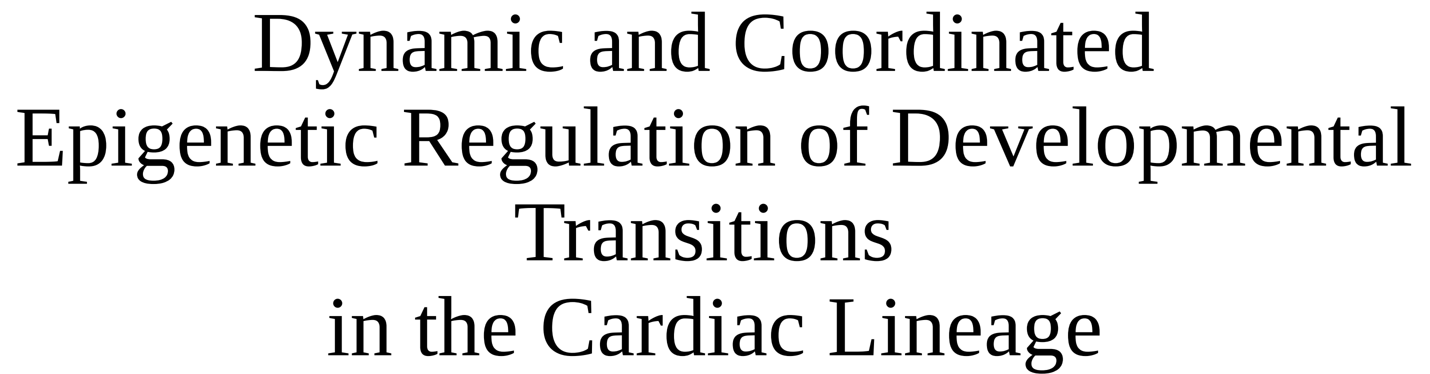
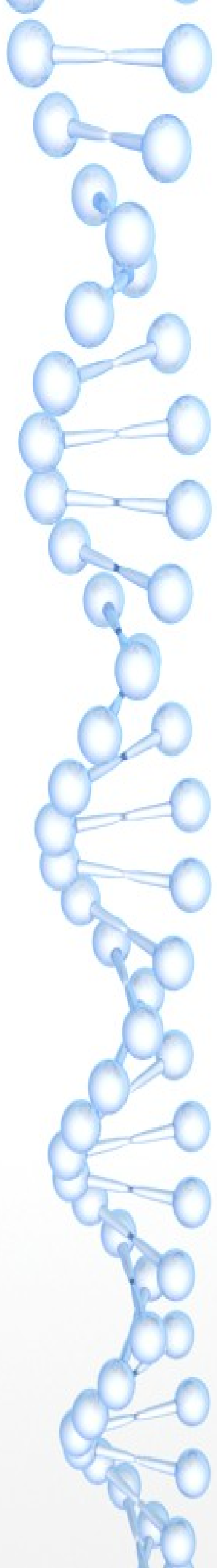
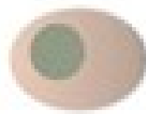
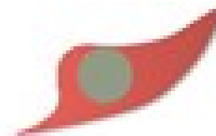
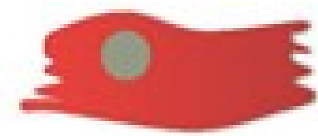

Dynamic and Coordinated
Epigenetic Regulation of Developmental
Transitions
in the Cardiac Lineage

Stages

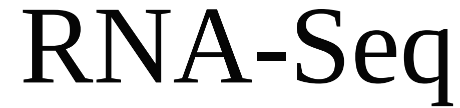
- ESC - undifferentiated embryonic stem cells expressing pluripotency genes (Pou5f1/Oct4, Nanog)
- MES - cells expressing mesodermal markers (Mesp1 , Brachyury)
- CP cells expressing cardiac transcription factors (Nkx2–5, Tbx5, Isl1), but not yet beating;
- CM functional cardiomyocytes with cardiomyocyte-specific gene expression (Myh6, Myh7).



Stages

RNA-Seq



13.500 – genes expressed in experiment

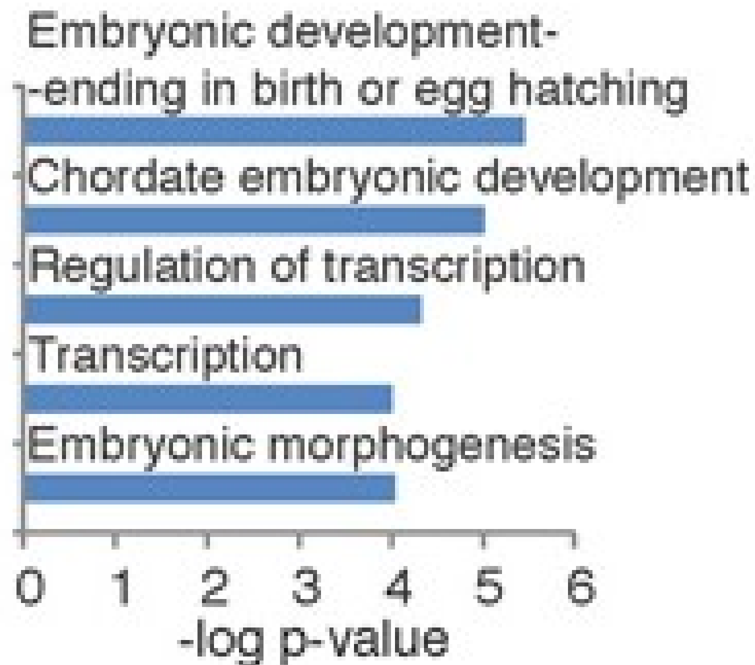
- reads per kilobase per million [RPKM] > 1

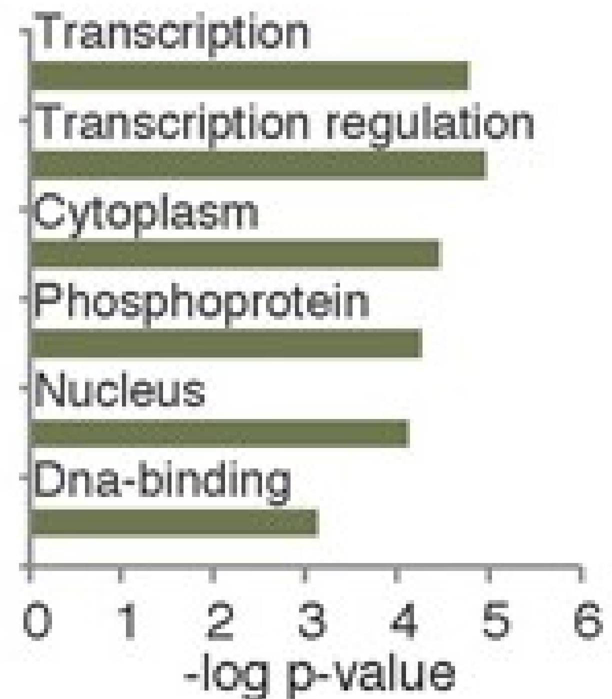
- microRNA show stage specific expression
- key ESC miRNAs, such as the miR290 cluster
- cardiac miRNAs (including miR-1, miR-208, and miR-143) in the CM stage

- Ensembl-annotated lncRNAs
- lncRNAs show striking stage-specific expression in our differentiation system
- lncRNAs identified in our data are significantly correlated in expression with their neighboring genes compared to randomly selected neighboring protein coding genes

GO Ternms near lncRNA

E





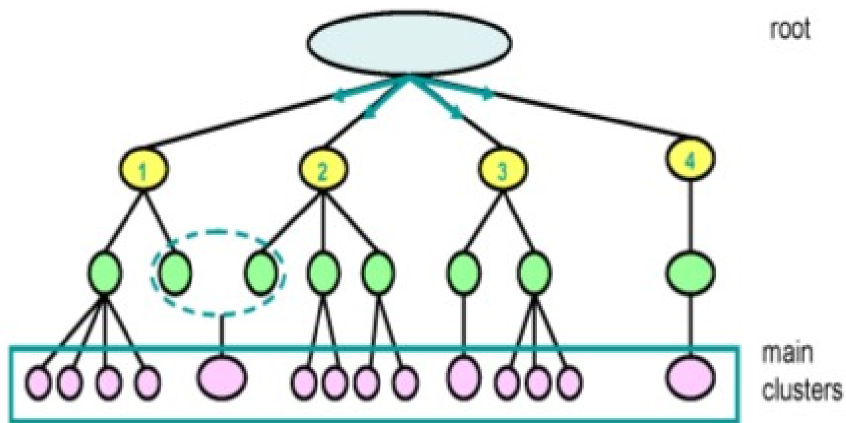
Chromatin state dynamics during cardiac differentiation

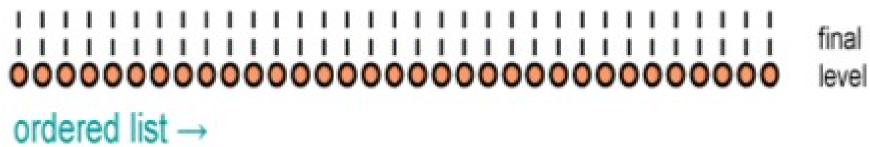
- H3K27me3 (inactive promoters)
- H3K4me3 (active promoters)
- H3K4me1 (associated with promoters and enhancers)
- H3K27ac (associated with promoters and enhancers)
- Binding of RNA polymerase II phosphorylated at serine 5 (RNAP), which is enriched at transcriptional start sites (TSS) (necessary for the initiation of transcription)

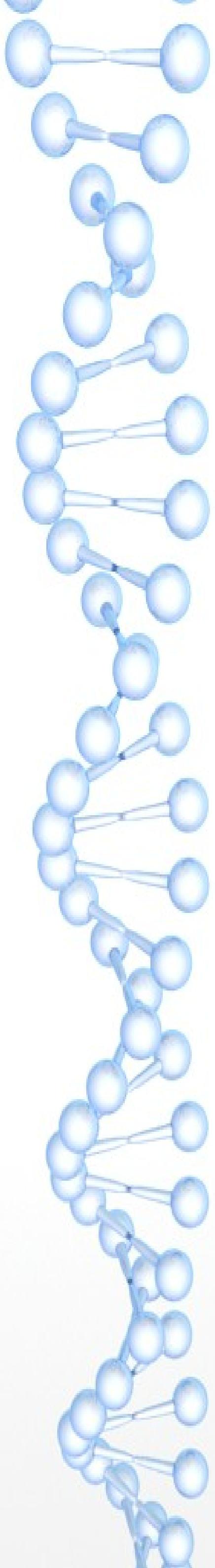
- To identify gene promoters with similar patterns, we performed unsupervised clustering of ChIP signal 2kb around the TSS of each gene
- A spearman distance metric was calculated as one minus the spearman correlation

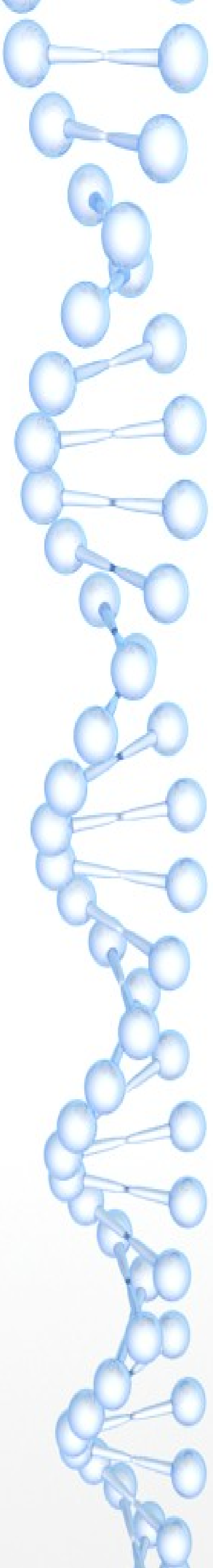
- Hybrid clustering method, Hierarchical Ordered Partitioning And Collapsing Hybrid (HOPACH), which is a hierarchical tree of clusters.
- The methodology combines the strengths of both partitioning (or divisive) and agglomerative clustering methods.
- At each node, a cluster is split into two or more smaller clusters with an enforced ordering of the clusters.

Clustering example









HOPACH-PAM

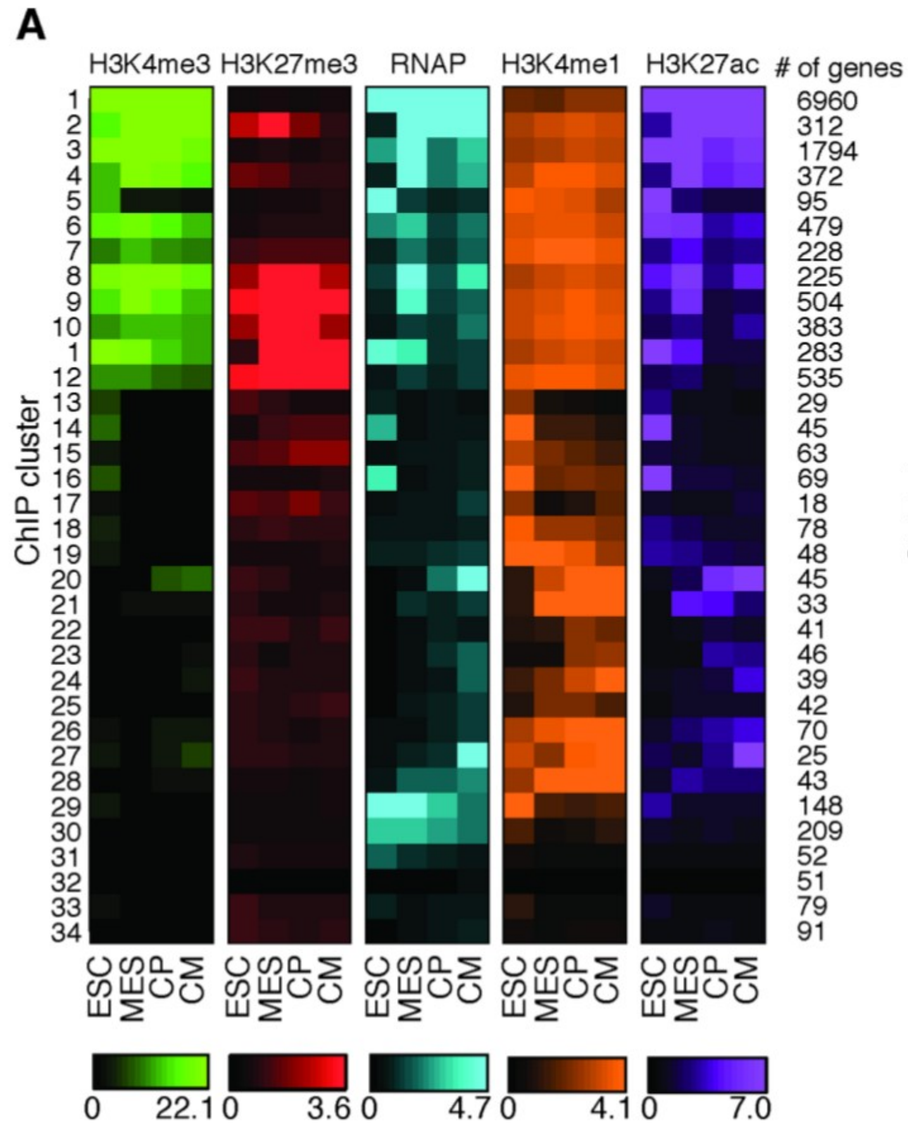
- Initial level of tree
- Step I: Next level of tree
- Step II: Visualization of an ordered distance matrix.
- Step III: Possibly collapse similar clusters

Median Split Silhouette

- For each gene j, calculate the average dissimilarity a_j of gene j with other elements of its cluster.
- For each gene j and each cluster l to which it does not belong, calculate the average dissimilarity b_{jl} of gene j with the members of cluster l. Let $b_j = \min b_{jl}$.

$$S = \frac{b_j - a_j}{\min(b_j, a_j)}$$

Clusters



Dynamic chromatin states correlate with distinct expression patterns

B

ChIP cluster

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

A B C D E F G H I J K L M N O P Q R S

Expression cluster

χ^2 residual

4 2 0 -2 -4

- Our promoter clustering revealed a group of genes that showed enrichment for H3K4me1 enrichment prior to enrichment of H3K4me3 and RNA Pol II and transcriptional activation.

A Novel Chromatin State Transition during CM Differentiation

D

Chromatin Pattern

H3K4me1 H3K4me3

# of genes

Cardiac genes

22

16

159

151

253

Group I

Group II

Group III

ESC MES CP CM

Depleted Enriched

E

H3K4me1 H3K4me3 Expression

Actc1

Ttn

Myh6

Tnni1

Ckm

Ckmt2

Myt1

Mrc1

Group I

Group II

Group III

ESC MES CP CM

A Novel Chromatin State Transition during CM Differentiation

- Group I gained H3K4me1 prior to H3K4me3 enrichment and transcriptional activation and was enriched exclusively for cardiovascular genes including those encoding contractile proteins associated with terminal differentiation and cardiomyocyte function. H3K4me1 was often maintained at these TSSs upon H3K4me3 enrichment and gene activation.
- Group II gained H3K4me1 over time, but failed to gain H3K4me3 or robust expression and included muscle lineage genes such as Ckm , Ckmt2 , and Tcap, whose expression is associated with cardiomyocyte maturation.
- Group III genes transiently gained H3K4me1 at specific stages, but showed no H3K4me3 enrichment during differentiation. These genes were not expressed above background levels throughout differentiation and function in non-cardiac lineages.
- Group II and III genes did not acquire H3K27me3 suggesting Polycomb-independent silencing.

Enhancer activity correlates with cardiac specific programs

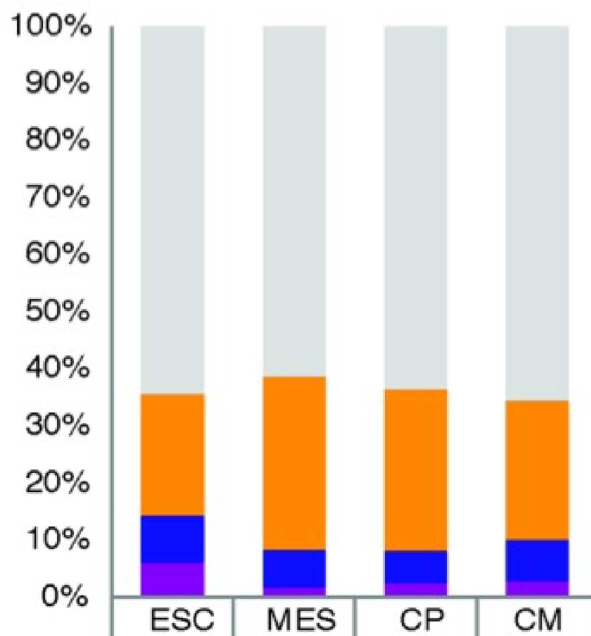
- H3K4me1 and H3K27ac demarcate enhancer elements in a wide range of cell types
- Using these modifications, we identified 81,497 putative distal enhancer regions
- For example, we identified networks for the GATA family of transcription factors that include GATA4, GATA5, and GATA6 in CP and CM stages. The GATA factors are known to orchestrate many developmental processes.

Enhancer activity correlates with cardiac specific programs

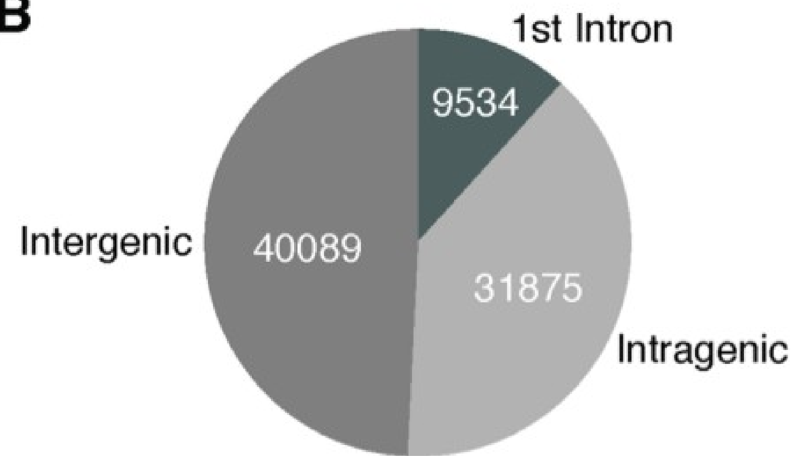
- We find that RNAP is highly enriched at active enhancers, consistent

  with transcription initiation at these regulatory elements

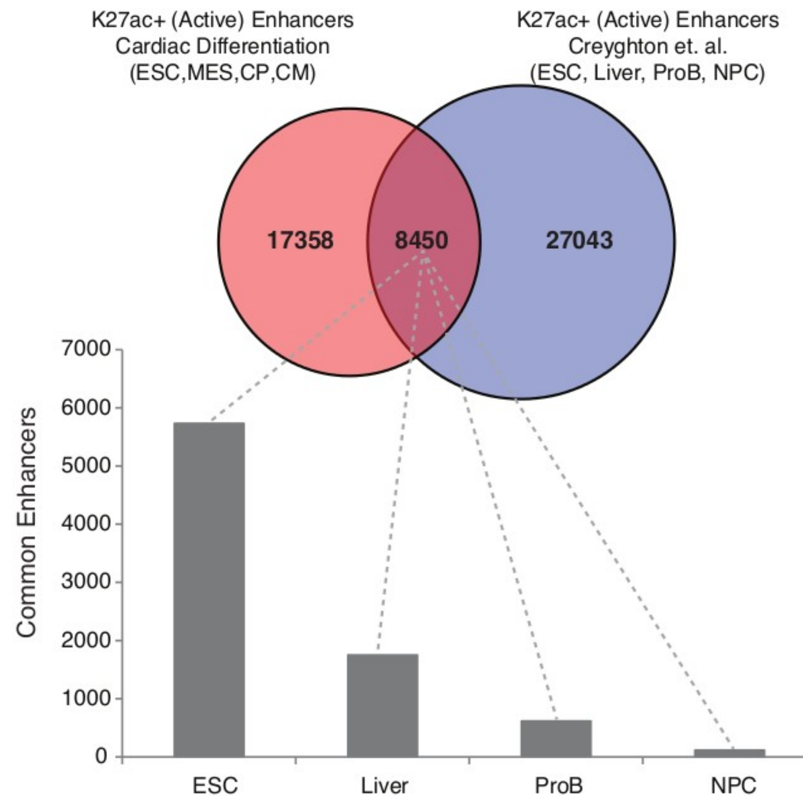
Enhancer activity



	ESC	MES	CP	CM
Unmarked	52473	49827	51843	53320
K4me1 Only	17227	24788	22914	19801
K27ac + K4me1	6803	5439	4641	6092
K27ac Only	4994	1443	2099	2284

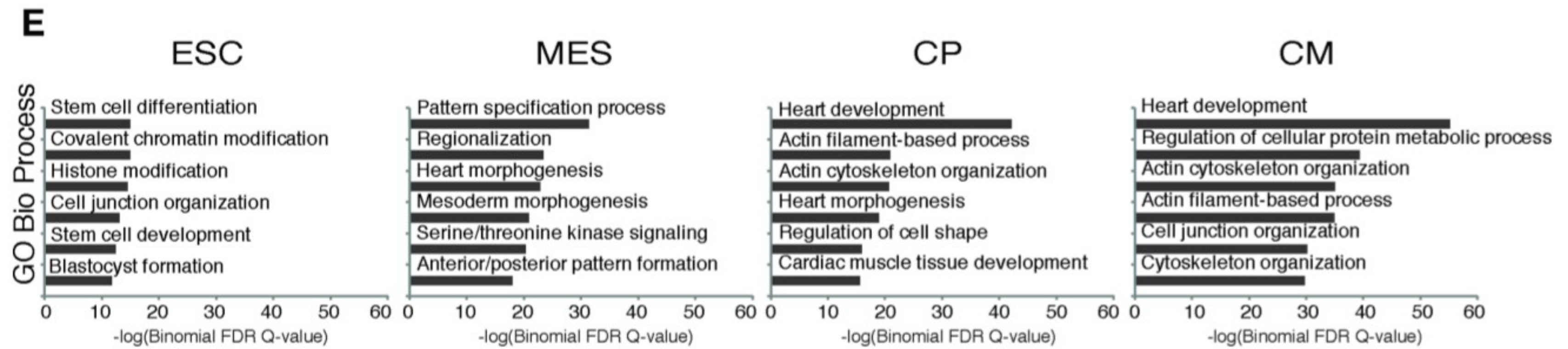


Enhancer intersection

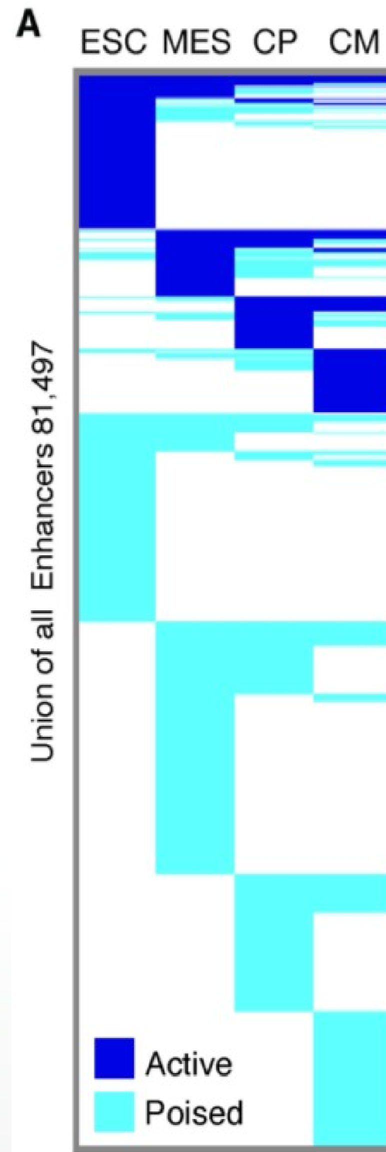


- We classified our enhancers as active (H3K27ac+, H3K4me1+/−) or poised (H3K4me1+ only) at each stage of differentiation

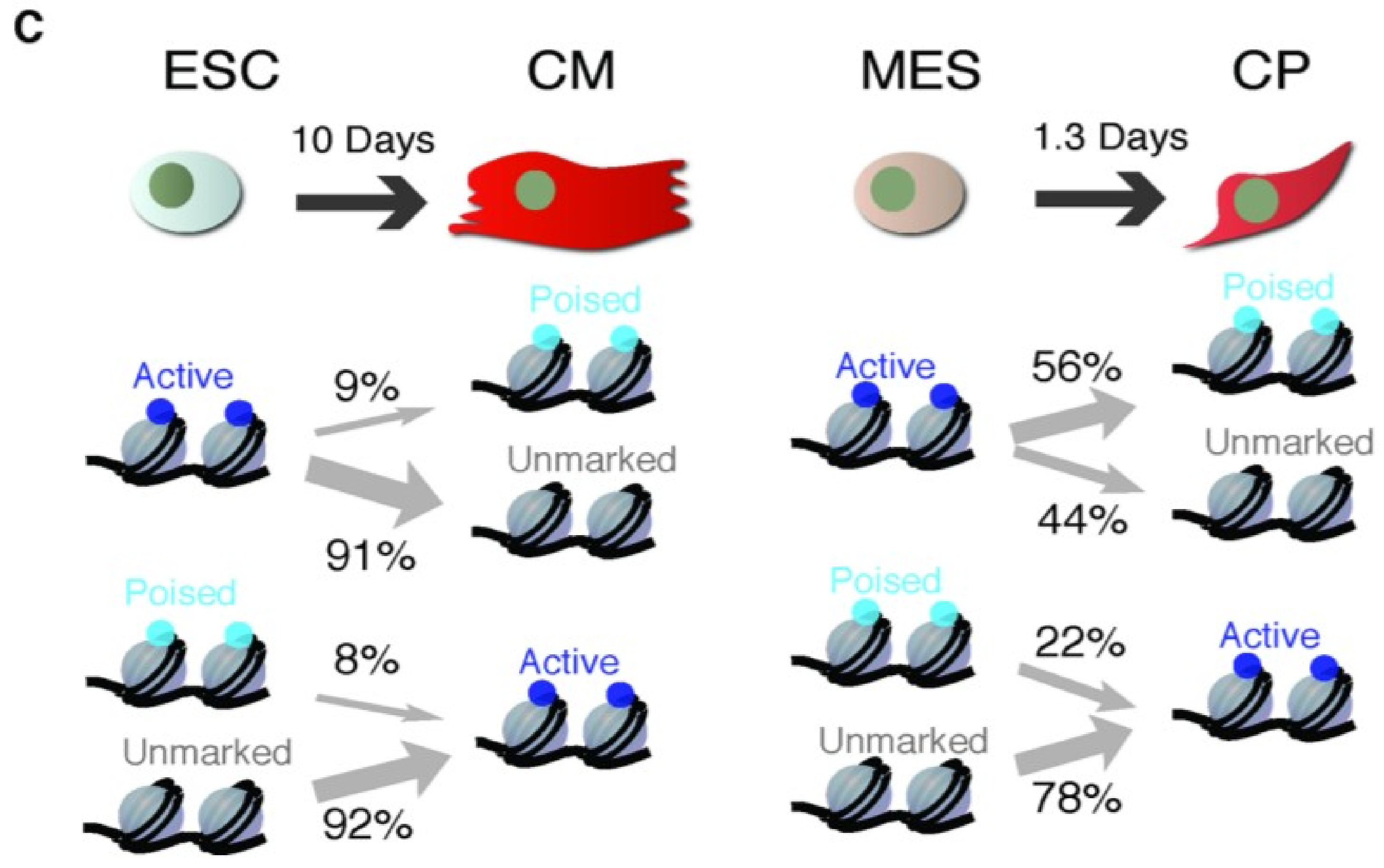
Gene associated with active enhancer



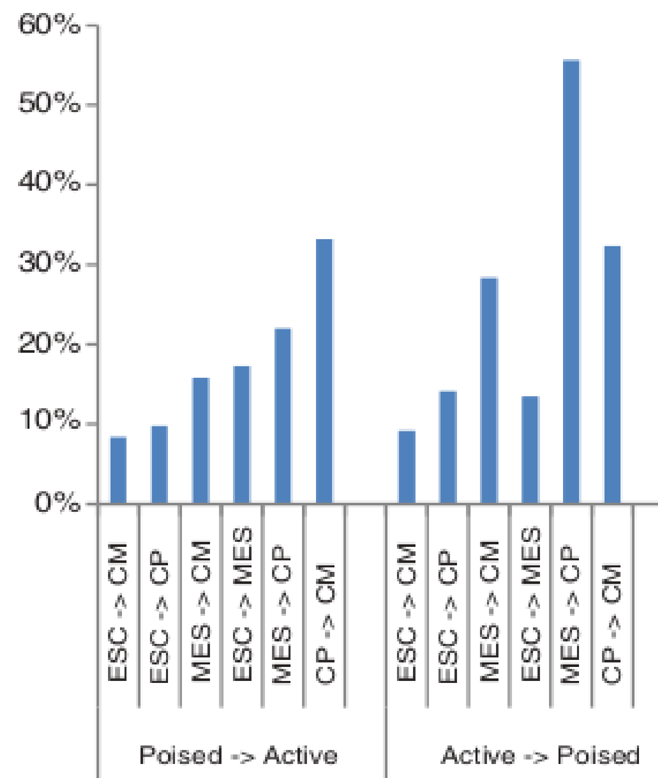
Enhancer transitions during CM differentiation



Enhancer transitions during CM differentiation



Enhancer transitions during CM differentiation

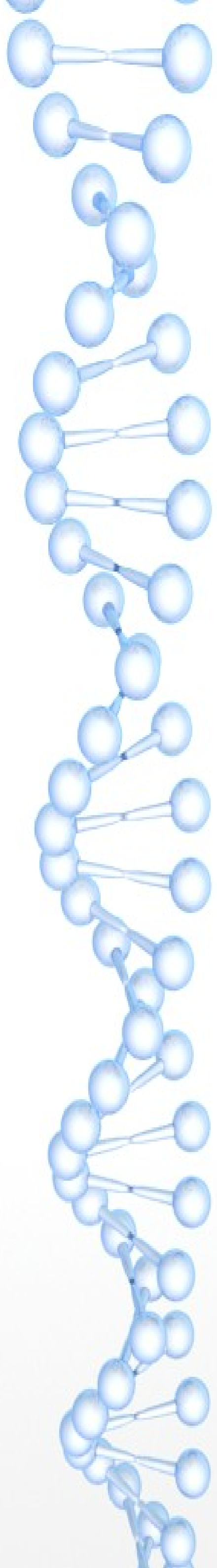


Integrating enhancers into gene networks

- We hypothesized that motifs for TFs that drive cardiac development would be enriched in active enhancers.
- Reasoning that TFs bind open chromatin regions we developed an algorithm to find depressions in the H3K27ac chromatin profile at active enhancer regions and used these regions to search for motifs

Dip Finding Algorithm

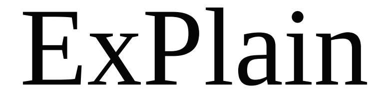
- To identify H3K27ac depressions or “dips” in active enhancers, the summed reads per million (RPM) of the H3K27ac ChIP replicates were used as input for each stage at a resolution of 25 bp.
-



- Slope Length
- Dips consist of a downward slope to the left of the minimum and an upward slope to the right of the minimum. If the slope length is set to N, then all passing dips must have a steady downward slope at least N bins long and a steady upward slope at least N bins long. A ‘‘steady’’ downward slope only requires that once a downward slope has been detected, an upward slope must not be detected for at least N bins; the slope profile is allowed to level-off temporarily. Similarly, a ‘‘steady’’ upward slope requires that once an upward slope has been detected, a downward slope must not be detected for at least N bins. A slope length of 5 was used.
- Dip Size Filter
- The dip size was measured as the height from the minimum of the dip to the lowest height at which the dip is interrupted. The dip is interrupted where the downward slope to the left of the minimum or the upward slope to the right of the minimum ceases to exist. A minimum dip size of 0.8 was required.
- Returned Dip Region Height
- The region returned as the location of the dip was a region centered around the minimum of the dip such that the height from the minimum to the height at both ends of the returned region was at least 0.3. If this region was longer than 200bp, a 200bp region centered at the minimum was returned instead.
- Zero Count Limit
- If in the process of searching for a dip, a stretch at least 250bp long (10 bins) for which the chromatin profile was zero, the dip did not pass.

Dip Ranking

- The chromatin depression or dip profiles were ranked by the height of the minimum of the dip, with the lowest minima ranking first. The total number of dips found in a given stage roughly correlated with the total number of active enhancers from that stage.
- The top 3500 dips of each stage were then selected for differential analysis between the stages. Selecting a fixed number of dips from each stage prevents biases in the p value resulting from certain stages having more dips (the larger the input data size, the stronger the p values).

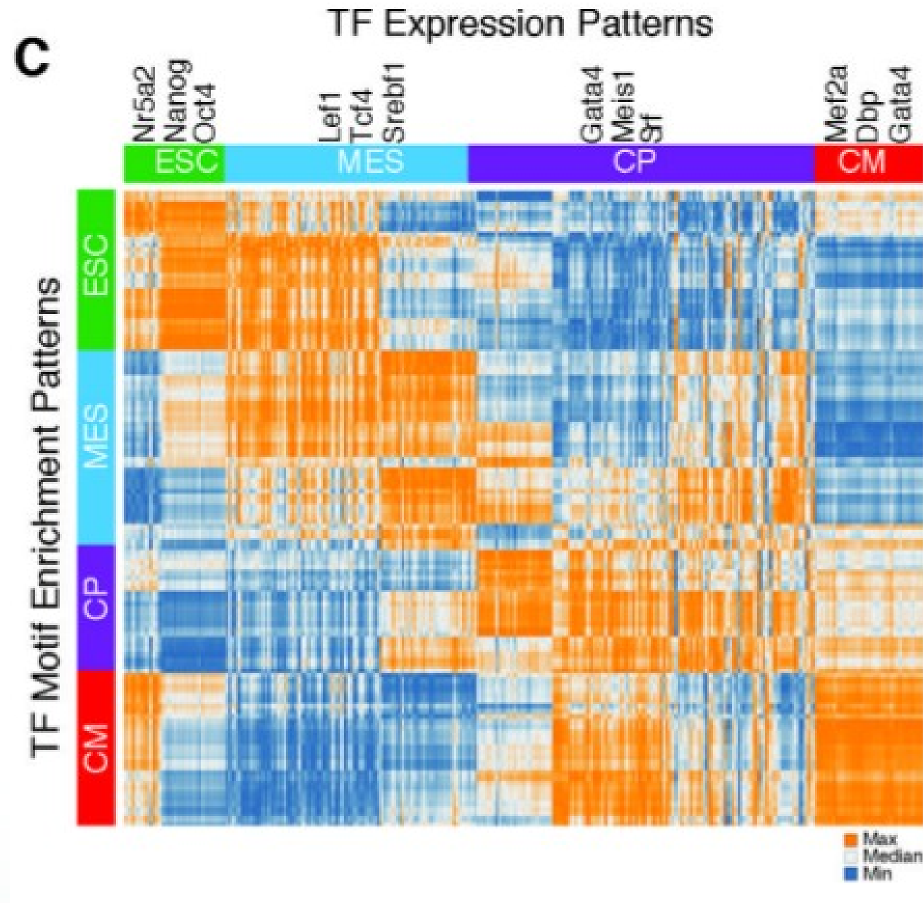


ExPlain

- ExPlain from BioBase (BioBase Biological Databases, Beverly, MA) was used to identify motifs in the dips from each stage. The profile of motif position-weight-matrices used to scan the dips was “vertebrate non-redundant,” and only high specificity matrices were used

- We compiled a list of TFs known to bind these highly conserved motifs and found strong correlations between TF expression and motif enrichment at each stage
- RPKM values of the 264 transcription factors known to bind the 124 enriched motifs
- The signed –log10 of the p value for motifs (124/196) over and under-enriched relative to the background in a least one stage of the time course

Pearson correlation matrix



- The nearest gene associated with the genomic interval, corresponding to the top 3500 H3K27ac dips containing an enriched motif, were identified.
- Target genes displaying a positive expression correlation (Pearson) with motif associated TF expression were used to compile a non-redundant target gene list for discovery of target gene networks using Ingenuity Pathway Analysis