specific transcriptome variation across mammals [28]. The third dataset was created by Liao et al. (Accession: E-MTAB-2596) to examine duplicate genes in the cattle genome and their expression divergence. This dataset includes nine samples from seven cattle tissues (adipose, duodenum, hypothalamus, kidney, liver, lung, muscle) of beef cattle from Canada [29]. Merkin's and Liao's datasets contain expression values for 21414 and 20688 cattle genes, respectively, which are given as transcripts per million (TPM) values.

Data processing

To deal with bias of genes with small TPM values during the TSG selection process, data processing analysis was conducted individually for each dataset. For this purpose, we first analyzed the distribution of TPM-values using density plots and established an expression threshold value of $\tau = 1.46$, which indicates for datasets the local minimum of the density (see S1 Fig for the density plot of the TPM-values). Consequently, the genes with TPM values $\leq \tau$

in all tissues are removed from Merkin's and Liao's RNA-seq datasets. The same threshold τ is also used for the selection of expressed TFs in Results section.

Tissue-specific gene selection

Following the TSG description strategy of Weber et al., [25] we identified the TSGs in Merkin's and Liao's RNA-seq datasets. As a result, we obtained the TSG-sets for 13 unique cattle tissues based on all three RNA-seq datasets. However, due to the inconsistency between the RNA-seq datasets, we eliminated all TSGs for the brain tissues from our analysis, since each experiment examined different parts of the brain (pituitary in Weber's dataset, hypothalamus in Liao's dataset and unspecified brain tissue in Merkin's dataset). Finally, following the study of Gusev et al. [30], we determined the common TSGs of a tissue, if it was included in more than one experiment/dataset, to compensate the effects of different experimental conditions and the variation between individual animals under study.

The selection of TSG-sets for each tissue was performed as follows:

1. The TSG-set for liver contains all liver specific genes common in all three datasets.
2. The TSG-sets for lung and kidney contain their corresponding specific genes common between Merkin's and Liao's datasets.
3. The TSG-sets for duodenum, muscle-, and adipose-tissue are from Weber's dataset.
4. The TSG-sets for heart, colon, spleen and testis are from Merkin's dataset.

Identification of cooperative TFs

We applied the PC-TraFF [1] and PC-TraFF⁺ [26] methodologies to the above defined TSG-sets for the identification of tissue-specific TF cooperations. The theory behind the approaches used in this sections are detailed in [1] and [26].

PC-TraFF is an information theory method that applies pointwise mutual information (PMI) for the detection of cooperative TFs based on the co-occurrence of their binding sites in the regulatory regions of genes. Its algorithm consists of six steps and requires regulatory sequences, a list of position weight matrices (PWMs) and pre-defined distance preferences of TFBSs as input parameters:

- **Regulatory sequences:** using the UCSC genome browser [31], we extracted for each gene in the TSG-sets its corresponding promoter sequence covering the -500 to 100 bp regions relative to transcription start sites.
- **PWMs and TFBS detection:** we used non-redundant vertebrate PWMs from the TRANSFAC database (release 2018.1) [32] and employed the Match [33] program with these PWMs by setting its profile parameter "minSum: cut-offs for minimizing the sum of false positive and false negative rate" for prediction of potential TFBSs in promoter sequences. The selection process of PWMs for each tissue is explained in the Results section.
- **Pre-defined distances:** as a prerequisite, the PC-TraFF algorithm needs user-specified minimum and maximum distances which are necessary for the construction of TFBS pairs. As a result, two TFBSs are considered to be able to form a dimer if their distance preferences satisfy the user-specified distances. In this study, we use the recommended distance values and set the minimum distance ≥ 5 and maximum distance ≤ 20 .

Significant TF cooperations. The PC-TraFF algorithm provides for each TF-pair t_a and t_b a $PMI(t_a; t_b)$ -value based on the co-occurrence frequencies of their TFBSs in the sequences of

interest. In a final step, the algorithm transforms the $\text{PMI}(t_a; t_b)$ -values into z -scores and the cooperation between t_a and t_b is considered to be *statistically significant* if they have a z -score ≥ 3 .

TSG-set-specific TF cooperations. For the division of significant TF cooperations into the two categories of TSG-set-specific and common (generally important) cooperations, we further applied the PC-TraFF⁺ algorithm, which is an extension of the PC-TraFF approach. The main methodology of PC-TraFF⁺ is a heuristic approach that estimates the level of background cooperation ($\text{AVG}(\text{PMI}(t_a; t_b))$ -value) of any TF pair based on simulated sequence sets. Afterwards, the $\text{AVG}(\text{PMI})$ -values are subtracted from the initial significant PMI-values:

$$\text{PMI}^{\text{specific}}(t_a; t_b) = \text{PMI}(t_a; t_b) - [(1 + \alpha) \times \text{AVG}(\text{PMI}(t_a; t_b))], \quad (1)$$

where $\alpha \in [-1, +1]$ is a preassigned scaling factor which is used to control the influence of the background level. If $\alpha = -1$, there will be no differentiation between the TSG-set-specific and common (generally important) TF cooperations. On the other hand, setting $\alpha \geq 0$ results in the enlargement of subtracted background level which will lead to a more strict separation process. Consequently, a positive $\text{PMI}^{\text{specific}}$ -value indicates the TSG-set-specific cooperation of a TF pair, whereas a $\text{PMI}^{\text{specific}}$ -value ≤ 0 refers to a common cooperation of corresponding TFs that could play a generally important role in regulatory programs.

In order to determine the most convenient/suitable value of α in our analysis, we followed the recommendation of the PC-TraFF⁺ approach and systematically tested different α -values for the assessment of their influence on the ratio between tissue-specific and common (generally important) TF cooperations. As shown in Fig 1, the impact of α itself is not linear and highly tissue (TSG-set) dependent. Going through different α -values, we established that setting $\alpha > 0.5$ results in a dramatic decrease in the number of identified specific pairs of multiple tissues (see Fig 1). For this reason, we set $\alpha = 0.5$ for our analysis.

It is important to note that the key concept of the PC-TraFF⁺ approach provides for those significant TF pairs, which are very sensitive to the nucleotide composition and the position of their binding sites in the sequences of interest, remarkably small $\text{AVG}(\text{PMI})$ -values and thus positive $\text{PMI}^{\text{specific}}$ -values. As a result, a significant TF pair is defined to be specific if its corresponding $\text{PMI}^{\text{specific}}$ -value is > 0 for the dataset under study.

Results

In this study, to identify tissue-specific transcription factor cooperations, we analyzed tissue-specific genes (TSGs) of ten cattle tissues by employing PC-TraFF and its extension (PC-TraFF⁺) [1, 26]. For this purpose, we first collected publicly available RNA-seq datasets and we defined sets of tissue-specific genes following the TSG description of Weber et al. An overview of the TSGs is given in Table 2 (for a list of TSGs see S1 Table). In the next step the PC-TraFF algorithm was applied to the promoter sequences of the TSG-sets to identify significant TF cooperations. However, due to common regulatory programs between tissues as well as the different properties of promoter sequences like their GC content, dinucleotide occurrence or mononucleotide composition—referring the order and frequencies of individual nucleotides in sequences [26]—, some TF pairs have been ubiquitously determined as significant for multiple tissues. In order to eliminate negligible pairs (ubiquitously appearing, generally important ones) and to emphasize the roles of tissue-specific TF pairs, we applied the extension approach PC-TraFF⁺ [26]. Consequently, we determined tissue specific TF-cooperations and focused on these TF pairs in our further analysis to understand the molecular mechanisms of tissue-specific regulatory processes. The overall analysis procedures are outlined in Fig 2.

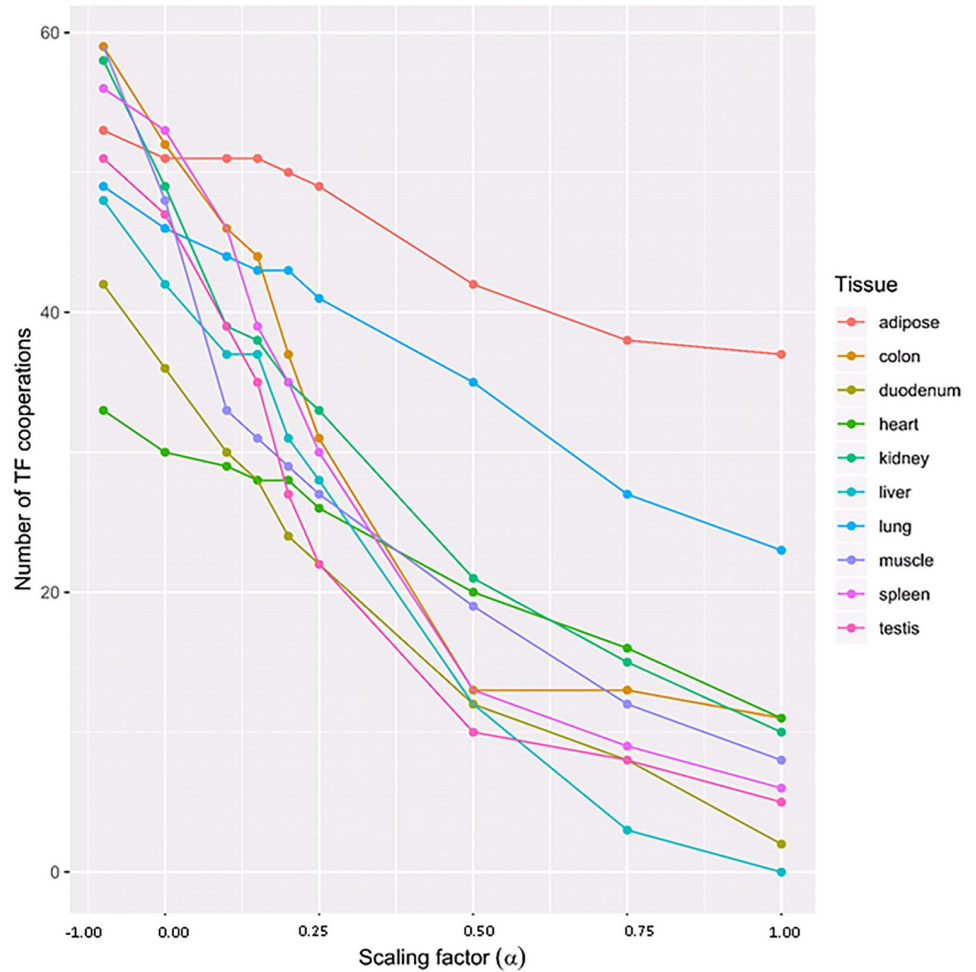


Fig 1. Number of tissue-specific TF cooperations identified by the PC-TraFF+ algorithm with different α -values. The subtracted background grows with α , thus reducing the number of specific cooperations.

<https://doi.org/10.1371/journal.pone.0216475.g001>

Selection of expressed TFs in tissues

As a prerequisite of the PC-TraFF approach, a library of position weight matrices (PWMs) is required to predict the putative binding sites of TFs in the promoter sequences. To fulfill this criterion, we first determined for the cattle TFs from AnimalTFDB 2.0 [34] the expression

Table 2. Numbers of TFs and tissue specific genes under study.

Tissues	Number of TSGs	Number of TFs
Heart	58	397
Lung	104	394
Liver	153	312
Kidney	163	395
Duodenum	213	285
Muscle tissue	220	407
Adipose tissue	33	383
Colon	213	369
Spleen	215	343
Testis	1958	297

<https://doi.org/10.1371/journal.pone.0216475.t002>

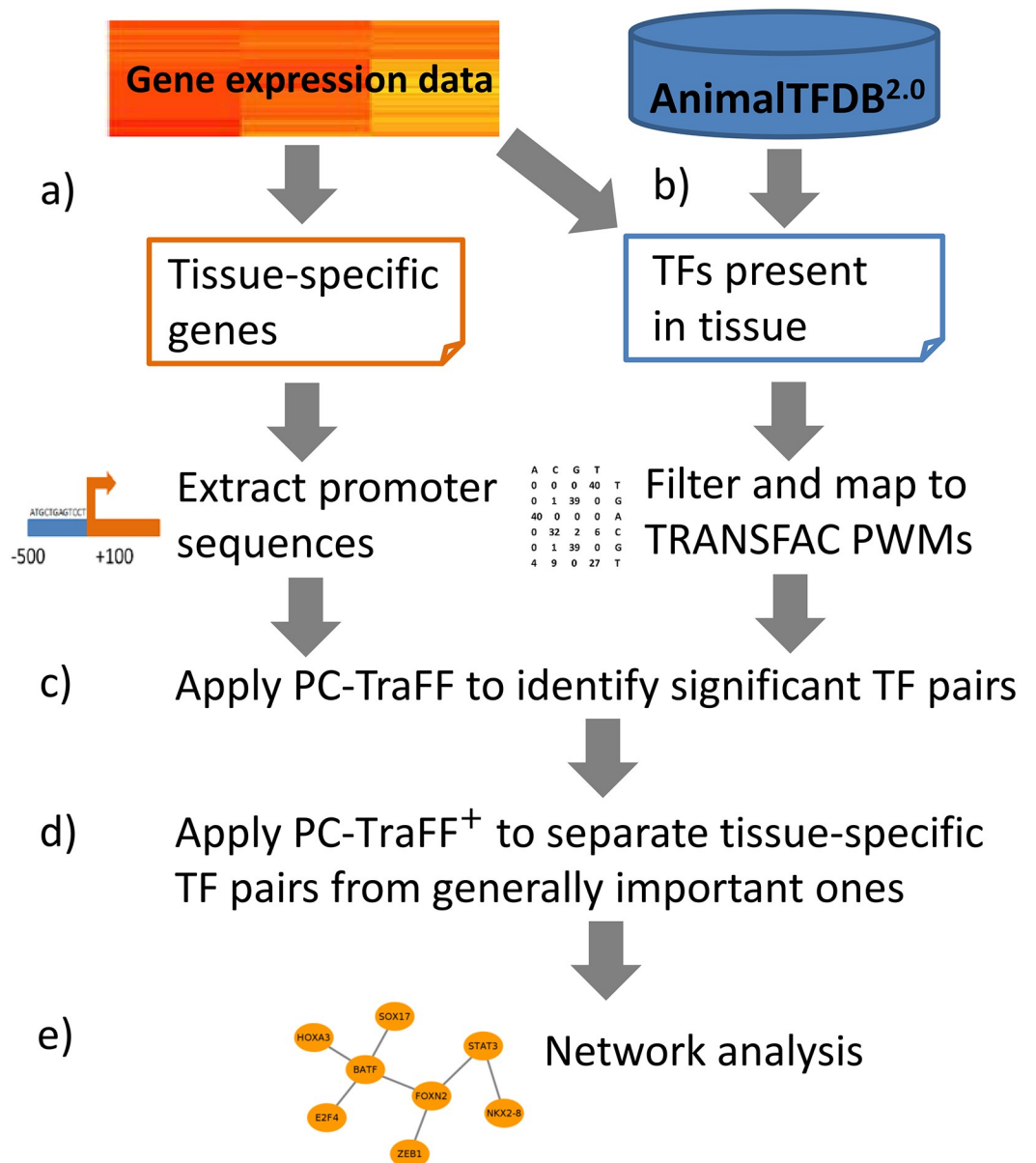


Fig 2. Flowchart of analysis procedures. (a) Identification of tissue-specific genes from RNA-seq data and extraction of promoter region of genes. (b) Identification of TFs expressed for each tissue in RNA-seq data. (c) Application of PC-TraFF [1]. (d) Application of PC-TraFF+ [26]. (e) Reconstruction of tissue-specific TF-TF cooperation networks.

<https://doi.org/10.1371/journal.pone.0216475.g002>

values (transcript per million (TPM) values) of the respective TF genes in the RNA-seq data of each tissue. Subsequently, using the threshold $\tau = 1.46$ established in Material and Methods section, we discarded all TFs with an expression value of smaller than τ . The remaining TFs were mapped to PWMs stored in the TRANSFAC database [32]. In order to avoid redundancies, Pearson correlations between these PWMs were computed, the PWMs were clustered with hierarchical clustering based on their correlation coefficients and the PWM with the highest information content in each cluster was chosen as representative. An overview of the number of TFs of interest is given in Table 2.

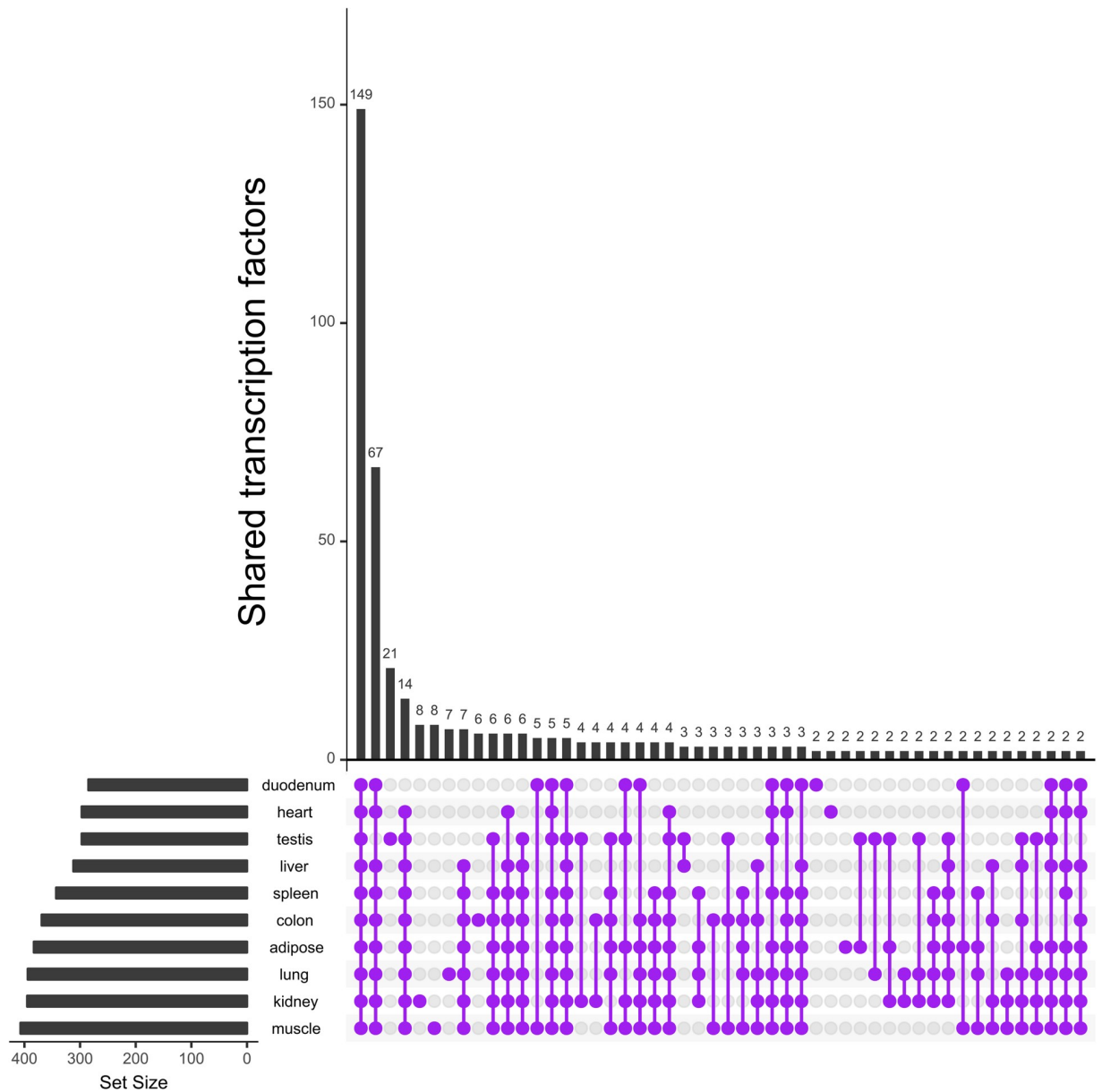


Fig 3. Occurrence of TFs present in the tissues. Number of TFs with an expression value $\geq \tau$ and their overlap between ten tissues represented in matrix layouts using the UpSet technique [35]. Purple circles in the matrix layout are related to the tissues that are part of the intersection. For the sake of clarity not all intersections are displayed.

<https://doi.org/10.1371/journal.pone.0216475.g003>

To gain more insight into the TFs of interest, we evaluated their overlap between tissues. Fig 3 shows that the majority of TFs are present in all ten tissues, and only a few are unique to each tissue.

Identification of tissue specific TF-cooperations

Applying the PC-TraFF algorithm to promoter sequences of TSGs, we identified for each tissue significant cooperative TF pairs. However, due to the common regulatory processes carried out by several tissues, the significant TF pairs are partially overlapping. In order to

Table 3. Numbers of cooperative TF pairs identified for each tissue as significant by PC-TraFF and TSG-set specific by PC-TraFF⁺.

Tissue	Number of cooperative TF pairs	
	Significant pairs	TSG-set-specific pairs
Heart	36	22
Lung	49	35
Liver	48	12
Kidney	62	25
Duodenum	50	13
Muscle tissue	63	21
Adipose tissue	59	48
Colon	68	15
Spleen	56	13
Testis	44	9

<https://doi.org/10.1371/journal.pone.0216475.t003>

separate TSG-set-specific TF cooperations from the common ones, we additionally employed the PC-TraFF⁺ [26] approach and eliminated 322 negligible pairs from our analysis. The numbers of significant and TSG-set-specific TF pairs are given in Table 3 (the whole list of pairs for each tissue can be found in S2 Table).

The analysis of TSG-set-specific TF pairs reveals that although the overlap of single TFs in multiple tissues is high (see Fig 3), the intersection of specific TF pairs is remarkably small (see Fig 4). In total, 213 out of 535 cooperative pairs have been assigned as tissue specific by PC-TraFF⁺. Further, 123 of these specific TF pairs have been determined as TSG-set-specific for a certain tissue and the remaining pairs have been determined as specific for at least two tissues. Interestingly, as shown in Fig 4, among all tissues only two combinations share more than one TF pair as TSG-set-specific. In particular, the transcription factor pairs *NR2C2*—*SP1* and *SP2*—*TEAD2* were found to be specific in colon, muscle, and kidney, while the TF pairs *BCL6B*—*SMAD5* and *GTF2I*—*STAT3* have been identified as specific for muscle and adipose, respectively. Among these transcription factors, *BCL6B* is a transcriptional repressor, which plays a role in immune responses of T-cells [36]. Its cooperation partner *SMAD5* is a family member of the *SMAD* factors [37] that plays a role in the transforming growth factor- β (*TGF- β*) signaling pathway, however they are also known to affect various other regulation mechanisms such as pathways controlling cell-cell adhesion [38]. Further, *SMAD5* is an important regulator in pathways controlling muscle mass, as it transmits the signal of the growth factor Bone morphogenetic protein (*BMP*) [39]. Additionally, *SMAD* factors are reported to affect adipocyte differentiation and metabolism [40].

The factor signal transducer and activator of transcription 3 (*STAT3*) is a family member of the *STAT* factors [37] and is involved in immune signaling, e.g. cytokine signaling [41]. In cattle *STAT3* is known to mediate signals during inflammatory response and to be relevant for fertility and embryo survival rate [42, 43]. It is activated by phosphorylation through Janus-kinases and potentially participates in adipogenesis and body weight homeostasis [44, 45]. Its cooperation partner *GTF2I* was at first identified to bind to the initiator element (*Inr*) of a multitude of promoters. Nowadays, the factor is known to be important for transcriptional initiation of *TATA*-less promoters and besides this general task, it is involved in cell-type specific regulatory processes of gene expression. The interplay between the *GTF2I* and *STAT* factors has been identified for the regulation of the *c-fos* promoter where the binding of *STAT* factors (*STAT1* or *STAT3*) is required to achieve the maximal activity of the promoter bound *GTF2I* factor. Further, *STAT3* and *GTF2I* have been shown to dimerize *in vivo* indicating that both

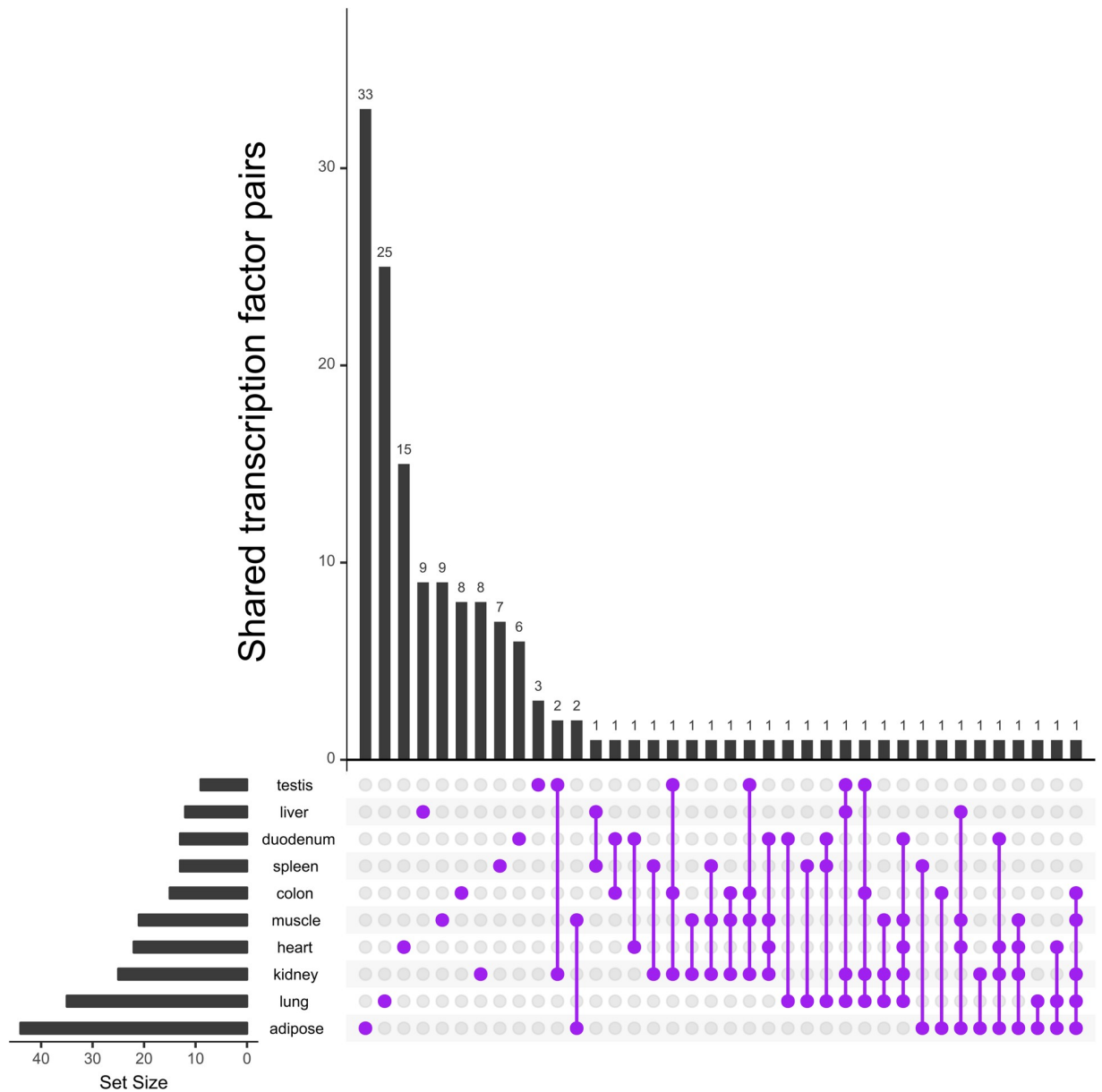


Fig 4. Occurrence of TSG-set-specific TF cooperations identified by PC-TraFF+ approach in ten tissues. Number of TF cooperations and their overlap between tissues represented in matrix layouts using the UpSet technique [35]. Lines with purple circles in the matrix layout show the tissues with overlapping TF cooperations. For the sake of clarity not all intersections are displayed.

<https://doi.org/10.1371/journal.pone.0216475.g004>

factors bind next to each other on DNA and thereby physically interact with each other in order to activate *c-fos* gene expression [46, 47].

The top three TSG-set-specific TF pairs for each tissue are presented in Table 4. Several TFs participate in different pairs across the different tissues and form specific cooperative dimers according to their biological functions. In particular, the transcription factors *NR2C2* and *STAT3* form cooperative pairs with different partners across the five different tissues. The factor *NR2C2* is a member of the *RXR*-related receptors family [37] and can act as a repressor or activator of various signaling pathways, specifically those of other TFs of the nuclear receptors with C4 zinc fingers class [48]. *NR2C2* preferentially binds to target sites in open chromatin

Table 4. Top three TSG-set-specific cooperative TF pairs of each tissue.

Tissue	Top three specific TF pairs
Heart	[<i>BCL6B</i> — <i>NR2C2</i>]; [<i>NKX2 - 5</i> — <i>HMGA1</i>]; [<i>FOXP2</i> — <i>TCF12</i>]
Lung	[<i>SP3</i> — <i>NR2C2</i>]; [<i>SMAD5</i> — <i>TEAD2</i>]; [<i>BCL6B</i> — <i>NR2C2</i>]
Liver	[<i>SMARCC2</i> — <i>STAT3</i>]; [<i>ZBTB7A</i> — <i>SP2</i>]; [<i>HOXB3</i> — <i>E2F4</i>]
Kidney	[<i>KLF4</i> — <i>NR2C2</i>]; [<i>SP1</i> — <i>NR2C2</i>]; [<i>SP3</i> — <i>NR2C2</i>]
Duodenum	[<i>SMAD1</i> — <i>HMGA1</i>]; [<i>KLF4</i> — <i>NR2C2</i>]; [<i>MYF6</i> — <i>NR2C2</i>]
Muscle	[<i>TCF3</i> — <i>KLF12</i>]; [<i>SP2</i> — <i>TEAD2</i>]; [<i>KLF4</i> — <i>NR2C2</i>]
Adipose	[<i>TCF12</i> — <i>STAT3</i>]; [<i>GTF2I</i> — <i>STAT3</i>]; [<i>KLF12</i> — <i>TGIF2</i>]
Colon	[<i>HAND1</i> — <i>STAT3</i>]; [<i>TCF3</i> — <i>STAT3</i>]; [<i>SMAD5</i> — <i>CTCF</i>]
Spleen	[<i>ZBTB7A</i> — <i>TEAD2</i>]; [<i>NKX2 - 5</i> — <i>IRX6</i>]; [<i>HMGA1</i> — <i>BATF</i>]
Testis	[<i>KLF4</i> — <i>SP1</i>]; [<i>KLF4</i> — <i>KLF4</i>]; [<i>KLF4</i> — <i>TEAD2</i>]

The pairs are sorted in ascending order based on the their *z*-scores provided by PC-TraFF⁺.

<https://doi.org/10.1371/journal.pone.0216475.t004>

regions and is known to affect cell type specific genes as well as RNA metabolism and protein translation [49].

In addition, other crucial TFs with varying partners have been identified: *SP1*, *SP2* and *SP3* are family members of the three-zinc finger Krüppel-related factors [37] and cooperate with *NR2C2*, *ZBTB7A*, and *TEAD2* in several tissues (see Table 4). The factor *SP1* is a well-known TF, which activates genes with functions in housekeeping, tissue-specificity, and cell cycle-regulation [50]. *SP2* is reported to bind to T-cell promoters, which could indicate an immune response related function [51]. *SP3* resembles *SP1* closely in its structure, but can affect the transcriptional regulation of genes in different ways [52]. For example during the *PKR* gene activation, *SP3* can modulate expression mediated by interferons, while *SP1* is responsible for the basal transcription [53]. Both *SP1* and *NR2C2* are reported to bind RB binding protein 4 (*RBB4*), a chromatin remodeling factor, which might mediate this cooperation [54, 55].

Further analysis of individual TF pairs in Table 4 demonstrates that there is literature support for the interaction between some cooperative pairs although our prediction of TF cooperations does not relate to their direct physical interactions. For example, the cooperative TFs *SMAD5* and *CTCF* in colon are reported to bind to the same promoter regions to interact in transcriptional regulation in mammals and *Drosophila melanogaster*, where *CTCF* is important for the recruitment of *SMAD* factors [56, 57]. Another remarkable pair found in cattle heart tissue is *HMGA1*—*NKX2-5*. The homeodomain factor *NKX2-5* is a well-studied TF which plays essential roles in healthy heart development and disease [58, 59]. Being a member of high-mobility group A (*HMGA*) non-histone chromosomal proteins, *HMGA1* is an important TF during heart development and growth as well as cardiomyocytic cell growth regulation [8, 60, 61].

Analysis of tissue-specific TF cooperation networks

To further establish the potential role of TSG-set-specific TF pairs, we constructed, similar to our previous studies [1, 8, 26], a cooperation network for each tissue, in which the nodes depict the TFs and edges refer to their cooperativity (see Fig 5). On the one hand, these networks are essential to monitor the preferential partner choice of TFs in different tissues. On the other hand, they highlight the functional dimers or high order complexes of TFs as well as the hub TFs which could provide crucial information for explaining the underlying mechanisms of regulatory events.

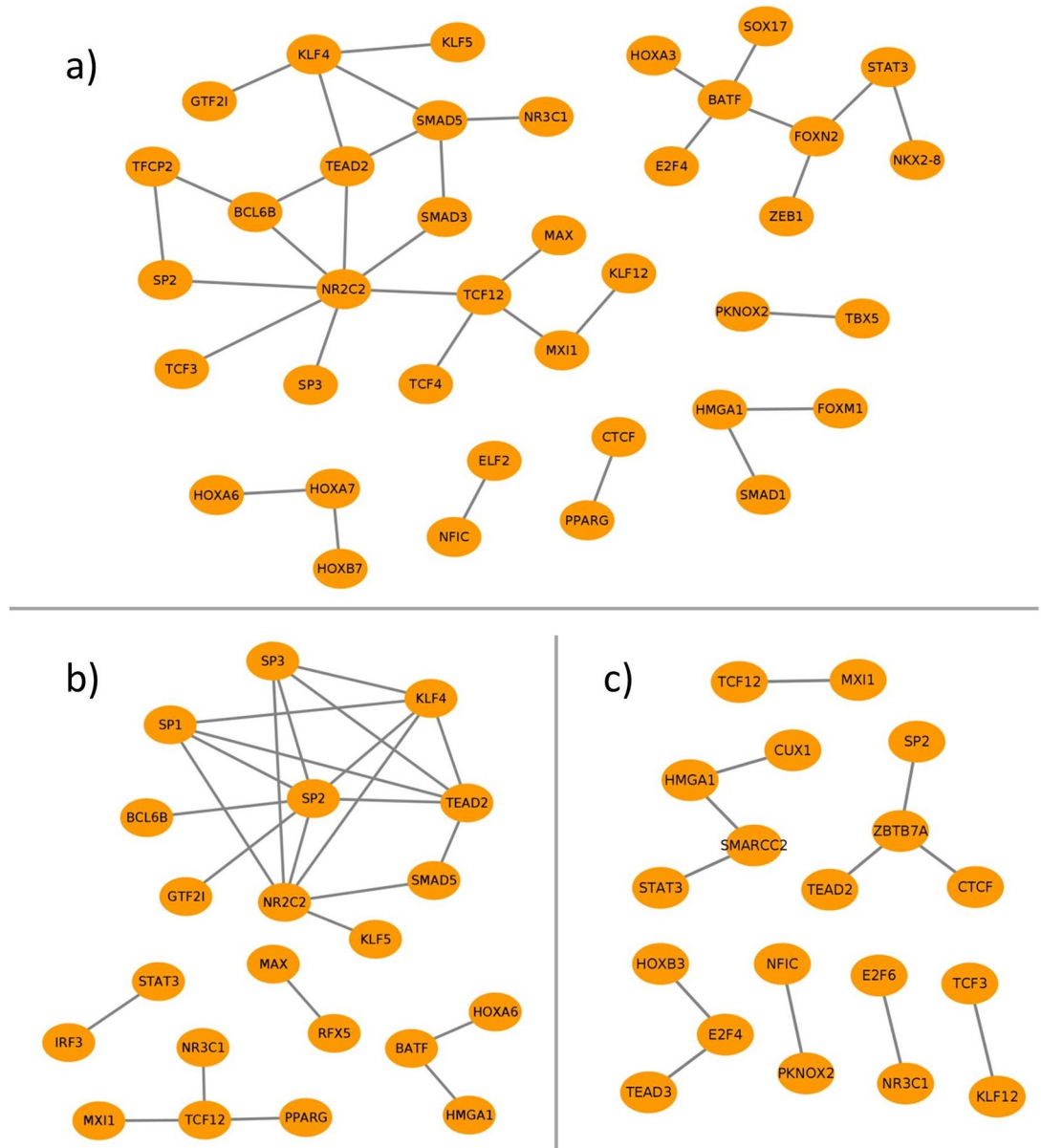


Fig 5. Cooperation networks for the TSG-set-specific TF pairs of (a) lung-, (b) kidney- and (c) liver-tissue.

<https://doi.org/10.1371/journal.pone.0216475.g005>

It is important to note that in the following analysis we have mainly focused on the TF cooperation networks of three cattle tissues, namely lung, kidney and liver, because of the higher confidence in their TSGs (see [Materials and methods](#)). The cooperation networks of the remaining seven tissues can be found in [S2 Fig](#).

[Fig 5](#) presents the cooperation networks of lung-, kidney-, and liver tissues based on their specific-TF pairs, which consist of 35, 25 and 12 cooperative pairs for each tissue, respectively. In comparison to liver and kidney, the network of lung tissue consists of more pairs, although the number of lung specific genes is smaller than those of kidney and liver. This result is in line with our previous studies [[1](#), [8](#), [26](#)] and indicates that the number of TSG-set-specific TF pairs

depends neither on the number of TSGs nor on the number of TFs under study (for the numbers see [Table 2](#)).

A general analysis of the networks indicates that the transcription factor *NR2C2* forms a hub in the networks of lung- and kidney specific-TF pairs, with a degree of seven and four respectively ([Fig 5a and 5b](#)). Among its partners, only its cooperations with *SP2* and *SP3* have been simultaneously identified as specific for both tissues. Additionally, an interesting cooperation is monitored between *NR2C2* and *SMAD* factor family members: while *NR2C2* cooperates with *SMAD3* in lung tissue, the pair *NR2C2-SMAD5* is determined as tissue-specific in kidney.

Other remarkable cooperations in the network of lung tissue are observed between the following TFs: *PPARG-CTCF*, *KLF4-SMAD5*, *KLF4-TEAD2*, and *SMAD5-TEAD2*. The factor peroxisome proliferator-activated receptor gamma (*PPARG*) is reported to be an important regulator in adipogenesis. The cooperation of *CTCF* with *PPARG* and its effects on the transcriptional activity of *PPARG* are well documented in [62, 63]. *SMAD* factors are known to form complexes with *KLF4* and to mediate its phosphorylation, which is necessary for the recruitment of other factors such as *PPARG* [64, 65]. In human stem cells transcriptional complexes, involving *TEADs* and *SMAD* factors that link the *TGF- β* and the hippo signaling pathway were described [66]. Additionally, both *KLF4* and *TEADs* are reported to be partners of yes-associated proteins (*YAPs*), which are important regulators in the hippo signaling pathway [66, 67]. This supports the hypothesis of a multi-factor complex that has a tissue specific role in the lung, potentially affecting cell proliferation or differentiation as the hippo and *TGF- β* regulatory pathways suggest.

A closer look at the cooperation network of kidney-specific TF pairs ([Fig 5b](#)) reveals that the largest hubs *NR2C2* and *SP2* form a cooperative dimer which also has been identified as specific in four other tissues (heart, lung, duodenum, and muscle) (see [S2 Fig](#)). *NR2C2* is a receptor of various hormones such as androgens and it is known to bind and orchestrate transcriptional regulation with the Hepatocyte nuclear factor 4 alpha (*HNF4A*), which plays an essential role in kidney cell proliferation and development [68, 69]. *SP2* could indicate an immune response related function [51]. *KLF4* was detected to cooperate with *NR2C2*, *SP1* and *SP2* and it belongs to the same family of three-zinc finger Krüppel-related factors as *SP1* and *SP2* [37]. The cooperation of *SP1* and *KLF4* is experimentally confirmed by a known protein-protein interaction (PPI) of these TFs, which cooperate when binding to the transforming growth factor *TGF- β* control element in promoters [65]. Another interesting hub in this network ([Fig 5b](#)) is *TCF12* that is cooperating with *PPARG*, *MAX* interactor 1 (*MXI1*) and the glucocorticoid receptor (*NR3C1*). *TCF12* in human is involved in the repression of E-cadherin [70], which plays essential roles in healthy animal tissue morphogenesis and development [71–74]. The role of *TCF12* as a hub in our network suggests participation in other regulation mechanisms as well, e.g. by its cooperation with *PPARG*, which is associated with insulin sensitivity and other metabolic pathways in kidney [75]. The cooperations of *TCF12* with *MXI1* as well as *MAX* were detected to be specific for cattle lung tissue as well. The factors *MAX* and *MYC* dimerize to act in the transcriptional regulation of cell proliferation and *MXI1* inhibits this mechanism by competing with *MYC* for *MAX* binding sites and additionally repressing *MYC* transcription [76, 77]. Through its cooperation with *MXI1* and *MAX*, *TCF12* might influence this important transcriptional repression mechanism in some cattle tissues.

The liver-specific TF pairs constitute the smallest cooperation network in this study, which includes 12 cooperative pairs in multiple unconnected subgroups (see [Fig 5c](#)). Having three different cooperations with *CTCF*, *TEAD2*, and *SP2*, the transcription factor *ZBTB7A* (zinc finger and BTB domain-containing protein 7A) forms the largest hub in this network. *ZBTB7A* is reported to bind to many promoter sequences in the human genome and it

cooperates with different TFs and proteins that influence DNA accessibility to modulate gene regulation [78]. Its cooperation partner *CTCF* interacts with the chromatin structure in various ways [79] and can bind to the co-repressor *SIN3A*, which plays a role in the recruitment of histone-deacetylases [80]. Interestingly, *SIN3A* can also bind to *ZBTB7A* in human, which might suggest a potential function of it as a co-factor in this cooperation [81].

Discussion

Today, it is widely known that the cooperation of TFs is crucial for the precise orchestration of tissue-specific genetic programs and/or transcriptional regulation in cells [1, 5]. Until now, several studies have shown that the TF partnership is accomplished through a non-random process which depends strongly on their specific roles in different biological processes as well as on the cellular context [1, 5, 6, 11, 14, 21, 26]. For instance, modeling gene regulatory networks of 38 human tissues, Sonawane et al. [5] recently pointed out the dimerization of TFs in the regulatory events of these tissues and recognized the specific partner alterations of TFs that contribute to tissue specificity by coordinating distinct regulatory processes. A comparable study has been performed by Rhee et al. [10] to address the crucial roles of TF cooperations in the transcriptional regulatory network of *Drosophila melanogaster* (for an overview see Table 1). To this end, Amoutzias et al. [6] discussed in their review that knowledge about cooperative TFs is helpful for understanding of different disease mechanisms as well as in drug development.

However, regarding cattle tissues only limited information about TF cooperativity is available. In this study, we addressed this need and performed a comprehensive study for ten cattle tissues which could aid researchers to create novel hypotheses for transcriptional regulation as well as provide comparability with model organisms. For this aim, we have analyzed three publicly available RNA-seq datasets and found that only three out of ten tissues have multiple TSG-sets with overlapping genes. To increase the quality and simultaneously eliminate the differences of their TSG-sets, we considered the intersection between the TSGs obtained for a certain tissue as suggested in [30].

For the identification of tissue-specific TF-cooperations, the PC-TraFF and PC-TraFF⁺ approaches have been applied to promoter sequences of TSGs. However, due to their underlying methodologies, the prediction performance of both algorithms is heavily influenced by the putative TFBSs predicted by using PWM libraries. Stormo et al. [82] as well as Whitfield et al. [4] have pointed out that computational TFBS predictions using PWMs is a very effective and established method, but it suffers from high rates of false positive predictions. To reduce this to some extent in our analysis, a fundamental step for the construction of tissue-specific PWM libraries is the inclusion of TF genes with expression values $\geq \tau$ in a tissue (see Results section). While the analysis of expressed TF genes provides crucial information about their comparability and presence in multiple tissues, the usage of their associated PWM libraries leads to the reduction of false positive TFBS predictions in the identification of tissue-specific TF cooperations.

Our results suggest that the consideration of single TFs, whose majority is present in all tissues under study, appears to be insufficient for the differentiation of common and specific regulatory programs (see Fig 3). Several studies show that in higher organisms TFs have to form cellular context-dependent cooperations rather than acting alone during, for instance, tissue differentiation and development [1, 5, 6]. Similar to previous studies based on model organisms, our results further show that TFs switch their partners to specify their biological functions depending on the tissue type, for example the transcription factor *STAT3* exhibits specific (different) cooperation partner preferences on the level of transcriptional regulation

of lung, kidney and liver (see Fig 5a, 5b and 5c). The partner choice of *STAT3* might explain its potentially specific functions in gene regulatory mechanisms controlling tissue specificity as well as development [41]. *Interferon regulatory factor 3 (IRF3)* was found to cooperate with *STAT3* in kidney. *IRF3* is a master regulator activating interferons or interferon stimulated genes, which activate Janus-kinases that subsequently phosphorylate *STAT* factors [83, 84]. *STAT* factors can bind to the response elements of interferon stimulated genes [85], which could indicate that *IRF3* and *STAT3* cooperate to regulate genes activating *STAT3*. In cattle liver, *STAT3* cooperates with *SMARCC2*, a member of the *SWI/SNF* chromatin remodeling complex that can bind to various mammalian promoters and could be important for the liver-specific role of *STAT3* [86]. Furthermore, *STAT3* collaborates in lung with *forkhead box N2 (FOXP2)* and *NK2 homeobox 8 (NKX2-8)*, which is a known lung developmental transcription factor that exhibits gene expression patterns related to *STAT* factors and Janus kinases pathways [87]. In general, *STAT3* demonstrates the advantage of the employed method: it is present in all tissues, but the identification of its cooperations is necessary to uncover its tissue specific roles. Another interesting TF in the cooperation networks is *NR2C2* which occurs in eight different tissues with multiple overlapping partners (see Fig 5a and 5b, and S2 Fig for the remaining six tissues). Despite the overlap between its partners, the *NR2C2*-dependent regulation processes of these tissues are characterized by a unique set of its partner TFs. On the one hand, it forms dimers with the three-zinc finger Krüppel-related factors (*SP1*, *SP2*, *SP3* and *KLF4*, *KLF5*) in seven different tissues and with *SMAD* factors in five tissues. On the other hand, *NR2C2* forms exclusive dimers, for example with: i) *TCF12* and *TCF3* in lung; ii) *MAF* in adipose tissue; iii) *IRF3* in muscle tissue; iv) *MYF6* in duodenum; v) *MXI1* in spleen.

Notwithstanding the limited availability of tissue specific RNA-seq datasets in cattle compared to other organisms such as human or mouse, taken together, our results provide a comprehensive overview of the specific regulatory processes in different cattle tissues. This knowledge is required to better understand the regulatory mechanisms of biological processes during development, cell cycle or different diseases. In addition, focusing on the specific partner choice of TFs (e.g., *NR2C2* and *STAT3*) highlights the need for our analysis to identify cooperative TF pairs in order to amplify the findings on the importance of single TFs for cattle tissues reported in [23–25].

Conclusion

In this study, we aim to contribute to the understanding of tissue-specific combinatorial gene regulation mechanisms by addressing the limited knowledge available about crucial biological functions of tissue-specific TF cooperations in cattle. Specific functions of different tissues and their expressional regulation are largely dependent on the complex interplay between TFs. By forming dimers or high order complexes, TFs have to act with their non-random cooperation partners in higher organisms to ensure coordinated cellular processes in response to environmental stimuli as well as tissue-specificity. In order to explore complex interplay between TFs, we performed a comprehensive analysis using PC-TraFF and PC-TraFF⁺ approaches and identified tissue-specific TF cooperations in ten cattle tissues that are essential for directing the specific transcriptional program of tissues. The results show that similar to the combinatorial regulatory events of model organisms, TFs switch their partners depending on their biological functions in activation or repression of tissue specific genes. Furthermore, we highlighted the preferential partner choice of TFs in different tissues using cooperation networks which could aid researchers to get a better understanding for the underlying mechanisms of regulatory events as well as to generate new hypotheses regarding the molecular

mechanisms of regulating processes in cattle tissues. On the top of that, the knowledge about TF cooperation could complement previous studies which mainly focused on the effect of single TFs.

Supporting information

S1 Fig. Density of expression values.

(PDF)

S2 Fig. Tissue-specific TF cooperation networks of seven cattle tissues.

(PDF)

S1 Table. Lists of tissue-specific genes.

(CSV)

S2 Table. Lists of significant and specific TF cooperations per tissue.

(CSV)

Author Contributions

Conceptualization: Mehmet Gültas.

Data curation: Lukas Steuernagel, Mehmet Gültas.

Formal analysis: Lukas Steuernagel, Felix Heinrich, Armin O. Schmitt, Mehmet Gültas.

Investigation: Lukas Steuernagel, Sebastian Zeidler, Armin O. Schmitt, Mehmet Gültas.

Methodology: Cornelia Meckbach, Mehmet Gültas.

Project administration: Mehmet Gültas.

Software: Cornelia Meckbach, Felix Heinrich, Armin O. Schmitt.

Supervision: Mehmet Gültas.

Validation: Cornelia Meckbach, Felix Heinrich, Sebastian Zeidler, Armin O. Schmitt.

Visualization: Lukas Steuernagel, Sebastian Zeidler.

Writing – original draft: Lukas Steuernagel, Mehmet Gültas.

Writing – review & editing: Mehmet Gültas.

References

1. Meckbach C, Tacke R, Hua X, Waack S, Wingender E, Gültas M. PC-TraFF: identification of potentially collaborating transcription factors using pointwise mutual information. *BMC Bioinformatics*. 2015; 16(400). <https://doi.org/10.1186/s12859-015-0827-2> PMID: 26627005
2. MacNeil LT, Pons C, Arda HE, Giese GE, Myers CL, Walhout AJ. Transcription factor activity mapping of a tissue-specific in vivo gene regulatory network. *Cell systems*. 2015; 1(2):152–162. <https://doi.org/10.1016/j.cels.2015.08.003> PMID: 26430702
3. de Souza N. Transcription factor interaction maps. *Nature methods*. 2010; 7(5):344. <https://doi.org/10.1038/nmeth0510-344b> PMID: 20440881
4. Whitfield TW, Wang J, Collins PJ, Partridge EC, Aldred SF, Trinklein ND, et al. Functional analysis of transcription factor binding sites in human promoters. *Genome Biology*. 2012 Sep; 13(9):R50. Available from: <https://doi.org/10.1186/gb-2012-13-9-r50> PMID: 22951020
5. Sonawane AR, Platig J, Fagny M, Chen CY, Paulson JN, Lopes-Ramos CM, et al. Understanding tissue-specific gene regulation. *Cell reports*. 2017; 21(4):1077–1088. <https://doi.org/10.1016/j.celrep.2017.10.001> PMID: 29069589

6. Amoutzias GD, Robertson DL, de Peer YV, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. *Trends in Biochemical Sciences*. 2008 may; 33(5):220–229. Available from: <https://doi.org/10.1016/j.tibs.2008.02.002> PMID: 18406148
7. Ament SA, Pearl JR, Cantle JP, Bragg RM, Skene PJ, Coffey SR, et al. Transcriptional regulatory networks underlying gene expression changes in Huntington's disease. *Molecular systems biology*. 2018; 14(3):e7435. <https://doi.org/10.15252/msb.20167435> PMID: 29581148
8. Zeidler S, Meckbach C, Tacke R, Raad FS, Roa A, Uchida S, et al. Computational Detection of Stage-Specific Transcription Factor Clusters during Heart Development. *Front Genet*. 2016; 7(33). <https://doi.org/10.3389/fgene.2016.00033> PMID: 27047536
9. Song L, Huang SSC, Wise A, Castanon R, Nery JR, Chen H, et al. A transcription factor hierarchy defines an environmental stress response network. *Science*. 2016; 354 (6312). Available from: <http://science.sciencemag.org/content/354/6312/aag1550>.
10. Rhee DY, Cho DY, Zhai B, Slattery M, Ma L, Mintseris J, et al. Transcription factor networks in *Drosophila melanogaster*. *Cell reports*. 2014; 8(6):2031–2043. <https://doi.org/10.1016/j.celrep.2014.08.038> PMID: 25242320
11. Nandi S, Blais A, Ioshikhes I. Identification of cis-regulatory modules in promoters of human genes exploiting mutual positioning of transcription factors. *Nucleic acids research*. 2013; 41(19):8822–8841. <https://doi.org/10.1093/nar/gkt578> PMID: 23913413
12. Laresgoiti U, Apraiz A, Olea M, Mitxelena J, Osinalde N, Rodriguez JA, et al. E2F2 and CREB cooperatively regulate transcriptional activity of cell cycle genes. *Nucleic acids research*. 2013; 41(22):10185–10198. <https://doi.org/10.1093/nar/gkt821> PMID: 24038359
13. Myšičková A, Vingron M. Detection of interacting transcription factors in human tissues using predicted DNA binding affinity. *BMC Genomics*. 2012 Jan; 13(1):S2. Available from: <https://doi.org/10.1186/1471-2164-13-S1-S2> PMID: 22369666
14. Yu X, Lin J, Zack DJ, Qian J. Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. *Nucleic Acids Research*. 2006; 34(17):4925–4936. Available from: <http://dx.doi.org/10.1093/nar/gkl595> PMID: 16982645
15. Hu Z, Gallo SM. Identification of interacting transcription factors regulating tissue gene expression in human. *BMC genomics*. 2010; 11(1):49. <https://doi.org/10.1186/1471-2164-11-49> PMID: 20085649
16. Kel-Margoulis O, Kel A, Reuter I, Deineko I, Wingender E. TRANSCOMP: a database on composite regulatory elements in eukaryotic genes. *Nucleic Acids Res*. 2002; 30:332–334. <https://doi.org/10.1093/nar/30.1.332> PMID: 11752329
17. Chatr-aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. *Nucleic Acids Research*. 2014; Available from: <http://nar.oxfordjournals.org/content/early/2014/11/26/nar.gku1204.abstract> PMID: 25428363
18. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*. 2014; 43(D1):D447–D452. <https://doi.org/10.1093/nar/gku1003> PMID: 25352553
19. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic acids research*. 2017; 46(D1):D380–D386. <https://doi.org/10.1093/nar/gkx1013>
20. Wakaguri H, Yamashita R, Suzuki Y, Sugano S, Nakai K. DBTSS: database of transcription start sites, progress report 2008. *Nucleic acids research*. 2007; 36(suppl_1):D97–D101. <https://doi.org/10.1093/nar/gkm901> PMID: 17942421
21. Girgis HZ, Ovcharenko I. Predicting tissue specific cis-regulatory modules in the human genome using pairs of co-occurring motifs. *BMC Bioinformatics*. 2012; 13(1):25. Available from: <https://doi.org/10.1186/1471-2105-13-25> PMID: 22313678
22. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, et al. Large-scale analysis of the human and mouse transcriptomes. *Proceedings of the National Academy of Sciences*. 2002; 99(7):4465–4470. Available from: <http://www.pnas.org/content/99/7/4465>.
23. Lim D, Lee SH, Cho YM, Yoon D, Shin Y, Kim KW, et al. Transcript profiling of expressed sequence tags from intramuscular fat, longissimus dorsi muscle and liver in Korean cattle (Hanwoo). *BMB Rep*. 2010; 43:151–121. <https://doi.org/10.5483/BMBRep.2010.43.2.115>
24. Bickhart DM, Liu GE. Identification of candidate transcription factor binding sites in the cattle genome. *Genomics, proteomics & bioinformatics*. 2013; 11(3):195–198. <https://doi.org/10.1016/j.gpb.2012.10.004>
25. Weber KL, Welly BT, Eenennaam ALV, Young AE, Porto-Neto LR, Reverter A, et al. Identification of Gene Networks for Residual Feed Intake in Angus Cattle Using Genomic Prediction and RNA-seq. *PLOS ONE*. 2016 mar; 11(3):e0152274. Available from: <https://doi.org/10.1371/journal.pone.0152274> PMID: 27019286

26. Meckbach C, Wingender E, Gültas M. Removing Background Co-occurrences of Transcription Factor Binding Sites Greatly Improves the Prediction of Specific Transcription Factor Cooperations. *Frontiers in Genetics*. 2018; 9:189. Available from: <https://www.frontiersin.org/article/10.3389/fgene.2018.00189> PMID: 29896218
27. Papatheodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, et al. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Research*. 2017 nov; 46 (D1):D246–D251. Available from: <https://doi.org/10.1093/nar/gkx1158>.
28. Merkin J, Russell C, Chen P, Burge CB. Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. *Science*. 2012 dec; 338(6114):1593–1599. Available from: <https://doi.org/10.1126/science.1228186> PMID: 23258891
29. Liao X, Bao H, Meng Y, Plastow G, Moore S, Stothard P. Sequence, Structural and Expression Divergence of Duplicate Genes in the Bovine Genome. *PLoS ONE*. 2014 jul; 9(7):e102868. Available from: <https://doi.org/10.1371/journal.pone.0102868> PMID: 25054921
30. Gusev Y, Riggins RB, Bhuvaneshwar K, Gauba R, Sheahan L, Clarke R, et al. In silico discovery of mitosis regulation networks associated with early distant metastases in estrogen receptor positive breast cancers. *Cancer informatics*. 2013; 12:CIN–S10329. <https://doi.org/10.4137/CIN.S10329> PMID: 23470717
31. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. *Nucleic acids research*. 2004; 32(suppl_1):D493–D496. <https://doi.org/10.1093/nar/gkh103> PMID: 14681465
32. Wingender E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Briefings in Bioinformatics*. 2008; 9(4):326–332. <https://doi.org/10.1093/bib/bbn016> PMID: 18436575
33. Kel AE, Gössling E, Reuter I, Chermushkin E, Kel-Margoulis OV, Wingender E. MATCH: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research*. 2003; 31 (13):3576–3579. <https://doi.org/10.1093/nar/gkg585> PMID: 12824369
34. Zhang HM, Liu T, Liu CJ, Song S, Zhang X, Liu W, et al. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Research*. 2014 sep; 43 (D1):D76–D81. Available from: <https://doi.org/10.1093/nar/gku887> PMID: 25262351
35. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 2017; 33(18):2938–2940. Available from: <http://dx.doi.org/10.1093/bioinformatics/btx364> PMID: 28645171
36. Manders PM, Hunter PJ, Telaranta AI, Carr JM, Marshall JL, Carrasco M, et al. BCL6b mediates the enhanced magnitude of the secondary response of memory CD8+ T lymphocytes. *Proceedings of the National Academy of Sciences*. 2005 apr; 102(21):7418–7425. Available from: <https://doi.org/10.1073/pnas.0501585102>.
37. Wingender E, Schoeps T, Haubrock M, Krull M, Dönitz J. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Research*. 2018; 46(D1): D343–D347. Available from: <http://dx.doi.org/10.1093/nar/gkx987> PMID: 29087517
38. Hoover LL, Kubalak SW. Holding Their Own: The Noncanonical Roles of Smad Proteins. *Science Signaling*. 2008 nov; 1(46):pe48–pe48. Available from: <https://doi.org/10.1126/scisignal.146pe48> PMID: 19018011
39. Sartori R, Schirwis E, Blaauw B, Bortolanza S, Zhao J, Enzo E, et al. BMP signaling controls muscle mass. *Nature Genetics*. 2013 sep; 45(11):1309–1318. Available from: <https://doi.org/10.1038/ng.2772> PMID: 24076600
40. Choy L, Skillington J, Derynck R. Roles of Autocrine TGF- β Receptor and Smad Signaling in Adipocyte Differentiation. *The Journal of Cell Biology*. 2000 may; 149(3):667–682. Available from: <https://doi.org/10.1083/jcb.149.3.667> PMID: 10791980
41. Levy DE, Lee Ck. What does Stat3 do? *J Clin Invest*. 2002 May; 109(9):1143–1148. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC150972/> PMID: 11994402
42. Cronin JG, Kanamarlapudi V, Thornton CA, Sheldon IM. Signal transducer and activator of transcription-3 licenses Toll-like receptor 4-dependent interleukin (IL)-6 and IL-8 production via IL-6 receptor-positive feedback in endometrial cells. *Mucosal Immunology*. 2016 jan; 9(5):1125–1136. Available from: <https://doi.org/10.1038/mi.2015.131> PMID: 26813342
43. Khatib H, Huang W, Mikheil D, Schutzkus V, Monson RL. Effects of signal transducer and activator of transcription (STAT) genes STAT1 and STAT3 genotypic combinations on fertilization and embryonic survival rates in Holstein cattle. *Journal of Dairy Science*. 2009 dec; 92(12):6186–6191. Available from: <https://doi.org/10.3168/jds.2009-2439> PMID: 19923622
44. Zhao P, Stephens JM. Identification of STAT target genes in adipocytes. *JAK-STAT*. 2013 apr; 2(2): e23092. Available from: <https://doi.org/10.4161/jkst.23092> PMID: 24058802

45. Richard AJ, Stephens JM. The role of JAK—STAT signaling in adipose tissue function. *Biochimica et Biophysica Acta (BBA)—Molecular Basis of Disease*. 2014 mar; 1842(3):431–439. Available from: <https://doi.org/10.1016/j.bbadis.2013.05.030>.
46. Kim DW, Cheryath V, Roy AL, Cochran BH. TFII-I Enhances Activation of the c-fos Promoter through Interactions with Upstream Elements. *Mol Cell Biol*. 1998 Jun; 18(6):3310–3320. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC108912/> PMID: 9584171
47. Roy AL. Biochemistry and biology of the inducible multifunctional transcription factor TFII-I: 10 years later. *Gene*. 2012 jan; 492(1):32–41. Available from: <https://doi.org/10.1016/j.gene.2011.10.030> PMID: 22037610
48. Huq MDM, Gupta P, Tsai NP, Wei LN. Modulation of Testicular Receptor 4 Activity by Mitogen-activated Protein Kinase-mediated Phosphorylation. *Molecular & Cellular Proteomics*. 2006 aug; 5(11):2072–2082. Available from: <https://doi.org/10.1074/mcp.m600180-mcp200>.
49. O'Geen H, Lin YH, Xu X, Echipare L, Komashko VM, He D, et al. Genome-wide binding of the orphan nuclear receptor TR4 suggests its general role in fundamental biological processes. *BMC Genomics*. 2010; 11(1):689. Available from: <https://doi.org/10.1186/1471-2164-11-689> PMID: 21126370
50. Suske G. The Sp-family of transcription factors. *Gene*. 1999 oct; 238(2):291–300. Available from: [https://doi.org/10.1016/s0378-1119\(99\)00357-1](https://doi.org/10.1016/s0378-1119(99)00357-1) PMID: 10570957
51. Kingsley C, Winoto A. Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. *Mol Cell Biol*. 1992 Oct; 12(10):4251–4261. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC360348/> PMID: 1341900
52. Li L, He S, Sun JM, Davie JR. Gene regulation by Sp1 and Sp3. *Biochemistry and Cell Biology*. 2004 aug; 82(4):460–471. Available from: <https://doi.org/10.1139/o04-045> PMID: 15284899
53. Ward SV, Samuel CE. The pkr kinase promoter binds both Sp1 and Sp3, but only Sp3 functions as part of the interferon-inducible complex with ISGF-3 proteins. *Virology*. 2003 sep; 313(2):553–566. Available from: [https://doi.org/10.1016/s0042-6822\(03\)00347-7](https://doi.org/10.1016/s0042-6822(03)00347-7) PMID: 12954221
54. Cui S, Kolodziej KE, Obara N, Amaral-Psarris A, Demmers J, Shi L, et al. Nuclear Receptors TR2 and TR4 Recruit Multiple Epigenetic Transcriptional Corepressors That Associate Specifically with the Embryonic -Type Globin Promoters in Differentiated Adult Erythroid Cells. *Molecular and Cellular Biology*. 2011 jun; 31(16):3298–3311. Available from: <https://doi.org/10.1128/mcb.05310-11> PMID: 21670149
55. Zhang Y. Silencing of Transcription of the Human Luteinizing Hormone Receptor Gene by Histone Deacetylase-mSin3A Complex. *Journal of Biological Chemistry*. 2002 jun; 277(36):33431–33438. Available from: <https://doi.org/10.1074/jbc.m204417200> PMID: 12091390
56. Bortle KV, Peterson AJ, Takenaka N, O'Connor MB, Corces VG. CTCF-dependent co-localization of canonical Smad signaling factors at architectural protein binding sites in *D. melanogaster*. *Cell Cycle*. 2015 jun; 14(16):2677–2687. Available from: <https://doi.org/10.1080/15384101.2015.1053670> PMID: 26125535
57. Bergström R, Savary K, Morén A, Guibert S, Heldin CH, Ohlsson R, et al. Transforming Growth Factor β Promotes Complexes between Smad Proteins and the CCCTC-binding Factor on the H19Imprinting Control Region Chromatin. *Journal of Biological Chemistry*. 2010 apr; 285(26):19727–19737. Available from: <https://doi.org/10.1074/jbc.m109.088385> PMID: 20427289
58. Harvey R, Lai D, Elliott D, Biben C, Solloway M, Prall O, et al. Homeodomain factor Nkx2-5 in heart development and disease. In: Cold Spring Harbor symposia on quantitative biology. vol. 67. Cold Spring Harbor Laboratory Press; 2002. p. 107–114.
59. Cambier L, Plate M, Sucov HM, Pashmforoush M. Nkx2-5 regulates cardiac growth through modulation of Wnt signaling by R-spondin3. *Development*. 2014; 141(15):2959–2971. Available from: <http://dev.biologists.org/content/141/15/2959> PMID: 25053429
60. Franco R, Esposito F, Fedele M, Liguori G, Pierantoni G, Botti G, et al. Detection of high-mobility group proteins A1 and A2 represents a valid diagnostic marker in post-pubertal testicular germ cell tumours. *The Journal of Pathology*; 214(1):58–64. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/path.2249> PMID: 17935122
61. Fedele M, Fidanza V, Battista S, Pentimalli F, Klein-Szanto AJP, Visone R, et al. Haploinsufficiency of the Hmga1 Gene Causes Cardiac Hypertrophy and Myelo-Lymphoproliferative Disorders in Mice. *Cancer Research*. 2006; 66(5):2536–2543. Available from: <http://cancerres.aacrjournals.org/content/66/5/2536> PMID: 16510570
62. Dubois-Chevalier J, Oger F, Dehondt H, Firmin FF, Gheeraert C, Staels B, et al. A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Research*. 2014 sep; 42(17):10943–10959. Available from: <https://doi.org/10.1093/nar/gku780> PMID: 25183525

63. Dubois-Chevalier J, Staels B, Lefebvre P, Eeckhoutte J. The ubiquitous transcription factor CTCF promotes lineage-specific epigenomic remodeling and establishment of transcriptional networks driving cell differentiation. *Nucleus*. 2015 Jan; 6(1):15–18. Available from: <https://doi.org/10.1080/19491034.2015.1004258>. PMID: 25565413
64. Li Hx, Han M, Bernier M, Zheng B, Sun Sg, Su M, et al. Krüppel-like Factor 4 Promotes Differentiation by Transforming Growth Factor- β Receptor-mediated Smad and p38 MAPK Signaling in Vascular Smooth Muscle Cells. *Journal of Biological Chemistry*. 2010 Apr; 285(23):17846–17856. Available from: <https://doi.org/10.1074/jbc.m109.076992> PMID: 20375011
65. Zhang Xh, Zheng B, Gu C, Fu JR, Wen Jk. TGF- 1 Downregulates AT1 Receptor Expression via PKC-Mediated Sp1 Dissociation From KLF4 and Smad-Mediated PPAR- Association With KLF4. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2012 Apr; 32(4):1015–1023. Available from: <https://doi.org/10.1161/atvbaha.111.244962> PMID: 22282354
66. Beyer TA, Weiss A, Khomchuk Y, Huang K, Ogunjimi AA, Varelas X, et al. Switch Enhancers Interpret TGF- β and Hippo Signaling to Control Cell Fate in Human Embryonic Stem Cells. *Cell Reports*. 2013 Dec; 5(6):1611–1624. Available from: <https://doi.org/10.1016/j.celrep.2013.11.021> PMID: 24332857
67. Imajo M, Ebisuya M, Nishida E. Dual role of YAP and TAZ in renewal of the intestinal epithelium. *Nature Cell Biology*. 2014 Dec; 17(1):7–19. Available from: <https://doi.org/10.1038/ncb3084> PMID: 25531778
68. Lin WJ, Li J, Lee YF, Yeh SD, Altuwajiri S, Ou JH, et al. Suppression of Hepatitis B Virus Core Promoter by the Nuclear Orphan Receptor TR4. *Journal of Biological Chemistry*. 2003 Jan; 278(11):9353–9360. Available from: <https://doi.org/10.1074/jbc.m205944200> PMID: 12522137
69. Lucas B, Grigo K, Erdmann S, Lausen J, Klein-Hitpass L, Ryyfel GU. HNF4 α reduces proliferation of kidney cells and affects genes deregulated in renal cell carcinoma. *Oncogene*. 2005 Jun; 24(42):6418–6431. Available from: <https://doi.org/10.1038/sj.onc.1208794> PMID: 16007190
70. Lee CC, Chen WS, Chen CC, Chen LL, Lin YS, Fan CS, et al. TCF12 Protein Functions as Transcriptional Repressor of E-cadherin, and Its Overexpression Is Correlated with Metastasis of Colorectal Cancer. *J Biol Chem*. 2012 Jan; 287(4):2798–2809. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3268437/> PMID: 22130667
71. Wong SHM, Fang CM, Chuah LH, Leong CO, Ngai SC. E-cadherin: Its dysregulation in carcinogenesis and clinical implications. *Critical Reviews in Oncology/Hematology*. 2018; 121:11–22. Available from: <http://www.sciencedirect.com/science/article/pii/S104084281730224X> PMID: 29279096
72. West JJ, Harris TJC. Cadherin Trafficking for Tissue Morphogenesis: Control and Consequences. *Traffic*; 17(12):1233–1243. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/tra.12407> PMID: 27105637
73. Lagendijk AK, Hogan BM. Chapter Ten—VE-cadherin in Vascular Development: A Coordinator of Cell Signaling and Tissue Morphogenesis. In: Yap AS, editor. *Cellular Adhesion in Development and Disease*. vol. 112 of *Current Topics in Developmental Biology*. Academic Press; 2015. p. 325–352. Available from: <http://www.sciencedirect.com/science/article/pii/S0070215314000258>.
74. Halbleib JM, Nelson WJ. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes & Development*. 2006; 20(23):3199–3214. Available from: <http://genesdev.cshlp.org/content/20/23/3199.abstract>.
75. Li Y, Wen X, Spataro BC, Hu K, Dai C, Liu Y. Hepatocyte Growth Factor Is a Downstream Effector that Mediates the Antifibrotic Action of Peroxisome Proliferator—Activated Receptor- γ Agonists. *Journal of the American Society of Nephrology*. 2006; 17(1):54–65. Available from: <http://jasn.asnjournals.org/content/17/1/54.abstract> PMID: 16291834
76. Lee TC, Ziff EB. Mxi1 Is a Repressor of the c-mycPromoter and Reverses Activation by USF. *Journal of Biological Chemistry*. 1999 Jan; 274(2):595–606. Available from: <https://doi.org/10.1074/jbc.274.2.595> PMID: 9872993
77. Zervos AS, Gyuris J, Brent R. Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. *Cell*. 1993; 72(2):223–232. Available from: <http://www.sciencedirect.com/science/article/pii/S009286749390662A> PMID: 8425219
78. Ramos Pittol JM, Oruba A, Mittler G, Sacconi S, van Essen D. Zbtb7a is a transducer for the control of promoter accessibility by NF-kappa B and multiple other transcription factors. *PLOS Biology*. 2018 05; 16(5):1–33. Available from: <https://doi.org/10.1371/journal.pbio.2004526>.
79. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med*. 2015 Jun; 47(6):e166. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4491725/> PMID: 26045254
80. Lutz M, Burke LJ, Barreto G, Goeman F, Greb H, Arnold R, et al. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res*. 2000 Apr; 28(8):1707–1713. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC102824/> PMID: 10734189

81. Choi WI, Jeon BN, Yun CO, Kim PH, Kim SE, Choi KY, et al. Proto-oncogene FBI-1 Represses Transcription of p21CIP1 by Inhibition of Transcription Activation by p53 and Sp1. *J Biol Chem*. 2009 May; 284(19):12633–12644. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2675992/> PMID: 19244234
82. Stormo GD. DNA binding sites: representation and discovery. *Bioinformatics*. 2000 jan; 16(1):16–23. Available from: <https://doi.org/10.1093/bioinformatics/16.1.16> PMID: 10812473
83. Darnell J, Kerr I, Stark G. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science*. 1994 jun; 264(5164):1415–1421. Available from: <https://doi.org/10.1126/science.8197455> PMID: 8197455
84. Wang Z, Ji J, Peng D, Ma F, Cheng G, Qin FXF. Complex Regulation Pattern of IRF3 Activation Revealed by a Novel Dimerization Reporter System. *The Journal of Immunology*. 2016; 196(10):4322–4330. Available from: <http://www.jimmunol.org/content/196/10/4322> PMID: 27045107
85. Ghislain JJ, Wong T, Nguyen M, Fish EN. The Interferon-Inducible Stat2:Stat1 Heterodimer Preferentially Binds In Vitro to a Consensus Element Found in the Promoters of a Subset of Interferon-Stimulated Genes. *Journal of Interferon & Cytokine Research*. 2001 jun; 21(6):379–388. Available from: <https://doi.org/10.1089/107999001750277853>.
86. Euskirchen GM, Auerbach RK, Davidov E, Gianoulis TA, Zhong G, Rozowsky J, et al. Diverse Roles and Interactions of the SWI/SNF Chromatin Remodeling Complex Revealed Using Global Approaches. *PLOS Genetics*. 2011 03; 7(3):1–20. Available from: <https://doi.org/10.1371/journal.pgen.1002008>.
87. Hsu DS, Acharya CR, Balakumaran BS, Riedel RF, Kim MK, Stevenson M, et al. Characterizing the developmental pathways TTF-1, NKX2-8, and PAX9 in lung cancer. *Proceedings of the National Academy of Sciences*. 2009 mar; 106(13):5312–5317. Available from: <https://doi.org/10.1073/pnas.0900827106>.