As one of the first organs to develop, the heart pumps nutrients, including oxygen, to the growing embryo. From embryonic stem cells, mesoderm, cardiac precursors to cardiomyocytes, these differentiating cells undergo movement, proliferation, and death. The mammalian heart undergoes intricate morphogenesis. At least 2 fields of cardiac progenitors migrate midline to form a linear heart tube, which subsequently loops rightward. Neural crest cells contribute to extensive remodeling to form the mature 4-chambered heart. Aberrations in cardiac development lead to congenital heart diseases (CHDs), a leading cause of morbidity and mortality in childhood.

Studying gene regulation of the developing cardiovascular system presents unique challenges, compared with other organ systems. As an essential organ system, environmental or genetic etiologies can cause abnormal cardiac morphogenesis and lead to embryonic demise, which can be difficult to prenatally diagnose and to study. Human heart tissue is not readily accessible and usually limited to pathological specimens. In mouse, the embryonic heart is relatively small, so only a small amount of in vivo cardiac tissue can be recovered. Despite these hurdles, tremendous progress has been made over the last 25 years to identify the genetic building blocks for cardiac development and the genetic culprits for CHDs, including transcriptional regulatory elements.

Investigating the Transcriptional Control of Cardiovascular Development

Irfan S. Kathiriya,* Elphège P. Nora,* Benoit G. Bruneau

Abstract: Transcriptional regulation of thousands of genes instructs complex morphogenetic and molecular events for heart development. Cardiac transcription factors choreograph gene expression at each stage of differentiation by interacting with cofactors, including chromatin-modifying enzymes, and by binding to a constellation of regulatory DNA elements. Here, we present salient examples relevant to cardiovascular development and heart disease, and review techniques that can sharpen our understanding of cardiovascular biology. We discuss the interplay between cardiac transcription factors, cis-regulatory elements, and chromatin as dynamic regulatory networks, to orchestrate sequential deployment of the cardiac gene expression program. (Circ Res. 2015;116:700-714. DOI: 10.1161/CIRCRESAHA.116.302832.)

Key Words: cell differentiation ■ heart diseases ■ transcription factors ■ transcriptional networks ■ transcriptional regulatory elements

This Review is part of a thematic series on Epigenetics, which includes the following articles:

Investigating the Transcriptional Control of Cardiovascular Development
Epigenetics and Metabolism
Long Noncoding RNAs in Cardiovascular Diseases
Long Noncoding RNAs and MicroRNAs in Cardiovascular Pathophysiology
HDACs and HATs in the Cardiovascular System
Role of Long Noncoding RNA in Cardiovascular Development
Application of Epigenetics to Human Studies
Epigenetics and Vascular Diseases
microRNAs as Cardiovascular Biomarkers
Reprogramming of Fibroblasts to Myocytes Using microRNA

Stefanie Dimmeler, Ali J. Marian, & Eric Olson, Editors
factors (TFs), their cofactors, and regulatory regions that precisely regulate cardiac gene expression. Insights from cardiac development in vivo has been instructive for establishing methods for directed cardiac differentiation in vitro from pluripotent stem cells, for transdifferentiation of mesoderm to the cardiac lineage and for cardiac reprogramming from fibroblasts.

A TF is a protein with sequence-dependent affinity for DNA that modulates transcriptional activity of target genes (Figure 1A). Cofactors, which interact with TFs but do not bind DNA directly, also regulate gene expression and participate in transcriptional networks. For this review, a cardiac TF is expressed in the progeny or descendants of the heart fields during development, although its expression may not be exclusive to or enriched in the heart. TF functions are targeted to specific sites of the genome called cis-regulatory elements via recognition of specific DNA motifs. Their affinity for a given DNA motif can be modulated by protein or RNA cofactors and by local physical properties of the DNA fiber and surrounding chromatin.

TFs act as a mechanistic link to the transcription apparatus for gene activation (Figure 1B). Their action culminates in recruitment of transcriptional machinery, including RNA polymerase II, to promoters and initiation of RNA synthesis. Many types of RNA species have an integral function in

<table>
<thead>
<tr>
<th>Nonstandard Abbreviations and Acronyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD</td>
</tr>
<tr>
<td>TAD</td>
</tr>
<tr>
<td>TF</td>
</tr>
</tbody>
</table>

Figure 1. Molecular players for transcriptional regulation. A. cis-regulatory elements containing DNA-binding sites are bound by transcription factors and (B) modulate the assembly of the preinitiation complex at promoters through (C) physical contacts driven by a 3-dimensional arrangement of chromatin, thereby acting as (D) a molecular platform between cellular signaling and gene activity.
control of transcriptional activity, through processes that often involve interaction with TFs or chromatin-modifying proteins, as reviewed in depth elsewhere. Many co- and post-transcriptional steps are also important in controlling gene expression, such as polymerase initiation, pausing, and elongation. We will focus on the mechanisms directly upstream of the assembly of the preinitiation complex that function during cardiac differentiation and development.

We hope to shed light on principles that govern cellular transitions relevant to cardiovascular biology and human disease. Therefore, we first describe how cardiac TFs were identified and what is known about their binding properties. We then discuss how their partnership with DNA, chromatin, and other TFs edifies regulatory elements that control transcription in cardiogenesis. Finally, we explain how these mechanisms integrate to form regulatory networks, allowing dynamic patterning of gene expression in cardiac development.

**Cardiac TFs and Cofactors**

In this section, we discuss how TFs and cofactors for cardiac development have been identified, based on evolutionary conservation, expression pattern, or function, and then discuss approaches to study their cardiac function.

**Identifying TFs by Evolutionary Conservation**

Evolutionary conservation can be leveraged to identify orthologous cardiac TF genes in invertebrates and mammalian species. _Drosophila melanogaster_ has a circulatory system composed of a rudimentary tube, the dorsal vessel, whose ontogeny is somewhat comparable to the mammalian embryonic linear heart tube. Many genes involved in specification and differentiation of the cardiac mesoderm in _Drosophila_ play a role in mammalian cardiac development. _Tinman_ (initially named _msh-2/NK4_) was first identified and shown to be required for specification of heart and visceral muscles in _Drosophila_. Soon thereafter, _Nkx2-5/Cxx_ was cloned using part of the _Drosophila msh2_ sequence as a probe for low stringency hybridization with embryonic mouse heart complementary DNA. As more mammalian TFs were identified, similar strategies based on homology were used to identify paralogous TF family members that share related structural domains. For example, _Hand2_ was cloned by phagemid library screening of mouse genomic DNA with a mouse _Hand1_ complementary DNA hybridization probe.

Sequence conservation might tempt us to think that the functions of those TFs in cardiac development are functionally conserved. As one significant example, mouse _Nkx2-5_ compensates for loss of _tinman_ in _Drosophila_, for visceral mesoderm specification and even cardio genesis, though only when appended to the _Drosophila_ specific _tinman_ N-terminal domain. These results illustrate that some degree of functional conservation may exist between orthologous cardiac TFs in mammals and flies. However, loss of function of _tinman_ or _Nkx2-5_ causes different phenotypes, demonstrating divergent functions of a cardiac TF in distinct organisms, despite an apparent conservation of its biochemical functions. This may be because of differences in mesoderm development between flies and mammals, or perhaps by targeting different regulatory elements containing similar DNA-binding motifs in a distinct genomic context.

**Identifying TFs by Expression Patterns**

Complementary methods to identify cardiac factors were based on screening for heart-specific expression. _Myocardin_ was identified by searching publicly available databases of expressed sequence tags for novel sequences present only in cardiac complementary DNA. Using a polymerase chain reaction–based modification of subtractive hybridization to enrich for embryonic chicken complementary DNAs common to the early heart field and linear heart tube but not the posterior noncardiogenic region, _Bop/Smyd1_ was isolated. In comparison, although _Baf60c_ was discovered in a screen as a target of _Wnt_ signaling, in situ hybridization demonstrated cardiac-restricted expression during development.

Important cardiac TFs were discovered by techniques based on expression in the heart. However, expression-based approaches are limited; they might miss widely expressed TFs and chromatin-modifying enzymes in cardiac development. Ubiquitous factors may function differently in various tissues, and a heart role may be important for a ubiquitous chromatin-modifying enzyme. For example, _Brg1_ modulates _myosin heavy chain_ switching during heart development and cardiac hypertrophy. Conversely, an expression-based approach may miss TFs that are not expressed in the heart, but their function and expression in other tissues are important for cardiac development.

**Identifying TFs by Function in Cardiovascular Development**

Methods to identify TFs implicated during heart development without a priori knowledge about expression pattern can be deployed, potentially revealing factors required transiently or nonautonomously to the cardiac lineage. In particular, forward genetic screens, based on circulatory system phenotypes in animal models, identified several factors necessary for cardiovascular development. For example, _N-ethyl-N-nitrosourea_–induced mutations in _zebrafish_ identified _Gridlock/Hey2/Hrt2_, which was also cloned using homology screening to the basic helix-loop-helix domain. Such loss-of-function screens are especially challenging in mice because heart dysfunction can cause early embryonic lethality. To overcome these hurdles, several groups used in utero ultrasound-based phentyping to uncover mutations with early cardiovascular developmental defects in mice. Yet, phenotype-based screens may not uncover factors that are also essential before heart formation. Many important heart TFs function in other tissues, complicating forward genetic approaches based on constitutive gene deletions. For example, a role for _Eomes_ in the specification of cardiac mesoderm is not evident, unless its early requirement for extraembryonic tissue growth and the epithelial-to-mesenchyme transition during gastrulation is bypassed.

By identifying the genetic basis of human cardiomyopathies, we might better understand human diseases. For example, linkage analysis revealed that heterozygous _TBX5_ mutations cause Holt–Oram syndrome, which includes CHDs and upper limb abnormalities, in humans. _Tbx5_ is necessary in mice and expressed in the heart. However, human linkage analysis of monogenic cardiac diseases, many of them autosomal dominant disorders, implicated only a few factors. Exome sequencing is a powerful alternative method to survey the integrity of TF-coding genes in patients with cardiovascular
Defects. Using exome sequencing of parent–offspring trios, de novo mutations of the ubiquitous TF SMAD2 and several chromatin-modifying enzymes were associated with nonsyndromic CHDs. These included the broadly expressed Trithorax-group histone methyltransferase MLL2, which was also identified in an exome study of unrelated probands with Kabuki syndrome, which includes CHDs and noncardiac anomalies. As congenital heart defects can be isolated or associated with syndromes, this may be explained by the observation that part of the molecular toolbox underlying cardiac development is also deployed in other developmental pathways. Consequently, there may be only a few, truly cardiac-specific factors.

**Investigating the Roles of TFs in Cardiac Biology**

To explore cardiac TFs in heart biology, investigators have typically studied developmental abnormalities from the loss of function of TFs. Yet, a TF can have multiple roles in controlling cardiac function, typically by acting at different developmental times or by distinct functions in different cell types of the heart. Because these functions are not readily revealed by analyzing constitutive deletions, exploring these roles often requires refined genetic analysis in animal models. For example, *Tbx5* is expressed in the left ventricle, atria, and conduction system in the developing heart. *Tbx5*-null mice die by embryonic day (E10.5) with defects in cardiac looping and hypoplasia of the left ventricle. Ventricle-restricted homozygous deletion of *Tbx5* results in a single, mispatterned ventricle and embryonic lethality by E11.5. Deletion of *Tbx5* by Mef2cAHF-Cre, which overlaps *Tbx5* in the interventricular septum, maintained a morphological and molecular distinction between left and right ventricles, but lacked an interventricular septum, consistent with a requirement of *Tbx5* for ventricular septal formation.

Endocardium-specific deletion of *Tbx5* did not cause embryonic lethality but caused 100% penetrance of atrial septal defects. Thus, TFs necessary for initial specification and expansion of the cardiac lineage, such as *Tbx5*, can also function later for heart morphogenesis.

Tissue-specific deletions of broadly expressed factors have revealed insights during heart development. Cardiac-specific deletion of *Ezh2* showed that *Ezh2* and the polycomb repressive complex-2 complex are necessary for cardiac morphogenesis and homeostasis. Despite broad expression and function, *Ezh2* has a specific role in cardiac precursors to repress a transiently expressed cardiac TF, *Six1*. This highlights a specific cardiac role for a broadly important chromatin regulator.

**TF–TF Protein Interactions**

A challenge is to understand how disruption of a TF can alter expression of hundreds or thousands of genes. How this works is unclear, but hints can be obtained from how TFs modulate transcription. TFs function within protein complexes, which include other TFs and chromatin-modifying enzymes. A candidate approach revealed several TF–TF interactions between *Gata4* and *Tbx5* or *Tbx5* and *Nkx2-5*. Unbiased approaches, such as expression cloning in mammalian or yeast 2-hybrid systems, uncovered interactions between *Tbx5* and *Nkx2-5* and *Gata* and *Fog* factors. These interactions may be crucial in heart development. For example, a single amino acid change in a mutant *Gata4* allele disrupts an interaction with the cofactor *Fog2* and leads to abnormal heart development.

Mass spectrometry can elucidate global protein interaction networks and identify post-translational modifications that modulate TF functions, such as the methylation of Gata4. With stable isotope labeling with amino acids in cell culture, the entire cellular proteome can be evaluated, including global post-translational modifications or for nuclear protein components from cardiomyocytes in *Xenopus*. These proteomic approaches provide entry points to identify other important factors in cardiac biology. However, not all TF–TF interactions are functionally relevant at all target genes. Insights into partner functionality came from characterizing how TFs contact the regulatory regions of their target genes.

**TF–DNA-Binding Elements**

Sequence specificity is conferred by properties of the DNA-binding domain of TFs. Additional factors participate in modulating TF affinity for a DNA motif, including protein or RNA cofactors, physical properties of DNA (eg, bending), post-translational modifications of the TF, and local chromatin structure. A TF typically contacts many DNA elements that can be distinct in motif composition, even in the same cell type. The general motif composition of these elements is similar within a TF family (eg, T-box versus Homeobox), with variations among TF family members.

A common approach to assess TF affinity is to perform in vitro binding assays and electromobility shift assays. Candidate sequences can be obtained using variations of a known motif of a TF from the same family. This approach identified a sequence that recruited *Tbx5* in vitro, based on similarity to a Brachyury-specific motif. In a less biased way, a polymerase chain reaction–based selective enrichment by affinity method was used to identify a similar binding motif for *Tbx5*. Complementary in vitro methods, such as mechanically induced trapping of molecular interactions, can determine quantitative TF-binding information.

Actual binding in vivo can be assessed by chromatin immunoprecipitation with sequencing and bioinformatics analysis for enriched motifs. This allows ab initio discovery of preferential binding motifs. Differences in cognate motifs for the same TF may be observed in different cell types, because a regulatory cofactor is expressed differentially, such as *Fog1* for the hematopoietic TF *Gata1*. Methods to precisely identify TF footprints, such as chromatin immunoprecipitation–exo (ChiP-exo), may greatly enhance our understanding of TF-binding specificity.

Strikingly, TFs bind a small fraction of all the DNA elements that harbor their consensus-binding site. The identity of bound sites can be different between cell types. Thus, features other than nucleotide sequences are major determinants of TF binding to their DNA elements. Chromatin structure and nucleosome positioning are major determinants of TF binding. Nucleosomes can mask the binding site of a given TF. Reciprocally, a TF can interfere with nucleosome sliding once bound to DNA, altering nucleosome positioning and affecting binding of another TF at the same site.

Another surprise from compiling chromatin immunoprecipitation with sequencing datasets is that distinct TFs often bind similar regions in a given cell type. Some cluster in stretches...
that span thousands of bases. In chromatin immunoprecipitation analysis of cardiac TFs in HL-1 cells, a mouse cardiomyocyte line, and adult mouse hearts, combinatorial binding of several TFs occurs at hundreds of regions across the genome, and binding of multiple TFs marks cardiac enhancer regions. Importantly, this suggests regulatory activity on transcription emerges from combinatorial binding and action of multiple TFs on the same genomic region. Although the underlying mechanism is still unclear, this may imply that different TFs may interact with each other once bound to the same DNA region, provided that arrangement of their binding sites favors such interactions. For example, 2 regulatory regions with the same motif composition but in a different organization (orientation, position, or distance) may have different regulatory outputs. Such a principle, referred to as motif grammar, enables cooperativity between clustered homo- or heterotypic TF-binding events. High-throughput binding assays, such as high-throughput systematic evolution of ligands by exponential enrichment (SELEX) or high-throughput sequencing-fluorescent ligand interaction profiling, allow progress toward a deeper understanding of how nucleotide context of a TF-binding site influences affinity of that TF for its cognate site. With information from massively parallel reporter assays in vivo, it is feasible to gain insight into how structural variation quantitatively affects the regulatory function of TF-binding sites. This will be invaluable for understanding how genetic alterations of regulatory sequences disrupt cardiac gene expression in human disease.

Finally, a given TF typically binds thousands to hundreds of thousands of genomic regions in a given cell type. Thus, the developmental defects on loss of function of a cardiac TF likely result from pleiotropic effects on many target genes. This may also explain why abnormal expression of a single TF (loss or gain of expression) sometimes leads to almost coherent switches of transcriptional programs. A significant example is the atrial-expressed TF, COUP-TFII, which controls a gene network that leads to ventricularized atria when disrupted during embryonic development, and atrIALIZED ventricles when ectopically expressed.

Connecting the dots between aberrant TF expression, target gene dysregulation, and phenotypic defects is a daunting task. Furthermore, binding events of cardiac TFs are mostly found remote from promoters of coding genes, further complicating identification of target genes. This suggests that TFs rely on mechanisms to convey regulatory information from their actual binding site at a distance to target promoters. How to identify functional regulatory elements and the mechanisms underlying long-range transcriptional control is discussed below.

Enhancer Elements
In this section, we discuss how cis-regulatory regions for cardiovascular development were identified based on structural properties, such as sequence conservation and chromatin features, or by functional assays. We then give contemporary examples illustrating how genomic approaches can benefit our understanding and discuss ways to study the function of these regulatory regions for cardiac biology. Regulatory sequences are defined by their ability to control gene expression. Different classes of regulatory elements, including promoters, enhancers, silencers, and insulators, were historically defined by functional behavior of these sequences in reporter assays. However, a strict distinction between these elements can be difficult: promoters can function as insulators, and enhancer elements can initiate transcription. An effect on gene expression is a feature of a regulatory element, but also of the target. In particular, we consider an enhancer as a functional qualifier for a DNA sequence, rather than an intrinsic property. Its ability as an enhancer depends on the nature of the target promoter, as well as the effect considered (eg, transcript level or spatiotemporal pattern of expression). Because an enhancer effect is context dependent, we use element or TF-binding site for the physical entity and reserve the term enhancer for situations of an implied regulatory function.

Most TF-binding events occur outside of promoters. Promoter refers to the region directly upstream of the transcription initiation site of a gene and typically cannot mediate efficient transcription alone. Transcriptional activity of promoters is under the control of cis-acting sequences, known as enhancers or cis-regulatory modules. These can be upstream or downstream from the target promoter, within introns or exons, and can reside tens to hundreds of kilobases away from their target promoter. Notably, distal cis-regulatory elements outnumber mRNA promoters by at least one order of magnitude.

Enhancers are more tissue-specific than promoters. Although most enhancers are active in several related cell types, enhancer activation seems to be tightly linked with TFs that define cell-type identity. This property was historically a defining feature of developmental enhancers: they drive expression of reporters in a tissue-restricted or developmental stage-specific fashion.

Because many regulatory regions are composed of several TF-binding sites, regulatory regions can act differently depending on the specific combination of TFs expressed in that cell type. This enables cis-regulatory modules to integrate input of multiple TFs, leading to a specific regulatory output depending on the TF combination. Because most signal transduction pathways culminate in modulating TF function by post-translational modifications, this also means that cis-regulatory modules represent genomic integration hubs for external signaling cascades.

We now review strategies to identify cis-regulatory regions for the emergence and differentiation of the cardiac lineage. We then discuss mechanisms by which they orchestrate transcriptional dynamics. Finally, we tie these insights to what is known and what remains to be understood about transcriptional regulation during heart development and homeostasis. Our discussion will mainly involve enhancer elements, as recent years have witnessed considerable progress in our understanding of this type of regulatory elements.

Identifying Enhancers by Noncoding Sequence Conservation
With the functional importance of cis-regulatory elements, evolutionary conservation can be leveraged to identify important noncoding regions, some of which may act as enhancers. A common approach has been to survey the genomic neighborhood of a gene of interest for conserved sequences (Figure 2A). Then, short fragments are assessed to determine whether they drive expression of a reporter in a fashion that is
reminiscent of the gene’s expression pattern. This strategy is performed in cell lines or transgenic animals.

Once a minimal element is defined, potential DNA-binding sites and putative TFs that bind them can be identified. By creating enhancer variants, nucleotides that matter most for the enhancer activity can be identified, a strategy commonly referred to as enhancer bashing. Identifying transduction pathways connected to these TFs can ultimately reveal the external signals that are responsible for controlling expression of a gene of interest. For example, a vascular endothelial growth factor/mitogen activated protein kinase (Vegf/MAPK)-dependent transcriptional pathway specifies arterial identity during cardiovascular development by activating an enhancer of Dll4, a Notch signaling component.11 In a reciprocal way, reporter assays containing enhancer elements and a basal promoter can be used to identify candidate TFs that regulate specific enhancers. This approach revealed that Myocardin coactivates Srf12 and Hey proteins repress Gata factors.72

A strength of this approach is that it relies on genomic sequence. Thus, it can uncover regions that are active transiently or in a handful of cells. However, it is also limited. First, the neighborhood can harbor tens or hundreds of conserved genomic regions. Second, it assumes genes with conserved expression patterns are controlled by conserved enhancers, which may not be true. Surprisingly, there is a loose correlation between enhancer activity and evolutionary constraint in the heart.73 Evolutionary conservation of heart enhancers correlates with the developmental stage at which they are active, with the tightest conservation during early cardiovascular development.74

Surveying regions of evolutionary conservation near cardiac genes genome-wide identified several enhancers, including human cardiac enhancers.75,76 However, a large population of cardiac enhancers marked by p300 are poorly conserved,73 and only developmental enhancers are ultraconserved.77 Thus, evolutionary conservation may not discover many cardiac developmental enhancers.

**A Structure-based identification of cardiac enhancers**

**Evolutionary conservation**

<table>
<thead>
<tr>
<th>Enhancer</th>
<th>Nucleotide identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart enhancers</td>
<td>Low</td>
</tr>
<tr>
<td>Stomach, Thyroid, Pharynx enhancer</td>
<td>High</td>
</tr>
</tbody>
</table>

**Epigenomic signature**

- TFs
- DHS
- H3K27Ac
- H4K4me1
- p300
- RNApol-II
- H3K4me3

**B Function-based identification of cardiac enhancers**

**In vivo transgenesis**

- Reporter expression in the heart?

**High-throughput screening**

- Integration of enhancer reporters
- ES cell library
- Cell sorting
- Sequencing
- Cardiomyocyte library

**Regulatory sensors**

- Endogenous enhancers
- in vivo transposition

**Figure 2. Studying cardiac regulatory elements.** A, Left, Evolutionary constraints on some regulatory elements render them identifiable by comparative genomics, as exemplified by enhancers upstream of mouse Nkx2-5.20 Right, Combinations of specific chromatin features can reveal potential regulatory elements active in a given cell type. B, Left, Enhancer activity is classically tested by an ability of candidate elements to drive tissue- or stage-specific activity of a reporter. Middle, New technologies, such as SIF-seq,87 allow screening of large genomic neighborhoods for tissue-specific enhancers. Right, Genome Regulatory Organization Mapping with Integrated Transposons92 is a technology that can reveal integrated regulatory inputs exerted at a locus.
Identifying Enhancers by Chromatin Structure

Other methods to identify cardiac enhancers are based on structural features of chromatin. A major difference with the evolutionary approach is that, unlike DNA sequences, chromatin structure depends on the cell type. Binding of TFs to cis-regulatory elements and transcriptional activation is accompanied by changes in chromatin features, such as histone modifications, nucleosome remodeling, DNA demethylation, and higher-order reorganization. All can be surveyed to identify regulatory elements across the genome and infer their activity state in a cell type or tissue being profiled.

For example, p300 catalyzes histone-3 lysine-27 acetylation (H3K27ac) at active promoters and enhancers, revealing the genomic location of these elements during heart development or cardiac differentiation. Likewise, histone-3 lysine-27 trimethylation (H3K27me3) enrichment denotes elements repressed by the polycomb machinery, histone-3 lysine-4 trimethylation (H3K4me3) marks promoters, and histone-3 lysine-4 mono-methylation (H3K4me1) represents a general signature of enhancers (Figure 2A). Acquisition of promoter and enhancer activity has been classified during developmental transitions for cardiac differentiation in vitro and in vivo.

TF binding to DNA creates local nucleosome-depleted regions, and general accessibility of their chromatin is another hallmark of cis-regulatory elements. In vitro DNaseI treatment preferentially cleaves accessible genomic DNA, with DNase-hypersensitive sites (DHS) highlighting TF-binding events and delineating regulatory elements. As steric hindrance protects nucleotides contacted by TFs from DNase digestion, DNaseI profiling can identify TF footprints and infer the underlying TFs from the protected motifs. Other methods to assess open chromatin include formaldehyde-assisted isolation of regulatory and sequencing (FAIRE-seq) and assay for transspose-accessible chromatin using sequencing. However, it is unclear how structural features relate to enhancer function, and our incomplete understanding of enhancer activity prompts a parallel use of functional screening assays.

Identifying Enhancers by Regulatory Function

Classical approaches consist of cloning candidate enhancer sequences next to a heterologous promoter driving the expression of a reporter (Figure 2B). The construct can be introduced in cultured cells or transgenic animals, as a transient episome or stably integrated in the genome and fully chromatinized. In an early example, a rat Nppa enhancer was identified by transient transfections in cardiomyocytes. Many enhancer regions of cardiac genes have been evaluated using this approach.

A promising technology to assess enhancer activity in a high-throughput fashion is based on self-transcribing active regulatory region sequencing (STARR-seq). Bacterial artificial chromosomes or whole genomes are fragmented and cloned downstream of a minimal promoter whose activity is measured by RNA-seq after nonintegrative introduction into cells. Surprisingly, about a third of sequences that drive high levels of reporter activity as ectopic episomes reside in closed chromatin in their genomic context. This is attributed to repressive mechanisms that prevent enhancer activity from these sequences in the genome and illustrates that caution should be taken when inferring endogenous regulatory activity from ectopic nonchromatinized reporters. STARR-seq remains to be applied genome-wide in mammals.

Additive transgenesis, meaning ectopic integration of DNA fragments, has been used to find regulatory elements within a region of interest (Figure 2A). This labor-intensive approach remains the gold-standard demonstration of regulatory activity in vivo and in a genomic context. It also revealed the integrated nature of cis-regulatory elements. Studies at the mouse Nkx2.5 locus revealed that enhancer activity of transgenes can be blocked or unleashed by including or excluding a few hundred base pairs. The underlying mechanisms remain unknown, but likely involve cross-talk of different TF-binding sites within the same regulatory regions and coexistence of activating and repressing TF-binding events in the same region.

Recent developments allowed scaling-up the throughput of additive transgenesis. One approach relies on fragmentation of bacterial artificial chromosomes into small fragments that are tested for their ability to drive expression of a fluorescent reporter integrated in a precise genomic location (site-specific integration fluorescence-activated cell sorting followed by sequencing, SIF-seq) (Figure 2B). Use of embryonic stem cells for such an assay allows differentiating them into other cell types. Differentiation into cardiomyocytes revealed unprecedented regulatory dynamics across 200 kb at the MYH7/MYH6 locus.

Many have also cloned larger DNA fragments into bacterial artificial chromosomes and integrated these constructs into the genome to assess reporter expression. This technique was used to identify distal enhancers for Nppa and Nppb. However, after decades of searching for enhancers of cardiac genes, including Nkx2.5, it remains challenging to find combinations of elements that recapitulate a gene’s endogenous cardiac expression. Some genes are likely controlled by multiple and sometime interdependent regulatory elements, with regulatory communication depending on genomic organization of the locus.

A converse approach to additive transgenesis consists of interrogating regulatory activity directly from the endogenous genomic context, by integrating regulatory sensors in a region to be surveyed. Genome Regulatory Organization Mapping with Integrated Transposons (GROMIT) is a method that allows in vivo transposon-assisted lacZ reporter integration at sites throughout the genome (Figure 2B). An ongoing effort to create a regulatory atlas of the genome has already revealed tens of noncoding genomic loci with heart-enriched regulatory activity that remain to be investigated.

Exploring the Role of cis-Regulatory Elements

The above assays monitor the potential of a regulatory region. Once a sequence is identified to drive the desired expression pattern in reporters, a fundamental question remains. Is the sequence necessary to pattern the expression of its target gene in its genomic context?

Functional testing in vivo for necessity can be assessed by loss of function of an enhancer element. In the limb, the zone of polarizing activity regulatory sequence, which is mutated in the Sasquatch and hemimelic extra toes (Hx) mice along...
with some patients with polydactyly, is 1 Mb from its target gene, Shh.94 Deletion of the branchial-arch enhancer of Hand2 demonstrates a necessary role for craniofacial development,95 and deletion of craniofacial enhancers for Mx1, Snail2, or Isl1 displays subtle yet quantifiable morphometric anomalies.96

Our view of enhancers may be biased: many enhancers were identified because their alterations lead to observable phenotypes. Loci-harboring genes important for development are typically populated by many cell-type–specific enhancers. What would happen by deletion of only one of them? Or combinations? This remains to be explored.

The field still awaits identification of cardiac enhancers necessary for heart development. Maybe too few candidates have been tested. Other biological reasons may be involved, including the distributed and redundant nature of regulatory elements.97 Alternatively, certain enhancers may control subtle traits that require thorough phenotyping. Clearly, deeper exploration of the functional relevance of cis–regulatory elements for cardiac development and homeostasis is an exciting research avenue.

How Does TF Binding Confer Regulatory Activity? The answer is unclear. A TF might act through different mechanisms, depending on many features, such as the nature of the bound DNA sequence, the presence of cofactors, and the nature of the target promoter.57 Most TFs rely on several intermediate steps to control core transcriptional machinery.

One intermediate, the Mediator complex, contains several subunits that bind TFs and recruit general TFs or RNA polymerase (Figure 1B). Some Mediator subunits are constitutive, and others are tissue specific.96 Missense mutations in some subunits (eg, MED13, MED15) are associated with cardiovascular abnormalities in humans,99 and mutations in Med30 lead to cardiomyopathies in the mouse.100 The reason for the cardiac-specific phenotype of these mutations is unclear.

As mentioned above, enhancer regions rarely contain a single TF–binding site. They often involve clusters of binding sites, suggesting that binding of one TF modulates the binding of others and the functional output on transcription. Several models have been proposed.57 In the enhanceosome model,101 TFs interact with strict cooperation. Their interactions are allowed by motif positioning within an enhancer, and all TFs need to be bound to activate the enhancer. DNA and TFs form a higher-order protein complex that provides an integrated activity of all TFs. However, most developmental enhancers do not work so rigidly. The billboard model posits that TF binding does not rely on physical interaction with each other, meaning that positioning of TF–binding sites is flexible.102 Cooperation in binding and output may occur for enhancer activity even if all binding sites are not occupied. A composite model, called TF collective, combines flexible grammar with TF–TF physical interaction. Thus, the TF cohort occupies each individual enhancer in many combinations of TF–DNA contacts or protein–protein interactions.103 This model may account for the occupancy and activity of cardiac enhancers in Drosophila, with a cardiac TF collective comprising Mad/Smad, Tcf, Tin/Nkx2-5, Dorsocross/Tbx5, and Pannier/Gata.101

**Importance of Cooperativity Between TFs** Although common, cooperativity in DNA binding is not observed at all regulatory elements.104 This is important because it drastically affects the transcriptional response. High cooperativity means that an enhancer exists in 2 states: bound by few TFs and inactive or bound by many TFs and active. High DNA-binding cooperativity is associated with a digital on/off response to changes in TF availability. However, DNA binding without cooperativity means that enhancers exist with a wide range of activity, based on how many TFs are bound. Noncooperative interactions provide a molecular basis for an analog or graded transcriptional response105 (see Figure 5). This aspect is especially relevant when considering heart diseases originating from heterozygous loss of function of cardiac TFs, where TF availability might be compromised. As well, it will be interesting to determine if the occupancy of many cardiac TFs is interdependent. Understanding what genes are dysregulated because of disruptions of regulatory elements may elucidate molecular mechanisms behind these cardiac defects.

Heterogeneity in the initial responsive population must also be considered (see Figure 5). The heart comprises 4 distinct chambers, atrioventricular and semilunar valves, composed of various cell types and subtypes, including cardiomyocytes, cardiac fibroblasts, endocardial, epicardial, conduction, and mesenchymal cells, each with different transcriptional programs. Using methods for single-cell gene expression profiling, analysis at single-cell resolution may clarify contributions of different cell types to cell populations of the heart, as well as heterogeneity within each cell type and subtype. This approach has yielded insights for the reprogramming field, with relevance to cardiac development: Single-cell analysis of reprogrammed human fibroblasts to induced cardiomyocytes showed gene expression variability among induced cardiomyocytes.10 In another example, single-cell analysis of reprogramming fibroblasts to induced pluripotent stem cells revealed stochastic gene expression changes early, followed by an hierarchical period of gene expression for activation of the endogenous pluripotency program.105 Consequently, application of single-cell analysis to cardiovascular development may elucidate how cardiac cell identity is acquired, perhaps by on/off (digital) or rheostatic (analog) transcriptional mechanisms106,108 at different stages of cardiac differentiation and heart development.

**Enhancer–Promoter Interactions** How does TF binding occur thousands to hundreds of thousands of bases from core promoter elements to impact transcriptional activity? How does a regulatory region affect specific promoters among many? Studies focused on large-scale organization of chromosomes are addressing these questions. Chromosome Conformation Capture (3C) technologies109 demonstrated that distal enhancers engage physical contacts with promoter(s) that they control. This property can be used to identify regulatory elements for a gene of interest, or conversely, target promoters of an enhancer of interest. By Circular Chromosome Conformation Capture (4C)-seq analysis, which evaluates multiple genomic loci to a fixed DNA location, the promoter of the Scn5a gene, encoding a sodium channel involved in cardiac
Developmental Dynamics of Transcriptional Regulation

Two questions are not answered by the temporally static view of transcriptional regulation: why does a TF only bind a subset of its cognate binding sites in a given cell type and why do these differ between cell types or developmental stages? We discuss how the regulatory function of DNA sequences is affected by cellular history of the genome. We then discuss how regulatory networks formed from TFs and cis-regulatory elements facilitate patterning and memory of transcriptional activity.

Nuclear transfer and cell fusion experiments using 2 somatic cell types revealed that transcriptional changes are generally limited (unlike when fusing a pluripotent and a somatic cell). This means that the TFs’ milieu from a somatic cell type is largely incapable of using the genomic template of another somatic cell type, highlighting that cis-acting information in addition to the DNA sequence guides how TFs control transcriptional activity.

Epigenetic Priming

Exposure to a TF may result in transcriptional activation if the chromatin context is permissive to TF binding and activity. In this situation, developmental patterning of chromatin is necessary for TFs to affect transcription. For example, TF binding is influenced by DNA methylation, histone modifications, nucleosome positioning, and cofactor anchoring to chromatin. Reorganizing chromatin can modulate TF activity by controlling physical proximity between distal regulatory regions and can be developmentally modulated.

Epigenetic priming refers to a prerequisite for transcriptional activity but may not result in activation (Figure 3). Priming is epigenetic in that a permissive state is an acquired property transmitted through cellular generations and does not necessarily require persistent expression of the priming factor. Conversely, occlusis is a developmentally acquired repression, leading to inability of a promoter to react to the TF milieu. Therefore, one must consider properties of the epigenome (the structural features of chromatin and chromosomes) to understand how TFs pattern transcriptional activity during development.

Epigenetic priming during cardiac development became evident on inspection of chromatin states during cardiomyocyte differentiation from mouse embryonic stem (ES) cells. Several hundreds of promoters gained histone-3 lysine-4 mono-methylation, a signature suggestive of TF binding at an early stage of differentiation, without signs of transcriptional activation. On cardiac progenitor and cardiomyocyte differentiation from this early population, a subset was transcriptionally activated, illustrating how priming can potentiate but not commit to transcriptional activation. Although such a phenomenon is not specific to the cardiac lineage, TFs responsible for such patterning remain to be identified, and their roles for cardiac differentiation are unexplored.

Chromatin State Modulates TF Activity and Vice Versa

Chromatin structure and TF binding are engaged in a 2-way relationship, with one affecting the other, positively or negatively, depending on the context. Post-translational histone modifications and chromatin-binding proteins are involved in...
such cross-talk. For example, polycomb-mediated histone-3 lysine-27 trimethylation prevents transcriptional activation by TFs, whereas transcriptional activation by TFs prevents polycomb-mediated repression and is important for gene expression during cardiac development and homeostasis in the adult heart.

What are the molecular initiators of these transitions in chromatin structure? Some TFs penetrate the chromatin barrier, and their binding represents a pioneering event in the molecular cascade that ultimately leads to transcriptional activation at later developmental stages. Importantly, pioneering activity is not completely intrinsic to these TFs but depends on DNA sequence around the direct binding site of the TF. For example, the same TF may initiate nucleosome remodeling at some sites but not others. So, pioneer refers more to its activity than to the TF itself.

An alternative yet nonexclusive mechanism is the tissue-specific expression of nucleosome remodelers that bind TFs, allowing TFs to access their binding sites. For example, Baf60c, a cardiac-specific subunit of the SWI/SNF Brg1/Brm-associated factor, is necessary for differentiation of cardiomyocytes in mouse and zebrafish. Baf60c physically interacts with Gata4 and stimulates the association of Tbx5 and Nkx2-5 with Brg1, suggesting that Baf60c modulates access of TFs to their binding sites in chromatin. Interestingly, proper cardiac differentiation requires an exquisite stoichiometry of TFs, exemplified by the genetic interaction between heterozygosity of Tbx5, Tbx20, or Nkx2-5 and Brg1. Other cardiac TFs physically interact with chromatin-modifying enzymes, mostly through candidate approaches. For example, Gata4 and Hoxp associate with Hdac2 and Gata4 interacts with polycomb repressive complex-2 subunits. From yeast 2-hybrid systems, Mef2c interacts with class II histone deacetylases (HDAC). How chromatin remodelers modulate TF action is unknown.

Relevance for Experimental Manipulation of Cellular States
Control of TF accessibility is a central question in the field of reprogramming, where ectopic TF expression drives transcriptional changes, ultimately resulting in the acquisition of a new cellular phenotype. Self-perpetuation of chromatin states exerts an impediment to the efficiency of TF-mediated reprogramming. Interestingly, transdifferentiation of fibroblasts into functional cardiomyocytes is possible by ectopic expression of Gata4, Mef2c, and Tbx5 in the mouse. Although it requires additional factors in human. In mouse embryos, adding Baf60c to Tbx5 and Gata4 facilitates ectopic cardiac transdifferentiation of mesoderm. Although it is possible to transdifferentiate cells, the efficiency is low and suggests ectopic TFs encounter barriers that remain to be overcome. Concomitant manipulation of the epigenome is a promising avenue under investigation.

TF-to-TF: Building a Regulatory Network
If a TF binds to a regulatory region of a gene encoding another TF and controls its expression, this relationship establishes a regulatory network (Figure 4). Aside from information stored in chromatin structure, the topology of a TF regulatory network itself is another way to instruct the execution of a transcriptional program and implement cellular memory. Deciphering the architecture of such regulatory networks allows pivotal factors to be identified. Some TFs may push developmental programs forward, and others may stabilize cell identity.

A positive feedback loop, when a TF reinforces its own expression (directly or indirectly), is the simplest regulatory motif and enables self-perpetuation and stability of cellular states (Figure 5A). However, positive feedback loops rarely exist in isolation because continuous autostimulation would result in an explosive increase in gene activity. Autostimulatory loops are often embedded in larger circuits that include negative regulation. This balance between positive and negative feedback determines the output of transcriptional activity (Figure 4B). With more intermediates in a pathway, there is more opportunity to influence and divert a cell from its stable state. For example, exposure to external signals can lead to directed differentiation of embryonic stem cells to cardiomyocytes. The reconstruction of regulatory networks through genome-wide assays opens an opportunity to investigate this topic in the context of cardiac cell differentiation.

Combinations of simple network motifs may be enough to explain many cellular behaviors (Figure 4C). When a TF controls a target directly and indirectly, it forms a feed-forward loop. Counterintuitively, feed-forward loops are often incoherent, meaning that the TF activates its target on the one hand and represses it on the other. Network motifs involving negative regulation can implement temporal dynamics, such as oscillations or pulses in gene expression. This situation is particularly relevant for developmental contexts where a pool of progenitors needs to arise and expand before subsequent differentiation. For example, Nkx2-5 represses genes important for cardiac progenitor proliferation while repressing its own expression through a negative feedback loop involving Bmp2 and Smad signaling. The time lag for the negative feedback to operate allows initial cardiac induction and ensures that a break in proliferation only operates after progenitors have arisen.

Developing quantitative approaches to study regulatory networks underlying cardiac development is a formidable challenge, experimentally and computationally. Moving forward will require gathering developmental, molecular, and computational biologists. Such integrated approaches have proved valuable in the fly, where information about TF occupancy and transgenic enhancer assays trained a machine-learning algorithm to predict spatiotemporal activity of cis-regulatory elements. The recent genome and epigenome engineering revolution offers an unprecedented way to functionally study these networks, in cellular or animal models. Although a global understanding of the transcriptional control of cardiac development is not yet within our grasp, pieces of a cardiovascular regulatory network are starting to be unraveled and are already yielding tremendous insights in the understanding of human cardiac defects.

Conclusions and Prospects
In summary, we present general principles to better understand transcriptional control of gene expression during cardiac development...
Many cardiac TFs bind to selective DNA sites in regulatory elements, with cofactors and chromatin, to activate cardiac gene expression. Mechanistic diversity for transcriptional regulation is such that each gene might constitute an exception to the general rule. Enhancers may go beyond a sequence-based code, likely involving contributions from long-range activation by enhancers acting through a binary mechanism, therefore contributing of cell-to-cell heterogeneity.

Figure 5. Population vs single-cell approaches for gene expression. Insights from cell population–based studies reveal global changes in fields of cells or tissue during cardiac differentiation and development. However, cell heterogeneity can obscure gene expression analysis from cell populations. Two distinct modes of transcriptional responses, rheostatic (or analog) and binary (or digital), give similar mean behaviors on a population scale. Yet, in a rheostatic regime, cells respond in a concerted fashion homogeneously, while a binary regime leads to the appearance of populations with expressers or nonexpressers. During development, this can have fundamental consequences on how cells differentiate. Of importance, long-range activation by enhancers acts through a binary mechanism, therefore contributing of cell-to-cell heterogeneity.
genomic organization. Cooperativity of TF binding through protein interactions at enhancers defines modes of target promoter activation (Figure 5). Higher-order chromosomal organization fosters functional cross-talk between control elements in regulatory neighborhoods within TADs. How TFs behave depends on chromatin changes at earlier developmental stages, and at the same time, influences future changes in gene expression. Understanding the nature of regulatory connections between TF proteins and TF genes is key to deciphering molecular mechanisms driving temporal dynamics during developmental patterning. This holds great promise for our comprehension of developmental defects, such as CHDs, where the disruption of one of these nodes may result in pathogenic consequences on many participants of the regulatory network.

For human disease, TF heterozygosity can lead to haploinsufficiency and is a frequent cause of inherited CHD. However, how reducing cardiac TF dosage leads to cardiac defects is not known. Investigation has been hampered by lack of relevant and tractable models in humans and animals. With a few notable exceptions, many mouse models of heterozygous deletions do not recapitulate phenotypic aspects of human disease. *Tbx5* heterozygosity causes ventricular septal defects (VSDs), atrial septal defects, and arrhythmias, along with upper limb anomalies reminiscent of Holt–Oram syndrome, consistent with TF haploinsufficiency. Tissue-specific heterozygous deletion of *Tbx5* in the dorsal menenchymal protrusion results in atrial septal defects, implicating *Tbx5* in the second heart field. Interestingly, discovery that *Tbx5* haploinsufficiency in mice causes diastolic dysfunction prompted an evaluation of a cohort of patients with Holt–Oram syndrome, which detected diastolic dysfunction that has implications for clinical management.

Genetic modifiers likely contribute as genetic buffers for human health or as perturbations in polygenic human disorders, including many human CHDs. Phenotypes of a few mouse models are strain dependent, including for *Tbx5* or *Nkx2-5*. Attempts to identify modifier genes of *Nkx2-5* for VSDs or atrial septal defects have revealed strain-dependent susceptibility loci that were shared or distinct based on the anomaly. However, specific polymorphisms for genetic modifiers have not yet been identified. The advent of induced pluripotent stem cell technology from human fibroblasts now allows study of a mixed genetic background derived from patients’ cells. In conjunction with directed cardiac differentiation of human and mouse* Nkx2-5* pluripotent stem cells, mechanisms for cardiac differentiation can be studied with a scalable method using cellular models.

An integrated understanding of TF networks during cardiac development can provide invaluable information to understand mechanisms of cardiac disease and vice versa. In the near future, it will be interesting and important to determine whether single nucleotide polymorphisms identified by genome-wide association studies for CHDs, coronary artery disease, cardiac hypertrophy, and arrhythmias might alter nodes for potential cardiac regulatory networks of different functions. Conversely, genome-wide association studies may reveal nodes of cardiac regulation that are clinically relevant to disease. These mechanistic insights may shed light on novel therapeutic targets and may lead to new treatment strategies including regenerative approaches for cardiovascular medicine.

**Acknowledgments**

We apologize to our colleagues whose work was not presented or cited because of space constraints. We thank members of the Bruneau laboratory for constructive comments, John Wylie for illustrations, Juan Perez-Bermejo for a diagram, and Gary Howard for editorial assistance.

**Sources of Funding**

I.S. Kathiriya is supported by the Foundation for Anesthesia Education and Research, and the John W. Severinghaus Endowment, E.P. Nora is supported by the European Molecular Biology Organization (ALTF523–2013) and Human Frontier Science Program, and B.G. Bruneau is supported by grants from the National Institutes of Health/National Heart, Lung, and Blood Institute (Bench to Bassinet Program U01HL098179), the Lawrence J. and Florence A. DeGeorge Charitable Trust/American Heart Association Established Investigator Award, and William H Younger, Jr.

**Disclosures**

None.

**References**


Lahn BT, Noordermeer D, Duboule D. Chromatin looping and or


Investigating the Transcriptional Control of Cardiovascular Development
Irfan S. Kathiriya, Elphège P. Nora and Benoit G. Bruneau

Circ Res. 2015;116:700-714
doi: 10.1161/CIRCRESAHA.116.302832
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/116/4/700

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org/subscriptions/