MANUSCRIPT TITLE

A pipeline to identify TF combinatorial binding uncovers TEAD1 as an antagonist of tissue-specific transcription factors in human organogenesis.

AUTHORS

Araceli Garcia-Mora[1], Joshua Mallen[1], Peyman Zarrineh[1], Neil Hanley[1], Dave Gerrard[2], Nicoletta Bobola[1 *]

[1]School of Medical Sciences, The University of Manchester, Manchester, United Kingdom

[2]School of Biological Sciences, The University of Manchester, Manchester, United Kingdom

* Correspondence: nicoletta.bobola@manchester.ac.uk

Present Address: [Araceli Garcia Mora], Centre for Developmental Neurobiology, King's College London, London, UK

ABSTRACT

Gene expression is largely controlled by transcription factors (TFs), which bind to distal enhancers to facilitate recruitment of RNA Pol II at promoters. TFs bind to enhancers in combination with other TFs, a mechanism referred to as combinatorial binding. Although TF combinatorial binding is well established, the functional tissue-specific combinations of TFs at active enhancers during human embryonic development are under-explored. Here, we developed cocoTF, a pipeline to identify co-occurring TF motifs at context-specific regulatory regions using comprehensive bioinformatic tools and widely available H3K27ac ChIP-seq and RNA-seq data as input. We use cocoTF to explore co-occurring TF motifs in tissue-specific developmental enhancers of 11 human embryonic tissues. We identify a significant enrichment of recognition motifs for ubiquitous TFs in the vicinity of tissue-specific sequence signatures, pointing at universal patterns of TF functional connectivity in organ-specific transcriptional networks. We focused on TEAD TFs to address the functional role of ubiquitous TFs on cell type-specific transcriptional programs. Our results indicate that TEAD1, together with its coactivator YAP, attenuates tissue-specific enhancer activation, pointing at a broad effect of TEAD on cell type-specific transcriptional programs.

INTRODUCTION

Gene expression programs determine and maintain cellular identity and function in embryonic development and adult tissue homeostasis, and their dysregulation causes a broad range of diseases (1). Gene expression is largely controlled by transcription factors (TFs), which bind to distal enhancers (in the linear genome) to facilitate recruitment of RNA Pol II at promoters (2,3).

TFs operate in highly interconnected networks. Examples from animal development or cellular reprogramming indicate that transcriptional regulation is commonly mediated by distinct combinations of TFs (3,4). This combinatorial mechanism allows the integration of multiple biological inputs at cis-regulatory elements, resulting in highly diverse regulatory outputs in space and time (2), as well as precise fine tuning of gene expression (5). Combinatorial TF binding is the effect of cooperative binding. Diverse mechanisms underlying TF cooperativity have been described. Direct cooperativity is the result of direct protein-protein interactions: TFs combine to form a hetero or homodimer, which displays more stable, higher affinity contacts with DNA (6). Indirect cooperativity is a less specific mechanism and probably underlies the widespread occurrence of combinatorial TF binding on chromatin. In this context, TFs rely on mutual interdependence while engaging with their respective sites on chromatin. Single TFs cannot effectively compete with nucleosomes, but multiple TFs that recognize closely spaced binding sites (within ~150 bp) synergistically act through 'mass action' to displace the same nucleosome. Consequently, they indirectly aid and enhance the binding of each other (6).

The notion of combinatorial binding of TFs is firmly established. However, most of the documented cases of combinatorial TF binding come from specific scenarios, like distinct cell types or developmental timeframes, and a systematic investigation of the tissue-specific TF combinations active in human embryonic development is yet to be undertaken. Consequently, our grasp of the

overall scope of TF combinatorial binding across different tissues and organs remains incomplete. Uncovering general principles of transcriptional networks is important to understand how cell states are instructed and maintained and to learn the precise sequence codes that determine when and where genes are expressed. These insights can advance the field of stem cell differentiation and cellular reprogramming by fine tuning epigenomic and transcriptomic events leading to cell-fate determination (7).

TF combinatorial binding can be identified in the genome in the form of co-localized TF binding sites (TFBS), sequence motifs within active regulatory regions. TFBS can be represented by numerical matrices called position weight matrices (PWMs), which enable predictions of TF binding based on DNA sequence (8). Several methods are available to explore the co-occurrence of TFs at regulatory regions (9–19). Some require ChIP-seq data of target TFs (10,12,18,19) or integrate TF and H3K27ac ChIP-seq experiments (13). These approaches are relatively context-specific as they rely on the availability of the target TF ChIP-seq, which relative to H3K27ac data, is less widely available. Other methods do not require TF ChIP-seq experiments, but largely focus on direct cooperativity, promoter regions rather than enhancers or use chromatin accessibility data as input instead of H3K27ac (11,14,15,17). Here, we asked if organ-specific transcriptional networks display universal patterns of TF functional connectivity. To systematically explore TF combinatorial binding at tissue-specific enhancers, active during human organogenesis, we developed cocoTF, a pipeline designed to identify <u>co</u>ntext-specific <u>co</u>-occurring TF motifs. CocoTF relies on well established bioinformatic tools to analyse widely available H3K27ac ChIP-seq and RNA-seq data as input.

Using cocoTF, we identify combinatorial binding of ubiquitous and tissue-specific TFs as a general feature of developmental enhancers. To address the functional role of ubiquitous TFs on cell type-specific transcriptional programs, we focused on broadly expressed TEAD TFs, identified as combinatorial partner of tissue-specific TFs in heart ventricle, retinal pigmented epithelium (RPE), upper and lower limb, kidney, brain and palate. Our results indicate that TEAD1, together with its coactivator YAP, attenuates tissue-specific enhancer activation, pointing at a broad, so far unrecognised effect of TEAD on cell type-specific transcriptional programs. We propose that TEAD may link transcriptional programmes controlling organ growth and differentiation in multiple tissues.

MATERIAL AND METHODS

Building cocoTF: a pipeline to identify context-specific co-occurring motifs.

Clustering PWMs by similarity. The HOMER motif library file (20) was used as a source of PWMs. A clustering method which groups PWMs by distance was used to group all motifs from the HOMER motif library. First, a similarity matrix using the R package *universalmotif* (21) was calculated for all pairs of PWMs, followed by the calculation of the corresponding distance matrix (1 – similarity matrix). Hierarchical clustering was performed using the base R function *hclust*, which uses the “complete” agglomeration method as default. A distance tree was plotted, and a cut-off height of 0.5 was used to obtain 79 clusters of motifs. A height of 0.5 was chosen based on the representation of known families. The clusters were named according to TF family systematics (22–26). The most prominent

family or subfamily within the group was used. TF subfamilies were used when possible. This was mainly the case for clades of basic Helix-Loop-Helix (bHLH) factors and some Zinc Finger (Zf) subfamilies. In the case where TFs belonging to the same family but spanning across different subfamilies were clustered together, the family and a number were assigned, e.g. Nuclear Receptors (NR) were NR_1, NR_2, NR_3, NR_4, NR_5. Clusters containing only two TFs which belonged to different families were labelled as “other”.

cocoTF: Identifying genomic coordinates of query motifs. Firstly, the query motif’s (referred to as “First Search” TF motif) PWM is obtained from the HOMER motif library file (users can obtain the PWM from other sources) and used as input for cocoTF. Next, cocoTF uses the *scanMotifGenomeWide.pl* script, from HOMER (v4.11, 10-24-2019) (20) to obtain a bed file of genomic coordinates belonging to the “First Search” motif in the genome build of interest (e.g. hg38). CocoTF uses bedtools *intersect* (27) to select input and background coordinates for motif enrichment analysis. Input coordinates consist of “First Search” TF motif coordinates that lie within context-specific regulatory regions (preferrably a set of H3K27ac ChIP-seq peaks in the case of enhancers). Bedtools *random* (27) is then used to generate a set of 1kb background regions from the same genome build. The background is then filtered to regions containing "First Search" TF motifs. The final background set contains approximately double the number of regions as the foreground set. HOMER uses a binomial distribution to score motifs, which works well when the background sequences outnumber the target sequences, which is why the starting number of random genomic bins is twice the number of tissue-specific bins available (20).

Background optimisation was performed before selecting random genomic coordinates as cocoTF background. Backgrounds tested were: 1) random “First Search” TF coordinates as background, 2) non-tissue-specific enhancers consisting of 1kb bins from the human genome acetylated in all tissues studied from Gerrard et al. (2020) (28) 3) HOMER default background consisting of random sequences with matching GC-content distribution as the target sequences and 4) all other putative tissue-specific enhancers from Gerrard et al. (2020) (28).

cocoTF: Motif enrichment analysis. Motif enrichment analysis is performed using the *findMotifGenomeWide.pl* script, from HOMER (20). As a default, cocoTF uses a size parameter of "200", to search for enrichment motifs 100nt either side of the “First Search” motif coordinates found within context-specific regulatory regions compared to random instances of the “First Search" motif. A p-value cut-off of $1x10^{-5}$ was used and only motifs occurring in 5% or more of foreground regions in HOMER results were considered. Significantly enriched motifs from the HOMER results are substituted by their corresponding cluster of HOMER PWMs. When HOMER results include more than one motif belonging to the same cluster, the motif with lowest p-value is used. The R package *ggplot* (29) is used to produce bubble plots for network visualization, where negative log p-value determine bubble size.

Case Study: characterisation of co-occurring TFs at human tissue-specific developmental enhancers.

Human RNA-seq data. RNA-seq data was originally presented in Gerrard et al. (2016) (30) and reanalysed in Gerrard et al. (2020) (28). It is available at www.ebi.ac.uk/ega/home under the accession EGAS00001003738. The RNA-seq data used in this study has been mapped to the hg38 genome.

Human H3K27ac ChIP-seq data. H3K27ac ChIP-seq data was obtained from Gerrard et al. (2020) (28) and is available at www.ebi.ac.uk/ega/home under the accession EGAS0001004335. The genome was parsed into 1kb bins and reads were mapped onto bins. Tissue-specific H3K27ac 1kb regions were identified using a thresholding method based on read counts. Candidate putative tissue-specific enhancers were bins replicated in both replicates of one tissue but not replicated in any other tissue. For details of the method used refer to the original paper (28).

Identification of "First search" TFs expressed during human organogenesis. RNA counts across tissues were down-sampled to the 75th percentile. Genes encoding TFs (5) were selected and their expression values across tissues was normalised by their median expression. The logarithm of the normalised expression was calculated. To identify tissue-restricted TFs, k-means clustering (k = 10) of TFs was performed using the R package *pheatmap* (31), and the largest cluster, whose expression remained mostly constant (determined visually) across samples was removed, leaving 9 clusters of TFs with tissue-restricted patterns of expression. For each tissue, TFs with a log normalised expression value > 5 in both replicates were classified as tissue-restricted.

Motif enrichment analysis of tissue-specific H3K27ac bins (28) was performed for each tissue individually using HOMER's script *findMotifsGenomeWide.pl* default setting. The top 20 enriched motifs were considered. TFs with tissue-restricted expression (as per the method above) and within the top 20 enriched motifs in the same tissue were selected as "First Search" TFs.

Validation of cocoTF ventricle results with VISTA enhancers. Human experimentally validated heart, limb and forebrain enhancers were downloaded from the VISTA enhancer browser (32). Only enhancers active in exactly one of the three tissues were selected for further analysis resulting in 254 heart enhancers, 400 forebrain enhancers and 258 limb enhancers. Instances of ventricle-specific, limb-specific and brain-specific co-occurring motifs predicted by cocoTF within each set of VISTA enhancers were identified by intersecting all instances of the motifs in the hg19 genome identified using the HOMER *scanMotifGenomeWide.pl* script with the VISTA enhancers' coordinates using bedtools *intersect.* The hypergeometric distribution was calculated using the R stats package, to identify whether a pair of co-occurring motifs significantly appeared within 100nt of each other more frequently in one tissue compared to the other two.

Validation of cocoTF GATA4 ventricle results with GATA4 ChIP-seq. An embryonic mouse ventricle GATA4 ChIP-seq (33) (GSE52123) was used to validate cocoTF's GATA4 results (peak calling is described in the following section). Motif enrichment analysis performed on mouse GATA4 200nt peak summits "peaks" and 100nt either side of GATA4 motifs within mouse GATA4 200nt peak summits "flanks" were compared to cocoTF's results using GATA4 and ventricle-specific H3K27ac bins as input ("cocoTF"). A fourth dataset was used for motif enrichment analysis and consisted of "flanks" within human heart ventricle-specific H3K27ac regions ("H3K27ac"). For "H3K27ac", where mouse GATA4 ChIP-seq and human ventricle-specific H3K27ac bins were used, human ChIP-seq

peaks were converted to mm10 using the UCSC liftOver conversion tool (34). A heatmap showing the enrichment p-values of motifs enriched within the top 20 known results in “peaks”, “flanks”, and “cocoTF” comparing cocoTF was created with R base function heatmap. P-values were centred and scaled across rows.

TF ChIP-seq analysis and downstream analysis. Fastq files from TEAD1 mouse adult liver (35) (n=3), TEAD1 mouse embryonic ventricle (36) (n=2), TEAD1 human primary keratinocyte ChIP-seq (37) (n = 2), GATA4 mouse E12.5 ventricle ChIP-seq (33) (n=2)), GATA6 mouse posterior branchial arch (PBA) ChIP-seq (38) (n=2) and CRX mouse P11 retina ChIP-seq (39) (n =2) were obtained from public databases. Raw fastq files were analysed using an adaptation of the TF ChIP-seq pipeline from Yu et al. (2021) (40) for single-end or paired-end. Trimmomatic (41) was used for trimming, Bowtie2 (42) for aligning to the mouse genome (mm10) or human genome (hg38), samtools (43) to remove the aligned reads with a mapping quality of Q10 and MACS2 for peak calling (44) using default narrow peak calling setting for TFs. Blacklisted peaks (45) were removed using bedtools *subtract* (27). A cut-off of fold enrichment (FE) > 3 was used to select for peak summits. Windows of 100nt either side of the peak summits were considered as peaks. HOMER (20) was used to check for the enrichment of the corresponding TF motif binding sites to ensure the high quality of the ChIP-seq experiment used. Finally, for the three TEAD1 ChIP-seq experiments, HOMER was used to re-centre the peaks around the corresponding top TEAD1 enriched binding site for each experiment. Any conversions from mm10 to hg38 and vice versa were performed using the UCSC liftOver conversion tool (34).

All peak intersections to identify candidate enhancers for experimental validation were done using bedtools *intersect* (27). R bioconductor packages *GenomicRanges* (46) and *ChIPpeakAnno* (47) were used to overlap 200nt peaks around the summits to visualise as Venn diagrams and *ChIPseeker* (47) was used for peak annotation. For gene ontology analysis, the Genomic Regions Enrichment of Annotations Tool (GREAT) (48) was used with default parameters. HOMER's script *findMotifsGenomeWide.pl* (20) was used for motif enrichment analysis.

Identification of candidate enhancers for analysis. To identify candidate putative heart ventricle-specific enhancers bound by GATA6 and TEAD1, mouse PBA GATA6 ChIP-seq summits (38), ventricle GATA4 ChIP-seq summits and (33) ventricle TEAD1 ChIP-seq summits (36) were intersected. Intersecting summits were converted from mm10 to hg38 using the UCSC liftOver conversion tool (34) and intersected with human H3K27ac ventricle-specific 1kb ChIP-seq regions, (28), resulting in a list of high confidence ventricle enhancers bound by TEAD1, Gata6 and GATA4. To further reduce the amount of candidate regions, only regions containing GATA4 (49) and TEAD1 (50) motifs (PWM downloaded from HOMER) within 100nt from each other were selected. Finally, the mouse genome informatics database (MGI) (51) was used to identify regions with known cardiovascular system phenotypes.To identify candidate putative RPE-specific enhancers bound by CRX and TEAD1, mouse retina CRX ChIP-seq summits (39) were converted to hg38 and intersected with human H3K27ac RPE-specific 1kb regions (28). To ensure high-confidence regions, only regions containing OTX2/CRX (52) motifs were selected. Only regions containing a TEAD1 (50) motif were

considered to select final candidate enhancers. Finally, candidate enhancers' closest genes had either a phenotype or expression profile in the visual system according to the MGI.

Enhancer cloning. Enhancers were amplified from human genomic DNA using the primers listed in table S1, cloned into pCR8/GW/TOPO vector (Life Technologies) and recombined using the LR Clonase-II Gateway System (ThermoFisher) into pGL4.23-GW (a gift from Jorge Ferrer; Addgene plasmid # 60323; http://n2t.net/addgene:60323RRID:Addgene_60323). Mutant versions of enh_NKX2.6 and enh_EYA2 described in table S1 were synthesized by GenScript and cloned as above.

Promoters were amplified from mouse genomic DNA using primers listed in table S1 and cloned into pGL3-Basic (E1751, Promega) using SacI and BglII.

Cell culture, transfections and luciferase assays. NIH3T3 cells were grown in DMEM (D6429) supplemented with 10% FBS and 5% penicillin/streptomycin, and seeded in 24-well plates at 100,000 cells in 1ml. Enhancers were co-transfected with pcDNA3.1 (V790-20, Thermo Fisher), pcDNA3.1(+)-TEAD1 (generated by Genscript), pcDNA-Flag-YAP (18881, Addgene) (53) and pcDNA3-bCrx (bovine *Crx*, a kind gift from Cheryl M. Craft) or pCMV-Gata6 (38). Cells were transfected with GeneJuice Transfection Reagent (Novagen) 24 hours after seeding, using 250ng of enhancer plasmids and a total of 300ng of expression plasmids per well. Cells were harvested 24 hours after transfection and luciferase measured using Luciferase Assay System and the GloMax Multi-Detection System (Promega). For Verteporfin (VP) experiments, either 0.25 uM VP (Sigma-Aldrich) resuspended in DMSO or the equivalent volume of DMSO (Sigma-Aldrich) was added during seeding. More details can be found in table S1.

Immunofluorescence and microscopy. For IF experiments, NIH3T3 cells were seeded in an eight-chamber slide at 30,000 cells in a final volume of 0.3ml per chamber, and fixed after 48 hours. Slides were fixed with 4% PFA for 10 minutes, washed with PBS (3x 15mins), and incubated overnight with Purified Mouse Anti-TEF-1 (1:500, BD-610922) or mouse anti-YAP1 H-9 (1:150, sc-271134) in a PBS solution containing donkey serum albumin and 0.1% Trizol (ThermoFisher). Cells were washed with PBS (3x for 15mins) and treated with Donkey Anti-Mouse IgG H&L (Alexa Fluor 555) (1:500, Ab-150106) and 0.1% Trizol for 2 hours. Cells were stained with 0.5ug/mL DAPI (D9542, Sigma Aldrich) for 5 mins, washed with PBS (2x 5mins), and the slide mounted using Prolong Gold Antifade (ThermoFisher). Experiment was carried out in duplicate. Images were collected on an Axio Imager.M2 Upright microscope (Zeiss) using a 10x objective and captured using a Coolsnap HQ2 camera (Photometrics) through Micromanager software v1.4.23. Specific band pass filter sets for DAPI and Cyf3 were used to prevent bleed through from one channel to the next. Images were analysed using ImageJ (54).

Statistical analysis. Enhancers were tested in three independent experiments (biological replicates n = 3) and three technical replicates per condition (experimental replicates n = 3). Verteporfin experiments were tested in two independent experiments (biological replicates n = 2) and three technical replicates per condition (experimental replicates n = 3). The median of technical replicates was used as biological replicates, and the average of three biological replicates and the SEM were plotted. Ordinary one-way ANOVA, two-way ANOVA and t-tests were performed using GraphPad

Prism version 9.3.1.471 for Windows (www.graphpad.com). Asterisks represent the following values: ns = P > 0.05, * = P < 0.05, ** = P < 0.01, *** = P < 0.001 and **** = P < 0.0001.

RESULTS

CocoTF: a pipeline to identify context-specific co-occurrence of TF motifs

We designed cocoTF to investigate the combinatorial TF landscape of tissue-specific enhancers. This workflow can be applied to any context of interest to identify co-occurring, context-specific TF motifs in two main steps (Fig 1A and Fig S1). First, context-specific TFs, whose motifs are enriched within active regulatory regions (context-specific, e.g. cardiac-specific) and whose expression levels are enriched in the same context (cardiac as in the example above), are identified. We refer to these TFs as 'First Search' TFs (this step can be skipped if the user already has a TF of interest). Second, co-binding TFs or 'Second Search' TFs, are identified using motif enrichment analysis in sequences flanking the 'First Search' TF motifs. The identification of 'Second Search' TFs is the main workflow and its design requires careful considerations on the search distance between co-occurring motifs. Previous methods created to identify functionally co-occurring TFs from DNA sequences have used tight distance and orientation constraints to identify pairs of TFs that form homo and heterodimers (14). CocoTF aims to identify any functional TF interaction, including those favoured by indirect cooperativity (6). As one nucleosome occupies roughly 150 nt, cocoTF allows a search distance of 100 nt between 'First' and 'Second Search' TF motifs, with no other constraints on space or orientation. The selection of regions to use as input and background is key to discriminate functional motifs from the high number of TF motifs scattered across the genome, the majority of which is unlikely to underlie functional regulatory events. For 'Second Search TF' motif enrichment analysis, cocoTF uses the genomic coordinates of the 'First Search' motif in context specific regions as foreground, and random 'First Search' TF motif coordinates as background. We evaluated different backgrounds when designing cocoTF and while we did not observe major differences in enrichment percentage and significance, random genomic coordinates did mask the query TF better than the others (table S2). As an additional step to minimise false positives, cocoTF integrates expression data to prioritise TFs expressed in the cell type or tissue of interest.

Here, we focused on tissue-specificity acquisition during embryonic organogenesis. We used H3K27ac ChIP-seq from 11 embryonic tissues (28) to identify TF motifs co-occurring at tissue-specific developmental enhancers (Fig 1B). H3K27ac histone modification marks active enhancers (55) and its genomic occurrence provides a landscape of active regulatory elements in cell-types or tissues of interest (56). Most importantly, H3K27ac data is accessible for many tissues and cell types in public databases (50) enabling a systematic investigation of functional co-occurring motifs in the absence of TF binding information. The number of candidate putative enhancers varied across the 11 tissues (Fig 1B). We first scanned each set of tissue-specific candidate enhancers for enriched motifs to identify 'First Search' TFs, and selected the top 20 motifs for further analysis (Fig 1C). Given that TFs belonging to the same families have conserved DNA-binding domains, which recognise very similar motifs, we collapsed all similar motifs in the HOMER motif database. Unsupervised classification resulted in 79 motif clusters (Fig S2, Table S3). In parallel, we analysed tissue-matching RNA-seq

experiments to identify tissue-restricted TFs (Fig 1D). Briefly, RNA-seq counts corresponding to TFs listed in Vaquerizas et al (5) were normalised by median expression across tissues and clustered into tissue-restricted (tissue-specific) and ubiquitous (non-specific) TF clusters. We selected 'First Search' TFs as i) TFs selectively expressed in a specific organ/tissue and ii) whose recognition motifs were represented in the top 20 enriched motifs in the same organ/tissue. This strategy led to the identification of the key transcriptional regulators governing cell fate in individual tissues/organs (table S4). In the example of the heart ventricle, GATA4, HAND1 and TBX20 were identified as 'First Search' TFs and their PWM motifs, obtained from the HOMER motif database, were used as input for 'Second Search' TF (Fig 1E). If multiple motifs belonging to the same cluster (table S3) appeared in the top 20 results, the PWM motif with the lowest enrichment p-value was used as cocoTF input. In some cases, multiple 'First Search' TFs were represented by the same motif (table S4).

The network of co-occurring motifs in enhancers active during human organogenesis

Applying cocoTF to human developmental enhancers resulted in a comprehensive network of co-occurring TF motifs across developing tissues in the human embryo (Fig 2A). The network was built using stringent cut-offs to select highly significant instances of TF motif co-occurrence (foreground enrichment > 5% and p-value < 1e-5). CocoTF predicted known TF functional interactions in cell differentiation. In liver, cocoTF identifies the well-established cooperation of FOXA2, GATA4 and HNF4a that primes enhancers for the activation of a liver-specific gene programmes (57,58), as well as co-binding of HNF4a with RXR nuclear receptor (59). In the brain, cocoTF detects instances of co-binding of bHLH and SOX TFs (ASCL1 and OLIG2 bind a glia-specific enhancer with SOX10 (60)), of NEUROD and bHLH TFs (NEUROD1 and TCF12 cooperate to drive corticogenesis (61)), and instances of another LIM homeodomain binding proximal to LHX motifs (consistent with LIM homeodomain LHX2 binding to DNA as a dimer (62)). Additional examples of functional interactions identified by cocoTF include GATA4 and HAND2 cooperativity to activate the expression of cardiac-specific genes (63), GATA4 and ER in cardiomyocyte differentiation (64) and TBX20, NKX2.5 and GATA4 during heart development (65). The identification of well-established functional TF-TF interactions in tissue specification and differentiation suggest that cocoTF is indeed capable of detecting instances of TF functional co-binding involved in tissue-specificity acquisition. As a more formal validation of cocoTF results, we used GATA4 occupancy in mouse embryonic heart (33) and compared motif enrichment analysis results of GATA4 ChIP-seq called peaks with cocoTF's GATA ventricle results. We performed motif enrichment analysis on GATA4 peaks ('peaks'), sequences flanking any GATA4 motifs contained in GATA4 peak ('flanks') and 'flanks' overlapping the set of human ventricle-specific enhancers converted to mouse coordinates ('H3K27ac'). We detected significant overlap within the top ten enriched motifs in 'peaks', 'flanks', 'H3K27ac' and cocoTF (Fig 2B). The heatmap shows negative log p-values for the top significantly enriched motifs in the three datasets (TEAD, MEF2, TALE, TBX, GATA). TEAD is the top co-occurring motif in 'peaks' and 'flanks', while MEF2 is the top motif in cocoTF and'H3K27ac'. As MEF2 also appears as the top motif in 'H3K27ac', this discrepancy is likely explained by an increased concentration of GATA-MEF2 sites in

active enhancers (relative to all GATA4 binding events), rather than an inaccuracy of cocoTF prediction. In sum, motif enrichment analyses indicate that MEF2, TALE, TBX, and TEAD often co-occur with GATA4 in the ventricle, and that GATA-MEF2 tend to bind together in active ventricle-specific enhancers. As an additional validation, we examined experimentally validated enhancers from the VISTA database (32) for motifs identified by cocoTF. We selected heart, forebrain and limb enhancers and performed hypergeometric tests to assess whether ventricle-specific motif pairs predicted by cocoTF co-occurred within 100nt of each other significantly more often in experimentally validated heart enhancers, compared to forebrain and limb enhancers. We focused on the following cocoTF predictions: GATA-TEAD, GATA-MEF2, GATA-TBX (TBX clustered within the Zf Snail group) and GATA-NKX. We also used cocoTF brain prediction bHLH-RFX, and limb prediction HOX-PITXEBOX. As expected, GATA-TEAD, GATA-MEF2 and GATA-NKX pairs (co-occurrence within 100nt) were significantly enriched in heart enhancers compared to limb and brain enhancers (p-value < 0.05) (Fig 2C). GATA-TBX pair was not enriched in the heart compared to brain and limb enhancers; however very few instances of TBX motifs were detected in the VISTA enhancers, which has an effect on hypergeometric p-values. Moreover, HOX-PITXEBOX was significantly enriched in the limb compared to brain and heart enhancers. We could not detect a significant enrichment of bHLH-RFX in brain enhancers compared to the heart and limb enhancers. This could be explained by the observation that the bHLH motif CAGCTG is enriched in enhancers active across all tissues (Fig 2A). Altogether, these results indicate that cocoTF can successfully identify functional co-occurring motifs in tissue-specific enhancers.

CocoTF provides insights into the regulatory landscape of tissue-specific developmental enhancers.

Homotypic motif co-occurrence is common at tissue-specific developmental enhancers. CocoTF identifies frequent co-binding of TFs belonging to the same family, particularly bHLH, FOX, HOX, GATA, LHX, NR and RFX TF families (Fig S3A). While TFs belonging to these families can bind DNA as homo- or hetero- dimers (66–71), these findings are consistent with previous studies on enhancer grammar, which have reported homotypic co-occurring motifs to be common and to occur independently of dimerization events (72,73). By analysing the overall combinatorial TF landscape of developmental enhancers during development, we observe that homotypic motif pairs are common across different TF families and, potentially, play a broad role in tissue specification.

Liver pioneer TF motifs co-occur with distinct sets of motifs. In the liver, FOXA and to a lesser degree GATA TFs function as pioneer factors (57,74) and bind to inactive, compact chromatin to promote accessibility for other secondary factors to bind DNA. Interestingly, cocoTF predicts FOX and GATA to bind with a distinct set of TFs in the liver. GATA and FOX motifs (potentially recognised by GATA4 and FOXA2 respectively) co-occur with motifs belonging to twelve (GATA) and thirteen (FOX) different motif clusters in liver enhancers, of which only four are shared between them (Fig. S3B). Both GATA and FOX motifs co-occur with AP-1 TFs. A recent study proposes that the observed pioneer ability of AP-1 is probably due to AP-1 cooperating with other factors capable of promoting DNA accessibility (76). CocoTF results suggest that FOX and GATA TFs could bind alongside AP-1 to

promote chromatin accessibility. FOX, but not GATA, co-occurs with other bZIP motif clusters, recognized by ATF, CEBP and DMRT, and with NR motifs, such as HNF4a recognition sequence (Fig S3B). Interestingly, FOXA1-3 have been shown to bind to nucleosomal DNA in the liver to enable subsequent binding of HNF4a and C\EBP (74). CocoTF predicts GATA, but not FOX, to co-occur with ZF motif clusters such as those recognized by GLI, KLF, SP, MAZ and ZIC (Fig S3B). Nevertheless, GATA and FOX are also predicted to bind together in liver enhancers. A likely explanation is that the set of putative liver-specific enhancers can be divided in three groups, enhancers bound by GATA, enhancers bound by FOX and enhancers bound by both.

E-box motifs are enriched close to 'First Search' TFs in most of the tissues studied. To classify 'Second Search' TF clusters based on the number of potential interactions, we created a heatmap of absolute and normalised number of interactions per 'Second Search' TF cluster (Fig. S3C). This resulted in four groups of 'Second Search' TF clusters, high co-occurring, medium co-occurring, low co-occurring and normalised medium co-occurring. Without normalization, the most common 'Second Search' TFs in developmental enhancers (based on the absolute number of predicted co-occurring partners) are members of the bHLH TF family, which bind to the consensus sequence CAGCTG, belonging to the cluster bHLH_C_A_B (Fig S3C). BHLH motifs co-occur with 'First Search' TFs in developmental enhancers in all tissues studied, except the pancreas (Fig 2A). BHLH TFs are known to be crucial during the establishment of tissue- and cell-identity in a wide range of tissues. These include pro-neural bHLH TFs such as NEUROD1, ATOH1 or ASCL1 amongst others (77–79), MYOD in muscle tissue differentiation (80), Atoh8 in skeletal tissue and retina (81) and TCF21 in cardiac fibroblast differentiation (82). Thus, cocoTF results highlight cooperativity of bHLH TF with other tissue-specific TFs as a central feature of human developmental enhancers.

Motifs recognized by families of ubiquitous TFs are a central component of the landscape of tissue-specific developmental enhancers. Tissue- and cell type-specific patterns of expression can be indicators of TF functions (5,83). Based on their expression patterns, TFs can be classified into two groups: ubiquitous TFs and tissue-specific TFs (5). Whether tissue-specific TFs interact with other tissue-specific TFs or whether they interact with ubiquitously expressed TFs to carry out their regulatory roles remains unanswered, with genetic evidence supporting both mechanisms (84,85). CocoTF identifies enrichment of motifs recognised by ubiquitous TF families, occurring proximal to tissue-specific TFs, across multiple tissues (Fig 2A), including TEAD (high co-occurring group, Fig. S3C), and families of widely expressed TFs TALE and ETS (medium co-occurring groups, Fig S3C). We find a significant enrichment of TEAD recognition motifs in the vicinity of tissue-specific sequence signatures in ventricle, lower and upper limbs, brain, kidney, RPE and palate (Fig 3A). Namely, cocoTF predicts co-binding of TEAD TFs with thirteen 'First Search' TFs (Fig 3A). Five of the 'First Search' TFs recognise a homeobox cluster motif (Homeobox_HOXC9, Homeobox_HOXC10, Homeobox_HOXD10, Homeobox_LHX8 and Homeobox_OTX2), suggesting that TEAD TFs widely co-bind with homeodomain TFs. In addition, cocoTF detects TFs belonging to the TALE homeodomain family as 'Second Search' TFs across ventricle, brain, upper limb and kidney (Fig 2A). We found TALE motifs in the vicinity of seven 'First Search' TFs, with the most highly significant predicted partner being GATA4 in the ventricle. Finally, TFs belonging to the ETS TF family are

predicted to functionally occur proximal to eleven 'First Search' TFs, in the lung, liver, upper and lower limb, adrenal gland and kidney (Fig 2A).

TEAD1 binds tissue- cell type-specific enhancers

We focused on TEAD TFs to explore the functional relevance of ubiquitous TF binding at tissue-specific enhancers. TEAD family is relatively small, with only four members, TEAD1-4 (86), which simplifies PWM assignments. In addition, the high number of interactions involving TEAD, even after normalization by the number of cluster members, suggests a broad and significant role of TEAD at developmental enhancers. With the exception of TEAD4, which is predominantly expressed in skeletal muscle later in development (87), TEAD TFs are largely ubiquitous. As a first step, we investigate if TEAD binds tissue-specific enhancers. We compared TEAD1 binding in adult mouse liver (35), mouse embryonic ventricle (36) and human primary keratinocytes (37). Indeed, overlap of high confidence TEAD1 peaks, centred on their top TEAD motif (Fig 3B), shows that most TEAD1 binding is tissue-or cell type-specific (Fig 3C). While the majority of tissue-specific TEAD1 peaks occur at intronic or distal intergenic regions, as expected for enhancer regions (Fig 3D), a high proportion of overlapping (non-tissue-specific; n=440) TEAD1 peaks occur at promoters. Tissue-specific TEAD1 peaks are associated with tissue-specific biological processes and mouse phenotypes (Fig 3E-G). In contrast, shared TEAD peaks are associated with the hippo signalling pathway and with tumorigenic mouse phenotypes (Fig 3H), consistent with the well-established role of TEAD, and its coactivator YAP, in the hippo pathway and in cancer progression (88,89). Moreover, tissue-specific peaks are enriched in motifs recognised by tissue-specific TFs, such as MEF2, NKX and TBOX in ventricle peaks (Fig S4A-C), matching cocoTF predictions of TEAD binding partners. Compared to tissue-specific peaks, shared non-tissue-specific TEAD1 peaks are enriched in high affinity TEAD site, suggesting that shared peaks may contain higher affinity versions of TEAD binding sites or multiple TEAD motifs (Fig S4D). Overall, these results support cocoTF predictions and suggest that TEAD act in concert with tissue-specific TFs at tissue-specific enhancers.

TEAD1 attenuates tissue-specific enhancer activation

We addressed the functional effects of TEAD cooperativity with tissue-specific regulators on high-confidence developmental enhancers. We selected two co-occurring partners of TEAD in the ventricle and in the RPE, GATA6 and CRX respectively (Fig 3A). Notably, *de novo* motif discovery performed on GATA6 peaks in the pharyngeal arches identifies TEAD amongst the top enriched motifs (38). We identified high confidence instances of GATA-TEAD and CRX-TEAD binding *in vivo*. For GATA-TEAD in the ventricle, we overlapped human ventricle-specific H3K27ac bins with TEAD1 (36) and GATA4 binding in mouse embryonic heart (33), and GATA6 binding in mouse posterior pharyngeal arches (which contribute to the anterior pole of the heart) (38). Overlapping regions containing TEAD and GATA motifs within a distance of 100nt and associated with genes with cardiac phenotypes (Fig S5A) were prioritised, resulting in five high confidence ventricle enhancers. For CRX-TEAD, given the lack of available CRX and TEAD1 binding information in the RPE, we selected high confidence RPE enhancers using RPE-specific H3K27ac bins and CRX peaks in the retina as a proxy, and association

with genes expressed in the visual system. Three enhancers were prioritised for downstream analysis, using less stringent selection criteria than those adopted for ventricle enhancers (Fig S5B).

All the selected enhancers, cloned upstream of a minimal promoter and reporter gene, respond to their tissue-specific activator (GATA6 or CRX) (Fig 4AB and Fig S5C). In all cases tested, and irrespectively of the activator, addition of TEAD1 attenuates enhancer activation (Fig 4AB and Fig S5C). Levels of inhibition by TEAD1 varied across enhancers (Fig S5D). YAP is a well-established coactivator of TEAD1 (86). We observed an additional decrease in enhancer activation when YAP and TEAD are co-transfected together (Fig 4AB). In contrast, disruption of the TEAD-YAP complex (90), resulted in higher GATA activation for all the enhancers tested (Fig 4C), consistent with endogenous nuclear TEAD/YAP (Fig S5E). Finally, ruling out transcriptional repression as a generalised effect of our experimental system, TEAD and YAP increase luciferase activity driven by the CCN2 promoter, a well-established TEAD target (Fig S5F). The identification of TEAD motifs close to tissue-specific TFs in developmental enhancers suggest that TEAD/YAP complex exerts its effect by binding to the enhancer. To address this, we generated mutant versions of one cardiac- and one RPE-specific enhancers (Fig 4DF). We observed that the repressive effect of TEAD/YAP complex is impaired in the absence of a canonical TEAD motif (Fig 4EG), suggesting that TEAD/YAP ability to modulate enhancer activity is dependent on TEAD binding to the enhancer. Overall, these results suggest that TEAD/YAP binding at developmental enhancers can negatively interfere with their activation by tissue-specific TFs and uncover a broad, so far unrecognised, effect of TEAD on cell type-specific transcriptional programs.

DISCUSSION

Gene expression is largely controlled by TFs, which determine the global state of the cell. TF cooperativity is well established and widespread across TF networks. However, our knowledge of TF combinatorial binding largely derives from specific contexts (e.g. cell types or developmental time points) and a global view of tissue-specific TF combinations at active developmental enhancers is lacking. We developed cocoTF to identify co-occurring TF motifs at active, context-specific regulatory regions. CocoTF is designed to detect instances of indirect cooperativity, which likely underlie the majority of TF combinatorial binding. It is run with software that is widely accessible to users and its input data consists of active enhancer and expression data (H3K27ac ChIP-seq and RNA-seq), which are also widely available. Due to its versatility, it can be used in numerous contexts. Notably, out of all other methods to identify functional co-occurring motifs at context-specific regions, TFcomb (16) is the most similar. Similar to cocoTF, TFcomb identifies co-occurring TF pairs in 100 nt sliding windows. However, while TFcomb uses market basket analysis to prioritise TF combinations, cocoTF uses a two-step approach to identify those motifs co-occurring with high confidence tissue-specific TF motifs. Potential applications of cocoTF include the identification of TF co-binding to optimise cell reprogramming or cell differentiation protocols (91). Moreover, as the single-cell genomics field advances, the increased availability of scH3K27ac (92,93) and scRNA-seq data will enable to run cocoTF on distinct, well-defined cell populations. We have provided a range of evidence that supports cocoTF's ability to detect functional co-occurring TF motifs. There are however limitations to using

bioinformatics methods that rely on PWMs to predict functional binding events. The main one is that co-expressed TFs belonging to the same TF family recognise the same or very similar sites, limiting the accuracy of the predictions.

A key finding of cocoTF is the recurrent co-occurrence of ubiquitous and tissue-specific TF motifs in active developmental enhancers. Namely, we find that tissue-specific enhancers (from unrelated tissues) are significantly enriched in motifs for the ubiquitous TFs TEAD, which occur nearby motifs recognised by tissue-specific TFs. By testing a subset of cardiac and RPE developmental enhancers, we find that TEAD1 together with its coactivator YAP, attenuates tissue-specific enhancer activation. The TEAD family of TFs consists of four proteins: TEAD1, TEAD2, TEAD3 and TEAD4. With the exception of TEAD4, TEAD TFs are largely ubiquitous; they do not contain an activation domain and provide a binding platform for coactivators, most notably YAP/TAZ. Formation of the TEAD-YAP complex is regulated by nuclear cytoplasmic shuttling of YAP; multiple inputs control YAP nuclear localization, most notably the Hippo pathway (94). In mouse organogenesis, hippo signalling regulates cell proliferation and organ growth (95,96) in heart, craniofacial, lung, eye, brain, kidney and bile duct development (97). Consistently, cocoTF identifies enriched TEAD motifs in developmental enhancers specifically active in the ventricle, palate, RPE, brain, kidney and upper and lower limb. The TEAD/YAP complex typically functions at early differentiation stages, in progenitor cells and is dowregulated at terminal stages of differentiation (98). As well as cell proliferation, the complex controls cell differentiation. TEAD or YAP loss of function experiments result in reduced cell proliferation and rapid differentiation (99) whilst gain of function experiments result in the opposite: aberrant differentiation and an increase in cell proliferation (35,96,100–102). In a developing organism, correctly timed differentiation of proliferating progenitors is essential for appropriate organ composition and size. Our observations point at a broad, so far unrecognised effect of TEAD on cell type-specific transcriptional programs. We propose that while promoting proliferation of precursor cells, TEAD TFs simultaneously attenuate their differentiation by targeting lineage specific enhancers (Fig 5). Attenuation of cell type differentiation programs by TEAD TFs may therefore be an essential mechanism to inhibit premature differentiation of progenitors and ensure the production of appropriate numbers of differentiated cells and the correct organ size.

The TEAD/YAP complex is well-known to function as an activator complex, but can also work as a transcriptional repressor, using a variety of mechanisms. These include NuRD complex recruitment (103), prevention of RNA pol II pause-release and inhibition of HNF4a binding at enhancers (35), or direct interaction with other TFs (99) (104). In our system, TEAD repressive function is dependent on the presence of an activator (TEAD/YAP complex does not repress enhancer activity per se) and requires TEAD binding to DNA. The observation that TEAD can act as a repressor on episomal DNA, which lacks chromatinization, indicates that this mechanism can occur, at least in part, independently from chromatin. However, additional studies on developmental enhancers in their genomic context, are required to fully elucidate TEAD role as a repressor.

In addition to TEAD, cocoTF identifies other commonly co-occurring clusters of motifs belonging to ubiquitous TF families such as TALE and ETS. Our observations suggest that, in human organogenesis, ubiquitously expressed TFs are a key component of the chromatin landscape of developmental enhancers. Uncovering general principles of transcriptional networks is important to understand how cell states are instructed and maintained, and to learn the precise sequence codes that determine when and where genes are expressed. The main focus of tissue-specific acquisition research revolves around the expression of tissue-specific TFs (105). Tissue-specific TF expression is central to stem cell differentiation protocols, reprogramming protocols and in developmental disease diagnostics. Analysis of co-occurring motifs enriched in promoters of differentially expressed genes uncovered a potential involvement of non-tissue-specific TFs in driving tissue-specificity (106). Given that enhancers have higher tissue-specificity than promoters (107), the recurring enrichment of ubiquitous TF binding sites at tissue-specific developmental regulatory elements is highly significant. This further reinforces that ubiquitous TFs should receive a larger spotlight in the process of tissue-specificity acquisition. It also opens the questions of how essential ubiquitous TFs are for correct cell differentiation and lineage determination, and what is their functional role in the acquisition of tissue-specificity.

DATA AVAILABILITY

CocoTF scripts and additional analysis scripts can be found in https://github.com/AGarciaMora/CocoTF.

SUPPLEMENTARY DATA

Supplementary Figures and Supplementary Table4.pdf

Supplementary Tables1-3.xlxs

ACKNOWLEDGEMENTS

The authors would like to thank members of the Bobola lab for their scientific and technical insights, and the Bioimaging Facility at the University of Manchester for help with the imaging performed in this project.

FUNDING

This work was supported by a Medical Research Council (MRC) (http:www.mrc.ukri.org) studentship to AGM, and Biotechnology and Biological Sciences Research Council (BBSRC) (http:www.bbsrc.ukri.org) grants BB/N00907X/1 and BB/T007761/1 to NB. The Bioimaging Facility microscopes used in this study were purchased with grants from BBSRC, Wellcome and the University of Manchester Strategic Fund.

CONFLICT OF INTEREST

No conflict of interest exists that might raise the question of bias in the work reported or the conclusions, implications, or opinions stated.

REFERENCES

1. Lee TI, Young RA. Transcriptional Regulation and Its Misregulation in Disease. Cell. 2013 Mar;152(6):1237–51.

2. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012 Sep 7;13(9):613–26.

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

4. Biggin MD. Animal Transcription Networks as Highly Connected, Quantitative Continua. Dev Cell. 2011 Oct;21(4):611–26.

5. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009 Apr;10(4):252–63.

6. Morgunova E, Taipale J. Structural perspective of cooperative transcription factor binding. Curr Opin Struct Biol. 2017 Dec;47:1–8.

7. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, et al. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell. 2017 Jan;168(3):442-459.e20.

8. Inukai S, Kock KH, Bulyk ML. Transcription factor–DNA binding: beyond binding site motifs. Curr Opin Genet Dev. 2017 Apr;43:110–9.

9. Ha N, Polychronidou M, Lohmann I. COPS: Detecting Co-Occurrence and Spatial Arrangement of Transcription Factor Binding Motifs in Genome-Wide Datasets. PLoS One. 2012 Dec 18;7(12):e52055.

10. Levitsky VG, Mukhin AM, Oshchepkov DYu, Zemlyanskaya E V., Lashin SA. Web-MCOT Server for Motif Co-Occurrence Search in ChIP-Seq Data. Int J Mol Sci. 2022 Aug 11;23(16):8981.

11. Vandenbon A, Kumagai Y, Akira S, Standley DM. A novel unbiased measure for motif co-occurrence predicts combinatorial regulation of transcription. BMC Genomics. 2012;13(Suppl 7):S11.

12. Vandel J, Cassan O, Lèbre S, Lecellier CH, Bréhélin L. Probing transcription factor combinatorics in different promoter classes and in enhancers. BMC Genomics. 2019 Dec 1;20(1):103.

13. Teng L, He B, Gao P, Gao L, Tan K. Discover context-specific combinatorial transcription factor interactions by integrating diverse ChIP-Seq data sets. Nucleic Acids Res. 2014 Feb 1;42(4):e24–e24.

14. Jankowski A, Szczurek E, Jauch R, Tiuryn J, Prabhakar S. Comprehensive prediction in 78 human cell lines reveals rigidity and compactness of transcription factor dimers. Genome Res. 2013 Aug;23(8):1307–18.

15. van Bömmel A, Love MI, Chung HR, Vingron M. coTRaCTE predicts co-occurring transcription factors within cell-type specific enhancers. PLoS Comput Biol. 2018 Aug 24;14(8):e1006372.

16. Bentsen M, Heger V, Schultheis H, Kuenne C, Looso M. TF-COMB – Discovering grammar of transcription factor binding sites. Comput Struct Biotechnol J. 2022;20:4040–51.

17. Jankowski A, Prabhakar S, Tiuryn J. TACO: a general-purpose tool for predicting cell-type–specific transcription factor dimers. BMC Genomics. 2014 Dec 19;15(1):208.

18. Whitington T, Frith MC, Johnson J, Bailey TL. Inferring transcription factor complexes from ChIP-seq data. Nucleic Acids Res. 2011 Aug;39(15):e98–e98.

19. Yang CC, Chen MH, Lin SY, Andrews EH, Cheng C, Liu CC, et al. Inferring condition-specific targets of human TF-TF complexes using ChIP-seq data. BMC Genomics. 2017 Dec 10;18(1):61.

20. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010 May;38(4):576–89.

21. Tremblay BJM. universalmotif: Import, Modify, and Export Motifs with R. 2022.

22. Skinner MK, Rawls A, Wilson-Rawls J, Roalson EH. Basic helix-loop-helix transcription factor gene family phylogenetics and nomenclature. Differentiation. 2010 Jul;80(1):1–8.

23. Holland PW, Booth HAF, Bruford EA. Classification and nomenclature of all human homeobox genes. BMC Biol. 2007 Dec 26;5(1):47.

24. Weikum ER, Liu X, Ortlund EA. The nuclear receptor superfamily: A structural perspective. Protein Science. 2018 Nov;27(11):1876–92.

25. Yanai H, Negishi H, Taniguchi T. The IRF family of transcription factors. Oncoimmunology. 2012 Nov 4;1(8):1376–86.

26. Krishna SS. Structural classification of zinc fingers: SURVEY AND SUMMARY. Nucleic Acids Res. 2003 Jan 15;31(2):532–50.

27. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.

28. Gerrard DT, Berry AA, Jennings RE, Birket MJ, Zarrineh P, Garstang MG, et al. Dynamic changes in the epigenomic landscape regulate human organogenesis and link to developmental disorders. Nat Commun. 2020 Dec 6;11(1):3920.

29. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. Available from: https://ggplot2.tidyverse.org

30. Gerrard DT, Berry AA, Jennings RE, Piper Hanley K, Bobola N, Hanley NA. An integrative transcriptomic atlas of organogenesis in human embryos. Elife. 2016 Aug 24;5.

31. Kolde R. pheatmap: Pretty Heatmaps. 2019.

32. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 2007 Jan 3;35(Database):D88–92.

33. He A, Gu F, Hu Y, Ma Q, Ye LY, Akiyama JA, et al. Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. Nat Commun. 2014 Sep 24;5:4907.

34. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002 Jun 1;12(6):996–1006.

35. Biagioni F, Croci O, Sberna S, Donato E, Sabò A, Bisso A, et al. Decoding YAP dependent transcription in the liver. Nucleic Acids Res. 2022 Aug 12;50(14):7959–71.

36. Akerberg BN, Gu F, VanDusen NJ, Zhang X, Dong R, Li K, et al. A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. Nat Commun. 2019;10(1):4907.

37. Li J, Tiwari M, Xu X, Chen Y, Tamayo P, Sen GL. TEAD1 and TEAD3 Play Redundant Roles in the Regulation of Human Epidermal Proliferation. Journal of Investigative Dermatology. 2020 Oct;140(10):2081-2084.e4.

38. Losa M, Latorre V, Andrabi M, Ladam F, Sagerström C, Novoa A, et al. A tissue-specific, Gata6-driven transcriptional program instructs remodeling of the mature arterial tree. Elife. 2017 Sep 27;6.

39. Andzelm MM, Cherry TJ, Harmin DA, Boeke AC, Lee C, Hemberg M, et al. MEF2D drives photoreceptor development through a genome-wide competition for tissue-specific enhancers. Neuron. 2015 Apr 8;86(1):247–63.

40. Yu CP, Kuo CH, Nelson CW, Chen CA, Soh ZT, Lin JJ, et al. Discovering unknown human and mouse transcription factor binding sites and their characteristics from ChIP-seq data. Proceedings of the National Academy of Sciences. 2021 May 18;118(20).

41. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

42. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr 4;9(4):357–9.

43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

44. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008 Nov 17;9(9):R137.

45. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep. 2019 Dec 27;9(1):9354.

46. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol. 2013 Aug 8;9(8):e1003118.

47. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015 Jul 15;31(14):2382–3.

48. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010 May 2;28(5):495–501.

49. van den Boogaard M, Wong LYE, Tessadori F, Bakker ML, Dreizehnter LK, Wakker V, et al. Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. Journal of Clinical Investigation. 2012 Jul 2;122(7):2519–30.

50. Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Ai R, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020 Jul 30;583(7818):699–710.

51. Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ, et al. Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology. Nucleic Acids Res. 2021 Jan 8;49(D1):D981–7.

52. Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell. 2014 Jun 5;14(6):838–53.

53. Levy D, Adamovich Y, Reuven N, Shaul Y. Yap1 Phosphorylation by c-Abl Is a Critical Step in Selective Activation of Proapoptotic Genes in Response to DNA Damage. Mol Cell. 2008 Feb;29(3):350–61.

54. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012 Jul 28;9(7):671–5.

55. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences. 2010 Dec 14;107(50):21931–6.

56. Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. Brain cell type–specific enhancer–promoter interactome maps and disease - risk association. Science (1979). 2019 Nov 29;366(6469):1134–9.

57. Cirillo LA, Zaret KS. An Early Developmental Transcription Factor Complex that Is More Stable on Nucleosome Core Particles Than on Free DNA. Mol Cell. 1999 Dec;4(6):961–9.

58. Hoffman BG, Robertson G, Zavaglia B, Beach M, Cullum R, Lee S, et al. Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver. Genome Res. 2010 Aug;20(8):1037–51.

59. Jiang G, Sladek FM. The DNA Binding Domain of Hepatocyte Nuclear Factor 4 Mediates Cooperative, Specific Binding to DNA and Heterodimerization with the Retinoid X Receptor α. Journal of Biological Chemistry. 1997 Jan;272(2):1218–25.

60. Gotoh H, Wood WM, Patel KD, Factor DC, Boshans LL, Nomura T, et al. NG2 expression in NG2 glia is regulated by binding of SoxE and bHLH transcription factors to a Cspg4 intronic enhancer. Glia. 2018 Dec;66(12):2684–99.

61. Singh A, Mahesh A, Noack F, Cardoso de Toledo B, Calegari F, Tiwari VK. Tcf12 and NeuroD1 cooperatively drive neuronal migration during cortical development. Development. 2022 Feb 1;149(3).

62. Subramanian L, Sarkar A, Shetty AS, Muralidharan B, Padmanabhan H, Piper M, et al. Transcription factor Lhx2 is necessary and sufficient to suppress astrogliogenesis and promote neurogenesis in the developing hippocampus. Proceedings of the National Academy of Sciences. 2011 Jul 5;108(27).

63. Dai YS, Cserjesi P, Markham BE, Molkentin JD. The Transcription Factors GATA4 and dHAND Physically Interact to Synergistically Activate Cardiac Gene Expression through a p300-dependent Mechanism. Journal of Biological Chemistry. 2002 Jul;277(27):24390–8.

64. Sakamoto T, Batmanov K, Wan S, Guo Y, Lai L, Vega RB, et al. The nuclear receptor ERR cooperates with the cardiogenic factor GATA4 to orchestrate cardiomyocyte maturation. Nat Commun. 2022 Apr 13;13(1):1991.

65. Stennard FA, Costa MW, Elliott DA, Rankin S, Haast SJP, Lai D, et al. Cardiac T-box factor Tbx20 directly interacts with Nkx2-5, GATA4, and GATA5 in regulation of gene expression in the developing heart. Dev Biol. 2003 Oct;262(2):206–24.

66. Torres-Machorro AL. Homodimeric and Heterodimeric Interactions among Vertebrate Basic Helix–Loop–Helix Transcription Factors. Int J Mol Sci. 2021 Nov 28;22(23):12855.

67. Wang X, Srivastava Y, Jankowski A, Malik V, Wei Y, del Rosario RC, et al. DNA-mediated dimerization on a compact sequence signature controls enhancer engagement and regulation by FOXA1. Nucleic Acids Res. 2018 Jun 20;46(11):5470–86.

68. Zhang Y, Larsen CA, Stadler HS, Ames JB. Structural Basis for Sequence Specific DNA Binding and Protein Dimerization of HOXA13. PLoS One. 2011 Aug 1;6(8):e23069.

69. Charron F, Paradis P, Bronchain O, Nemer G, Nemer M. Cooperative Interaction between GATA-4 and GATA-6 Regulates Myocardial Gene Expression. Mol Cell Biol. 1999 Jun;19(6):4355–65.

70. Butt TR, Walfish PG. Human nuclear receptor heterodimers: opportunities for detecting targets of transcriptional regulation using yeast. Gene Expr. 1996;5(4–5):255–68.

71. Jurata LW, Pfaff SL, Gill GN. The Nuclear LIM Domain Interactor NLI Mediates Homo- and Heterodimerization of LIM Domain Transcription Factors. Journal of Biological Chemistry. 1998 Feb;273(6):3152–7.

72. Gotea V, Visel A, Westlund JM, Nobrega MA, Pennacchio LA, Ovcharenko I. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. Genome Res. 2010 May;20(5):565–77.

73. Georgakopoulos-Soares I, Deng C, Agarwal V, Chan CSY, Zhao J, Inoue F, et al. Transcription factor binding site orientation and order are major drivers of gene regulatory activity. Nat Commun. 2023 Apr 22;14(1):2333.

74. Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, et al. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol Cell. 2016 Apr;62(1):79–91.

75. Biddie SC, John S, Sabo PJ, Thurman RE, Johnson TA, Schiltz RL, et al. Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. Mol Cell. 2011 Jul;43(1):145–55.

76. Pop RT, Pisante A, Nagy D, Martin PCN, Mikheeva LA, Hayat A, et al. Identification of mammalian transcription factors that bind to inaccessible chromatin. Nucleic Acids Res. 2023 Sep 8;51(16):8480–95.

77. Lee DG, Kim YK, Baek KH. The bHLH Transcription Factors in Neural Development and Therapeutic Applications for Neurodegenerative Diseases. Int J Mol Sci. 2022 Nov 11;23(22):13936.

78. Dennis DJ, Han S, Schuurmans C. bHLH transcription factors in neural development, disease, and reprogramming. Brain Res. 2019 Feb;1705:48–65.

79. Imayoshi I, Kageyama R. bHLH Factors in Self-Renewal, Multipotency, and Fate Choice of Neural Progenitor Cells. Neuron. 2014 Apr;82(1):9–23.

80. Kim JA, Shon YH, Lim JO, Yoo JJ, Shin HI, Park EK. MYOD mediates skeletal myogenic differentiation of human amniotic fluid stem cells and regeneration of muscle injury. Stem Cell Res Ther. 2013 Dec 11;4(6):147.

81. Yao J, Zhou J, Liu Q, Lu D, Wang L, Qiao X, et al. Atoh8, a bHLH Transcription Factor, Is Required for the Development of Retina and Skeletal Muscle in Zebrafish. PLoS One. 2010 Jun 3;5(6):e10945.

82. Acharya A, Baek ST, Huang G, Eskiocak B, Goetsch S, Sung CY, et al. The bHLH transcription factor Tcf21 is required for lineage-specific EMT of cardiac fibroblast progenitors. Development. 2012 Jun 15;139(12):2139–49.

83. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018 Feb;172(4):650–65.

84. Bridoux L, Zarrineh P, Mallen J, Phuycharoen M, Latorre V, Ladam F, et al. HOX paralogs selectively convert binding of ubiquitous transcription factors into tissue-specific patterns of enhancer activation. PLoS Genet. 2020 Dec 14;16(12):e1009162.

85. Shiojima I, Komuro I, Oka T, Hiroi Y, Mizuno T, Takimoto E, et al. Context-dependent Transcriptional Cooperation Mediated by Cardiac Transcription Factors Csx/Nkx-2.5 and GATA-4. Journal of Biological Chemistry. 1999 Mar;274(12):8231–9.

86. Vassilev A, Kaneko KJ, Shu H, Zhao Y, DePamphilis ML. TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. Genes Dev. 2001 May 15;15(10):1229–41.

87. Jacquemin P, Hwang JJ, Martial JA, Dollá P, Davidson I. A Novel Family of Developmentally Regulated Mammalian Transcription Factors Containing the TEA/ATTS DNA Binding Domain. Journal of Biological Chemistry. 1996 Sep;271(36):21775–85.

88. Pobbati A V., Hong W. Emerging roles of TEAD transcription factors and its coactivators in cancers. Cancer Biol Ther. 2013 May 27;14(5):390–8.

89. Moroishi T, Hansen CG, Guan KL. The emerging roles of YAP and TAZ in cancer. Nat Rev Cancer. 2015 Feb 16;15(2):73–9.

90. Wang C, Zhu X, Feng W, Yu Y, Jeong K, Guo W, et al. Verteporfin inhibits YAP function through up-regulating 14-3-3σ sequestering YAP in the cytoplasm. Am J Cancer Res. 2016;6(1):27–37.

91. Wang H, Yang Y, Liu J, Qian L. Direct cell reprogramming: approaches, mechanisms and progress. Nat Rev Mol Cell Biol. 2021 Jun 22;22(6):410–24.

92. Bartosovic M, Kabbe M, Castelo-Branco G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. Nat Biotechnol. 2021 Jul 12;39(7):825–35.

93. Bartosovic M, Castelo-Branco G. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. Nat Biotechnol. 2022 Dec 19;

94. Ota M, Sasaki H. Mammalian Tead proteins regulate cell proliferation and contact inhibition as transcriptional mediators of Hippo signaling. Development. 2008 Dec 15;135(24):4059–69.

95. Wu Z, Guan KL. Hippo Signaling in Embryogenesis and Development. Trends Biochem Sci. 2021 Jan;46(1):51–63.

96. Camargo FD, Gokhale S, Johnnidis JB, Fu D, Bell GW, Jaenisch R, et al. YAP1 Increases Organ Size and Expands Undifferentiated Progenitor Cells. Current Biology. 2007 Dec;17(23):2054–60.

97. Yu FX, Zhao B, Guan KL. Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. Cell. 2015 Nov;163(4):811–28.

98. Goode DK, Obier N, Vijayabaskar MS, Lie-A-Ling M, Lilly AJ, Hannah R, et al. Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. Dev Cell. 2016 Mar;36(5):572–87.

99. Yuan Y, Park J, Feng A, Awasthi P, Wang Z, Chen Q, et al. YAP1/TAZ-TEAD transcriptional networks maintain skin homeostasis by regulating cell proliferation and limiting KLF4 activity. Nat Commun. 2020 Dec 19;11(1):1472.

100. Osman I, He X, Liu J, Dong K, Wen T, Zhang F, et al. TEAD1 (TEA Domain Transcription Factor 1) Promotes Smooth Muscle Cell Proliferation Through Upregulating SLC1A5 (Solute Carrier Family 1 Member 5)-Mediated Glutamine Uptake. Circ Res. 2019;124(9):1309–22.

101. Lian I, Kim J, Okazawa H, Zhao J, Zhao B, Yu J, et al. The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. Genes Dev. 2010 Jun 1;24(11):1106–18.

102. Fan W, Jurado-Arjona J, Alanis-Lobato G, Péron S, Berger C, Andrade-Navarro MA, et al. The transcriptional co-activator Yap1 promotes adult hippocampal neural stem cell activation. EMBO J. 2023 Jun 21;42(11).

103. Kim M, Kim T, Johnson RL, Lim DS. Transcriptional Co-repressor Function of the Hippo Pathway Transducers YAP and TAZ. Cell Rep. 2015 Apr;11(2):270–82.

104. 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 Rep. 2013 Dec;5(6):1611–24.

105. Zhou Q, Liu M, Xia X, Gong T, Feng J, Liu W, et al. A mouse tissue transcription factor atlas. Nat Commun. 2017 Apr 21;8(1):15089.

106. 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 Res. 2006 Oct;34(17):4925–36.

107. Ko JY, Oh S, Yoo KH. Functional Enhancers As Master Regulators of Tissue-Specific Gene Regulation and Cancer Development. Mol Cells. 2017;40(3):169–77.

108. Jimenez P, Saner K, Mayhew B, Rainey WE. GATA-6 Is Expressed in the Human Adrenal and Regulates Transcription of Genes Required for Adrenal Androgen Biosynthesis. Endocrinology. 2003 Oct 1;144(10):4285–8.

109. Martin LJ, Taniguchi H, Robert NM, Simard J, Tremblay JJ, Viger RS. GATA Factors and the Nuclear Receptors, Steroidogenic Factor 1/Liver Receptor Homolog 1, Are Key Mutual Partners in the Regulation of the Human 3β-Hydroxysteroid Dehydrogenase Type 2 Promoter. Molecular Endocrinology. 2005 Sep;19(9):2358–70.

FIGURE LEGENDS

Figure 1. CocoTF: a pipeline to identify context-specific co-occurrence of TF motifs. A. CocoTF identification of 'Second Search' TF. Motif enrichment analysis performed as part of cocoTF uses "First Search" TF coordinates within tissue-specific regulatory regions as foreground, and "First Search" TF coordinates within random regions as background. Motif enrichment analysis is performed 100nt either side of the "First Search" TF's motif to search for TF motifs that co-occur with the "First Search" TF motif exclusively in tissue-specific regulatory regions. Co-occurring motifs are labelled as "Second Search" TFs. B. Human embryo. RNA-seq and H3K27ac data were obtained from Gerrard et al. (2016) and Gerrard et al. (2020) (28,30). Black lines connect the organs/tissues to their corresponding number of H3K27ac marked 1kb bins replicated only in one tissue, used as putative tissue-specific enhancers. Details of the thresholding method used to identify tissue-specific H3K27ac 1kb bins used as putative tissue-specific enhancers can be found in the original paper (28). Each tissue is labelled with the same colour code across the manuscript. C. The top 20 results of HOMER motif enrichment analysis of heart ventricle-specific H3K27ac bins searching for known motifs. D. List of TFs from Vaquerizas et al. (2009) (5) represented in the RNA-seq data from Gerrard et al. (2016) (30). All tissues were duplicated except for pituitary, muscle and skin. Raw counts were down sampled based on the 75th percentile, expression of each TF was divided by their median expression across all tissues and log median-normalised expression was used to plot the heatmap. Hoerarchical clustering was performed (k = 10). Ubiquitous TFs, with expression constant across tissues clustered in a large cluster labelled non-specific TF cluster. The remaining clusters were considered tissue-restricted. Ventricle-specific clusters are enclosed by a red rectangle. E. Identification of 'First Search' TFs. Ventricle-restricted TFs, whose recognition motifs (or a motif belonging to a member of the same cluster) are within the top 20 motif enrichment analysis results shown in D. Their corresponding motifs are used as 'First Search' TF motifs.

Figure 2. A comprehensive network of TF co-binding at tissue-specific developmental enhancers. A. Network of co-occurring TF motifs at active tissue-specific enhancers in human embryonic development. A cut-off of $p < 1x10^{-5}$ and presence in > 5% of foreground regions, was used. On the y-axis are the 'First Search' TFs. On the x-axis are TFs with binding sites motifs

enriched within 100nt either side of the tissue-specific TFs ('Second Search' TF), clustered by motif similarity (as seen in Fig S2 and table S3). Colours correspond to the ones assigned in Fig 1B. The size of the bubbles matches the negative log p-value: larger bubbles represented smaller p-values. B. Motif enrichment analysis of GATA4 mouse heart ventricle ChIP-seq validates cocoTF GATA results. Three datasets were analysed and compared to results obtained by cocoTF for validation. Dataset 'peaks' consists of 200nt GATA4 ChIP-seq peaks, dataset "flanks" consists of 100nt flanks either side of any GATA motifs found within GATA4 'peaks' and dataset 'H3K27ac' consists of GATA4 'flanks' overlapping human ventricle-specific H3K27ac 1kb bins converted to the mouse genome. MEF2, TALE, TEAD and TBOX motifs are among the top 10 enriched motifs in all datasets. Heatmap shows enrichment p-values of MEF2, TALE, TEAD and TBOX motifs obtained from motif enrichment analysis of 'peaks, 'flanks', 'cocoTF' and 'H3K27ac' performed by HOMER. Neg log p-values are scaled across experiments, to show patterns of significance rather than absolute values. GATA motifs were excluded from the results as the GATA motifs was not masked in some of the motif enrichment analysis. C. Heatmap showing p-values of the probability of GATA-TEAD, GATA-MEF2, GATA-TBX, GATA-NKX, bHLH-RFX, HOX-PITXEBOX motif pairs co-occurring (within 100nt) in heart enhancers compared to brain and limb specific enhancers (from VISTA). P-values were calculated using a hypergeometric test and significant results ($p < 0.05$) are indicated with an asterisk.

Figure 3. TEAD1 binding occurs mainly at tissue-specific intronic or distal intergenic regions. A. 'First Search' TFs (y-axis) co-occurring with TEAD binding sites in the tissues shown (x-axis) as predicted by cocoTF. B. TEAD *de novo* motifs enriched in adult mouse liver, embryonic mouse ventricle and human primary keratinocyte ChIP-seq peaks. Each motif was used to centre the corresponding ChIP-seq 200nt regions around the peak summits (referred to as 'TEAD1 peak' in the result section) C. Intersection of Liver, Ventricle and Keratinocyte TEAD1 peaks shows a high percentage of tissue-specific binding. D. Annotation of TEAD1 specific peaks in mouse liver, ventricle and primary keratinocytes and overlapping TEAD1 peaks in all three ChIP-seq experiments show a higher percentage of promoters in non-tissue-specific peaks compared to overall TEAD1 binding events. E-H. Top five mouse phenotype and biological process gene ontologies associated with ventricle-specific (E), liver-specific (F), keratinocyte-specific (F) and non-tissue-specific (H) TEAD1 peaks, using the whole mouse genome as background.

Figure 4. TEAD1 binding attenuates enhancer activation by lineage-specific TFs. A. Luciferase activity driven by three ventricle-specific enhancers (enh_PDLIM5, enh_NKX2.6, enh_RAMP1) co-transfected with TEAD1 alone (T1), Gata6 alone (G6), YAP alone, TEAD1 and Gata6 (G6 + T1), TEAD1 and YAP (T1 + YAP), G6 and YAP and G6, T1 and YAP in NIH3T3 cells, normalised against basal enhancer activity driven by an empty pcDNA3 vector (pcD). Co-expression of GATA6, YAP and TEAD1 significantly reduces enhancer activity. B. Luciferase activity driven by two RPE-specific enhancers (enh_PCMTD2, enh_EYA2) co-transfected with TEAD1 alone (T1), CRX alone, YAP alone (YAP), TEAD1 and CRX (CRX + T1), TEAD1 and YAP (T1 + YAP), CRX and YAP and CRX, T1 and YAP in NIH3T3 cells, normalised against basal enhancer activity driven by an empty pcDNA3 vector (pcD). Co-expression of CRX, YAP and TEAD1 reduces enhancer activity down to almost pcDNA3

levels. A+B. Values are presented as the mean +/- SEM of three biological replicates, each obtained from the median of three experimental replicates. Asterisk denotes significant results. Significance was calculated using a one-way ANOVA with multiple comparisons for each of the enhancer constructs, **** = P < 0.0001, *** = P < 0.001, ** = P < 0.01, * = P < 0.05, ns = P > 0.05. C. Luciferase activity driven by enh_MEIS2, enh_RAMP1 and enh_PDLIM5 co-transfected with GATA6 in NIH3T3 cells and treated with verteporfin (VP) or DMSO, normalised against basal enhancer level driven by an empty pcDNA3 vector (pcD). Inhibiting formation of the TEAD/YAP complex increases GATA6-dependent enhancer activation. Values are presented as the mean +/- SEM of two biological replicates, each obtained from the median of three experimental replicates. Asterisk denotes significant results. Significance was calculated using an unpaired t-test, * = P < 0.05. D. Diagram of the wild-type and mutant versions of enh_NKX2.6 in pGL4 plasmid (WT, G6-, T1- and G6-/T1-). E. Luciferase activity driven by wild-type enh_NKX2.6 and enh_NKX2.6 mutants. Each enh_NKX2.6 plasmids was co-transfected with TEAD1 alone (T1), with Gata6 alone (G6) and with TEAD1 and GATA6 (G6 + T1) in NIH3T3 cells and luciferase values normalised against basal enhancer activity driven by an empty pcDNA3 vector (pcD). Compared to the wild-type enhancer, TEAD1 failed to significantly affect Gata6-induced enhancer activity upon mutation of TEAD1 binding site. F. Diagram of the wild-type and mutant versions of enh_EYA2 pGL4 plasmid constructs (WT, CRX-, T1- and CRX-/T1-). G. Luciferase activity driven by wild-type enh_EYA2 and three enh_EYA2 mutants: T1-, CRX/T1- and CRX-. Each of the four _EYA2 plasmids was co-transfected with TEAD1 alone (T1), with Crx alone (CRX) and with TEAD1 and Crx (CRX + T1) in NIH3T3 cells, normalised against basal enhancer activity driven by an empty pcDNA3 vector (pcD). When TEAD1 binding site is mutated, there is no significant inhibition of Crx-induced enhancer activity. E+G. Values are presented as the mean +/- SEM of 3 biological replicates, each obtained from the median of 3 experimental replicates. Asterisk denotes significant results. Significance was calculated using a one-way ANOVA with Tukey test for multiple comparisons of the means calculated for each enhancer plasmid (significance results are shown in black). In addition, unpaired t-tests were used to compare two conditions across enhancers (significance results shown in red/cyan), **** = P < 0.0001, ** = P < 0.01, * = P < 0.05, ns = P > 0.05.

Figure 5. TEAD as an antagonist of tissue-specific enhancer activity during human organogenesis. In progenitor cells, TEAD and YAP complex promotes proliferation (96). At the same time, TEAD binds tissue-specific enhancers within 100nt of the tissue-specific activator binding site and attenuates enhancer activity, reducing the expression of tissue-specific genes and slowing down cell differentiation. Reduced effects of TEAD in differentiated cells, possibly due to an increase in the levels of tissue-specific TFs and/or activation of the Hippo signalling pathway, result in full enhancer activity and sustained activation of tissue-specific genes. This mechanism could be part of a wider role of TEAD/YAP to maintain appropriate pools of progenitor-like cells to achieve correct organ size during organogenesis.

Supplementary Data is available in additional files.

Figure 1

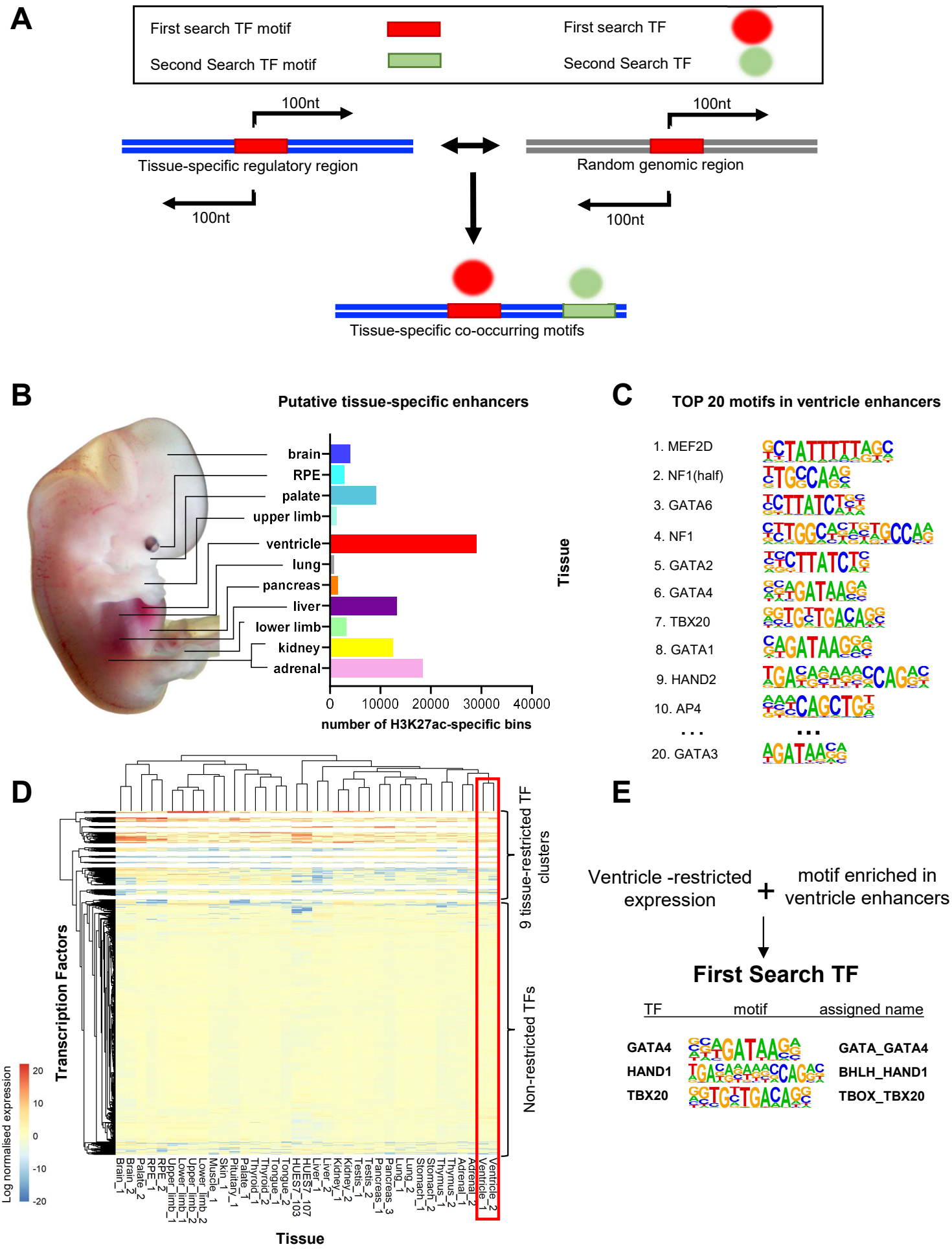

A
First search TF motif
Second Search TF motif
First search TF
Second Search TF
100nt
Tissue-specific regulatory region
100nt
100nt
Random genomic region
100nt
Tissue-specific co-occurring motifs
B
Putative tissue-specific enhancers
brain
RPE
palate
upper limb
ventricle
lung
pancreas
liver
lower limb
kidney
adrenal
Tissue
0
10000
20000
30000
40000
number of H3K27ac-specific bins
C
TOP 20 motifs in ventricle enhancers
1. MEF2D
2. NF1(half)
3. GATA6
4. NF1
5. GATA2
6. GATA4
7. TBX20
8. GATA1
9. HAND2
10. AP4
...
20. GATA3
D
Transcription Factors
Log normalised expression
20
10
0
-10
-20
9 tissue-restricted TF clusters
Non-restricted TFs
Brain_1
Brain_2
Palate_2
RPE_1
RPE_2
Upper_limb_1
Upper_limb_2
Lower_limb_1
Lower_limb_2
Muscle_1
Skin_1
Pituitary_1
Palate_1
Thyroid_1
Thyroid_2
Tongue_1
Tongue_2
HUES7_103
HUES7_107
Liver_1
Liver_2
Kidney_1
Kidney_2
Testis_1
Testis_2
Pancreas_1
Pancreas_3
Lung_1
Lung_2
Stomach_1
Stomach_2
Thymus_1
Thymus_2
Adrenal_1
Adrenal_2
Ventricle_1
Ventricle_2
Tissue
E
Ventricle -restricted expression
+
motif enriched in ventricle enhancers
First Search TF
TF
motif
assigned name
GATA4
GATA_GATA4
HAND1
BHLH_HAND1
TBX20
TBOX_TBX20

Figure 2

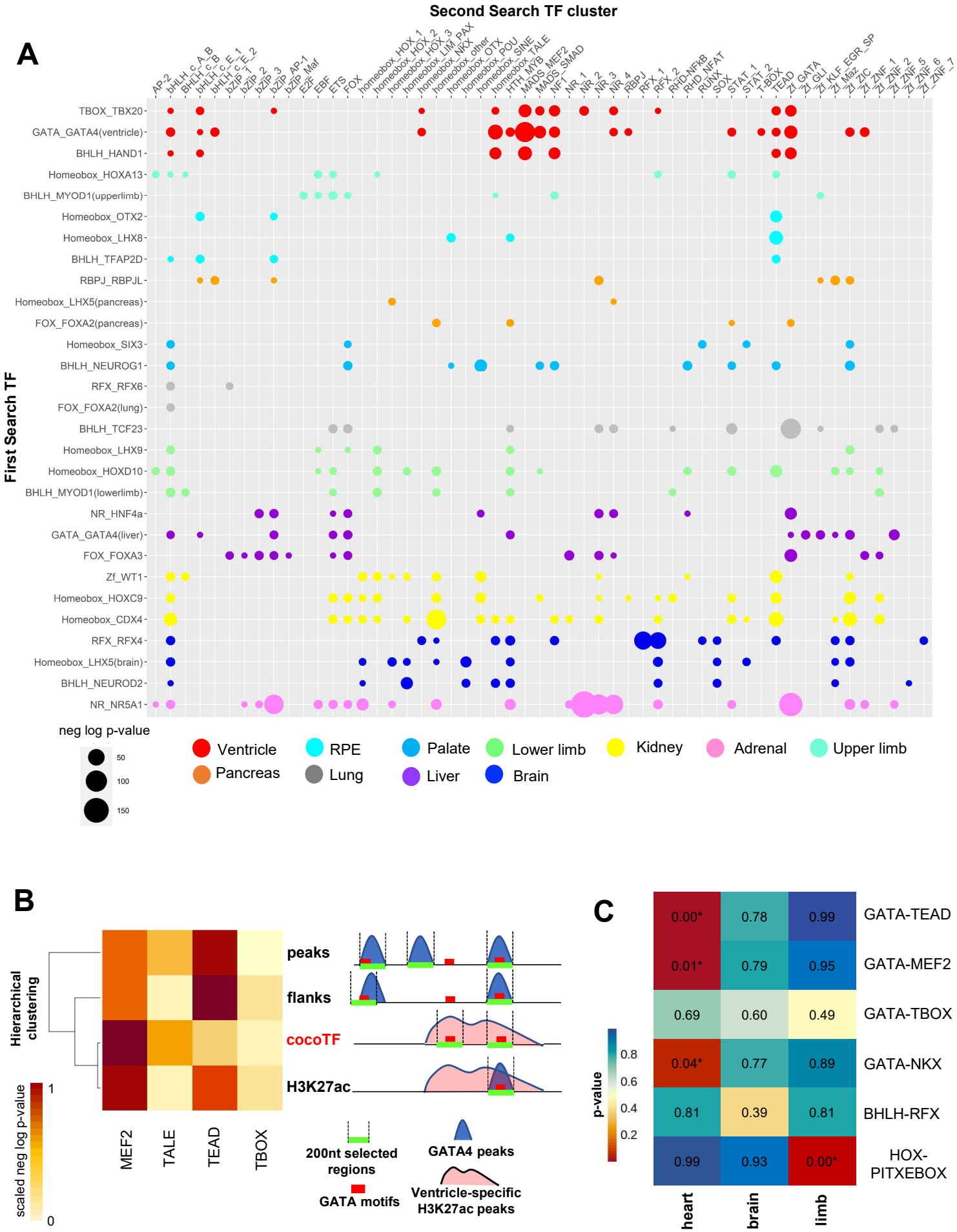

A
Second Search TF cluster
First Search TF
AP-2
bHLH_c_A_B
bHLH_c_B
bHLH_c_E_1
bHLH_c_E_2
bZIP_1
bZIP_2
bZIP_3
bZIP_AP-1
bZIP_Maf
E2F
EBF
ETS
FOX
homeobox_HOX_1
homeobox_HOX_2
homeobox_HOX_3
homeobox_LIM_PAX
homeobox_NKX
homeobox_other
homeobox_OTX
homeobox_POU
homeobox_SINE
HTH_MYB
homeobox_TALE
MADS_MEF2
MADS_SMAD
NF1
NR_1
NR_2
NR_3
NR_4
RBPJ
RFX_1
RFX_2
RHD-NFkB
RHD_NFAT
RUNX
SOX
STAT_1
STAT_2
T-BOX
TEAD
Zf_GATA
Zf_GLI
Zf_KLF_EGR_SP
Zf_Maz
Zf_ZIC
Zf_ZNF_1
Zf_ZNF_2
Zf_ZNF_5
Zf_ZNF_6
Zf_ZNF_7
TBOX_TBX20
GATA_GATA4(ventricle)
BHLH_HAND1
Homeobox_HOXA13
BHLH_MYOD1(upperlimb)
Homeobox_OTX2
Homeobox_LHX8
BHLH_TFAP2D
RBPJ_RBPJL
Homeobox_LHX5(pancreas)
FOX_FOXA2(pancreas)
Homeobox_SIX3
BHLH_NEUROG1
RFX_RFX6
FOX_FOXA2(lung)
BHLH_TCF23
Homeobox_LHX9
Homeobox_HOXD10
BHLH_MYOD1(lowerlimb)
NR_HNF4a
GATA_GATA4(liver)
FOX_FOXA3
Zf_WT1
Homeobox_HOXC9
Homeobox_CDX4
RFX_RFX4
Homeobox_LHX5(brain)
BHLH_NEUROD2
NR_NR5A1
neg log p-value
50
100
150
Ventricle
Pancreas
RPE
Lung
Palate
Liver
Lower limb
Brain
Kidney
Adrenal
Upper limb
B
Hierarchical clustering
scaled neg log p-value
1
0
MEF2
TALE
TEAD
TBOX
peaks
flanks
cocoTF
H3K27ac
200nt selected regions
GATA motifs
GATA4 peaks
Ventricle-specific H3K27ac peaks
C
p-value
0.8
0.6
0.4
0.2
0.00*
0.78
0.99
GATA-TEAD
0.01*
0.79
0.95
GATA-MEF2
0.69
0.60
0.49
GATA-TBOX
0.04*
0.77
0.89
GATA-NKX
0.81
0.39
0.81
BHLH-RFX
0.99
0.93
0.00*
HOX-PITXEBOX
heart
brain
limb

Figure 3

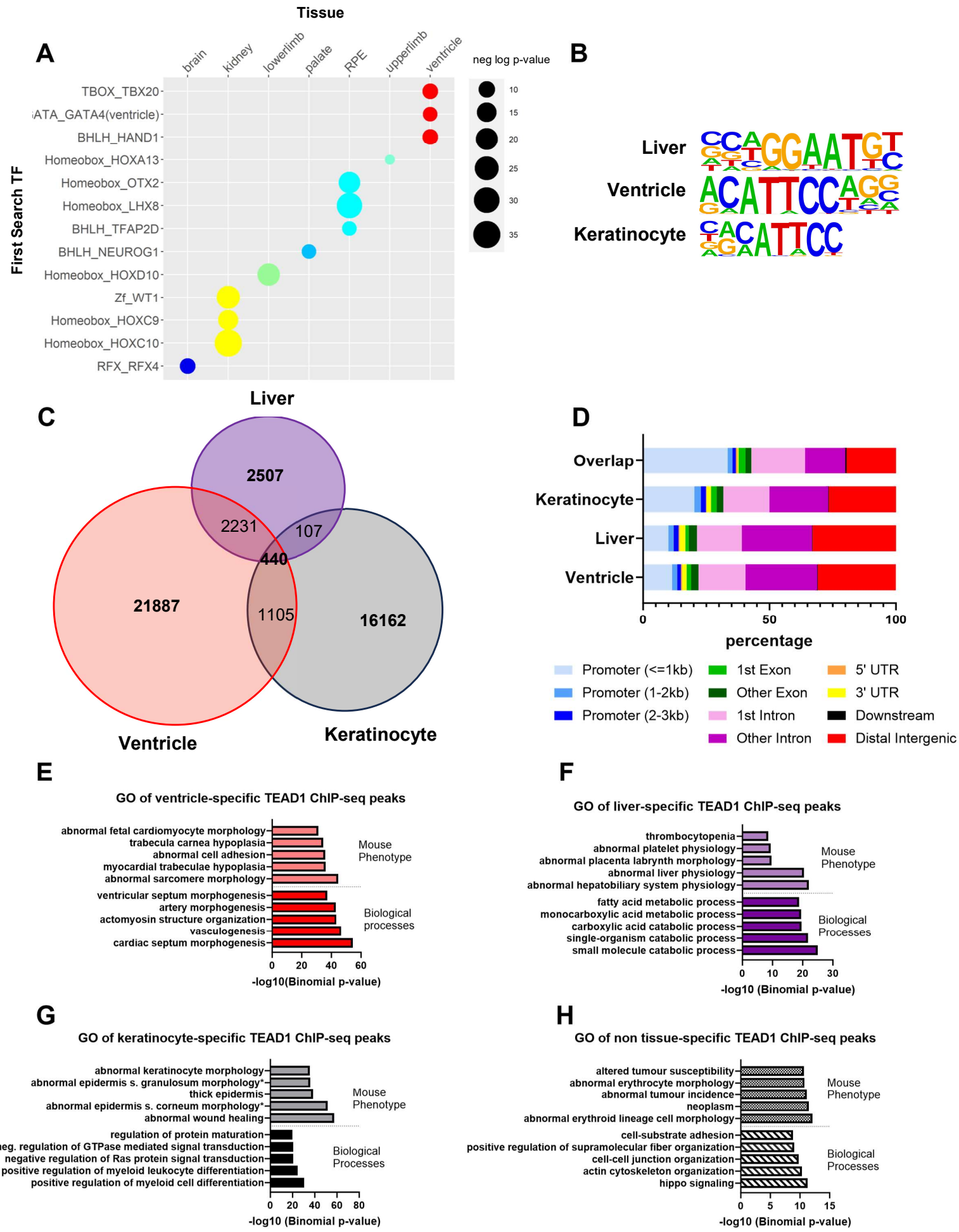

A
Tissue
brain
kidney
lowerlimb
palate
RPE
upperlimb
ventricle
neg log p-value
10
15
20
25
30
35
First Search TF
TBOX_TBX20
GATA_GATA4(ventricle)
BHLH_HAND1
Homeobox_HOXA13
Homeobox_OTX2
Homeobox_LHX8
BHLH_TFAP2D
BHLH_NEUROG1
Homeobox_HOXD10
Zf_WT1
Homeobox_HOXC9
Homeobox_HOXC10
RFX_RFX4
B
Liver
Ventricle
Keratinocyte
C
Liver
2507
2231
107
440
21887
1105
16162
Ventricle
Keratinocyte
D
Overlap
Keratinocyte
Liver
Ventricle
0
50
100
percentage
Promoter (<=1kb)
Promoter (1-2kb)
Promoter (2-3kb)
1st Exon
Other Exon
1st Intron
Other Intron
5' UTR
3' UTR
Downstream
Distal Intergenic
E
GO of ventricle-specific TEAD1 ChIP-seq peaks
abnormal fetal cardiomyocyte morphology
trabecula carnea hypoplasia
abnormal cell adhesion
myocardial trabeculae hypoplasia
abnormal sarcomere morphology
ventricular septum morphogenesis
artery morphogenesis
actomyosin structure organization
vasculogenesis
cardiac septum morphogenesis
Mouse Phenotype
Biological processes
0
20
40
60
-log10(Binomial p-value)
F
GO of liver-specific TEAD1 ChIP-seq peaks
thrombocytopenia
abnormal platelet physiology
abnormal placenta labrynth morphology
abnormal liver physiology
abnormal hepatobiliary system physiology
fatty acid metabolic process
monocarboxylic acid metabolic process
carboxylic acid catabolic process
single-organism catabolic process
small molecule catabolic process
Mouse Phenotype
Biological Processes
0
10
20
30
-log10 (Binomial p-value)
G
GO of keratinocyte-specific TEAD1 ChIP-seq peaks
abnormal keratinocyte morphology
abnormal epidermis s. granulosum morphology*
thick epidermis
abnormal epidermis s. corneum morphology*
abnormal wound healing
regulation of protein maturation
neg. regulation of GTPase mediated signal transduction
negative regulation of Ras protein signal transduction
positive regulation of myeloid leukocyte differentiation
positive regulation of myeloid cell differentiation
Mouse Phenotype
Biological Processes
0
20
40
60
80
-log10 (Binomial p-value)
H
GO of non tissue-specific TEAD1 ChIP-seq peaks
altered tumour susceptibility
abnormal erythrocyte morphology
abnormal tumour incidence
neoplasm
abnormal erythroid lineage cell morphology
cell-substrate adhesion
positive regulation of supramolecular fiber organization
cell-cell junction organization
actin cytoskeleton organization
hippo signaling
Mouse Phenotype
Biological Processes
0
5
10
15
-log10 (Binomial p-value)

Figure 4

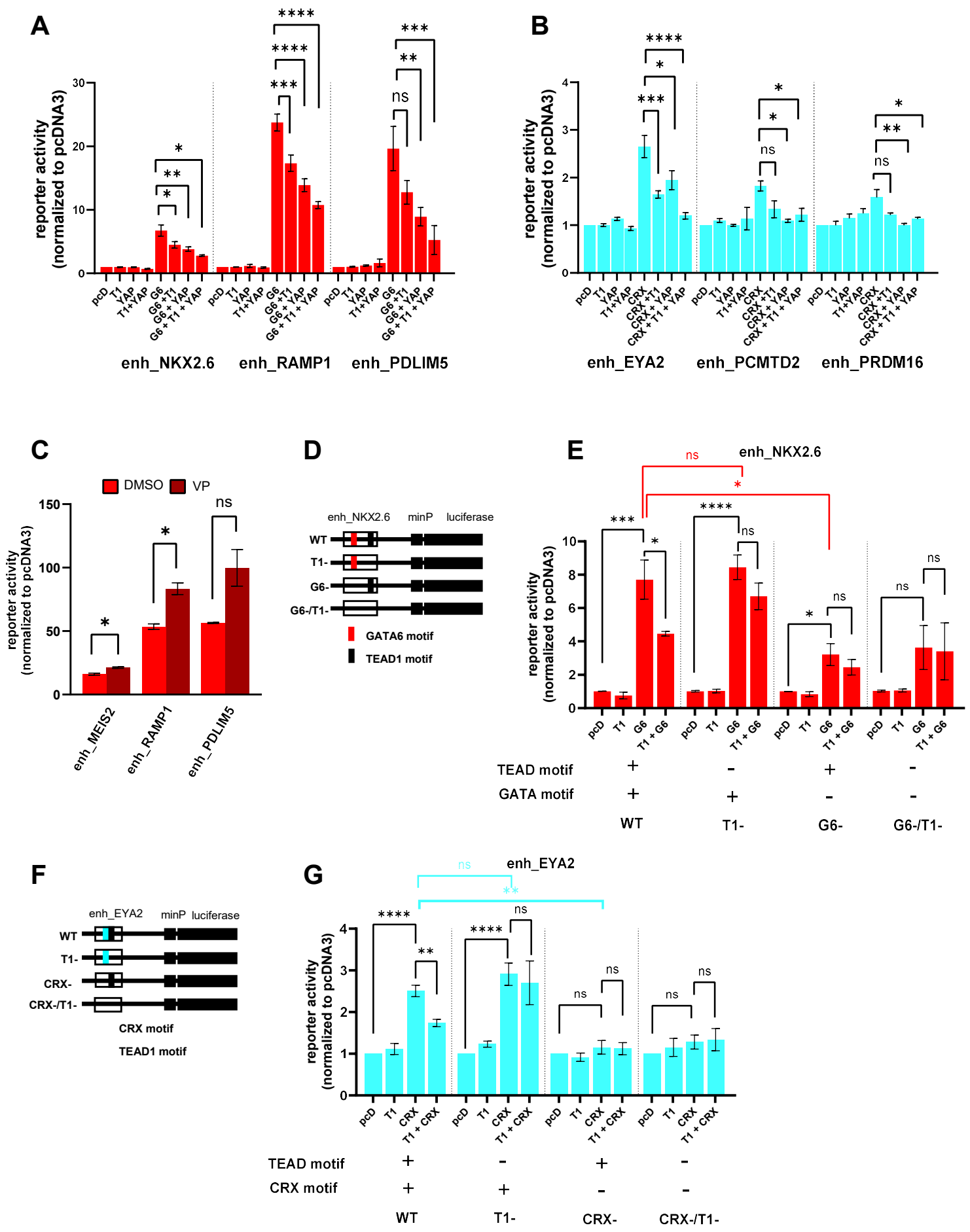
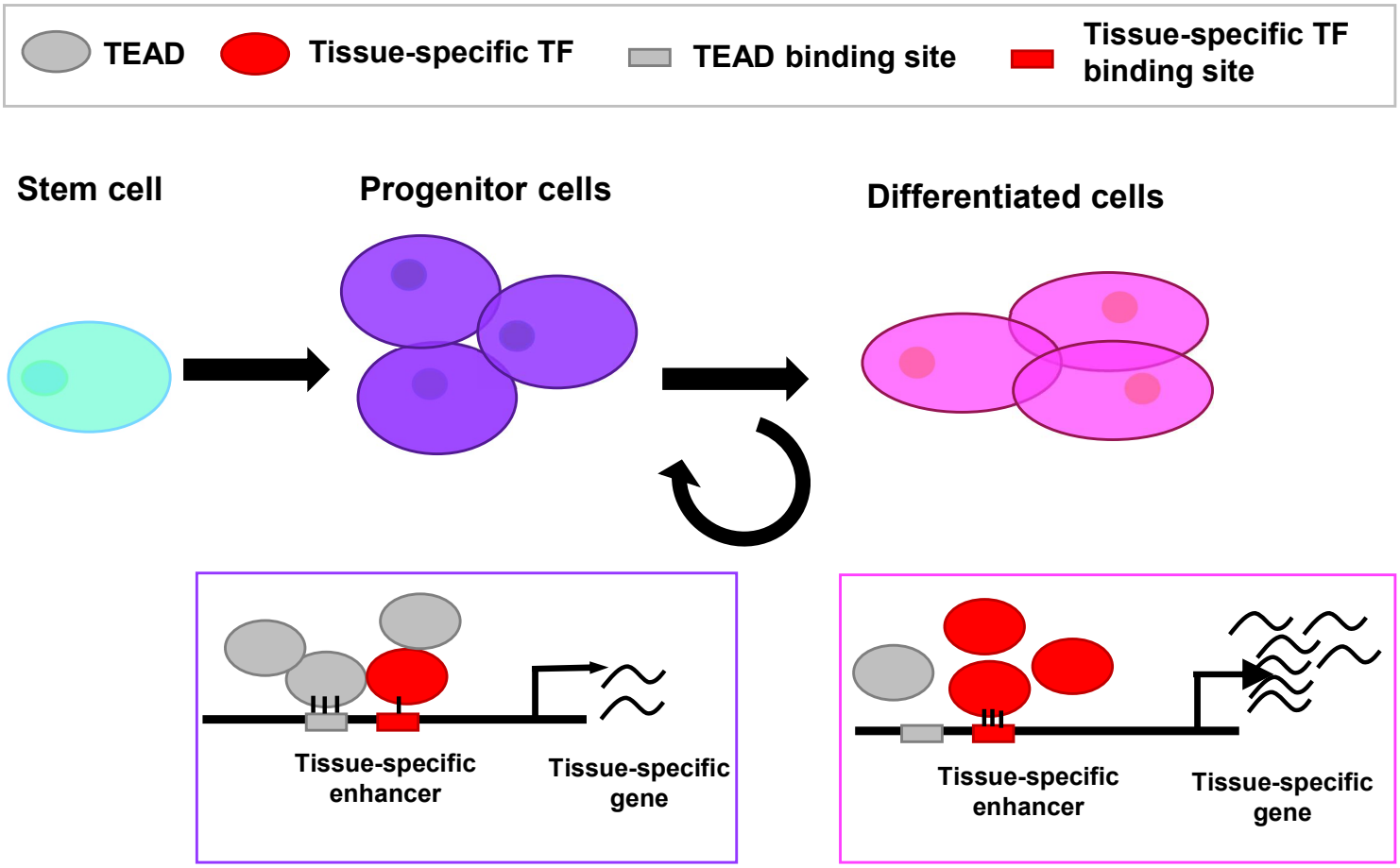

A
reporter activity (normalized to pcDNA3)
pcD
T1
YAP
T1+YAP
G6
G6 +T1
G6 + YAP
G6 + T1 + YAP
enh_NKX2.6
enh_RAMP1
enh_PDLIM5
B
CRX
CRX +T1
CRX + YAP
CRX + T1 + YAP
enh_EYA2
enh_PCMTD2
enh_PRDM16
C
DMSO
VP
enh_MEIS2
D
enh_NKX2.6
minP
luciferase
WT
T1-
G6-
G6-/T1-
GATA6 motif
TEAD1 motif
E
enh_NKX2.6
TEAD motif
GATA motif
F
enh_EYA2
CRX-
CRX-/T1-
CRX motif
TEAD1 motif
G
enh_EYA2
CRX motif

Figure 5


TEAD
Tissue-specific TF
TEAD binding site
Tissue-specific TF binding site
Stem cell
Progenitor cells
Differentiated cells
Tissue-specific enhancer
Tissue-specific gene
Tissue-specific enhancer
Tissue-specific gene