Cell

A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History

Guillaume Junion,[1,3] Mikhail Spivakov,[1,2,3] Charles Girardot,[1] Martina Braun,[1] E. Hilary Gustafson,[1] Ewan Birney,[2] and Eileen E.M. Furlong[1,*]

[1]Genome Biology Unit, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany
[2]European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[3]These authors contributed equally to this work
*Correspondence: furlong@embl.de
DOI 10.1016/j.cell.2012.01.030

SUMMARY

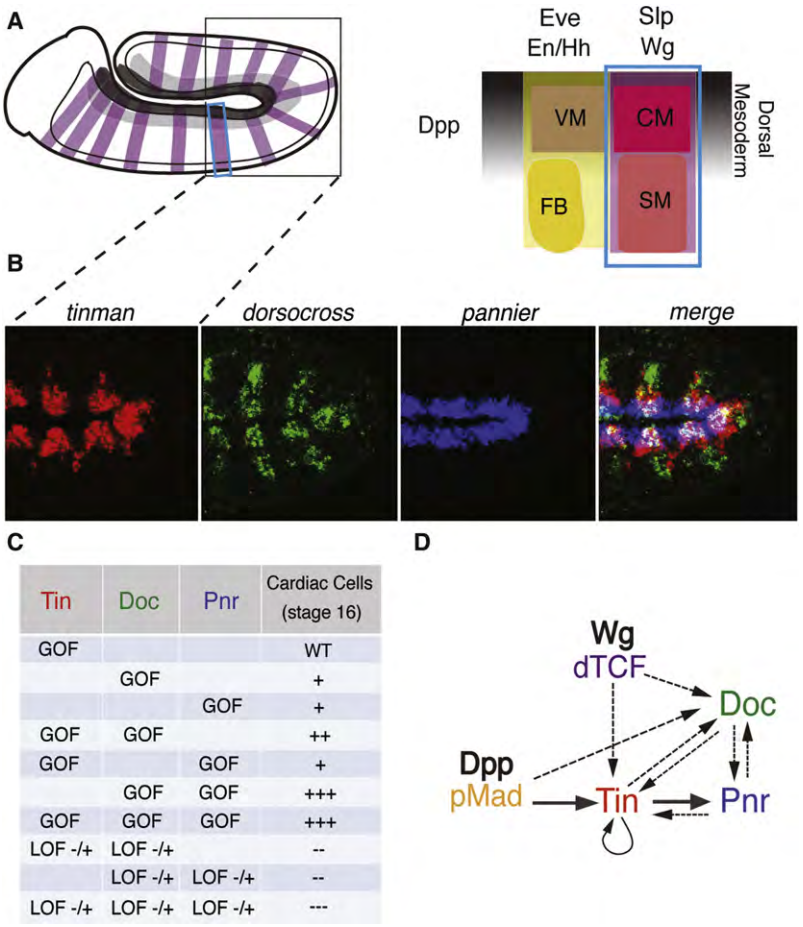
Cell fate decisions are driven through the integration of inductive signals and tissue-specific transcription factors (TFs), although the details on how this information converges in *cis* remain unclear. Here, we demonstrate that the five genetic components essential for cardiac specification in *Drosophila*, including the effectors of Wg and Dpp signaling, act as a collective unit to cooperatively regulate heart enhancer activity, both in vivo and in vitro. Their combinatorial binding does not require any specific motif orientation or spacing, suggesting an alternative mode of enhancer function whereby cooperative activity occurs with extensive motif flexibility. A fraction of enhancers co-occupied by cardiogenic TFs had unexpected activity in the neighboring visceral mesoderm but could be rendered active in heart through single-site mutations. Given that cardiac and visceral cells are both derived from the dorsal mesoderm, this "dormant" TF binding signature may represent a molecular footprint of these cells' developmental lineage.

INTRODUCTION

Pluripotent cells become progressively restricted in their cell fate through the action of inductive signals from surrounding tissues and specific cohorts of transcription factors (TFs). This multilevel information converges on *cis*-regulatory modules (CRMs, or enhancer elements) to elicit specific developmental programs. Information at some CRMs is integrated through cooperative TF binding, mediated via direct protein-protein interactions between TFs or common cofactors. Cooperative occupancy often requires a specific orientation, relative spacing, and helical phasing of TF-binding sites (Senger et al., 2004), referred to as motif grammar, to facilitate the appropriate protein interactions. A classic example of this is the enhanceosome model of enhancer activation (Panne, 2008). However, this stringent enhanceosome mode of regulation may represent only a small fraction of enhancers. Many developmental enhancers operate under more flexible conditions in which a subset of factors may bind cooperatively while the remaining factors are recruited independently and thus require little or no motif grammar. The billboard model, for example, suggests that TFs do not function in a single concerted manner at enhancers; rather, submodules interact independently and/or redundantly with the basal transcriptional machinery (Kulkarni and Arnosti, 2003). In some cases, enhancer flexibility appears even more extreme—not only can the relative location of binding sites vary, but also the identity of the TFs that are involved in regulating a specific pattern of expression (Brown et al., 2007; Zinzen et al., 2009).

The specification of the *Drosophila* dorsal mesoderm into visceral mesoderm (VM) and cardiac mesoderm (CM) cell fates represents an excellent paradigm for complex enhancer integration (Halfon et al., 2000; Kelly and Buckingham, 2002; Xu et al., 1998; Zaffran and Frasch, 2002). Here, cell fate decisions are induced through the intersection of ectodermal Wingless (Wg, a Wnt protein) and Decapentaplegic (Dpp, a TGF-β family protein) signaling (Figure 1A). Pluripotent cells that receive both signals within the underlying dorsal mesoderm are specified to become CM, and the neighboring cell population that only receives Dpp signal becomes VM (Lee and Frasch, 2000; Lockwood and Bodmer, 2002) (Figure 1A). Tinman (Tin, an Nkx factor) and pMad (the effector of Dpp signaling) provide the competence for these "precursor cells" to acquire either a VM or CM cell fate (Xu et al., 1998). In particular, Tin acts together with Pannier (Pnr, a GATA factor) and Dorsocross (Doc, a T box factor) to specify CM cell fate (Reim and Frasch, 2005), whereas the VM fate is actively repressed in these cells (Lee and Frasch, 2005) (Figures 1A and 1B).

Genetic studies in both *Drosophila* and mice suggest that the *cis*-regulatory network driving cardiac specification is highly cooperative. For example, although Nkx, GATA and T box factors are essential for heart development in all species studied to date (Cripps and Olson, 2002; Frasch, 1999; Olson, 2006; Reim and Frasch, 2005), neither factor alone is sufficient to



Tin	Doc	Pnr	Cardiac Cells (stage 16)
GOF			WT
	GOF		+
		GOF	+
GOF	GOF		++
GOF		GOF	+
	GOF	GOF	+++
GOF	GOF	GOF	+++
LOF -/+	LOF -/+		--
	LOF -/+	LOF -/+	--
LOF -/+	LOF -/+	LOF -/+	---

Figure 1. Dorsal Mesoderm Specification into Cardiac and Visceral Mesoderm during *Drosophila* Embryogenesis

(A) Diagram of a *Drosophila* embryo showing *wg* expression in 14 parasegments. Area indicated by blue rectangle is enlarged in the right panel, showing a schematic representation of mesoderm subdivision in one hemisegment. The dorsal domain, which has high levels of Decapentaplegic (Dpp) signaling (black), gives rise to visceral mesoderm (VM) and cardiac mesoderm (CM), whereas ventral regions become fat body (FB) and somatic muscle (SM). CM is specified at the intersection of Wingless (Wg, purple) and Dpp signaling in the posterior part of each parasegment. Wg activates *sloppy paired* (*slp*) expression, and together they promote CM and repress VM specification.

(B) Triple-fluorescent in situ hybridization showing *tinman*, *dorsocross*, and *pannier* expression in the dorsal mesoderm during early stage 11, when cardiac specification takes place. All three genes are coexpressed exclusively in the cardiogenic mesoderm (pink-white area of coexpression). The region of the embryo shown is depicted by the black square in (A).

(C) Summary of the genetic interaction between Tinman (Tin), Dorsocross (Doc), and Pannier (Pnr) to CB specification (Reim and Frasch, 2005). GOF, gain-of-function; LOF, loss-of-function (−/+ = heterozygous) genetic backgrounds. + and − denote an increase or decrease in the number of cardioblasts, respectively.

(D) Recursive regulation between the key factors essential for CM specification. Solid lines indicate direct regulation; dashed lines represent a genetic interaction (direct or indirect).

See also Figure S1.

induce a cardiac cell fate. Rather, the ectopic expression of combinations of TFs is required to drive the cardiogenic program in both flies (Reim and Frasch, 2005) and mice (Durocher et al., 1997; Sepulveda et al., 1998) (Figure 1C). Moreover, combinations of GATA4, Tbx5, and a third factor are sufficient to drive transdifferentiation of cell types into a CM cell fate (Takeuchi and Bruneau, 2009) and to direct reprogramming of fibroblasts into cardiomyocytes (Ieda et al., 2010), and yet, the molecular nature of this cooperativity is very poorly understood. Despite the extensive genetic characterization of CM specification, only a handful of enhancers are known to regulate early stages of heart development (Figures S1A–S1G available online), precluding any general hypotheses on how the input from multiple TFs (Tin, Pnr, Doc, and the effectors of Wg and Dpp signaling) converges in *cis*. For example, it is not known if the cooperativity observed between these factors at a genetic level (Figures 1C and 1D) is reflected at the *cis*-regulatory level and requires a specific motif grammar at the sequence level.

To address these issues, we examined the genome-wide occupancy of pMad, dTCF, Doc, Pnr, and the mesoderm-specific factor Tin during dorsal mesoderm specification in *Drosophila* embryos. We find that all five TFs are recruited to shared enhancers to a much higher degree than expected by chance and do so in a mesoderm-specific context, matching their only domain of coexpression (Figure 1B). These regions function as heart enhancers in vivo and require the presence of all five TFs for their cooperative regulation and maximal enhancer activity in vitro. The collective enhancer occupancy, which we further confirm using a cell culture model and mutagenesis analysis in vivo, occurs in the absence of any consistent motif grammar, revealing an alternative mode of cooperative regulation using very flexible motif content. Our analysis also uncovered an additional property of developmental enhancers, whereby dormant TF binding signatures reflect a developmental footprint of a cell's lineage. ''Cardiac'' TFs occupy enhancer elements that are active in the neighboring VM, echoing the fact that both cell populations are derived from the dorsal mesoderm.

RESULTS

Building a TF Binding Atlas for Enhancers Active in the Dorsal Mesoderm

To generate a TF binding atlas of regulatory regions active in the dorsal mesoderm, we performed genome-wide ChIP-on-chip experiments with antibodies directed against Doc, Pnr, dTCF, and pMad, the activated phosphorylated form of the Dpp effector Mad. The experiments were performed at two

consecutive stages of development: 4–6 hr after egg lay (stages 8 and 9) and 6–8 hr (stages 10 and 11), corresponding to the subdivision of the dorsal mesoderm and its subsequent specification into CM and VM (Campos-Ortega, 1997). A high confidence set of TF-bound regions was defined for each factor, identifying thousands of occupied sites per TF (Table S1 and Extended Experimental Procedures). These data were combined with Tin occupancy data generated under the same conditions at the same stages of development (Zinzen et al., 2009). The pairwise occupancy patterns of all five TFs showed highly significant overlap (Figures S2E and S2F), providing an initial indication that these factors occupy common *cis*-regulatory elements.

To convert TF binding peaks into co-occupied enhancers, binding events that clustered in close proximity to each other were merged to define putative *cis*-regulatory regions, as described previously (Zinzen et al., 2009). In this way, the combined 55,423 significant TF binding peaks clustered into 11,286 nonredundant CRMs, approximately one-third of which (4,041) have significant levels of Tinman binding. Though *tinman* expression is restricted to the mesoderm, the expression of the other four TFs is not, even within these narrow time windows of development (Figure 1B). We used Tin binding to limit our analysis to CRMs more likely to be active in mesodermal lineages and therefore focused for the remainder of this study on the 4,041 Tin-bound CRMs, the majority of which also recruited other factors (Table S2).

The TF occupancy patterns fully recapitulated *all* known binding to previously characterized dorsal mesoderm enhancers and the four known early cardiac enhancers active at the analyzed stages (Figures S1A–S1H) and in many cases revealed additional regulatory connections, demonstrating the sensitivity and resolution of the data. To extend this further, we selected 50 genes with at least Tin, Doc, and Pnr binding in their vicinity and examined their expression patterns by double-fluorescent in situ hybridization. Forty-two of these genes gave specific spatiotemporal expression, 38 of which (90%) are expressed in the dorsal mesoderm (26 genes) and/or cardiac mesoderm (20) and/or visceral mesoderm (7) (Figures S1I–S1K and Table S3). This supports our reasoning that the integration of binding signatures for four nonmesoderm-specific TFs (pMad, dTCF, Pnr, and Doc) with Tin is a valid approach to focus on transcriptional regulation within the dorsal mesoderm and its derivatives.

A Regulatory Collective of Cardiogenic TFs Is Recruited to Tin-Bound Enhancers

To relate TF binding signatures to specific *cis*-regulatory function, we applied an unbiased clustering approach to assess general TF preferences for enhancer co-occupancy, followed by extensive in vivo transgenic reporter analyses to assess enhancer activity. The maximum moving average ChIP signal for each TF at each enhancer was used as a quantitative input for enhancer classification. As enhancers were defined based on high-confidence binding signals for at least one TF, this procedure ensured that subthreshold signals for all other TFs were taken into account for enhancer classification (Extended Experimental Procedures). The Bayesian clustering algorithm Autoclass (Cheeseman, 1996) was used to partition enhancers based on their similarity in TF binding signatures across all experiments, computing a probability score for each enhancer to belong to each cluster. This approach produced confident single-cluster assignments for 77% (3,099) of the 4,041 Tin-bound enhancers (Figure 2A, left heatmap, and Table S4), and the robustness of this classification was confirmed by bootstrap analysis (Extended Experimental Procedures).

Examining the signal distribution of TF occupancy in each cluster revealed six broad enhancer classes that are qualitatively distinct from each other (Figures 2A, left, S2A, and Extended Experimental Procedures). The first class harbors enrichment for all five TFs (Figures 2A, left, "All TF" CRMs labeled with shades of red, and S2A, upper-left). The second class, in contrast, is depleted in binding signal for all TFs except Tin and represents ~20% of CRMs ("Tin only," labeled with shades of gray in Figure 2A). The four remaining classes are defined by elevated signals for Tin and one additional TF, with generally medium to low signals for other factors. We loosely refer to these as "two TF" classes as follows: "pMad+Tin" (~2% CRMs), "dTCF+Tin" (~8%), "Doc+Tin" (~4%), and "Pnr+Tin" (~20% CRMs) (Figures 2A, left, and S2A). Individual clusters within each of these classes differ in the quantitative levels of TF binding signals but generally not in the identity of the TFs themselves (Figures 2A and S2A and Extended Experimental Procedures). CRM clusters with the most prominent binding profiles from each class were used for further analysis (Figures 2C and S2A, boxed histograms, and Extended Experimental Procedures).

This unbiased grouping of enhancers, based on their similarity in TF occupancy, revealed two unexpected findings. First, the most prominent binding signature at enhancers is the recruitment of all five TFs. Depending on the threshold of the mean TF binding signal per class (Extended Experimental Procedures), between 22% and 46% of classified enhancers have highly correlated signals for Doc, dTCF, pMad, Pnr, and Tin across one or both developmental times (Figures 2A, left, labeled with shades of red from high TF binding signal [top] to low [bottom], and S2A). Second, there are few enhancers bound at high levels by three or four TFs; instead, the majority of regions are either occupied by all five factors or have high enrichment for only two factors (TF+Tin). This suggests that all five factors bind to these elements as a collective unit, which may require a specific mesodermal context to anchor their binding. To test this further, we applied the same clustering procedure to enhancers that are significantly bound by one or more TF but are not bound by Tin (the mesoderm-specific factor) at the analyzed stages of development (1,209 CRMs with near-zero Tin signal; Table S2). On these Tin-negative regions, there is very little correlated co-occupancy of the other four TFs (Figure 2A, right). This was further confirmed on a stringent set of enhancers that are highly enriched for two or more TFs other than Tin, whereas the signal for the remaining analyzed TFs is below the lower 50% of the background signal distribution (Figure 2B). The occupancy of Doc, dTCF, pMad, and Pnr at these "all or nothing" CRMs is strikingly different depending on the presence of Tin binding (Figure 2B). More globally, the degree of TF co-occupancy is significantly higher at Tin-bound regions, but not Tin-negative regions, compared to that expected at random (Figures S2B–S2D).

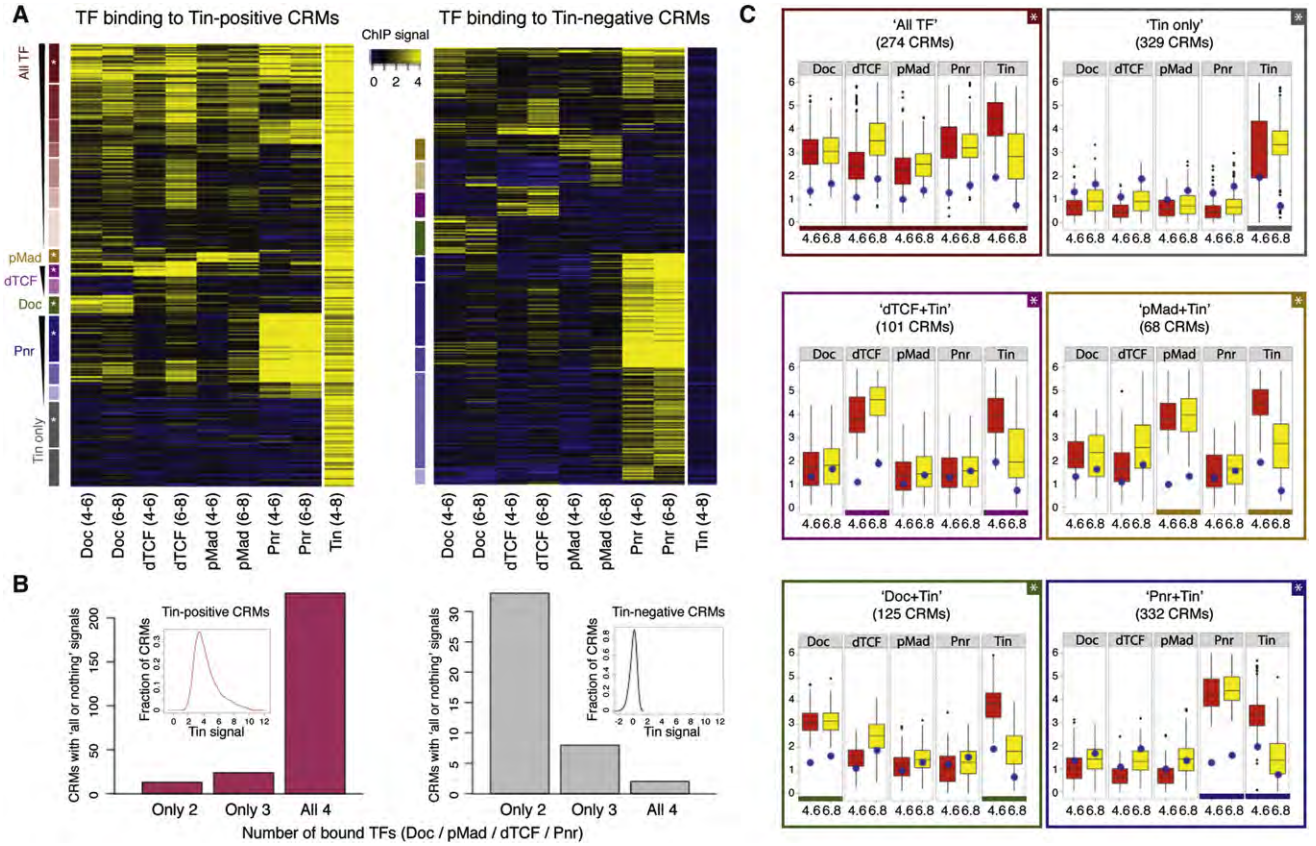


Figure 2. Co-Occupancy of Cardiogenic TFs at Tin-Bound CRMs

(A) Unsupervised classification of Tin-positive (left) and Tin-negative (right) CRMs using Autoclass Bayesian clustering. Rows correspond to defined CRMs, and columns correspond to the maximum moving average ChIP signal for a transcription factor (TF) at the indicated time points. Yellow represents high and blue background signal. Rectangles to the left of the heatmaps indicate subclasses of CRMs with related TF binding signals; white asterisks indicate subclasses with the most prominent TF binding signal from each class, which was selected for further analysis (shown in Figure 2C).

(B) Assessment of TF co-occupancy on a subset of CRMs that have either strong or background ("all or nothing") signals for each of the four analyzed TFs. Bar charts show the number of such CRMs occupied by two to four TFs on Tin-positive (left) or Tin-negative (right) CRMs. Density plots show the distribution of Tinman signal in each of the two CRM subsets (inset).

(C) Representative subclasses of each binding signature (marked with asterisks in Figure 2A) used for further analysis. Boxplots show the distributions of ChIP signals for the five TFs at two time points (4–6 and 6–8 hr). Blue dots show median signal for each TF across all CRMs. Comparing the position of the blue dot to the median area of the box plot indicates whether a TF's binding is specifically enriched on this group of CRMs.

See also Figure S2.

Taken together, these data indicate that all five TFs tend to be corecruited to regulatory regions in a concerted manner (as further confirmed using a cell-based system below), which occurs in a mesoderm-specific context (Tin-bound CRMs), in keeping with their only domain of coexpression (Figure 1B).

Enhancers Occupied by All Five TFs Regulate Expression in the Dorsal Mesoderm or Its Derivatives

Having defined specific classes of enhancers with qualitatively different TF occupancy patterns, we assessed which of these represent active enhancers in vivo and drive expression in the dorsal mesoderm and/or in cardiac cells. ChIP-defined enhancers (average size 550 bp) were cloned upstream of a GFP reporter gene and stably integrated into the *Drosophila* genome. Enhancer spatiotemporal activity throughout embryonic development was assayed by in situ hybridization in transgenic embryos to provide accurate temporal resolution for when the enhancer is active. Importantly, the selection of enhancers was based purely on representative binding signatures, without prior knowledge concerning the function of neighboring genes or the motif content of the enhancers. In total, the activities of 55 regions were examined in transgenic embryos, almost half of which correspond to the All TF binding class (47%), as this represents the most predominant TF binding signature (Figures 3, 6, and S3 and Table S5).

A striking 92% of enhancers tested from the All TF class (24 of 26 regions) were sufficient to function as enhancers in vivo. The vast majority of these (91.6%; 22/24) regulate expression in mesodermal lineages (Figure 3A), of which the most prominent expression signature (50%; 12 CRMs) is activity within the cardiogenic mesoderm (Figures 3B and S3A). These complex spatial patterns of enhancer activity cannot be achieved through

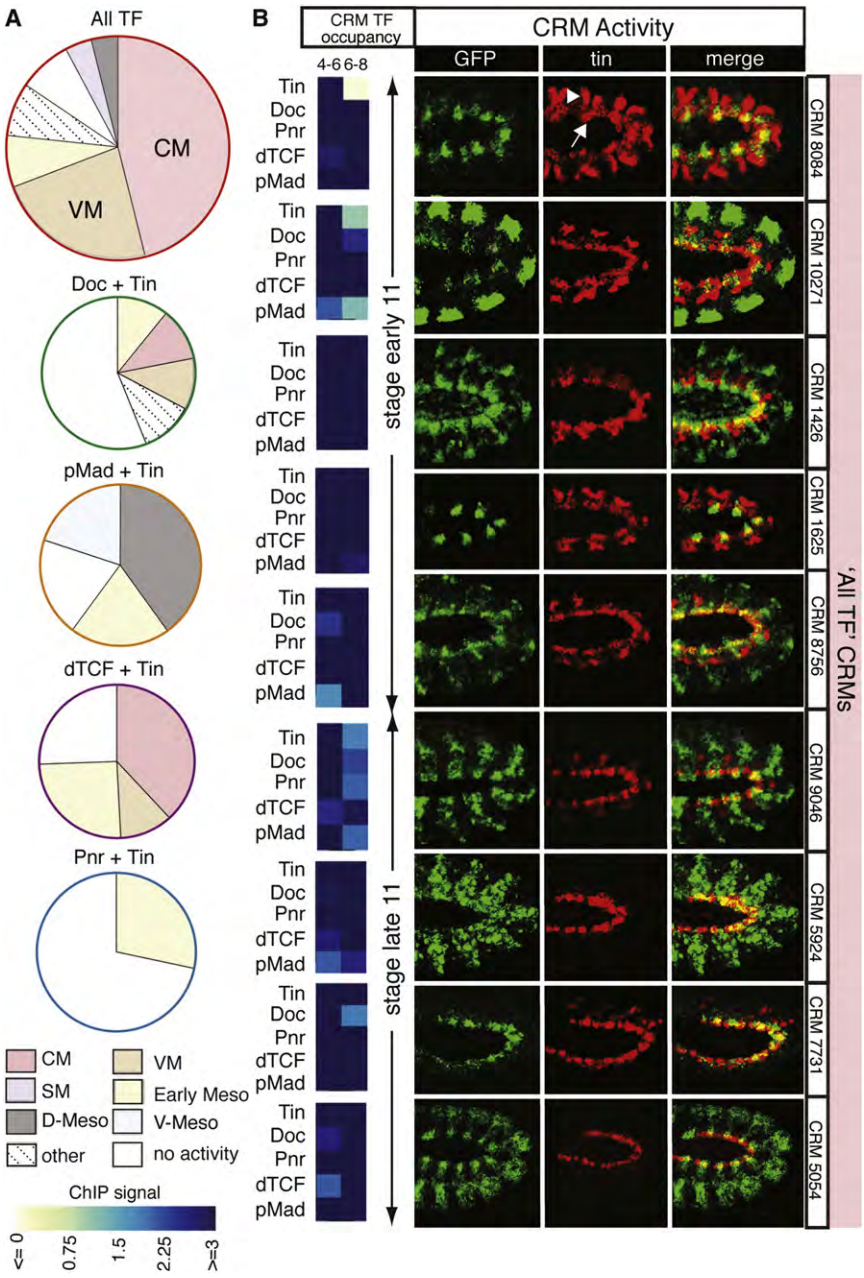


Figure 3. Collective TF Occupancy Correlates with Enhancer Activity in Cardioblasts

(A) Summary of the activity of 55 CRMs tested in vivo by transgenic reporter assays. Pie charts represent the proportion of CRMs driving expression in different tissues for each Autoclass-derived subclass. CRMs active in two (or more) mesodermal tissues are indicated in both. All TF CRMs had the highest percentage of regions that functioned as enhancers in vivo; 84.6% regulate expression in the mesoderm and/or its derivatives, with cardiac mesoderm (CM) expression being predominant (46%). VM, visceral mesoderm; SM, somatic muscle; Early Meso, early mesoderm; D-Meso, dorsal mesoderm; V-Meso, ventral mesoderm; other, nonmesodermal tissues.

(B) CRM spatiotemporal activity assayed by in situ hybridization of embryos with a transgenic reporter. (Left) The TF binding signals for each factor on each CRM at both time points (mean moving average ChIP signal per CRM; blue represents high levels). (Right) In situ hybridization using antisense RNA probes directed against the *GFP* reporter (green) and *tin* (red) as a marker of dorsal mesoderm and its derivatives. At stage early 11, *tinman* is expressed in visceral (arrowhead) and cardiac mesoderm (arrow); by late 11, only cardiac expression remains. CRMs show activity restricted to CM (1625, 7731) or more complex patterns in CM and other cell types (1426, 9046, 5054), reflecting the nonexclusive expression of many heart genes. All embryos shown laterally; anterior, left; dorsal, up; region depicted is indicated by the black square in Figure 1A. The remaining tested CRMs are shown in Figures 5, 6, and S3.

See also Figure S3.

the action of any one TF alone but, rather, reflect the intersection in expression domains of many of these factors, in line with their observed collective occupancy. The second most prominent activity (25%) was VM expression, which was surprising given the collective occupancy of all five "cardiogenic" TFs at these CRMs and is dissected in detail below.

The activity of approximately seven CRMs was tested for each of the four two-TF classes, of which 59% (17/29) function as enhancers in vivo (detailed results are shown in Figure S3). Eighty-eight percent (15/17) of active regions regulate activity in mesodermal tissues, including the early trunk, ventral, dorsal, and visceral mesoderm. However, in contrast to the All TF CRMs, only four regions (23%) regulate activity in CM, with all but one CRM belonging to the dTCF+Tin class (Figures 3A and S3).

In summary, regions co-occupied by all five TFs were much more likely to direct expression in the dorsal mesoderm (or its derivatives) compared to any of the two-TF classes, with 75% (18/24) of active All TF CRMs driving specific expression in CM or VM. It is important to note that Tin binding in the absence of Doc, Pnr, pMad, and dTCF is not sufficient to regulate enhancer activity in cardiac cells, as demonstrated by extensive analyses of enhancers bound by Tin in combination with other TFs (Liu et al., 2009; Zinzen et al., 2009). Therefore, the activity of Tin within the cardiogenic TF collective has unique properties in terms of its functional output.

Relaxed Sequence Requirements at Enhancers Occupied by All Five Factors

Given the extensive corecruitment of the five TFs, we asked whether the motif content of the All TF enhancers explains their collective occupancy and activity. We first determined the

general sequence preferences of each TF using de novo motif discovery on all regions bound by that factor (Extended Experimental Procedures). The identified position weight matrices (PWMs), which were similar to published models (Figure S4A), were then used to assess differential motif enrichment between All TF CRMs and two-TF CRMs (Figure 4). This analysis revealed two classes of TFs, suggesting different modes of their recruitment to DNA. Doc and Pnr transcription factor binding sites (TFBS) are preferentially found in All TF CRMs compared to their respective two-TF CRMs, whereas, in contrast, the numbers of TFBSs for pMad, dTCF, and Tin are lower in All TF CRMs compared to their respective two-TF CRMs (Figure 4B). This holds true regardless of which specific PWM score threshold is used for the motif detection (data not shown) or when using an unthresholded approach summarizing both high- and low-affinity TFBSs (TRAP, Roider et al., 2007; Figure 4C). The differential motif enrichment of Doc and Pnr compared to pMad, dTCF, and Tin was further confirmed using de novo motif analysis (Figure S4B).

The enrichment of Doc and Pnr TFBSs suggests that these factors are preferentially recruited to All TF CRMs in a sequence-specific fashion (Figures 4B and 4C), and consistent with this, their motifs are more conserved in All TF CRMs compared to their respective two-TF classes (data not shown). Conversely, the number of pMad, dTCF, and Tin sites are lower in All TF CRMs compared to their respective two-TF CRMs, which is particularly striking for dTCF (Figures 4B and 4C), suggesting that heterotypic cooperative binding may play a role in their recruitment to All TF CRMs. A role for cooperativity in this system is supported by direct protein-protein interactions between almost all of these TFs in both *Drosophila* and vertebrates (Brown et al., 2004; Bruneau et al., 2001; Durocher et al., 1997; Gajewski et al., 2001; Garg et al., 2003; Nishita et al., 2000; Zaffran et al., 2002).

Protein-protein interactions between TFs can often introduce sequence constraints within enhancers, where the relative spacing and orientation of motifs must maintain a certain configuration to facilitate protein interaction and binding (Panne, 2008). We searched for this type of motif grammar, examining the relative motif spacing and orientation of Doc, Pnr, pMad, dTCF, and Tin TFBS within All TF CRMs. Surprisingly, we found no evidence of consistent grammar as a characteristic signature of All TF CRMs (“CRM Grammar Analysis” in Extended Experimental Procedures and Figures S4C and S4D). Moreover, the motif content itself is highly diverse, whereby the occurrence of pMad, dTCF, and Tin sites and distance between them varies between each All TF CRM. Despite this motif heterogeneity, however, these enhancers recruit all five TFs and function as heart enhancers in vivo, mirroring the cooperative function of these TFs during heart development.

The Presence of Pnr and Doc Is Essential for Tin-pMad-dTCF-Mediated Enhancer Activation

The collective occupancy and activity of the All TF enhancers suggests that the high level of cooperativity observed between these factors at a genetic level extends to their downstream *cis*-regulatory network (Figures 1C and 1D). To examine this further, we generated a cell culture-based model that expresses all five TFs in their active forms. Although this system lacks the spatial and temporal context of the developing embryo, it provides a more homogenous cell population. Based on extensive RNA-seq data (Cherbas et al., 2011), we found that DmD8 cells (an established *Drosophila* cell line derived from dorsal mesothoracic disc) express *pnr* and *doc*, but not *tin*, which we also confirmed at the protein level (Figure S5A). Although all components of the Wg and Dpp signaling cascades are expressed, the ligands are not; therefore, these signaling pathways are inactive in this cell line. To obtain activated dTCF and pMad, we generated conditioned DmD8 medium containing secreted Wg and Dpp. Applying this conditioned medium to fresh DmD8 cells resulted in the phosphorylation of Mad and the activation of the Wg signaling pathway (Figure S5B). Therefore, upon *tin* transfection, all five TFs were active in this cell culture system.

We used this cell culture system to examine: (1) the co-occupancy of all TFs by ChIP followed by quantitative PCR and (2) the requirement of each TF for enhancer activity by luciferase assay. The results for one enhancer (CRM 3436) are highlighted in Figure 5. CRM 3436 is bound by all five TFs in vivo (Figure 5A) and is sufficient to regulate expression in a segmentally repeated pattern encompassing part of the cardiogenic mesoderm (Figure 5B). Performing ChIP for all five factors in cell culture revealed significant occupancy of each TF on the endogenous enhancer locus compared to an unbound negative region (Figure 5C). A similar significant enrichment in the occupancy of all TFs was observed for all six enhancers analyzed (Figures S5C and S5D), confirming the collective occupancy observed in vivo.

To examine the regulatory input of these TFs, three All TF CRMs were placed upstream of a minimal promoter driving a luciferase reporter and transfected into DmD8 cells where both Pnr and Doc were depleted using RNAi to obtain a basal level of the enhancer’s activity. The presence of either Pnr or Doc alone had no significant effect on enhancer activity, whereas both together caused a marginal increase (Figures 5D, S5E, and S5F). Addition of Tin in the presence of Pnr and Doc, however, caused a significant increase in activity, whereas the presence of all five activated TFs had the most dramatic effect, leading to a 15-fold increase over the basal level (Figure 5D). These results demonstrate that all five TFs contribute to the enhancers’ activity and are required for maximal enhancer activation (Figures 5D, S5E, and S5F).

The clear differences in the enrichment and conservation of Pnr and Doc motifs compared to those of Tin, dTCF, and pMad suggest that these two TFs may preferentially serve as anchors for the collective TF binding. Taking advantage of this cell system, we systematically tested this hypothesis by removing Doc alone, Pnr alone, or both in the presence of the other three TFs. As shown in Figure 5D (red asterisk), removal of Doc had a significant effect, whereas the removal of Pnr alone reduced the enhancers activity back to its basal level, despite the presence of Tin, pMad, and dTCF. Therefore, the presence of Pnr and Doc is required for the ability of Tin, dTCF, and pMad to activate the enhancer. The fact that Pnr alone or in combination with Doc is not sufficient for significant enhancer activation suggests that these TFs are essential for the collective recruitment of all five TFs, consistent with the motif content of these CRMs.

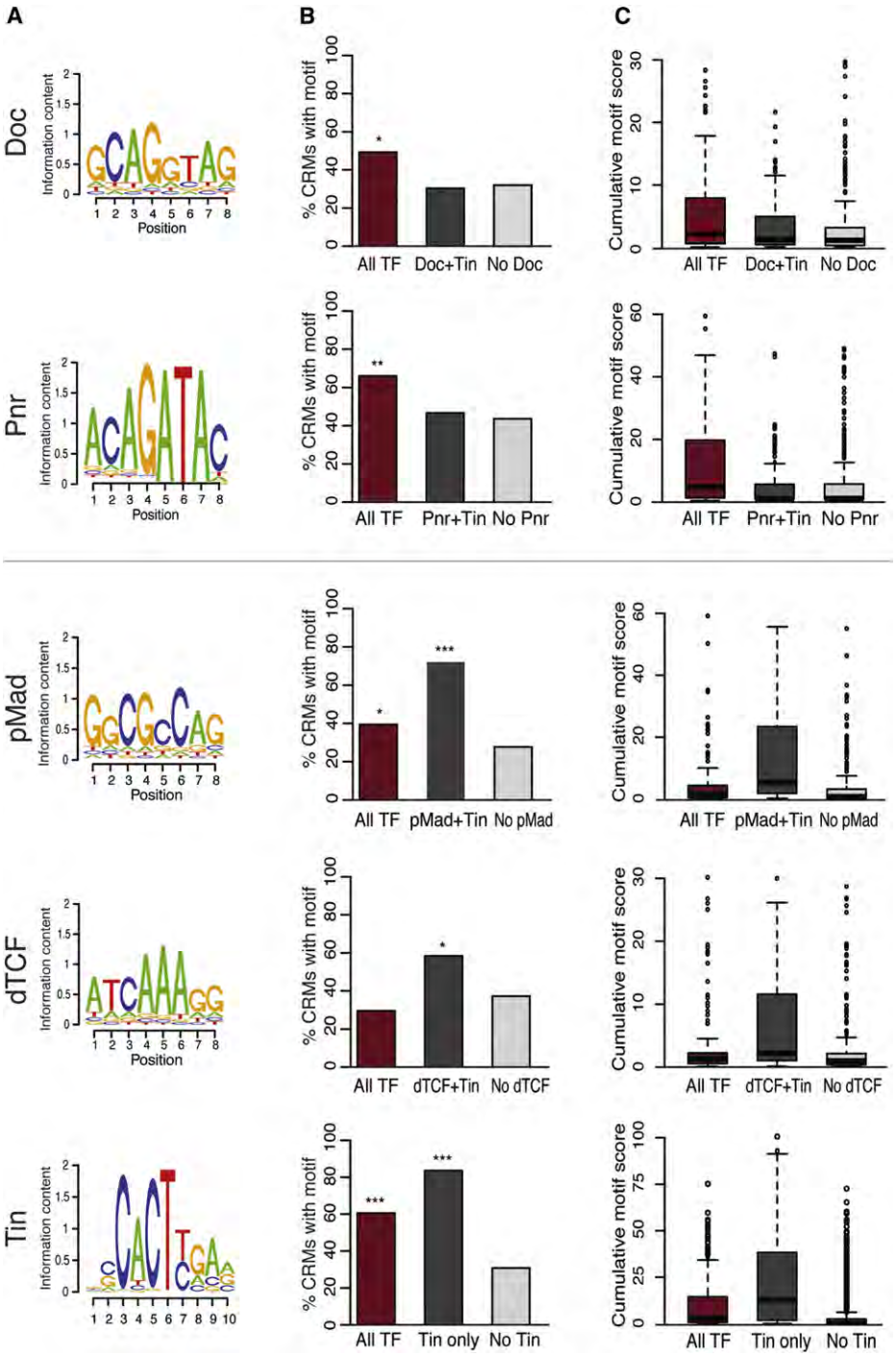


Figure 4. Sequence Properties of All TF CRMs versus Two TF CRMs

(A) Motifs discovered de novo for Doc, Pnr, pMad, dTCF, and Tin in all regions bound by the respective TF are similar to those reported previously (Figure S4A).

(B) Enrichment of TF-binding sites in different CRM classes. Doc and Pnr motifs are more frequently found in All TF CRMs compared to their two TF classes, whereas pMad, dTCF, and Tin motifs are more frequently found in their respective two-TF CRMs, compared to All TF CRMs.

(C) Cumulative motif enrichment scores (computed without score thresholds; TRAP) confirm the differential motif enrichment: Doc and Pnr motifs have elevated cumulative scores in All TF CRMs compared to their respective two TF CRMs (Wilcoxon test $p = 0.02$ and $p = 3.8 \times 10^{-12}$, respectively) and those not bound by the analyzed TF ($p = 2.5 \times 10^{-6}$ and $p = 6.7 \times 10^{-14}$). In contrast, pMad, dTCF, and Tin have lower cumulative motif scores in All TF CRMs compared to their respective two-TF CRMs (pMad $p = 2.3 \times 10^{-8}$, dTCF $p = 6.7 \times 10^{-6}$, Tin $p = 8.8 \times 10^{-11}$). Cumulative motif scores (computed using TRAP) are normalized to the median value for each TF.

See also Figure S4.

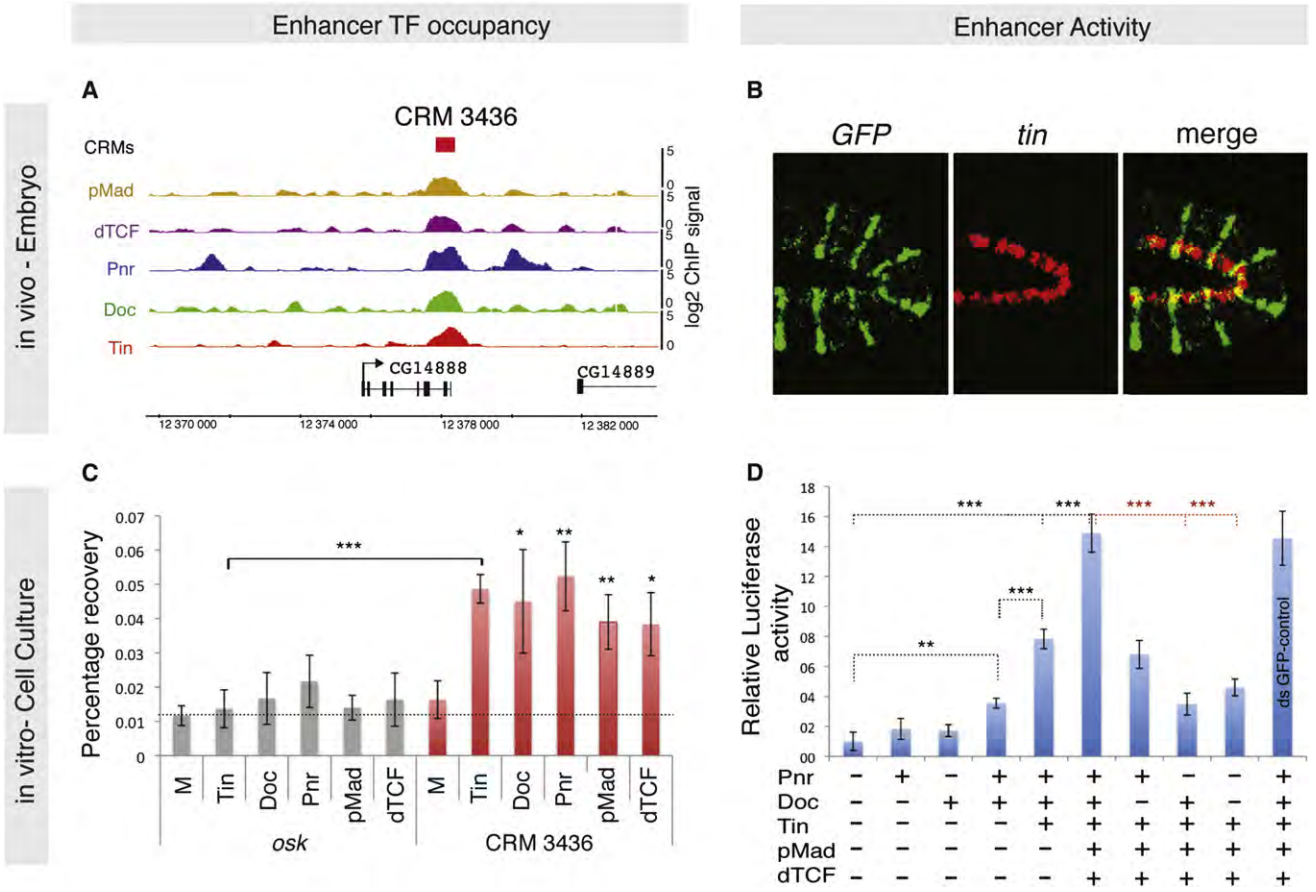


Figure 5. The Presence of Pnr and Doc Is Essential for the Ability of Tin, pMad, and dTCF to Activate Heart Enhancers

(A) CRM 3436 is bound by all TFs in vivo. Shown is log_2 ChIP signal for each TF at embryonic stages 9–11 (merged 4–8 hr data).

(B) CRM 3436 spatiotemporal activity assayed by in situ hybridization of transgenic embryos containing a stable insertion of the ChIP-bound region (red rectangle in A) regulating a GFP reporter. Antisense RNA probes directed against *GFP* (green) and *tin* (red) as a marker of dorsal mesoderm and its derivatives reveal overlapping expression in cardiac mesoderm, indicated by the yellow area of coexpression (merge panel).

(C) CRM 3436 is occupied by all TFs in DmD8 cells. DmD8 cells containing activated forms of all five TFs were used for ChIP experiments followed by real-time PCR of the endogenous enhancer. Gray histograms represent percentage recovery of input for a negative unbound region (maternal gene *oskar*, *osk*); red histograms represent occupancy on CRM 3436. The binding of each TF is significantly enriched on CRM 3436 compared to that TF's enrichment on the negative region (indicated by solid line for Tin). *M*, mock reaction. Error bars show the standard deviations of triplicate experiments. p values (one-tailed type 2 *t* test): *p = < 0.05; **p = < 0.01; ***p = < 0.001.

(D) Luciferase assay of CRM 3436 activity in DmD8 cells. The first column indicates the basal level of the enhancer's activity, using dsRNAi to remove Pnr and Doc (Figure S5A). Error bars show the standard deviations of two biological replicates, each conducted in triplicate. p values (two-tailed type 3 *t* test): *p = < 0.05; **p = < 0.01; ***p = < 0.001. Results from all CRMs tested are shown in Figure S5.

See also Figure S5.

Biniou Binding Is Predictive of VM Activity for CRMs Co-Occupied by Cardiogenic TFs

Examining the activity of the All TF CRMs revealed that, while 50% regulate expression in the cardiogenic mesoderm (Figure 3), an additional 25% have specific activity in the visceral mesoderm (Figure 6A). This VM activity was unexpected given the collective binding of all five cardiogenic TFs (that are not coexpressed in the VM), which we further confirmed for three CRMs in our cell culture-based system (Figure S5D). Of note, the CM and VM activity were mutually exclusive, suggesting a "CM-VM" regulatory switch. To dissect the mechanism of this bimodality, we first assessed whether a central VM-specific regulator, Biniou, is bound to these enhancers based on our previously published data (Zinzen et al., 2009). Biniou is a FoxF TF that is specifically expressed in VM, where it is essential for its specification and subsequent differentiation (Jakobsen et al., 2007; Zaffran et al., 2001). Consistent with our expectation, Biniou ChIP signal is significantly higher at characterized enhancers with VM-specific activity compared to those active in CM (Figure 6B, top; Wilcoxon test p = 0.01). Biniou occupies these CRMs only at the early stages of dorsal mesoderm specification into VM and CM (6–8 hr) and not at later development stages (8–10 hr), mirroring their transient activity (Figure 6B, bottom). Extending this analysis to the entire All TF class

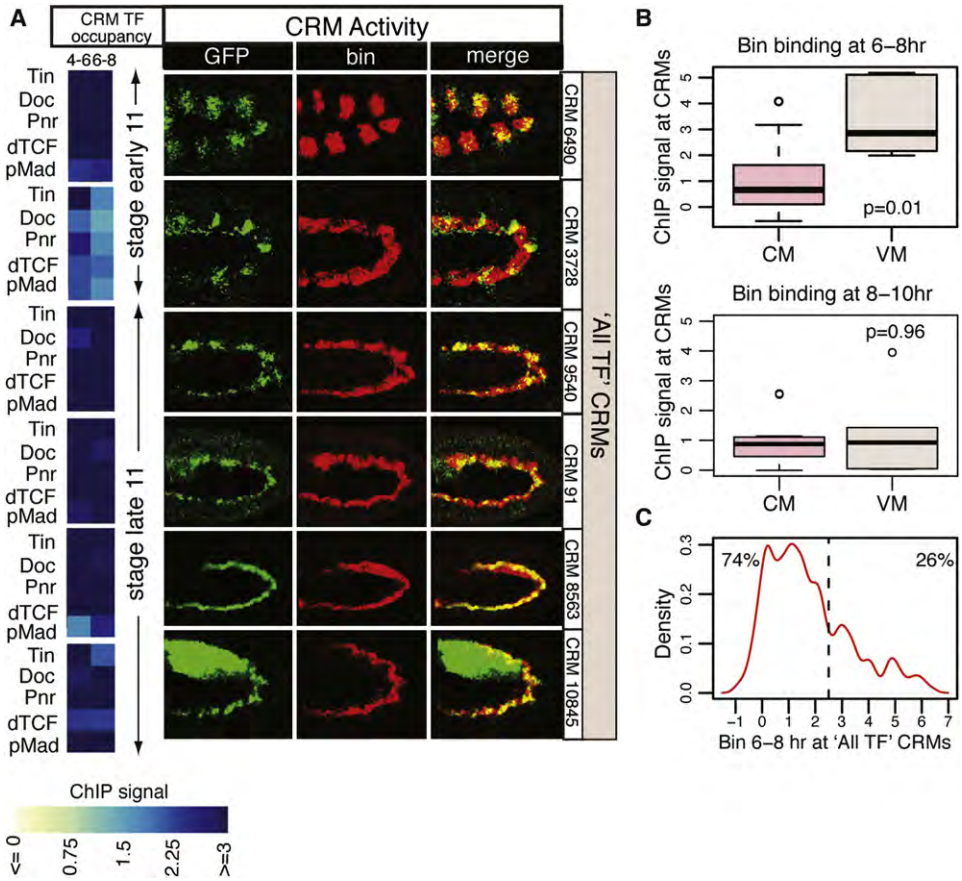


Figure 6. Biniou Occupancy Predicts Visceral Muscle Activity for Enhancers Collectively Bound by Cardiogenic TFs

(A) Unanticipated CRM activity in visceral mesoderm and not cardiac mesoderm for 25% of All TF CRMs tested. (Left) TF binding signals (mean moving average ChIP signal per CRM, wherein blue represents high enrichment). (Right) CRM activity by in situ hybridization using antisense RNA probes directed against the *GFP* reporter gene (green) and *biniou* (red) as a specific marker for VM. The CRMs drive expression in trunk VM (CRMs 6490, 91, 8563, 9540, and 10845) or in restricted populations of VM cells (CRM 3728).

(B) Box plots showing significantly higher levels of Biniou (Bin) occupancy at All TF CRMs driving VM (visceral mesoderm) expression compared to CM (cardiac mesoderm) (Wilcoxon test p = 0.01). This enrichment is only present at 6–8 hr, the stages of dorsal mesoderm specification (stages 10 and 11, top) and not at later stages (bottom).

(C) Density of all All TF CRMs with high or low Biniou (Bin) occupancy at 6–8 hr (x axis). Twenty-six percent of All TF CRMs have high levels of Bin binding, consistent with the proportion of tested CRMs with VM activity (25%).

revealed high levels of Biniou binding at ~25% of enhancers (Figure 6C), consistent with the proportion of tested CRMs showing VM activity. Therefore, a high level of Biniou binding at 6–8 hr (stages 10 and 11) is highly predictive of VM-specific activity, as indicated by the largely nonoverlapping distributions in ChIP signals (Figure 6B, top), and is consistent with the model of Biniou as an instructive regulator of VM cell fate (Jakobsen et al., 2007; Zaffran et al., 2001).

A Lineage Switch Motif Occupied by Two Fox Transcription Factors

Based on our current knowledge of how VM enhancers function, the binding signatures of Biniou, Tin, pMad, and another regulator Bagpipe (Azpiazu and Frasch, 1993) fully explain enhancer activity in the trunk visceral mesoderm at stage 10 (Lee and Frasch, 2005; Lee et al., 2006). However, this model does not explain the observed collective occupancy of cardiogenic TFs on these enhancers in the juxtaposed heart field or the fact that this binding signature is not sufficient to induce CM transcription, whereas other enhancers with similar binding signatures exhibit CM activity (compare Figure 6A to 3B). We reasoned that a transcriptional repressor likely binds to these "complex-VM" enhancers in cardioblasts and blocks the collective activity of pMad, dTCF, Tin, Pnr, and Doc. Sloppy paired (Slp) is a good candidate, as it is expressed in the cardiogenic mesoderm at these stages (Lee and Frasch, 2000) and is required to repress the activity of a VM enhancer in the *bagpipe* locus (*bap3*) in the cardiogenic domain (Lee and Frasch, 2005).

To investigate a potential role of Slp, we performed genome-wide ChIP-on-chip experiments against Slp at the same stages of development as the other TFs and then examined Slp recruitment to the 4,041 Tin-bound CRMs (Table S6). In

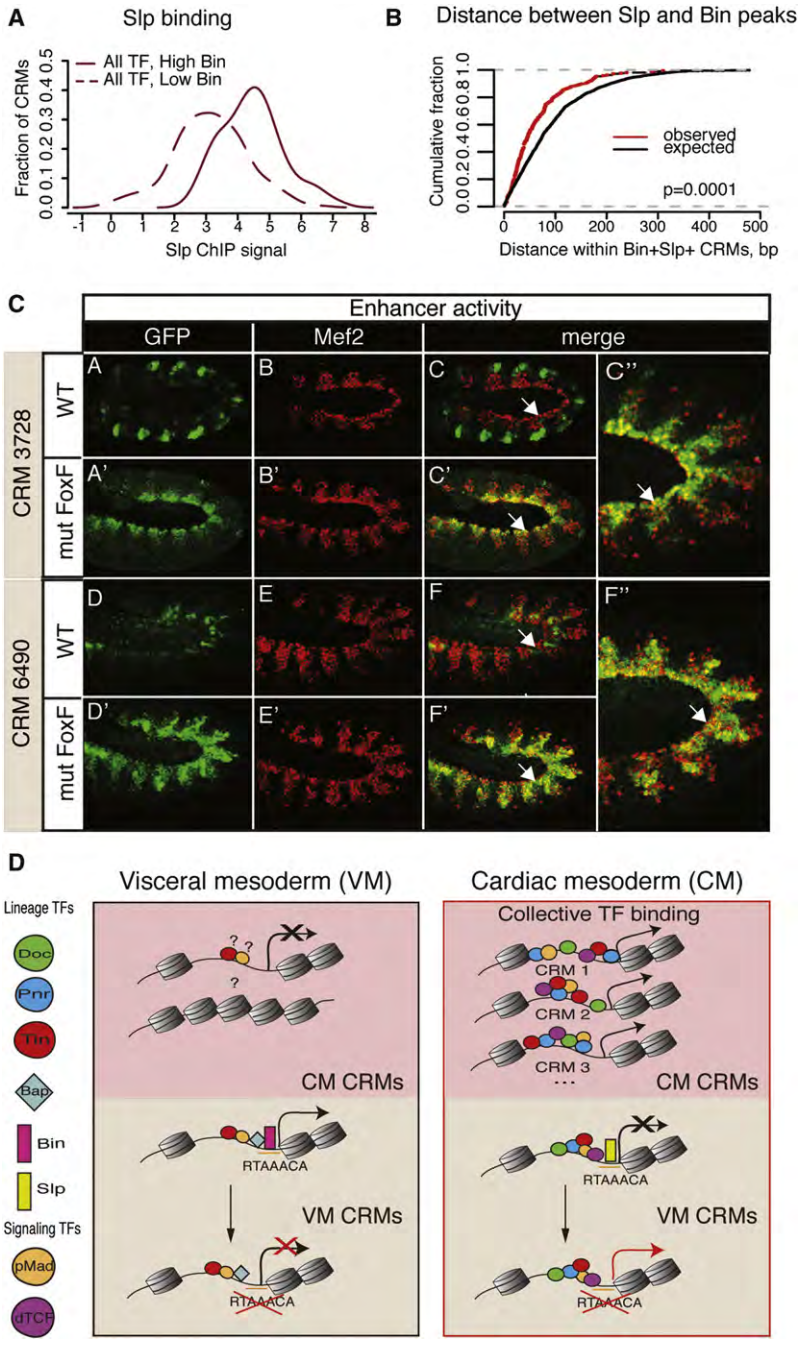


Figure 7. Sloppy Paired Represses the Activity of Dormant TF Binding in Cardiac Cells

(A) The distribution of Sloppy paired (Slp) binding signal at All TF CRMs depending on the levels of Biniou binding. Highest Slp signals were observed at All TF CRMs with high Biniou levels (visceral muscle enhancers) compared to low-Bin All TF CRMs (Wilcoxon test $p = 5.7 \times 10^{-12}$).

(B) Distance between Slp and Bin ChIP peaks within Bin-Slp-Tin-positive CRMs. Cumulative density distributions of observed distances (red) are shifted to the left compared to those expected at random (black), indicating that these peaks nonrandomly localize in proximity to each other. Wilcoxon test p values, p = 0.0001 (observed versus expected).

(C) Mutation of Slp-FoxF motif facilitates enhancer activity in heart and dorsal mesoderm. Immunostaining of transgenic embryos containing the wild-type (WT) and mutant (mut FoxF) enhancers using anti-GFP (enhancer reporter, green) and anti-Mef2 (a mesodermal marker, red) antibodies. Mutated Slp-FoxF sites are shown in Figure S6E. CRM 3728 and 6490 are active in the VM (Figure 6A), but not in CM (A–C and D–F). Mutation of the Slp FoxF sites leads to new activity in CM (A′–C′ and D′–F′, arrow). C″ and F″ are higher magnification images of C′ and F′, respectively.

(D) Proposed model for the regulation of cardiac and visceral mesoderm CRMs in both cell types. VM enhancers (left) contain FoxF motifs that recruit Biniou (Bin) in VM and Slp in cardiac cells, whereas all five heart TFs occupy these enhancers in cardiac cells. Slp counteracts the activity of the cardiogenic TF collective by repressing transcription. In contrast, enhancers that recruit the five heart TFs but lack FoxF motifs drive expression in cardiac cells (right).

See also Figure S6.

addition to the previously described binding on the *bap3* enhancer (Figure S1H), Slp binding is enriched at all enhancers within the All TF class with characterized VM activity (Figure S6A), as well as at those with predicted VM activity based on high levels of Biniou occupancy (VM CRMs) (Figure 7A). Moreover, Slp and Biniou binding peaks nonrandomly localize in close proximity to each other (Figures 7B and S6D) and to the Biniou-FoxF motifs (Figure S6C). Both results suggest that Biniou and Slp are recruited to enhancers via the same motif, globally extending the model of the *bap3* enhancer (Lee and Frasch, 2005). However, in contrast to the *bap3* enhancer, the early VM enhancers (Biniou-high CRMs) identified here are collectively bound by the five cardiogenic TFs, in addition to Slp. This complex binding signature promoted us to ask whether the cardiogenic TFs are capable of activating these enhancers once the repressive influence of Slp is removed. To test this, we mutated the Slp-Biniou FoxF motifs in three of the All TF CRMs that regulate expression in VM (Figure 6A, top three enhancers). In two out of three cases examined, mutation of this site was sufficient to facilitate expression in CM and, interestingly, also in the somatic muscle while attenuating activity in VM (Figures 7C and S6F). These results demonstrate that the "dormant" TF occupancy of cardiac factors has the capacity to direct CM activity. FoxF motifs within these enhancers are therefore used to activate transcription within the VM (mediated by Biniou; Figure S6F) and repress CM activity in the cardiogenic mesoderm (mediated by Slp; Figure 7C). These motifs thereby serve as a "lineage switch," ensuring exclusive enhancer activity in one of the two tissues derived from the dorsal mesoderm (Figure 7D).

DISCUSSION

Dissecting transcriptional networks in the context of embryonic development is inherently difficult due to the multicellularity of the system and the fact that most essential developmental regulators have pleiotropic effects, acting in separate and sometimes interconnected networks. Here, we present a comprehensive systematic dissection of the *cis*-regulatory properties leading to cardiac specification within the context of a developing embryo. The resulting compendium of TF binding signatures, in addition to our extensive in vivo and in vitro analysis of enhancer activity, revealed a number of insights into the regulatory complexity of developmental programs.

Cardiogenic TFs Form a Coherent Functional Module during Cardiac Specification

Nkx, GATA, and T box factors regulate each other's expression in both flies and mice (Lien et al., 1999; Molkentin et al., 2000; Reim and Frasch, 2005; Sun et al., 2004), where they form a recursively wired transcriptional circuit (Figure 1D) that acts cooperatively at a genetic level to regulate heart development across a broad range of organisms. Our data demonstrate that this cooperative regulation extends beyond the ability of these TFs to regulate each other's expression. All five cardiogenic TFs (including dTCF and pMad) converge as a collective unit on a very extensive set of mesodermal enhancer elements in vivo (Tin-bound regions) and also in vitro (in DmD8 cells). Importantly, this TF co-occupancy occurs in *cis*, rather than being mediated via crosslinking of DNA-looping interactions bringing together distant sites. Examining enhancer activity out of context, for example, in transgenic experiments and luciferase assays, revealed that the TF collective activity is preserved in situations in which these regions are removed from their native genomic "looping" context.

In keeping with the conserved essential role of these factors for heart development, the integration of their activity at shared enhancer elements may also be conserved. Recent analyses of the mouse homologs of these TFs (with the exception of the inductive signals from Wg and Dpp signaling) in a cardiomyocyte cell line support this, revealing a signifcant overlap in their binding signatures (He et al., 2011; Schlesinger et al., 2011), although interestingly not in the collective "all-or-none" fashion observed in *Drosophila* embryos. This difference may result from the partial overlap of the TFs examined, interspecies differences, or the inherent differences between the in vivo versus in vitro models. Examining enhancer output for a large number of regions indicates that this collective TF occupancy signature is generally predictive of enhancer activity in cardiac mesoderm or its neighboring cell population, the visceral mesoderm—expression patterns that cannot be obtained from any one of these TFs alone.

TF Collective: Cooperative Enhancer Regulation Using Flexible Sequence Context

There are currently two prevailing models of how enhancers function. The enhanceosome model suggests that TFs bind to enhancers in a cooperative manner directed by a specific arrangement of motifs, often having a very rigid motif grammar (Panne, 2008). An alternative, the billboard model, suggests that each TF (or submodule) is recruited independently via its own sequence motif, and therefore the motif spacing and relative orientation have little importance (Kulkarni and Arnosti, 2003). Our results indicate that cardiogenic TFs are corecruited and activate enhancers in a cooperative manner, but this cooperativity occurs with little or no apparent motif grammar to such an extent that the motifs for some factors do not always need to be present. This is at odds with either the enhanceosome (cooperative binding; rigid grammar) or billboard (independent binding; little grammar) models and represents an alternative mode of enhancer activity, which we term a "TF collective" (cooperative binding; no grammar), and likely constitutes a common principle in other systems.

Our data suggest that the TF collective operates via the cooperative recruitment of a large number of TFs (in this case, at least five), which is mediated by the presence of high-affinity TF motifs for a subset of factors initiating the recruitment of all TFs. The occupancy of any remaining factor(s) is most likely facilitated via protein-protein interactions or cooperativity at a higher level such as, for example, via the chromatin activators CBP/p300, which interact with mammalian GATA and Mad homologs (Dai and Markham, 2001; Feng et al., 1998). This model allows for extensive motif turnover without any obvious effect on enhancer activity, consistent with what has been observed in vivo for the *Drosophila spa* enhancer (Swanson et al., 2010) and mouse heart enhancers (Blow et al., 2010).

Dormant TF Occupancy Reflects the Developmental History of a Cell's Lineage

Integrating the TF occupancy data for all seven major TFs involved in dorsal mesoderm specification (the five cardiogenic factors together with Biniou and Slp) revealed a very striking observation: the developmental history of cardiac cells is reflected in their TF occupancy patterns. VM and CM are both derived from precursor cells within the dorsal mesoderm. Once specified, these cell types express divergent sets of TFs: Slp, activated dTCF, Doc, and Pnr function in cardiac cells, whereas Biniou and Bagpipe are active in the VM (Figures 1A and 7D). Despite these mutually exclusive expression patterns, the cardiogenic TFs are recruited to the same enhancers as VM TFs in the juxtaposed cardiac mesoderm (Figure 7D). Moreover, dependent on the removal of a transcriptional repressor, these combined binding signatures have the capacity to drive expression in either cell type. This finding provides the exciting possibility that dormant TF occupancy could be used to trace the developmental origins of a cell lineage. It also explains why active repression in *cis* is required for correct lineage specification, which is a frequent observation from genetic studies.

At the molecular level, it remains an open question why the VM-specific enhancers are occupied by the cardiac TF collective. We hypothesize that this may occur through chromatin remodeling in the precursor cell population. An "open" (accessible) chromatin state at these loci in dorsal mesoderm cells, which is most likely mediated or maintained by Tin binding prior to specification, could facilitate the occupancy of cell type-specific TFs in both CM and VM cells. Such early "chromatin priming" of regulatory regions active at later stages has been

observed during ES cell differentiation (Liber et al., 2010; Walter et al., 2008). Our data provide evidence that this also holds true for TF occupancy and not just chromatin marks. On a more speculative level, this developmental footprint of TF occupancy may reflect the evolutionary ancestry of these two organs (Pérez-Pomares et al., 2009). Visceral and cardiogenic tissues are derived from the splanchnic mesoderm in both flies and vertebrates. These complex VM-heart enhancers may represent evolutionary relics containing functional binding sites that reflect enhancer activity in an ancestral cell type.

Taken together, the collective TF occupancy on enhancers during dorsal mesoderm specification illustrates how the regulatory input of cooperative TFs is integrated in *cis*, in the absence of any strict motif grammar. We expect this more flexible mode of cooperative *cis* regulation to be present in many other complex developmental systems.

EXPERIMENTAL PROCEDURES

Chromatin Immunoprecipitation

Chromatin immunoprecipitations (ChIPs) were performed as described previously (Sandmann et al., 2006). The following antibodies were used here: rabbit anti-dTCF (M. Bienz), rabbit anti-pMad (C.-H. Heldin), rabbit anti-Doc2 (M. Frasch), and guinea-pig anti-Slp (H. Jackle). The rabbit anti-Pannier serum was generated in this study and raised against amino acids 125–294 and 206–336. The quality of each antibody was assessed by immunostains (data not shown) and western blot (Figure S5), and all ChIP data was integrated with our previously published Tin data, which was based on two independent anti-Tin antibodies. *Doc2* and *Doc3* have almost identical expression patterns and are functionally redundant and are therefore expected to occupy the same sites. Although we used an antibody directed against Doc2, we refer to the data as Doc binding to reflect the redundancy between these TFs. ChIP DNA was amplified and hybridized to Affymetrix GeneChip *Drosophila* high-density Tiling array1.0R. ChIP of endogenous loci in DmD8 cells was performed using a similar protocol 4 days posttransfection of pRM-Tin and 1 day postincubation with Wg+Dpp-conditioned medium (Figure S5B); signal was detected by real-time PCR. See Extended Experimental Procedures for more details.

Defining TF Binding Events and ChIP-Defined CRMs

Quantile normalization (Bolstad et al., 2003) was applied to the four data sets for each TF (two ChIP experiments and two mock controls) for each of the 14 conditions (seven TFs at two time points). High-confidence binding events (shown in Tables S1 and S7) were defined using TileMap (Ji and Wong, 2005). CRMs (listed in Table S2) were defined as neighboring clusters of high-confidence TF binding peaks, as described previously (Zinzen et al., 2009). Slp and Bin signals at CRMs are shown in Table S6. All ChIP data are available in ArrayExpress with accession number E-TABM-1184 and on the Furlong lab web page. See Extended Experimental Procedures for greater detail.

Autoclass Clustering of TF Binding Signals

Clustering was performed using Autoclass-C (Cheeseman, 1996) based on maximum moving average probe-wise ChIP signals (Wilczyński and Furlong, 2010) for each TF/time per CRM (window size = 200 bp). The results were filtered to exclude CRMs with maximum posterior probabilities of cluster assignment less than 0.5 and/or probabilities of best and second-best cluster assignment differing by less than 2-fold. See Table S4 for the list of classified CRMs. More details in Extended Experimental Procedures.

Transgenic Reporter Assays

CRM activity was assayed using transgenic reporter assays by placing the ChIP-defined genomic region upstream of a minimal promoter driving a *GFP* reporter gene in a modified version of pDuo2n-attB (Zinzen et al., 2009); see Extended Experimental Procedures. All constructs were targeted to chromosomal arm 3L via attB/phiC31-mediated integration (Bischof et al., 2007). Transgenic lines were balanced, homozygosed, and tested by double-fluorescent in situ hybridization using probes directed against the *GFP* reporter gene (green) and *tin* (red). CRM activity in dorsal mesoderm, cardiac mesoderm, or visceral mesoderm is readily apparent via the coexpression of *GFP* and *tin* at specific developmental stages. Images were taken using a Zeiss LSM510meta confocal microscope and were processed in Adobe Photoshop. Results are listed in Table S5 (results of double-fluorescent in situ hybridizations for selected endogenous genes are summarized in Table S3).

Motif Analysis

De novo motif discovery was performed using Weeder (Pavesi et al., 2004) on 400 bp regions surrounding the positions of the 100 highest-scoring TileMap peaks for each data set (defined as described above) and RSAT (Thomas-Chollier et al., 2008) on CRMs of the All TFs class. Motif scanning was performed using Patser (Hertz and Stormo, 1999), applying thresholds defined on the basis of specificity-sensitivity criteria (data not shown). See Extended Experimental Procedures for details and additional analyses.

ACCESSION NUMBERS

Data have been deposited under ArrayExpress accession number E-MTAB-1184.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and seven tables and can be found with this article online at doi:10.1016/j.cell.2012.01.030.

ACKNOWLEDGMENTS

We are extremely grateful to M. Bienz, C.-H. Heldin, M. Frasch, and H. Jackle for antibodies. This work was technically supported by the EMBL Genomics Core facility, with specific thanks to Jos de Graaf for array hybridizations. We thank all members of the Furlong lab for discussions and comments on the manuscript, Stijn Van Dongen for help with assessing the robustness of clustering methods, and Thomas Sandmann for designing dsRNA probes. This work was supported by a Deutsche Forschungsgemeinschaft (DFG FU 750/1) grant and Human Frontier Science Program (HFSP) grant to E.E.M.F. and postdoctoral fellowships to G.J. from EMBO and to M.S. from the EMBL EIPOD program.

Received: August 17, 2010
Revised: August 16, 2011
Accepted: January 17, 2012
Published: February 2, 2012

REFERENCES

Azpiazu, N., and Frasch, M. (1993). tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila. Genes Dev. *7*(7B), 1325–1340.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. USA *104*, 3312–3317.

Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2010). ChIP-Seq identification of weakly conserved heart enhancers. Nat. Genet. *42*, 806–810.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Brown, C.O., III, Chi, X., Garcia-Gras, E., Shirai, M., Feng, X.H., and Schwartz, R.J. (2004). The cardiac determination factor, Nkx2-5, is activated by mutual

cofactors GATA-4 and Smad1/4 via a novel upstream enhancer. J. Biol. Chem. *279*, 10659–10669.

Brown, C.D., Johnson, D.S., and Sidow, A. (2007). Functional architecture and evolution of transcriptional elements that drive gene coexpression. Science *317*, 1557–1560.

Bruneau, B.G., Nemer, G., Schmitt, J.P., Charron, F., Robitaille, L., Caron, S., Conner, D.A., Gessler, M., Nemer, M., Seidman, C.E., and Seidman, J.G. (2001). A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. Cell *106*, 709–721.

Campos-Ortega, J.A. (1997). The Embryonic Development of Drosophila Melanogaster (New York: Springer-Verlag).

Cheeseman, P. (1996). Bayesian Classification (AutoClass): Theory and Results. In Advances in Knowledge Discovery and Data Mining, U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds. (Cambridge, MA: AAAI Press/MIT Press).

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B.D., Carlson, J.W., Landolin, J.M., Kapranov, P., Dumais, J., et al. (2011). The transcriptional diversity of 25 Drosophila cell lines. Genome Res. *21*, 301–314.

Cripps, R.M., and Olson, E.N. (2002). Control of cardiac development by an evolutionarily conserved transcriptional network. Dev. Biol. *246*, 14–28.

Dai, Y.S., and Markham, B.E. (2001). p300 Functions as a coactivator of transcription factor GATA-4. J. Biol. Chem. *276*, 37178–37185.

Durocher, D., Charron, F., Warren, R., Schwartz, R.J., and Nemer, M. (1997). The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. EMBO J. *16*, 5687–5696.

Feng, X.H., Zhang, Y., Wu, R.Y., and Derynck, R. (1998). The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for smad3 in TGF-beta-induced transcriptional activation. Genes Dev. *12*, 2153–2163.

Frasch, M. (1999). Intersecting signalling and transcriptional pathways in Drosophila heart specification. Semin. Cell Dev. Biol. *10*, 61–71.

Gajewski, K., Zhang, Q., Choi, C.Y., Fossett, N., Dang, A., Kim, Y.H., Kim, Y., and Schulz, R.A. (2001). Pannier is a transcriptional target and partner of Tinman during Drosophila cardiogenesis. Dev. Biol. *233*, 425–436.

Garg, V., Kathiriya, I.S., Barnes, R., Schluterman, M.K., King, I.N., Butler, C.A., Rothrock, C.R., Eapen, R.S., Hirayama-Yamada, K., Joo, K., et al. (2003). GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature *424*, 443–447.

Halfon, M.S., Carmena, A., Gisselbrecht, S., Sackerson, C.M., Jiménez, F., Baylies, M.K., and Michelson, A.M. (2000). Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. Cell *103*, 63–74.

He, A., Kong, S.W., Ma, Q., and Pu, W.T. (2011). Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. Proc. Natl. Acad. Sci. USA *108*, 5632–5637.

Hertz, G.Z., and Stormo, G.D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics *15*, 563–577.

Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375–386.

Jakobsen, J.S., Braun, M., Astorga, J., Gustafson, E.H., Sandmann, T., Karzynski, M., Carlsson, P., and Furlong, E.E. (2007). Temporal ChIP-on-chip reveals Biniou as a universal regulator of the visceral muscle transcriptional network. Genes Dev. *21*, 2448–2460.

Ji, H., and Wong, W.H. (2005). TileMap: create chromosomal map of tiling array hybridizations. Bioinformatics *21*, 3629–3636.

Kelly, R.G., and Buckingham, M.E. (2002). The anterior heart-forming field: voyage to the arterial pole of the heart. Trends Genet. *18*, 210–216.

Kulkarni, M.M., and Arnosti, D.N. (2003). Information display by transcriptional enhancers. Development *130*, 6569–6575.

Lee, H.H., and Frasch, M. (2000). Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target sloppy paired. Development *127*, 5497–5508.

Lee, H.H., and Frasch, M. (2005). Nuclear integration of positive Dpp signals, antagonistic Wg inputs and mesodermal competence factors during Drosophila visceral mesoderm induction. Development *132*, 1429–1442.

Lee, H.H., Zaffran, S., and Frasch, M. (2006). In Development of the Larval Visceral Musculature, H. Sink, ed. (New York: Springer).

Liber, D., Domaschenz, R., Holmqvist, P.H., Mazzarella, L., Georgiou, A., Leleu, M., Fisher, A.G., Labosky, P.A., and Dillon, N. (2010). Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage. Cell Stem Cell *7*, 114–126.

Lien, C.L., Wu, C., Mercer, B., Webb, R., Richardson, J.A., and Olson, E.N. (1999). Control of early cardiac-specific transcription of Nkx2-5 by a GATA-dependent enhancer. Development *126*, 75–84.

Liu, Y.H., Jakobsen, J.S., Valentin, G., Amarantos, I., Gilmour, D.T., and Furlong, E.E. (2009). A systematic analysis of Tinman function reveals Eya and JAK-STAT signaling as essential regulators of muscle development. Dev. Cell *16*, 280–291.

Lockwood, W.K., and Bodmer, R. (2002). The patterns of wingless, decapentaplegic, and tinman position the Drosophila heart. Mech. Dev. *114*, 13–26.

Molkentin, J.D., Antos, C., Mercer, B., Taigen, T., Miano, J.M., and Olson, E.N. (2000). Direct activation of a GATA6 cardiac enhancer by Nkx2.5: evidence for a reinforcing regulatory network of Nkx2.5 and GATA transcription factors in the developing heart. Dev. Biol. *217*, 301–309.

Nishita, M., Hashimoto, M.K., Ogata, S., Laurent, M.N., Ueno, N., Shibuya, H., and Cho, K.W. (2000). Interaction between Wnt and TGF-beta signalling pathways during formation of Spemann's organizer. Nature *403*, 781–785.

Olson, E.N. (2006). Gene regulatory networks in the evolution and development of the heart. Science *313*, 1922–1927.

Panne, D. (2008). The enhanceosome. Curr. Opin. Struct. Biol. *18*, 236–242.

Pavesi, G., Mereghetti, P., Mauri, G., and Pesole, G. (2004). Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. Nucleic Acids Res. *32*(Web Server issue), W199–W203.

Pérez-Pomares, J.M., González-Rosa, J.M., and Muñoz-Chápuli, R. (2009). Building the vertebrate heart - an evolutionary approach to cardiac development. Int. J. Dev. Biol. *53*, 1427–1443.

Reim, I., and Frasch, M. (2005). The Dorsocross T-box genes are key components of the regulatory network controlling early cardiogenesis in Drosophila. Development *132*, 4911–4925.

Roider, H.G., Kanhere, A., Manke, T., and Vingron, M. (2007). Predicting transcription factor affinities to DNA from a biophysical model. Bioinformatics *23*, 134–141.

Sandmann, T., Jakobsen, J.S., and Furlong, E.E. (2006). ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos. Nat. Protoc. *1*, 2839–2855.

Schlesinger, J., Schueler, M., Grunert, M., Fischer, J.J., Zhang, Q., Krueger, T., Lange, M., Tönjes, M., Dunkel, I., and Sperling, S.R. (2011). The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. PLoS Genet. *7*, e1001313.

Sepulveda, J.L., Belaguli, N., Nigam, V., Chen, C.Y., Nemer, M., and Schwartz, R.J. (1998). GATA-4 and Nkx-2.5 coactivate Nkx-2 DNA binding targets: role for regulating early cardiac gene expression. Mol. Cell. Biol. *18*, 3405–3415.

Senger, K., Armstrong, G.W., Rowell, W.J., Kwan, J.M., Markstein, M., and Levine, M. (2004). Immunity regulatory DNAs share common organizational features in Drosophila. Mol. Cell *13*, 19–32.

Sun, G., Lewis, L.E., Huang, X., Nguyen, Q., Price, C., and Huang, T. (2004). TBX5, a gene mutated in Holt-Oram syndrome, is regulated through a GC box and T-box binding elements (TBEs). J. Cell. Biochem. *92*, 189–199.

Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev. Cell *18*, 359–370.

Takeuchi, J.K., and Bruneau, B.G. (2009). Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. Nature *459*, 708–711.

Thomas-Chollier, M., Sand, O., Turatsinze, J.V., Janky, R., Defrance, M., Vervisch, E., Brohee, S., and van Helden, J. (2008). RSAT: Regulatory sequence analysis tools. Nucleic Acids Res *36*, W119–W127.

Walter, K., Bonifer, C., and Tagoh, H. (2008). Stem cell-specific epigenetic priming and B cell-specific transcriptional activation at the mouse Cd19 locus. Blood *112*, 1673–1682.

Wilczyński, B., and Furlong, E.E. (2010). Dynamic CRM occupancy reflects a temporal map of developmental progression. Mol. Syst. Biol. *6*, 383.

Xu, X., Yin, Z., Hudson, J.B., Ferguson, E.L., and Frasch, M. (1998). Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the Drosophila mesoderm. Genes Dev. *12*, 2354–2370.

Zaffran, S., and Frasch, M. (2002). Early signals in cardiac development. Circ. Res. *91*, 457–469.

Zaffran, S., Küchler, A., Lee, H.H., and Frasch, M. (2001). biniou (FoxF), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in Drosophila. Genes Dev. *15*, 2900–2915.

Zaffran, S., Xu, X., Lo, P.C., Lee, H.H., and Frasch, M. (2002). Cardiogenesis in the Drosophila model: control mechanisms during early induction and diversification of cardiac progenitors. Cold Spring Harb. Symp. Quant. Biol. *67*, 1–12.

Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., and Furlong, E.E. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature *462*, 65–70.