**Abstract**

Even though the foundation of systems biology approaches to cardiac function was laid more than fifty years ago, there has been slow progression over the last few decades. Systems biology studies were mainly focused on lower organisms, frequently on yeast. With the boost of high-throughput technologies, systems level analyses, building one backbone of systems biology, started to complement the single-gene focus in the fields of heart development and congenital heart disease. A challenge is to bring together the many uncovered molecular components driving heart development and eventually to establish computational models describing this complex developmental process. Congenital heart diseases represent overlapping phenotypes, reflecting the modularity of heart development. The aetiology of the majority of congenital heart disease is still unknown, and it is suggestive that understanding the biological network underlying heart development will enhance our understanding for its alteration. This review provides an overview of the framework for systems biology approaches focusing on the developing heart and its pathology. Recent methodological developments building the basis for future studies are highlighted and the knowledge gained is specified.

**Keywords**

Systems biology • Heart development • Congenital heart disease • High throughput • Computational modelling

---

**1. Introduction**

In 1952 Hodgkin and Huxley\(^1\) first identified the system of interactions between cell electrical potential and ion channel activity that enabled them to formulate equations and reconstruct the nerve impulse and its conduction (Nobel Prize in Physiology or Medicine 1963). Today one could name this a systems biology approach, which shows that the foundations for interpreting the genome and proteome data in respect to higher level biological function were laid long ago. Ten years later, Noble developed a mathematical model by modifying the previous Hodgkin and Huxley model to describe the long-lasting action and pacemaker potential of the Purkinje fibres of a heart.\(^2\) With the success of high-throughput technologies for genome sequencing, expression profiling, proteomics, structural analysis and \emph{in vivo} gene targeting, integrative biological approaches such as systems biology or multi-scale bioengineering developed at a broader scale.\(^3,4\) In the last decade, integrative biology started to successfully complement the reductionist single-gene focused biology.\(^5\) Single-gene or single-factor focused approaches have uncovered many of the components of biological systems as well as their properties and interactions. However, to understand dynamic properties of a system, the individual parts need to be studied in context with each other. System biology integrates and combines data within or across molecular levels of biological systems for analysing input–output relationships.\(^6–9\) The framework of systems biology consists of (i) experimental data derived from different levels of a system representing its components and interactions, (ii) integration of these heterogeneous data, and (iii) iterative adjustment of experiments and consecutive modelling of the system. One of the major topics of systems biology is the construction of dynamic networks in order to understand how network properties perform biological function.\(^10–15\) A network is technically speaking a graph consisting of nodes connected by edges. In biological networks, nodes typically represent molecules like gene transcripts or proteins. Edges define the mode of connection between nodes, e.g., interactions or relationships, such as protein–protein interactions or transcription factor-DNA binding. Examples of biological networks are gene networks, transcription networks, protein–protein interaction networks, and metabolomic networks.

\(^*\) Corresponding author. Tel: +49 30 8413 1232; fax: +49 30 8413 1699, Email: silke.sperling@charite.de

Published on behalf of the European Society of Cardiology. All rights reserved. © The Author 2011. For permissions please email: journals.permissions@oup.com.
This review attempts to give an overview of the developing field of cardiac systems biology with a focus on heart development. Although >3000 publications of systems biology studies exist in PubMed currently, only a limited number of these refer to cardiogenesis, which is conceptional and experimental challenging due to the complexity of this developmental process involving multiple cell lineages and diverse layers of regulation.

The mammalian heart is composed of diverse cell types and structures such as cardiac and smooth muscle cells contributing to the contractile apparatus and vascular system, the conduction system as well as endothelial, valvular, and interstitial mesenchymal fibroblast cells. During evolution the heart developed from a two-chambered organ in fish to a four-chambered one in mammals, where its anatomy enables the separation of low- and high-oxygen blood flow. Along with this macro-structural development a full panel of changes in its physiology occurred. It is highly suggestive that the interaction between environmental factors (e.g. physiological conditions such as oxygen content or pressure load) and the molecular makeup (e.g. transcript levels, protein modifications, protein–protein interactions) of the cardiomyocytes has played a crucial role in this evolutionary developmental process, which is in part mimicked during the embryonic development of the vertebrate heart.

The study of heart development is naturally confronted with very limited material for experimental research. At birth 1 mm³ of heart muscle in rat consists of ~3.3 × 10⁵ cardiomyocytes and a mouse heart at E11.5 consists of ~1 × 10⁶ cells, including all different cell types and cardiac structures. Therefore, the development of methods suitable to gather sufficient tissue- and cell-selective material (e.g. ES or iPS cells, GFP marked cells) and the establishment of technologies enabling the study of limited biological material up to single cells is essential to systems biology approaches of heart development. A crucial issue remains in the application of high-throughput phenotyping technologies starting from static phenotyping at a cellular level (e.g. histology) up to the study of dynamic functional parameters (e.g. pressure load, shear stress, oxygen supply, and heart beat) under in vivo conditions.

2. The framework of systems biology approaches to heart development

Systems biology is an integrated approach and, therefore, requires a distinct set of tools, namely an applicable model system enabling perturbations, suitable phenotype assessment, biotechniques measuring molecular states, and interactions; and computational methods and resources providing the analysis platform for probabilistic and mathematical modelling (Figure 1).

2.1 Applicable biological systems

2.1.1 Animal models

Zebrafish (Danio rerio) is a valuable model for the study of heart formation as it permits in vivo high-resolution inspection of heart size, shape, and function based on the transparency of the zebrafish embryo. In addition, zebrafish represents a good model for perturbation screens due to its small size, fecundity and brief generation time. During embryogenesis the zebrafish embryo does not require a functional cardiovascular system, which enables the study of otherwise lethal cardiac dysfunctions. Large zebrafish screens found >100 mutations affecting heart development and function.

The frog (Xenopus leavis) has been at the forefront of research in developmental biology for the second half of the last century, which was based on its ease of use for the generation of embryos and surgical manipulation. However, its tetraploid genome and long maturation time for breeding have been unfavourable for its use as a genetic model. More recently, chemical genetic screens were successfully performed in Xenopus and open the window to high-throughput screenings. Thus compared with zebrafish it represents a model for investigations of a higher vertebrate system.

The fruit fly (Drosophila melanogaster) has been a frequently used model to study the regulation of early heart development. As an insect, drosophila is characterized by a simple tube-shaped heart located at the dorsal midline that pumps the haemolymph along its body. The heart, also called dorsal vessel, consists of a single layer of cardiomyocytes, which are surrounded by non-contractile pericardial cells. The main advantages of drosophila are the easy use for large genetic screens and for crossing mutants of interest with collections of other mutants out of a large available collection. The cardiac anatomy and less-frequent genetic redundancies made drosophila a widely used model organism for the identification of essential players and regulatory pathways of the heart developmental process. However, this might disqualify drosophila for studies of buffering effects, which seem to be essential to understand cardiac alterations in human.

The mouse (Mus musculus) has been extensively used as a model organism for cardiac developmental malformations in humans. It shares a high degree of homology with humans; but there are frequently no direct correlates to human genetic variants such as single nucleotide or copy number variants. A recent study of heterozygous Nkx2-5 knockout mice in systematically varied genetic backgrounds showed the impact of (so far unknown) modifier genes on the penetrance of cardiac malformations. Such model settings are eligible to systems biology approaches. The generation of a probabilistic model based on genome-wide sequence variations might eventually enable to understand the interaction between genetic modifiers and Nkx2-5 and the phenotypic outcome. Through the
International Knockout Mouse Consortium a major resource of knockout mouse embryonic stem cells (ES cells) is being generated.\textsuperscript{46} To this date, mutated ES cell lines are available for over 15,000 protein-coding genes.\textsuperscript{47}

The laboratory rat (\textit{Rattus norvegicus}) is the most widely used model organism to decipher how the human organ system functions and is frequently employed for pharmaceutical testing.\textsuperscript{48,49} Because of its size, ease of manipulation and high similarity to the human cardiovascular physiology, it remained the preferred choice for respective studies, while the mouse became the leading mammal for experimental genetics in the last decades. Recently, three groups overcame the genetic manipulation limits for rat by developing gene-targeting approaches for gene knockout as well as gene replacement.\textsuperscript{50–52}

2.1.2 Cell culture and single cell analysis

The heart is built of a mixture of distinct cells from different lineages, which moreover crosstalk to each other.\textsuperscript{53} Cardiac fibroblasts represent \textasciitilde 50% of the adult mammalian heart and several types of cardiomyocytes can be distinguished (e.g. atrial vs. ventricular cardiomyocytes). As the heart develops during embryogenesis, cardiac cells migrate a considerable distance and an ever-changing environment influences their fate. An example is neural crest-derived cells contributing to the developing heart.\textsuperscript{54,55} One way to study largely homogenous cell populations according to spatio-temporal changes is the use of cells expressing particular developmental markers of representative developmental points. For example, cells could be sorted by fluorescence-activated cell sorting (FACS) for GFP-positive cells like in Gata4p[5kb]-GFP mouse embryos or using cell layer-specific expression of GFP in transgenic zebrafish.\textsuperscript{20,56} Furthermore, cardiomyocytes derived from ES cells or induced pluripotent stem cells (iPSCs) are novel potential sources for systems biology approaches studying differentiation stages of cardiomyocytes.\textsuperscript{19,21,57–59}

In addition, the mouse atrial-derived cell line HL-1, the rat cell line H9C2 derived from embryonic rat heart tissue and the mouse P19 ES cells are frequently used.\textsuperscript{60–62} The contracting HL-1 cardiomyocytes maintain a gene expression pattern similar to adult cardiomyocytes, whereas H9C2 cells can have both cardiac muscle as well as skeletal muscle properties. The pluripotent P19 embryonal carcinoma cells can differentiate into mesoderm lineages, including cardiomyocytes upon the presence of 1% DMSO. Additionally, a P19Cl6 GFP reporter line, where GFP is expressed under the control of the MLC-2v promoter, has been established enabling the quantification of cardiomyocytes.\textsuperscript{63}

2.1.3 Patients with congenital heart disease

It is clear that genetic manipulations cannot be carried out in man; however, patients with congenital heart disease (CHD) are a unique resource to gain insights into cardiac functional properties and molecular pathways.\textsuperscript{64} The genetic alterations underlying the phenotype of these patients lead to disturbances in the molecular read-out, which can be studied, for example, by gene expression analysis and...
potentially provide insights into regulatory relationships. Collections of biomaterial are available through national registries like the CONCOR registry in the Netherlands or the National Registry for Congenital Heart Defects in Germany.\textsuperscript{65,66}

Besides the immediate use of patient-material, the generation of patient-specific iPSCs offers a new and very promising way to study human diseases.\textsuperscript{67,68} Different groups have already shown the power of iPSCs for the analysis of congenital arrhythmia and malformation.\textsuperscript{69–71} An alternative but more challenging approach is the generation of genetically modified human ES cells using so-called zinc-finger technology to introduce chromosome deletions and translocations.\textsuperscript{72,73} The advantage of recombiant human ES cell is the availability of the parental genetically matched ES cell line as a built-in control.

2.2 Cardiac phenotyping approaches
In order to understand biological systems, it is necessary to understand the relationship between the genome and the phenotype. The aim would be to monitor the molecular background and the cellular and physiological phenotype at a comparable level of profundity and in a machine-readable manner. In respect to the information gathered, phenotyping approaches can monitor static profiles (e.g. confocal microscopy), or dynamic processes [e.g. echocardiography (ECHO)]. A set of phenotyping approaches relevant for heart development is described in the following.

2.2.1 Systemic phenotyping of mouse models
Thousands of mouse disease models are now available by the collaborative effort of the International Mouse Knockout Consortium.\textsuperscript{46}

The main characterization, archiving, and dissemination of mouse disease models are realized by phenotyping centres (mouse clinics) established in Germany (GSF), France (ICS), and the UK (MRC Harwell and Sanger Institute). These provide large-scale phenotyping platforms and collaborate for the standardization of phenotyping protocols in the EUMODIC consortium (http://www.eumodic.org/). Their current cardiovascular phenotyping consists of blood pressure analysis, electrocardiography, ECHO, and the quantification of biomarkers such as ANP in serum.\textsuperscript{25} These phenotyping approaches are performed on viable adult mice, whereas characterization of structural cardiac or other developmental malformations potentially occurring in embryonic lethal recessive mutations is currently not realized. However, these mouse models would be of particular interest to gain insights into CHD.

2.2.2 Imaging cardiogenesis in live animals
In vivo imaging of the processes that underlie cardiogenesis—cell division, migration and differentiation, or haemodynamic parameters—is central to understand the structural formation and function of the heart and its molecular background. Since the heart is beating from early stages onward, live time-lapse imaging is particularly challenging. A current lack of technology remains for monitoring of haemodynamic parameters like pressure load or mechanical parameter like shear-stress during cardiogenesis.

ECHO is one imaging modality extensively used for cardiac assessments in clinical settings and has become invaluable for the study of mouse models. Prenatal ultrasound imaging allows longitudinal studies and monitoring of the same animal in vivo.\textsuperscript{74} However, the small size of the foetal mouse heart (<5 mm at birth) makes it technically challenging. Mouse foetal ECHO can be carried out using standard clinical ultrasound systems as well as ultra-high frequency ultrasound systems; also referred to as ultrasound biomicroscopes.\textsuperscript{75,76} The latter have a two-dimensional spatial resolution that allows monitoring of mouse embryos from embryonic Day 5.5 on.

A further method to identify gross cardiac malformations in the mouse embryo is magnetic resonance imaging (MRI). Recent technical advances based on 3D MRI and computational enable the identification of subtle phenotypic differences at E15.5.\textsuperscript{29,77} However, for routine high-throughput application the scan times are very long with 9–36 h and the spatial resolution is limited. High-resolution magnetic resonance histology and high-resolution epicam imaging (HREM) start to overcome this limitation.\textsuperscript{78,79} Collections of high-resolution HREM images of mouse and human embryos as well as MRI images of mouse are available online (http://embryoimaging.org, http://mouseatlas.caltech.edu).

A routine method to study the transparent zebrafish embryo is live-stream imaging using confocal microscopy. Multiple fluorescent transgensics and mutants have been established like the Tg(cmlc2:eGFP) and the Tg(β1:eGFP) zebrafish expressing green fluorescent protein in the cardiac myocytes and endothelial cells, respectively.\textsuperscript{80–82} Advanced imaging techniques have been developed which enable the tracking of cardiomyocyte trajectories during cardiac contraction and 3D reconstruction.\textsuperscript{26,27} Real-time imaging has further been applied to monitor the heart tube of the drosophila embryo providing insights into the mechanics of cardiac function.\textsuperscript{28}

2.2.3 Imaging fixed hearts and monitoring single cells
Classical methods also suitable for high-throughput studies are histological assessment using haematoxylin and eosin staining for morphological inspection, in-situ hybridization to visualize transcript expression or immunohistochemistry for protein expression. The gold standard for the visualization of subcellular structures is electron microscopy. A recent interesting technological development is the approach to monitor the interaction between a cell and its physical microenvironnement using mechanical tractions exerted by cells in three-dimensional matrices.\textsuperscript{83} Further investigations of this method might eventually be used to gain insights into the role of cellular forces for the developing heart.

2.2.4 Patients with congenital heart disease
In clinical settings, patients with cardiac malformations are studied using ECHO, MRI, or cardiac catheterization. In 2005 the International Pediatric and Congenital Cardiac Code (IPCCC, www.ipccc.net) was launched, which enables a detailed description of anatomic and haemodynamic features in CHD in human and provides a unique source for a standardized phenotype description.\textsuperscript{84}

2.2.5 Computational coding of cardiac phenotypes
Computational coding of phenotypes and machine-readable representation of biological models is essential for any systems biology approach. In contrast to the computational coding system of CHD in human, a defined ontology describing the precise cardiac anatomy, morphology, and haemodynamic features of model organisms is currently missing. The establishment of such phenotype ontology remains, therefore, a central goal for respective systems biology studies.
2.3 Biotechniques for systems biology

Techniques used in systems biology approaches can in general be grouped into techniques measuring molecular states and techniques measuring molecular interactions. A number of technical developments have been used as the basis for systems level studies building a ground floor for systems biology. One major impact emanated from the microarray technology to analyse gene expression profiles (DNA microarray) and as a general application to measure nucleotide sequence or protein abundances derived by a panel of different primary selection and enrichment processes.

2.3.1 Genome, epigenome, and transcriptome analysis

Genomic sequences can be enriched based on antibody-specific selection using chromatin-immunoprecipitation (ChIP) to measure transcription factor binding events (ChIP-chip) or occurrence of histone modifications, or the presence of DNA methylation (MeDIP-chip). A further selection method for DNA methylation is based on the hypersensitivity to DNaseI cleavage (DNase-chip). DamID chromatin profiling, here the bacterial Dam-methylase is fused to a protein of interest and subsequently marks its genomic binding site. Other applications are the analysis of sequence variations by genotyping microarrays, copy number variations (Array-CGH) or RNA profiling (mRNA, microRNA). So far, these methods have been the main basis for systems biology studies of heart development and its alterations. Nowadays, microarrays are replaced more and more by next generation sequencing (NGS) technologies. The currently most frequently used NGS platforms are 454 (Roche Diagnostics), Solexa (Illumina), and SOLiD (Applied Biosystems). Each of these technologies has their advantages and disadvantages. Four hundred and fifty-four sequencing is based on pyrosequencing and enables sequencing of roughly 400–600 Mb of DNA per 10-h run with an average read length of 400 bases on the Genome Sequencer FLX. The long read lengths are the main advantage of this technique, whereas it is also the most expensive platform. The Solexa Genome Analyzer IIx platform based on sequencing-by-synthesis process requires ~2 days to generate read lengths of 35–150 bases and an overall sequence output of 10–12 Gb per run. The SOLiD (Supported Oligonucleotide Ligation and Detection) platform is also a short-read (35–75 bases) sequencing technology based on ligation. The 5500 SOLiD system generates up to 10–15 Gb per day. The first single-molecule sequencing platform (HeliScope) is available from Helicos BioSciences (http://www.helicosbio.com) with an output of 1 Gb/day and read lengths of up to 45–50 bases.

2.3.2 Proteome and metabolome analysis

Mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) are the core technologies for proteome and metabolome studies. The majority of approaches is based on the analysis of peptides, which are frequently generated by enzymatic digestion of proteins. The selection and enrichment of the fraction of proteins/peptides of interest is a key issue for MS analysis. Selection strategies can be applied for subcellular fractioning (e.g. membrane enrichment, nucleus precipitation, or mitochondria separation), for a particular protein and its interaction partners (e.g. immunoprecipitation) or proteins harbouring particular modifications (e.g. phosphorylation). A particular interesting approach is SILAC, which relies on the incorporation of isotopically labelled amino acids into proteins during their synthesis. This technique can be applied to cell culture studies and more recently also to study mouse and drosophila models, which opens a novel way to study proteome differences in respective model systems.

NMR is a common method in metabolomics and in contrast to MS does not require analyte separation. It can provide detailed information on the molecular structure of compounds found in complex mixtures, such as biofluids and cell and tissue extracts, which can be e.g. analysed through the use of 1H NMR spectroscopy. Long-standing methods for the analysis of protein–protein interactions are the yeast-two-hybrid (Y2H) and the mammalian-two-hybrid (M2H) systems. In these systems, a read-out is achieved via the protein–protein interaction of the two proteins to be tested, which are expressed as hybrid fusion proteins. Automated and high-throughput Y2H experiments identified over 3000 interactions among 1705 human proteins resulting in a large and highly connected network and a M2H study provided a map of physical interactions among 762 human and 877 mouse DNA-binding transcription factors. Besides the two-hybrid systems, peptide microarrays have been employed to study protein–protein interactions; however, their implementation has been much slower compared with DNA arrays because of technical challenges such as high-throughput synthesis of peptides at a low cost.

2.3.3 Techniques to induce perturbations

In vivo gene targeting techniques are essential to enable the study of cause–effect relationships. Gene targeting in ES cells is widely used to generate designed mouse mutants. Knockout mice harbouring a null allele in their germline represent essential genetic models for the study of inherited disease, whereas conditional gene inactivation is extensively used for the study of gene function in adult mice as well as in specific cell types. The latter relies on the DNA recombinase Cre and its recognition (loxP) sites. In the last decade, a large collection of Cre transgenic lines has been generated and can be used in a combinatorial manner to attain gene inactivation in many different cell types or starting at a defined embryonic stage. This approach is particularly useful for models targeting cardiac essential genes, which otherwise might lead to very early embryonic lethal phenotypes. The application of CreER(T2) transgenic mice enables the inactivation of floxed alleles in adult mice upon administration of tamoxifen. A widely used approach for gene knockdown is RNA interference (RNAi) and large genome-scale RNAi high-throughput screens have been carried out in both drosophila and mammalian cultured cells. Zebrfish, Morpholino oligonucleotides (MOs) are the most common anti-sense ‘knockdown’ technique. MOs bind RNA, thereby facilitating steric hindrance of proper transcript processing or translation.

Small molecules and chemical screens have played an important role in delineating molecular pathways involved in embryonic development and disease pathology. In contrast to directed in vivo gene targeting approaches, chemical screens enable the discovery of gene functions in an unbiased manner. In mice, chemical mutagenesis induced by N-ethyl-N-nitrosourea (ENU) provides a unique resource for mutants and displays a range of mutant effects from complete or partial loss of function to exaggerated function. A furthermore well-suited model for use in chemical screens is zebrafish, where small molecules are typically added to the aquatic environment in which
they live, allowing absorption into the fish without the need for invasive and time-consuming injection.\textsuperscript{38,109,110}

Besides gene targeting approaches, environmental alterations governing perturbations of molecular networks allow the analysis of gene–environment interactions, which might play a particular role in the development of CHD with incomplete penetrance.\textsuperscript{32}

2.4 Computational analysis

Computational approaches have mainly universal applicability, which is in contrast to the specifications of data gathering (models system, molecular level, organ focus, etc.). Recent developments of computational analysis strategies, modelling approaches, and software packages are reviewed elsewhere.\textsuperscript{4,111,112}

A particular aspect for analysing a developmental process such as cardiogenesis is the fact that information flow may use different pathways at different developmental time points. Thus, tools that can model dynamic regulatory networks phasing over time are important for the study of heart development.\textsuperscript{113} For example, binding of a particular transcription factor to DNA may only modulate gene expression in concert with an epigenetic modification, which occurs only at a specific developmental window, or a defined outcome of gene expression may depend on the impact of a panel of different factors acting in concert on this regulatory region.\textsuperscript{114} Therefore, nodes in a transcription network do not necessarily represent essential points at a particular developmental stage, but might rather be elements enabling fine-tuning of transcriptional and thus developmental responses to environmental signals. Computational modelling might provide the opportunity to define the genetic and epigenetic impact in the context of variable internal and external, environmental, and stochastic influences.

3. Knowledge gained by recent studies

In the last decade, deep insight was gained on how a number of cardiac-restricted transcription factors act in concert to govern cardiogenesis.\textsuperscript{115–117} However, at latest with the discovery of the impact of chromatin remodelling factors (e.g. Baf60 or Baf45c/DPF3) or histone-modifying enzyme (e.g. Jmjd6) it is obvious that the molecular network underlying heart development and CHD is far more complex.\textsuperscript{118–120} A challenge is bringing together the many uncovered molecular components driving heart development and eventually to establish computational models with predictive power. This is even complicated by the unknown level of the environmental and stochastic impact as well as the nature and capacity of buffering mechanism of the system (Figure 2). Besides the map of the physical components and interactions driving cardiogenesis, insights have to be gained on how the flow of information propagates and responds to perturbations. Given an example, maternal high-fat diet of Cited2 haploinsufficient mice increases the risk (penetrance) of CHD in their offspring.\textsuperscript{32} if this holds through in human is an open question.\textsuperscript{121} The reduced expression of Pitx2c in these mice might be a key driver for the observation, but we currently lack understanding of the mechanism behind. It is unknown how the observed maternal hyperglycaemia is transduced to the genome leading to expression changes of respective genes and, moreover, which cascades are initiated by it resulting in the phenotypic endpoint. Therefore, critical mass of data need to be gathered in appropriate model systems, and computational approaches must utilize the biological information at the different levels by integrating the heterogeneous data and finally enable prediction of molecular and phenotypic outcome induced by perturbations. Insights into cardiogenesis are primarily generated by the study of embryonic models and with constraints using cell–culture systems.

Figure 2 Simplified overview of the aetiology of congenital heart disease (CHD) reflecting multifactorial impacts on heart development. Biological networks (orange) mediate interactions between particular DNA sequences (blue; e.g. genes, miRNAs), epigenetic modifications (grey; DNA methylation, histone modification), and distinct phenotypes (red). The aetiology of CHD ranges from simple relationships between single genes and a particular type of CHD up to an underlying complex network influenced by multiple genes, epigenetic, environmental influences, (green) and potentially stochastic events (purple). Environmental influences can be ‘external’ as well as ‘internal’ to the system. Congenital heart diseases are depicted as in part overlapping phenotypes, which is one characteristic feature of cardiac alterations and reflects the modularity of heart development.
A cardiovascular development signalling network constructed based on gene expression profiles of ES cells undergoing guided differentiation to cardiomyocytes showed that selected nodes orchestrate coordinated recruitment of specialized ontological classes to secure a developmental theme. Integrins and WNT/β-catenin were the most significant cascades identified. Their relevance was further shown by a proteomic-based study of mouse embryos developing congenital heart defects after exposure to high glucose and correspondingly increased protein levels could be observed in the amniotic fluid of human foetuses with CHD. A further study based on protein interaction networks showed that distinct clusters of interacting proteins (so-called ‘functional modules’) underlie heart development characterized by the development of defined anatomical substructures. The study suggests that higher order functional networks are achieved by the integration of discrete protein complexes in the way that evolutionary younger anatomical structures are potentially parts of networks already used in more ancient structures. Here, 255 proteins grouped into 19 morphological subgroups where used as an indicator for spatio-temporal function of individual genes. Thus, the work points to the concept of combinatorial regulation at the protein level, which was confirmed recently in a study investigating the transcription factors Gata4, Mef2a, Nkx2.5, and Srf. These factors showed a high degree of combinatorial regulation and moreover, can partially compensate each other’s function (Figure 3). This has direct implication for our understanding of the molecular background of CHD as it might explain incomplete penetrance and disease variability depending on the degree of functional alterations of discrete factors or in combination with secondary insults (e.g., environmental effects). Besides focusing on one molecular level such as the genome, transcriptome, or proteome independently, it is essential to interrogate different levels in order to understand the full biological architecture. It was shown that the functional read-out generated by distinct cardiac transcription factors such as Srf and Gata4 can be modulated by complementary histone acetylation and that regulatory circuits between the transcription factor Srf and microRNAs might contribute a high degree of indirectly affected targets genes in the situation of reduced function.

Further insights into regulatory cardiac transcription networks have been gained by interrogating gene expression in hearts of CHD patients and the anatomical, morphological, and haemodynamic features of these malformed hearts. Networks of correlated gene groups could be identified pointing once more to the modular architecture underlying heart development.

In addition to the immediate study of heart development, the analysis of cardiac disease harbouring remodelling features can also provide insights into network modules. Dilated cardiomyopathy is one such

Figure 3  Regulation of the cardiac transcriptome by combinatorial binding of the transcription factors Gata4 (green), Mef2a (red), Nkx2.5 (dark blue), and Srf (light blue). Shown are in total 1671 genes, which are targeted by at least one factor. All four factors share 91 target genes and sets of three and two factors share 121 and 286 genes, respectively. The combinatorial binding by these four key cardiac transcription factors is in line with their combinatorial regulation of shared down-stream targets, which is moreover interrogated by a number of indirect regulatory effects through microRNAs as well as epigenetic marks such as histone modifications. This illustrates the complex genetic, epigenetic, and environmental background underlying heart development and congenital heart disease.
exemplary cardiac disorder. A recent study of dilated cardiomyopathy analysed protein–protein interaction networks in combination with condition-specific co-expression information.\(^{64,126}\) Two dynamic modules, consisting of proteins essential for muscle contraction and organ morphogenesis, respectively, could be identified. Even though the importance of edges and nodes might differ during normal cardiogenesis, the general network properties identified should be considered with the disease state as a perturbation model. In a further high-throughput proteome study with integrated network biology, the consequences of Kir6.2 KATP channel pore deletion were analysed in respective knockout mice showing cardiomyopathic predisposition.\(^{127}\) Based on the integration of proteome and gene expression data, a KATP channel-associated subproteome could be identified and further assessment of this network showed the primary impact of the Kir6.2 KATP channel on metabolic pathways associated with cardiovascular disease.

4. Limitations of systems biology

Modelling of pieces that are interlinked in a way we do not understand yet is challenging and further complicated by the need to separate background noise from signal, which is a common problem of high-throughput data. The analysis of high-throughput data is based on the application of suitable statistical models, which estimate the significance of observed correlations and other dependencies. Thus, it is most crucial to select appropriate models and to carefully deal with the influence of noise and the high number of performed tests, e.g. by suitable correction for multiple testing. Although systems biology approaches is highly suggestive to derive new hypotheses, well-defined gene/protein-focused or pathway-focused experiments are essential to complement it and define the extent to which insights gained at a global scale hold true for specific segments of the system.

5. Concluding remarks and future perspective

The heart is the first organ fully functioning during development, and CHDs are the most common birth defect in human. Despite the heart being one of the central organs, we still have limited understanding about the molecular background driving heart development and its interferences leading to the burden caused by alterations of this process. The heart develops based on evolutionary conserved sequences of functional modules. Over the last decades important key regulators have been identified and insights have been gained by the study of animal models, as well as patients with CHD. Cardiac malformations represent a broad panel of in part overlapping phenotypes, potentially reflecting the modular background of cardiogenesis. However, the biological network modifying the impact of key regulators is still widely a black box. One important lesson learned from the study of animal models, as well as patients with complex CHD is that our primary hope that one gene would simply refer to one phenotype is not fulfilled. Even one particular mutation can be associated with a panel of different cardiac malformations and the majority of CHD is not following Mendelian inheritance. Nevertheless, it is undoubtful that there is a clear genetic impact.\(^{128,129}\) We currently enter a novel era of research, which is characterized by fast evolving technological advances enabling the generation, study and interpretation of more and more complex data. It is highly suggestive that this will allow the study of heart development in a systems biology manner that provides insights into the biological network influenced by genetic, epigenetic and environmental factors, and stochastic events. Two strategies can be envisaged: approaches utilizing already generated data (e.g. gene expression profiles and transcription factor–DNA binding)\(^{89,130–132}\) and the set-up of interdisciplinary projects where data generation (in particular perturbations) is driven by the input needs of modelling. The extraordinary conservation of the heart developmental process and its genes between human and model organisms enables the integration of data from different species.\(^{133,134}\) Hopefully, we can one day understand the causative impacts leading to the majority of CHD at some point, and thereof push forward the development of novel therapeutic and, even more important, preventive options such as e.g. folic acid treatment in place today.

Acknowledgements

I am grateful to Markus Schüler and Rainer Dunkel for preparation of graphics and thank Marcel Grunert, Cornelia Dorn and Markus Schüler for discussion and review of the manuscript. I apologize to those who have made contributions to our understanding of this research area and are not cited in this review.

Conflict of interest: none declared.

Funding

This work was supported by the European Community’s Seven Framework Program contract (‘CardioGeNet’) 2009-223463.

References

Systems biology of heart development and malformation


