Rui Sotero Rodrigues Machado

Identification of novel key factors of heart development using a systems biology approach

UNIVERSIDADE DO ALGARVE

Universidade do Algarve
Center for Biomedical Research
2021

Rui Sotero Rodrigues Machado

Identification of novel key factors of heart development using a systems biology approach

Doutoramento em Medicina
Regenerativa e Doenças Crónicas
ProRegeM

Trabalho realizado sob a orientação de:
Matthias Futschik, Investigador Principal
José Bragança, Investigador Principal

Universidade do Algarve
Center for Biomedical Research
2021

Identification of key factors of heart regeneration using a systems biology approach

Declaração de autoria do trabalho

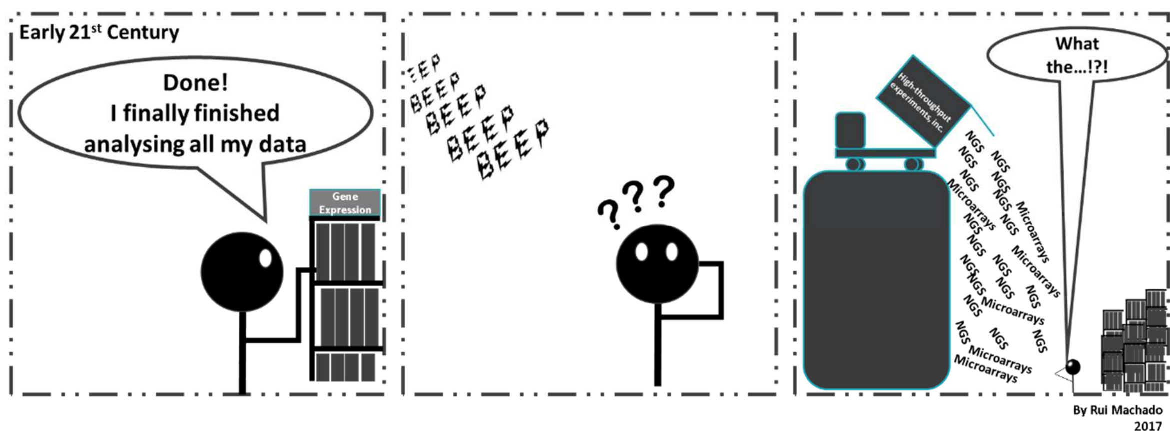
Declaro ser o autor deste trabalho, que é original e inédito.
Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluídas.
Declaro ainda que toda a recolha, tratamento e análise dos dados foram realizadas por mim.
As ferramentas *online* apresentadas e idealizadas neste doutoramento foram desenvolvidas em conjunto com o Investigador Principal Matthias Futschik e o Pós-Doc José Pedro Pinto.

____

(Rui Sotero Rodrigues Machado)

Copyright© Rui Sotero Rodrigues Machado

A Universidade do Algarve tem o direito, perpetuo e sem limites geográficos, de arquivar e publicitar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.



Microarrays and next generation sequencing are rapidly pervading all areas of human genetics: improving speed, precision, and last but not least breadth of diagnostics. In parallel, it profoundly affected basic biology discovery by unravelling mutation mechanisms, the causes, and consequences in gene regulation and what goes wrong in genetic disease and cancer.

Acknowledgements

The writing of this thesis was one of the most difficult challenge that I ever had to overcome during my life. I would like to say a “big thank you” to everyone that supported me during all this process and helped me overcome this hurdle.

I would like to start by thanking my family for the support during quest. They understood how important and difficult this was to me, providing me all the encouragement, love and patience to finally finish this task.

To Patrícia and my little Mafalda, for always being there for me in my ups and downs, good and bad moments, in these challenging years of my life. They gave me the strength to continue and improve myself little by little.

To my supervisor Matthias Futschik for his endless support and patience. I cannot thank him enough for all the encouragement that he provided me throughout the years to make me keep going. It was a pleasure to have him as my supervisor during this chapter of my life.

To José Pedro, for the good times and the programming skills that helped develop HeartmiR.

To every member that stopped by the SysBioLab group, I learned a lot from them throughout the years.

To José Bragança, for his support and scientific input during the thesis.

To Agapios Sachinidis, for the amazing collaboration, for having me in his lab and provide the microarray data that helped me develop this work.

To all the colleagues from ProRegeM PhD programme, especially the 2nd edition “ProRegeMers”, for all the fun moments and memories which I will never forget.

To Isabel Duarte, for the afternoons spent talking about a bit of everything.

To Ana Teresa Maia, for pushing me to do the final part of the thesis.

For last, but not least, to everyone that will have the chance to read this work and to the ones that might use it as a steppingstone for their own research.

Sincerely,
Rui Machado

This work was supported by the Portuguese Foundation for Science (FCT) through the doctoral grant PD/BD/105894/2014 from the ProRegeM PhD Programme (PD/00117/2012).

Abstract

Heart diseases are the leading cause of death worldwide. Although surgical interventions can provide valuable options for treatment, current therapies in cardiovascular medicine only delay disease progression. A main reason for this shortcoming is the limited regenerative capacity of the adult human heart. In contrast to many other tissues and organs, the mammalian heart has very limited regenerative capacity. However, it has been observed that neonatal hearts in mice show remarkable capacity to regenerate lost functional muscle tissue, a capacity that rapidly disappears after the first post-natal week. Therefore, the study of heart development might give crucial cues for cardiac regenerative medicine. Heart development or cardiogenesis is a highly complex process with many components that are finely tuned in a precise manner across time and space. Regulation of gene expression plays an important role in this process. To capture this level of regulation, technologies such as microarrays or next generation sequencing provide powerful tools, as they enable the simultaneous measurement of expression levels of thousands of coding and non-coding genes. Although, it is possible to obtain the expression information of several thousand of genes, there is a clear lack of platforms in which research can scan through this information to develop or generate insightful biological questions in the field of heart study.

Hence, this doctoral research work has tried to provide different systems biology approaches in order to offer new insights into gene expression events that occur mainly during heart development. These approaches include: (i) the integration of more than 20 published microarray studies related to cardiogenesis and the development of the HeartEXpress database (http://heartexpress.sysbiolab.eu/) to provide an easy and public access to the integrated data; (ii) the integrative analysis of a genome-wide study profiling coding and non-coding genes during embryonic heart development *in vivo*; (iii) the assessment of transcription factors and miRNAs previously associated to heart development; (iv) the integration and prioritisation of miRNA-mRNA interactions to identify novel miRNAs, mRNAs or miRNA-interactions with potential impact on cardiogenesis; (v) the development of a web-server called HeartmiR (http://heartmir.sysbiolab.eu/), which enables independent query and visualisation of miRNA-mRNA interactions obtained from the *in vivo* study; and (vi) comparative analysis of *in vivo* and *in vitro* studies to obtain further insights into mRNA, miRNA

and miRNA-mRNA interactions during embryonic stem cell differentiation and to clarify how the *in vitro* experiment can be used as a faithful model to study embryonic heart formation.

The main contributions of this research work for the study in the heart field are:

1. The development of HeartEXpress, which is a database that integrates the expression of more than 16400 genes and 130 experimental conditions in both human and mouse;
2. The development of HeartmiR, which is a database profiling the expression of 9211 mRNA and 386 microRNAs during the heart development period (from E10.5 to E19.5) and additionally in adult and old murine heart tissue;
3. Identification of the potential of 165 miRNAs to be involved in heart development using different methods of miRNA candidate prioritisation;
4. Identification of 102 miRNA and 214 putative novel miRNA-mRNA interactions relevant for cardiac cell development in vivo and in vitro. Furthermore, from the top20 miRNA with most interactions, 12 of the miRNA (60%) had been already associated to heart related events, indicating promising results for the remaining 8 miRNAs (40%)

In summary, I have developed, implemented and applied different systems biology approaches to analyse both publicly available and new generated experimental data. As result, I was able to identify potential novel coding and non-coding key factors important for cardiogenesis that might be utilised as markers or targets in future cardiac regenerative medicine strategies.

Keywords: Systems Biology; Transcriptomics; Gene Regulation; Heart Development; Molecular Interactions.

Resumo

As doenças cardíacas são uma das principais causas de morte a nível mundial. Apesar das intervenções cirúrgicas serem uma opção de tratamento viável, as terapias atuais apenas conseguem atrasar a progressão da doença. A principal razão para esta insuficiência é a capacidade regenerativa limitada do coração humano adulto. Em contraste com muitos outros tecidos e órgãos, o coração do mamífero tem uma capacidade de regeneração muito limitada. No entanto, foi observado, em corações de ratinhos neonatais, que ainda existe a incrível capacidade de regenerar tecido cardíaco perdido, a qual desaparece rapidamente após a primeira semana depois do nascimento. Portanto, o estudo do desenvolvimento cardíaco poderá providenciar pistas cruciais para posteriormente aplicar à medicina regenerativa cardíaca. O desenvolvimento cardíaco, também designado por cardiogénese, é um processo altamente complexo que envolve vários intervenientes que são coordenados de uma forma precisa no espaço e no tempo. Para tornar possível a captura deste nível complexo de regulação, tecnologias como os *microarrays* e *next generation sequencing* são ferramentas poderosas que permitem a medição simultânea dos níveis de expressão de milhares de genes codificantes e não codificantes.

No decorrer deste trabalho, diferentes métodos de biologia de sistemas foram implementados com o objetivo de fornecer um vislumbre da expressão genética que ocorre principalmente durante o desenvolvimento cardíaco. Como tal, foram realizadas: (i) a análise de mais de 20 estudos publicados de *microarrays* relacionados com o desenvolvimento cardíaco e, consequentemente, desenvolvida a ferramenta *online* HeartEXpress (http://heartexpress.sysbiolab.eu/) que permite o acesso público à base de dados tratados; (ii) a análise de um estudo do genoma que perfila os genes codificantes e não codificantes durante o desenvolvimento cardíaco embrionário *in vivo*; (iii) a análise de fatores de transcrição previamente associados com o desenvolvimento cardíaco; (iv) a apreciação de interações miRNA-mRNA para revelar novos miRNAs, mRNAs ou interações miRNA-mRNA que possam potencialmente ter um papel preponderante durante a cardiogénese; (v) o desenvolvimento de uma ferramenta *online* nomeada HeartmiR (http://heartmir.sysbiolab.eu/), que possibilita a consulta independente das interações obtidas no estudo do desenvolvimento cardíaco embrionário *in vivo*; e (vi) a análise comparativa de um estudo *in vitro* com o estudo *in vivo* para fornecer pistas adicionais sobre

os miRNAs, mRNAs e interações miRNA-mRNA durante a diferenciação de células estaminais embrionárias e compreender como o modelo *in vitro* pode ser utilizado para o estudo do desenvolvimento do coração.

As principais contribuições deste trabalho de investigação doutoral na área de estudo do coração são:

1. O desenvolvimento do HeartEXpress, sendo esta uma base de dados que integra a expressão de mais de 16400 genes e 130 condições experimentais tanto em humano como em ratinho;
2. O desenvolvimento do HeartmiR, que é uma base de dados que perfila a expressão de 9211 mRNA e 386 miRNAs durante o desenvolvimento cardíaco (de E10.5 a E19.5) e adicionalmente tecido cardíaco de ratinho em jovem adulto e velho;
3. Identificação de 165 miRNAs que podem estar potencialmente envolvidos no desenvolvimento cardíaco utilizando métodos de prioritização de candidatos;
4. Identificação de 102 miRNAs e 214 potenciais interações miRNA-mRNA relevantes para o desenvolvimento cardíaco celular *in vivo* e *in vitro*. Adicionalmente, dos top20 miRNAs com mais interações, 12 deles (60%) já foram associados a eventos cardíacos, indicando que os restantes 8 miRNAs (40%) serão promissores para investigações futuras.

Os resultados demonstraram que os diferentes métodos de biologia de sistemas aplicados forneceram resultados interessantes de diferentes formas para estudos adicionais ao desenvolvimento cardíaco e, possivelmente, no futuro aplicar a estudos de medicina regenerativa. Para além dos mRNAs e miRNAs que vão ser apresentados neste trabalho e que já foram posteriormente associados ao desenvolvimento cardíaco de algum modo, muitos outros foram descobertos durante o decorrer deste trabalho com o potencial de serem inovadores na área da regeneração e desenvolvimento cardíaco. Por exemplo, a descoberta de genes co-expressos relacionados com desenvolvimento cardíaco poderão providenciar pistas sobre os padrões de expressão que ocorrem durante os eventos relacionados com o coração. Como é no caso de estudo do *Isl1*, um conhecido fator de transcrição cardíaco, que apresenta como padrão de co-expressão outros três conhecidos fatores ligados ao desenvolvimento cardíaco tais com o *Gata3*, *Wnt2* e *Wnt11*. Apresenta ainda a co-expressão

com outros genes que potencialmente poderão ser importantes no desenvolvimento cardíaco, nomeadamente três genes Riken (um codificante e dois não-codificantes longos). Em contrapartida, investigar interações miRNA-mRNA também é um passo importante, visto que o desenvolvimento cardíaco é um processo altamente regulado, sendo assim possível indiciar potenciais candidatos miRNAs que possam contribuir para o desenvolvimento cardíaco. Neste estudo foram descobertos 3 miRNAs candidatos no estudo *in vivo* (Let-7i, mir-3472 e miR-490-3p) que apresentam um grande número de genes "alvo", sendo indicados como potenciais reguladores do desenvolvimento cardíaco. Por último, a comparação dos estudos de *microarrays in vitro* e *in vivo* também foi um ponto importante, uma vez que forneceu informação fulcral sobre as semelhanças e disparidades entre estas duas distintas experiências laboratoriais, contribuindo para entender se o modelo *in vitro* será um paradigma apropriado para estudar o desenvolvimento cardíaco. Não obstante, os resultados revelaram ser encorajadores, visto terem sido encontradas interações miRNA-mRNA sobrepostas entre os dois estudos, contendo miRNAs que poderão ser promissores para validações experimentais futuras, como acontece, por exemplo, com o miR-608 que foi encontrado como diferencialmente expresso nos dois estudos – *in vitro* e *in vivo* – e tem como alvo seis genes que já foram previamente associados ao desenvolvimento cardíaco.

Em suma, este trabalho utilizou diferentes metodologias da biologia de sistemas que permitiram utilizar e analisar informação publicamente disponível e dados laboratoriais que permitiram identificar potenciais fatores-chave codificantes e não-codificantes para a cardiogénese e, possivelmente, aplicar no futuro essa informação em validações laboratoriais e na medicina regenerativa.

Palavras-chave: Biologia de sistemas; Transcriptómica; Regulação de genes; Desenvolvimento cardíaco; Interações moleculares

INDEX

Figure Index ... i

Supplementary Figure Index ... v

Table Index ... vii

Supplementary Table Index ... ix

Abbreviations ... xi

1. General Introduction ... 1

1.1 The formation of a functional heart ... 3

1.1.1 Mammalian early embryonic development ... 3

1.1.2 Early cardiac development ... 6

1.1.3 First and second heart field ... 7

1.1.4 Transcription factors ... 9

1.1.5 miRNAs ... 14

1.2 ESCs as model for heart study ... 18

1.3 Stem cells as a therapeutic approach for heart regenerative medicine ... 20

1.4 Technologies in molecular analysis and diagnostics ... 22

1.4.1 Microarrays in heart study ... 24

1.4.2 Identification of novel heart related candidates through microarrays ... 26

2. Bioinformatic Methods ... 27

2.1 Microarray data processing ... 29

2.2 Microarray data quality analysis ... 31

2.3 Microarray differential expression analysis ... 32

2.4 Clustering of differentially expressed genes and miRNAs ... 33

2.5 Gene ontology enrichment analysis of differentially expressed genes and miRNAs ... 34

3. HeartEXpress – A web database for exploration and visualisation of integrated cardiac gene expression data ... 35

3.1 Introduction ... 37

3.2 HeartEXpress assembly ... 38

3.2.1 Data collection and integration on HeartEXpress ... 38

3.2.2 Clustering in HeartEXpress ... 41

3.3 HeartEXpress content & structure 41
3.3.1 HeartEXpress data processing and quality control 41
3.3.2 HeartEXpress functionality.. 47
3.3.3 HeartEXpress utility.. 50
3.4 Conclusion 52
4. Genome-wide profiling of coding and non-coding RNAs to identify novel regulatory components in embryonic heart development 53
4.1 Introduction.. 55
4.2 Experimental design.. 57
4.3 Results and discussion 58
4.3.1 Analysis of microarray profiles of mRNA and miRNA transcriptomes.. 58
4.3.2 Identification of differentially expressed genes in the developing heart.. 59
4.3.3 Inspection of genes associated with heart development.. 59
4.3.4 Differential miRNA expression 64
4.3.5 Dynamics of gene expression during embryonic heart development.. 66
4.3.6 Genes expressed in an age- or gender-dependent manner 67
4.3.7 Gene Expression of transcripts of unknown function during the heart development and adult heart 69
4.4 Conclusion 71
5. HeartmiR: A web tool for visualisation of miRNA-mRNA regulatory interactions with potential relevance for heart development 73
5.1 Introduction.. 75
5.2 Curation and integrative analysis of potential regulatory miRNA-mRNA interaction data for HeartmiR 76
5.3 *In vivo* data dual transcriptome analysis 79
5.3.1 Integrative analysis of dual transcriptome data and miRNA gene targets.. 79
5.3.2 Prioritisation of miRNA candidates involved in heart development and maturation.. 83
5.4 HeartmiR structure and querying.. 86
5.5 Conclusion 89

6. Detection of novel potential regulators of stem cell differentiation and cardiogenesis through combined genome-wide profiling of protein coding transcripts and microRNA 91

6.1 Introduction 93

6.2 Experimental design 94

6.3 Results and discussion 95

6.3.1 Differential gene expression during *in vitro* differentiation 95

6.3.2 Dynamic expression of marker genes and miRNAs 96

6.3.3 Identification of miRNA that might serve as genetic switches 98

6.3.4 Embryonic bodies as *in vitro* model for cardiogenesis on transcriptional level 102

6.3.5 Detection of shared *in vitro* and *in vivo* miRNA-mRNA interactions 105

6.3.6 Putative novel candidate miRNA-mRNA interaction relevant for cardiac cell development *in vivo* and *in vitro* 107

6.4 Conclusion 109

7. Summary and Outlook 111

8. Bibliography 119

Figure Index

Figure 1.1.1 First stages of human and mouse embryonic development........................ 4

Figure 1.1.2 A sketch of the different mammalian cell lineages that compose the organs........................ 5

Figure 1.1.3 Embryology of the human heart........................ 6

Figure 1.1.4 Heart fields in cardiogenesis........................ 7

Figure 1.1.5 Transcription factors involved in cardiogenesis........................ 9

Figure 1.1.6 Cardiac transcription factors interactions based on textmining, experimental validation and gene co-occurrence........................ 12

Figure 1.1.7 Multistep process for the formation of mature miRNAs........................ 14

Figure 1.1.8 miRNAs in stem cells and heart development........................ 16

Figure 1.2.1 The ESCs-EBs system that provides a model for commitment of progenitors to specific fates and subsequent identification of multipotent progenitors for cardio-vascular derivatives........................ 19

Figure 1.3.1 Stem cell biology and reprogramming........................ 20

Figure 1.3.2 Simplified applications of iPSC for therapy........................ 21

Figure 1.4.1 Simplified overview of workflow for microarrays and NGS technology data processing........................ 22

Figure 1.4.2 General overview about the advantages and disadvantages of microarrays and RNA-seq (NGS)........................ 23

Figure 2.1.1 Pre-processing for microarray intensity signals to obtain the expression levels of the transcripts........................ 30

Figure 3.3.1 PCA for the study by Gaspar *et al*., 2012........................ 42

Figure 3.3.2 Density histogram for the study by Gaspar *et al*., 2012........................ 43

Figure 3.3.3 Dendrogram clustering for the study by Gaspar *et al*., 2012........................ 43

Figure 3.3.4 Key marker genes expression in the study Gaspar *et al*., 2012........................ 44

Figure 3.3.5 PCA of heart related expression data for mouse........................ 45

Figure 3.3.6 PCA of human heart related expression data........................ 46

Figure 3.3.7 HeartEXpress Intro page........................ 47

Figure 3.3.8 HeartEXpress Integrated dataset page........................ 48

Figure 3.3.9 Data sets can be selected on the *Gene Expression* page........................ 48

Figure 3.3.10 HeartEXpress gene expression analysis page... 49

Figure 3.3.11 HeartEXpress utility applied to the murine *in vivo* cardiogenesis data-set.. 51

Figure 4.2.1 Experimental design of parallel monitoring of mRNA and miRNA during heart development and in two adult stages.. 57

Figure 4.3.1 Principal component analysis of the transcriptome data.......................... 58

Figure 4.3.2 Number of differentially expressed genes (DEGs) in developing heart and older adult heart tissues, compared with young mature heart tissue.............................. 59

Figure 4.3.3 Temporal expression profiles of selected genes with existing association with heart development.. 60

Figure 4.3.4 Temporal expression profiles of selected genes with existing association with heart development (II)... 61

Figure 4.3.5 Temporal expression profiles of selected genes with existing association with heart development and ion channel function.. 62

Figure 4.3.6 Temporal expression profiles of selected genes with cell cycle.................. 63

Figure 4.3.7 Number of differentially expressed miRNAs (DEmiRs) in developing heart and older adult heart tissues, compared with young mature heart tissue.................. 64

Figure 4.3.8 Temporal expression profiles of selected differentially expressed microRNAs (DEmiRs) for embryonic and mature heart tissues.. 65

Figure 4.3.9 Number of differentially expressed genes and miRNAs in developing heart and older adult heart tissues, compared with E19.5 heart tissue......................... 66

Figure 4.3.10 Comparison of gene expression in male and female samples.................... 67

Figure 4.3.11 Differentially expressed genes encoding for mRNA without current functional annotation... 69

Figure 4.3.12 Differentially expressed long non-coding genes without current functional annotation... 70

Figure 5.3.1 Integrative analysis of clusters of DEmiRs and DEGs................................. 80

Figure 5.3.2 Prioritisation of cardiac miRNAs by regulatory activity............................. 83

Figure 5.3.3 Prioritisation of cardiac miRNAs by functional annotation analysis............ 84

Figure 5.3.4 Template-based prioritisation of cardiac miRNAs...................................... 85

Figure 5.4.1 Workflow scheme for HeartmiR.. 87

Figure 6.2.1 Time series experiment layout, to profile simultaneously mRNAs and miRNAs during *in vitro* stem cell differentiation... 94

Figure 6.3.1 mRNA and miRNA profile analysis of *in vitro* ESCs differentiation.............. 95
Figure 6.3.2 Pluripotency and differentiation gene markers... 97
Figure 6.3.3 Pluripotency and differentiation miRNAs markers................................... 98
Figure 6.3.4 miRNAs target expression profiles.. 100
Figure 6.3.5 Temporal expression profiles of genes associated to heart development and ion channel function.. 103
Figure 6.3.6 Correlation of *in vitro* and *in vivo.* expression of miRNA-mRNA interactions.. 106
Figure 7.1 General overview of the work and investigations undertaken to better understand heart development and uncover novel regulatory mechanisms and factors during cardiogenesis in this thesis 114

Supplementary Figure Index

Figure S4.3.1 Clustering dendrograms produced based on mRNA and miRNA micro-array profiles.. 1

Figure S4.3.2 Density plots of log2-tranformed expression intensities for mRNA and miRNA.. 1

Figure S4.3.3 Expression profiles of *Casq2* and the intronic lncRNA 4632404M16Rik.. 1

[1] To consult the supplementary figure data, please visit: "https://drive.google.com/drive/folders/1GEOOe0uaTQvTHXdnmmpRkatKWVDU7x8s?usp=sharing"

Table Index

Table 3.2.1 HeartEXpress integrated datasets for human and mouse........................... 40

Table 5.3.1 Anti-correlated miRNAs to the specific gene subcategories that are present in Chapter IV – Section 4.3.2: "Markers for heart development and function show distinct expression changes".. 82

Table 6.3.1 Top 20 miRNAs targeting heart related genes.. 107

Supplementary Table Index

Table S3.2.1 An overview of all studies integrated in HeartEXpress together with detailed information... 2

Table S4.3.1 List of DEGs at different developmental stages of the mouse heart and in old heart tissue using as reference young adult heart tissue.. 2

Table S4.3.2 DEGs Annotated to GO: 0007507 "Heart Development"........................... 2

Table S4.3.3 DEGs Annotated to GO: 0005216 "Ion Channel Activity"........................... 2

Table S4.3.4 List of DEmiRs at different developmental stages of the mouse heart and in old heart tissue using as reference young adult heart tissue...................................... 2

Table S4.3.5 List of DEGs at different developmental stages of the mouse heart using E19.5 and young adult heart as reference... 2

Table S4.3.6 List of DEmiRs at different developmental stages of the mouse heart using E19.5 and young adult heart as reference.. 2

Table S4.3.7 List of DEGs expressed in an age- or gender-dependent manner............... 2

Table S4.3.8 Differentially expressed Riken genes at different heart developmental stages in mouse.. 2

Table S5.2.1 HeartmiR interaction list database... 2

Table S5.3.1 miRNAs profiling regarding their regulatory activity................................. 2

Table S5.3.2 List of miRNAs targeting mRNAs annotated to Heart Development (GO: 0007507)... 2

Table S6.3.1 DEGs at different stem cell differentiation stages using as reference ESCs.. 2

Table S6.3.2 DEmiRs at different stem cell differentiation stages using as reference ESCs.. 2

Table S6.3.3 List of marker genes and miRNAs for pluripotency maintenance and differentiation.. 2

Table S6.3.4 List of miRNAs with peak expression patterns... 2

Table S6.3.5 List of miRNAs targets expression profiles... 2

[2] Since the supplementary tables are too large to display in the appendix, to consult the supplementary data, please visit: "https://drive.google.com/drive/folders/1GEOOe0uaTQvTHXdnmmpRkatKWVDU7x8s?usp=sharing"

Table S6.3.6 List of miRNA-mRNA interactions overlapping in the *in vitro* and *in vivo* (chapter 5) studies.. 3[3]

Table S6.3.7 List of miRNA-mRNA interaction of miRNAs that target genes associated with heart development (GO: 0007507)... 3

Table S6.3.8 List of miRNA-mRNAs interacting pairs with negative correlation in *in vitro* and *in vivo* studies.. 3

Table S6.3.9 miRNAs targeting heart related genes... 3

[3] Since the supplementary tables are too large to display in the appendix, to consult the supplementary data, please visit:
"https://drive.google.com/drive/folders/1GEOOe0uaTQvTHXdnmmpRkatKWVDU7x8s?usp=sharing"

Abbreviations

Actc1 – Cardiac alpha actin gene

CF – Cardiac fibroblasts

CCS – Cardiac conduction system

CM – Cardiomyocytes

CPC – Cardiac progenitor cells

CSC – Cardiac stem cells

CV – Cardinal veins

DAW – Dorsal atrial walls

DEmiRs – Differentially expressed miRNAs

DEGs – Differentially expressed genes

DETGs – Differentially expressed target genes

DETGNCs – Differentially expressed target genes negatively correlated

EBs – Embryoid bodies

ESCs – Embryonic stem cells

eSet – Expression set

FDR – False discovery rate

FHF – First heart field

Gata4 – Transcription factor, Gata binding protein 4

GEO – Gene Expression Omnibus

GO – Gene Ontology

Hand2 – Transcription factor, heart and neural crest derivatives expressed transcript 2

hESCs – Human embryonic stem cells

(h)iPSC – (Human) induced pluripotent stem cells

iCM – Induced cardiomyocytes or cardiomyocytes derived from iPSC

iPSC – Induced pluripotent stem cells

lncRNAs – Long noncoding RNAs

LV – Left ventricle

Mef2c – Transcription factor, myocyte enhancer factor 2c

mESCs – Mouse embryonic stem cells

miRNAs – microRNAs

mRNAs – Messenger RNAs

NGS – Next-Generation Sequencing

Nkx2-5 – NK2 Homeobox 5

Nt – Nucleotides

OT – Outflow tract

PCA – Principal component analysis

PEO – Pro-epicardial organ

PV – Pulmonary vein

RMA – Robust multiarray average

RV – Right ventricle

SHF – Second heart field

Tbx5 – Transcription factor T-box 5

TFs – Transcription factors

Chapter I

General Introduction

1.1 The formation of a functional heart

This thesis aims to improve our understanding of the molecular and cellular events that occur during the formation of the heart, the first functioning organ in the developing embryo. As heart development is highly conserved, observations made in model organisms can provide valuable insights into cardiogenic mechanisms with potential clinical relevance for humans.

In the introduction, I will give a brief overview about early embryonic development and heart development, followed by a presentation of key regulators that are responsible for the correct formation of a functional heart. Subsequently, I will elaborate on the use embryonic stem cells (ESCs) as a model for heart study and how they can be potentially used in therapeutic approaches. Finally, I will present a perspective how genomic approaches such microarray analysis can be used to study heart development in systems biology.

1.1.1 Mammalian early embryonic development

After fertilisation, the zygote undergoes through a series of cleavage, divisions to form a structure called morula (Figure 1.1.1 – embryonic day 2.5 in mouse and day 3 in human). At this early stage, the outer and inner cells of the morula start to acquire different features. The inner cells of this embryonic structure give rise to the epiblast, from which the adult organism originates, while the outer cell lineage supplies the extraembryonic structures that will provide support and nutrients for embryo development[1]. This leads to the formation of a spherical structure called blastocyst (Figure 1.1.1 - embryonic day 4 in mouse and day 6 in human) formed by an outer layer of cells termed the trophectoderm (TE) and an aggregate of cells localised inside of the blastocyst structure termed the inner cell mass (ICM). The ICM is composed by progenitor cells that give rise to the epiblast (primitive ectoderm) and the hypoblast (primitive endoderm)[2].

The hypoblast contributes to the formation of extraembryonic membranes, while the epiblast will give rise to the embryo. As the blastocyst invades the uterine epithelium, it undergoes morphological modifications associated with the gastrulation process that turns the bilaminar disk into a tri-layered embryo with the three germ layers i.e. the ectoderm, endoderm and mesoderm[2].

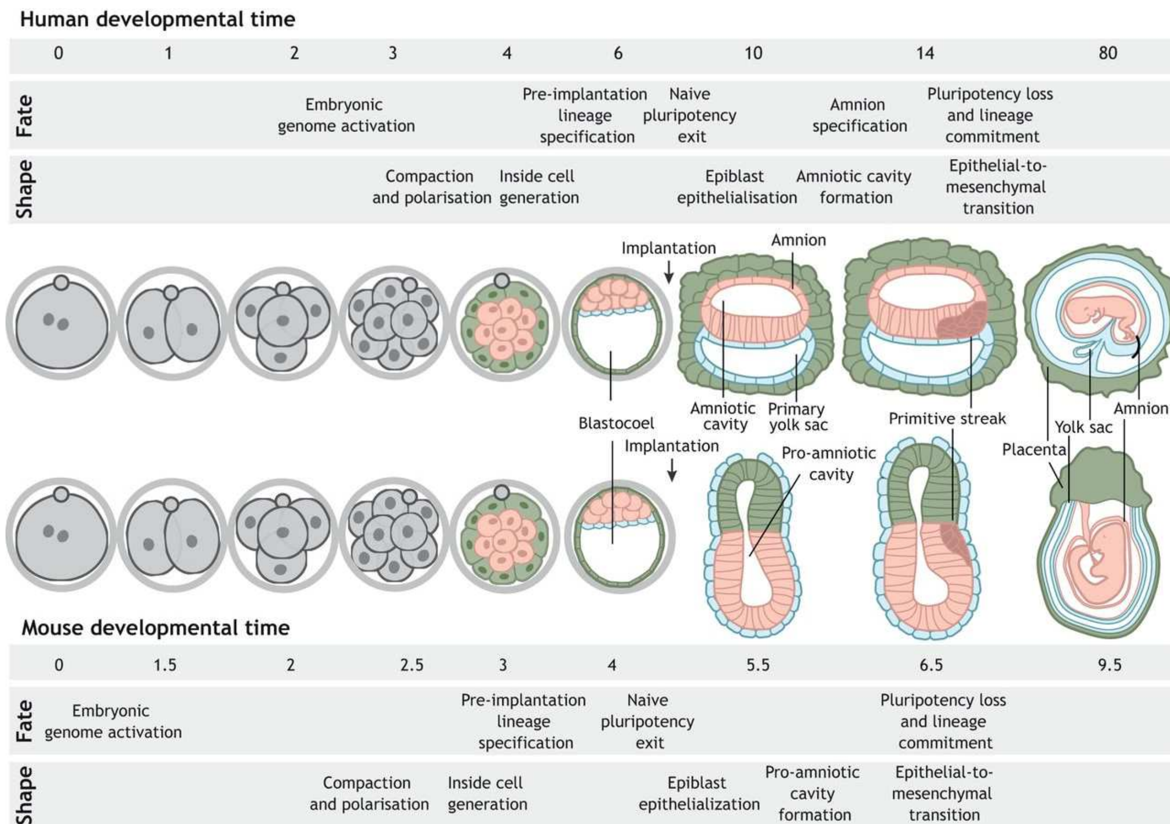


Figure 1.1.1 First stages of human and mouse embryonic development (adapted from Shahbazi *et al.,* 2020)[3]. The formation of the blastocyst occurs around embryonic day 4 in mouse and day 6 in human. Around day 5 (mouse) and day 7 (human), the embryonic structure undergoes global morphological transformation. On embryonic days 6.5 (mouse) and 14 (human), gastrulation is initiated in the posterior epiblast. The cells start to lose their pluripotent character and commit to one of the germ layers. Epiblast-derived tissues are shown in pink, hypoblast-derived tissues are shown in blue and trophectoderm-derived tissues are shown in green[3].

Each germ layer will give rise to specific structures of the embryonic body. The ectoderm differentiates to form the central nervous system and the epidermis. The endoderm provides the gastrointestinal and respiratory tract and is responsible for development of many endocrine glands and organs, such as the liver and pancreas.

The mesoderm transitionally differentiates into the following four compartments known as: axial, paraxial, intermediate, and lateral plate mesoderm. The axial mesoderm gives rise to the notochord, which is responsible for formation of the intervertebral disks. From the paraxial mesoderm the somites will give rise to the axial skeleton, skeletal musculature, and the dermis.

Furthermore, the branchial arches will develop from this compartment, which will form the facial muscle and cartilage. The intermediate mesoderm develops into the urogenital system comprising the kidneys and gonads. Finally, the lateral plate mesoderm is involved in the formation of the body wall, viscera and the circulatory system including the heart. Figure 1.1.2 provides an overview of the different cell lineages that can be originate from these three germ layers.

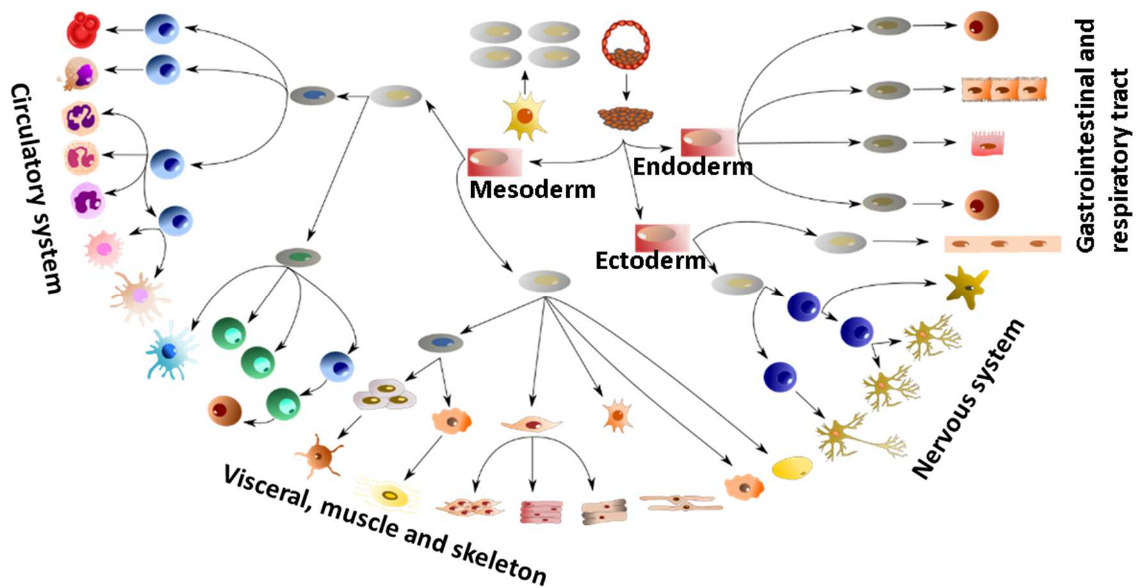


Figure 1.1.2 A sketch of the different mammalian cell lineages that compose the organs (Adapted from StemMapper logo)[4].

Heart development in human starts around day 19 (while in mouse it begins around 7 days post fecundation (dpf)) in the lateral plate mesoderm region[5]. which has been already specified by the upregulation of certain genes related to early cardiac specification events, such as, *Nkx2-5* and *Gata4*[6,7]. The lateral plate divides into somatic and splanchnic mesoderm. The latter is located in both sides of the primitive streak, which will fuse to form a primary heart tube with a venous and an arterial pole[8]. In human, at day 23 (or 8 dpf for mice) the heart tube starts to loop due to the influence of genes like *Nodal* and *Lefty*[9] that drive left and right specification. At this stage of looping, the heart is composed by a primitive ventricle, from which the right ventricle and left ventricle will originate, and single primordium atrium. At subsequent stages, the septation, outgrowth and remodelling of the individual heart chambers will eventually lead to a four-chambered mature heart around day 35 in human[10] (Figure 1.1.3).

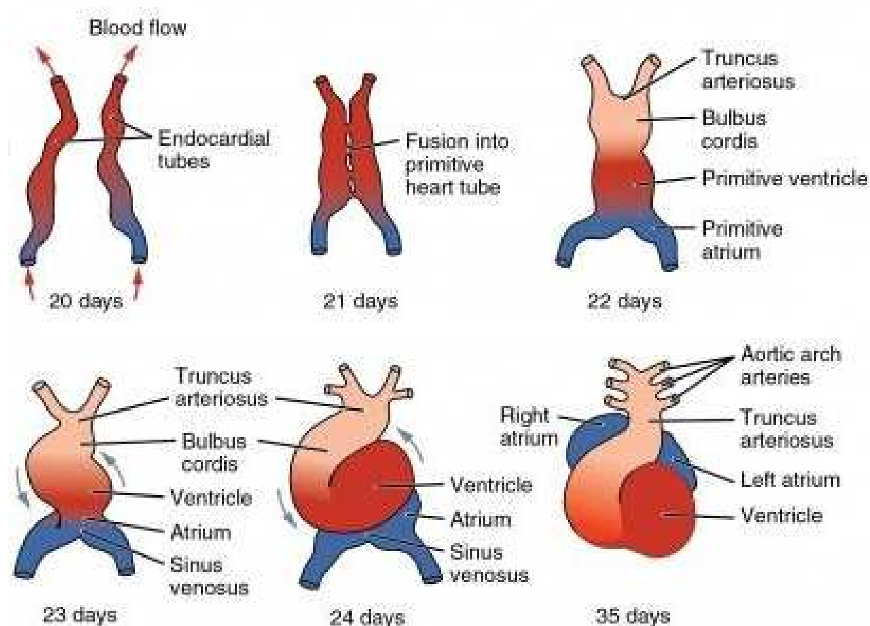


Figure 1.1.3 Embryology of the human heart (adapted from Anatomy and Physiology book)[10].

All structural cells present in the heart are derived from cardiac progenitor cells (CPCs) which are able to differentiate into different cell types, such as smooth muscle cells, endothelial cells and cardiomyocytes[11]. The heart is composed by three main layers. The innermost layer – the endocardium – which is composed mainly by endothelial cells, smooth muscle cells and cardiac neural crest cells (these latter cells derive from ectoderm).

The intermediate layer – the myocardium – mostly harbours cardiomyocytes and provides the contractile power of heart. The outermost layer – the epicardium – which is composed by proepicardium cells, which are multipotent cells with the capability of differentiate not only into epicardium cells, but also into cardiomyocytes, fibroblast as coronary vessels[12].

1.1.3 First and second heart field

To uncover the developmental route of individual or groups of cells is a superb challenge. Sophisticated research techniques enabled the discovery that not all cardiac cells have a unique developmental origin. In 2001, it was confirmed that there exists two main and distinct groups of CPCs: (1) the progenitors derived from the first heart field (FHF) and (2) the progenitors derived from the second heart field (SHF)[13]. The FHF can be considered as a first wave of CPCs that give rise to the initial heart tube, while the SHF CPCs initially continue to proliferate without differentiating further[8]. The FHF largely contributes for the atria and left ventricle (LV) formation while the SHF generates the right ventricle (RV), outflow tract (OT), and contributes to some extent to the formation of the atria and several other structures in the heart (Figure 1.1.4).

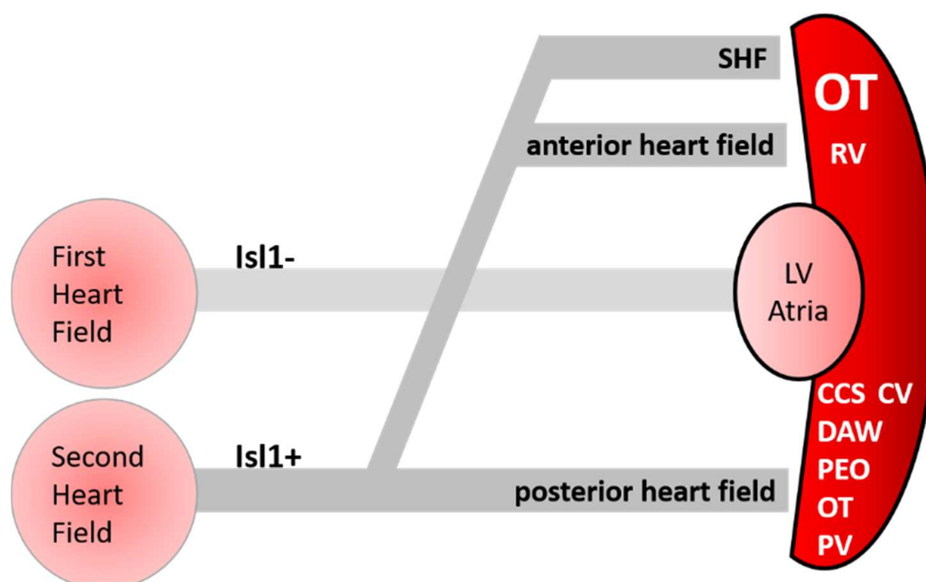


Figure 1.1.4 Heart fields in cardiogenesis. This schematic gives a general view about the contributions of different heart fields during cardiogenesis. The FHF gives rise to the left ventricle (LV) and atria. In contrast, the contribution of the SHF is more divers, being responsible for the formation of the outflow track (OT), right ventricle (RV), cardiac conduction system (CCS), cardinal veins (CV) dorsal atrial wall (DAW), pro-epicardial organ (PEO); pulmonary vein (PV). Figure adapted and simplified from Gittenberger-de Groot AC *et al*., 1998[14].

The segregation of these two different populations of CPCs occurs very early during gastrulation. This was demonstrated by a study, which showed two temporally distinct pools of progenitors that were already committed either to FHF or SHF during gastrulation[15].

The first known gene to play a critical role in SHF CPC differentiation into the different heart structures is the transcription factor Islet1 (*Isl1*). Lineage tracing of this transcription factor revealed that *Isl1* positive cells contributed to part of the atria development and full development of the OT[16]. Although other studies also verified the presence of low levels of *Isl1* in the FHF[17]. *Isl1* down regulation during development did not explicitly affect the structures originated from the FHF. In contrast, its absence in the SHF caused major impairments in the formation of heart structures derived from the SHF[16,18]. Hence, *Isl1* is considered as a key regulator and marker of the SHF cells. The cells present in the SHF are highly proliferative to ensure that there is no lack of CPC, thus the trade-off between proliferation and differentiation has to be maintained in check through balancing CPC proliferation promoting signalling pathways, such as *Wnt* and Hedgehog signalling pathways, with bone morphogenic proteins, Notch and non-canonical *Wnt* driving differentiation and maturation[19].

In the next section, I will introduce some key transcription factors that are downstream of cardiogenic signalling pathways and responsible for correct anatomical and functional heart development.

Heart development is an event that is tightly regulated by the occurrence of spatial and temporal specific events resulting in the correct formation of a functional heart. These cardiogenic events are highly conserved across mammals as documented in several *in vitro* and *in vivo* studies (see review by Günthel *et al*., 2018)[20]. One main reason for the high degree of evolutionary similarity is the conserved use of a defined set of transcription factors (Figure 1.1.5).

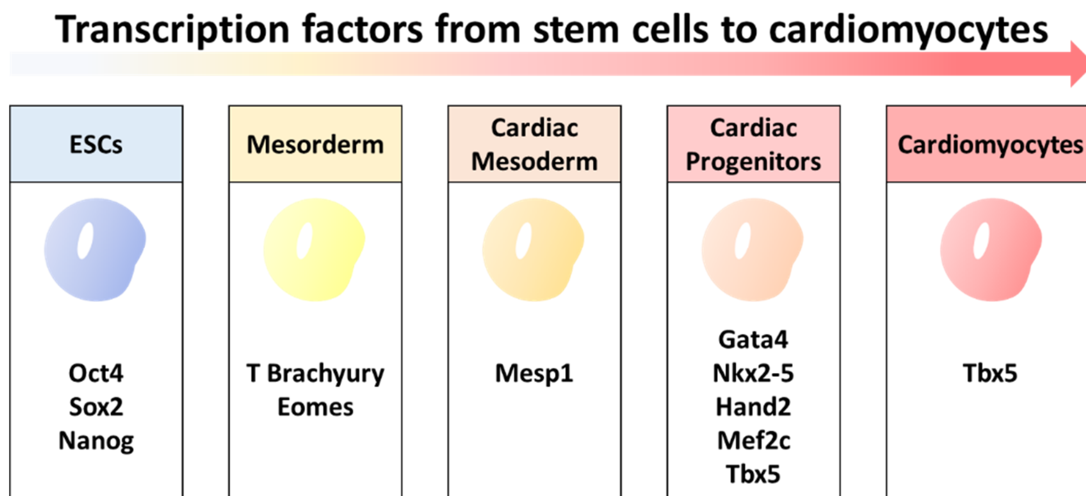


Figure 1.1.5 Transcription factors involved in cardiogenesis. Pluripotent cells are mainly characterised by the expression of *Oct4, Sox2,* and *Nanog.* Early mesoderm cells express *T Brachyury* and *Eomes*, while the expression of *Mesp1* marks a more advanced stage of specification into cardiac mesoderm. Cardiac progenitors highly express *Gata4, Nkx2-5,* and *Hand2.* In mature stage, cardiomyocytes express *Tbx5*[21] (Adapted from Kamps *et al.,* 2016).

In essence, a transcription factor is a protein with the capability of binding the DNA transcriptional regulatory elements of their target genes, such as promotor or enhancer regions. Furthermore, they are able to interact with RNA polymerase II or other transcription factors to regulate the number of transcripts produced by the target gene. Transcription factors can be found in all cells of an organism. Some transcription factor are essential for initiating patterns of gene expression that underly major developmental changes[22].

Pluripotency Markers

The most well-known and best described master regulators of pluripotency are *Nanog*, *Sox2* and *Oct4* (also known as *Pou5f1*). A fine regulation of the expression of these transcription factors is required to maintain cells in a pluripotent state. Pluripotent stem cells expressing these TFs, maintain the capability to differentiate into the three different germ layers[23], which is the basic definition of pluripotency.

When *Nanog* expression levels start to decrease, cells become more prone for differentiation. Despite not having by itself a weight in the cell fate decision[24], *Nanog* is essential for embryonic development, since *Nanog* null embryos do not developed beyond implantation[25]. Furthermore, overexpression of *Nanog* contributes towards cell pluripotency status maintenance and self-renewal[26].

Similarly to *Nanog*, *Sox2* ablation leads to early differentiation, causing embryonic lethality shortly after blastocyst implantation, due to the lack of the epiblast development[27]. Furthermore, low expression levels of *SOX2* in hESCs may be bias for the differentiation of cells towards the primitive streak structure[28]. Contrarily to what was observed for *Nanog*, *Sox2* overexpression in ESCs appears to push cells towards a neuroectodermal cell fate and to inhibit primitive streak differentiation[29].

Finally, *Oct4* was the first key transcription factor to be related to pluripotency. It is encoded by the gene *Pou5f1*, a member of the POU family. *Oct4* is known to have an impact in cell fate commitment, as higher expression levels in ESCs of this TF lead to the formation of mesendoderm. Furthermore, *Sox2* and *Oct4* were reported to play antagonistic roles in the differentiation of ESCs towards the neuroectodermal and mesendodermal fates, respectively[30]. Knockout (KO) of *Oct4* in mice resulted in the impairment of the epiblast development and consequent embryonic lethality[31]. *Oct4* depletion in ESCs led to spontaneous differentiation and incapability of differentiating into mesoderm cell fates[30]. Thus, the fine balance of expression between *Nanog*, *Oct4* and *Sox2* involved in pluripotency dictates ESCs behaviour, by controlling their pluripotency state or by promoting different cell fates.

Mesoderm Markers

Upon differentiation, stem cells begin to lose their pluripotency and become cells of one of the three germ layers, i.e., ecto-, meso- or endoderm. Early mesoderm is marked by the expression of two essential TFs *T-Brachyury* and *Eomes*odermin (*Eomes*). Early mesoderm specification is essentially controlled by *T-Brachyury*, and it is known that the lack of expression of this gene causes multiple defects in development and its ablation is lethal[32]. *T-Brachyury* expression peaks, around day 3 of ESCs differentiation. Furthermore, *T-Brachyury* is highly and specifically expressed in mesoderm[32,33].

One of the earliest known cardiac precursors to be expressed is the transcription factor *Eomes*[34], which promotes cardiovascular fate during ESCs differentiation. Human embryonic stem cells (hESCs) with *EOMES* knockout fail to differentiate into CMs. Furthermore, *EOMES* knockout hESCs failed to express any of the pan-cardiac genes such as *NKX2-5*, *MYH6*, or *CTNT*[35]. It is also known that *Eomes* is responsible for activating *Mesp1*, which is a key regulator for cardiovascular cell fate differentiation in mouse[36,37].

Cardiac Mesoderm Markers

Mesp1 was the first identified marker for early cardiac mesoderm. However this gene does not solely contribute to heart cell lineage formation, but also contributes to the formation of mesenchyme and limbs[38]. Inactivation of *Mesp1* results in an abnormal heart morphogenesis due to defective migration of heart precursor cells[38]. In contrast, *Mesp1* gain of function greatly increases cardiac differentiation both *in vitro* and *in vivo*, and *Mesp1* is essential for ESCs to exit the pluripotent state, migrate and promote cardiovascular specification[15,39]. A study performed in 2016, proved that cardiac progenitor cells expressing *Mesp1* exhibit the potential to repair infarcted mouse hearts, demonstrating that these cells differentiate into vascular smooth muscle cells, endothelial cells and cardiomyocytes in post-myocardial infarction of mice hearts and remained proliferative until 12-week post injection[40].

Cardiogenesis involves the activation of multiple transcription factors at precise timeframes during development. The core cardiogenic transcription factors include *Gata4*, *Mef2c*, *Hand2*, *Nkx2-5*, *Isl1*, amongst others, some of which have been mentioned above. These factors regulate each other's expression and control the expression of common downstream targets (Figure 1.1.6).

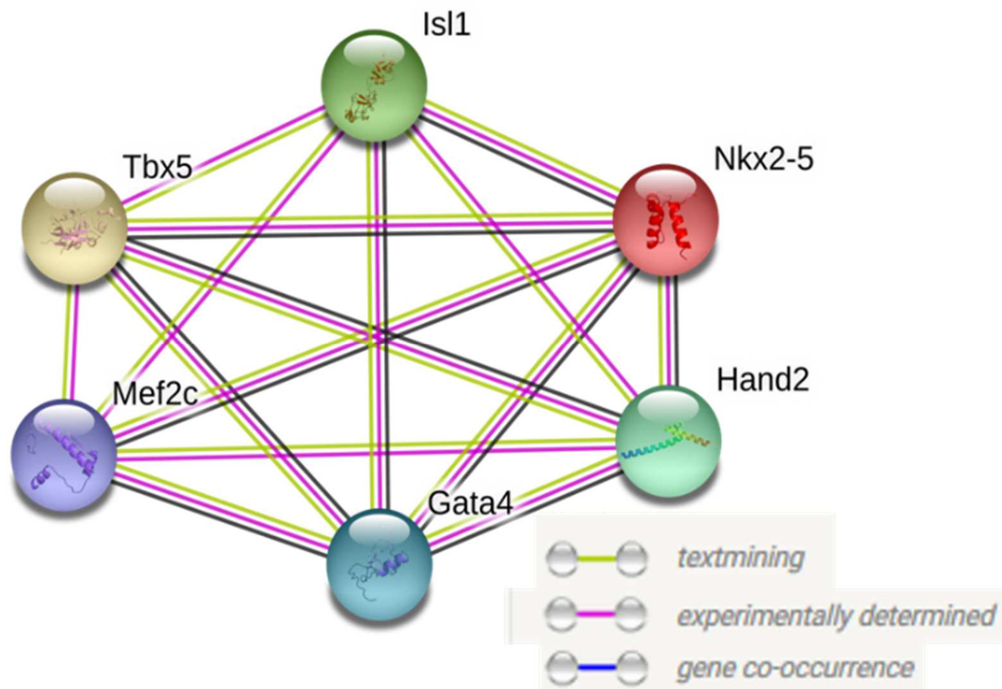


Figure 1.1.6 Cardiac transcription factors interactions based on textmining, experimental validation and gene co-occurrence. The selected cardiac transcription factors form a highly connected network indicated by the retrieved multiple interactions from the web application called STRING-DB[41].

As displayed in Figure 1.1.4, *Isl1* has been strongly associated with the potency of the SHF, since different studies revealed that *Isl1*$^+$ cells mostly contribute to the formation of the RV, OT, and atria. Indeed, the KO of *Isl1* in mice embryos results in their death during gestation with an absence of OT and RV and reduced atria formation[16]. *Isl1* transcriptional activity greatly diminishes as soon as the cardiac progenitor cells integrate into the primitive embryonic heart, suggesting that this TF is essential for proliferation, expansion and migration of these cells though not for cell differentiation[42].

Nkx2-5 has been attributed a role for later stages of cardiac morphogenesis during development of atrial, ventricular and conduction cell lineages[43]. *Nkx2-5* is expressed in both the primary and secondary heart fields and plays a pivotal role in the early steps of mammalian cardiogenesis. *Nkx2-5* KO in early development leads to embryonic lethality and defects in cardiac looping morphogenesis can be observed, which are consistent with the role of *Nkx2-5* in the early stages of cardiogenesis[44].

Gata4 (GATA biding protein 4) is expressed in the cardiac cell lineage and plays an important role throughout cardiac development. *Gata4* promotes cardiogenesis through several mechanisms and interacts with other known cardiac players to activate and stimulate the secretion of cardiac inducing factors[45]. Upon deletion of *Gata4*, cardiac malformations and embryo lethality can be observed, as described by Kuo *et al*., 1997[46]. Interestingly, overexpression of *Gata4* in ESCs directs cells towards endoderm rather than cardiac mesoderm. However, endoderm cells that overexpress *Gata4*, produce paracrine factors that stimulate adjacent cells to differentiate into cardiac mesoderm[45].

Tbx5 has been primarily known for its role in cardiac and forelimb development since mutations have been associated with Holt-Aram syndrome and defects on cardiac septation[47]. Mice embryos that lack *Tbx5* die in utero around E10.5 with abnormal heart tube and defective LV[47]. In mice, *Tbx5* is expressed around E8.0 throughout the cardiac crescent and around E8.5 it is expressed in cells that will give rise to the atria. At E9.0 it becomes expressed in the LV[48]. Furthermore, during late cardiac patterning, *Tbx5* is required for the proper formation of the cardiac conduction system and maintenance of mature cardiomyocyte functions[47].

Besides TFs, other factors are involved in cardiac gene regulation. In the next part, I will introduce some of the non-coding RNAs that have been described in literature to be implicated in gene regulation in cardiac cells and heart development.

In the last decades, small non-coding RNAs, known as microRNAs (miRNAs) have been recognised as key regulators of organ development in several organisms, such as, *Caenorhabditis elegans*, *Drosophila melanogaster*, zebrafish, mouse and humans[49-51]. miRNAs (21–25 nucleotides in length) commonly regulate gene expression at post-transcriptional level through imperfect base pairing to target mRNAs, leading to translational inhibition or mRNA degradation[52].

Production of mature miRNAs occurs in four steps: (1) transcription of a primary miRNA (pri-miRNA) via RNA polymerase II; (2) processing in the nucleus of the pri-miRNA by Drosha (a RNase III type endonuclease) and the double-stranded RNA binding protein DGCR8 (DiGeorge syndrome critical region gene 8) complex to produce a hairpin precursor miRNA (pre-miRNA), consisting of ~70nt; (3) export of the pre-miRNA to the cytosol via the protein exportin-5; and (4) cleavage by Dicer (a RNAse III type endonuclease) to produce ~22-bp double-stranded miRNA (Figure 1.1.7).

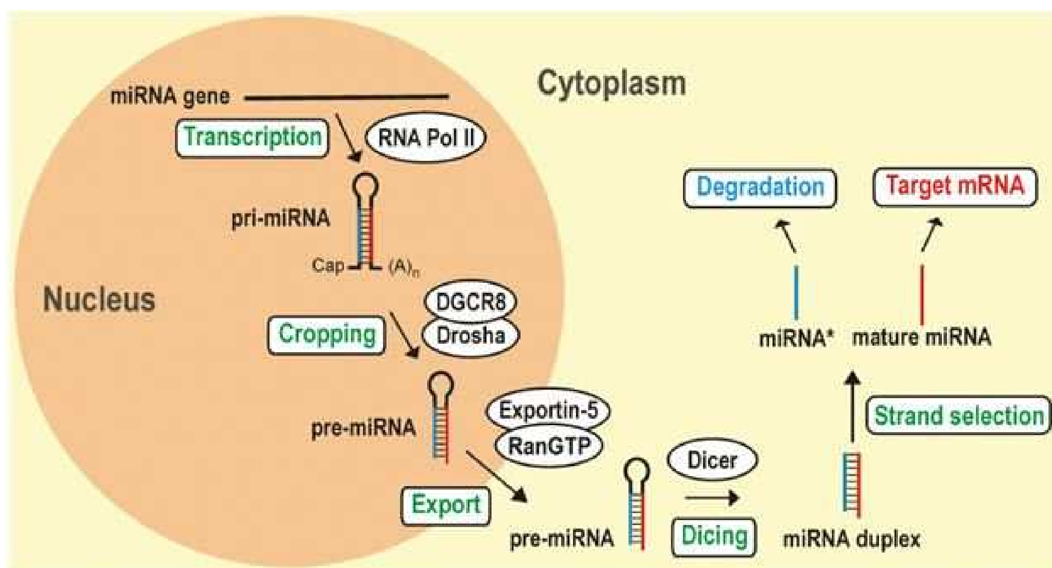


Figure 1.1.7 Multistep process for the formation of mature miRNAs. (i) miRNA genes are primarily transcribed by RNA polymerase II (RNA Pol II) in the nucleus; (ii) the resulting primary transcript (pri-miRNA) is cleaved by Drosha and DGCR8 and the resulting pre-miRNA is transported to the cytoplasm; (iii) the pre-miRNA undergoes cleavage by Dicer to produce a duplex molecule; (iv) the duplex molecule is separated. One strand is selected as mature miRNA and is directed to a target-specific mRNA[53] (Adapted from Schneider *et al.,* 2012).

Subsequently, one strand of the miRNA is typically degraded, whereas the other strand is integrated into the RNA-induced silencing complex (RISC). The RISC complex targets complementary sequences in the mRNAs that are mainly located in the 3′ UTR but can also be found in the 5′ UTR and coding region, resulting in degradation of the mRNA or inhibition of its translation.

miRNAs acting in the heart

The human heart has limited capacity for completely regenerating lost or damaged cardiomyocytes. Instead, heart recovery is characterised by extensive cardiac remodelling by fibroblasts, resulting in posterior deterioration of overall cardiac function. Besides the well-known transcription factors related to heart development and function, miRNAs have been showing promising results in the heart development and regeneration area. Several researchers have recently utilised miRNAs to promote cardiomyocytes derivation through different experimental approaches (Figure 1.1.8). For instance, several miRNAs can act in combination to promote direct conversion of cardiac fibroblasts into cardiomyocytes[54]. In addition, there are miRNAs that can guide ESC differentiation towards cardiomyocyte-like cells[55]. Furthermore, there are also other studies that focused on the capability of miRNAs enabling cardiomyocytes to re-enter the cell cycle to replicate and generate new muscle cells or even about targeting CPC niches and make them proliferate[54-57].

One well-studied heart-related microRNA is miR-1. This miRNA has been described as being co-expressed with miR-133 in cardiac and skeletal muscle, and both miRNA are transcriptionally regulated by myogenic differentiation factors[50,58,59]. So far, a number of studies have suggested crucial roles for these miRNAs in heart development and function. For instance, overexpression of miR-1 in a developing mouse heart inhibits the proliferation of CPCs and forces them towards cardiomyocytes differentiation, causing developmental arrest at E13.5 as a result of thinning of the ventricle walls and heart failure[52]. Similarly, miR-133 has been shown to participate in the proliferation, differentiation, and electrical conduction of cardiac cells[60].

MiR-499 has been associated with the expression of mature cardiomyocytes proteins/markers such as *MHC6, MHC7*, *MLC2*, and *TNNT2*[61]. Furthermore, it has been reported that miR-499 is co-expressed with miR-1 in ESCs-derived cardiomyocytes and controls the electrical/conduction system through targeting transcripts encoding for the ion channels[62].

MiR-208a is an interesting miRNA since it is encoded by an intron of the α-cardiac muscle myosin heavy chain gene (*Myh6*) and it is specifically expressed in the heart. Furthermore, this miRNA is sufficient by itself to induce cardiac remodelling and regulate the expression of hypertrophy pathway components, such as, the up regulation of the βMHC protein. On the other hand, deletion of miR-208a was sufficient to induce aberrant cardiac conduction showing its importance for the development of the conduction system[63].

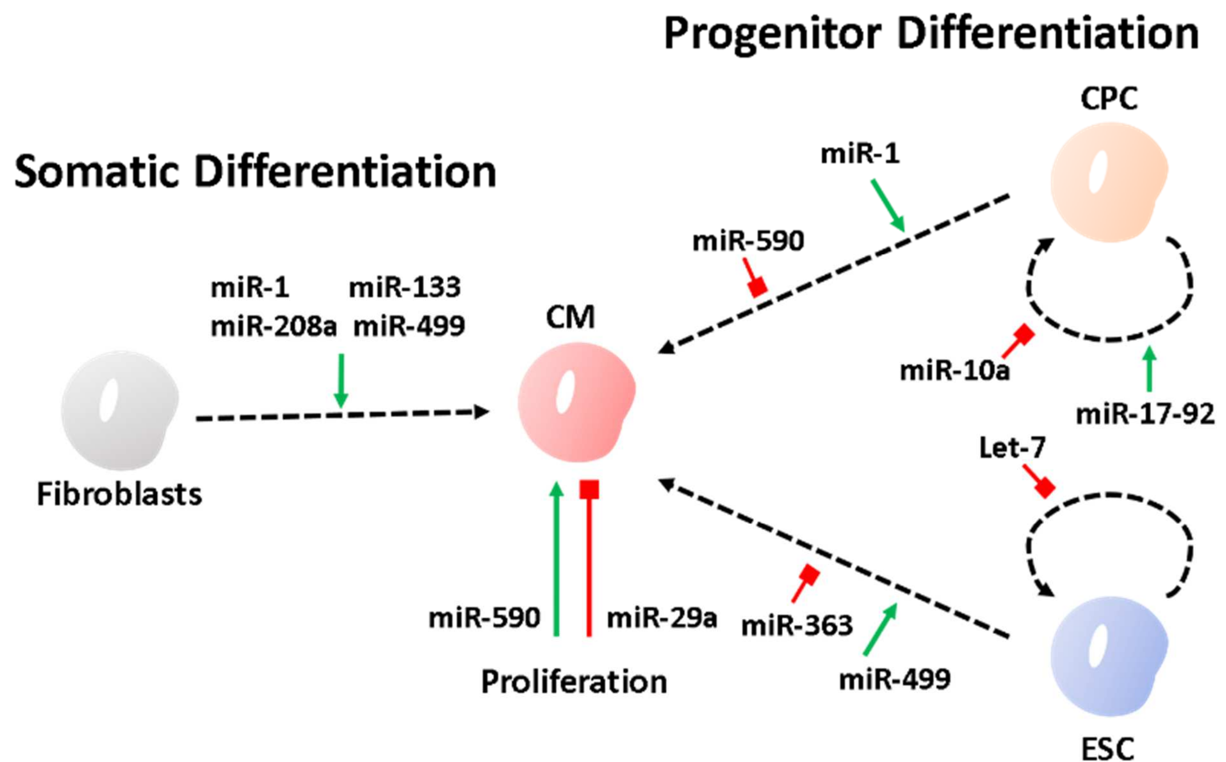


Figure 1.1.8 miRNAs in stem cells and heart development. miRNAs promote the generation of cardiomyocytes through different mechanisms. miRNAs promote CPC and ESCS cardiac differentiation. miRNAs can also promote or inhibit cardiomyocyte proliferation. Adapted from Hodgkinson *et al.,* 2016[55].

It is worth mentioning that the combined induction of miR-1, miR-133, miR-499 and miR-208a ("miR combo") demonstrated high efficiency for reprogramming cardiac fibroblasts into cardiomyocytes-like cells[55-57,64] (Figure 1.1.8 – somatic differentiation). Moreover, the induced cardiomyocyte-like cells, by the "miR combo", presented functional properties which are

characteristic of cardiomyocytes, such as L-type channel expression, spontaneous calcium oscillations and contractility, although their physiological properties appeared to be less mature than those of neonatal cardiac myocytes[57].

Another important set of miRNAs that helps to understand heart development is the miR-17-92 cluster. Initially this cluster has been reported as an oncogene and classified as oncomir1[65]. However, it was later found that these miRNAs were important for normal heart development, since mouse miR-17-92 cluster mutants died postnatally, displaying multiple organ malformations[66]. Since then, it has been reported that miR-17-92 cluster overexpression induces proliferation of neonatal and adult cardiomyocytes with the same intensity. Furthermore, the overexpression in adult cardiomyocytes can have a protective role against myocardial infarction (MI) associated injuries, by reducing the expression of PTEN, which is a cell cycle regulator. Loss of function of miR-17-92 cluster leads to the aberrant differentiation of CPCs into normal cardiomyocytes by repressing the function of cardiac progenitor *Isl1* factor during embryonic cardiac development[67].

Another miRNA regulating CPCs is miR-10a, as its overexpression or repression can inhibit or promote the proliferation of CPCs without interfering with the normal differentiation towards functional cardiomyocytes. For instance, Gata6 which is an important gene for proper heart development[68], is targeted by miR-10a that greatly reduces Gata6 expression, affecting the proliferation of CPCs[69]. This result was confirmed by Gata6 overexpression, which attenuated the effects of miR-10a on CPCs proliferation[69].

Finally, Let-7 is an important miRNA family that was reported to be capable of inhibiting ESCs proliferation and lead towards cardiomyocyte-like cells maturation. Furthermore, members of the Let-7 family are capable of modulating CM metabolism, cell size and contractibility force, proving to be one of the most well established factors to promote cell maturity of stem cells towards cardiomyocytes-like cells[70,71]. Inhibition of Let-7 family factors led to acute cardiomyocyte apoptosis and induced cardiac hypertrophic phenotype post myocardial infarction. On the other hand, Let-7 family miRNA overexpression reduced the hypertrophic phenotypes and improved cardiac function[72].

1.2 ESCs as model for heart study

Stem cell research holds a great potential for understanding the fundamental mechanisms of human development as well as for prevention or treatment of several cardiovascular pathogenesis including myocardial infarction (MI). However, obtaining and using hESCs constitutes a major ethical dilemma which appears to be a major factor that hinders the application of hESCs in clinical therapies[73,74]. Furthermore, there are also safety issues when considering hESCs-based therapies. As hESCs have special specific features, such as high proliferation rate and unlimited self-renewal, hESCs-based therapies have been associated with tumorigenicity[73,75].

Nevertheless, ESCs represents a homogeneous cell population with the potential to generate any of the hundreds of distinct cell types present in an organism. This enables the development of experimental approaches to direct lineage differentiation towards cardiomyocytes by using for example forced expression of regulatory genes, defined signalling proteins, or small molecules. From a translational point of view, ESCs allows the generation of potentially unlimited numbers of progenitors at various defined stages of developmental potency for testing cellular therapies to ameliorate cardiac related problems.

Transfer of knowledge obtained from developmental studies to *in vitro* ESC systems remains however challenging. For instance, the generation of cardiomyocytes by undirected differentiation *in vitro* from ESCs is inefficient. Thus, several experimental methods have been developed to generate cardiomyocytes from ESCs, including monolayers of cardiomyocytes using defined culture media[76]. Alternatively, conditions and procedures that more closely reflect *in vivo* development have been introduced. For instance, formation of embryoid bodies (EBs) can mimic normal gastrulation and allow the specification of the primary germ layers, including mesoderm which is responsible for deriving cardiomyocytes[32,77,78]. This EB method has been proven useful to study development of cardiac progenitors and signalling pathways that can influence cardiogenesis process. According to Evans and colleagues, although being an heterogeneous cell population during differentiation, the ESC-derived EBs development remains synchronised and mostly recreates the temporal transitions that lead to the specification of defined progenitor fates[79].

Therefore, we can expect the concordant expression of genetic markers in the population of EBs, which eases the purification of progenitor subsets that appear at defined developmental stages (Figure 1.2.1).

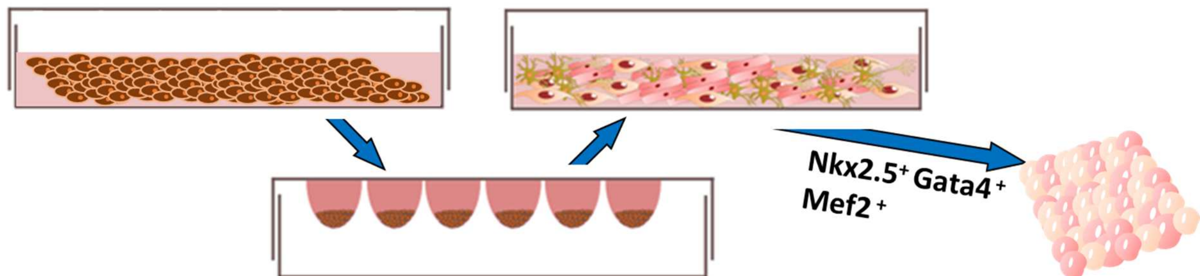


Figure 1.2.1 The ESCs-EBs system that provides a model for commitment of progenitors to specific fates and subsequent identification of multipotent progenitors for cardio-vascular derivatives (Adapted from Evans *et al*., 2008[79]).

This methodology may be essential in a practical sense for the isolation of sufficient quantities of such early progenitors for experimental testing and validation[79].

Finally, CPCs are another alternative option as a model system to study cardiogenesis. As multipotent cells, they have the capability of self-renewal for multiple passages, but differentiation is restricted towards cardiovascular cells only. Indeed, CPCs begin to spontaneously differentiate into cardiomyocytes and smooth muscle cells in mouse hearts after MI injury, and can improve heart function[80]. Nonetheless, production and maintenance of large quantities of CPCs is a challenging task, since culture requirements are not well defined and the necessary steps to maintain CPCs are poorly understood. Therefore, efforts have been undertaken to uncover the necessary conditions to maintain inducible expandable CPCs with potential to differentiate into cardiac related lineages[81].

1.3 Stem cells as a therapeutic approach for heart regenerative medicine

Regenerative medicine approaches are based on the study of how organisms replace lost or damaged tissue with new functional tissue. Advances in understanding of stem cell biology can lay a foundation for the potential repair of damaged tissue that was previously uncapable of regenerating or even repair with functional tissue. As aforementioned, ESCs may have many applications due their capability of unlimited self-renewal and high proliferation, but these cells are associated with ethical issues. Thus, alternative stem cell populations might be of advantage for the translation of approaches of stem cell biology to clinical application.

Promising candidate cell types for cardiac regenerative medicine include adult stem cells (ASC). These type of stem cells have the capability of self-renewal but are generally restricted in their cell potency i.e. they differentiate only to a very limited range of cell types. Therefore, there was excitement to use adult cardiac stem cells, as this concept could lead to autologous cell transplantation. However, in contrast to bone marrow ASC, cardiac stem cell are harder to identify and typically are represented in heterogeneous population that have low capability of differentiation into cardiomyocyte like cells[82].

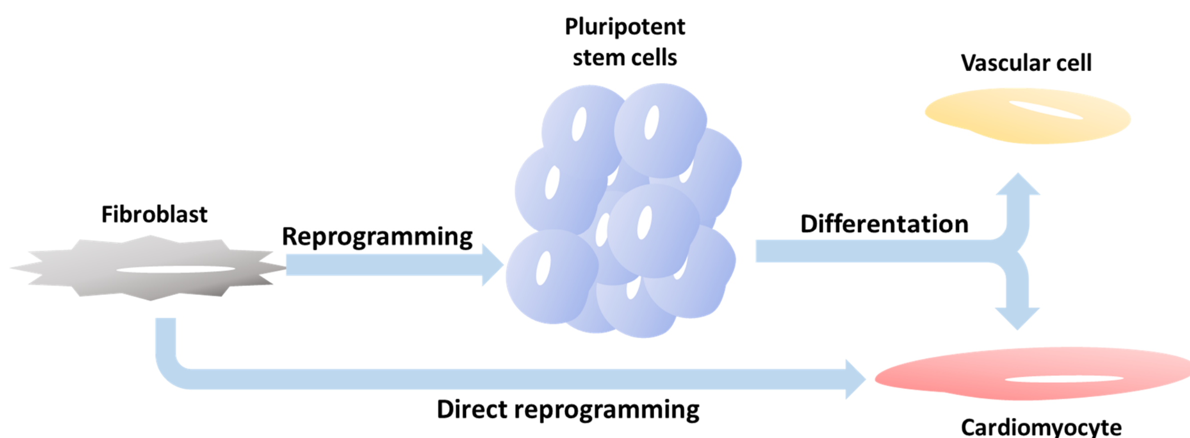


Figure 1.3.1 Stem cell biology and reprogramming. Increasing knowledge about ESCs has enable the capability of generating any cell type and with this understanding it was possible to uncover that cells can be reprogrammed into other cells, creating a wide variety of possibilities in regenerative medicine[83] (adapted from Lee *et al.*, 2016).

With the discovery of the processes which allow the generation of induced pluripotent stem cells (iPSCs) a new paradigm in stem cell biology was established. These cells displayed ESCs properties and were generated from terminally differentiated cells by using reprogramming factors[84].

This discovery made possible the generation of pluripotent stem cells able to generate all cell types of the adult organism, but without the use of ESCs (Figure 1.3.1). Hence, reprogramming technology has enabled a possible route for generation of cardiomyocytes via iPSCs[85-87]. The technique also eased the ethical consequences of using stem cells, since iPSCs generation does not require cells originating from embryos (Figure 1.3.2). Even more, the related technique, called direct cell reconversion (i.e. changing a differentiated cell type into another differentiated cell with different functions, without passing by a pluripotent state – Figure 1.3.1 – lower arrow) reduces risk of potential generation of teratomas.

Indeed, *in vivo* cardiac fibroblast resultant from heart injury can be reprogrammed into functional cardiomyocytes without reverting to a pluripotent state[88]. Another alternative could be the stimulation of endogenous cardiac regenerative mechanisms through gene or protein delivery, resulting in the formation of new cardiomyocytes without exogenous cell delivery[89].

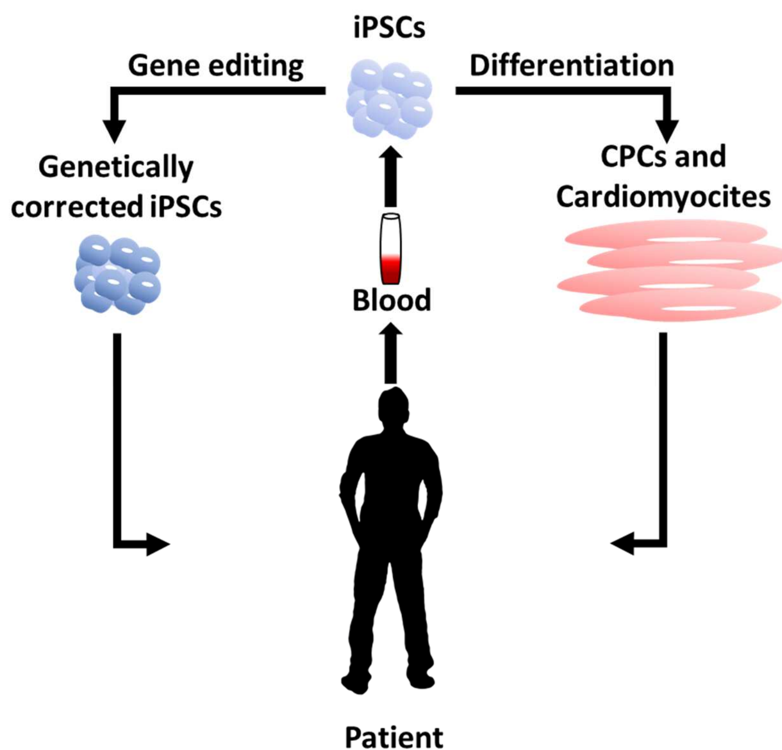


Figure 1.3.2 Simplified applications of iPSC for therapy. Somatic cells from patients can be reprogrammed into iPSC. iPSC genome can be corrected through gene editing for correcting a specific disease and/or be used as a regenerative and recuperative therapy.

Although the outlined advances indicate the great potential of stem cell-based products for cardiac regenerative medicine, a better understanding of the complex underlying molecular mechanisms will be required for the safe transition from bench to bedside. Some of the necessary knowledge can be gained through the application of transcriptomic technologies.

1.4 Technologies in molecular analysis and diagnostics

Technologies capturing gene expression analysis on a genome level have been under rapidly development in the last few decades. Two of the most eminent technologies are currently microarray and next-generation sequencing (NGS) technologies.

Microarrays are a well-established tool for whole-transcriptome analysis[90]. They are based on the hybridisation of labelled transcripts to complementary sequence (i.e. probes) fixed on a solid surface. mRNA levels can be monitored by conversion to labelled cDNA or cRNA and subsequently compared against a reference (Figure 1.4.1). For instance, mRNA levels in scaring heart tissue or regenerative heart tissue can be compared with transcript levels in normal heart tissue. For such experiments, microarrays are well suited, as they enable measurements of thousands of the expression of genes simultaneously, thanks to the high density placing of probes on a single array.

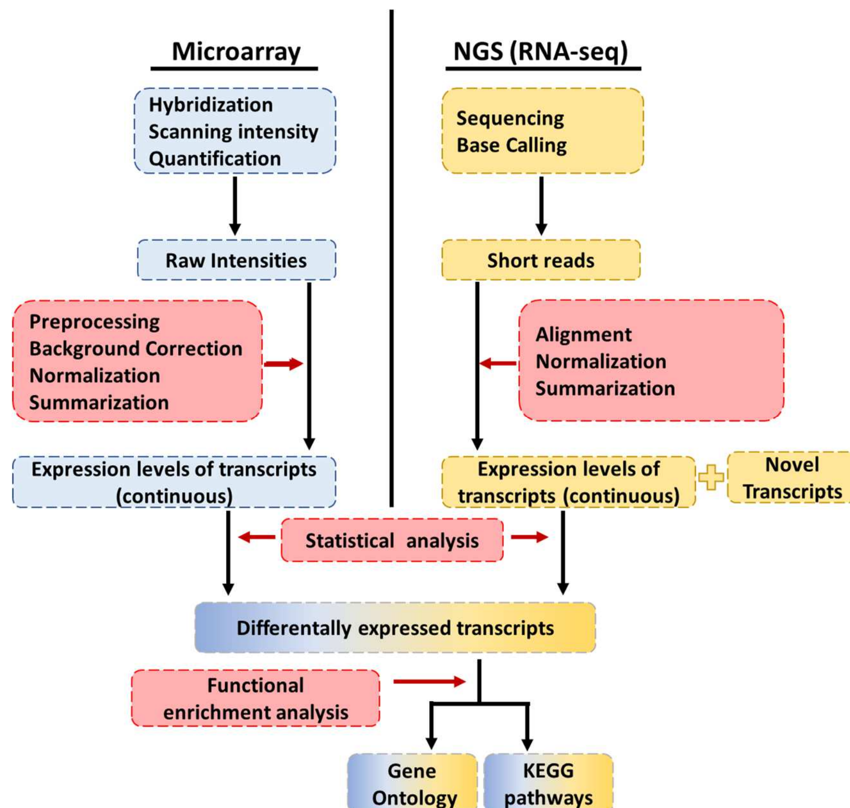
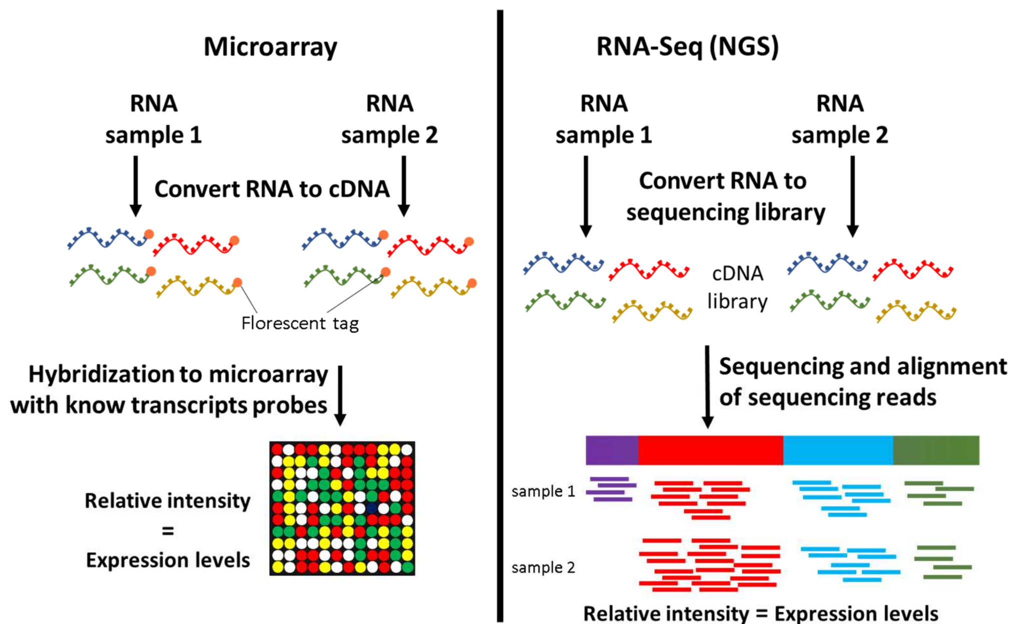


Figure 1.4.1 Simplified overview of workflow for microarrays and NGS technology data processing. Data and information obtained by using microarray technology is represented in light blue; while data obtained by using NGS (RNA-seq) is indicated by yellow. Steps for data processing and analysis are coloured in red. Common information obtained in microarray and RNA-seq experiments is shown in blue to yellow colour gradient.

Different transcript isoforms can additionally be measured through the use of multiple probes per gene. Furthermore, allele-specific probes can be used to detect mutations such as single

nucleotide polymorphisms (SNPs). As microarrays have been applied for over two decades, experimental procedures and data analysis are nowadays well developed, and researchers tend to be familiar with the methods and their outcomes. The relative ease of their use has help maintain their place as an important tool for monitoring expression changes in well annotated genomes. This technology is expected to be still applied in the near future due to its relatively low cost combined with its low data handling burden and the well-established pre-processing, quality control and analysis methods for microarray data.

Nonetheless, microarrays have their known limitations. Generally, microarray technology depends of probe selection and has been generally restricted to species with sequenced genomes unless the researcher developed his own library[91]. For organisms, which have not been sequenced so far, microarrays are of limited use.



Advantages	Disadvantages	Advantages	Disadvantages
Well-stablished data analysis protocols	*A priori* knowledge of the genome	*A priori* knowledge of the genome not required	Data analysis methodology are more complex
Low computational burden	Moderate sensitivity and dinamyc range	High sensitivity and dynamic range	High computational burden
Low data storage	Low alternative splicing information	Novel transcript sequences identified	High data storage
Low cost	Specific data would require specific microarrays format	Alternative splicing information	Expensive (but it is getting cheaper)
Can be used for disease profiling		Unlimited sample comparison	

Figure 1.4.2 General overview about the advantages and disadvantages of microarrays and RNA-seq (NGS).

Next generation sequencing (NGS) refers to a range of techniques that analyse sections of the genome to a single-nucleotide resolution (Figure 1.4.1). NGS is capable of performing efficient whole-genome, whole-exome, transcriptome, and targeted sequencing. NGS measure the transcript abundance through massive parallel sequencing. NGS single nucleotide resolution, enables the possibility to detect the smallest of mutations without having to necessarily know the mutation beforehand. Furthermore, NGS does not require the design of probes and the transcriptome can be *de novo* assembled if no reference genome is available. NGS experiments however generate huge volumes of data, which currently present challenges for data management, storage, and analysis (Figure 1.4.2).

The use of NGS is fast growing in oncology, complex diseases, and infectious diseases. With the produced data, researchers are able to inspect particular details such small mutations (SNPs or inserts and deletions "*indels*") or larger abnormalities in the genome. NGS is particularly attractive for cancer studies, as many mutations need to be captured. Eventually, the application of NGS in translational research will depend on sequencing costs as well as developments in the regulatory landscape. While NGS is likely to dominate the field of transcriptomics, I have used microarray studies for my research work because of the existing data available for heart development.

1.4.1 Microarrays in heart study

Microarrays have been widely used for heart studies regarding its development, regeneration, reprograming, responses to injury and cardiovascular associated diseases[92-96]. The ability to access many expression data sets provides a unique opportunity of elucidating the gene regulatory mechanisms. The gained knowledge can help to increase the understanding of heart development and might provide new insights into cardiovascular diseases[93].

An important microarray experiment in the area of heart development was performed by Domian and colleagues[92], in which they used a two-coloured fluorescent reporter construct to identify and separate mouse FHF and SHF progenitors from developing mouse embryos. Although progenitors have been previously shown to give risen to the different structures of the heart, the exact mechanisms were poorly understood.

This microarray experiment revealed distinct molecular signatures of progenitor populations and gave indications of combining tissue-engineering with stem cell biology for the generation of functional ventricular tissue[92].

Another area, in which microarrays can play a pivotal role, is in study of heart diseases. Cardiac hypertrophy, fibrotic remodelling and heart failure progression represent key areas for genomic discovery using microarrays. Little was known about the gene expression changes in cardiomyocytes on a genome level during cardiac diseases. Microarray analysis helped to identify multiple genes involved in the response to cardiac injuries and to uncover the associated changes in expression[97]. For instance, one key event in heart failure is the programmed cardiomyocyte cell death which has been associated with chronic *Akt* activation in the heart. Microarray analysis indicated that the upregulation of the Nix protein could be responsible for cardiomyocyte apoptosis and for myocardial hypertrophy[98]. Microarrays were also used to capture molecular profiles for different cardiomyopathies using heart tissue of patients and to compare them with control healthy tissue to identify up- and downregulated genes in affected individuals[99].

Another use for microarrays can be the profiling of genome-wide expression of somatic cells when reprogrammed towards a cardiomyocyte-like state. Through the utilisation of four well-described cardiac-related transcription factors (*Gata4*, *Hand2 Mef2c* and *Tbx5*), for instance, Song[96] and colleagues could reprogram adult mouse tail-tip and cardiac fibroblasts into beating cardiac-like myocytes *in vitro*. Furthermore, the overexpression of these factors in mice that suffered myocardial infarction led to conversion of non-cardiomyocytes cells into cardiac-like myocytes and improved cardiac function while reducing fibrotic ventricular remodelling[96]. Through the use of microarrays, it was possible to characterise the molecular processes that lead to this outcome. For instance, gene expression patterns of cardiomyocyte-like cells displayed an activation of a broad range of cardiac genes in fibroblasts transduced with *Gata4*, *Hand2*, *Mef2c* and *Tbx5*[96].

Although many genes and molecular regulators are well-described and associated with heart development, there is still a great amount of information that needs to be uncovered in order to fully understand how normal heart development occurs.

For this work, I designed and employed several different methodological approaches to discover and reveal novel heart related candidates by exploring the potentialities of microarray technology.

(1) Data collection and integration: One of the focus of this work was to build a comprehensive database of transcriptomic measurements of heart development, regeneration and reprogramming *in vivo* and *in vitro* in mouse and *in vitro* in human. All microarray data contained in this database was retrieved in raw format, filtered, and manually curated to reduce the risk of artifacts or technical errors. The analysis of microarray data from multiple experimental condition enables gene expression profiling and comparative analysis, such as, differentiation of stem cells towards cardiomyocytes and reprogramming of somatic cells towards cardiomyocytes.
(2) Expression profiling *in vitro* and *in vivo*: Another approach taken was the analysis of dual parallel genome wide expression profiling between mRNAs and miRNAs *in vivo* and *in vitro*. The mRNA and miRNA transcriptome profiling not only provided a comprehensive characterisation of murine heart development, but also promises to be an excellent basis to establish the regulatory molecular networks involved in heart development.
(3) Prioritisation of regulatory interactions: I led the development of a comprehensive database that contains all relevant expression values obtained in the *in vivo* microarray experiment and a vast number of mRNA-miRNAs interactions that have been manually integrated from 5 different databases. The microarray data can be used to prioritise regulatory interactions.

The work revealed novel heart related genes and miRNAs that have not been described to be associated with heart so far. The identified candidates constitute attractive targets for experimental validation and generate new biological questions that could lead to ground-breaking discoveries in study of heart development and regeneration.

Chapter II

Bioinformatic Methods

This chapter describes the bioinformatic tools and methods for data processing that were applied to multiple microarray data sets across different chapters. Bioinformatic approaches, which were applied only to a specific data set, are described in more detail in the relevant chapter. Most of the computational analysis were performed in R, a programming environment that is specifically suitable for this work, as many add-on R packages for microarray data analysis exist on the Bioconductor platform[100].

2.1 Microarray data processing

Microarray measurements are based on the hybridisation of dye-labelled targets to complementary probes. Excitation of the dye by a laser provides a fluorescence signal which constitutes a proxy for the abundance of the labelled target. Before changes in signal intensities can be derived, microarray data needs to be pre-processed to reduce technical variability. Typically, this means that the data undergo background correction, log transformation, normalisation and potentially summarisation. To perform these steps in R, Bioconductor packages *affy*[101] or *oligo*[102] can be applied depending on the microarray version.

Affymetrix microarray (also called GeneChips or simply chips) include both perfect match (PM) probes i.e. probes that contain sequences to match exactly the sequence of interest, and mismatch probes i.e. probes that do not match exactly the sequence of interest. The tools applied for this work use only PM probes. The data are presented as *.CEL* files that contain probe intensities derived from the scan of the GeneChip. *CEL* data are of very simple structure storing all probe intensities from the chip. Information about the identity, location and the sequence of each probe are stored in a *CDF* file object.

As scanned microarray images typically contain background noise, a background correction is applied to correct for non-specific signals based on a global model for the distribution of probe intensities in such way that all background corrected intensities remain positive[103].

The data is conventionally log-transformed (with base-2) in order to balance the data distribution. The log transformation typically renders the data less skewed and more normally distributed. The transformation also provides a similar spread of values for up- and downregulated genes across the intensity scale[103]. The Log_2 of the background corrected intensity is calculated for each probe.

Normalisation procedures attempt to correct for systematic differences between chips, so data from different microarrays can be directly compared[101]. Normalisation is required because of variation introduced during the microarray experiment. For instance, the quantity of RNA hybridised to a microarray can vary slightly from microarray to microarray and thus introduces technical variability. A popular normalisation procedure is quantile normalisation, which is a mathematic approach for rendering two or more distributions identical in statistical properties by ranking each distribution and calculating the arithmetic mean for the ranked list of values. Then, the highest value in a distribution is set to the mean of the highest values, the second highest value set to the mean of the second highest values, and so on. These transformed values from different arrays have the very same overall distribution while the values for a particular gene can be still vary across the different arrays[104].

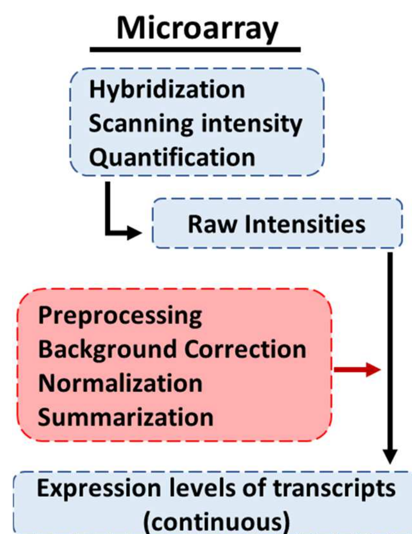


Figure 2.1.1 Pre-processing for microarray intensity signals to obtain the expression levels of the transcripts. RMA normalisation approach pre-processes microarray data so that luminous raw intensities can be interpreted as transcript abundance.

Finally, summarisation is required to average over multiple probe sets targeting the same transcript. Simultaneously, it can be used to correct for variations within probe sets. This process balances the behaviours of the probes between the arrays and combines the normalised data values of probes from a probe set into a single value for the whole probe set[103]. A standard method for normalisation is robust multi-array average (RMA) which uses the PM probes. After RMA is performed, expression levels of the transcripts are obtained.

2.2 Microarray data quality analysis

To verify quality of data used for subsequent analysis, different quality assessments were applied to each individual study included in this work:

(i) Evaluation of the Principal Component Analysis (PCA), which is a statistical procedure that converts a set of observations with possibly correlated variables into a set of values of linearly uncorrelated variables. The transformation is defined in such a way that the first principal component has the largest possible variance, and each succeeding component has the maximum possible variance while being orthogonal to the previous ones. This procedure was used to detect and visualise the variability between samples;

(ii) Density plots of expression values to verify whether the experimental conditions of a determined study resulted in similar distributions;

(iii) Generation of clustering dendrograms to assess whether the replicates of experimental conditions are placed in the same cluster or present outliers which can subsequently be removed before performing gene expression analysis;

(iv) Inspection of expression of established marker genes (e.g. for pluripotency, mesoderm specification or cardiac cell development) to confirm whether they follow the expression expected based on previous literature reports.

2.3 Microarray differential expression analysis

In *chapter IV* and *VI*, the mRNA and miRNA differential expression analysis were performed by using the R/Bioconductor package *limma*[105] and multivariant empirical Bayes statistics[106]. *Limma* is a Bioconductor package for differential expression analysis for microarray experiments. It is designed to analyse complex comparisons between many RNA targets simultaneously that are present in two or more experimental conditions in order to identify differential expression between the experimental conditions[107]. *Limma* is based on a moderated t-statistics following an empirical Bayes approach to estimate the variance of a gene's expression. Using additionally the observed variance of expression of all genes, it provides a more robust estimate of the significance of differential expression especially for small number of replicate samples. Such approach aids in controlling the false discovery rate (FDR) and thus assists in extracting reliable information from microarray data sets[108,109]. FDR is defined as the expected proportion of false positives among all genes for which the null hypothesis (of not being differentially expressed) was rejected. It is a statistical measure to correct for multiple hypothesis testing and popular in the analysis of omics experiments such as microarray experiments. It can be derived using the Benjamin-Hochberg procedure, which is less conservative as the traditional Bonferroni method for correction of significance[110]. In general, multiple testing correction is needed to adjust the statistical confidence measures based on the number of tests performed[111].

To detect expression changes, it was necessary to compare the experimental conditions against a reference/control condition. Therefore, for each experimental study described in *chapter IV* and *VI*, I have selected a suitable experimental condition as reference/control based on the experimental design and research question. For instance, in chapter IV the reference sample was the sample for "young adult heart" to compare expression levels during development with the adult stage. Alternatively, I defined developmental stage "E19.5" as reference to obtain a finer assessment of the dynamics of gene expression changes during embryonic heart development. The different designs are further discussed in *chapter IV* and *VI.*

2.4 Clustering of differentially expressed genes and miRNAs

To cluster the obtained differentially expressed genes (DEGs) and differentially expressed miRNAs (DEmiRs) (*Chapter IV* and *VI*), the fuzzy c-means algorithm was used as implemented in the R/Bioconductor package Mfuzz[112]. Fuzzy c-means clustering has two features which make it particularly suitable for gene expression analysis: (1) objects such as genes can be assigned to multiple clusters and (2) the degree of association of an object with a cluster is represented as a graded membership value from 0 to 1. The latter indicates whether the object is well represented by the cluster centroid or can be rather considered as an outlier in the cluster. This contrasts more traditional hard clustering approaches such as k-means, which assigns an object to exactly one cluster without indication about the fidelity of clustering assignment. It also increases the robustness against noise, which is a favourable characteristic for transcriptomic data[113].

For clustering, expression profiles of DEGs and DEmiRs were standardised, i.e., the mean value was set to zero and the standard deviation (SD) was scaled to one for each DEG and DEmiR. The default value of 2 for the fuzzification parameter m, which determines the spread of genes across clusters, was kept. The number of clusters was selected based on the minimum distance between cluster centroids (as implemented in the Dmin function of the Mfuzz package), as well as on the inspection of the visualised cluster patterns. To facilitate interpretation, the cluster index was sorted based on the overall cluster profile, starting with downregulated clusters.

2.5 Gene ontology enrichment analysis of differentially expressed genes and miRNAs

Gene ontology (GO) enrichment analysis for these clusters was carried out in R applying an *in-house* built script that uses the Bioconductor packages org.Mm.eg.db[114] and GOstats[115]. For testing the overrepresentation (enrichment) of genes associated with a GO category, the hypergeometric test was used. This test is implemented as the function hyperGtest in the GOstats package[115]. The function allows to use a set of different parameters. These parameters comprise information such as (i) the specific gene list to be tested (i.e. genes of biological interest); (ii) the "gene universe" defined by the set of possible genes (i.e. all genes present in the annotation package for the microarray platform); (iii) the Gene Ontology (GO) that will be queried (i.e. Biological Processes) and (iv) the gene annotation data package. These parameters are used to infer the overrepresentation of GO associations among the genes of biological interest when compared to randomly retrieved gene sets from the "gene universe"[115]. Additional inputs can be added, such as test direction, conditioning and p-value cut-off. Test direction accounts for over- or underrepresented GO terms.

One challenge in the use of GO for enrichment analysis is the detection of highly overlapping categories due to the hierarchical structure of GO. A specific GO category is generally part of more general GO categories and all genes found in the specific category are also included in the more general categories. In practical terms, a significant enrichment in specific categories can result in the significance of a large number of more general categories simply because of the sharing of genes. To reduce the number of similar categories detected as enriched, a hypergeometric conditioned test on the GO tree structure was applied[115]. Conditioning is responsible for the narrowing of the GO tree structure, by reducing the number similar generic GO categories. For the (1) hypergeometric test it was a cut-off p-value ≤ 0.05 testing if a gene would be enriched in a determined GO category and; (2) it was set a p-value cut off of ≤ 0.01 to define the GO categories tree structure. The latter test estimates for each category whether there exists evidence for enrichment beyond that provided by the significance of the more specific category that it is related to.

Chapter III

HeartEXpress

A web database for exploration and visualisation of integrated cardiac gene expression data

3.1 Introduction

Heart failure and cardiac congenital malformations are main causes of death worldwide. Although surgical interventions have provided valuable options for treatment, current therapies in cardiovascular medicine only delay disease progression. One of the main reasons for this shortcoming is the limited regenerative capacity of the adult human heart. Injuries that occur in adult heart muscle have lasting consequences for its normal function, as damaged or dead heart muscle cells are not or only to a low rate replaced by newly generated functional cells.

To better understand the plasticity of cardiac cells, a large number of gene expression studies were performed using microarray technology. In the context of the study of cardiogenesis, microarray based gene expression profiling has been applied to the study of *in vitro* stem cell differentiation to cardiomyocytes[92,116,117] direct reprograming of fibroblasts into cardiomyocytes[86-88,118] and heart development *in vivo* by various groups[119,120].

In principle, these microarray experiments constitute a powerful basis for the study of gene regulation during cardiogenesis. However, there is a major limitation in their usage for the broader research community. Although most of the microarray data sets are publicly available, they are stored in different repositories restricting the study of expression patterns across different data sets. Moreover, several microarray data sets are only available in raw format requiring further processing before their exploration. Caution also needs to be taken in interpretation of the transcriptomic data. It is well known that biological as well as technical variability can compromise expression measurements and that expression profiling studies might lead to divergent results[121]. Finally, a more comprehensive picture of the regulation of gene expression can be obtained when expression changes can be inspected across a broad range of experimental conditions.

For all these reasons, it is important to carry out comparative analyses of published datasets to assess the reliability and reproducibility of the observed expression changes. As a dedicated resource for cardiac gene expression data is lacking, I established the HeartEXpress database (http://heartexpress.sysbiolab.eu/).

To enable comprehensive analysis of transcriptomic studies in the context of cardiogenesis, I collected and curated data from a variety of studies to integrate them in HeartEXpress.

This chapter first describes the collection and processing of the data. Then, multiple bioinformatic steps were taken to profile and compare gene expression during differentiation of stem cells into cardiomyocytes, reprogramming of somatic cells *in vitro* both in mouse and in human, and during heart development *in vivo* in mouse.

Finally, the utility of HeartEXpress is illustrated in a case study, with the aim to identify potential key regulators of cardiomyogenesis and heart regeneration, which can provide possible candidates for subsequent validation.

3.2 HeartEXpress assembly

3.2.1 Data collection and integration on HeartEXpress

Publicly available datasets were retrieved from ArrayExpress database[122] or NCBI Gene Expression Omnibus (GEO)[123], being both functional genomics repositories that are continuously updated and maintained. For query, I used the terms “heart expression”, “cardiomyocyte expression”, “heart development”, “cardiac development”, “cardiomyocyte time”, “myocardial development”, “cardiac fibroblast reprogramming”, “cardiac regeneration”, “*in vivo* cardiac reprogramming”. After downloading, each data set was locally stored and information about authors, biological origin, respective experimental description, unified reference system for expression changes, retrieved database and accession number and publication ID (PubMed ID) was curated. An overview of all studies integrated in HeartEXpress together with detailed information can be found in Table S3.2.1. Further information about the data sets can be found in HeartEXpress *help* page.

Data processing was carried out using the R/Bioconductor platform[100,124,125]. Datasets were individually processed and normalised by application of *affy* and *oligo* libraries[101,102]. Both libraries provide a series of statistical procedures for data pre-processing applied prior to downstream analysis, as described in “Chapter 2 - Bioinformatic methods”, sections 2.1 and 2.2. As result, these methods produced an intensity matrix with rows and columns holding the expression values for genes and samples, respectively. The intensity values in the matrix were log_2 transformed to reduce the skewness of the data distribution. The expression matrices from different studies were merged based on common Entrez Gene IDs.

To support identification of co-expression and potential co-regulation, a hierarchical clustering of genes was carried out on the filtered data set using the software Cluster 3.0[126]. This filtering of the data set consisted in the removal of genes with more than 60% of the expression values missing. To facilitate more focused inspections, clustered expression matrices were generated for several sub-sets of the microarrays and genes (Table 3.2.1):

For human gene expression, two sub-sets were generated with microarray measurements for (i) *in vitro* reprogramming into cardiomyocytes[86,127,128] or (ii) *in vitro* differentiation into cardiomyocytes[117,129-132];

For mouse expression, three sub-sets were generated covering (i) *in vitro* reprogramming into cardiomyocytes[85,87,88,96,118,133,134], (ii) *in vitro* differentiation into cardiomyocytes[92,116,117,135,136] or (iii) *in vivo* development[119,120,137,Unpublished Data];

Integrated human and mouse expression data, in which mouse and human data were merge based on the mapping of homolog genes provided NCBI homologene file (https://www.ncbi.nlm.nih.gov/homologene). This integrated data set contains 16445 genes and 131 experimental conditions.

While the human and mouse specific sub-sets encompass expression data for particular experimental protocols, the integrated dataset includes microarray data for all procedures enabling a comparison of human and mouse expression values in a wide range of experimental conditions. The clustered expression matrices can be queried and visualised through a Common Gateway Interface (CGI)-based application, which is a modified version of the GeneXplorer software[138]. HeartEXpress is currently running on an Apache web server under an Ubuntu Linux operating system.

Table 3.2.1 HeartEXpress integrated datasets for human and mouse.

Human

	Authors	Conditions	Accession ID	Publication
In vitro* *Differentiation	Babiarz *et al*., 2012	20	GSE35671	22050602
	Gu *et al*., 2013	1	GSE48257	23982742
	Uosaki *et al*., 2012	6	GSE28191	21876760
	Poon *et al*., 2013	3	GSE50704	24204964
	Sun *et al*., 2012	4	GSE35108	22517884
***In vitro* Reprogramming**	Fu JD *et al*., 2013	9	GSE49192	24319660
	Muraoka *et al*., 2014	4	GSE56874	24920580
	Nam *et al*., 2013	2	GSE43588	23487791

Mouse

	Authors	Conditions	Accession ID	Publication
In vitro* *Differentiation	Domain *et al*., 2009	3	GSE17513	19833966
	Faustino *et al*., 2008	3	GSE6689	18184438
	Gan *et al*., 2014	4	GSE58300	25058891
	Gaspar *et al*., 2012	9	E-TABM-672	22420508
***In vitro* Reprogramming**	Addis *et al*., 2013	6	GSE45274	23591016
	Christoforou *et al*., 2013	3	GSE44401	23704920
	Fu Y *et al*., 2015	2	GSE69924	26292833
	Ieda *et al*., 2011	5	GSE22292	20691899
	Ifkovits *et al*., 2014	6	GSE54022	24586958
	Song *et al*., 2012	3	GSE37057	22660318
	Zhou *et al*., 2012	7	GSE27329	23144723
In vivo	Cattaneo *et al*., 2014	2	GSE44829	25526092
	Harvard data, 2004	12	GSE1479	NA
	Li *et al*., 2014	13	GSE51483	24803680
	Qian *et al*., 2012	2	GSE49192	22522929

3.2.2 Clustering in HeartEXpress

After removal of genes with more than 60% of expression values missing, genes were clustered based on the obtained expression changes for each of the defined experimental sets. For clustering, the software *Cluster 3.0* was used, which is a program for analysing data from microarray experiments[126]. *Hierarchical clustering* was performed with Pearson correlation as similarity metric. For construction of dendrograms, *complete linkage clustering* was chosen in which the distances between two clusters x and y are defined as the maximum of all pair-wise distances between genes contained in x and y. Results from this analysis were indexed in the webtool HeartEXpress, so that the user can inspect the obtained gene clusters.

3.3 HeartEXpress content & structure

Over 120 microarray measurements were analysed and compared in various ways to obtain a comprehensive view of transcriptional changes during cardiomyogenesis under different experimental conditions. Since most of the microarrays were generated using the same platform (i.e. Affymetrix), data treatment tended to be similar, but with minor variations due to differences in experimental design. To make the data easily accessible, HeartEXpress provides a query interface with different options.

3.3.1 HeartEXpress data processing and quality control

One of the main steps in any microarray-based study is to validate and confirm the quality and robustness of the collected data. Quality control and data validation are crucial to exclude artefacts or measurement errors that might derive from the experimental procedure. After performing RMA for summarizing and normalizing gene expression values, PCA was therefore carried out to assess data variability. Other types of inspections to assess data quality were based on dendrograms, density histograms and analysis of key markers to determine cell stage. They were performed individually to each study (data not shown).

To illustrate the implemented quality control scheme, the individual types of checks are shown here for Gaspar *et al.,*2012 study[116], which assessed murine stem cell differentiation induced through hanging droplet procedure in a time series experiment over the period of 10 days.

The PCA of normalised gene expression data segregates samples according to the respective time points. Importantly, we can observe that replicated samples are placed in close proximity indicating a high reproducibility of measurement for each time point (Figure 3.3.1).

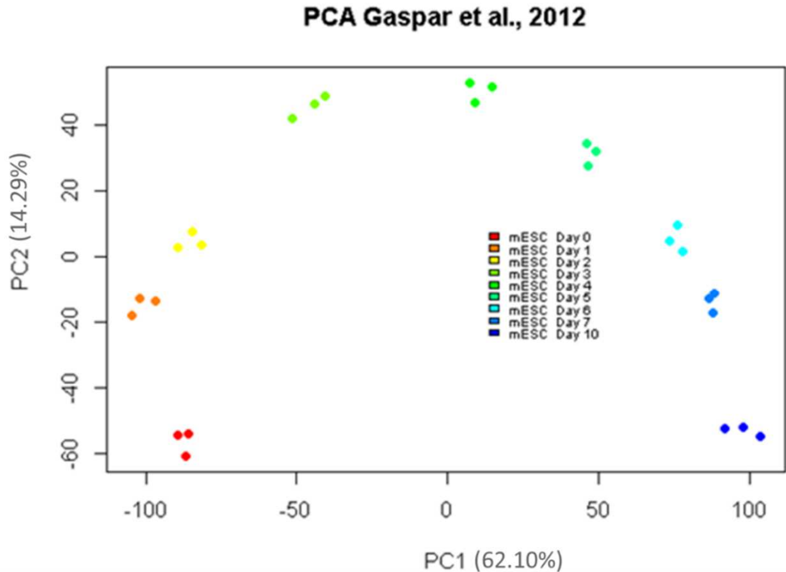


Figure 3.3.1 PCA for the study by Gaspar *et al.*, 2012. Biological replicates are coloured according to the sampling times.

To assess if the overall gene expression intensity is comparable between samples, density plots were generated. The resulting plot showed that very similar distribution for logged intensities were obtained after normalisation which indicates an absence of major artefacts and a comparability of samples (Figure 3.3.2).

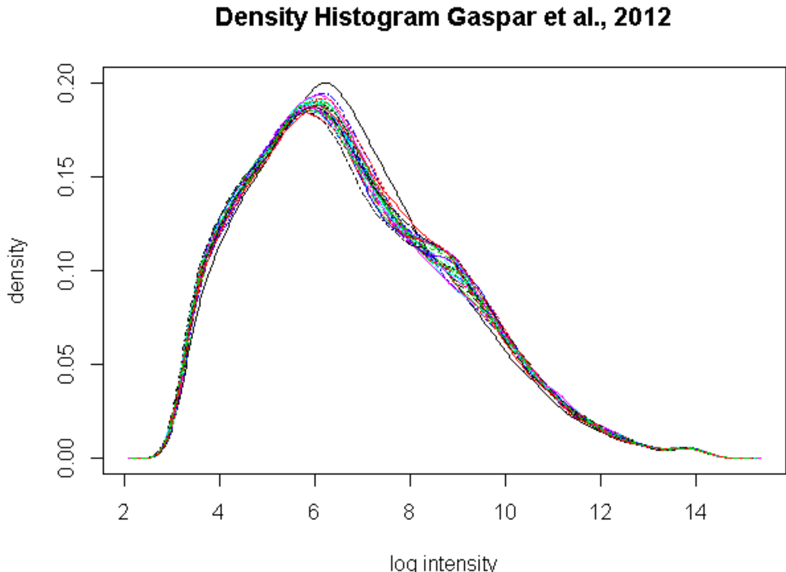


Figure 3.3.2 Density histogram for the study by Gaspar *et al.*, 2012. Distributions of normalised signal intensities are shown for the different microarrays included.

To confirm that the study did not include outliers, a dendrogram cluster was generated (Figure 3.3.3). Once again, replicates are clustering according to the time point of sampling. Moreover, we observe 3 distinct clusters, that represent the undifferentiated cell state from day 0 to day 2, the phase of cell fate commitment from day 3 to day 5, and the period for organ and tissue development from day 6 to day 10.

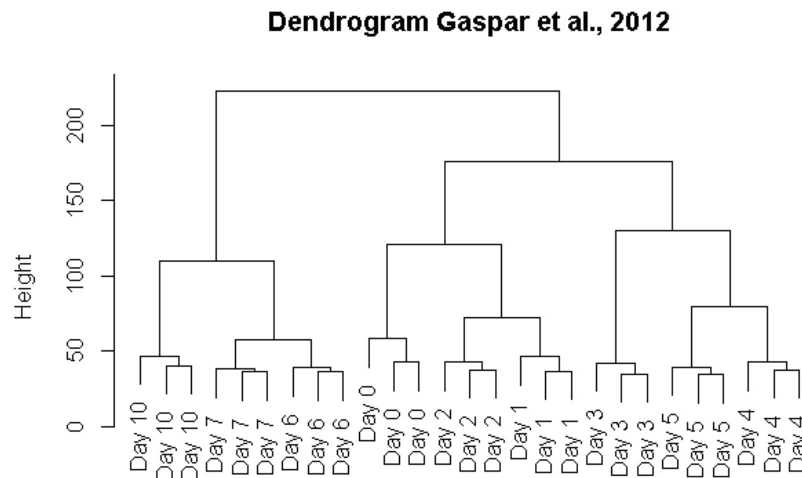


Figure 3.3.3 Dendrogram clustering for the study by Gaspar *et al*., 2012.

Finally, key marker gene expression for cell stages were assessed whether their gene expression follows the patterns expected based on previous reports. For this study, we can observe that expression of genes linked to pluripotency (*Nanog*, *Sox2* and *Pou5f1*)[139,140]. were downregulated over time while genes linked to heart development (*Nkx2-5*, *Gata4*, *Mef2c*)[141]. were upregulated as time progresses (Figure 3.3.4). For genes linked to mesoderm differentiation (T-brachyury and *Mesp1*) we can detect a peak expression at day 3 as expected for ESC differentiation[33].

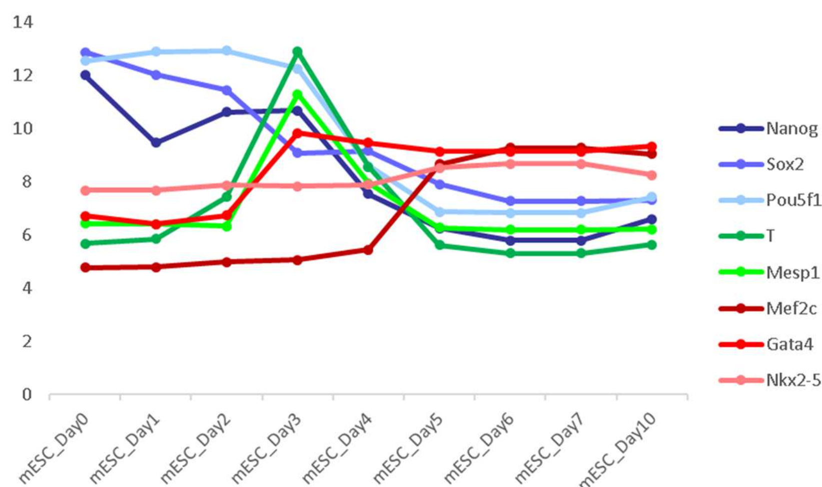


Figure 3.3.4 Key marker genes expression in the study Gaspar *et al*., 2012. Expression of pluripotency, mesoderm or cardiac related markers are shown in shades of blue, green or red tones, respectively.

As none of the quality controls indicated a flaw in the data, the full dataset by Gaspar *et al*.,2012 was subsequently included in HeartEXpress. If the quality control would have indicated a lack of coherency of replicate samples or other data artefacts, affected samples or even the whole study would have been excluded from further analysis and integration on HeartEXpress. The same assessments were carried out for all datasets before integration in HeartEXpress.

The described quality control focused on the assessment of individual studies and their internal consistency. While these checks/assessments can exclude individual studies or samples with spurious data from further analysis, it remained to be verified whether measurements taken in different studies were comparable after integration. To assess the inter-study comparability, a global PCA for mouse and human samples was performed, respectively. Both PCAs were performed using genes that have been associated in Gene Ontology (GO) with heart development (GO:007507). These genes should provide an assessment of relevant gene expression patterns given the biological focus of the studies. In contrast, inclusion of other genes might reduce the power of comparison, especially if they were not strongly regulated in the experiments or were affected by confounding experimental conditions such as culture medium.

In the PCA plot for mouse samples (Figure 3.3.5), we can observe that samples from the similar tissue, cell types or conditions tend to be placed in proximity to each other. In particular, we obtain cluster for *in vivo* experiments, *in vitro* mESCs differentiation, cardiac cells, and programmed cells as highlighted in Figure 3.3.5. This suggests that the gene expression values for specific cell type or from similar samples are consistent across different studies, when considering known heart related genes. In contrast, if all genes were included in the PCA, samples clustered by experiments indicating potential confounding factors which likely affect genes that were not strongly regulated (data not shown).

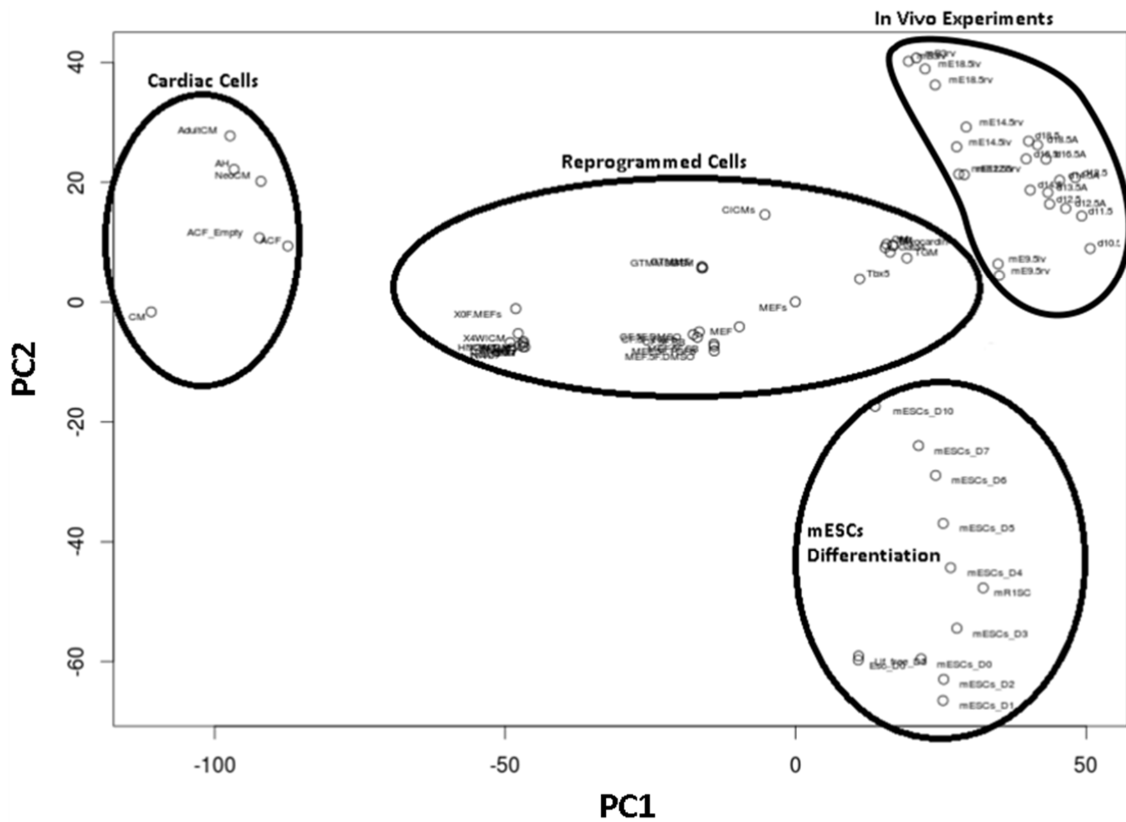


Figure 3.3.5 PCA of heart related expression data for mouse.

The same reduction to genes associated with heart development was applied for the PCA of the human expression data. I observed clusters consisting of microarray measurements of mature cardiac cells, cells that undergo differentiation and reprogrammed cells (Figure 3.3.6). Notably, hESCs showed a very distinct profile and were placed at the opposite end of the PC spectra in plot. The grouping of samples from different experiments in biologically relevant clusters indicate the comparability of the integrated data across studies.

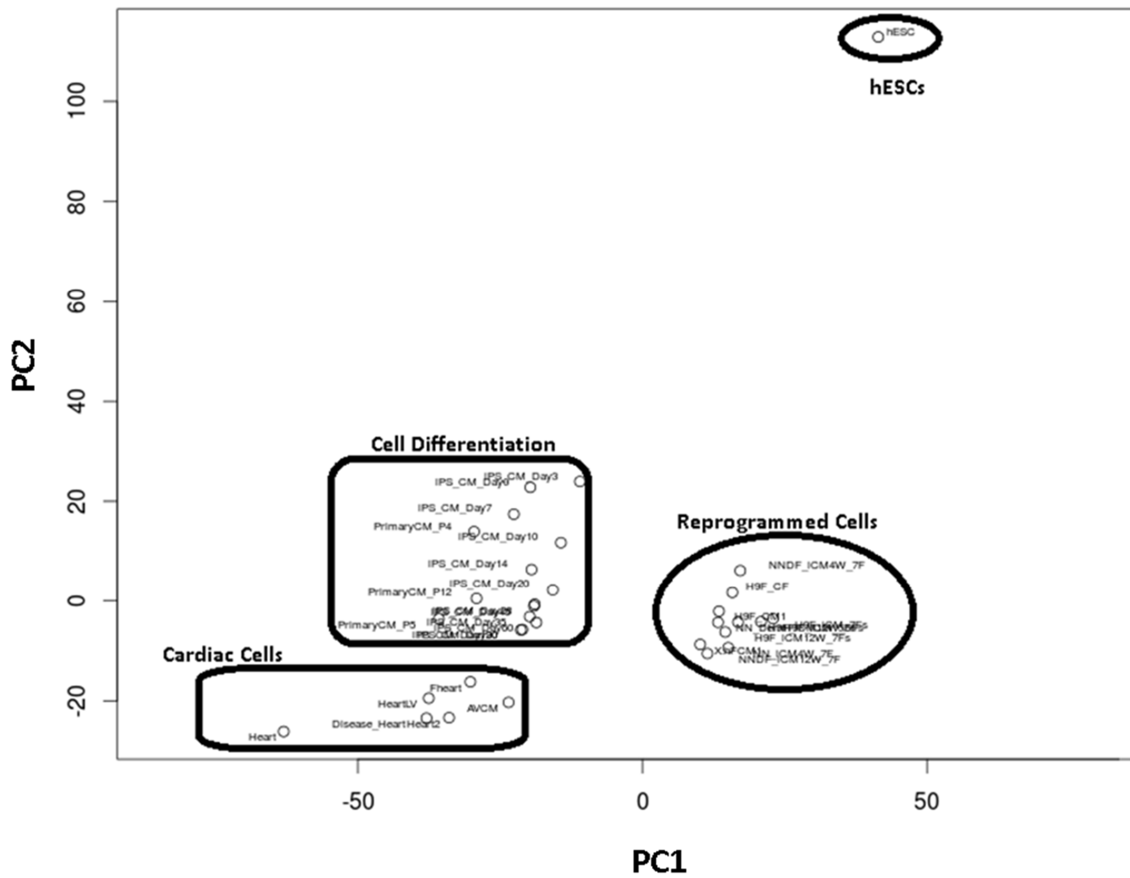


Figure 3.3.6 PCA of human heart related expression data.

Overall, global PCAs indicated that studies assessing similar conditions or cell types tend to provide similar expression patterns and support the feasibility of data integration in HeartEXpress. The next sections will describe the functionality and utility of the tool.

HeartEXpress website can be accessed via URL: http://heartexpress.sysbiolab.eu/. The ***Intro Page*** (Figure 3.3.7) contains a brief explanation about the webtool and enables the user to access the different sections available in the tool.

Figure 3.3.7 HeartEXpress Intro page.

On the ***Integrated datasets*** page (Figure 3.3.8) the microarray data integrated in HeartEXpress are listed. This section contains information about microarray labels, experiments, the origin, and processing of the expression data as well as respective accession numbers and PubMed IDs.

The ***Help*** page provides direct links to the repositories from where the data sets were obtained and where these experiments were described and published. Furthermore, the ***Help*** page includes a manual to help navigate and understand HeartEXpress.

AUTHOR	LABEL	DESCRIPTION	REFERENCE	DATABASE	ACESSION	PUBLICATION
Babiarz et al., 2012	iPS_derived_CM_D0	iPSC-derived cardiomyocytes, Day 0	Gene-wise Expression Average	GEO	GSE35671	PMID: 22050602
	iPS_derived_CM_D3	iPSC-derived cardiomyocytes, Day 3	Gene-wise Expression Average	GEO	GSE35671	PMID: 22050602
	iPS_derived_CM_D7	iPSC-derived cardiomyocytes, Day 7	Gene-wise Expression Average	GEO	GSE35671	PMID: 22050602
	iPS_derived_CM_D10	iPSC-derived cardiomyocytes, Day 10	Gene-wise Expression Average	GEO	GSE35671	PMID: 22050602
	iPS_derived_CM_D14	iPSC-derived cardiomyocytes, Day 14	Gene-wise Expression Average	GEO	GSE35671	PMID: 22050602

Figure 3.3.8 HeartEXpress Integrated dataset page.

The main functionality of HeartEXpress is accessible via the ***Gene Expression Analysis*** page. Here, the user can choose one of the integrated datasets available and visualise the clustered expression data (Figure 3.3.9). At this point, the user selects the dataset most relevant for the "biological question" in hands, e.g., *in vitro* differentiation of human cells.

Figure 3.3.9 Data sets can be selected on the *Gene Expression* page.

Upon dataset selection, a miniature expression heatmap is displayed on the left-hand side of the page. Using this miniature map, a range of genes can be selected, and their expression pattern magnified in a main frame (Figure 3.3.10A). The names and identifiers of the genes are given on the right side of the enlarged expression heat map. Gene identifiers are linked to the GeneCards database[142]. for follow up inspection of the corresponding genes.

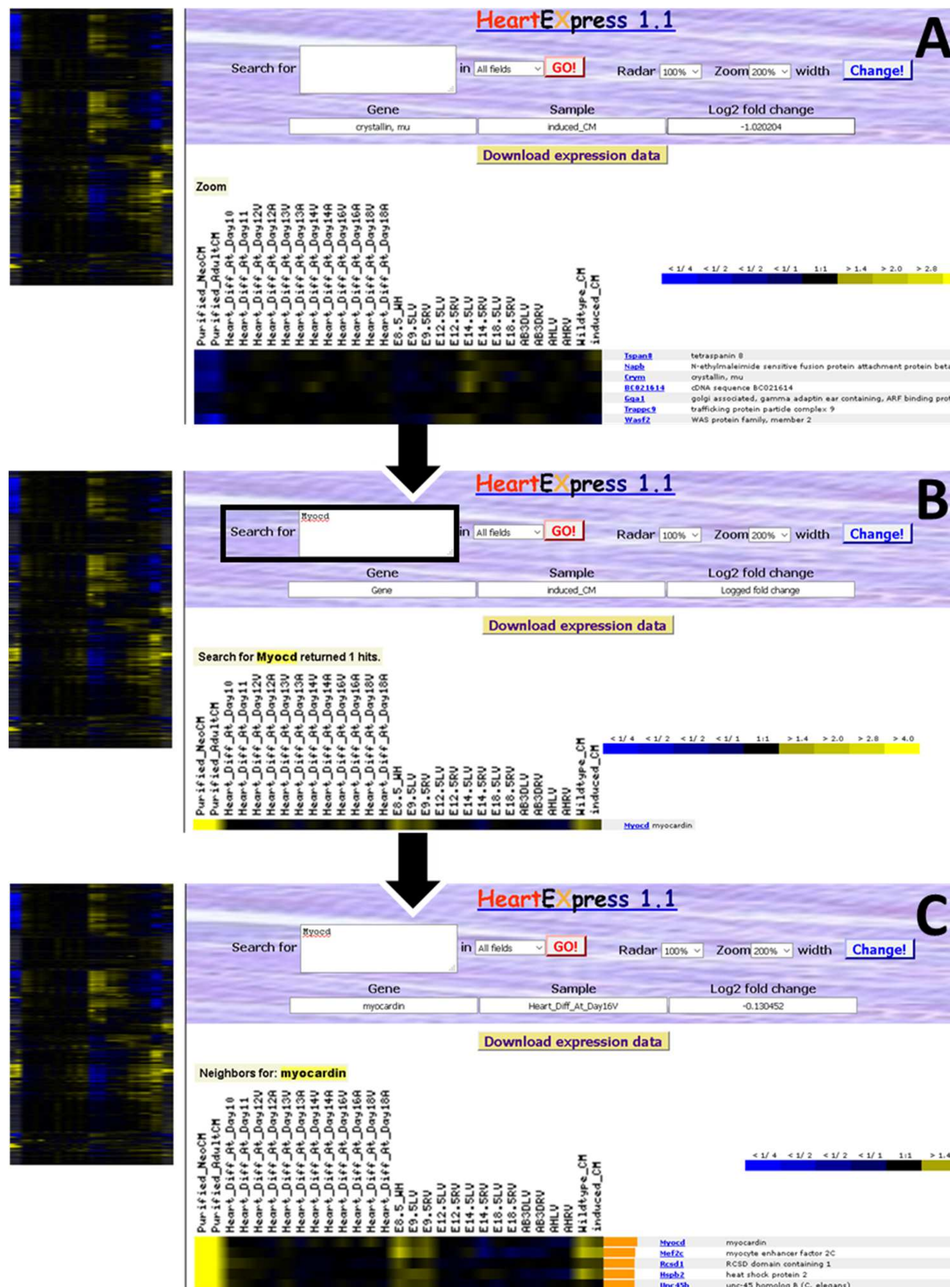


Figure 3.3.10 HeartEXpress gene expression analysis page. (A) HeartEXpress displays gene expression changes across different experimental conditions for a set of clustered genes. (B) Single or multiple genes can be queried, and upon gene selection (C) co-expressed genes and the correlation are displayed. Expression data can be downloaded as table in tab-delimited format. Detailed information about HeartEXpress functionalities can be found on the *Help* page.

Alternatively, it is possible for the user to query HeartEXpress for single or multiple genes, for which expression changes will subsequently be displayed. To execute such query, the user enters Entrez Gene identifiers or gene symbols in the ***Search*** box (Figure 3.3.10B, black rectangle). The retrieved expression data is provided as a heatmap formed by rows representing genes, and columns representing the experimental conditions. Expression changes are colour-coded using blue to yellow colour gradient, representing decreased (blue) or increased (yellow) expression compared to the reference sample or mean expression value across time series (as indicated individually for each experiment in the "integrated dataset" tab on the website). No differences in expression are represented by black colour, while grey indicate missing data. In the top panel, the user can obtain the gene description (*Gene* text box), experimental condition (*Sample* text box) and differential expression (*Log2 fold change* text box) of a gene by hovering the mouse pointer over the respective position in the heatmap. Finally, clicking on a specific gene in the heatmap will lead to the display of up to 50 genes that have the highest co-expression with the selected gene (Figure 3.3.10C) with a minimum coefficient correlation of 0.3. Results can be download as tab-delimited text file by clicking on *Download Expression* Data. The *Help* page provides further instructions about HeartEXpress functionalities.

3.3.3 HeartEXpress utility

The HeartEXpress webtool can serve many purposes such as the query for the expression profile of a gene of interest, the inspection of co-expressed genes under different experimental conditions or the formulation of biological hypothesis for further experimental validations. In the following case study, I illustrate the use of HeartEXpress to identify novel potential cardiac genes and/or markers. To this end, I selected the murine cardiogenesis dataset and consulted HeartEXpress' integral heatmap (Figure 3.3.11A) to identify a sub-cluster of genes that appeared to be differentially regulated during early heart development. In the selected gene cluster, I could identify established regulators of heart development such as *Isl1* and *Gata3*. Here, I decided to focus on *Isl1* since it is a cardiac transcription factor that is involved in the formation and maturation of cardiac cells from the second heart field[16]. Indeed, *Isl1* had a stronger expression in the right ventricle indicated by the heatmap (Figure 3.3.11B)[92].

Subsequently, co-expressed genes of *Isl1* were retrieved by simply clicking on the *Isl1*'s expression profile. The obtained co-expressed genes were examined and amongst them, I found genes that have been previously linked to heart related processes such as Wnt2[143], *Gata3*[144], Dkk1[145] and Wnt11[146] (Figure 3.3.11C). This result supports the use of co-expression identify functionally related genes, a principle which is also known as guilt-by-association.

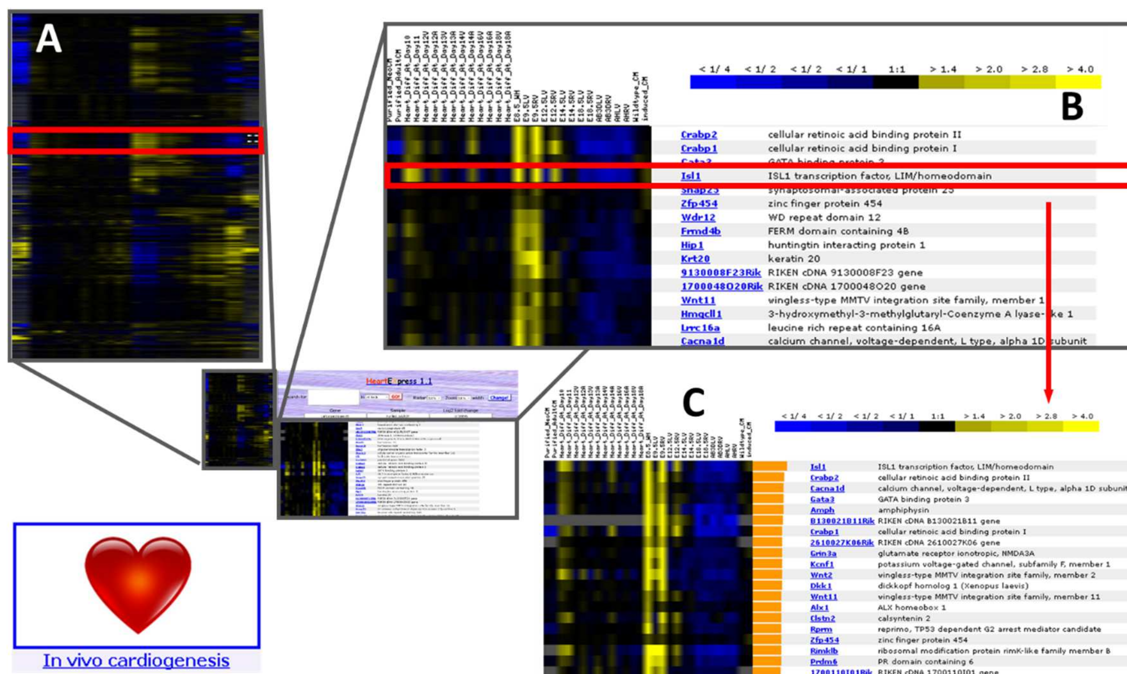


Figure 3.3.11 HeartEXpress utility applied to the murine *in vivo* cardiogenesis dataset. (A) A cluster of interest was selected based on its expression profile displayed in the heatmap on the left-hand side. (B) *Isl1* was selected from the group of genes that are present in the selected cluster. (C) Further inspection reveals genes that are highly correlated with *Isl1* and present similar expression patterns.

Interestingly, amongst the co-expressed genes I also found three so-called Riken genes, which originated from sequencing of full-length cDNA libraries collected at the Riken Institute and have frequently remained to be fully characterised. The observed co-expression of these genes (B130021B11Rik, 2610027K06Rik and 17110I01Rik) with *Isl1*, Wnt2 and Wnt11 suggests a potential relevance for heart development, cardiac cell formation and maturation. Importantly, none of these Rikens have been associated to any biological process, molecular function nor cellular process according to gene annotation provided by the GO Consortium[147]. The only information available from the NCBI gene annotation database for the found Riken genes is that B130021B11Rik is a protein coding gene, while 2610027K06Rik and 17110I01Rik are long non-coding mRNAs.

Based on the strong co-expression with key regulator *Isl1* and their novelty, they constitute attractive candidates for further study.

3.4 Conclusion

For this work, I collected and curated the data from several different transcriptomic studies. As a result, I developed the interactive webtool HeartEXpress which empowers the research community to explore and query the integrated gene expression data. Correlation and clustering analyses enable detection of co-expression patterns that might indicate common regulation and function during differentiation, reprogramming or heart development. Furthermore, HeartEXpress might be valuable for assigning functions to genes that have not been yet associated to heart development or morphology, prioritise genes for follow-up experimental validation and generate new hypothesis for investigations. Thus, the customised HeartEXpress functions can assist researchers in the quest of querying the gene expression profiles for heart development and can help to pinpoint novel potential candidate markers for further experimental validations.

Chapter IV

Genome-wide profiling of coding and non-coding RNAs to identify novel regulatory components in embryonic heart development

4.1 Introduction

The use of microarrays for systematic studies of gene expression has enable the simultaneous measurement of expression levels of thousands of genes. This kind of approach is especially advantageous for the analysis of complex processes such as heart development, as many events are occurring at the same time and only comprehensive techniques can provide a holistic view. While *in vivo* studies of heart development focused frequently on the characterisation of individual gene functions[119,120,137], only few comprehensive studies exist for the transcriptome profiling of *in vivo* cardiogenesis. In fact, only Li and co-workers *et al.,*2014 have published a profiling of transcriptomic changes during embryonic heart development and after birth[119].

Besides TFs, miRNAs are known to play important roles in gene regulation during development. This appears also to be the case for heart development. For instance, miR-1 is expressed at early stages of heart development, particularly at the onset of cardiomyocyte differentiation and during the formation of the heart tube[148]. Expression of miR-1 can be found in the ventricular myocardium and also in the interventricular septum from embryonic stage E8.5 until adulthood[149]. The expression of miR-1 has to be finely tuned, as it has a strong impact on cardiomyocyte differentiation and maturation. Both excess and loss of expression are detrimental for a proper heart development. For example, the overexpression of miR-1 in the embryonic heart blocks cardiomyocyte proliferation and pushes cells towards premature differentiation resulting in heart malformations due to lack of sufficient expansion of the cardiac walls[150].

Despite the strong indications of the importance of miRNA-based gene regulation, no comprehensive profiling of miRNA abundance in the developing heart had been carried out. This motivated us to measure the levels of miRNAs during murine heart development using microarray technology. To enhance the study, mRNA was measured in parallel enabling the alignment of changes in miRNA levels with changes in transcript levels of potential target genes (as described in detail in the next chapter). This unbiased systems approach made it possible to capture the dynamics of gene expression throughout the course of the experiment, to determine which genes and miRNAs are expressed or repressed in a comprehensive manner and to identify new genetic elements with potential functional relevance for murine

heart development. Elements of interest included protein-coding genes, miRNAs, and long non-coding RNAs. All the experiments were performed by collaboration partners at the University of Cologne (Germany), where also the microarray data were produced. The pre-processing, analysis and visualisation of the data was carried out by me.

The time-series experiment consisted of monitoring gene expression during the foetal heart development (E10.5 to E19.5) and in two different adult (10 weeks and 10 months) stages. As with other high-throughput data, the microarray measurements needed to be critically examined. To assess the reliability of the microarray experiment, established marker genes and miRNA were inspected as well as concordance of biological replicates. Both young adult and developmental stage E19.5 were used as reference timepoints to obtain DEGs. This approach was chosen to distinguish genes which are differentially expressed in the foetal heart compared to the adult heart and genes which show changes in expression in earlier time points compared to the ultimate time point (E19.5) in the foetal heart development. Interestingly, I detected various genes without functional annotation amidst DEGs which could constitute missing components of heart development and thus provide interesting candidates for further validation. Additionally, the experimental set-up enabled comparing male and female samples and observe how gender specific gene expression changes occur during heart development.

Overall, the combined assessment of gene and miRNA expression profiles provided insight about the dynamic expression changes throughout heart development and offered some clues about potentially novel genes partaking in this process. This chapter is based on the publication “Parallel Genome-wide Profiling of Coding and Non-coding RNAs to Identify Novel Regulatory Elements in Embryonic and Maturated Heart”[94].

4.2 Experimental design

Foetal hearts were isolated from E10.5 to E19.5 embryos dissected from OF1 pregnant mice. Young (10-week old) and old (10-month old) adult hearts of OF1 mice were isolated as well. For each time point, three samples were profiled by microarrays. For the time point from E12.5 onward, they included hearts from one female pup, from one male pup, and one mixed sample of male and female hearts. The same strategy for sampling of hearts was performed for young and old adult mice (Figure 4.2.1). Technical procedures of the tissue treatment and RNA extraction can be found in detail in the Davood *et al.*, 2018[94]. The microarray labelling and hybridisation techniques were performed as previously reported[151].

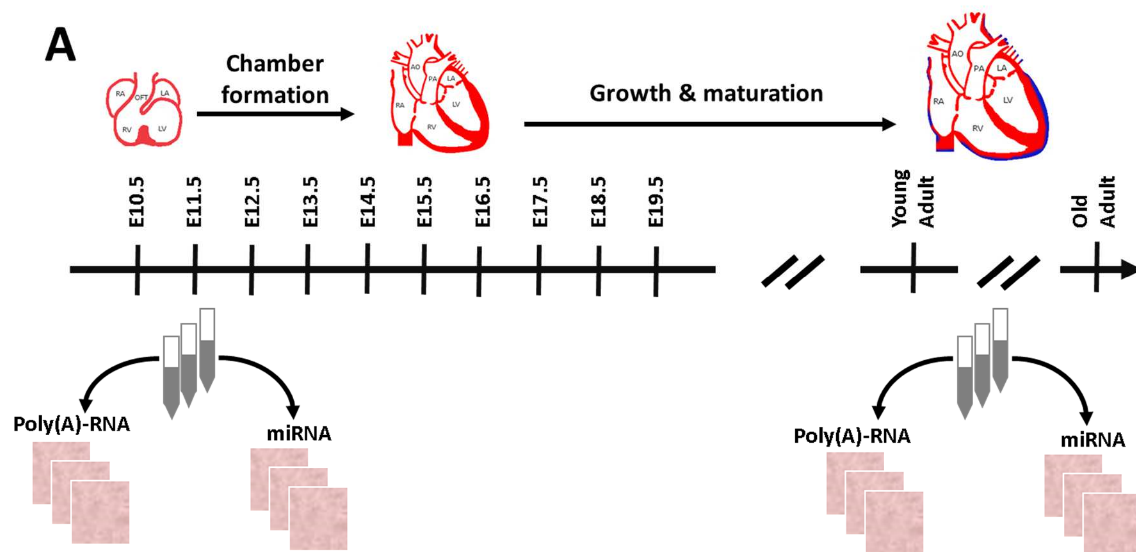


Figure 4.2.1 Experimental design of parallel monitoring of mRNA and miRNA during heart development and in two adult stages.

This experimental design allows us to assess the dynamic expression of mRNA and miRNA during murine heart development *in vivo* from embryonic stage E10.5 until E19.5, young adult, and old adult. Hence, this experiment broadly covers the early to final stage of heart development. In mouse, chamber formation and remodulation of the heart occur from E10.5 to E15.5 followed by the growth and maturation of the cardiac structure[152]. In total, seventy-two samples were profiled using Affymetrix Mouse Genome 430 2.0 Arrays, which includes probes against protein-coding as well as long non-coding transcripts, and Affymetrix miRNA 3.0 GeneChips. Samples from young and old adult hearts were used as reference for the developing heart. The dissection of heart tissue from both male and female embryos enabled the search for gender-specific effects on gene expression.

4.3 Results and discussion

4.3.1 Analysis of microarray profiles of mRNA and miRNA transcriptomes

For both mRNAs and miRNAs, clustering of the full microarray profiles was carried out. In general, samples from the same developmental status were grouped, suggesting a robust temporal expression signature (Figure S4.3.1). Expression profiles of young and old adult hearts formed a distinct cluster. Similar patterns also emerged in the PCA. Global expression during development (E10.5–E19.5) gradually approximated those of mature hearts, although a clear segregation of embryonic and mature samples remained (Figures 4.3.1A and 4.3.1B). This indicates substantial differences in expression between developing and mature cardiac tissue. Interestingly, the distribution of mRNA levels for E10.5–E13.5 showed a notable "shoulder", suggesting an underlying bimodal distribution, whereas samples from later time points displayed a gradual decrease in the number of genes with high expression values (Figure S4.3.2). Such bimodality might suggest a greater tendency toward an "on or off" mode of expression during early development, with more gradual adjustment during later stages of development.

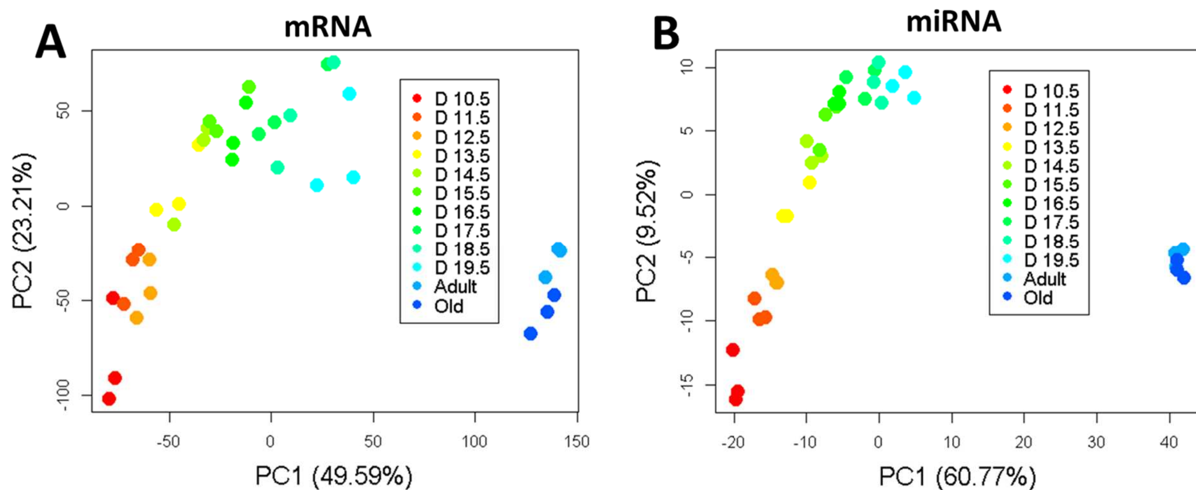


Figure 4.3.1 Principal component analysis of the transcriptome data. (A) Principal component analysis of the mRNA expression data; (B) Principal component analysis of the miRNA expression data.

4.3.2 Identification of differentially expressed genes in the developing heart

To detect DEGs, normalised Affymetrix GeneChip signal intensities of developmental stages were compared with intensities for young adult hearts. Because a large number of genes displayed changes in expression, a very stringent threshold for differential expression was set, using an adjusted p value of $\leq 10^{-5}$ and an absolute log2 fold change of ≥ 2 (4-fold change). In total, it was found 2,708 non-redundant genes to be differentially expressed for at least one time point. These represent 13% of the genes covered by the array. Notably, the number of DEGs reduced drastically with ongoing developmental time, reducing to 2,080 at E10.5 and further to 495 at E19.5 (Figure 4.3.2). DEGs at each time point are listed in Table S4.3.1.

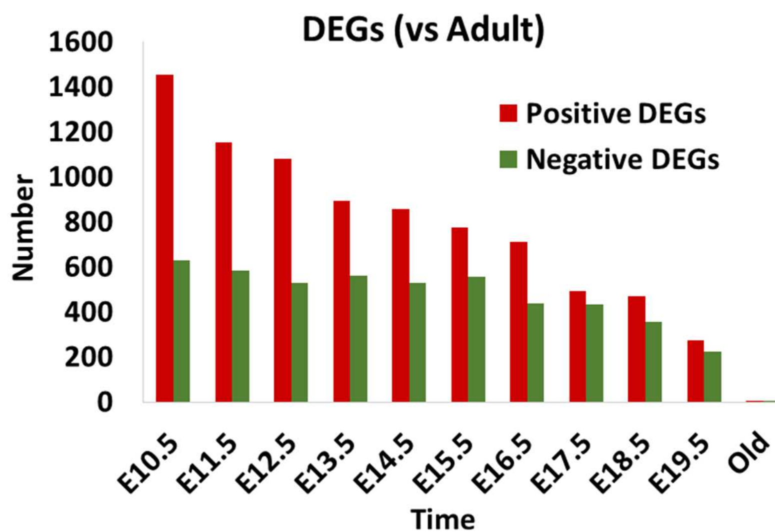


Figure 4.3.2 Number of differentially expressed genes (DEGs) in developing heart and older adult heart tissues, compared with young mature heart tissue.

4.3.3 Inspection of genes associated with heart development

To evaluate the dynamics of gene expression during cardiogenesis, the transcript levels of genes associated with heart development in Gene Ontology (GO: 0007507; Table S4.3.2) were initially inspected. TFs essential for heart development, such as *Gata3*, *Gata4* (Figure 4.3.3A), *Nkx2-5*, and *Hand2*, showed gradual downregulation during embryonic development, whereas *Myocd* (transcriptional co-activator of serum response factor) displayed an initial increase in expression but plateaued from E15.5 to E18.5 (Figure 4.3.3B). All these TFs were only weakly expressed in mature hearts.

Expression of *Foxc1*, *Foxc2*, and *Foxp1*of the forkhead family of TFs, known to play an important role in embryonic heart development,[153] decreased with progressive development, reaching lowest levels in adult tissue (Figure 4.3.3C).

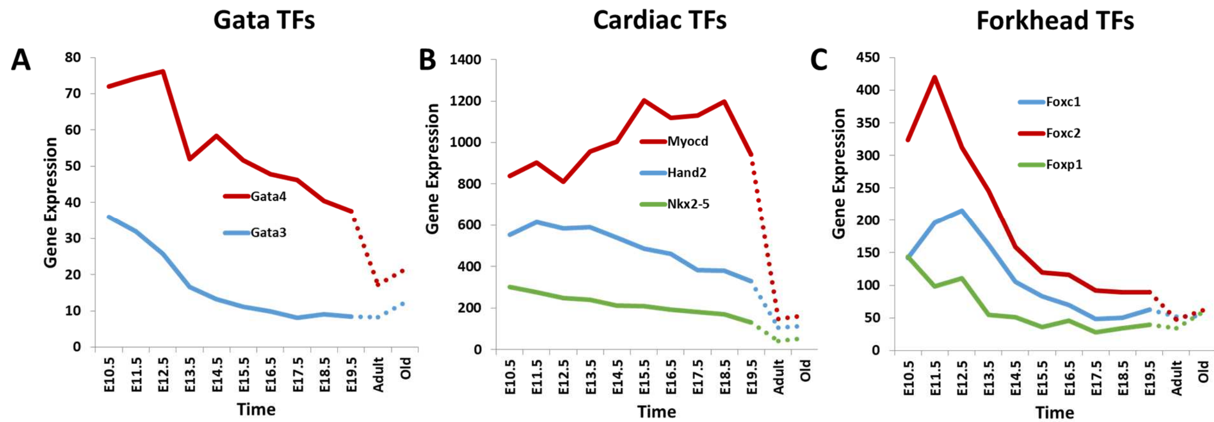


Figure 4.3.3 Temporal expression profiles of selected genes with existing association with heart development. (A) Gata transcription factors (TFs); (B) Known key TFs for cardiogenesis; (C) Fox family TFs. Dotted lines were plotted between E19.5, young, and old adult stages to emphasize the largely increased time intervals between these time points compared to the previous time points. Gene expression is represented in absolute intensity values.

Consistent with the established role of bone morphogenetic protein (BMPs) in transforming growth factor beta and Wnt signalling pathways during cardiac development[151], the expression of *Bmp2*, *Bmpr1a* (bone morphogenetic protein receptor type 1A), *Tgfβ2*, and *Wnt5a* was downregulated over time (Figure 4.3.4A).

The expression of genes associated with contraction of cardiac muscle, affecting the primary function of the heart are displayed in Figure 4.3.4B. Strikingly, cardiac alpha actin gene (*Actc1*), a known marker for early myogenesis[154], had the highest signal intensity of all genes at E10.5 but was subsequently downregulated, having minimum expression in the adult heart. Likewise, *Myh7*, which encodes for the b-myosin heavy chain, was highly expressed at all developmental stages but weakly expressed in adult hearts. This finding is consistent with previous observations of postnatal downregulation of *Myh7* in mice and other rodent hearts[155]. In contrast, the expression of cardiac troponin I (*Tnni3*) was gradually upregulated during development, with maximum expression occurring in mature hearts. An even more extreme case, showing almost switch-like upregulation, was observed for titin-cap (*Tcap*), linked to sarcomere assembly. Throughout most of the monitored developmental stages, it was expressed at marginal levels but began to accumulate at E19.5 only and was more than 10-fold upregulated in adult hearts.

The expression of epigenetic regulators was also investigated. Interestingly, the expression of histone deacetylase 2 (*Hdac2*) was strongly reduced in a linear manner during development, exhibiting a greater than 10-fold change (Figure 4.3.4C). *Hdac2* deacetylates lysine residues at the N-terminal regions of the core histones H2A, H2B, H3, and H4, playing an important role in transcriptional regulation and plasticity[13].

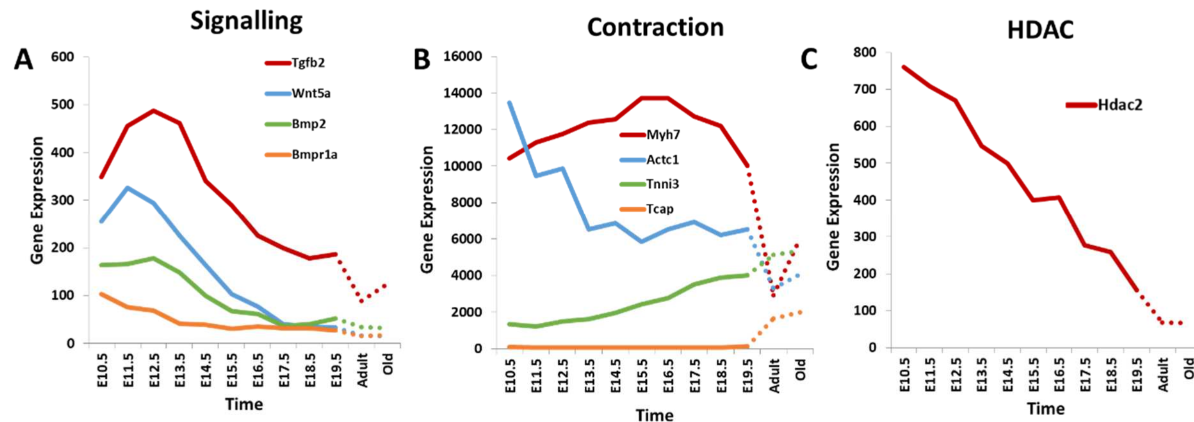


Figure 4.3.4 Temporal expression profiles of selected genes with existing association with heart development (II). (A) Signalling pathway molecules linked for cardiac development; (B) Genes encoding for structural cardiac proteins; (C) Histone deacetylase. Dotted lines were plotted between E19.5, young, and old adult stages to emphasize the largely increased time intervals between these time points compared to the previous time points. Gene expression is represented in absolute intensity values.

Ion channels are known to be essential for the correct functioning of the heart. **Table S4.3.3** lists differentially expressed genes associated with the activity of ion channels in GO identified in our study. Close examination suggests that ion channels undergo major remodelling during cardiac development. For instance, *Cacna1h* is encoding for α_1 subunits and *Cacna2d2* is encoding for α_2-δ subunits of voltage-gated calcium (Ca^{2+}) channels were gradually downregulated during development and were expressed at relatively low levels in adult hearts. In contrast, the expression of *Cacna1g* showed reciprocal expression changes during heart development. Such remodelling may have physiological relevance, as *Cacna1g* and *Cacna1h* encode for the distinct subtypes of T-type Ca2+ channels: Cav3.1 (a1G) and Cav3.2 (a1H), which are crucial for electrical conduction in the atria[156]. Furthermore, it has been reported that disruption of *Cacna2d2* results in abnormalities in heart development[157] (Figure 4.3.5A).

Among the genes corresponding to sodium channels, it was found that the α subunit encoded by the *Scn7a* gene remained very weakly transcribed in embryonic hearts, but had up to 10-fold higher expression in mature hearts, displaying a switch-like expression pattern (Figure 4.3.5B).

The genes for potassium channels, *Kcna5* and *Kcnb1*, were very lowly expressed in all developmental stages, but induced in both young and old adult hearts (Figure 4.3.5C). *Kcna5* encodes for the voltage-gated K^+ channel (Kv1.5), which has emerged as a promising target for the treatment of atrial fibrillation[158,159]. The voltage-gated K^+ channel *Kcnb1* is mainly expressed in the heart, brain muscle, and pancreas[160] and is a key player in apoptotic programs associated with oxidative stress in the cardiovascular system[161].

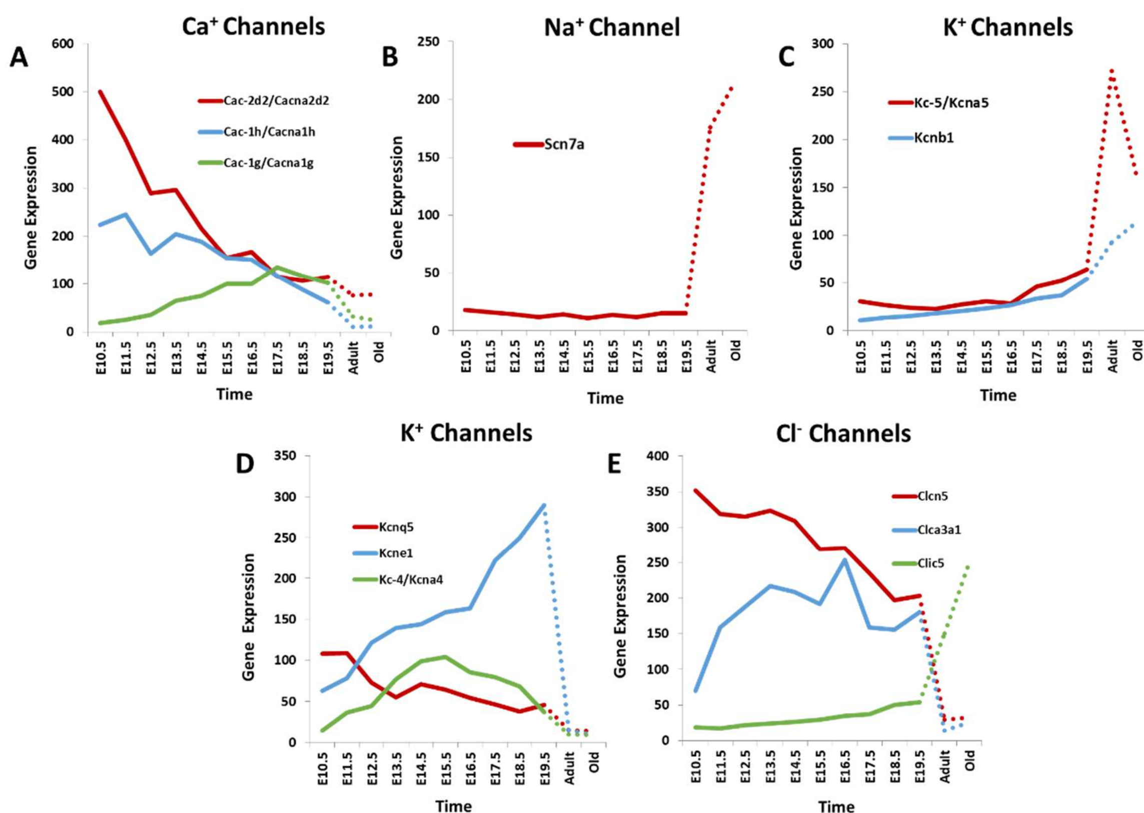


Figure 4.3.5 Temporal expression profiles of selected genes with existing association with heart development and ion channel function. (A) Calcium channel genes; (B) Sodium channel gene; (C and D) Potassium channel genes upregulated (C) and downregulated (D) in young mature heart tissue; (E) Chloride channel genes. Dotted lines were plotted between E19.5, young, and old adult stages to emphasize the largely increased time intervals between these time points compared to the previous time points. Gene expression is represented in absolute intensity values.

In contrast, transcripts of *Kcne1* gradually accumulated during development, but were markedly depleted in adult hearts (Figure 4.3.5D). Expression of *Kcnq5* tended to be more strongly repressed as development occurred, whereas expression of *Kcna4* showed a

parabolic expression pattern, with maximal expression at E15.5. This transcript displayed a distinct pattern that warrants further investigation, as its role during heart development has not yet been explored (Figure 4.3.5D).

A remarkable transcriptional plasticity was also observed for chloride (Cl^-) channels. Expression of *Clinc5* progressively decreased between E10.5 and E19.5, resulting in very low levels in adult hearts (Figure 4.3.5E).

A transient pattern was observed for *Clca3a1.* After an initial increase from E10.5 to E13.5, an almost stable expression level was maintained until E19.5, before a drastic reduction occurred, yielding very low levels in both young and old adult hearts. *Clca3a1* encodes for a calcium-activated chloride channel contributing to the regulation of cellular volume. It has been speculated that *Clca3a1* is involved in arrhythmogenesis of the heart[162].In contrast, *Clic5,* which encodes for a chloride channel located in cardiomyocyte mitochondria of different rodents,[163]. displayed a linear increase in expression during development, with a considerable boost in expression in adult heart tissue.

These findings suggest that the composition of cardiac ion channels undergo major changes during development, as many of their components were differentially expressed during the various stages of heart development, compared with adult hearts.

Finally, the transcript level of cell-division cycle (*Cdc*) genes such as *Cdc27*, *Cdc20* and *Cdca3* decreased during embryonic development and were further reduced in mature cardiac tissue reflecting the ceasing of cell cycle activity with progressing heart development (Figure 4.3.6). Interestingly, I found that *Rgcc* (regulator of cell cycle) displayed a contrasting pattern with low transcript levels during earlier development but high levels in adult heart. This suggests that *Rgcc* might contribute to post-natal cell cycle arrest in cardiomyocytes (Figure 4.3.6).

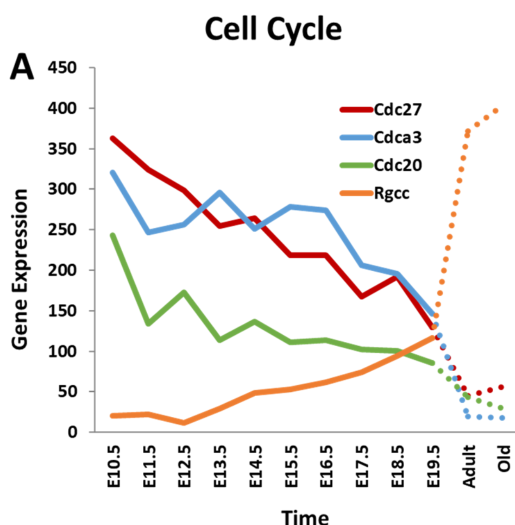


Figure 4.3.6 Temporal expression profiles of selected genes with cell cycle. (A) Cell cycle genes. Dotted lines were plotted between E19.5, young, and old adult stages to emphasize the largely increased time intervals between these time points compared to the previous time points. Gene expression is represented in absolute intensity values.

The same method and thresholds for DEGs were used to identify differentially expressed miRNAs (DEmiRs). In total, I identified 217 DEmiRs. Similar to DEGs, the number of DEmiRs gradually decreased from 191 at E10.5 to 98 at E19.5, as heart development progressed (Table S4.3.4; Figure 4.3.7).

Comparing the number of upregulated and downregulated DEmiRs, upregulated DEmiRs were 2.5–3.5 times more frequent than downregulated DEmiRs during heart development. This contrast with the observed ratios of up- to downregulated DEGs, which is more balanced in later stages of development (Figure 4.3.2).

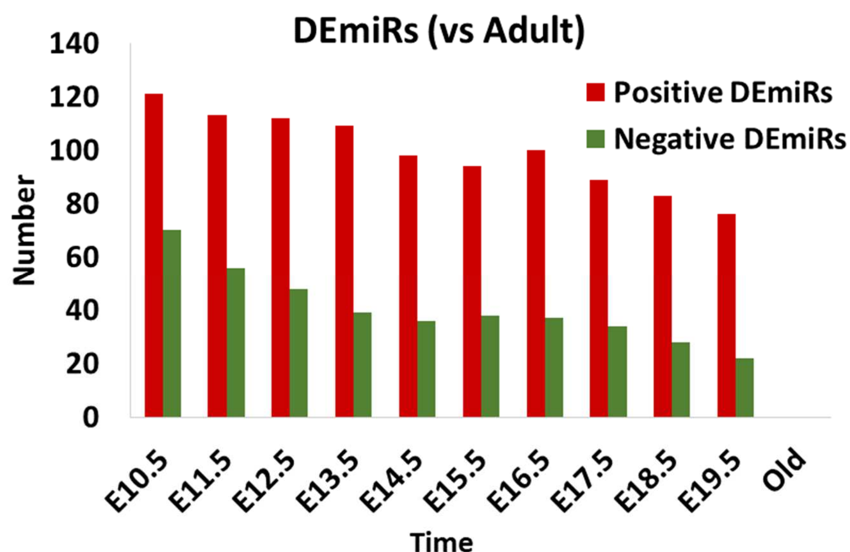


Figure 4.3.7 Number of differentially expressed miRNAs (DEmiRs) in developing heart and older adult heart tissues, compared with young mature heart tissue.

A number of DEmiRs associated with the cardiac cell lineage were identified, including miR-17[58], miR-29-3p[164,165], miR-199a-3p[166], miR190-3p[167], and miR-1[56,57,64].

The expression of miR-17-5p was gradually downregulated from E10.5 to E19.5, with a subsequent 3-fold drop in expression in adult heart tissue (Figure 4.3.8). This supports a role in the regulation of proliferation of cardiac cells, as recently reported[58].

In contrast, the expression levels of miR- 29a-3p, miR-195a-5p, and miR-1a were relatively low during all developmental stages but highly upregulated in adult hearts. More recently, it has been shown that miR-29-3p is highly upregulated in adult hearts, as well as under pathological conditions, such as hypertrophic cardiomyopathy[164,165]. The expression of miR-1 was also high in mature hearts compared with all embryonic stages.

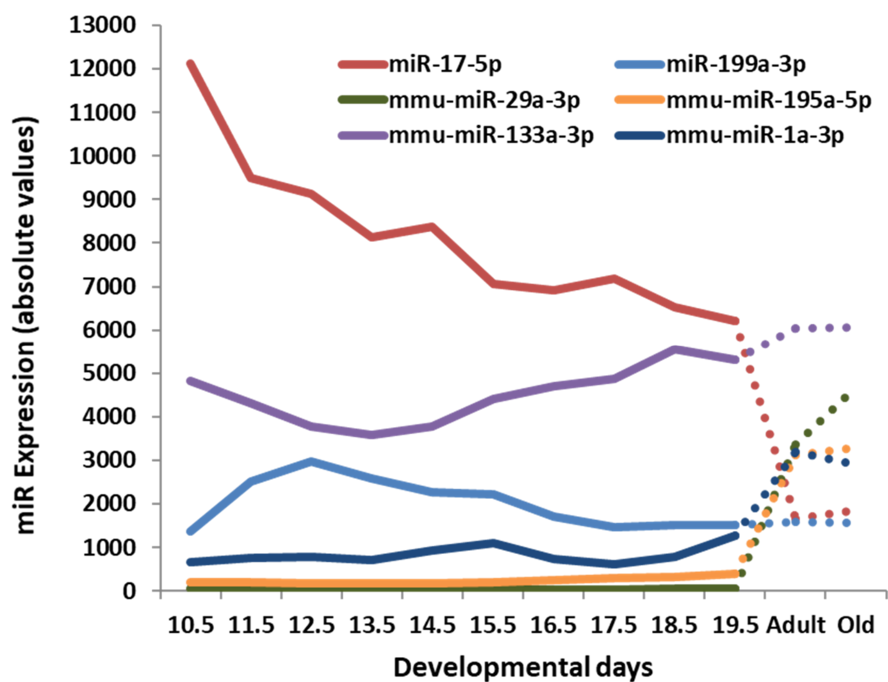


Figure 4.3.8 Temporal expression profiles of selected differentially expressed microRNAs (DEmiRs) for embryonic and mature heart tissues.

More transient expression patterns were displayed by miR-199a-3p and miR-133a, for which the maximum and minimum signal intensities was recorded at developmental stages of E12.5 or E13.5, respectively. Expression patterns for miR-199a-3p and miR-133a showed maxima and minima at developmental stages E12.5 and E13.5, respectively. Expression of miR-199a-3p has been reported in the adult human heart and is correlated with heart failure[166,168]. In addition, miR199a-3p expression plays a pivotal role for cardiomyocytes survival[167].

The use of young adult heart tissue as reference delivered a large number of genes that were differentially expressed in embryonic tissue compared to mature tissue. To demarcate genes that showed significant expression changes during the embryonic development from genes which only change expression postnatal, I reanalysed the transcriptomic with E19.5 embryonic heart as new reference. As shown in Figures 4.3.9A and 4.3.9B, the number of the newly detected DEGs and DEmiRs rapidly decreased in the early stages of the development (E10.5–14.5) in exponential manner. Compared to the larger numbers of genes and miRNAs observed in the previous comparison (Figures 4.3.2 and 4.3.7), the numbers of genes and miRNAs with significant dynamic expression during development were considerably smaller. In particular, only a few genes and no miRNAs were found differentially expressed for E15.5–E18.5 when compared to E19.5. Analysing the miRNA expression in mature tissue with E19.5 as reference confirmed the imbalance between positively and negatively regulated miRNAs. There were 3.5 times more negatively regulated than positively regulated miRNAs in mature tissue when compared to E19.5. Tables S4.3.5 and S4.3.6 list the genes and miRNAs that were found differentially regulated using E19.5 and/or young adult heart as reference, respectively. The tables also show genes and miRNAs that were found as differentially expressed in both references.

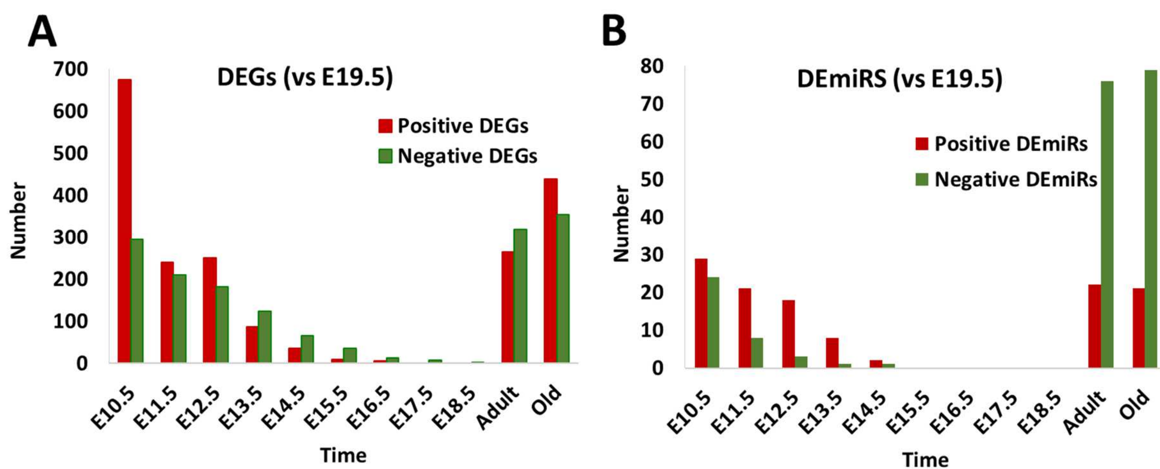


Figure 4.3.9 Number of differentially expressed genes and miRNAs in developing heart and older adult heart tissues, compared with E19.5 heart tissue. (A) DEGs and (B) DEmiRs.

This specific experimental design enabled us to search for sex and age specific changes. PCA and clustering analysis (Figures 4.3.1A and S4.3.1) indicated that gene expression in the hearts of older adult mice (10 months old) closely resembles that in hearts of younger adult (10 weeks old) mice. Only seven genes were found to be differentially expressed (Table S4.3.1).

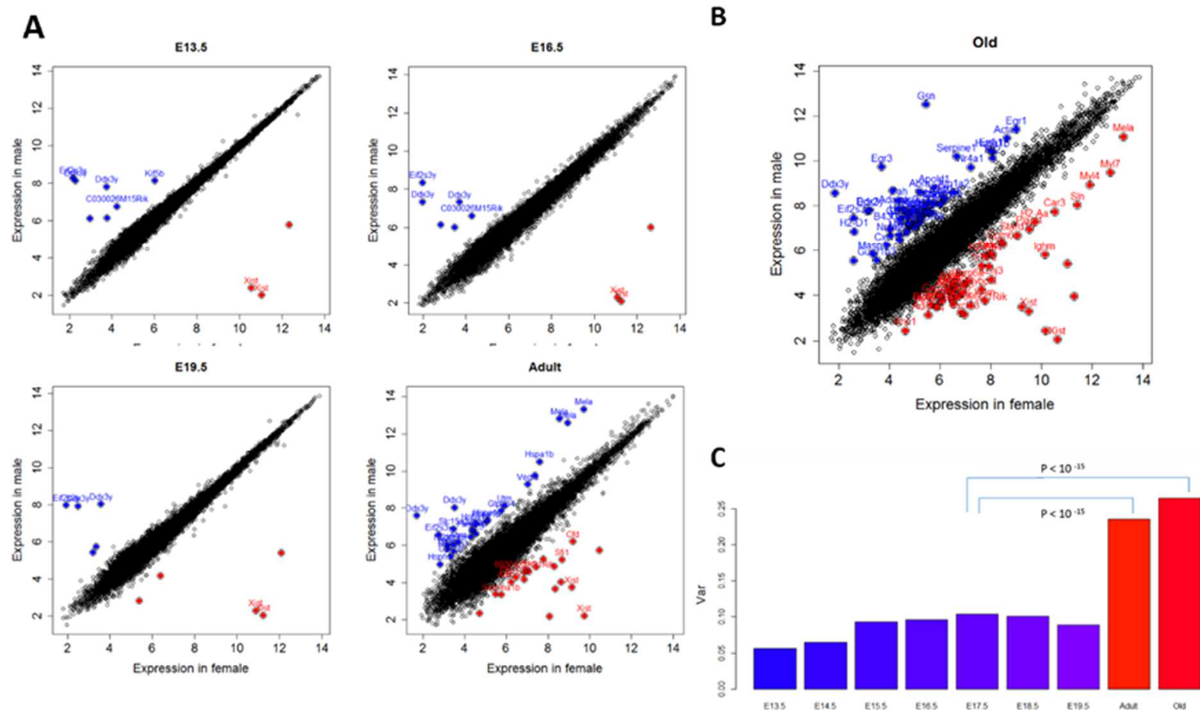


Figure 4.3.10 Comparison of gene expression in male and female samples. (A) Scatterplots of logged signal intensities for male vs female samples of E13.5, E16.5, E19.5 and young adult mice. Red dots mark probe sets with signal intensities that are more than 4-fold larger in female than in male samples. Blue dots indicate probe sets with at least 4-fold larger expression in male samples; (B) Scatterplot for old heart tissue. C) Variance of differential expression between male and female samples for different time points. P-values were derived using the F-test.

The most upregulated gene in older hearts was *Adamts9*, which is a member of the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) family. It encodes for a secreted protease, acting as an angiogenesis inhibitor[169]. The most downregulated gene was the coiled-coil domain containing 141 (*Ccdc141*).

In relation to sex specific expression patterns, I only found three genes that were consistently differentially expressed between male and female tissue samples over the time series. As expected, *Xist* was detected only in female samples, whereas *Eif2s3y* and *Ddx3y*, both located on the Y chromosome, were detected only in male samples. Clearly, the extent of gender-specific expression was marginal during development, although significantly larger variability

in expression between male and female samples existed for mature hearts, especially for older hearts (Figure 4.3.10). The number of genes having more than a 2-fold expression change between male and female samples increased from 9 at E19.5 to 38 in adult tissue and up to 110 in older mature tissue (Table S4.3.7). Of these, 60 genes were upregulated and 50 were downregulated in old male heart tissue.

The gene *Atp1a2* was found to be upregulated in old adult male compared with old adult female heart tissue. It belongs to the subfamily of Na+/K+-ATPase membrane proteins, regulating the electrochemical gradient in the cardiomyocytes and other cell types by transporting Na^+ and K^+ against their intracellular and extracellular concentrations. Similar upregulation was detected for gelsolin, *Gdf15*, members of the *Egr* and *Nr4a* protein families, and *Pah* in our study. Increased protein levels of gelsolin have previously been found in several organs of old rats[170]. In senescent human fibroblasts, an enhanced aging-associated resistance to apoptosis was observed for increased gelsolin expression. *GDF15* is a cytokine having increased expression levels in heart degenerative diseases[171] and is induced under various stress conditions by the early growth response protein-1 (EGR-1), whose transcript increased in abundance over time in our study, together with those of its paralogs *Egr-2* and *Egr-3*. Members of the nuclear receptor subfamily 4, group A (*Nr4a*) are transcriptional regulators of metabolism and energy balance that are induced by a pleiotropy of stimuli and processes. Recently, it was hypothesised that they could serve as targets for anti-aging interventions[172]. Analysis of different microarray studies showed that aging led to consistently increased expression of phenylalanine hydroxylase (*Pah*) in murine hearts[173].

Finally, a GO enrichment analysis was performed for up- and downregulated genes. The upregulated genes showed significant overrepresentation of cytoskeletal genes, especially those associated with actin filament and organisation. In fact, the most upregulated gene was gelsolin, which encodes for a protein regulating the actin cytoskeleton. In contrast, no enriched GO categories were associated with downregulated genes.

As in the case of mRNAs, I searched for differentially expressed miRNAs in older versus younger adult heart tissue, as well as for gender-specific changes. However, no significant differential expression was detected in these comparisons.

We were particularly interested in genes listed as Riken genes or predicted in the Mouse Genome Informatic (MGI) database due the potential novelty. The former originates from sequencing of full-length cDNA libraries collected at the Riken Institute, and many of these remain poorly characterised. In our study, I found 127 Riken and predicted genes among the DEGs. After removal of genes that were functionally annotated in GO, 107 Riken or predicted genes remained (Table S4.3.8). Remarkably, almost a quarter (23%) was curated as (long) non-coding RNA genes in the MGI database. The protein-coding Riken and predicted genes were divided into three classes, based on whether their expression patterns showed (1) a gradual decrease in expression over time (29 genes; Figure 4.3.11A); (2) a transient higher expression during development (37 genes; Figure 4.3.11B); or (3) a gradual increase in expression over time (16 genes; Figure 4.3.11C).

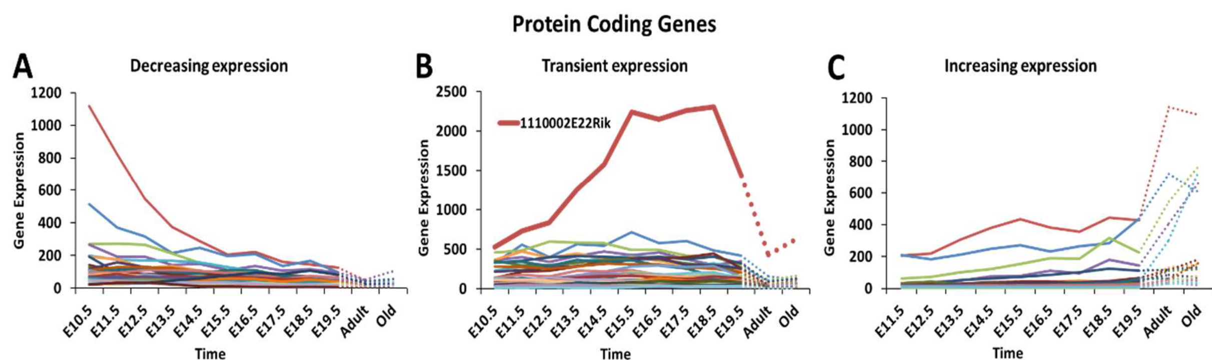


Figure 4.3.11 Differentially expressed genes encoding for mRNA without current functional annotation. (A–C) Protein-coding genes with decreased (A), transient (B), or increased (C) expression.

A particularly strong increase of over 6-fold change and a high absolute expression level at E15.5 was recorded for 1110002E22Rik. This was also one of the few Riken genes, for which a putative protein domain was detected. Close to the C terminus, a histone deacetylase superfamily domain (IPR000286) was predicted by Interpro, providing an interesting clue to the function of 1110002E22Rik.

Similarly, the differentially expressed non-coding genes were grouped into three classes, reflecting their expression patterns. In this case, six genes were downregulated during development (Figure 4.3.12A) and ten genes displayed transiently higher expression (Figure 4.3.12B), whereas nine genes were gradually induced (Figure 4.3.12C).

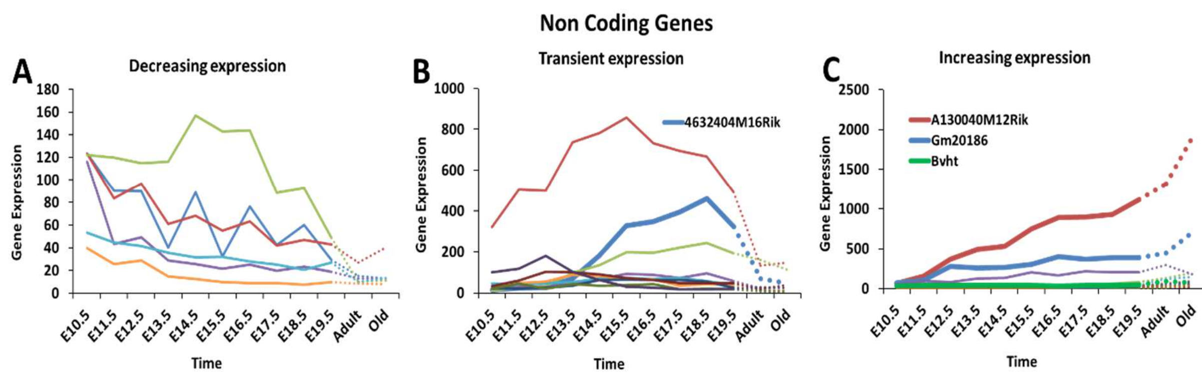


Figure 4.3.12 Differentially expressed long non-coding genes without current functional annotation. (A-C) Non-coding genes with decreased (A), transient (B), or increased (C) expression during embryonic development.

Two lncRNAs showed an especially strong increase in expression during development: Riken A130040M12 (29-fold) and Gm20186 (10-fold). The former is a 3,366nt long transcript derived from Viral-like 30 elements (VL30s). The latter is a 741nt long transcript and was detected as widely expressed across multiple organs in mouse embryos by *in situ* hybridisation (MGI-Gene Expression database). The expression of both lncRNAs increased in old compared with young adult hearts. Remarkably, a similar expression pattern was found for Braveheart, a lncRNA that was recently shown to modulate the expression of cardiac TFs[174]. Whereas Braveheart was not included in the list of DEGs because of our stringent threshold, its previously reported expression is consistent with our data[174]. In our study, Braveheart displayed low transcript levels during developmental stages but a two-fold upregulation in mature heart tissue (Figure 4.3.12B). A more transient expression pattern was observed for 4632404M16Rik (Figure 4.3.12C), which is an intronic lncRNA of 2,882nt length. It initially follows the upregulation of cardiac calsequestrin 2 (*Casq2*), in which 4632404M16Rik is located. *Casq2* is the most abundant Ca^{2+}-binding protein in the sarcoplasmic reticulum and integral to the high Ca^{2+} capacity of the sarcoplasmic reticulum. Although *Casq2* remained highly expressed after development, 4632404M16Rik expression was only weak in mature tissue, suggesting that its function is evoked mainly during cardiogenesis (Figure S4.3.3).

4.4 Conclusion

The identification of differentially expressed mRNA and miRNA during heart development gave us a detailed view of the underlying dynamics of this process. Many differentially expressed genes and miRNAs detected in this study have been described earlier in the literature, which supports the results of our study. An interesting by-product of the applied experimental design was the identification of genes that showed gender- or age-dependent expression. The gender/age-dependent analysis indicated no significant difference between male and female during embryonic heart development. However, hearts from old males appear to have higher levels of known age-related marker genes, such as, gelsolin, *Gdf15*, and *Pah*, when compared to female hearts.

One of the most interesting results obtained were the genes encoding for lncRNAs, which are broadly defined as non-coding transcripts of length of 200nt or more. Several lncRNAs are known to play key roles during heart development and have been associated with the regulation of gene expression through a variety of mechanisms on transcriptional, post-transcriptional, and epigenetic levels[175]. In this context, a recent study identified a mouse-specific lncRNA termed Braveheart, expressed in ESCs and in mature heart tissue[174]. It was shown that Braveheart strongly enhances the cardiac commitment by acting as a decoy for *SUZ12* (a member of the polycomb repressive complex 2 [*PRC2*]), thereby releasing *MESP1* (a master regulator of cardiac differentiation) from *PRC2* suppression. In our study, I detected 25 lncRNAs with significant differential expression. Given the relevance of lncRNAs in gene regulation, this set of lncRNAs provides an attractive list of candidates for future study.

In addition to TFs and lncRNAs, miRNAs constitute another important class of regulators of gene expression. They have pivotal roles in developmental processes and are implicated in various diseases. Despite the known importance of miRNAs for regulation of heart-specific expression, a comprehensive overview of miRNA abundance in the developing heart has never been undertaken. Our study provided the first genome wide capture of the dynamics of miRNAs in the developing murine heart and faithfully captured the expression of miRNAs that were previously associated with the heart supporting their involvement in heart development and cardiac tissue maintenance[55,57,58,176,177]. The inclusion of profiles of young and old heart tissue in our dataset enabled us to identify miRNAs that undergo characteristic postnatal

changes and might drive the maturation process. In addition to providing fundamental biological knowledge, understanding the expression dynamics of miRNAs captured in our study might therefore provide crucial clues for regenerative treatment of heart diseases[178].

The transcriptome analysis contribution towards the deciphering of embryonic heart development is a key element to better understand this process. However, there are several limitations regarding the profiling platform and experimental design. Our study was carried out using Affymetrix microarray platform due to its ready availability for our research project. An alternative approach could have been RNA sequencing (RNA-seq). A major advantage of RNA-seq is the capacity of identifying previously unknown transcripts in contrast to the microarrays, which can identify only transcripts for which probes have been included on the microarray. However, there are also drawbacks of the newer transcriptomic technology. It is still relatively costly and requires a more complex and time-consuming process of storing and analysing of the generated data. In the context of meta-analysis such as those performed in this dissertation, one important asset of microarrays is also the availability of numerous published datasets, which allows an easy reanalysis and direct comparison with newly generated data. For comparison across different paradigms such as microarrays and RNA-seq, the use of data transformation methods is required[179].

A limitation of the study imposed by the design was the lack of neonatal and postnatal transcriptome profiles, which could have captured important stages in the maturation process of the murine heart. This gap in monitoring is reflected by the large differences of expression profiles between the last day of embryonic development (E19.5) and of the adult stages. Therefore, future integrative studies adding transcriptome profiles for neonatal and postnatal heart tissue can help to delineate the dynamics of gene expression during maturation and to cover the neonatal time period in which the murine hearts still display regenerative capacities.

The next chapter will focus profiling potential miRNA-mRNA interactions and inferring one novel miRNA candidates that might be crucial for embryonic heart development.

Chapter V

HeartmiR:

A web tool for visualisation of miRNA-mRNA regulatory interactions with potential relevance for heart development

5.1 Introduction

To assess the role of miRNAs, an essential step to take is to identify gene targets of the miRNAs. Currently, most targets are delivered by computational tools that identify promising miRNA-mRNA interaction candidates, which subsequently need to be subjected to experimental validation[180]. Due to the importance of identifying these interactions, numerous algorithms have been developed to perform miRNA target prediction and functional characterisation[181]. As technology evolves, new models have been created based on experimental advances as well as novel insights into the mechanisms of mRNA target regulation. Several features have been identified to improve the prediction of miRNA targets, such as perfect base pairing of the miRNA 5'-end to the target site, or low level of GC content in the target resulting in increased accessibility for miRNA binding[182,183]. Although there has been substantial progress in the field of miRNA target prediction, the available algorithms still suffer from suboptimal performance and restricted coverage. To obtain a comprehensive basis for the analysis of miRNA-based gene regulation during heart development, I collected therefore miRNA-mRNA interactions data from five independent resources, one of which is based on experimental data[184] while the other four were obtained through computational prediction of miRNA-mRNA interaction[182,185-187]. This merged miRNA-mRNA interaction dataset was then combined with *in vivo* gene expression data for embryonic heart development (*Chapter IV*) to create a data structure linking mRNA and miRNA expression through miRNA-mRNA interactions. The integrated data were used to identify and characterise relevant miRNA-mRNA interactions with relevance for the late heart development in mouse. The data also provided the basis for HeartmiR (http://heartmir.sysbiolab.eu), a publicly accessible webserver for identification, visualisation, and assessment of cardiac miRNA-mRNA interactions.

5.2 Curation and integrative analysis of potential regulatory miRNA-mRNA interaction data for HeartmiR

The data integrated in HeartmiR was obtained from several resources of potential miRNAs-mRNAs interactions and the parallel measurement of mRNA and miRNA levels in the murine heart during development. Five different publicly available datasets of miRNA-mRNA interactions were included. They are either based on experimental data[184] or computational predictions[182,185-187]. The experimentally obtained miRNA-mRNA interactions were previously collected by Chao and co-workers from 21 independent studies generating a set of 13027 with strongly supporting experimental evidence[184]. The other computationally based interaction data sets were created using different prediction algorithms. These algorithms utilise several types of molecular features combined in different ways. For instance, the predictors take into account (i) the free energy for the formation of the miRNA-mRNA interaction[182,186]; (ii) target complementary region match and mismatch[182,186,187]; (iii) target seeding sites[182,185-187]; and (iv) the presence of an homologous miRNA-binding site in the mRNA sequence[186]. Detailed description of the approaches to identify miRNA-mRNAs interactions can be found in the original publications[182,184,185,187].

Through combination of the results from these five data sets, a comprehensive miRNA-mRNA interaction dataset was obtained. However, to ensure a high level of confidence, I analysed each dataset individually and applied quality filters according to authors' recommendations:

- **microRNA.org:** Two files labelled as "Good mirSVR score, Conserved miRNA" and the "Good mirSVR score, Non-conserved miRNA" were obtained from (www.microrna.org/microrna/getDownloads.do). Interactions were subsequently merged into a single list and duplicated interactions were removed. To increase confidence all interactions that had a mirSVR score higher than -0.1 were removed following the recommendations stated in Betel *et al., 2010*[186]. After applying these criteria, 92141 miRNA-mRNA interactions were retained for integration;

- **Pita:** miRNA targets were obtained from the Pita Download Predictions Page (https://genie.weizmann.ac.il/pubs/mir07/catalogs/PITA_sites_mm9_0_0_ALL.tab.gz). A minimal score of -10 was set according to the authors' specifications indicated in the FAQ/Notes page (https://genie.weizmann.ac.il/pubs/mir07/index.html). Setting this criterion, I obtained 19615 miRNA-mRNA interactions;
- **mirDB:** The most recent version of miRDB (http://www.mirdb.org/miRDB/download/miRDB_v5.0_prediction_result.txt.gz) was downloaded. According to the authors' suggestions[187] stated in the download page, a gene with a score higher that 50 was considered to be likely a miRNA target. Using only interactions that had a score higher than 50, 68035 miRNA-mRNA interactions were retained;
- **TargetScan:** *Mus musculus* miRNA-mRNA interactions were obtained from the TargetScan (http://www.targetscan.org/mmu_71/) database by downloading (http://www.targetscan.org/mmu_71/mmu_71_data_download/Summary_Counts.all_predictions.txt.zip). Only interactions with "Cumulative weighted context++ score" lower than -0.1 were used for the analysis to increase prediction confidence as suggested by the authors (https://elifesciences.org/content/4/e05005). This cut-off resulted in 18612 interactions;
- **MirTarBase:** miRNA-mRNA interactions based on experimental data were downloaded from MirTarBase (Version 6.1). No filtering was carried out and the complete set of 13027 miRNA-mRNA interactions of the MirTarBase was used.

For all datasets (microRNA.org, Pita, miRDB, TargetScan and MirTarBase) duplicated interactions were removed and miRNA targets were mapped to their gene Symbol and Entrez Gene ID for identification.

After combining miRNA interactions, I integrated the miRNAs interactions with our mRNA and miRNA expression data for heart development in mouse. To reduce noise, only interactions were included for which the corresponding mRNAs and miRNAs had expression signals larger than 32 units for at least one time point. To simplify interpretation of interaction scores obtained for the resources, they were ranked from 1 to 100 according to their relative value for each of the computationally derived interaction sets.

Thus, interactions, with a score within the top 1%, were given the rank of 1 while interactions with scores within the top 1-2%, were attributed a rank of 2 and so on. In total, this integration procedure resulted in the identification of 102083 potential interactions between 386 miRNAs and 9211 target genes (Supplementary Table S5.2.1). Correlation between mRNAs and miRNAs was calculated using the Kendall rank correlation, which provides a more robust measure of the similarity of expression profiles than a standard Pearson correlation as it does not require normally distributed data.

Additionally, cluster analysis of the DEGs and DEmiRs was carried out using the fuzzy c-means algorithm implemented in the R/Bioconductor package Mfuzz[112]. In order to simplify interpretation, the cluster index was sorted based on the cluster profile, starting with downregulated clusters. Furthermore, GO enrichment analysis for these clusters was carried out in R based the Bioconductor annotation packages org.Mm.eg.db[114] and GOstats[115]. To reduce the number of overlapping categories detected as enriched, a hypergeometric conditioned test on the GO tree structure was applied[115]. All these methods are fully described in chapter "2. Bioinformatic Methods" in section 2.4 and 2.5.

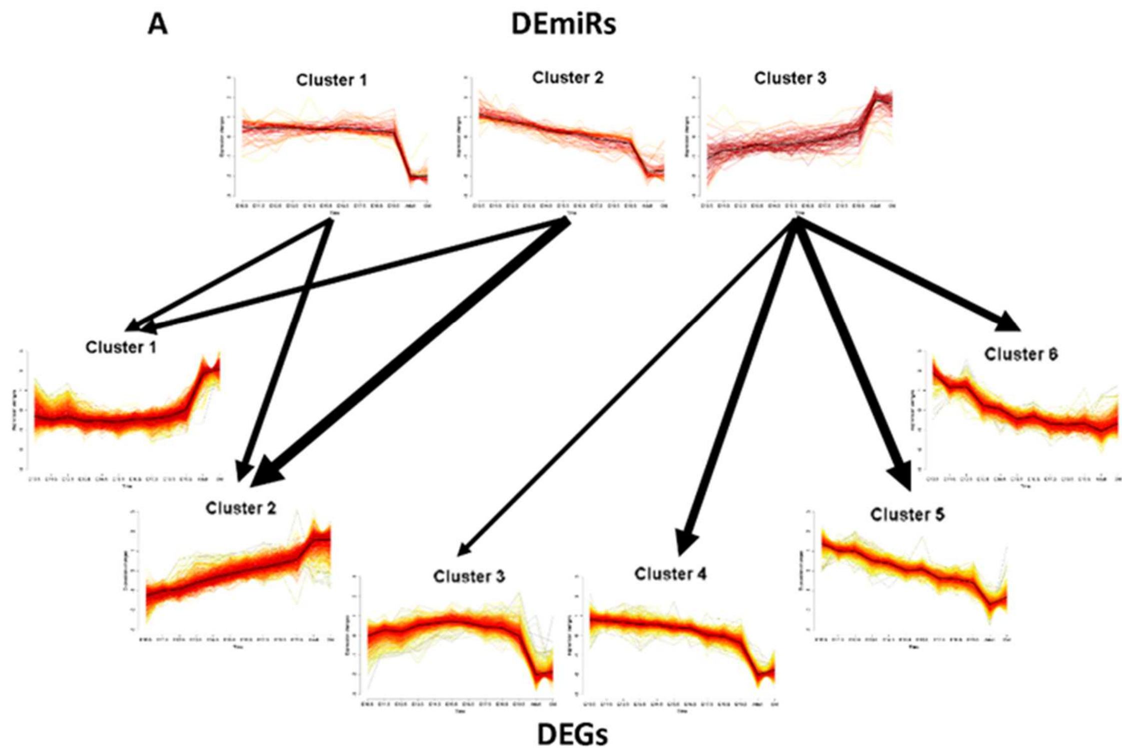
5.3 *In vivo* data dual transcriptome analysis

5.3.1 Integrative analysis of dual transcriptome data and miRNA gene targets

To cope with complexity and evaluate major expression trends, I clustered the DEGs and DEmiRs separately, based on their standardised expression profiles using fuzzy c-means clustering[113]. implemented in the R/Bioconductor package Mfuzz[112]. Expression profiles of DEGs and DEmiRs were standardised, i.e., the mean value was set to zero and the SD was scaled to one for each DEG and DEmiR. The default value of 2 for the fuzzification parameter *m* was kept.

These procedures resulted in the generation of six clusters of DEGs and three clusters of DEmiRs. To link observed miRNA and mRNA expression patterns, the cluster were connected by the integrated miRNA-mRNA interactions and the frequency of connections evaluated. More specifically, the number of putative miRNA-gene interactions for each possible pair of clusters of DEGs and DEmiRs was counted and entered in an interaction matrix ($\mathbf{M_{all}}$). As miRNAs are commonly considered to act as post-transcriptional repressors, I also determined an interaction matrix $\mathbf{M_{neg}}$, which included only interactions between miRNAs and their anti-correlated target genes with a Kendall ≤ -0.1. In this way, I combined the clustering approach and interaction data to generate a compact network of miRNA and gene clusters displayed in Figure 5.3.1. Given that in conventional models' miRNA binding leads to repression of their targets, I focused on the links between DEmiR and DEG clusters, in which putative negative regulatory interactions dominated.

DEmiRs in cluster 1 were strongly downregulated in mature heart tissue compared with samples of developing hearts. Only minor differences in their expression levels were observed between stages of embryonic development. Compared with DEmiR cluster 1, DEG cluster 1 showed the strongest anti-correlation in expression. Accordingly, DEG cluster 1 showed little change in gene expression from E10.5 to E19.5 but a considerable increase in mature tissue. Remarkably, miRNAs of DEmiR cluster 1 targeted 241 of the 643 genes included in DEG cluster 1 i.e. over one third of the genes in DEG cluster 1 are under the potential control of miRNAs in DEmiR cluster 1. A significant enrichment of target genes among biological processes associated with metabolism, such as oxidation-reduction processes (FDR = $2.0E^{-5}$) and the immune system (FDR = 0.009) was detected.



M_{all}	mRNA Cl1	mRNA Cl2	mRNA Cl3	mRNA Cl4	mRNA Cl5	mRNA Cl6
miRNA Cl1	241	179	188	377	342	281
miRNA Cl2	248	198	207	393	369	280
miRNA Cl3	231	168	188	365	338	267

M_{neg}	mRNA Cl1	mRNA Cl2	mRNA Cl3	mRNA Cl4	mRNA Cl5	mRNA Cl6
miRNA Cl1	227	170	39	79	112	134
miRNA Cl2	230	197	55	15	9	7
miRNA Cl3	34	20	166	350	329	253

Figure 5.3.1 Integrative analysis of clusters of DEmiRs and DEGs. Clusters were linked based on the relative number of miRNA-target gene interactions. Links are only displayed when dominated by interactions with negatively correlated miRNAs and target gene. The tables in the bottom panel display the total number of potential regulatory interactions between pairs of DEmiR and DEG clusters (**M_{all}**) and the number of interactions between negatively correlated miRNAs and target genes (**M_{neg}**) with a negative correlation ≤ -0.1. The width of the arrow indicates the number of anticorrelated targeted genes.

DEmiR cluster 2 showed a gradual decrease in expression during embryonic development and further downregulation in mature heart tissue. These changes in expression levels suggest a dynamic adjustment of the regulatory function of miRNAs included in DEmiR cluster 2 during embryonic development and are in contrast to miRNAs included in DEmiR cluster 1, which showed little expression change over this period. The largest number of interactions was with DEG cluster 2 that displayed anti-correlated gene expression behaviour of a gradual increase in gene expression during embryonic heart development and further upregulation in mature tissue. Notably, a large number of genes in DEG cluster 2 were targeted by miRNAs in DEmiR cluster 2, i.e., 198 out of 517 genes.

Further functional enrichment analysis of the targeted genes in DEG cluster 2 revealed a significant association with developmental biological processes, including angiogenesis (FDR = $3.7E^{-5}$) and cardiovascular system development (FDR = $9.9E^{-6}$).

Finally, the expression of most miRNAs in DEmiR cluster 3 increased during embryonic heart development and reached maximum expression levels in mature hearts. DEG clusters 3, 4, 5, and 6 showed the opposite pattern of gene expression over time, comprising genes with maximum expression during heart development and minimum expression in young and old adult hearts. Members of DEmiR cluster 3 targeted 188 genes belonging to DEG cluster 3, 365 genes of DEG cluster 4, 338 genes of DEG cluster 5, and 267 genes of DEG cluster 6. GO analysis of the targeted DEGs from clusters 3, 4, 5, and 6 indicated statistically overrepresented biological processes, such as regulation of gene expression, cell cycle (FDR = $7.3E^{-5}$), heart development (FDR = $2.8E^{-5}$), and cytoskeleton organisation (FDR = 0.0006). These results suggest that members of cluster 3 DEmiRs act as key regulators of cardiogenesis, given their potential to induce progressive downregulation of genes that were especially active during development and postnatal maturation.

Alternative to the data-driven clustering approach connecting the temporal profiles of DEGs and DEmiRs, I aimed to detect individual miRNAs regulating genes that are known to be involved in heart development and that were examined in previous chapter. To this end, I searched for miRNAs that target genes associated with heart development and show anti-correlated expression during embryonic development. I excluded mature samples in the calculation of correlation to avoid confounding effects through large postnatal expression changes that frequently were observed.

Table 5.3.1 displays anti-correlated miRNAs that target genes in the specific sub-categories shown in. For the majority of inspected genes, miRNAs with a moderately or strongly anti-correlated expression could be identified. However, I did not find indications that the same miRNA targets different genes of the same functional sub-categories. Further inspection of the role of the identified miRNAs in regulating heart developmental genes is vindicated.

Table 5.3.1 Anti-correlated miRNAs to the specific gene subcategories that are present in Chapter IV – Section 4.3.2: "Markers for heart development and function show distinct expression changes"

Gata TFs		Kendall
Gata4	miR-3474	*-0.4*
	miR-22-5p	*-0.26*
Gata3	miR-27a-3p	*-0.64*
	miR-34a-5p	*-0.4*
Cardiac TFs		
MyoCD	miR-495-3p	*-0.6*
	miR-31-3p	*-0.54*
Hand2	miR-671-5p	*-0.26*
	miR-181a-5p	*-0.18*
Nkx2-5	miR-207	*-0.02*
Forkhead TFs		
Foxc1	miR-133b-3p	*-0.56*
	miR-680	*-0.46*
Foxc2	miR-133a-3p	*-0.61*
	miR-133b-3p	*-0.61*
Foxc3	-	-
Signalling		
Tgfb2	miR-301a-3p	*-0.57*
	miR-193b-3p	*-0.52*
Wnt5a	miR-378a-5p	*-0.71*
	miR-378b	*-0.66*
Signalling		
Bmp2	let-7b-5p	*-0.48*
	miR-27a-3p	*-0.44*
Bmpr1a	miR-194-5p	*-0.72*
	miR-362-3p	*-0.46*
Contraction		
Myh7	-	-
Actc1	miR-30a-5p	*-0.46*
	miR-30e-5p	*-0.44*
Tnni3	-	-
Tcap	miR-128-3p	*-0.23*
	miR-324-3p	*-0.17*

Hdac		Kendall
Hdac2	-	-
Ca^{2+} Channels		
Cacna2d2	miR-490-3p	-0.67
	miR-671-5p	-0.13
Cacna1h	miR-28a-3p	-0.63
	miR-466a-3p	-0.6
Cacna1g	miR-139-5p	-0.78
	miR-187-3p	-0.69
Na^{+} Channel		
Scn7a	-	-
K^{+} Channel		
Kcna5	miR-450b-3p	-0.2
Kcnb1	miR-409-3p	-0.8
	miR-17-5p	-0.6
K^{+} Channel		
Kcnq5	miR-292-5p	-0.54
	miR-324-3p	-0.41
Kcne1	miR-466g	-0.55
	miR-466f-3p	-0.42
Kcna4	miR-193b-3p	-0.46
	miR-181a-5p	-0.12
Cl^{-} Channel		
Clcn5	miR-27a-3p	-0.62
	miR-194-5p	-0.58
Clca3a1	-	-
Clic5	miR-182-5p	-0.53
	miR-485-5p	-0.45
Cell Cycle		
Cdc27	miR-362-5p	-0.46
	miR-195a-5p	-0.45
Cdca3	miR-195a-5p	-0.48
Cdc20	-	-
Rgcc	-	-

The microarray experiment revealed a surprisingly large number of DEmiRs with the vast majority of them not linked to heart development or cardiac tissue maturation yet. To assist future studies of their functional relevance, I applied several complementary approaches to prioritise miRNAs:

1. Analysis of the correlation between miRNAs and their targets to capture their regulatory activity: Assuming that the dominant model of miRNAs is post-transcriptional repression, we can expect that target genes show a negative correlation with miRNA expression profiles. Conversely, it should then be possible to evaluate the regulatory activity of a miRNA by calculating how many of its differentially expressed target genes (DETGs) are negatively correlated (DETGNC). Here, I classified target genes of miRNAs as negatively correlated if the Kendall's correlation coefficient of the miRNA and target gene was smaller than ≤ -0.4. Target genes were also classified as associated with heart development or with transcriptional regulation based on their current GO annotation[147]. This data can be found in Table S5.3.1. Plotting the number of DETGs and DETGNCs against each other for the different miRNAs (Figure 5.3.2) remarkably showed that many of the miRNAs having a high percentage of negatively correlated target genes are already known to be involved in the regulation of heart development or function.

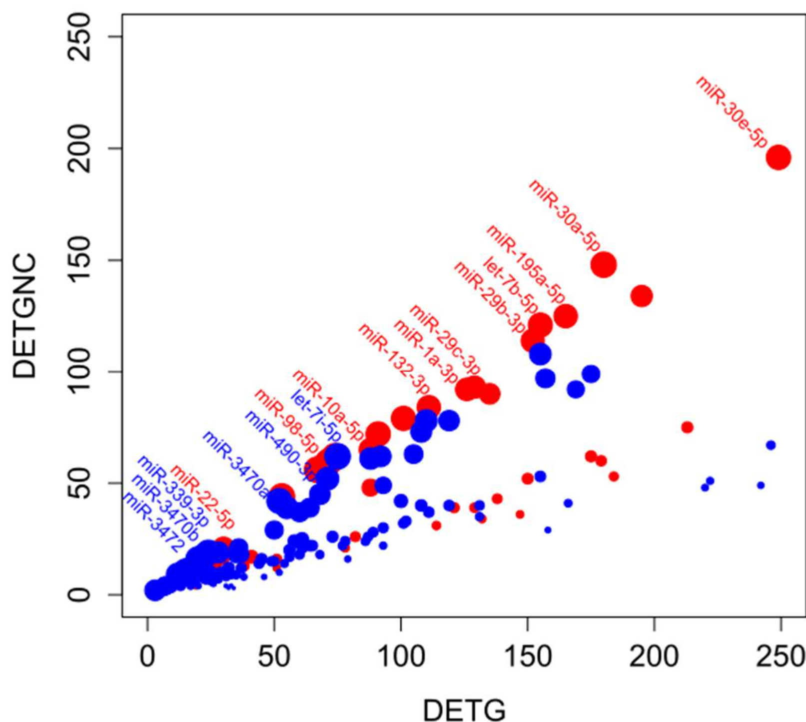


Figure 5.3.2 Prioritisation of cardiac miRNAs by regulatory activity. Number of differentially expressed targets (DETGs) and differentially expressed genes that are negatively correlated (DETGNCs). Red dots label miRNAs that have been previously associated with heart development. Blue dots label miRNAs that have not been associated with heart development to date. Dot size indicates the percentage of anti-correlated targets, compared to all differentially expressed targets.

For instance, at least 70% of the DETGs of miR-1a-3p and miR-195a-5p (a member of the miR-15 family) were negatively correlated. This supports the conjecture that the percentage of DETGNCs could be indicative of a miRNA's regulatory activity. The highest percentage of DETGNCs (~83%) was recorded for let-7f-5p, let-7g-5p, and miR-499-5p, which belong to the so-called intronic myomiRs. Based on their high percentage of DETGNCs, I also identified miRNAs that are not yet linked to heart development, namely let-7i (82%), mir-3472 (79%), and miR-490-3p (73%; Figure 5.3.2). These miRNAs can provide attractive leads for further studies.

2. Evaluation of the existing functional annotation of the targets: Another approach of prioritisation is to consider the regulatory capacity of the target genes. Of particular interest are transcriptional regulators as miRNA targets since their interaction can have wide ranging indirect effects through controlling the abundance levels of the targeted regulators. Thus, I inspected the number of transcriptional regulators among the targets of miRNAs. I found 13 miRNAs that target 10 or more transcriptional regulators that showed anti-correlated expression patterns. Differential expression of these miRNAs may have an especially broad effect on gene expression at a systems level. Strikingly, miR-1a-3p was the miRNA with the largest number of anti-correlated transcription regulators among its potential targets. Among the 18 transcription regulators possibly targeted by miR-1a-3p were six TFs (Foxp1, Gata6, *Hand2*, *Myocd*, Rarb, and Srf) associated with heart development, supporting a key role for miR-1 in the cardiac cell lineage (Figure 5.3.3).

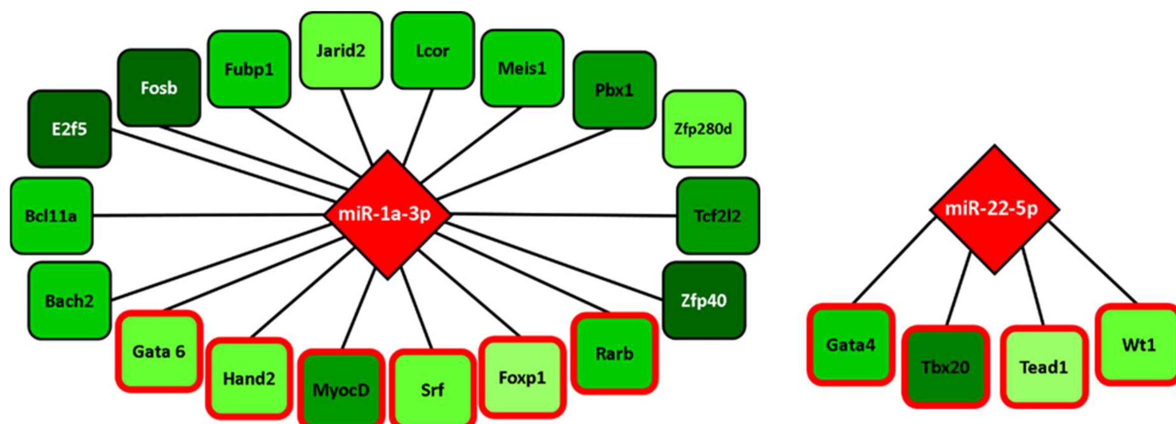


Figure 5.3.3 Prioritisation of cardiac miRNAs by functional annotation analysis. Examples of two miRNAs (symbolised as diamonds) targeting TFs are shown (squares). Red borders highlight TFs associated with heart development. The colours of the symbols indicate expression changes in young mature tissue (shades of red: upregulated; shades of green: downregulated).

All of the targeted TFs were found to be downregulated in the mature heart, while miR-1a-3p expression was increased. Other miRNAs targeting multiple transcriptional regulators linked to heart development included miR-362-3p, miR-132-3p, miR-22-3p, miR-22-5p, and miR-221-3p. Four of the six transcriptional regulators potentially targeted by miR-22-5p are associated with heart development in GO, indicating that downstream regulatory effects of miR-22-5p may be specific to cardiogenesis (Figure 5.3.3). This approach not only prioritised miRNAs, but also elucidated their role in the gene regulatory networks that drive cardiogenesis.

3. Template-based detection of relevant miRNAs: As a third procedure for selecting relevant miRNAs, I analysed in more detail the expression patterns to find miRNA which showed clear anticorrelation with relevant target genes. To this end, I filtered putative interactions by requiring that both miRNAs and target genes show at least a 4-fold differential expression change at one time point and were anti-correlated (Kendall correlation < -0.4). I only included target genes that were already associated with heart development in GO, so that the functional relevance of the miRNA interaction can be readily established. Supplementary Table S5.3.2 lists 51 miRNAs that were identified as targeting 48 genes annotated as heart developmental genes. Given these stringent thresholds, I retrieved only a few miRNAs that targeted several genes (e.g., miR-30e-5p targeted eight genes); the majority targeted only a single gene. However, miR-22-5p targeted both Tbx20 and *Gata4*, i.e., two TFs that are essential for heart development (Figure 5.3.4)[178]. A literature review for these 51 miRNAs showed that 24 (47%) are currently linked to cardiogenesis or heart functions (Table S5.3.2). In this context, their identified putative interactions with heart developmental genes may indicate a specific role in cardiogenic regulation. Thus, I suggest that the remaining 27 miRNAs should be considered as candidates for cardiac regulators. These include miR-139-3p targeting *Dicer1*, *Nedd4*, *Sox11*, *and Ednra*, as well as miR-540-3p targeting *Nrp2*, *Caxadr*, *Heg1*, and *Ttn* (Figure 5.3.4).

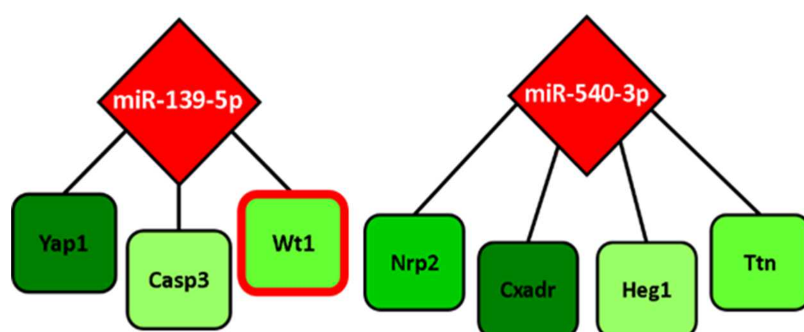


Figure 5.3.4 Template-based prioritisation of cardiac miRNAs. Examples of candidates of differentially expressed miRNAs (DEmiRs) targeting differentially expressed genes (DEGs) associated with heart development are shown.

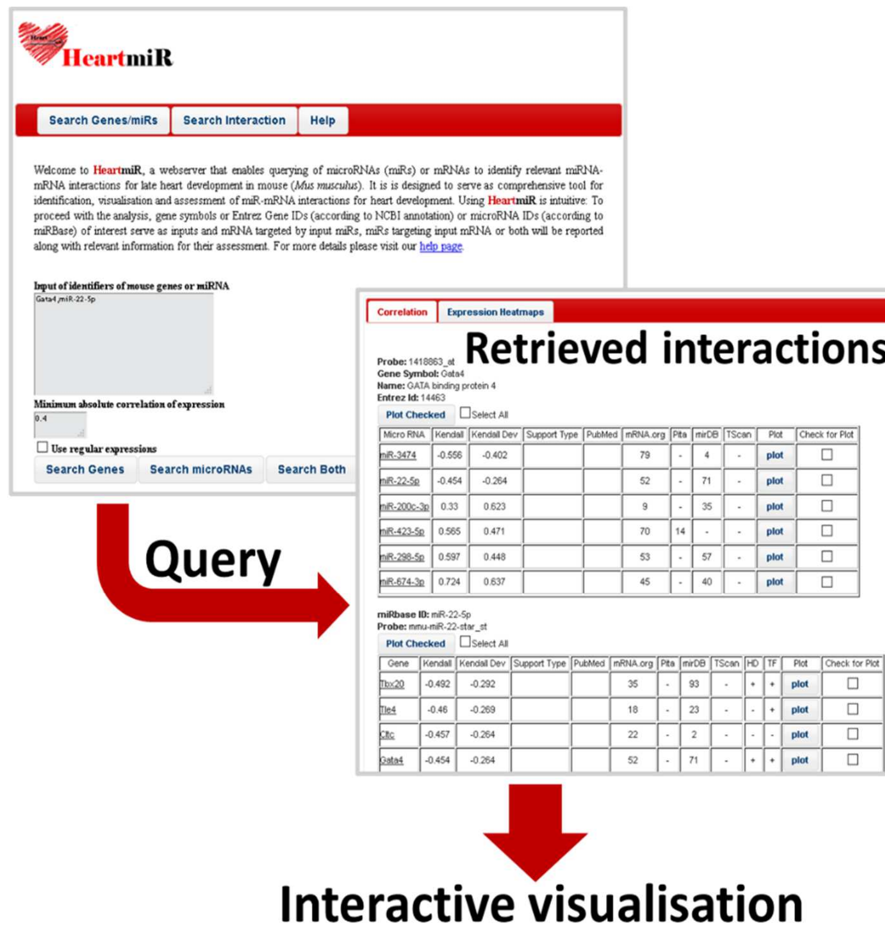
5.4 HeartmiR structure and querying

We implemented a database called HeartmiR, which is freely accessible at http://heartmir.sysbiolab.eu. It aims to be a comprehensive tool for query, visualisation, and identification of regulatory interactions of miRNAs with potential relevance in heart development and maturation. Its use is intuitive and is outlined in Figure 8.

For querying HeartmiR, three main options are available: (i) search by a set of genes or miRNAs; (ii) search for a specific interaction and; (iii) templated-based search. The search by genes/miRNAs displays a web page where mRNA or microRNA identifiers can be used as input. Currently, mRNAs can be searched by Gene Name, Gene Symbol, Entrez Gene ID (according to NCBI annotation) and by the Affymetrix ID. The miRNAs can be searched by miRBase ID, Affymetrix ID and Transcript ID (Figure 5.4.1).

Additionally, a threshold for correlation can be set, resulting in the exclusion of interactions with low absolute correlation of expression between miRNAs and their target genes. Correlation of expression is measured using Kendall's tau coefficient, as we found that Pearson correlation tended to be overly sensitive to large expression changes occurring between E19.5 and the young adult stage. Two Kendall correlation coefficients can be used for filtering: one based the complete time series and one based on developmental time points only ("Kendall Dev"). The latter enables a more sensitive examination of how strongly the expression of miRNAs and their targets are (anti-) correlated during embryonic development (E10.5–E19.5). The user can set a threshold for absolute correlation, resulting in the exclusion of interactions with lower correlation between miRNA and the corresponding target mRNA. Thus, using for instance, a threshold of 0.4, only miRNA-mRNA interactions that have a Kendall correlation higher than 0.4 or lower than -0.4 will be shown. By default, this value is set to 0.4. The maximal value of the Kendall correlation could be +1 (perfect correlation) or -1 (perfect anti-correlation) (Figure 5.4.1).

Once the search parameters are set, the user can search either for genes or miRNAs or simultaneously for both genes and miRNAs that match the input identifiers in the database.



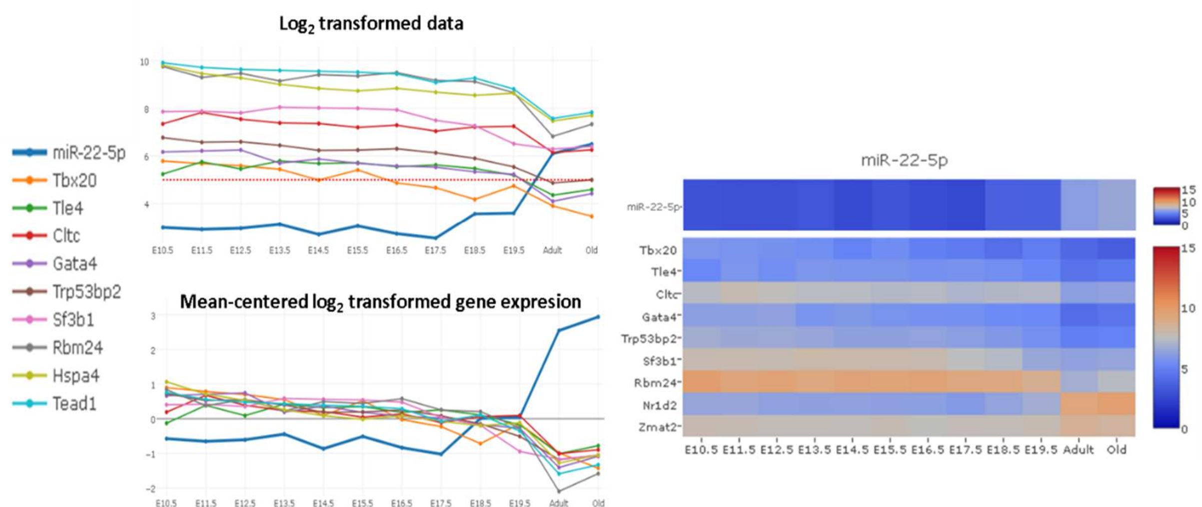


Figure 5.4.1 Workflow scheme for HeartmiR. Following the input of miRNA or gene identifiers (here: *Gata4* and mirR-22-5p), HeartmiR can be queried for putative miRNA-target gene interactions. All retrieved interactions are listed in tables, along with additional information. Subsequently, miRNA (mirR-22-5p) and target genes (Tbx20 and *Gata4*) can be selected and their profiles visualised and interactively analysed. Gene expression data are plotted as log2 intensities with or without mean centring. The former provides a more informative presentation of the absolute expression levels, whereas the latter shows expression changes of transcripts and their (anti-) correlation more clearly.

The results of all queries are displayed in table format (Figure 5.4.1 – Retrieved Interactions). For each of the queried genes and miRNAs, a separate table presents these interactions, along with additional information. For experimentally derived interactions, the relevant PubMed references are given. To facilitate assessment of predicted interactions, their scores were ranked and converted to percentiles for each resource, separately. Thus, values between 1 and 100 are displayed for computationally predicted interactions included in HeartmiR. For gene targets, the result table also indicates whether they were assigned to processes related to heart development (GO: 0007507) or transcription regulation (GO: 0003700) in GO. All tables can be interactively sorted and filtered by increasing the threshold for absolute correlation. Finally, the expression profiles of miRNAs and target genes can be visualised as interactive line plots or heatmaps (Figure 5.4.1 – Interactive visualisation).

5.5 Conclusion

One reason for the incomplete understanding of heart development lies in the lack of a comprehensive knowledge of how miRNAs contribute to the development of a fully mature and functional heart. As miRNAs are an essential class of gene expression regulators, it is important to gain insights in their actions and roles during heart development. This may also provide us with crucial cues for regenerative treatment of heart diseases. While some miRNAs have already been associated with heart-specific expression, a comprehensive study of the effects of miRNAs in the developing mouse heart had not been undertaken. The integration of transcriptome data and miRNAs targets led to the identification of 102083 potential interactions between 386 miRNAs and 9211 target genes. To enable other research to explore the collected data and to gain greater insights in heart development, HeartmiR was developed. It is constituted by the integration of known or predicted miRNA-mRNA interactions with expression data from the *in vivo* experiment described in the previous chapter.

The integrated data helped to elucidate some of the miRNA-mRNA regulatory interactions that may occur during cardiogenesis. Exploiting the curated information, I identified several miRNAs that were be already reported to be associated with heart development across different studies[55,177], and more interestingly, I discovered novel candidates which warrant further exploration[188].

To gain an overview of the temporal dynamics, the DEmiRs and DEGs were clustered based on their expression profiles and posteriorly performed enrichment analysis. Through Gene Ontology enrichment analysis, it was possible to assign miRNA clusters to potential roles in heart development. To complement this data driven clustering approach that connects temporal expression profiles of DEGs and DEmiRs, I inspected individual miRNAs that target specific genes known to be involved in cardiogenesis. Accordingly, I carried out three different complementary approaches for prioritisation of miRNA candidates for future validation. Overall, this integrative analysis of miRNAs expression profiles and miRNAs targets led to identification of 27 miRNA candidates.

Finally, we developed an open access database and interactive visualisation tool called HeartmiR that facilities the identification of anti-correlated gene-miRNA expression patters during the developing heart, in young and old adult mice heart tissues. Besides being a platform to explore gene regulation by miRNAs, HeartmiR can be useful for a wider scientific community, especially for developmental and clinical follow-up investigations in the context of heart development, congenital heart diseases or cardiac regenerative medicine.

Chapter VI

Detection of novel potential regulators of stem cell differentiation and cardiogenesis through combined genome-wide profiling of protein coding transcripts and microRNA

6.1 Introduction

ESCs are derived from the inner cell mass and differentiate into different specialised cell lineages, such as, endoderm, ectoderm and mesoderm[189]. In later stages, these cells will further specialise into somatic cell types. Using the hanging drop technique, it is possible to simulate the early embryonic development process, since the cells aggregate in one point of the droplet through gravitational force, allowing for these group of cells to differentiate and form embryoid bodies. The embryoid bodies contain non-patterned and organised cell clusters that retain organoids characteristics such as multiple specific cell types with the capability of recapitulating some specific organ functions[190]. To understand these differentiation processes, functional genomic technologies, such as, high throughput assays and microarrays have contributed to identify key transcription regulators, such as transcription factors, non-coding mRNAs and miRNAs[191,192]. Besides the transcriptional regulation, miRNAs commonly regulate gene expression at post-transcriptional level through imperfect base pairing to target mRNAs[52], leading to translational inhibition and/or mRNA degradation[193]. The study of *in vitro* differentiation remains important due to the ease of experimental set up, the possibility of rapid upscaling compared to *in vivo* experiments and avoidance of ethical considerations. Furthermore, achieving *in vitro* differentiation that faithfully reflects *in vivo* developmental processes will be crucial for many applications of stem cell biology in a clinical or pharmaceutical context.

Despite evidence of the impact of miRNAs on ESCs differentiation[194], no comprehensive profiling of their parallel expression of both miRNAs and mRNAs during *in vitro* differentiation has been undertaken to date. To address this salient lack of knowledge, we performed parallel genome-wide profiling of miRNAs and genes from differentiating ESCs from day 0, day 4, day 8, day 12 and day 18. Profiles of mRNA and miRNA were analysed independently, and their differential expression was determined. A focus was set on the determination of miRNAs displaying characteristic peaks indicting potential roles as molecular switches. Additionally, due to the stem cell ability to differentiate into cardiac lineages, it was within our interest to unveil novel markers that might be present both in *in vitro* differentiation and in *in vivo* during heart development[94,141,195]. Therefore, I performed comparative analysis to connect *in vitro* stem cell differentiation with *in vivo* heart development. In particular, I attempted to cross validate genes and miRNAs indicated in Chapter IV in the *in vitro* settings.

This analysis revealed potential novel genes and miRNAs that might be drivers for cardiac differentiation in an *in vitro* environment. These genes and miRNAs might provide novel vehicles to guide *in vitro* cardiac differentiation that more accurately reflects developmental processes *in vivo*.

6.2 Experimental design

This experimental design aimed to provide a biological insight of the expression of mRNAs and miRNAs during stem cell differentiation, by using the hanging drop method in order to obtain EB structures. To this end, 10µl drops of ESCs suspension (approx. concentration of 2500 cell per drop) were placed on the inner surface of a Petri dish lid[196]. Drops were placed sufficiently apart so as not to touch each other. The Petri dishes contained 5ml of PBS in the bottom to act as a hydration chamber when the lid is closed. Cells were incubated under normal culture conditions. Cells were monitored daily to observe when they form cell aggregates. Besides the ESCs collected at day 0, it was also collected at days 4, 8, 12 and 16 (Figure 6.2.1). Cells were detached from the petri dish lid using 0.05% trypsin/2 mM calcium to preserve cadherin function. All samples were collected for RNA extraction. Total RNA from these time points were taken for the transcriptome study and for each time point, biological triplicates were made. The microarray labelling and hybridisation techniques were performed as previously reported[151]. The experiments were carried out by our collaboration partners at the University of Cologne.

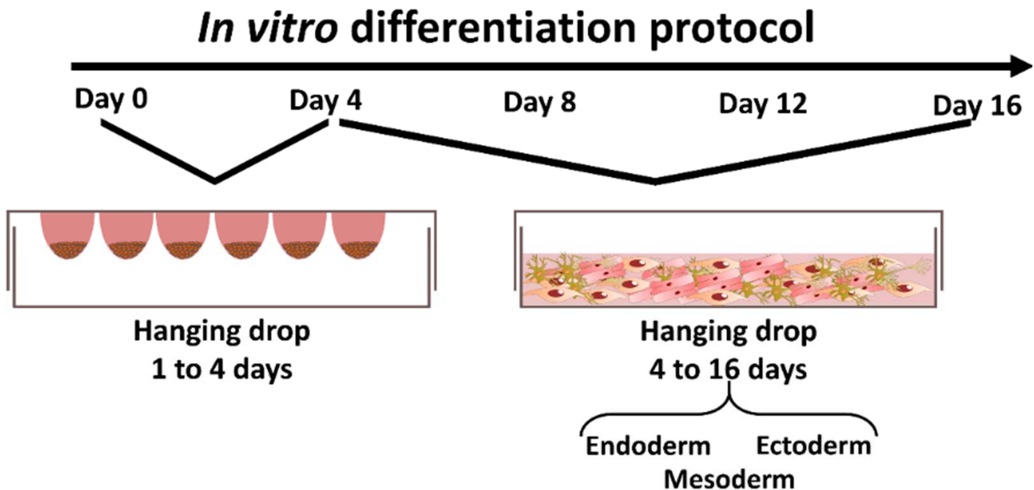


Figure 6.2.1 Time series experiment layout, to profile simultaneously mRNAs and miRNAs during *in vitro* stem cell differentiation.

6.3 Results and discussion

6.3.1 Differential gene expression during *in vitro* differentiation

In total, fifteen samples were profiled using Affymetrix Mouse Genome 430 2.0 Arrays and Affymetrix miRNA 3.0 GeneChips. To gain a first overview of the expression changes measured in the experiment, I clustered the full mRNAs and miRNAs microarray profiles and carried out a PCA analysis. Samples taken at the same time point were in proximity to each other while samples from different time points were placed apart indicating the reproducible transcriptomic profiles and a robust temporal expression signature (Figure 6.3.1A and B). The largest differences between time points were observed for the first half of the time series i.e. between day 0 and day 8. After day 8, the variation diminished and the transcriptome profiles for mRNAs and miRNAs tended to stabilise.

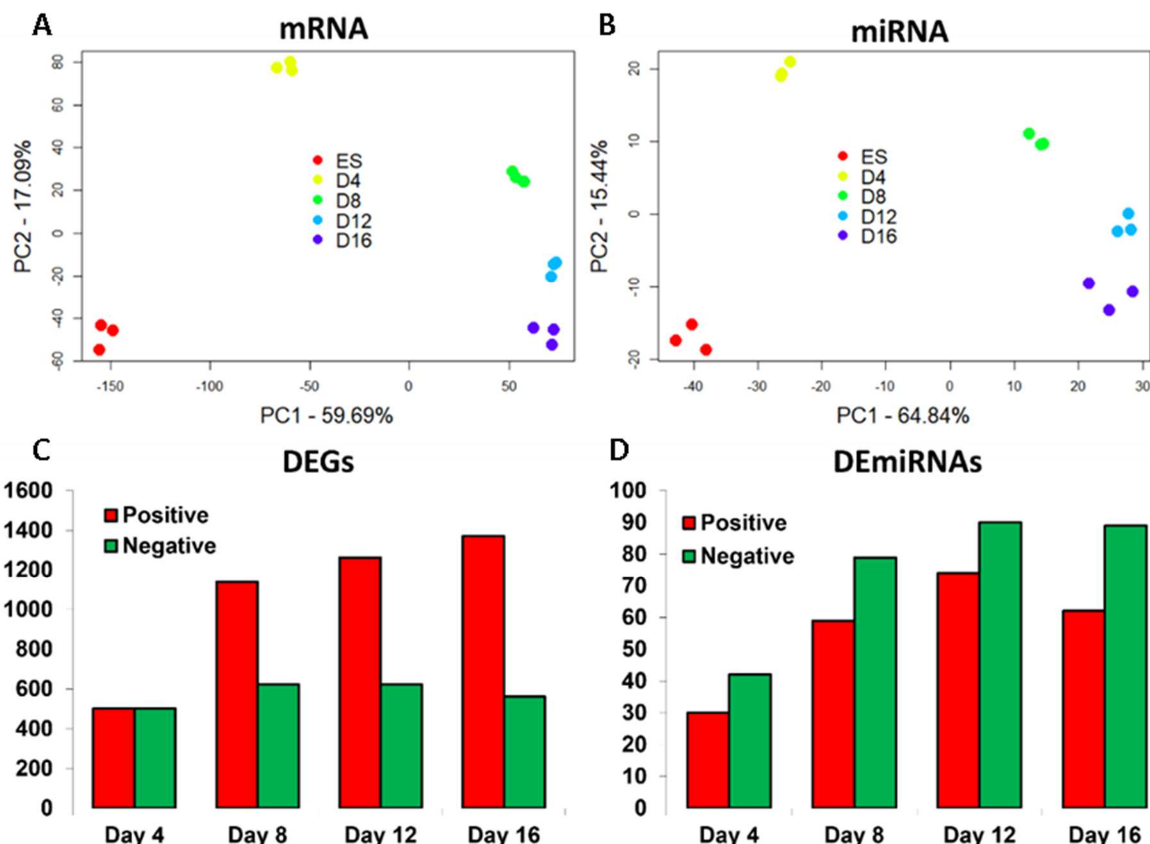


Figure 6.3.1 mRNA and miRNA profile analysis of *in vitro* ESCs differentiation. (A, B) Principal component analysis of mRNA and miRNAs profiles. The two principal components capturing the largest variance for mRNA and miRNA, respectively. (C, D) Number of differentially expressed mRNAs and miRNAs with positive or negative changes in expression compared to day 0.

To identify DEGs, transcribed as mRNAs, normalised Affymetrix GeneChip signal intensities measured at day 4, 8, 12 and 16 were compared to intensities of day 0.

A threshold for differential expression was set with an adjusted p-value of $\leq 10^{-5}$ and an absolute log2 fold change of ≥ 2 (4-fold change). The stringent threshold was motivated by the large number of genes that displayed changes in expression. In total, 2718 non-redundant genes were detected as differentially expressed for at least one time point, representing a total of 13% of all genes covered by the array. The number of DEGs doubled from 1003 at day 4 to 1933 at day 16 (Figure 6.3.1C). Notably, I found that considerably more genes were upregulated than downregulated at day 8, 12, and 16 compared to day 0. For day 4, no such tendency was observed. A full list of DEGs at each time point is provided in Table S6.3.1.

For detection of DEmiRs, I used the same thresholds and reference time point as for the identification of DEGs. This led to the detection of 196 DEmiRs in total. Similarly to DEGs, the number of DEmiRs increased during ongoing differentiation from 41 DEmiRs on day 4 to 122 DEmiRs on day 16 (Figure 6.3.1D). Contrary to DEGs, more negatively than positively regulated miRNAs were detected. The list of DEmiRs can be found in Table S6.3.2.

6.3.2 Dynamic expression of marker genes and miRNAs

Differentiation of ESCs is characterised by the diminishing expression of master regulators of pluripotency and the sequential activation of transcription factors establishing specific germ layers and cell lineages[197]. To validate whether our data faithfully reproduces these features, I inspected the expression of known marker genes. To facilitate this analysis, transcript values were linearly transformed into a 0 to 1 scale (Min-Max normalisation). As pluripotency makers, *Nanog*, *Sox2*, *Pou5f1* (*Oct4*), *Klf4* and *Lefty1* were chosen[139,197]. For these markers, maximum expression was observed at day 0 followed by rapid decline in expression (Figure 6.3.2A). At day 4, none of the pluripotency markers showed more than 10% of its expression level measured at day 0 demonstrating an efficient abolishing of the pluripotency state by the hanging drop method. Subsequently, I assessed the expression of transcription factors associated with the formation of the three germ layers: *Otx2* and *Fgf8* for ectoderm[198,199], *T-Brachyury*, *Eomes* and *Mesp1* for mesoderm[77,200], and *Gsc*, *Sox17* and *Mixl1* for endoderm[33].

Most of the selected transcription factor exhibited a prominent peak in expression at day 4 and lower expression at the time points before and after indicating early specification into germ lineages in our experiments (Figure 6.3.2B-D). Supplementary data for the mRNA's expression values can be found in Table S6.3.3.

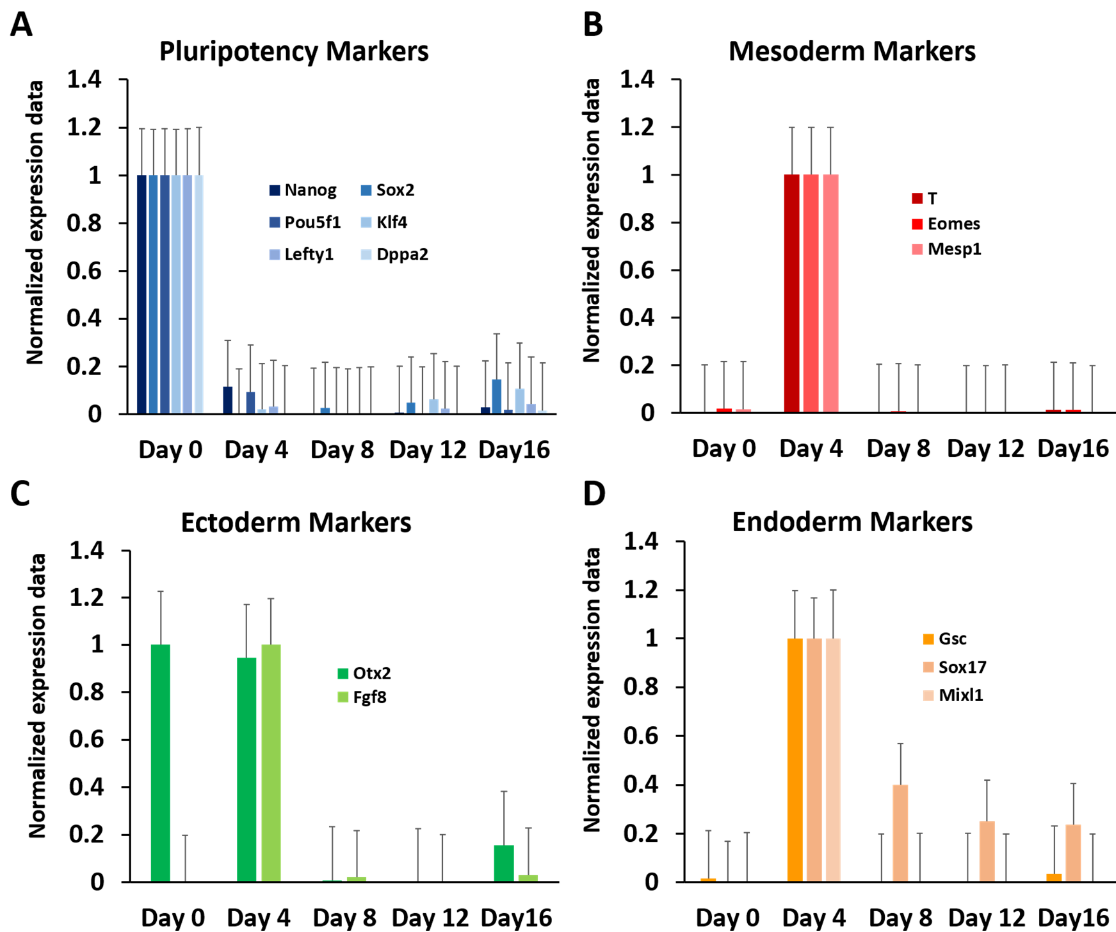


Figure 6.3.2 Pluripotency and differentiation gene markers. (A) Pluripotency markers such as *Nanog*, *Pou5f1* and *Sox2* were used to verify cells pluripotency state at day 0; **(B-D)** Mesoderm, ectoderm and endoderm markers were used to verify if the three main germ layers were present in the differentiating cells.

To assess the data quality of miRNA measurements I inspected the expression values of miRNA that have previously been linked to stem cell maintenance and differentiation. In particular, I examined the expression patterns of miR-290a-3p, miR-290a-5p, miR-295-5p, miR-363-3p and miR-363-5p as they have been associated with stem cell maintenance and proliferation[71,201,202]. Similar to the pluripotency transcription factors, our data shows that they are strongly expressed at day 0 but subsequently show a gradual decrease in expression (Figure 6.3.3A). Supplementary data for the miRNAs expression values can be found in Table S6.3.3.

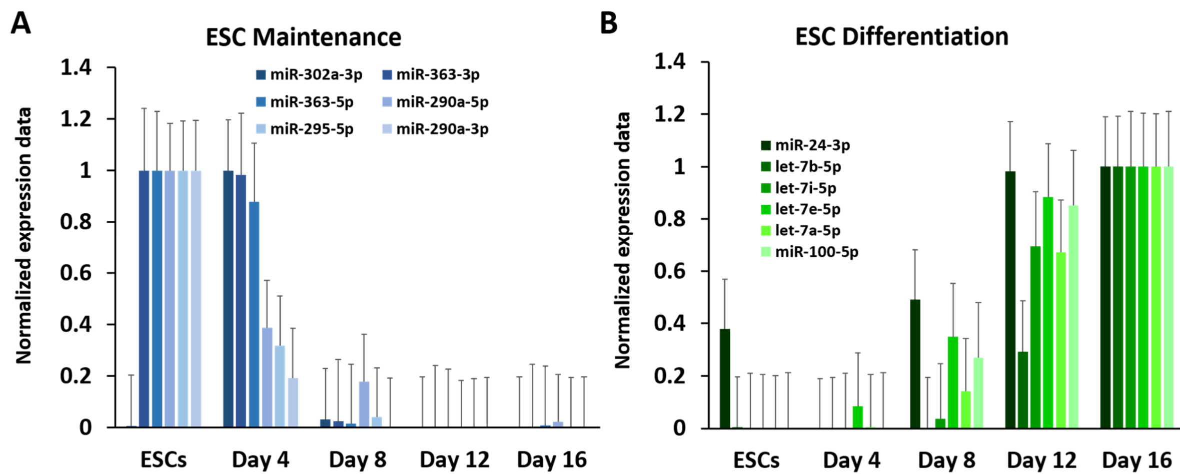


Figure 6.3.3 Pluripotency and differentiation miRNAs markers. (A) Pluripotency miRNAs markers that have been described in literature to be important for stem cell pluripotency maintenance; **(B)** Differentiation miRNA markers that have been described in literature to be important in generic embryogenesis.

In contrary, miR-24-3p that has been reported to impinge on stem cell differentiation[203] and Let-7 family members that are highly involved in embryogenesis[204], brain development[205] and hematopoietic stem cell fate[206]. displayed a gradual increase in expression along the timeline and reached maximum abundance at day 16 (Figure 6.3.3B). Thus, both mRNA and miRNA levels measured in our microarray experiment agreed well with results expected based on previous expression studies.

6.3.3 Identification of miRNA that might serve as genetic switches

A striking feature of many key transcription factors for pluripotency maintenance or stem cell differentiation is a well-defined transient activity which is displayed as sharp expression peaks at a specific time point in our time series experiments. For instance, *Mesp1* was expressed 8 times more at day 4 compared to any other time points. Such binary or 'on-off' expression patterns might reflect the role of these transcription factors as genetic switches during stem cell differentiation with key importance for embryonic development. The observation motivated me to examine whether such patterns also exist in the miRNA time series data as they could point to miRNAs that might equally serve as genetic switches.

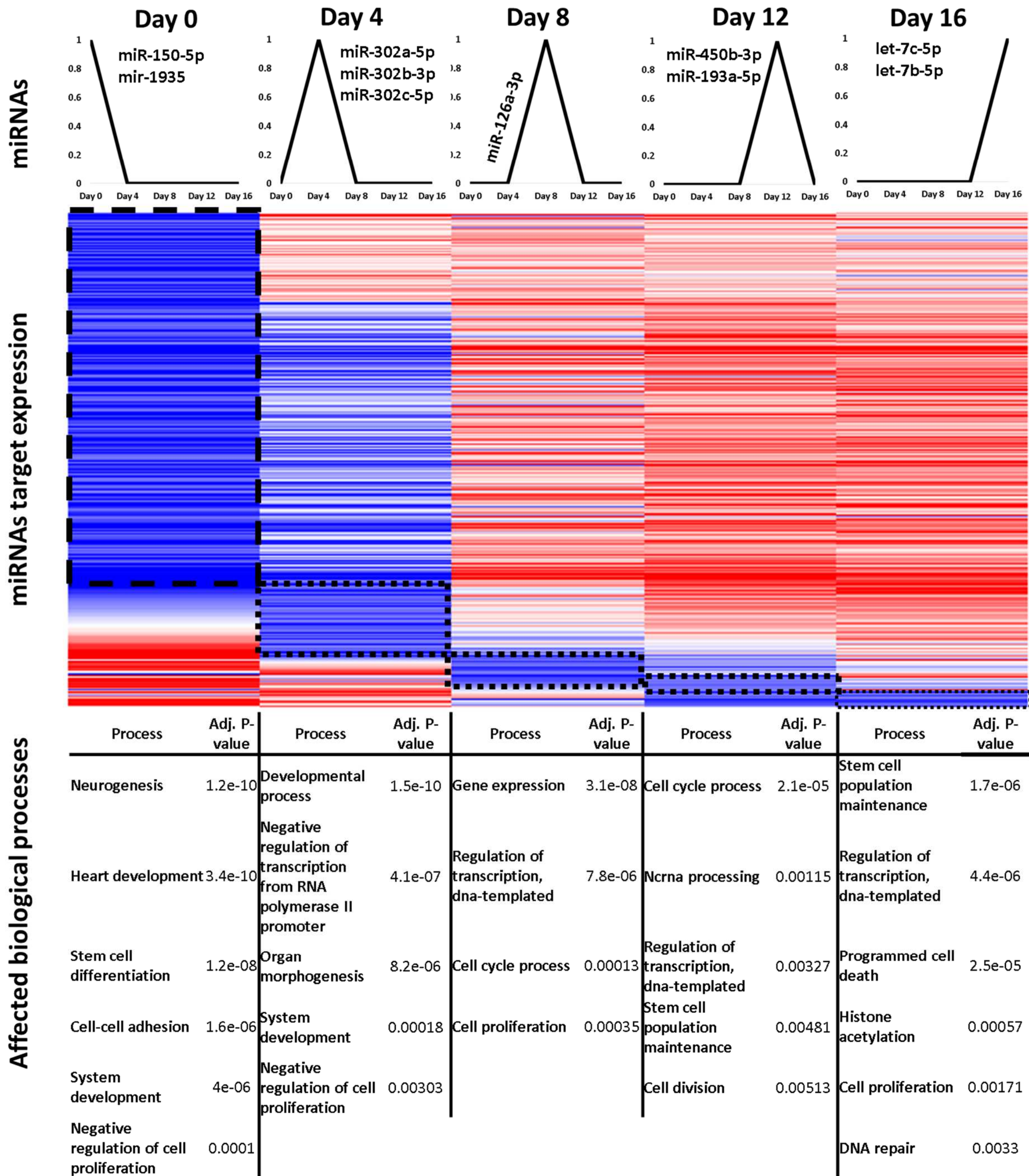
To identify prominent peaks, the correlation between miRNA expression values and binary vectors were calculated. The later had the length of the time series and included "1s" and "0s". "1" represent the peak expression and "0" is the base expression.

To find miRNAs peaking at day 0, I correlated the observed expression values with the binary vector (1,0,0,0,0). A large correlation coefficient indicates that a miRNA displays a peak at day 0. Similarly, expression values were correlated with the vector (0,1,0,0,0) to find miRNAs peaking at day 4. Corresponding binary vectors were used for the other time points. The results of the analysis can be found in supplementary Table 6.3.4. Requiring a minimum correlation of 0.6 with the aforementioned binary vector, the search for miRNAs with prominent peak patterns led to the identification of 88 miRNAs with an expression peak at day 0, 16 miRNAs at day 4, 7 miRNAs at day 8, 4 miRNAs at day 12 and 13 at day 16 (Figure 6.3.4, top panel – Table S6.3.4).

For each peak expression pattern displayed in the top panel of Figure 6.3.4, some of the correlated miRNAs are listed. Interestingly, several of these miRNAs have already been associated with development processes or used as biomarkers for the identification of different pathologies. For instance, miR-150-5p presents a peak expression at day 0. This miRNA has been associated with pathologies such as (i) cancer, where a capability of repressing tumour suppressor P53 and promoting uncontrolled cell proliferation was found for colon cancer[207] and for breast cancer[208], or (ii) advanced heart failure, where it can serve as a biomarker[209].

Other notable miRNAs belong to the miR-302 family and showed a prominent peak at day 4 in our time series. This miRNA family has been described to be highly active during the onset of differentiation, since these miRNAs govern a wide variety of functions during development. For instance, the miR-302 family is known to target several genes that regulate cell cycle[210]. Furthermore, it has been stated that the decreased expression of members of this miRNA family can have an effect in early embryonic development[210].

In the context of heart development, miRNAs with peaks at day 8, 12 and 16 seem especially relevant. For instance, miR-126-3p peaking at day 8 was identified as a potential biomarker for coronary artery disease[211]; miR193a-5p with a peak at day 12 was shown to repress proliferation and migration of cardiac stem cells (CSCs) through the downregulation of c-kit in c-kit positive CSCs[212]; while members of the let-7 family were detected for day 16 and were previously linked to the regulation of multiple processes including heart development and cardiac differentiation as well as cardiovascular diseases[213].



Process	Adj. P-value	Process	Adj. P-value	Process	Adj. P-value	Process	Adj. P-value	Process	Adj. P-value
Neurogenesis	1.2e-10	Developmental process	1.5e-10	Gene expression	3.1e-08	Cell cycle process	2.1e-05	Stem cell population maintenance	1.7e-06
Heart development	3.4e-10	Negative regulation of transcription from RNA polymerase II promoter	4.1e-07	Regulation of transcription, dna-templated	7.8e-06	Ncrna processing	0.00115	Regulation of transcription, dna-templated	4.4e-06
Stem cell differentiation	1.2e-08	Organ morphogenesis	8.2e-06	Cell cycle process	0.00013	Regulation of transcription, dna-templated	0.00327	Programmed cell death	2.5e-05
Cell-cell adhesion	1.6e-06	System development	0.00018	Cell proliferation	0.00035	Stem cell population maintenance	0.00481	Histone acetylation	0.00057
System development	4e-06	Negative regulation of cell proliferation	0.00303			Cell division	0.00513	Cell proliferation	0.00171
Negative regulation of cell proliferation	0.0001							DNA repair	0.0033

Figure 6.3.4 miRNAs target expression profiles. (Top panel) Peak patterns used to identify miRNA by correlation analysis. A selection of identified miRNAs are shown; (Middle panel) Heatmap for expression data of gene targets of the top 10 miRNA. Shades of blue indicate expression values smaller than the gene-wise average of expression, while shades of red indicate expression values greater than the average expression (Lower panel) Results of GO enrichment analysis of the top 10 miRNAs targets.

To gain further insight into the potential roles of the identified miRNA, I analysed the functions of their targets based on the high-confidence miRNA-mRNA interactions integrated in the HeartmiR database (Chapter 5). To this end, I calculated the Kendall rank correlation coefficient τ for the gene expression levels of each miRNA-mRNA interaction and retained only miRNA-mRNA interactions with anti-correlation of expression ($\tau < -0.2$), as we assume effective regulation by miRNAs should be reflected by diminished mRNA levels of the target genes. For visualisation as heatmaps in Figure 6.3.4, expression values of a target gene were subtracted by their average. For each time point, GO enrichment analysis was subsequently performed for the 10 miRNA with the most prominent peak patterns i.e. the 10 miRNAs showing the highest correlation with the binary vectors (Figure 6.3.4, additional information in supplementary Table S6.3.5).

This approach revealed a remarkable sequential order of repressed processes. Day 0 exhibits the greatest number of negatively regulated targets, suggesting wide-ranging effects of the selected miRNAs through repressing genes that might cause stem cells to differentiate. Indeed, miRNAs with peak expression at day 0 tended to silence processes related to morphogenesis and development such as neurogenesis or heart development. Furthermore, cell-cell adhesion and negative regulation of cell proliferation were preferentially targeted by the selected miRNA. Most of these processes were still affected by miRNA with peak expression at day 4. The set of potentially repressed processes however changed drastically from day 8 onwards. For the later timepoints, I observed a strong tendency of the selected miRNA to target genes associated with cell cycle, cell proliferation, transcription, and stem cell maintenance. Intriguingly, ncRNA processing is targeted by miRNA peaking at day 12 suggesting potential negative feedback control on themselves itself. Finally, histone acetylation which leads to opening of chromatin structures, appeared as significantly suppressed at day 16.

6.3.4 Embryonic bodies as *in vitro* model for cardiogenesis on transcriptional level

Embryonic bodies generated through hanging drop technique are a well-established model to study differentiation of ESCs into more specialised cells. In particular, the formation of beating foci presents an easily trackable system for the study of cardiogenesis[78]. To elucidate this feature on molecular level, I examined the expression of genes that are known to participate during heart morphogenesis, development, maturation and/or maintenance in our time series experiment.

Taking into consideration the previous results, I decided to evaluate the dynamics of *in vitro* gene expression for the cardiac genes, for which I described the *in vivo* expression in Chapter 4. To facilitate this analysis, transcript values were linearly transformed into a 0 to 1 scale (Min-Max normalisation). First, I inspected the expression of cardiac TFs: *Gata3*, *Gata4*, *Nkx2-5* peaked at day 4 of stem cell differentiation marking early cardiac progenitor cells in a precocious development sate (Figure 6.3.5A-B), while *Hand2* and *Myocd* (transcriptional co-activator with serum response factor (*Srf*)) displayed an increase of expression at later timepoints (Figure 6.3.5B). The latter indicates for the presence of smooth muscle cells and cardiac muscle in the differentiating stem cell population[214].

Subsequently, I examined the expression of genes associated with cardiac muscle contraction, which is the primary function of the heart (Figure 6.3.5C). *Myh7*, which encodes for the β-Myosin Heavy Chain, was most strongly expressed at day 8 and gradually decreasing at later days, while *Actc1* and *Tnni3* displayed expression peaks at later stages of differentiation. This finding is consistent with previous observations of postnatal downregulation of *Myh7* in mice and other rodent hearts[155]. Strikingly, *Actc1*, a known marker for early myogenesis had the highest signal intensity at later differentiation timepoints. Likewise, the expression of cardiac troponin I (*Tnni3*) was gradually upregulated during differentiation, with maximum expression occurring at day 16.

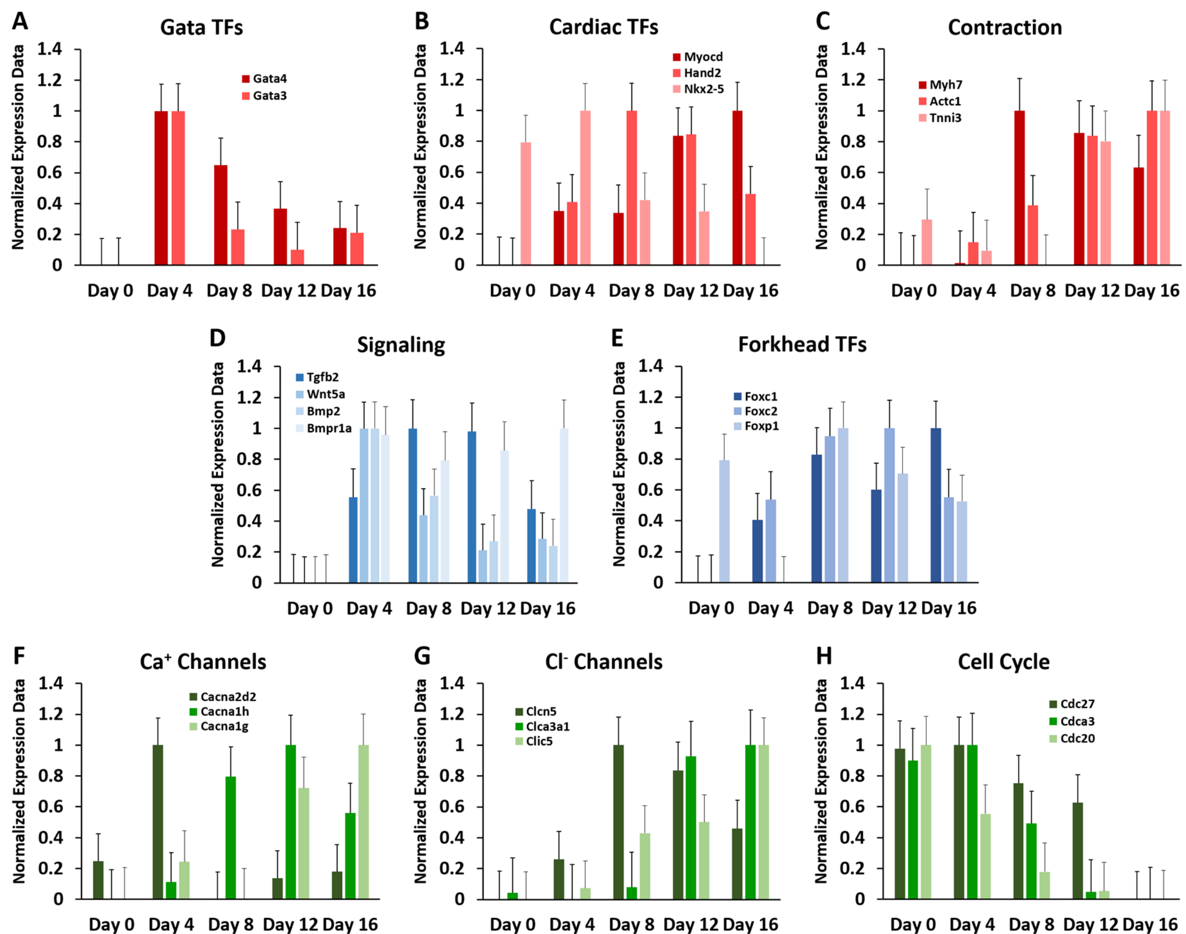


Figure 6.3.5 Temporal expression profiles of genes associated to heart development and ion channel function. Expression values were transformed to the [0,1] range. The scale included a larger range to capture the errors intervals. In the graphics are shown (A) Gata transcription factors (TFs); (B) Other known key TFs for cardiogenesis; (C) Genes encoding for structural cardiac proteins; (D) Signalling pathway molecules linked for cardiac development; (E) Forkhead gene family; (F) Calcium channel genes; (G) Chloride channel genes and; (H) Cell cycle genes.

Members of BMP, Wnt and Tgf-β signalling pathways were also observed to be expressed during *in vitro* stem cell differentiation. These pathways have well established roles in heart development. *Tgf-β* is an important regulator of a wide range of cellular responses during cell differentiation[215], while BMP and WNT signalling pathways contribute, for instance, to specification of the FHF and SHF progenitors in pluripotent stem cells[216]. All of the inspected signalling genes (*Tgf-β2*, *Wnt5a*, *Bmp2*, *Bmpr1a*) were not expressed in undifferentiated cells but displayed characteristic expression patterns at later time points. Both *Wnt5a* and *Bmp2* reached maximum expression at day 4 and were downregulated over time afterwards. *Tgf-β2* displayed an increased expression from day 4 until day 12 and decreased expression afterwards, while *Bmpr1a* showed stable expression levels from day 4 on (Figure 6.3.5D).

Expression of *Foxc1* and *Foxc2* of the forkhead family of TFs, which are known to be important factors in embryonic heart development[217], which were observed to peaked at day 16 and day 12, respectively. *Foxc1* and *Foxc2*, are also important for heart development and present redundant functions, since the elimination of both genes present a more severe phenotype[218]. Furthermore, they contribute for the proper formation of the heart outflow tract and the right ventricle[219]. In contrast to *Foxc1* and *Foxc2*, *Foxp1*was expressed at day 0 but not expressed at day 4. Its expression exhibited a peak at day 8 and was decreased progressively afterwards (Figure 6.3.5E). *Foxp1* is known to be downregulated during early differentiation[220] while being still essential at later stages for heart development, particularly in the myocardial area, valve formation and ventricular septation[221].

Components of ion channels, which were examined in the *in vivo* study, showed also dynamic expression patterns *in vitro. Cacna2d2*, which encodes for α2-δ subunits of voltage-gated calcium (Ca2+) channels, peaked at day 4 and gradually decreased during differentiation, while *Cacna1h* encoding for α1 subunits peaked at day 12 and then decreased in expression (Figure 6.3.5F). In contrast, the gene *Cacna1g*, which is a paralog of *Cacna1h* and encodes an alternative α1 subunit, displayed a small increase in expression at day 4 and reached a maximum at day 18 only. Similarly, the chloride channels showed distinct dynamics with Clcn5, *Clca3a1* and *Clic5* displaying the strongest increase in expression at day 4, day 8 and day 12, respectively (Figure 6.3.5G). *Clic5* is known to be expressed in cardiomyocyte mitochondria, therefore it expected to observe its peak expression at later differentiation stages[163].

Finally, transcript levels of cell-division cycle associated genes (Figure 6.3.5H) decreased during stem cell differentiation with minimal expression at day 16. This reflects the loss of cell cycle activity with the progression of differentiation and might contribute to cell-cycle arrest of cardiomyocytes.

Overall, the expression patterns of genes known to be associated with heart morphogenesis and normal heart function in this section indicated the existence of a robust cardiac expression signature despite the EBs' heterogeneity. This supports the use of EBs as a model for heart development studies[78,222]. However, I also observed divergent expression patterns. For instance, *Actc1* showed its highest expression during early cardiogenesis and was repressed progressively during development *in vivo*, while higher levels of *in vitro* expression of *Actc1*

were only detected at later stages of differentiation. This might indicate that EB cells were at least partially still in an early myogenesis phase even at the final time point of the *in vitro* experiment.

6.3.5 Detection of shared *in vitro* and *in vivo* miRNA-mRNA interactions

The detection of a prominent cardiac gene expression supports the suitability of EBs for the study of cardiogenesis. In principle, therefore, EBs should provide means for detection of heart-relevant miRNA and their targets genes through the analysis time series data. Nevertheless, the heterogeneity of cell population in embryonic bodies imposes a challenge for interpretation and prioritisation of indicated interactions for further study, as anticorrelation of expression of miRNAs and its targets could derive from regulatory activity in non-cardiac cells in the EB. To alleviate this problem, I used the compiled *in vivo* miRNA-mRNA interaction data described in the previous chapter for stratification. While heterogeneous cell population in the *in vitro* experiment can impact on gene/miRNA expression, I expect that the *in vivo* experiment gene/miRNA expression should be strictly associated to heart development and maturation. The aim of the following analysis is to reveal which miRNAs interactions are potentially active both *in vitro* or *in vivo* during differentiation or development, respectively. Conserved interactions can then be prioritised for further study.

The analysis was based on the previously integrated set of miRNA-mRNA interactions (described in Chapter 5 – Section 5.2, Table S5.2.1) comprising 104001 potential interactions between 319 miRNAs and 10596 target genes. Firstly, I extracted all overlapping miRNA-mRNA interactions for both studies and obtained a table of 76343 interactions (Table S6.3.6). Notably, not all of the interactions that showed negative correlation *in vitro* were also negatively correlated *in vivo* and vice-versa, meaning that if an interaction is negatively correlated in one study, it is not necessarily negatively correlated in the other study. To identify relevant interactions, only interactions were kept for which miRNAs targeted genes associated with heart development in GO[147].

The additional filtering reduced the number of miRNA-mRNA interactions to 3983 interactions. The distribution of correlation coefficients for the *in vitro* and *in vivo* experiments in shown in Figure 6.3.6 (see also Table S6.3.7).

Finally, to identify potential interactions with a conserved correlation between miRNA and mRNA, I set a maximal threshold for the differences in correlation coefficients of 0.2.

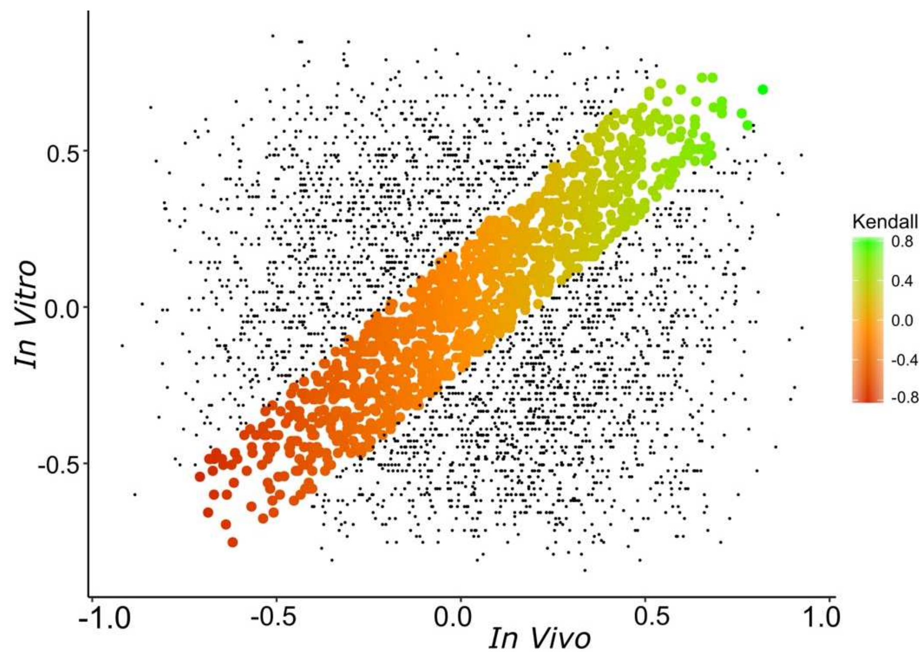


Figure 6.3.6 Correlation of *in vitro* and *in vivo.* expression of miRNA-mRNA interactions. Each dot represents a miRNA-mRNA interaction. The values of the x-axis are the correlation coefficients for *in vivo* data; the values of the y-axis are the correlation coefficients for the *in vitro* data. Highlighted by colour are miRNA-mRNA interactions with similar correlation coefficients in both *in vivo* and *in vitro*, i.e., with a difference in correlation coefficient < 0.2.

Most of the miRNA-mRNA interactions did not display conserved interaction (black dots in Figure 6.3.6), which might be expected due to the experimental differences, i.e., sampling from the heart vs sampling from heterogeneous cell populations differentiation. Nonetheless, 1129 relevant interactions with conserved correlation patterns were retained. Assuming that the miRNAs tend to convey a repressive effect on target mRNAs, I retrieved miRNA-mRNA interaction negatively correlated in both studies. As a threshold, a negative correlation of less than -0.2 in both studies was applied. This final step yielded a set of 214 miRNA-mRNAs interacting pairs with negative correlation in both studies (Figure 6.3.6 orange-red dots, Table S6.3.8) that warranted further investigations described in the next section.

6.3.6 Putative novel candidate miRNA-mRNA interaction relevant for cardiac cell development *in vivo* and *in vitro*

For further prioritisation, I first identified the miRNAs with the largest number of targets and verified whether they had been previously described in the literature to be associated with heart development. Table 6.3.1 lists the 20 miRNAs with the largest number of targets. For a complete view of all miRNAs, consult Table S6.3.9.

Table 6.3.1 Top 20 miRNAs targeting heart related genes. Some of the miRNA-mRNA interactions have been already experimentally tested, but have not been associated with heart related processes (for more information consult Table S6.3.9). **PMID column** indicates the publication where an association has been already described.

miRID	Count	GeneSymbol	PMID
mmu-miR-27a-3p	7	*Robo1, Gata3, Zfp36l1, Flrt2, Mef2c, Slc8a1, Bmpr1a*	28293796
mmu-miR-34a-5p	7	*Pofut1, Fat4, Tab2, Maml1, Angpt1, Pdgfra, Kdm2a*	30474870
mmu-miR-680	6	*Ctnnb1, Zfp36l1, Foxc1, Gnaq, Zmiz1, Tgfbr3*	-
mmu-let-7b-3p	6	*Sfrp2, Lmo4, Vegfa, Hectd1, Acvr1, Frs2*	27072074
mmu-miR-23b-3p	5	*Has2, Pcsk5, Adam19, Whsc1, Nox4*	31103821
mmu-miR-705	5	*Ncam1, Sh3pxd2b, Kdm2a, Casp7, Gpc3*	-
mmu-miR-23a-3p	5	*Isl1, Sox11, Robo1, Aldh1a2, Med12*	31130720
mmu-miR-762	5	*Ncam1, Pdpn, Tgfbr1, Zfp36l1, Nedd4*	-
mmu-miR-27b-3p	4	*Smad4, Pdgfra, Ccm2, Tgfbr3*	27072074
mmu-miR-20b-5p	4	*Heg1, Pkd1, Nrp2, Tgfbr2*	25898012
mmu-miR-150-5p	4	*Efnb2, Prkar1a, Trps1, Erbb2*	31122130
mmu-miR-152-3p	4	*Dicer1, Ube4b, Gys1, Sos1*	-
mmu-miR-3473a	4	*Vegfa, Gab1, Camk2d, Mdm4*	-
mmu-miR-466f-3p	4	*Pitx2, Wisp1, Zfpm2, Kif3a*	-
mmu-miR-106b-5p	4	*Pkd1, Heg1, Tgfbr2, Grhl2*	-
mmu-miR-148a-3p	3	*Dicer1, Ube4b, Adam19*	25630970
mmu-miR-30c-5p	3	*Nfatc3, Adam19, Bcor*	30279543
mmu-miR-19b-3p	3	*Mef2a, Tgfbr2, Sgcb*	29664809
mmu-miR-3472	3	*Sema3c, Tmem100, Foxc1*	-
mmu-miR-106a-5p	3	*Heg1, Pkd1, Tgfbr2*	30004470

From the top 20 miRNAs, 12 miRNAs (60%) have already been described as being associated to heart related events, either in development, maturation, differentiation, or disease supporting the chosen approach for prioritisation. Despite existing association with heart related events, their role in heart development has not been yet fully characterised, since the majority of found miRNAs-mRNA interactions await experimentally validation (for more details see Table S6.3.8).

For instance, a relevance miR-27a-3p for the heart was already described in the literature[223,224], but an interaction with *Mef2c* has not been experimentally validated yet. Another potentially highly relevant regulation is given by the interaction of miR-23a-3p and *Isl1*, since *Isl1* is one of the most important transcription factors during early cardiac differentiation[16].

The remaining 40% of the miRNAs in table 6.3.1 have not been yet associated with cardiac events in the literature. One such miRNA is miR-680, despite having predicted interactions with six genes involved in heart development (GO:0007507). These targets include *Foxc1*, which is a known and well-established regulator early cardiogenesis and controls the generation of functional cardiomyocytes together with other early cardiac transcription factors such as *Mef2c*, *Isl1* and *Nkx2-5*[225]. Furthermore, *Foxc1* regulates early cardiomyogenesis by actuating in a very specific time frame during differentiation, potentiating the efficiency of cardiomyocyte production from ESCs[225]. The targeting of *Tgfbr3* (transforming growth factor type III receptor, also known as betaglycan) by miR-680 would be a further interesting interaction for further validation, since *Tgfbr3* is an accessory co-receptor of *Tgf-β* and important for the maintenance and protection of cardiac fibroblasts[226]. While the regulatory role of *Tgfbr3* in the Tgf-β signalling pathway is still not fully clear, recent results indicated *Tgfbr3* as a potential regulator of Tgf-β signalling with anti-apoptotic properties in the heart[226]. In cardiac fibroblasts exposed to hypoxia, Tgf-β signalling accounts predominantly for apoptosis of cardiac fibroblasts. However, the overexpression of *Tgfbr3* effectively prevented hypoxia-induced apoptosis of cardiac fibroblasts[226].

The presented examples indicate that the miRNAs listed in table 6.3.1 provide attractive candidates for further experimental validation and characterisation. The list includes 8 miRNAs, which have not yet been associated with cardiogenesis and thus could be novel cardiac miRNAs to be involved in cardiogenesis. Further research is also warranted for the remaining 12 miRNAs associated already with cardiac events, as only a minor part of their interactions has been experimentally validated.

6.4 Conclusion

How cardiac miRNAs contribute to stem cell differentiation towards cardiac cell formation is a formidable question. An answer could provide crucial clues not only for our understanding of heart development but also of *in vitro* models such as EBs in cardiac research. Using embryoid bodies as organoid models to understand heart formation, development and maturation is not without challenges due to their cellular heterogeneity. To dissect the data for EB formation and gain relevant information for research focusing on cardiogenesis, I have applied several procedures in the data analysis.

Firstly, only transcripts that were up- or downregulated by at least 4-fold were kept. This basic filtering step reduced the number of false positives among DEGs. Inspection of marker genes demonstrated that ESCs differentiation was occurring within the expected time frames. Markers of pluripotency were only expressed at day 0 of the experiment, while the majority of the markers for the three different germ layers only appear to be expressed at day 4. From day 8 onwards, all those markers were not or lowly expressed, indicating that the majority of the cells are already committed to the three different cells and are undergoing further specification[227]. A similar analysis approach was performed for miRNAs, where I inspected miRNAs that have been previously described to pluripotency maintenance and cell differentiation[71,201,202]. According to literature, miR-290a-3p, miR-290a-5p, miR-295-5p, miR-363-3p and miR-363-5p are involved in stem cell maintenance and proliferation. This conjecture was confirmed by the results obtained here, since these miRNAs were expressed at day 0 and increasingly repressed at the remaining time points. In contrast, miR-24 displayed increasing expression levels which agrees with its role in stem cell differentiation described in the literature[203]. Similarly, I observed an upregulation of Let-7 family members that are involved in differentiation, embryogenesis and development[228].

I also analysed the functions of miRNA targets. For this purpose, miRNAs with expression peaks were identified by correlation with a binary pattern and their targets were functionally assessed though enrichment analyses. The results were consistent with expected progression of ESCs differentiation.

At day 0, processes associated to organogenesis and morphogenesis were inhibited, while affected processes could be linked to both maintenance and differentiation of the ESCs at day 4. Processes related to cell cycle and pluripotency were effectively repressed after day 8.

As the aim of my study was to elucidate the utility of *in vitro* ESC differentiation for the study of cardiogenesis, I assessed specific cardiac markers and their expression during ESC differentiation. Importantly, I observed the presence of transcription factors linked for heart development in the early time points, which indicates that a part of cell population was differentiating towards to a cardiac fate. The expressed transcription factors included the Gata family, *Hand2*, *Nkx2-5* and *Myocd*. I also identified the expression of genes associated with of heart contraction and ion channels indicating the initiation of cardiac specification of cells. Collectively, the observations supported the use of embryoid bodies for studying cardiogenesis despite the challenges and restrictions.

Comparative analysis of *in vitro* and *in vivo* expression data allowed to detect the similarities between studies and to identify those miRNAs, which warrant further experimental validation. To increase their relevance for cardiogenesis, I only include miRNAs as targets if they have been already associated to heart development according to GO. This type of analysis led to the detection of 1129 miRNA-mRNAs interactions with conserved correlation of expression between studies. Remarkably,214 miRNA-mRNAs pairs showed negative correlation of expression in both studies.

Finally, a discriminatory list of the obtained 214 interactions was compiled to prioritise novel miRNA-mRNA interactions for further experimental analysis. From the 102 miRNAs detected in total, I focused on the top 20 miRNAs with the largest number of targets and verified through literature review if they have been previously associated to heart related events, i.e., development, morphogenesis, maturation, or disease. Notably, 12 of the top miRNAs have already been associated to heart related event. The remaining 8 miRNAs appear as particularly attractive candidates for novel cardiac regulators. One of the miRNAs that stood out was miR-680, which targeted 6 mRNAs that have been previously described as participating in heart development. The results demonstrate how the parallel profiling of the miRNAs and mRNA and comparison of *in vitro* and *in vivo* data can provide novel cues about ESC differentiation and eventually of heart development and their inherent molecular networks.

Chapter VII

Summary and Outlook

Summary

Heart development is a highly complex, dynamic and three-dimensional process that comprise numerous events at molecular, cellular, and tissue level. Over the last decades, genetic studies and the characterisation of the spatial activity of genes during heart development have provided new insights into the orchestration of the multitude of mechanisms. However, we are still far from a complete and holistic understanding of heart development and its regulation.

To gain further insights into gene expression and gene regulation underlying heart development, I have applied a systems biology approach based on publicly available omics data sets. Systems biology analysis aims to focus on the biological systems in a comprehensive manner and thereby enables the interpretation of genomic data within a more holistic framework of biological events. Furthermore, the integrative approach of systems biology complements the more reductionist studies focusing on single genes.

Using such systems biology framework, I was able to address some salient limitations that existed for the study of heart development, such as the lack of dedicated analysis platforms for the study of cardiogenesis and the missing capture of the cardiogenic role of miRNAs on genome wide level. The efforts included (i) the collection and integration of existing transcriptomics data in a publicly accessible database; (ii) the development of a platform that allowed to study, analysis and visualisation of miRNAs-mRNA relationship on a genomic level, by using expression data of a parallel profiling study *in vivo* and experimentally and computationally derived miRNA-mRNA interactions from 5 different tools; and (iii) comparing and connecting *in vivo* with *in vitro* observations, in order to understand the differences and similarities between the two systems and understand the potential application for studying heart development related events. The different approaches together with the main results are displayed in Figure 7.1.

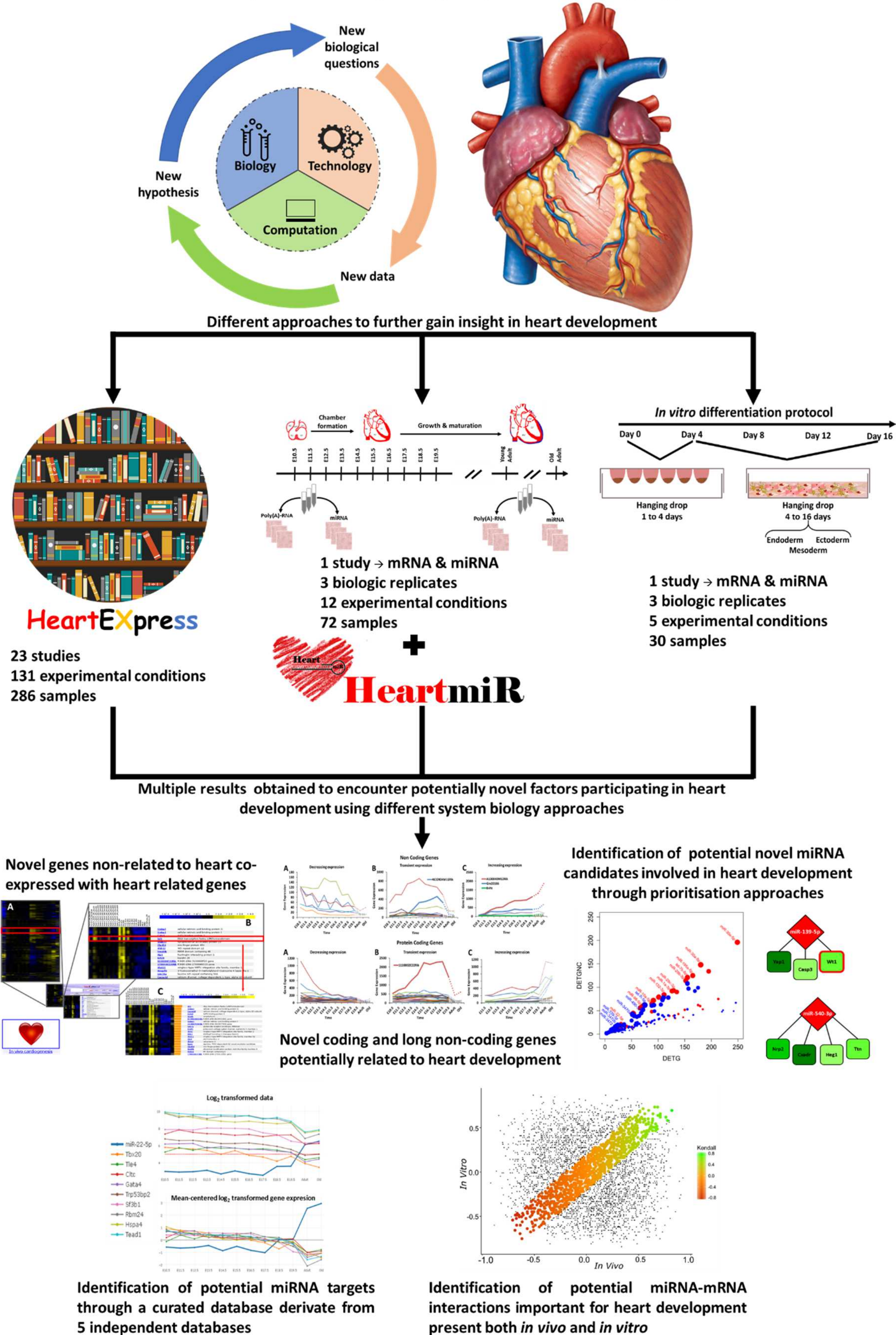


Figure 7.1 General overview of the work and investigations undertaken to better understand heart development and uncover novel regulatory mechanisms and factors during cardiogenesis in this thesis.

Main Contributions

(1) Curation and integration of cardiac transcriptomic data in HeartEXpress

The development of HeartEXpress allowed to tackle one of the main current challenges in scientific research, which is the lack of dedicated tools that compile some or all the data available regarding specific fields of study that in this case is on the subject of the heart study field. HeartEXpress comprises more than 20 experimental studies, containing the expression of more than16400 genes and 130 experimental conditions in both human and mouse. By bringing together the information of multiple studies, this platform will empower researchers querying gene expression profiles underling heart development and enable to uncover novel potential candidate genes for further experimental validations. In particular, this tool can be useful for assigning functions to genes that have not been yet associated to heart development or morphology, for prioritizing genes for experimental validation and for generating new biological hypothesis.

(2) Integration of miRNA-gene interactions and expression data in HeartmiR

A genome-wide dual profiling of mRNA and miRNA during heart development *in vivo* enabled the development of HeartmiR which address the lack of resources regarding miRNA-mRNA interactions that are specifically linked to study of heart development. The creation of HeartmiR opens new possibilities to analyse, study and correlate miRNA and mRNA by using the expression data of the parallel profiling study *in vivo* and gathered experimentally and computationally derived miRNA-mRNA interactions from 5 different databases/tools. This resource gathers the expression of 9211 mRNA and 386 miRNAs during murine heart development period (from E10.5 to E19.5) and additionally in adult and old murine heart tissue and identification of 102083 potential miRNA-mRNA interactions.

I strongly believe that HeartEXpress and HeartmiR webtools can significantly contribute to cardiac research, since they are unique webtools that provide specific features that will enable experimentalists and researchers to dissect expression patterns underlying cardiogenesis.

(3) Connecting miRNAs and mRNAs on transcriptomic level *in vivo* and *in vitro*

Genome-wise dual profiling of *in vivo* and *in vitro* data enabled the integration of miRNA and mRNA within regulatory networks. Based on the *in vivo* study, I identified 165 miRNAs that might be potentially involved in heart development. This was achieved through the analysis of individual miRNAs that target specific genes known to be involved in cardiogenesis. Accordingly, I carried out three different complementary approaches that intended to prioritise miRNA candidate's selection by evaluating how these candidates might influence heart development. The evaluation of the miRNAs through these prioritisation methods allowed to pinpoint certain miRNAs from a larger group of miRNAs and enable the creation of a more concise list of potential candidates.

Complementarily to the *in vivo* study, I also analysed the genome-wise profiling of coding and non-coding genes *in vitro.* One of the most interesting results in the *in vitro* study consisted in observing the action of miRNA transitory activity across timepoints during differentiation. I used the cross-correlation with binary vectors to retrieve miRNAs that had strong transient expression peaks for each experimental day. Potential mRNA targets indicated an astonishing sequential order of repressed events during differentiation. Notably, at early timepoint genes related to development and morphogenesis were being targeted, while at later timepoints the obtained miRNAs targeted genes linked to cell cycle, proliferation, transcription, and stem cell maintenance.

To go further, I have combined the previously gathered information from the *in vivo* miRNA-mRNA interaction and expression data with the *in vitro* data in order to detect novel potential regulators of stem cell differentiation and cardiogenesis through combined genome-wide profiling of protein coding transcripts and microRNA from both studies. This approach aimed to address the challenges in the interpretation of results for EBs and prioritizing miRNA and mRNAs interactions that might be relevant for heart development due the underlying heterogeneric cell population in EBs. Since the gene/miRNA expression observed *in vivo* is assumed to be strictly associated to heart development and maturation, I selected *in vitro* miRNAs interactions that are potentially active both *in vitro* or *in vivo* during differentiation or development (respectively). Astoundingly, 1129 interactions that showed conserved correlation patterns for both studies were retained. In line with the assumption that the

miRNAs tend to convey a repressive effect on target mRNAs, I solely gathered the miRNA-mRNA interactions that had a negative expression correlation. This final step yielded a set of 214 miRNA-mRNAs interacting pairs with negative correlation in both studies.

For further scrutiny of these 214 miRNA-mRNA interactions, I first identified the miRNAs with the largest number of targets and verified whether they had been previously described in the literature to be associated to heart development. Remarkably, from the top 20 miRNAs that had most targets, 12 of them (60%) had already been described to be associated to heart related processes supporting the validity of my approach and suggesting that the remaining miRNAs can serve as promising candidates for further studies.

The combination of both studies for *in vivo* and *in vitro* for embryonic heart development and embryonic stem cell differentiation may serve as a rational basis for study of cardiac miRNAs. The analysis encompassing the dual profiling of mRNA and miRNA transcriptome not only filled a key missing part in the comprehensive characterisation of murine heart development and stem cell differentiation, but also promises to be an excellent foundation to establish the regulatory molecular networks for heart development.

Outlook

All of the transcriptomic data analysed in the present work are based on RNA samples gained from tissues or cell populations. Over recent years, new fast evolving technologies will help understand the gene expression at the "single cell" level. Especially, single cell RNA-seq (scRNA-seq) technology has started to prove to be very valuable for understanding the different cardiac cell types that are present in the heart. Indeed, with the possibility to analyse at the single cell level, detailed insights across the repertoire of cardiac cells in different regions of the heart can be gained[229], which will help to uncover the molecular foundations of cardiac physiology and the cellular response to stress and disease. Nevertheless, studying cardiac functions will remain challenging as the heart is a complex organ with multiple types of cells that interact with each other and work as a whole to make this structure function. Thus, besides snapshots of the single cells' states provided by scRNA-seq, it will be crucial to capture the interactions between the different cardiac cell populations.

Another strategy that could prove to be very powerful for study of heart development could be the application of "multi-omics" approaches, which consist in the synchronised integration of genomics, epigenomics, transcriptomics, proteomics, and metabolomics. Such approaches can help us to understand better the mechanisms driving heart development and ideally to capture causal relationships within a statistical model framework. In recent years, high-throughput technology has further facilitated the integration of omics data for the identification of causal genes and molecular mechanisms involved in the development of cardiovascular events. Although "multi-omics" methods have been applied in heart investigation, the challenge to be address are formidable and the number of multi-omics studies are still limited.

Final conclusions

Overall, this work developed and applied computational tools within a systems biology framework to elucidate the complex process of cardiogenesis. Many interesting results for heart development and potential applications have been gained such as the detection of novel mRNAs, non-annotated protein-coding, lncRNA, miRNAs and miRNA-mRNA interactions. These results provide new means for experimental biologists to formulate novel hypothesis and to gain new insights. In addition, two webtools were developed so that researchers can perform their own independent research. I hope that the fruits of my labour will contribute to progress in the field of cardiogenesis and eventually to future success of cardiac regenerative medicine.

Chapter VIII

Bibliography

1 Tam, P. P. & Loebel, D. A. Gene function in mouse embryogenesis: get set for gastrulation. Nature reviews. Genetics 8, 368-381, doi:10.1038/nrg2084 (2007).

2 Zhang, S. et al. Physiological and molecular determinants of embryo implantation. Molecular aspects of medicine 34, 939-980, doi:10.1016/j.mam.2012.12.011 (2013).

3 Shahbazi, M. N. Mechanisms of human embryo development: from cell fate to tissue shape and back. Development 147, dev190629, doi:10.1242/dev.190629 (2020).

4 Pinto, J. P. et al. StemMapper: a curated gene expression database for stem cell lineage analysis. Nucleic acids research 46, D788-D793, doi:10.1093/nar/gkx921 (2018).

5 Gilbert, S. F. Developmental biology, the stem cell of biological disciplines. PLoS biology 15, e2003691, doi:10.1371/journal.pbio.2003691 (2017).

6 He, A., Kong, S. W., Ma, Q. & Pu, W. T. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. Proceedings of the National Academy of Sciences of the United States of America 108, 5632-5637, doi:10.1073/pnas.1016959108 (2011).

7 Schlesinger, J. et al. The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. PLoS genetics 7, e1001313, doi:10.1371/journal.pgen.1001313 (2011).

8 Kelly, R. G. Molecular inroads into the anterior heart field. Trends in cardiovascular medicine 15, 51-56, doi:10.1016/j.tcm.2005.02.001 (2005).

9 Meno, C. et al. lefty-1 Is Required for Left-Right Determination as a Regulator of lefty-2 and nodal. Cell 94, 287-297, doi:https://doi.org/10.1016/S0092-8674(00)81472-5 (1998).

10 Betts, J. G. et al. Anatomy and physiology. (2014).

11 Martin-Puig, S., Wang, Z. & Chien, K. R. Lives of a heart cell: tracing the origins of cardiac progenitors. Cell stem cell 2, 320-331, doi:10.1016/j.stem.2008.03.010 (2008).

12 Brade, T., Pane, L. S., Moretti, A., Chien, K. R. & Laugwitz, K. L. Embryonic heart progenitors and cardiogenesis. Cold Spring Harbor perspectives in medicine 3, a013847, doi:10.1101/cshperspect.a013847 (2013).

13 Kelly, R. G., Brown, N. A. & Buckingham, M. E. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. Developmental cell 1, 435-440 (2001).

14 Gittenberger-de Groot, A. C., Vrancken Peeters, M. P., Mentink, M. M., Gourdie, R. G. & Poelmann, R. E. Epicardium-derived cells contribute a novel population to the myocardial wall and the atrioventricular cushions. Circulation research 82, 1043-1052, doi:10.1161/01.res.82.10.1043 (1998).

15 Lescroart, F. et al. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. Nature cell biology 16, 829-840, doi:10.1038/ncb3024 (2014).

16 Cai, C. L. et al. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Developmental cell 5, 877-889 (2003).

17 Dorn, T. et al. Direct nkx2-5 transcriptional repression of isl1 controls cardiomyocyte subtype identity. Stem cells 33, 1113-1129, doi:10.1002/stem.1923 (2015).

18 Ma, Q., Zhou, B. & Pu, W. T. Reassessment of Isl1 and Nkx2-5 cardiac fate maps using a Gata4-based reporter of Cre activity. Developmental biology 323, 98-104, doi:10.1016/j.ydbio.2008.08.013 (2008).

19 Francou, A. et al. Second heart field cardiac progenitor cells in the early mouse embryo. Biochimica et biophysica acta 1833, 795-798, doi:10.1016/j.bbamcr.2012.10.003 (2013).

20 Günthel, M., Barnett, P. & Christoffels, V. M. Development, Proliferation, and Growth of the Mammalian Heart. Molecular Therapy 26, 1599-1609, doi:https://doi.org/10.1016/j.ymthe.2018.05.022 (2018).

21 Kamps, J. A. & Krenning, G. Micromanaging cardiac regeneration: Targeted delivery of microRNAs for cardiac repair and regeneration. World journal of cardiology 8, 163-179, doi:10.4330/wjc.v8.i2.163 (2016).

22 Vicente-García, C. & Carvajal, J. REFERENCE MODULE IN BIOMEDICAL SCIENCES: TRANSCRIPTION FACTORS IN MAMMALIAN MYOGENESIS (REVIEW) (2018).

23 Olariu, V., Lövkvist, C. & Sneppen, K. Nanog, Oct4 and Tet1 interplay in establishing pluripotency. Scientific reports 6, 25438, doi:10.1038/srep25438 https://www.nature.com/articles/srep25438#supplementary-information (2016).

24 Wray, J., Kalkan, T. & Smith, A. G. The ground state of pluripotency. Biochemical Society transactions 38, 1027-1032, doi:10.1042/BST0381027 (2010).

25 Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113, 631-642, doi:10.1016/s0092-8674(03)00393-3 (2003).

26 Silva, J. et al. Nanog is the gateway to the pluripotent ground state. Cell 138, 722-737, doi:10.1016/j.cell.2009.07.039 (2009).

27 Campolo, F. et al. Essential role of Sox2 for the establishment and maintenance of the germ cell line. Stem cells 31, 1408-1421, doi:10.1002/stem.1392 (2013).

28 Wang, Z., Oron, E., Nelson, B., Razis, S. & Ivanova, N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell stem cell 10, 440-454, doi:10.1016/j.stem.2012.02.016 (2012).

29 Zhang, S. & Cui, W. Sox2, a key factor in the regulation of pluripotency and neural differentiation. World J Stem Cells 6, 305-311, doi:10.4252/wjsc.v6.i3.305 (2014).

30 Thomson, M. et al. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell 145, 875-889, doi:10.1016/j.cell.2011.05.017 (2011).

31 Scholer, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. The EMBO journal 9, 2185-2195 (1990).

32 Willison, K. The mouse Brachyury gene and mesoderm formation. Trends in genetics : TIG 6, 104-105, doi:10.1016/0168-9525(90)90106-g (1990).

33 Kim, P. T. & Ong, C. J. Differentiation of definitive endoderm from mouse embryonic stem cells. Results and problems in cell differentiation 55, 303-319, doi:10.1007/978-3-642-30406-4_17 (2012).

34 Costello, I. et al. The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. Nature cell biology 13, 1084-1091, doi:10.1038/ncb2304 (2011).

35 Pfeiffer, M. J. et al. Cardiogenic programming of human pluripotent stem cells by dose-controlled activation of EOMES. Nat Commun 9, 440, doi:10.1038/s41467-017-02812-6 (2018).

36 Saga, Y. et al. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Development 126, 3437-3447 (1999).

37 van den Ameele, J. et al. Eomesodermin induces Mesp1 expression and cardiac differentiation from embryonic stem cells in the absence of Activin. EMBO reports 13, 355-362, doi:10.1038/embor.2012.23 (2012).

38 Yoshida, T., Vivatbutsiri, P., Morriss-Kay, G., Saga, Y. & Iseki, S. Cell lineage in mammalian craniofacial mesenchyme. Mechanisms of development 125, 797-808, doi:10.1016/j.mod.2008.06.007 (2008).

39 Bondue, A. et al. Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. Cell stem cell 3, 69-84, doi:10.1016/j.stem.2008.06.009 (2008).

40 Liu, Y. et al. Mesp1 Marked Cardiac Progenitor Cells Repair Infarcted Mouse Hearts. Scientific reports 6, 31457, doi:10.1038/srep31457 (2016).

41 Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research 47, D607-D613, doi:10.1093/nar/gky1131 (2019).

42 Khattar, P. et al. Distinction between two populations of islet-1-positive cells in hearts of different murine strains. Stem cells and development 20, 1043-1052, doi:10.1089/scd.2010.0374 (2011).

43 Pashmforoush, M. et al. Nkx2-5 pathways and congenital heart disease; loss of ventricular myocyte lineage specification leads to progressive cardiomyopathy and complete heart block. Cell 117, 373-386, doi:10.1016/s0092-8674(04)00405-2 (2004).

44 George, V., Colombo, S. & Targoff, K. L. An early requirement for nkx2.5 ensures the first and second heart field ventricular identity and cardiac function into adulthood. Developmental biology 400, 10-22, doi:10.1016/j.ydbio.2014.12.019 (2015).

45 Zhou, P., He, A. & Pu, W. T. Regulation of GATA4 transcriptional activity in cardiovascular development and disease. Current topics in developmental biology 100, 143-169, doi:10.1016/B978-0-12-387786-4.00005-1 (2012).

46 Kuo, C. T. et al. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes & development 11, 1048-1060, doi:10.1101/gad.11.8.1048 (1997).

47 Steimle, J. D. & Moskowitz, I. P. TBX5: A Key Regulator of Heart Development. Current topics in developmental biology 122, 195-221, doi:10.1016/bs.ctdb.2016.08.008 (2017).

48 Bruneau, B. G. et al. Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome. Developmental biology 211, 100-108, doi:10.1006/dbio.1999.9298 (1999).

49 Katz, M. G., Fargnoli, A. S., Kendle, A. P., Hajjar, R. J. & Bridges, C. R. The role of microRNAs in cardiac development and regenerative capacity. American journal of physiology. Heart and circulatory physiology 310, H528-541, doi:10.1152/ajpheart.00181.2015 (2016).

50 Kwon, C., Han, Z., Olson, E. N. & Srivastava, D. MicroRNA1 influences cardiac differentiation in Drosophila and regulates Notch signaling. Proceedings of the National Academy of Sciences of the United States of America 102, 18986-18991 (2005).

51 Stainier, D., Lee, R. K. & Fishman, M. C. Cardiovascular development in the zebrafish. I. Myocardial fate map and heart tube formation. Development 119, 31-40 (1993).

52 Zhang, H. M. et al. Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. Briefings in bioinformatics 16, 45-58, doi:10.1093/bib/bbt085 (2015).

53 Schneider, M. R. MicroRNAs as novel players in skin development, homeostasis and disease. The British journal of dermatology 166, 22-28, doi:10.1111/j.1365-2133.2011.10568.x (2012).

54 Dal-Pra, S. & Mirotsou, M. Reprogramming approaches in cardiovascular regeneration. Current treatment options in cardiovascular medicine 16, 327, doi:10.1007/s11936-014-0327-0 (2014).

55 Hodgkinson, C. P., Kang, M. H., Dal-Pra, S., Mirotsou, M. & Dzau, V. J. MicroRNAs and Cardiac Regeneration. Circulation research 116, 1700-1711, doi:10.1161/CIRCRESAHA.116.304377 (2015).

56 Jayawardena, T., Mirotsou, M. & Dzau, V. J. Direct reprogramming of cardiac fibroblasts to cardiomyocytes using microRNAs. Methods in molecular biology 1150, 263-272, doi:10.1007/978-1-4939-0512-6_18 (2014).

57 Jayawardena, T. M. et al. MicroRNA-mediated in vitro and in vivo direct reprogramming of cardiac fibroblasts to cardiomyocytes. Circulation research 110, 1465-1473, doi:10.1161/CIRCRESAHA.112.269035 (2012).

58 Chen, J. et al. mir-17-92 cluster is required for and sufficient to induce cardiomyocyte proliferation in postnatal and adult hearts. Circulation research 112, 1557-1566, doi:10.1161/CIRCRESAHA.112.300658 (2013).

59 Espinoza-Lewis, R. A. & Wang, D. Z. MicroRNAs in heart development. Current topics in developmental biology 100, 279-317, doi:10.1016/B978-0-12-387786-4.00009-9 (2012).

60 Li, N., Zhou, H. & Tang, Q. miR-133: A Suppressor of Cardiac Remodeling? Frontiers in pharmacology 9, 903, doi:10.3389/fphar.2018.00903 (2018).

61 Xin, M., Olson, E. N. & Bassel-Duby, R. Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. Nature reviews. Molecular cell biology 14, 529-541, doi:10.1038/nrm3619 (2013).

62 Ling, T. Y. et al. Regulation of the SK3 channel by microRNA-499--potential role in atrial fibrillation. Heart rhythm 10, 1001-1009, doi:10.1016/j.hrthm.2013.03.005 (2013).

63 Callis, T. E. et al. MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. The Journal of clinical investigation 119, 2772-2786, doi:10.1172/JCI36154 (2009).

64 Jayawardena, T. M. et al. MicroRNA induced cardiac reprogramming in vivo: evidence for mature cardiac myocytes and improved cardiac function. Circulation research 116, 418-424, doi:10.1161/CIRCRESAHA.116.304510 (2015).

65 He, L. et al. A microRNA polycistron as a potential human oncogene. Nature 435, 828-833, doi:10.1038/nature03552 (2005).

66 Ventura, A. et al. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. Cell 132, 875-886, doi:10.1016/j.cell.2008.02.019 (2008).

67 Wang, J. et al. Bmp signaling regulates myocardial differentiation from cardiac progenitors through a MicroRNA-mediated mechanism. Developmental cell 19, 903-912, doi:10.1016/j.devcel.2010.10.022 (2010).

68 Xin, M. et al. A threshold of GATA4 and GATA6 expression is required for cardiovascular development. Proceedings of the National Academy of Sciences 103, 11189-11194 (2006).

69 Liang, D. et al. miR-10a regulates proliferation of human cardiomyocyte progenitor cells by targeting GATA6. PloS one 9, e103097, doi:10.1371/journal.pone.0103097 (2014).

70 Kuppusamy, K. T. et al. Let-7 family of microRNA is required for maturation and adult-like metabolism in stem cell-derived cardiomyocytes. Proceedings of the National Academy of Sciences 112, E2785-E2794, doi:10.1073/pnas.1424042112 (2015).

71 Kuppusamy, K. T., Sperber, H. & Ruohola-Baker, H. MicroRNA regulation and role in stem cell maintenance, cardiac differentiation and hypertrophy. Current molecular medicine 13, 757-764 (2013).

72 Chen, C. Y. et al. MicroRNA let-7-TGFBR3 signalling regulates cardiomyocyte apoptosis after infarction. EBioMedicine 46, 236-247, doi:10.1016/j.ebiom.2019.08.001 (2019).

73 Lo, B. & Parham, L. Ethical issues in stem cell research. Endocr Rev 30, 204-213, doi:10.1210/er.2008-0031 (2009).

74 Poulos, J. The limited application of stem cells in medicine: a review. Stem cell research & therapy 9, 1-1, doi:10.1186/s13287-017-0735-7 (2018).

75 Volarevic, V. et al. Ethical and Safety Issues of Stem Cell-Based Therapy. Int J Med Sci 15, 36-45, doi:10.7150/ijms.21666 (2018).

76 Batalov, I. & Feinberg, A. W. Differentiation of Cardiomyocytes from Human Pluripotent Stem Cells Using Monolayer Culture. Biomark Insights 10, 71-76, doi:10.4137/BMI.S20050 (2015).

77 Kitajima, S., Takagi, A., Inoue, T. & Saga, Y. MesP1 and MesP2 are essential for the development of cardiac mesoderm. Development 127, 3215-3226 (2000).

78 Wang, X. & Yang, P. In vitro differentiation of mouse embryonic stem (mES) cells using the hanging drop method. Journal of visualized experiments : JoVE, doi:10.3791/825 (2008).

79 Evans, T. Embryonic Stem Cells as a Model for Cardiac Development and Disease. Drug Discov Today Dis Models 5, 147-155, doi:10.1016/j.ddmod.2009.03.004 (2008).

80 Senyo, S. E. et al. Mammalian heart renewal by pre-existing cardiomyocytes. Nature 493, 433-436, doi:10.1038/nature11682 (2013).

81 Birket, M. J. et al. Expansion and patterning of cardiovascular progenitors derived from human pluripotent stem cells. Nature Biotechnology 33, 970-979, doi:10.1038/nbt.3271 (2015).

82 Van Berlo, J. H. & Molkentin, J. D. An emerging consensus on cardiac regeneration. Nature medicine 20, 1386-1393 (2014).

83 Lee, R. T. & Walsh, K. The Future of Cardiovascular Regenerative Medicine. Circulation 133, 2618-2625, doi:10.1161/CIRCULATIONAHA.115.019214 (2016).

84 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. cell 126, 663-676 (2006).

85 Christoforou, N. et al. Transcription factors MYOCD, SRF, Mesp1 and SMARCD3 enhance the cardio-inducing effect of GATA4, TBX5, and MEF2C during direct cellular reprogramming. PloS one 8, e63577, doi:10.1371/journal.pone.0063577 (2013).

86 Fu, J. D. et al. Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state. Stem cell reports 1, 235-247, doi:10.1016/j.stemcr.2013.07.005 (2013).

87 Fu, Y. et al. Direct reprogramming of mouse fibroblasts into cardiomyocytes with chemical cocktails. Cell research 25, 1013-1024, doi:10.1038/cr.2015.99 (2015).

88 Ieda, M. et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell 142, 375-386, doi:10.1016/j.cell.2010.07.002 (2010).

89 Bollini, S., Smits, A. M., Balbi, C., Lazzarini, E. & Ameri, P. Triggering Endogenous Cardiac Repair and Regeneration via Extracellular Vesicle-Mediated Communication. Frontiers in physiology 9, 1497-1497, doi:10.3389/fphys.2018.01497 (2018).

90 Consortium, M. et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nature biotechnology 24, 1151-1161, doi:10.1038/nbt1239 (2006).

91 Foong, Y. M., Fu, J., Yao, S. Q. & Uttamchandani, M. Current advances in peptide and small molecule microarray technologies. Current Opinion in Chemical Biology 16, 234-242, doi:https://doi.org/10.1016/j.cbpa.2011.12.007 (2012).

92 Domian, I. J. et al. Generation of functional ventricular heart muscle from mouse ventricular progenitor cells. Science 326, 426-429, doi:10.1126/science.1177350 (2009).

93 Napoli, C. et al. Microarray analysis: a novel research tool for cardiovascular scientists and physicians. Heart 89, 597-604, doi:10.1136/heart.89.6.597 (2003).

94 Sabour, D. et al. Parallel Genome-wide Profiling of Coding and Non-coding RNAs to Identify Novel Regulatory Elements in Embryonic and Maturated Heart. Molecular therapy. Nucleic acids 12, 158-173, doi:10.1016/j.omtn.2018.04.018 (2018).

95 Sharma, H. S., Peters, T. H., Moorhouse, M. J., van der Spek, P. J. & Bogers, A. J. DNA microarray analysis for human congenital heart disease. Cell biochemistry and biophysics 44, 1-9, doi:10.1385/cbb:44:1:001 (2006).

96 Song, K. et al. Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature 485, 599-604, doi:10.1038/nature11139 (2012).

97 Steenbergen, C. et al. Alterations in apoptotic signaling in human idiopathic cardiomyopathic hearts in failure. American journal of physiology. Heart and circulatory physiology 284, H268-276, doi:10.1152/ajpheart.00707.2002 (2003).

98 Cook, S. A., Matsui, T., Li, L. & Rosenzweig, A. Transcriptional effects of chronic Akt activation in the heart. Journal of Biological Chemistry 277, 22528-22533 (2002).

99 Friddle, C. J., Koga, T., Rubin, E. M. & Bristow, J. Expression profiling reveals distinct sets of genes altered during induction and regression of cardiac hypertrophy. Proceedings of the National Academy of Sciences 97, 6745-6750 (2000).

100 Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology 5, R80, doi:10.1186/gb-2004-5-10-r80 (2004).

101 Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20, 307-315 (2004).

102 Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. Bioinformatics 26, 2363-2367 (2010).

103 Irizarry, R. A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249-264, doi:10.1093/biostatistics/4.2.249 (2003).

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

105 Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 43, e47, doi:10.1093/nar/gkv007 (2015).

106 Tai, Y. C. & Speed, T. P. A multivariate empirical Bayes statistic for replicated microarray time course data. The Annals of Statistics 34, 2387-2412 (2006).

107 Smyth, G. K. in Bioinformatics and computational biology solutions using R and Bioconductor 397-420 (Springer, 2005).

108 Efron, B. & Tibshirani, R. Empirical bayes methods and false discovery rates for microarrays. Genet Epidemiol 23, doi:10.1002/gepi.1124 (2002).

109 Efron, B., Tibshirani, R., Storey, J. D. & Tusher, V. Empirical Bayes analysis of a microarray experiment. Journal of the American statistical association 96, 1151-1160 (2001).

110 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57, 289-300 (1995).

111 Rouam, S. in Encyclopedia of Systems Biology (eds Werner Dubitzky, Olaf Wolkenhauer, Kwang-Hyun Cho, & Hiroki Yokota) 731-732 (Springer New York, 2013).

112 Kumar, L. & M, E. F. Mfuzz: a software package for soft clustering of microarray data. Bioinformation 2, 5-7 (2007).

113 Futschik, M. E. & Carlisle, B. Noise-robust soft clustering of gene expression time-course data. Journal of bioinformatics and computational biology 3, 965-988 (2005).

114 Carlson, M. org.Mm.eg.db: Genome wide annotation for Mouse. (2014).

115 Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. Bioinformatics 23, 257-258, doi:10.1093/bioinformatics/btl567 (2007).

116 Gaspar, J. A. et al. Gene expression signatures defining fundamental biological processes in pluripotent, early, and late differentiated embryonic stem cells. Stem cells and development 21, 2471-2484, doi:10.1089/scd.2011.0637 (2012).

117 Uosaki, H. et al. Efficient and scalable purification of cardiomyocytes from human embryonic and induced pluripotent stem cells by VCAM1 surface expression. PloS one 6, e23657, doi:10.1371/journal.pone.0023657 (2011).

118 Addis, R. C. et al. Optimization of direct fibroblast reprogramming to cardiomyocytes using calcium activity as a functional measure of success. Journal of molecular and cellular cardiology 60, 97-106, doi:10.1016/j.yjmcc.2013.04.004 (2013).

119 Li, X. et al. Transcriptional atlas of cardiogenesis maps congenital heart disease interactome. Physiological genomics 46, 482-495, doi:10.1152/physiolgenomics.00015.2014 (2014).

120 Qian, L. et al. In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. Nature 485, 593-598, doi:10.1038/nature11044 (2012).

121 Russ, J. & Futschik, M. E. Comparison and consolidation of microarray data sets of human tissue expression. BMC genomics 11, 305, doi:10.1186/1471-2164-11-305 (2010).

122 Kolesnikov, N. et al. ArrayExpress update--simplifying data submissions. Nucleic acids research 43, D1113-1116, doi:10.1093/nar/gku1057 (2015).

123 Barrett, T. et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic acids research 41, D991-995, doi:10.1093/nar/gks1193 (2013).

124 R: A language and environment for statistical computing (Foundation for Statistical Computing, Vienna, Austria, Vienna, Austria, 2015).

125 Team, R. RStudio: integrated development for R. RStudio, Inc., Boston, MA URL http://www. rstudio. com (2015).

126 de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. Bioinformatics 20, 1453-1454, doi:10.1093/bioinformatics/bth078 (2004).

127 Muraoka, N. et al. MiR-133 promotes cardiac reprogramming by directly repressing Snai1 and silencing fibroblast signatures. The EMBO journal 33, 1565-1581, doi:10.15252/embj.201387605 (2014).

128 Nam, Y. J. et al. Reprogramming of human fibroblasts toward a cardiac fate. Proceedings of the National Academy of Sciences of the United States of America 110, 5588-5593, doi:10.1073/pnas.1301019110 (2013).

129 Babiarz, J. E. et al. Determination of the human cardiomyocyte mRNA and miRNA differentiation network by fine-scale profiling. Stem cells and development 21, 1956-1965, doi:10.1089/scd.2011.0357 (2012).

130 Gu, Y. et al. Global DNA methylation and transcriptional analyses of human ESC-derived cardiomyocytes. Protein & cell, doi:10.1007/s13238-013-3911-2 (2013).

131 Poon, E. et al. Transcriptome-guided functional analyses reveal novel biological properties and regulatory hierarchy of human embryonic stem cell-derived ventricular cardiomyocytes crucial for maturation. PloS one 8, e77784, doi:10.1371/journal.pone.0077784 (2013).

132 Sun, N. et al. Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. Science translational medicine 4, 130ra147, doi:10.1126/scitranslmed.3003552 (2012).

133 Ifkovits, J. L., Addis, R. C., Epstein, J. A. & Gearhart, J. D. Inhibition of TGFbeta signaling increases direct conversion of fibroblasts to induced cardiomyocytes. PloS one 9, e89678, doi:10.1371/journal.pone.0089678 (2014).

134 Zhou, L., Liu, Y., Lu, L., Lu, X. & Dixon, R. A. Cardiac gene activation analysis in mammalian non-myoblasic cells by Nkx2-5, Tbx5, Gata4 and Myocd. PloS one 7, e48028, doi:10.1371/journal.pone.0048028 (2012).

135 Faustino, R. S., Behfar, A., Perez-Terzic, C. & Terzic, A. Genomic chart guiding embryonic stem cell cardiopoiesis. Genome biology 9, R6, doi:10.1186/gb-2008-9-1-r6 (2008).

136 Gan, L., Schwengberg, S. & Denecke, B. Transcriptome analysis in cardiomyocyte-specific differentiation of murine embryonic stem cells reveals transcriptional regulation network. Gene expression patterns : GEP 16, 8-22, doi:10.1016/j.gep.2014.07.002 (2014).

137 Cattaneo, P. et al. DOT1L-mediated H3K79me2 modification critically regulates gene expression during cardiomyocyte differentiation. Cell death and differentiation 23, 555-564, doi:10.1038/cdd.2014.199 (2016).

138 Rees, C. A., Demeter, J., Matese, J. C., Botstein, D. & Sherlock, G. GeneXplorer: an interactive web application for microarray data visualization and analysis. BMC Bioinformatics 5, 141, doi:10.1186/1471-2105-5-141 (2004).

139 Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nature cell biology 9, 625-635, doi:10.1038/ncb1589 (2007).

140 Niwa, H. How is pluripotency determined and maintained? Development 134, 635-646, doi:10.1242/dev.02787 (2007).

141 van Vliet, P., Goumans, M. J., Doevendans, P. A. & Sluijter, J. P. Human cardiomyocyte progenitor cells: a short history of nearly everything. Journal of cellular and molecular medicine 16, 1669-1673, doi:10.1111/j.1582-4934.2012.01535.x (2012).

142 Safran, M. et al. GeneCards Version 3: the human gene integrator. Database 2010, baq020 (2010).

143 Onizuka, T. et al. Wnt2 accelerates cardiac myocyte differentiation from ES-cell derived mesodermal cells via non-canonical pathway. Journal of molecular and cellular cardiology 52, 650-659, doi:10.1016/j.yjmcc.2011.11.010 (2012).

144 Raid, R. et al. Lack of Gata3 results in conotruncal heart anomalies in mouse. Mechanisms of development 126, 80-89, doi:10.1016/j.mod.2008.10.001 (2009).

145 Phillips, M. D., Mukhopadhyay, M., Poscablo, C. & Westphal, H. Dkk1 and Dkk2 regulate epicardial specification during mouse heart development. International journal of cardiology 150, 186-192, doi:10.1016/j.ijcard.2010.04.007 (2011).

146 Cohen, E. D., Miller, M. F., Wang, Z., Moon, R. T. & Morrisey, E. E. Wnt5a and Wnt11 are essential for second heart field progenitor development. Development 139, 1931-1940, doi:10.1242/dev.069377 (2012).

147 The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic acids research 45, D331-D338, doi:10.1093/nar/gkw1108 (2017).

148 Chen, J.-F. et al. microRNA-1 and microRNA-206 regulate skeletal muscle satellite cell proliferation and differentiation by repressing Pax7. Journal of Cell Biology 190, 867-879 (2010).

149 Liu, N. et al. microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart. Genes & development 22, 3242-3254, doi:10.1101/gad.1738708 (2008).

150 Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. Nature 436, 214-220, doi:10.1038/nature03817 (2005).

151 Meganathan, K., Sotiriadou, I., Natarajan, K., Hescheler, J. & Sachinidis, A. Signaling molecules, transcription growth factors and other regulators revealed from in-vivo and in-vitro models for the regulation of cardiac development. International journal of cardiology 183, 117-128, doi:10.1016/j.ijcard.2015.01.049 (2015).

152 Savolainen, S. M., Foley, J. F. & Elmore, S. A. Histology atlas of the developing mouse heart with emphasis on E11.5 to E18.5. Toxicologic pathology 37, 395-414, doi:10.1177/0192623309335060 (2009).

153 Zhu, H. Forkhead box transcription factors in embryonic heart development and congenital heart disease. Life sciences 144, 194-201, doi:10.1016/j.lfs.2015.12.001 (2016).

154 Sassoon, D. A., Garner, I. & Buckingham, M. Transcripts of alpha-cardiac and alpha-skeletal actins are early markers for myogenesis in the mouse embryo. Development 104, 155-164 (1988).

155 England, J. & Loughna, S. Heavy and light roles: myosin in the morphogenesis of the heart. Cellular and Molecular Life Sciences 70, 1221-1239 (2013).

156 Curran, J. et al. Eps15 Homology Domain-containing Protein 3 Regulates Cardiac T-type Ca2+ Channel Targeting and Function in the Atria. The Journal of biological chemistry 290, 12210-12221, doi:10.1074/jbc.M115.646893 (2015).

157 Ivanov, S. V. et al. Cerebellar ataxia, seizures, premature death, and cardiac abnormalities in mice with targeted disruption of the Cacna2d2 gene. The American journal of pathology 165, 1007-1018, doi:10.1016/s0002-9440(10)63362-7 (2004).

158 Olson, T. M. et al. Kv1.5 channelopathy due to KCNA5 loss-of-function mutation causes human atrial fibrillation. Human molecular genetics 15, 2185-2191, doi:10.1093/hmg/ddl143 (2006).

159 Schumacher-Bass, S. M. et al. Role for myosin-V motor proteins in the selective delivery of Kv channel isoforms to the membrane surface of cardiac myocytes. Circulation research 114, 982-992, doi:10.1161/circresaha.114.302711 (2014).

160 Sesti, F., Wu, X. & Liu, S. Oxidation of KCNB1 K(+) channels in central nervous system and beyond. World journal of biological chemistry 5, 85-92, doi:10.4331/wjbc.v5.i2.85 (2014).

161 Roder, K. & Koren, G. The K+ channel gene, Kcnb1: genomic structure and characterization of its 5'-regulatory region as part of an overlapping gene group. Biological chemistry 387, 1237-1246, doi:10.1515/bc.2006.153 (2006).

162 Duan, D. Y. et al. Functional role of anion channels in cardiac diseases. Acta pharmacologica Sinica 26, 265-278, doi:10.1111/j.1745-7254.2005.00061.x (2005).

163 Ponnalagu, D. et al. Molecular identity of cardiac mitochondrial chloride intracellular channel proteins. Mitochondrion 27, 6-14, doi:10.1016/j.mito.2016.01.001 (2016).

164 Fang, L. et al. Circulating microRNAs as biomarkers for diffuse myocardial fibrosis in patients with hypertrophic cardiomyopathy. Journal of translational medicine 13, 314, doi:10.1186/s12967-015-0672-0 (2015).

165 Li, M. et al. MicroRNA-29a-3p attenuates ET-1-induced hypertrophic responses in H9c2 cardiomyocytes. Gene 585, 44-50, doi:10.1016/j.gene.2016.03.015 (2016).

166 Ovchinnikova, E. S. et al. Signature of circulating microRNAs in patients with acute heart failure. European journal of heart failure 18, 414-423, doi:10.1002/ejhf.332 (2016).

167 Park, K. M. et al. Carvedilol-responsive microRNAs, miR-199a-3p and -214 protect cardiomyocytes from simulated ischemia-reperfusion injury. American journal of physiology. Heart and circulatory physiology 311, H371-383, doi:10.1152/ajpheart.00807.2015 (2016).

168 Vegter, E. L. et al. Use of biomarkers to establish potential role and function of circulating microRNAs in acute heart failure. International journal of cardiology 224, 231-239, doi:10.1016/j.ijcard.2016.09.010 (2016).

169 Koo, B. H. et al. ADAMTS9 is a cell-autonomously acting, anti-angiogenic metalloprotease expressed by microvascular endothelial cells. The American journal of pathology 176, 1494-1504, doi:10.2353/ajpath.2010.090655 (2010).

170 Ahn, J. S. et al. Aging-associated increase of gelsolin for apoptosis resistance. Biochemical and biophysical research communications 312, 1335-1341 (2003).

171 Adela, R. & Banerjee, S. K. GDF-15 as a Target and Biomarker for Diabetes and Cardiovascular Diseases: A Translational Prospective. Journal of diabetes research 2015, 490842, doi:10.1155/2015/490842 (2015).

172 Paillasse, M. R. & de Medina, P. The NR4A nuclear receptors as potential targets for anti-aging interventions. Medical hypotheses 84, 135-140, doi:10.1016/j.mehy.2014.12.003 (2015).

173 Swindell, W. R. Genes and gene expression modules associated with caloric restriction and aging in the laboratory mouse. BMC genomics 10, 585, doi:10.1186/1471-2164-10-585 (2009).

174 Klattenhoff, C. A. et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell 152, 570-583, doi:10.1016/j.cell.2013.01.003 (2013).

175 Schonrock, N., Harvey, R. P. & Mattick, J. S. Long Noncoding RNAs in Cardiac Development and Pathophysiology. Circulation research 111, 1349-1362, doi:doi:10.1161/CIRCRESAHA.112.268953 (2012).

176 Dimmeler, S. MicroRNAs in cardiovascular diseases and aging. Febs J 281, 23-23 (2014).

177 Thum, T., Catalucci, D. & Bauersachs, J. MicroRNAs: novel regulators in cardiac development and disease. Cardiovascular research 79, 562-570, doi:10.1093/cvr/cvn137 (2008).

178 Yan, S. & Jiao, K. Functions of miRNAs during Mammalian Heart Development. International journal of molecular sciences 17, doi:10.3390/ijms17050789 (2016).

179 Lê Cao, K.-A., Rohart, F., McHugh, L., Korn, O. & Wells, C. A. YuGene: A simple approach to scale gene expression data derived from different platforms for integrated analyses. Genomics 103, 239-251, doi:https://doi.org/10.1016/j.ygeno.2014.03.001 (2014).

180 Liu, W. & Wang, X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. Genome biology 20, 18, doi:10.1186/s13059-019-1629-z (2019).

181 Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F. & Procopio, A. D. Bioinformatic tools for microRNA dissection. Nucleic acids research 44, 24-44, doi:10.1093/nar/gkv1221 (2016).

182 Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. Nature genetics 39, 1278-1284, doi:10.1038/ng2135 (2007).

183 Khorshid, M., Hausser, J., Zavolan, M. & van Nimwegen, E. A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. Nature methods 10, 253-255, doi:10.1038/nmeth.2341 (2013).

184 Chou, C. H. et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic acids research 44, D239-247, doi:10.1093/nar/gkv1258 (2016).

185 Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. eLife 4, doi:10.7554/eLife.05005 (2015).

186 Betel, D., Wilson, M., Gabow, A., Marks, D. S. & Sander, C. The microRNA.org resource: targets and expression. Nucleic acids research 36, D149-153, doi:10.1093/nar/gkm995 (2008).

187 Wong, N. & Wang, X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic acids research 43, D146-152, doi:10.1093/nar/gku1104 (2015).

188 Ding, W. et al. miR-30e targets IGF2-regulated osteogenesis in bone marrow-derived mesenchymal stem cells, aortic smooth muscle cells, and ApoE-/- mice. Cardiovascular research 106, 131-142, doi:10.1093/cvr/cvv030 (2015).

189 Keller, G. M. In vitro differentiation of embryonic stem cells. Current opinion in cell biology 7, 862-869 (1995).

190 Lancaster, M. A. & Knoblich, J. A. Organogenesis in a dish: modeling development and disease using organoid technologies. Science 345, 1247125, doi:10.1126/science.1247125 (2014).

191 Chen, L. & Daley, G. Q. Molecular basis of pluripotency. Human molecular genetics 17, R23-27, doi:10.1093/hmg/ddn050 (2008).

192 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews. Genetics 10, 57-63, doi:10.1038/nrg2484 (2009).

193 Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281-297 (2004).

194 Hadjimichael, C., Nikolaou, C., Papamatheakis, J. & Kretsovali, A. MicroRNAs for Fine-Tuning of Mouse Embryonic Stem Cell Fate Decision through Regulation of TGF-beta Signaling. Stem cell reports 6, 292-301, doi:10.1016/j.stemcr.2016.01.004 (2016).

195 Bai, F. et al. Directed Differentiation of Embryonic Stem Cells Into Cardiomyocytes by Bacterial Injection of Defined Transcription Factors. Scientific reports 5, 15014, doi:10.1038/srep15014. https://www.nature.com/articles/srep15014#supplementary-information (2015).

196 Hamazaki, T., Oka, M., Yamanaka, S. & Terada, N. Aggregation of embryonic stem cells induces Nanog repression and primitive endoderm differentiation. Journal of Cell Science 117, 5681, doi:10.1242/jcs.01489 (2004).

197 Kashyap, V. et al. Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs. Stem cells and development 18, 1093-1108, doi:10.1089/scd.2009.0113 (2009).

198 Macatee, T. L. et al. Ablation of specific expression domains reveals discrete functions of ectoderm- and endoderm-derived FGF8 during cardiovascular and pharyngeal development. Development 130, 6361-6374, doi:10.1242/dev.00850 (2003).

199 Pankratz, M. T. et al. Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage. Stem cells 25, 1511-1520, doi:10.1634/stemcells.2006-0707 (2007).

200 Faial, T. et al. Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. Development 142, 2121-2135, doi:10.1242/dev.117838 (2015).

201 Li, N., Long, B., Han, W., Yuan, S. & Wang, K. microRNAs: important regulators of stem cells. Stem cell research & therapy 8, 110, doi:10.1186/s13287-017-0551-0 (2017).

202 Mathieu, J. & Ruohola-Baker, H. Regulation of stem cell populations by microRNAs. Advances in experimental medicine and biology 786, 329-351, doi:10.1007/978-94-007-6621-1_18 (2013).

203 Ma, Y. et al. Functional screen reveals essential roles of miR-27a/24 in differentiation of embryonic stem cells. The EMBO journal 34, 361-378, doi:10.15252/embj.201489957 (2015).

204 Schulman, B. R., Esquela-Kerscher, A. & Slack, F. J. Reciprocal expression of lin-41 and the microRNAs let-7 and mir-125 during mouse embryogenesis. Developmental dynamics : an official publication of the American Association of Anatomists 234, 1046-1054, doi:10.1002/dvdy.20599 (2005).

205 Wulczyn, F. G. et al. Post-transcriptional regulation of the let-7 microRNA during neural cell specification. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 21, 415-426, doi:10.1096/fj.06-6130com (2007).

206 Lee, H., Han, S., Kwon, C. S. & Lee, D. Biogenesis and regulation of the let-7 miRNAs and their functional implications. Protein & cell 7, 100-113, doi:10.1007/s13238-015-0212-y (2016).

207 Liu, F. & Di Wang, X. miR-150-5p represses TP53 tumor suppressor gene to promote proliferation of colon adenocarcinoma. Scientific reports 9, 6740, doi:10.1038/s41598-019-43231-5 (2019).

208 Lu, Q., Guo, Z. & Qian, H. Role of microRNA-150-5p/SRCIN1 axis in the progression of breast cancer. Exp Ther Med 17, 2221-2229, doi:10.3892/etm.2019.7206 (2019).

209 Scrutinio, D. et al. Circulating microRNA-150-5p as a novel biomarker for advanced heart failure: A genome-wide prospective study. The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation 36, 616-624, doi:10.1016/j.healun.2017.02.008 (2017).

210 Dorfeshan, P., Ghaffari Novin, M., Salehi, M. & Farifteh, F. Expression of miR-302 in human embryo derived from in-vitro matured oocyte. International journal of reproductive biomedicine 17, 405-412, doi:10.18502/ijrm.v17i6.4812 (2019).

211 He, X. & Xu, J. miR-126: A Potential Biomarker for Coronary Artery Disease? Cardiology 142, 2-3, doi:10.1159/000496613 (2019).

212 Sun, Y. et al. Insulin-like growth factor-1-mediated regulation of miR-193a expression promotes the migration and proliferation of c-kit-positive mouse cardiac stem cells. Stem cell research & therapy 9, 41, doi:10.1186/s13287-017-0762-4 (2018).

213 Bao, M. H. et al. Let-7 in cardiovascular diseases, heart development and cardiovascular differentiation from stem cells. International journal of molecular sciences 14, 23086-23102, doi:10.3390/ijms141123086 (2013).

214 Miano, J. M. Myocardin in biology and disease. Journal of biomedical research 29, 3-19, doi:10.7555/JBR.29.20140151 (2015).

215 Sakaki-Yumoto, M., Katsuno, Y. & Derynck, R. TGF-beta family signaling in stem cells. Biochimica et biophysica acta 1830, 2280-2296, doi:10.1016/j.bbagen.2012.08.008 (2013).

216 Andersen, P. et al. Precardiac organoids form two heart fields via Bmp/Wnt signaling. Nature Communications 9, 3140, doi:10.1038/s41467-018-05604-8 (2018).

217 Papanicolaou, K. N., Izumiya, Y. & Walsh, K. Forkhead transcription factors and cardiovascular biology. Circulation research 102, 16-31, doi:10.1161/CIRCRESAHA.107.164186 (2008).

218 Kume, T., Jiang, H., Topczewska, J. M. & Hogan, B. L. The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. Genes & development 15, 2470-2482, doi:10.1101/gad.907301 (2001).

219 Seo, S. & Kume, T. Forkhead transcription factors, Foxc1 and Foxc2, are required for the morphogenesis of the cardiac outflow tract. Developmental biology 296, 421-436, doi:10.1016/j.ydbio.2006.06.012 (2006).

220 Leishman, E. et al. Foxp1 maintains hair follicle stem cell quiescence through regulation of Fgf18. Development 140, 3809, doi:10.1242/dev.097477 (2013).

221 Wang, B. et al. Foxp1 regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation. Development 131, 4477-4487, doi:10.1242/dev.01287 (2004).

222 Lei, I. L., Bu, L. & Wang, Z. Derivation of cardiac progenitor cells from embryonic stem cells. Journal of visualized experiments : JoVE, 52047, doi:10.3791/52047 (2015).

223 Ieda, M. Key Regulators of Cardiovascular Differentiation and Regeneration: Harnessing the Potential of Direct Reprogramming to Treat Heart Failure. Journal of cardiac failure, doi:10.1016/j.cardfail.2019.09.005 (2019).

224 Wan, G. X. et al. MiR-15b-5p is Involved in Doxorubicin-Induced Cardiotoxicity via Inhibiting Bmpr1a Signal in H9c2 Cardiomyocyte. Cardiovascular toxicology 19, 264-275, doi:10.1007/s12012-018-9495-6 (2019).

225 Lambers, E. et al. Foxc1 Regulates Early Cardiomyogenesis and Functional Properties of Embryonic Stem Cell Derived Cardiomyocytes. Stem cells 34, 1487-1500, doi:10.1002/stem.2301 (2016).

226 Chu, W. et al. TGFBR3, a potential negative regulator of TGF-beta signaling, protects cardiac fibroblasts from hypoxia-induced apoptosis. Journal of cellular physiology 226, 2586-2594, doi:10.1002/jcp.22604 (2011).

227 Shparberg, R. A., Glover, H. J. & Morris, M. B. Modeling Mammalian Commitment to the Neural Lineage Using Embryos and Embryonic Stem Cells. Frontiers in physiology 10, 705, doi:10.3389/fphys.2019.00705 (2019).

228 Roush, S. & Slack, F. J. The let-7 family of microRNAs. Trends in cell biology 18, 505-516, doi:10.1016/j.tcb.2008.07.007 (2008).

229 Litviňuková, M. et al. Cells of the adult human heart. Nature 588, 466-472, doi:10.1038/s41586-020-2797-4 (2020).