Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis

Graphical Abstract

Highlights
- Complex interdependency between transcription factors controls cardiac differentiation
- Cooperative interaction between heterotypic factors relies on interdependent binding
- Preferential motif arrangements dictate heterotypic factor genomic interactions
- Interdependent factor binding prevents distribution to inappropriate sites

Authors
Luis Luna-Zurita, Christian U. Stirnimann, Sebastian Glatt, ..., Katherine S. Pollard, Christoph W. Müller, Benoit G. Bruneau

Correspondence
cmueller@embl.de (C.W.M.), benoit.bruneau@gladstone.ucsf.edu (B.G.B.)

In Brief
Genome-scale cooperative interactions between cardiac transcription factors coordinate gene expression during cardiac differentiation and morphogenesis. Cooperative DNA binding depends on preferential motif arrangements and serves not only to activate lineage-appropriate genes, but also to prevent transcription factors from redistributing to other genomic sites and activate lineage-inappropriate genes.

Accession Numbers
GSE72223
5FLV

Luna-Zurita et al., 2016, Cell 164, 999–1014
February 25, 2016 ©2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2016.01.004
Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis

Luis Luna-Zurita, Christian U. Stirimann, Sebastian Glat, Bogac L. Kaynak, Sean Thomas, Florence Baudin, Md Abul Hassan Samee, Daniel Ho, Eric M. Small, Maria Mileikovsky, Andras Nagy, Alisha K. Holloway, Katherine S. Pollard, Christoph W. Müller, and Benoit G. Bruneau

SUMMARY

Transcription factors (TFs) are thought to function with partners to achieve specificity and precise quantitative outputs. In the developing heart, heterotypic TF interactions, such as between the T-box TF TBX5 and the homeodomain TF NKX2-5, have been proposed as a mechanism for human congenital heart defects. We report extensive and complex interdependent genomic occupancy of TBX5, NKX2-5, and the zinc finger TF GATA4 coordinately controlling cardiac gene expression, differentiation, and morphogenesis. Interdependent binding serves not only to co-regulate gene expression but also to prevent TFs from distributing to ectopic loci and activate lineage-inappropriate genes. We define preferential motif arrangements for TBX5 and NKX2-5 cooperative binding sites, supported at the atomic level by their co-crystal structure bound to DNA, revealing a direct interaction between the two factors and induced DNA bending. Complex interdependent binding mechanisms reveal tightly regulated TF genomic distribution and define a combinatorial logic for heterotypic TF regulation of differentiation.

INTRODUCTION

Transcriptional regulation during differentiation and organogenesis relies on the fine quantitative regulation of thousands of genes. Within a particular cell type, defined sets of DNA-binding transcription factors (TFs) functioning on defined genomic elements are involved in determining cellular identity (Davidson, 2010; Levo and Segal, 2014). Combinatorial interactions of small numbers of TFs have been proposed to underlie tissue-specific gene expression in Drosophila embryonic mesoderm development and in mammalian organogenesis (De Val et al., 2008; He et al., 2011; Junion et al., 2012; Siersbæk et al., 2014; Steffova et al., 2013; Tijssen et al., 2011; Tsankov et al., 2015; Wilson et al., 2010; Zaret and Grompe, 2008; Zinzen et al., 2009). Despite the suggestion of combined function from occupancy or gain-of-function biochemical studies (e.g., Jolma et al., 2015; Spitz and Furlong, 2012), the functional output of interactions between developmentally important TFs and their genomic and structural basis have not been explored on a broad scale or in sufficient detail. The nature and importance of heterotypic TF interactions are therefore incompletely understood.

The T-box TF TBX5 and the homeodomain TF NKX2-5 regulate several aspects of heart development (Bruneau et al., 1999; Lyons et al., 1995; Pashmforoush et al., 2004; Tanaka et al., 1999). TBX5 and NKX2-5 can interact physically, and their potential for co-activation of target genes is apparent in their synergistic action in vitro on reporter constructs (Bruneau et al., 2001; Hiroi et al., 2001). This interaction has been proposed to be the basis for the similar congenital heart defects caused by mutations in either TBX5 or NKX2-5 in humans (Basson et al., 1997; Li et al., 1997; Schott et al., 1998). Similarly, mutations in the zinc finger TF gene GATA4 result in congenital heart defects, and GATA4 has been proposed to be a functional partner of TBX5 and NKX2-5 (Durocher et al., 1997; Garg et al., 2003). Indeed, mice doubly heterozygous for null alleles of Tbx5 and Nkx2-5 or Tbx5 and Gata4 have defects in heart formation that are more severe than those caused by each individual
mutation (Maitra et al., 2009; Moskowitz et al., 2007). Relative doses of TBX5, NKKX2-5, and GATA4 are clearly important, but it is not known what the overlapping roles are for these important cardiac TFs or how this interaction is coordinated at the genomic or molecular level. TBX5 and GATA4 can, together with additional factors, induce cardiac gene expression programs de novo (Ieda et al., 2010; Qian et al., 2012; Song et al., 2012; Takeuchi and Bruneau, 2009). Therefore, they are clearly positioned in a robust gene-regulatory network. *Drosophila* orthologs of TBX5, NKKX2-5, and GATA4 function together as a collective unit with shared binding that does not rely on motif grammar (Junion et al., 2012). Whether this mechanism is retained in mammals is not known. A mechanistic understanding of the collaborative function of TBX5, NKKX2-5, and GATA4 will have important implications for understanding congenital heart disease and for developing strategies for cardiac regeneration and, more broadly, for understanding fundamental principles of heterotypic TF interactions.

Here, we examined the overlapping function of TBX5 and NKKX2-5 in mouse cardiac differentiation and morphogenesis and their relationship with GATA4. We find that they function cooperatively across the genome, participating in a complex and dynamic TF network regulating programs of gene expression required for expansion of cardiac progenitors and for cardiac differentiation. This cooperative interaction relies on the interdependent binding of the three TFs. The crystal structure of the ternary complex of TBX5 and NKKX2-5 and a bona fide DNA target provides an atomic structural basis for this interdependent interaction. We also find that, surprisingly, in the absence of TBX5 or NKKX2-5 (or both), the remaining TFs redistribute to other genomic sites associated with genes with increased expression in the TF-null cells. Therefore, in addition to cooperative binding for gene activation, the interdependent binding of heterotypic factors serves to prevent TF distribution to lineage-inappropriate sites. This redistribution relies on the establishment of new DNA/protein interactions able to induce ectopic gene activation and unveils a tightly regulated genomic distribution that may underlie the dysregulation of gene expression in congenital heart defects. The genomic interactions between TBX5 and NKKX2-5, dictated by a preferential motif arrangement, define a molecular and genomic logic for the regulation of mammalian differentiation by heterotypic TFs.

**RESULTS**

**Impaired Cardiac Differentiation in Tbx5/Nkx2-5-Null Embryos**

Perinatal death of mice doubly heterozygous for Tbx5- and Nkx2-5-null alleles (Moskowitz et al., 2007) precludes a breeding scheme to generate embryos lacking both factors. Using mice carrying a null allele of Nkx2-5 (Tanaka et al., 1999) and a conditional allele of Tbx5 (Mori et al., 2006), we derived wild-type (WT), Nkx2-5-null (NKO), Tbx5-null (TKO), and Tbx5/Nkx2-5-double null (DKO) embryonic stem cells (ESCs) (Figure S1A). These ESCs were used to generate entirely ESC-derived WT, single, and double knockout (KO) embryos (Figure S1B). TKO and NKO embryonic day (E) 9.5 embryos recapitulated phenotypes reported previously (Bruneau et al., 2001; Tanaka et al., 1999; Figures 1A–10). DKO ESCs yielded embryos that died by E9.5 and had, in the most extreme cases, only a small patch of quiescent or very slowly beating heart tissue (Figures 1D, 1E, 11, 1J, 1N, and 10). Marker analysis confirmed the significant increase in phenotypic severity in the DKO embryos (Figures 1P–1W). Therefore, combined loss of Tbx5 and Nkx2-5 results in a dramatic restriction in the potential for the embryonic heart to form beyond its initial differentiation.

**A Genetic Program Regulated by TBX5 and NKKX2-5**

We examined gene expression by microarray analysis of heart tissues from E8.75 ESC-derived embryos (Figure 1X; Tables S1 and S2) and confirmed the microarray results by qRT-PCR of representative genes (Figure S1C). Concordant with the role of Tbx5 and Nkx2-5 as transcriptional activators (Bruneau et al., 2001; Durocher et al., 1996), a majority of transcripts were downregulated in DKO hearts, spread among subgroups with different patterns of response to each genetic perturbation (clusters 10–17, blue bracket) (Bruneau et al., 2001; Hiroi et al., 2001). A second major group with upregulated gene expression in DKO hearts was identified, highlighting the existence of Tbx5- and/or Nkx2-5-dependent repressed genes (clusters 1–8, yellow bracket). Although major groups showed some regulatory contexts depending on only one or the other TF (clusters 1, 6–8, 11, 15, and 16), complex patterns were also identified (e.g., cluster 9), reflecting complexity in the genetic interaction between Tbx5 and Nkx2-5.

**Dynamic Regulation of Gene Expression by TBX5 and NKKX2-5 during CM Differentiation**

For a detailed understanding of the genetic programs regulated by these factors during cardiomyocyte (CM) differentiation and to overcome the early lethality of the mutant embryos, we used a directed differentiation system (Kattman et al., 2011; Wamstad et al., 2012) to generate cardiac precursors (CPs) and CMs from WT, TKO, NKO, and DKO ESC lines (Figure S1A). The directed differentiation protocol yielded cultures enriched (80%–90%) in cardiac Troponin T (cTnT) (+) beating CMs for all genotypes. The dynamics of differentiation were different across the four genotypes. NKO cells begin spontaneous contractions 12 hr earlier than WT cells, and TKO cells do so 12 hr later. DKO cells begin beating approximately 6 hr later than WT cells, suggesting a major role for TBX5 in the induction of CM differentiation and for Nkx2-5 as a repressor of premature cardiac differentiation (Figure 2A; Prall et al., 2007). We examined gene expression by RNA sequencing (RNA-seq) at the CP and CM stages and clustered transcripts by expression pattern at each specific stage (Figures 2B and 2C; Tables S1 and S2). Although no correlation was found between TKO cells and embryonic gene expression analysis (likely because of the TKO embryonic phenotype, which is largely a hypoplasia of Tbx5-expressing tissue; Bruneau et al., 2001 and this report), for the Nkx and DKO cells, expression changes in genes detected in the in vivo microarray data and across each stage correlated well (Figures S2C and S2D).

The analysis of changes in gene ontology (GO) terms between clusters/stages allowed us to understand the temporal dynamics of Tbx5/Nkx2-5-dependent regulation of gene expression (Figures 2B and 2C; Table S2). Concordant with early
**NKO differentiation**, two major groups of **Nkx2-5-dependent** genes were identified at CPs: repressed genes (CP1 and CP2) associated with cell division and cycle progression and upregulated genes (CP5–CP9) associated with cardiac differentiation and function. Two clusters in this second group (CP8 and CP9) presented a different behavior between **NKO** and **DKO** cells, suggesting that genes in these two clusters could be responsible for the different beating initiation between **NKO** and **DKO** cells.

The dynamics of expression of these clusters showed predominant early expression of downregulated cell proliferation genes (CP1 and CP2) in contrast to later predominant expression of upregulated CM differentiation-specific genes (CP7 and CP9) (Figures 2D and S2A), finding a similar temporal pattern in genes associated with these functions identified at CMs (Figures 2E and S2B). This stage specificity is more obvious in cluster CP9, whose genes are highly expressed in WT cells at CMs, activated by both NKX2-5 and TBX5, but are repressed by NKX2-5 at CPs. These results confirm the role of NKX2-5 as a repressor of premature cardiac differentiation (Prall et al., 2007). At the CM stage, although NKX2-5 seems to still promote cell proliferation (CM10), **NKO** cells showed upregulated (CM5, CM9, and CM11) and downregulated (CM6 and CM7) cardiac differentiation genes. In contrast, TBX5 functions as a repressor of proliferative genes (cluster CM10), and is primarily required for CM differentiation (CM3, CM5, CM9, and CM11). We also found clusters with increased or ectopic gene expression in single and double
Genomic Occupancy of TBX5, Nkx2-5, and GATA4 during CM Differentiation

To understand mechanisms underlying the dynamic gene regulation described above, we examined the genomic localization of TBX5, Nkx2-5, and GATA4 at the CP and CM stages by chromatin immunoprecipitation with exonuclease (ChiP-exo) (Rhee and Pugh, 2011; Figure 3A). As negative controls for Nkx2-5 and TBX5, we used NKO, TKO, and DKO cells. Any footprint (fp) remaining in the KO of the respective factor was discarded. We identified 4,985, 8,718, and 11,000 fps at CPs and 8,952, 25,381, and 10,641 fps at CMs for TBX5, Nkx2-5, and GATA4, respectively. Concordant with the dynamics of Tbx5 and Nkx2-5 expression, stage-specific occupancy analysis revealed varied temporal patterns, including single-stage occupancy events and sites occupied in CPs and CMs (Figure 3B). Taking advantage of the high resolution of ChiP-exo, we characterized TBX5, Nkx2-5, and GATA4 binding at each stage, performing de novo motif discovery on the fps for each TF. In both stages, the most significantly enriched motifs in Nkx2-5 and GATA4 fps matched motifs characterized previously (Figure 3C). For TBX5, the most enriched motif in both stages has a 6-bp consensus sequence, CTG(T/C)A, corresponding to the complementary strand of the MEIS1 motif core (TGACAG), which does not match in vitro-derived, 8-bp TBX5 consensus motifs (AGGTGTGA) (Ghosh et al., 2001; Jolma et al., 2013) but is highly similar to the 3′ end of alternate motifs identified for TBX3, TBX5, and TBX20 (He et al., 2011; Sakabe et al., 2012; Shen et al., 2011; van den Boogaard et al., 2012; Figure 3C). We also found a second significant TBX5 motif (GAGGTG) that shares 5 bp with the reported 8-bp consensus motifs. We refer to this secondary motif as 5p-TBX5 and the primary motif as 3p-TBX5. Although 5p-TBX5 and 3p-TBX5 appear in the same fps, they are most often not located directly next to each other. Instead, they occur without any consistent pattern of spacing, except in the strongest TBX5 fps (top 500 bound loci), where they appear together as a longer combined motif (He et al., 2011; Figure 3D) considered previously a single motif. This new TBX5 motif composition suggests the existence of more complex mechanisms regulating TBX5 occupancy during CM differentiation.

TBX5, Nkx2-5, and GATA4 Co-occupy Genomic Loci during CM Differentiation

Our occupancy analysis revealed dynamic and extensive co-binding of TBX5, Nkx2-5, and GATA4 at the CP and CM stages (Figures 3B and 3E), and motif analysis revealed a statistically significant enrichment of motifs for the other two TFs within fps for TBX5, Nkx2-5, and GATA4 (Jolma et al., 2013). Examining heterotypic motif spacing and/or orientation pattern between the instances of the co-occurring motifs, we found consistent spacing of Nkx2-5 and TBX5 motifs in fps for both TFs more often than expected by chance. Specifically, 3p-TBX5-Nkx2-5 (3p-TBX5 the primary motif) appears frequently with 4-bp (10% of cases) or 0-bp (i.e., immediately adjacent, 9%) spacing, and Nkx2-5–3p-TBX5 (Nkx2-5 the primary motif) often occurs with 4-bp spacing (12% of cases) (Figure 3F, E values < 10−5; Jolma et al., 2013). In both cases, we found that the motifs tended to co-occur on the same strand with a head-to-head orientation, but we did not find any bias in ordering. No other TF pairs (GATA4-TBX5 and GATA4-Nkx2-5) have conserved spacing across fps, supporting the notion of cooperative binding of Nkx2-5 and TBX5 and indicating a preferential use of the 3p-TBX5 motif in these instances.

Sites where TBX5, Nkx2-5, and GATA4 co-occur (TN and TNG sites; Figure 3G) corresponded with the highest signal. These sites also have the highest degree of overlap with active enhancers (Wamstad et al., 2012) at both stages (Figure 3H) and with proximal regions (±20 kb) to differentially expressed genes (DEGs) upon loss of either factor (Figure 3H). Therefore, shared binding of multiple heterotypic TFs is associated with the most active cardiac enhancers and the most active regulation of target genes. An important question about co-occupancy is whether multiple TFs indeed bind the same piece of DNA. We addressed this question by performing sequential ChiP-exo (re-ChiP-exo) for TBX5 and Nkx2-5 at CMs in WT cells (Figure S3A; Experimental Procedures). TBX5(+)-Nkx2-5(+) single ChIP-exo fps (TN and TNG sites) overlapping Re-ChiP-exo fps (11.1% of the TBX5-Nkx2-5 sites, 6% of TN and 15.2% of TNG sites) correspond to the highest ChiP-exo signal intensity for each TF (Figure S3B), and TNG Re-ChiP-exo-positive sites have the highest association with DEGs (Figure S3C).

These results reveal the high relevance of sites with the highest probability of co-binding by TBX5, Nkx2-5, and GATA4 in the control of gene expression during cardiac differentiation.

Interdependent Genomic Occupancy of TBX5, Nkx2-5, and GATA4

The complex gene expression patterns in single and double KOs suggested the existence of complex interrelationships between TBX5, Nkx2-5, and GATA4. To understand how the occupancy of one factor functionally relates to the binding of others, we performed ChiP-exo for TBX5, Nkx2-5, and GATA4 in NKO, TKO,
Figure 3. Dynamic DNA Occupancy of TBX5, NKX2-5, and GATA4 during Cardiomyocyte Differentiation

(A) Visualization of TBX5, NKX2-5, and GATA4 ChIP-exo data at the CP and CM stages in representative loci (positive for acetylation of lysine 27 of histone H3 (H3K27ac)) at CPs and/or CMs; Wamstad et al., 2012).

(B) Clustering of TBX5, NKX2-5, and GATA4 footprints at the CP and CM stages.

(C) The most enriched binding motifs for TBX5, NKX2-5, and GATA4 at the CP and CM stages compared with published motifs.

(D) The most enriched motifs for TBX5 in the top 500 and the entire set of peaks identified for TBX5 in this report and by He et al. (2011).

(legend continued on next page)
and DKO cells (Figure S4A). We found various behaviors for TBX5 and NKX2-5: independent binding, one TF dependent on the other’s presence, and interdependent binding (both require the other’s presence). Similarly, GATA4 binding was often dependent on the presence of TBX5, NKX2-5, or both. Unexpectedly, we observed significant ectopic binding of the remaining TFs when the other was ablated (e.g., TBX5 ectopic binding at the Corin and Tbx3 loci in NKO cells; Figure S4A).

To better classify these behaviors, and interrogate TF binding relationships, TF fps were clustered according to the WT/KO signal ratio (Figures 4A, TBX5 WT/NKO; 4B, NKX2-5 WT/TKO; 4C, GATA4 WT/NKO; 4D, GATA4 WT/TKO; and 4E, GATA4 WT/ DKO; red frames), plotting over them the occupancy density for each partner factor for each genotype (Figures 4A–4E). The highest partner WT co-occupancy densities (NKX2-5 for TBX5, Figures 4A; TBX5 for NKX2-5, Figure 4B; and TBX5 and NKX2-5 for GATA4, Figures 4C–4E; dotted black frames) were found around WT fps independently of their behavior in KO cells, and the lowest was around the ectopic KO occupancy sites. A similar result was found by plotting the TBX5 and NKX2-5 ChIP-exo signal over fps only found in WT cells, in WT and KO cells, and only in KO cells (NKX2-5 signal over TBX5 fps, Figure S4B; TBX5 signal over NKX2-5 fps, Figure S4C), where TBX5-NKX2-5 overlap was found mostly in WT and WT/KO occupancy sites (Figures S4B and S4C, dotted lines), but most of the ectopically occupied regions were unoccupied in WT cells by the partner factor.

Of note, TBX5/NKX2-5 NKO/TKO ectopic occupancy was frequently accompanied by GATA4 (Figures 4A and 4B, green frames), and, in WT cells, the highest WT/KO GATA4 signal ratio was associated with the highest partner occupancy (Figures 4C–4E), highlighting the close partnership between GATA4 and TBX5/NKX2-5.

Taken together, these occupancy patterns revealed that lost or preserved partner occupancy in null cells cannot be explained only by the joint/single occupancy of both factors and that ectopic occupancy cannot be explained by direct TBX5-NKX2-5-GATA4 competition, leaving the mechanisms underlying these behaviors still unclear.

Interdependent Transcription Factor Genomic Occupancy Is Essential for CM Differentiation

To understand how the different behaviors in TF occupancy affect gene expression, we defined 15 groups of occupied regions: five groups defined by the changes in TBX5 occupancy in NKO cells (Tn), NKX2-5 in TKO cells (Nt), GATA4 in NKO cells (Gn), GATA4 in TKO cells (Gt), and GATA4 in DKO cells (Gd). Each of these groups was subdivided into ectopic occupancy in KO cells (e), unaffected occupancy (u), and loss of occupancy (l) (Figures 4A–4E, scheme at right). We analyzed the statistical enrichment of each group in proximal regions (1, 5, 10, 20, 50, and 100 kb) to DEGs identified by RNA-seq (Figures 4F and S4D). Groups corresponding to ectopic occupancy of TBX5 (Tn*), NKX2-5 (Nt*), or GATA4 (Gn*, Gt*, and Gd*) were represented significantly near DEGs corresponding to clusters CM1–CM3, CM5–CM7, CM9, and CM11 (Figures 4F, 4G, and S4D), which are characterized by upregulated gene expression in those genotypes where the ectopic TF occupancy was found to be significant (Figure 4G). In particular, the expression cluster characterized for the presence of upregulated genes in all KO backgrounds (CM1) was enriched for all groups of sites exhibiting ectopic TF binding (Tn*, Nt*, Gn*, Gt*, and Gd*) (Figure 4G, red arrow). These observations strongly suggest that abnormally increased gene expression in KO cardiomyocytes can be explained in multiple loci by ectopic TF occupancy and, therefore, highlight the importance of understanding the mechanisms preventing this ectopic occupancy.

Interestingly, the three different groups for GATA4 ectopic occupancy (Gn*, Gt*, and Gd*) showed the highest direct correlation (GATA4 ectopic occupancy in the same location in different genotypes; Figure 4H, black asterisks). This implies the existence of multiple sites where GATA4 ectopic occupancy proximal to DEGs is suppressed in the presence of NKX2-5 and TBX5. The complexity of this functional interdependency is confirmed by the significant correlation between the groups where GATA4 binding was lost in single or double KO cells (Figure 4H, red asterisks) and found to be enriched significantly around DEGs. These results pointed to a complex interdependence between TFs. Clear examples for this complexity are clusters 6 and 11 (Figure 4I), where opposite TF occupancy behaviors in different KO backgrounds result in opposite changes in gene expression.

Among CM RNA-seq clusters, only CM4, CM10, and CM12–CM14 showed no enrichment in any group of occupancy sites (Figure S4D). GO analysis of these clusters did not show enrichment in specific processes of cardiomyocyte differentiation (Figure 2C, Table S2), indicating that we are able to distinguish between those genes (and processes) where TBX5, NKX2-5, and GATA4 act directly or indirectly.

Altogether these data suggest complex interplay between TBX5, NKX2-5, and GATA4, where phenotypes associated with absence of TBX5/NKX2-5 are not only due the lack of the mutated TF, or the concomitant loss of cooperative binding of partner TFs, but also to the redistribution of other TFs across the genome.

TBX5-NKX2-5 Interdependent Occupancy

To understand the mechanisms regulating TF co-binding, interdependence, and ectopic TF occupancy, we attempted to uncover a potential specific binding motif arrangement responsible for the various TF interrelationships by analyzing the
relative prevalence of binding motifs. For co-occurring TBX5 and NKX2-5 sites, we analyzed independently the different behaviors associated with single KO contexts (Figure 5A, first through fourth columns). Independent TBX5-NKX2-5 sites had a broad range of represented motifs, although with less prevalence of consensus NKX (CACTT core) or TBX (GGTGTGA) motifs, and high occurrence of GATA, MEF2, or MEIS (3p-TBX5) recognition sequences (Figure 5A, first column). This suggests that, although TBX5 and NKX2-5 consensus motifs are present at some of these sites, TBX5-NKX2-5 independent co-occupancy might depend on interactions with third partners. The opposite pattern was found in interdependent co-binding sites, with each factor’s

Figure 4. Interdependent Transcription Factor DNA Occupancy Regulates Gene Expression during Cardiomyocyte Differentiation
(A–E) TF occupancy density over TBX5 (A), NKX2-5 (B), and GATA4 (C–E) footprints aligned along the ChIP-exo signal gradient between WT and NKO (A and C), TKO (B and D) or DKO (E) CMs. Ectopic (e), unaffected (u), or lost (l) occupancy of TBX5 in NKO (Tn), NKX2-5 in TKO (Nt), GATA4 in NKO (Gn), GATA4 in TKO (Gt), or GATA4 in DKO (Gd) CMs is indicated with brackets. Right: scheme of each TF occupancy pattern.
(F) Overlap of each occupancy pattern within 1, 5, 20, 50, and 100 kb proximal to DEGs of each RNA-seq cluster at the CM stage. Red intensity represents q value enrichment.
(G) Representation of each ectopic occupancy pattern (Tne, Nte, Gne, Gte, and Gde) overlap around RNA-seq heatmap clusters identified in (F). The red arrow indicates a significant overlap of the five ectopic occupancy patterns around cluster CM1.
(H) Occupancy patterns direct correlation (yellow, positive correlation; blue, inverse correlation). The best correlation was between the Gne-Gte-Gde groups (black asterisks) and the Gn-Gt-Gd groups (red asterisks). Gata4-unaffected occupancy groups significantly overlap with most of the Gata4 occupancy groups (lines).
(I) The most represented occupancy patterns in two representative gene expression clusters (CM6 and CM11).
See also Figure S4.
Figure 5. Specific Binding Motif Composition Regulates TBX5-NKX2-5 Genome-wide Distribution
(A) Heatmap showing the relative prevalence of selected motifs in TBX5-NKX2-5 co-binding sites where occupancy is unaffected in KO cells (first column), lost in both (second column), and lost only in TKO (third column) or NKO (fourth column) CMs and sites with ectopic occupancy of TBX5 in NKO (fifth column) or NKX2-5 in TKO (sixths column) CMs.
consensus being enriched, suggesting that, even in presence of these strong binding motifs, the binding of one factor relies on the presence of the other (Figure 5A, second column). This observation was supported by the different relative motif composition that characterized single-dependency sites. Consensus TBX or NKX motifs were highly overrepresented in sites where NKX2-5 is lost in TKO cells (Figure 5A, third column) or in sites where NKX2-5 is retained in TKO but TBX5 is lost in NKO cells (Figure 5A, third column). These results clearly indicate that varied TF co-occupancy behaviors can be explained by potential anchor factors determined by a specific binding motif composition.

Ectopic occupancy sites were characterized by a similar motif composition as single-dependency co-occupancy sites but with opposite relative prevalence (Figure 5A, third and fifth columns for TBX5 and fourth and sixth columns for NKX2-5). Therefore, secondary motifs represented in single-dependency co-occupancy sites, MEIS/3p-TBX5 and MEF2 A/T-rich motifs for TBX5 and NKX2-5, respectively (Figure 5A, third and fourth columns), were the most prevalent in TBX5 and NKX2-5 ectopic occupancy (Figure 5A, fifth and sixth columns). This pattern was even more obvious when analyzing each factor independently. For TBX5 and NKX2-5 general occupancy (Figures 5B and 5C), although consensus motifs are highly represented in TF-independent sites (Figures 5B and 5C, center columns) and equally low prevalent in dependent and ectopic sites (Figures 5B and 5C, left and right columns), secondary motifs were highly represented in both independent and ectopic sites (Figures 5B and 5C, center and right columns). Similar results were obtained comparing GATA4 TBX5-NKX2-5-dependent/independent occupancy (Figure S5A). These observations suggest that secondary motifs are more dependent on the genomic context than consensus motifs. This context dependency becomes more obvious when analyzing differences in motif prevalence for ectopic GATA4 occupancy, where we detected different motif enrichments depending on the lack of Tbx5, Nkx2-5, or both (Figure S5B).

Together, these results suggest that ectopic TF redistribution relies on the establishment of new motif/partner interactions, interactions that directly depend on the complex behavior of different partners in each TBX5/NKX2-5 KO context. This implies a tightly co-regulated TF co-binding where, in many sites, the interaction between different TFs prevents them, directly or indirectly, from redistributing to less favored genomic interactions and inducing ectopic gene activation. This mechanism is likely to be relevant to many heterotypic TF interactions.

**Binding Motif Conformation Regulates Heterotypic TF Occupancy**

Complex TF interactions seem to rely on a specific binding motif composition. To identify specific motif arrangements that might regulate the establishment of TF interactions, we performed a spacing/orientation/order analysis of the most represented pair of motifs for each factor (3p-TBX5 and NKX2-5 consensus motif) in TBX5-NKX2-5 co-occurrence groups (Figure 5D).

In contrast to interdependent co-binding sites, where varied spacing/orientations were allowed, single-dependency TBX5-NKX2-5 co-binding sites showed, respectively, a very specific enriched configuration of this motif pair. Therefore, a 12-bp gap/5’-5’ orientation was found in TBX5-dependent, NKX2-5-independent sites and a 4-bp gap/3’-5’ in NKX2-5-dependent TBX5-independent sites (Figure 5D). These findings strongly suggest the existence of a preferential motif distribution facilitating heterotypic TF interactions for specific contexts and interdependent behaviors.

We extended our spacing/orientation analysis to the general occupancy patterns for each factor. Among others, specific arrangements for consensus TBX, NKX, MEF2, and MEIS pairs were identified to be associated with different TBX5/NKX2-5 dependent or independent occupancy patterns and also with ectopic TBX5/NKX2-5 sites (Figures 5E–5G and S6).

Together, these results support the existence of specific permissive binding motif pair distributions allowing heterotypic TF interactions. The same analysis was performed distinguishing between sites proximal (20 kb) to DEGs, distal to DEGs, or the entire set of occupancy sites, obtaining the same statistical enrichments for the same motif pairs. This result showed the strength of this TF occupancy-regulatory mechanism.

**Crystal Structure of NKX2-5/TBX5 Fusion Protein Bound to the Nppa Promoter Region**

A good example of TF interactions allowed by specific binding motif arrangement is the proximal promoter of Nppa, which, consistent with previous gain-of-function studies that demonstrated co-activation of this regulatory element (Bruneau et al., 2001; Hiroi et al., 2001), is co-occupied by TBX5 and NKX2-5 (Figures 3A and 6A). TBX5 binding at this site depends on NKX2-5, showing cooperative binding, and has the most preferential motif arrangement defined for loci with this interdependent behavior (Figures 6A and S3). We structurally characterized the ternary complex of TBX5 and NKX2-5 on the Nppa promoter by X-ray crystallography. After unsuccessful initial trials to crystallize the TBX5 T-box domain (TBD, amino acids (aa) 51–251) and the NKX2-5 homeodomain (HD, aa 134–197) bound to the Nppa promoter region (positions −252 to −234), we engineered a fusion protein between NKX2-5HD and TBX5TBD with a poly-serine linker (Figure 6B). The NKX2-5(aa134-197)155RecTBX5(aa51–251) construct (NKX-TBXlinked) expressed well in E. coli, was purified to homogeneity and associated stably with DNA (Figure 6B). We obtained crystals of the TBX5-NKX2-5-DNA complex fractions by vapor diffusion that diffracted to 3.0-Å resolution. The crystal structure...
was solved by molecular replacement with the known individual crystal structures and refined to $R_{\text{work}}/R_{\text{free}}$ values of 19.0%/24.1% (Table S3).

The structure shows that TBX5 and NKX2-5 specifically bind to their respective recognition sequences in the Nppa promoter, which harbors both sites in succession (Figure 6C). The linker is not visible in our electron density. It is flexible and does not influence the positioning of the individual proteins. We superimposed individual DNA binding regions of TBX5 and NKX2-5 with known crystal structures of the individual proteins bound to DNA (Pradhan et al., 2012; Stirnimann et al., 2010) and could not observe changes in the orientation of the residues involved in DNA binding.
Figure 7. Detailed Analysis of the Interactions between NKX2-5 and TBX5 on Nppa DNA

(A) Schematic summary of protein-DNA interactions. Only direct contacts between protein and DNA molecules are labeled because no water molecules were built.
binding (Figures S7A and S7B). In detail, z-helix 3 of TBX5 TBD (aa 232–237) is structured, and specific residues make direct contact with the DNA bases (F236 and F232). For NKX2-5, we also found almost identical interactions with residues F144, Q180, K182, W184, Q186, N187, R189, and Y190 contacting the DNA (Figure 7A). Some of the contacts in the high-resolution structures of the binary TBX5-DNA and NKX2-5-DNA complexes (Pradhan et al., 2012; Stirnimann et al., 2010) are mediated by water molecules that cannot be positioned reliably at the resolution obtained for the TBX5-NKX2-5-DNA complex. Nonetheless, the orientations of the individual side chains of the residues involved in these water-mediated contacts are almost identical to their orientations in high-resolution structures. We therefore conclude that, in the ternary complex, the protein-nucleic acid interactions of TBX5 and NKX2-5, with their respective DNA recognition sequences, are identical to the previously determined binary complex structures (Pradhan et al., 2012; Stirnimann et al., 2010). Although we did not observe any global structural rearrangements of the TBX5 and NKX2-5 proteins in comparison with their individual DNA-bound species or unbound TBX5 (Figure 7B), the overall conformation of the Nppa promoter DNA significantly deviates from idealized B-DNA or DNA bound to a Tbx3 dimer (Figure 7C). In the ternary complex, the Nppa promoter DNA shows strong bending that might be induced by a small (~150 Å²) and previously unknown protein interaction surface between the TBX5 and NKX2-5 DNA binding domains of the two TFs (Figure 7D). These findings strongly support our observation of specific binding motif distribution allowing stable TF interactions.

**TBX5 and NKX2-5 Interaction Interface**

We identified three pairs of amino acid residues involved in a direct TBX5-NKX2-5 interaction (Figure 7D) and investigated the importance of this newly described interaction on their binding to the Nppa promoter. Two adjacent surfaces in TBX5 have high evolutionary sequence conservation. One patch is conserved among TBX2/3/4/5/6 proteins, and the other is only conserved among TBX5 proteins (Figure S7C). The involved residues in TBX5 reside close to helix a2, and the residues in NKX2-5 HD are located at its highly conserved C terminus. We used a mutational approach to assess the contribution of the interaction itself and individual residues to the binding of NKX2-5 and TBX5 to the Nppa promoter site. Using a filter-binding assay, we show that TBX5 TBD and NKX2-5 HD proteins separately bind the Nppa promoter with apparent K_D values of ~500 nM, whereas, together, TBX5 TBD and NKX2-5 HD bind the Nppa promoter with one order of magnitude-enhanced affinity (apparent K_D, ~50 nM). The sigmoidal binding curve suggests cooperative binding. The fusion construct NKX-TBX5_HD shows the same enhanced binding affinity as the combination of TBX5 TBD and NKX2-5 HD, demonstrating no effect of covalently linking the two factors (Figure 7E, top). Mutating individual residues in the TBX5-NKX2-5 interface into alanine (K157, Q195, and R196 in NKX2-5 and P139, D140, R150, and Q151 in TBX5) reduced the binding of NKX2-5 HD and TBX5 TBD to the Nppa promoter, whereas combining mutations of several residues further enhanced this effect (Figures 7E, bottom, and S7D).

These results suggest that the newly observed TBX5-NKX2-5 interface greatly contributes to the cooperative binding of both factors to the Nppa promoter. We therefore provide a possible mechanism for the specific interplay between TBX5 and NKX2-5 on this region and, likely, for other interdependent TBX5 and NKX2-5 binding elements.

**DISCUSSION**

We defined the genomic and structural basis for dynamic control of gene expression during cardiomyocyte differentiation by the interdependent heterotypic TFs TBX5 and NKX2-5. Although many studies have reported the capacity of these (and other) TFs to cooperate in the control of gene expression, our results showed that this phenomenon occurs in vivo, with significantly more complexity and fine-tuning than anticipated. We find that heart development and cardiac gene expression necessitate tight integration between TBX5, NKX2-5, and, among others, GATA4. This integration depends directly on TF binding motif arrangement, which facilitates physical interactions between TFs.

The combined loss of TBX5 and NKX2-5 in vivo shows that their shared regulatory programs are essential for critical aspects of heart development, including expansion of cardiac progenitors and important elements of the cardiac differentiation program. This genetic interaction has been proposed to be at the root of human congenital heart defects (Brunneau et al., 2001, 2008; Hiroi et al., 2001), and our results support this genetic link, also demonstrating the implication of additional partners in this interaction. A complex but defined motif arrangement dictates the interrelationships between these TFs, resulting in a gene-regulatory network deployed during the early stages of heart development to ensure robust tissue-specific gene expression.

In accordance with their dynamic regulation of gene expression, genomic occupancy of TBX5, NKX2-5, and GATA4 follows combinatorial rules that neatly predict their regulatory consequences. Occupancy data from He et al. (2011) for TBX5, NKX2-5, and GATA4 do not agree with the results presented (B) Structures of TBX5 TBD bound to DNA (yellow, PDB ID 2X6V) and unbound TBX5 TBD (brown, PDB ID 2X6U) superimposed on TBX5 TBD from NKX-TBX5 linked (red) are shown in ribbon representation. The structure of NKX2-5 HD bound to DNA (cyan, PDB ID 3RHQ) superimposed on NKX2-5 HD from NKX-TBX5 linked (red) is shown in ribbon representation. DNA from the NKX-TBX5 complex is shown in schematic representation (gray).

(C) Structural comparison between DNA molecules from NKX-TBX5 linked (red/grey/blue) and individually bound TBX5 (yellow), NKX2-5 (cyan), TBX3 (purple, PDB ID 1H6F), and idealized B-DNA of the Nppa promoter (orange). Lines indicate the central tracks of each individual DNA molecule.

(D) Close-up view of the interface between NKX2-5 HD (blue, schematic) and TBX5 TBD (red, surface) highlighting the position of contributing and mutated residues.

(E) Graphs of filter-binding assays using radioactively labeled DNA in combination with different purified proteins (indicated below).

(F) Same as (E) using indicated point mutants. NKX2-5 M1 corresponds to NKX2-5_1_22–230 K157A, NKX2-5 M2 to NKX2-5_1_22–230 Q195A/R196A, and TBX5 M3 to TBX5_1_251 P139A/D140A/R150A/Q151A. See also Figure S7.
here, perhaps because of their use of a highly divergent transformed atrial cell line and ectopic binding events from over-expression. Nonetheless, a common theme is the correlated binding of the three TFs with active cardiac enhancers. Importantly, we identify a strong relationship between differentially expressed genes and altered TF binding in the absence of TBX5, NKX2-5, or both. Our results establish the existence of broadly relevant cooperative interactions between heterotypic TFs essential in defining the cardiac differentiation program.

We uncovered a motif logic relying on the different interdependent relationships between both factors. Besides the recognition of specific binding motifs, TF occupancy depends to a high degree on the interaction with other TFs, giving rise to TF complexes where the strength of interaction is dictated by the combination of TF-TF and TF-DNA affinities. We also observed permissive binding motif configurations, highlighting the role of DNA as an instructive partner for pairs of heterotypic TFs (Jolma et al., 2015). Our structural studies proved to be a biophysical basis for these interactions and, additionally, suggest important TF-TF protein interactions away from the DNA-binding interface. Studies in Drosophila of TBX5, NKX2-5, and GATA4 orthologous genes suggest that TF-co-binding could be mediated by motifs without specific orientation/order (Junion et al., 2012). The motif logic that we observe, correlating with shared binding behavior and gene regulation, suggests a complex but well defined TF binding site arrangement that regulates mammalian cardiac gene expression.

Heterotypic TF complexes also prevent the establishment of less favored interactions based on the stabilization of new complexes by DNA sequence recognition and/or physical TF interaction. The loss of one or more factors unbalances the optimal co-activation of target loci and results in concomitant redistribution to less favored genomic interactions and ectopic gene expression. Therefore, we propose a dual role of co-binding where interactions between anchor TFs and their partners serve not only to induce gene expression but also to maintain partner TF localization away from undesired genomic locations. Maintenance of this equilibrium appears to be a crucial mechanism for controlling gene expression and cell fate acquisition.

In summary, we showed that TBX5 and NKX2-5 are intercrosses following standard methodology (Nagy et al., 2003). The Tbx5floxed/Nkx2-5floxed and Tbx5floxed/Nkx2-5flox2flox lines were transiently transfected with a Cre-expressing construct to generate the following sub-lines: Tbx5floxed/Nkx2-5floxed (WT), Tbx5floxed/Nkx2-5flox2flox (NKO), Tbx5floxed/Nkx2-5flox2flox (TKO), and Tbx5floxed/Nkx2-5flox2flox (DKO) (Figure S1A). Directed differentiation into cardiomyocytes was performed according to methods published previously (Kattman et al., 2011; Wamstad et al., 2012). Tetraploid or morula aggregation was used to generate ESC-derived embryos (Nagy et al., 2003).

**EXPERIMENTAL PROCEDURES**

**Transgenic ESCs**

The transgenic mouse lines have been described previously (Bruneau et al., 2001; Mori et al., 2006; Tanaka et al., 1999). ESCs were derived from mouse interspecies following standard methodology (Nagy et al., 2003). The Tbx5floxed/Nkx2-5flox2flox and Tbx5floxed/Nkx2-5flox2flox lines were transiently transfected with a Cre-expressing construct to generate the following sub-lines: Tbx5floxed/Nkx2-5floxed (WT), Tbx5floxed/Nkx2-5flox2flox (NKO), Tbx5floxed/Nkx2-5flox2flox (TKO), and Tbx5floxed/Nkx2-5flox2flox (DKO) (Figure S1A). Directed differentiation into cardiomyocytes was performed according to methods published previously (Kattman et al., 2011; Wamstad et al., 2012). Tetraploid or morula aggregation was used to generate ESC-derived embryos (Nagy et al., 2003).

Analyses of Gene Expression

Affymetrix mouse gene ST 1.0 arrays were hybridized and scanned according to the manufacturer’s recommendations. Linear models were fitted for each gene comparison of paired genotypes to derive the mutant effect using the Limma package in R/Bioconductor. For RNA-seq, total RNA was isolated from 3 × 10⁶ cells using TRIzol reagent. RNA-seq libraries were prepared according to Life Technologies Ion Total RNA-seq Kitv2 and paired end 100 bp-sequenced on an Illumina HiSeq 2500 instrument.

Chromatin Immunoprecipitation

ChIP-exo was performed according to methods published previously (Boyer et al., 2005; Serandour et al., 2013; Wamstad et al., 2013) using specific antisera to TBX5 (sc-17866 XS, lot no. B1213), NKX2-5 (sc-8697 XS, lot no. B1213), and GATA4 (sc-1237 XS, lot no. B1213) on chromatin isolated from 10⁶ cells. Re-ChIP-exo was performed from 4 × 10⁵ cells, combining methods published previously (Serandour et al., 2013; Shankaranarayanan et al., 2011).

Structure Determination

A fusion protein (NKK-TBX15) was produced that included the NKX2-5 homeodomain (aa 134-197), the TBX5 T-box domain (aa 51-251), and a 15-aa-long poly-serine linker between the NKX2-5 C terminus and the TBX5 N terminus. Crystals were obtained using the vapor diffusion method. Diffraction data to 3.0 Å were collected at D14-1 at the European Synchrotron Radiation Facility Grenoble, France. The structure was solved using several iterative molecular replacement and manual rebuilding steps.

**ACCESSION NUMBERS**

ChIP-exo and RNA-seq data are available in experiments 353R (RNA-seq) and 189R1 (ChIP-exo) at https://b2b.hci.utah.edu/gnomex/. The GEO/SRA series accession numbers for the microarray, ChIP-exo, and RNA-seq data are GSE72223 and SRP062699. The atomic coordinates and structure factors of the NKX2-5-GBA2 DNA complex have been deposited with the European Protein Data Bank (PDB) under accession code 5FLV.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.01.004.

**AUTHOR CONTRIBUTIONS**

J.L.Z. generated most of the experimental data, performed most of the analyses, and wrote the manuscript. C.U.S obtained crystals and solved the crystal structure. C.U.S. and S.G. refined and analyzed the crystal structure. E.M.S., B.L.K., and M.M. isolated ESCs under the supervision of A.N. B.L.K. obtained ESC-derived embryos, performed in situ hybridizations, and extracted embryonic mRNA for microarray analysis. S.T. performed computational analyses. F.B. performed filter-binding assays. M.A.H.S. participated in the motif analyses. D.H. participated in the ChIP experiments. K.S.P. supervised the computational analyses with participation from A.K.H. C.W.M. supervised the structural and biophysical analyses. B.G.B. imaged the embryos, supervised and coordinated the project, and wrote the manuscript. All authors participated in the writing of the manuscript.
ACKNOWLEDGMENTS

We thank S. Izumo for providing the Nkx2.5lacZ/+ mice. We are grateful to P. Swinton (Gladeck Tranagenic Core) for morula injections. We also thank L. Ta and R. Chadwick (Gladeck Genomics Core), A. Williams (Gladeck Bioinformatics Core), C. Miller (Gladeck Histology Core), D. Hawkins (Gladeck Summer Scholars Program), and G. Capitani (Paul Scherrer Institut, Switzerland) for technical support and G. Howard for editorial assistance. This work was funded by NIH/NHLBI P01 HL089707 and R01 HL114948 (to B.G.B.), NHLBI Bench to Bassinet Program U01 HL098179/UM1 HL098179 (to B.G.B. and K.S.P.), the California Institutes for Regenerative Medicine (RN2-00903 to B.G.B.), and the Lawrence J. and Florence A. DeGeorge Charitable Trust/American Heart Association Established Investigator Award (to B.G.B.). L.L.Z. was supported by a California Institutes for Regenerative Medicine fellowship. C.U.S. is grateful for financial support from Marie Curie Framework Program 7 and the European Molecular Biology Organization long-term fellowship program. This work was also supported by an NIH/NCRR grant (CO6 RR018526) to the J. David Gladstone Institutes, by BioFulcrum: A Gladstone Institutes Enterprise, and by William H. Younger, Jr. (to B.G.B.).

Received: August 23, 2015
Revised: November 21, 2015
Accepted: December 22, 2015
Published: February 11, 2016

REFERENCES


