

3D reconstruction of genome-wide gene expression and regulation in mouse hearts

Submitted by

Nathalia May-Yen Tan (BSc/BCompSc)

Supervised by

Dr. Mirana Ramialison and A/Prof. Jose M. Polo

Submitted in partial requirement for

0051: Bachelor of Science (Honours)

Australian Regenerative Medicine Institute

15 Innovation Walk, Monash University Clayton Campus

Clayton, Victoria 3800

Australia

Thursday, 15th October 2015

Student ID: 23403381

Word count: 13,725



MONASH
University

 **ARMI**
AUSTRALIAN REGENERATIVE
MEDICINE INSTITUTE

Authorship declaration

This thesis contains no material that has been accepted for the award of any other degree in any university. To the best of my knowledge and belief, this thesis contains no material previously published or written by any other person, except where due reference is given in the text.

Signed:

Date:

Acknowledgements

I would like to extend my expression of sincere gratitude and appreciation towards the following persons whom without, this project would not have been possible.

Thank you to **Dr. Mirana Ramialison** and **A/Prof. Jose M. Polo** for their guidance, support, patience, understanding, invaluable advice, and fantastic senses of humour throughout the year, making honours a memorable and irreplaceable experience.

From the **Romialison group**, thank you to Dr. Michael Eichenlaub, Jeannette Hallab, Markus Tondl, Louis Dang, and Henry Chiu.

From the **Polo group**, thank you to Dr. Christian Nefzger, Dr. Fernando Rossello, Dr. Anja Knaupp, Dr. Sue-Mei Lim, Jaber Firas, Sara Alaei-Shehni, Joseph Chen, Melissa Holmes, and Ethan Liu.

From the **Nilsson group**, thank you to A/Prof. Susie Nilsson, Jessica Hatwell-Humble, Andrea Reitsma, and Brenda Williams for their help in the tissue dissociation process.

From the **Rosenthal group**, thank you to Dr. Milena Furtado, Dr. Mauro Costa, Dr. Alexander Pinto, and Anjana Chanandran.

From the **Heart Failure Research Centre** in the Netherlands, thank you to A/Prof. Jan M. Ruijter and Dr. Antoon F. Moorman for their generosity in sharing their model of the heart.

From **ARL**, thank you to Victoria Brown for her help in obtaining mice.

Thank you to the mice.

Table of contents

1. Abstract	9
1. Introduction	11
1.1. The importance of spatiotemporal gene expression in development and disease	12
1.1. Gene transcription regulation.....	14
1.1.1. Histone modifications	15
1.1.2. Regulatory elements.....	16
1.1.3. DNA methylation.....	17
1.1.4. Non-coding RNA.....	17
1.2. Gene regulatory networks.....	18
1.3. Investigating gene expression and regulation	18
1.3.1. <i>In situ</i> hybridisation.....	19
1.3.2. RNA-seq	19
1.3.3. ChIP-seq.....	20
1.3.4. ATAC-seq	20
1.4. 3D modeling and gene expression	21
1.4.1. Tomo-seq	22
1.4.2. Drop-seq.....	23
1.4.3. <i>In situ</i> hybridisation on 3D models.....	23
1.5. Murine hearts as a model for investigating gene regulatory networks	24
1.6. Conclusion.....	25
2. Project outline	27
2.1. Background.....	27
2.2. Aims.....	28
2.3. Experimental design and methodology.....	28
2.4. Anticipated outcomes	29
3. Materials and Methods	30
3.1. Heart extraction and microdissection.....	30
3.1.1. Extraction.....	30
3.1.2. Dissection into slices	30
3.1.3. Dissection into pieces.....	32
3.2. Tissue preparation	33
3.2.1. Enzymatic digestion.....	33
3.2.2. Tissue homogeniser	34
3.3. FACS analysis	34
3.4. RNA purification and RNA-seq	35
3.5. ATAC-seq	35

3.6. Computational model.....	35
3.6.1. Software used.....	36
3.6.2. <i>In silico</i> slicing and sectioning	37
3.6.3. Programming	37
3.7. RNA seq pipeline	38
3.7.1. Batch analysis setup.....	38
3.7.2. Quality check I	39
3.7.3. Trimming and filtering.....	39
3.7.4. Quality check II	39
3.7.5. Alignment/mapping.....	40
3.7.6. Conversion into BAM.....	40
3.7.7. Read counts	41
3.8. Analysis	42
3.8.1. Cluster analysis	42
3.8.2. MDS analysis.....	42
4. Results.....	43
4.1. Adaptation of 3D model of adult murine heart	43
4.1.1. Sectioning.....	43
4.1.2. Preparation of 3D model to accept quantitative datasets	47
4.2. Sectioning of adult murine heart <i>ex vivo</i>	48
4.3. Tissue dissociation	49
4.4. Transcriptome of sub-anatomical compartments of the heart	51
4.4.1. Quantification of RNA.....	52
4.4.2. RNA quality assurance.....	53
4.5. Attempt at acquiring chromatin accessibility of sub-anatomical compartments	55
4.5.1. Quality assurance of ATAC-seq samples	55
4.5.2. Quantification of chromatin.....	56
4.5.3. Quality assurance of chromatin	57
4.6. Unsupervised cluster analysis	58
4.6.1. Hierarchical clustering	58
4.6.2. MDS plots	59
4.7. Integration of RNA-seq datasets with 3D model.....	61
5. Discussion	64
5.1. Technical issues	64
5.1.1. Computational model of the heart	64
5.1.2. Mice used for experiments	64
5.1.3. Heart extraction	65
5.1.4. Tissue dissociation	66

5.1.5. RNA-seq	66
5.1.6. Technical replicates	67
5.1.7. ATAC-seq	67
5.2. Discussion of results	68
5.2.1. Computational model	68
5.2.2. Transcriptomic data of the heart	68
5.2.3. Comparisons between 3D model and ISH	69
6. Conclusions and future directions	70
7. Appendices	72
Appendix A: Phosphate Buffered Solution recipe	72
Appendix B: RNA-seq pipeline	73
Appendix C: C# code for mapping transcriptomic data onto 3D model	77
Appendix D: Tondl protocol for tissue dissociation	81
8. References	83

1. Abstract

An understanding of spatial and temporal gene expression and regulation is key to uncovering developmental and physiological processes, and accordingly, disease processes. Numerous techniques exist to gain gene expression and regulation information, but very few utilise intuitive true-to-life imaging methods to analyse and present their results. Mapping gene expression and regulation information onto a three-dimensional model will greatly aid in the visualisation of gene expression patterns and allow researchers to see which areas in an organ or organism express which genes in the genome, and vice versa. Additionally, analysing this data using cluster analysis may uncover novel spatially restricted genes, relating to development and disease relating to specific areas in the heart. Hearts make a good model for this study as they are structurally and developmentally complex with distinct sub-compartments. Furthermore, mice as an animal model are ideal as they are genetically similar to humans and may have clinical translation. We hypothesise that deciphering the three-dimensional position and regulation of every gene at a given time during development and in disease conditions is a powerful approach to reveal the specific subset of genes that play an essential role in specific organ sub-compartments.

1. Introduction

A sound understanding of spatial and temporal gene expression and regulation is key to uncovering developmental and physiological processes, and accordingly, disease processes. A necessity in discovering these processes is in gaining information on gene expression and regulation. Spatio-temporal control of gene expression plays a major factor in the specification of tissues and organs within an organism from the first stages of development through to growth and maturation. Subsequently, disruption of normal gene expression can have detrimental or even fatal results for embryonic development or homeostasis.

To control the expression of genes temporally and spatially, gene regulation is achieved through epigenetic modulation: that is, changes relating to chromatin accessibility that are not directly related to the genome sequence itself. Chromatin comprising of DNA and histones must be appropriately structured to allow or prevent access of transcriptional regulators to specific regions. Epigenetic changes such as histone modifications and DNA methylation ultimately result in changes in gene expression, by affecting the accessibility of regulatory regions. Dissecting the dynamics of gene expression and regulation may lead to a better understanding of development and disease.

A wide variety of techniques exist to investigate gene expression and regulation patterns. Amongst these are *in situ* hybridisation, RNA-seq, Chromatin Immunoprecipitation (ChIP)-seq, and Assay for Transposase-Accessible Chromatin (ATAC)-seq, each with a unique set of characteristics, advantages, and disadvantages, that are a trade off between qualitative versus quantitative information or high-throughput and low resolution versus low-throughput with high spatial information.

To visualise information derived from these methods, 3D modelling proves to be a visually intuitive aid. For example, it has been shown that a simple way to overcome these limitations is to scrutinise sections or parts of an organism to gain information on a smaller scale, and then digitally reconstruct the sample in 3D again as exemplified using *Drosophila melanogaster* embryos (Combs & Eisen 2013). In this way, it was demonstrated that complex structures may be analysed quantitatively whilst retaining spatial information on gene function.

Hearts are a structurally and developmentally complex organ, making it a good model for the investigation of spatial and temporal gene expression and regulation in different organ sub-compartments. Murine hearts specifically make a good model for development and physiology, as they are genetically and physically highly similar to human hearts. This likeness provides a potential opportunity for the mouse model to be used for translational medicine in diagnostics.

1.1. The importance of spatiotemporal gene expression in development and disease

Animal development begins from the single fertilised cell, or zygote. This lone cell contains the DNA responsible for the formation of the entire organism. As the zygote replicates itself to grow into a multicellular embryo, each cell of the organism still retains the exact same DNA as its cell of origin. Given this, one of the most astonishing facets of developmental biology is the fact that a wide range of cell, tissue, and organ specificity can arise from the one set of ‘genetic blueprint’. To achieve this, developmental processes are orchestrated through precise spatial and temporal gene expression.

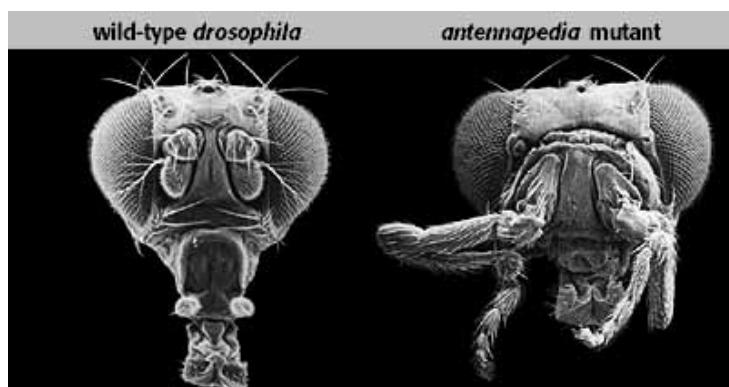


Figure 1.1. The importance of spatial gene expression, illustrated by the ectopic expression of the *labial* gene in *Drosophila melanogaster* resulting into the abnormal formation of limbs in the head region (Merrill et al. 1989).

When spatial gene expression is affected, the effects are often evident. A distinct example that displays the importance of the necessity for spatial gene control is the homeobox mutant using

the *Drosophila melanogaster* animal model. The homeobox genes are responsible for formation of the segments of the thorax in *Drosophila*. When the most proximal gene in the Antennapedia complex, *labial*, is ectopically activated in the location of antenna growth, limbs grow in place of antennae as seen in figure 1.1 (Merrill et al. 1989). In some cases, complete knockouts of genes result in embryonic lethality, but have more specific effects when the deletion is restricted to certain areas of the embryo. Conditional knockouts allow researchers to closely follow the effects of gene knockouts that would otherwise result in lethality in specific tissues and organs, a good indication that genes expressed in different areas have different consequences (Friedel et al. 2011).

Temporal control is equally as important as spatial control of gene expression. For example, if the *Notch1* gene is not expressed during embryonic days (E) E9 – E10.5 in mouse development, this results in embryonic fatality (Copp 1995). Because of this, similar to conditional knockouts, the temporal expression of a gene can be controlled if a complete knockout of a gene has fatal implications. Techniques such as Cre-Lox inducible gene knockouts mean that a gene knockout that would otherwise cause embryonic lethality at a particular timepoint in development can be passed without fatal results. Once the known lethality period has lapsed, the addition of a ligand to the Cre receptor results in the splicing of the gene and so allowing the investigation of a knockout (Friedel et al. 2011). Additionally, there are a number of genes in certain tissues and organs that need to be expressed constantly for purposes such as housekeeping, survival, growth, and proliferation (Eisenberg & Levanon 2013). If mutated, the function of the genes can change and result in embryonic fatality, malformation, cancer, or other diseases (Vecoli et al. 2014; Costa et al. 2013; Copp 1995).

A multitude of developmental and physiological processes can go awry when genes fail to be expressed at the right time and place. Therefore, uncovering the mechanisms that regulate gene expression can further our understanding of spatiotemporal gene expression and could explain the misexpression of genes.

1.1. Gene transcription regulation

Although gene expression is key in controlling growth and development, gene regulation holds the key to controlling the physical structure of chromatin in numerous ways. Gene regulation is the control of the activation or repression of a gene, and within gene regulation, epigenetics is the study of alterations that do not directly change the DNA sequence.

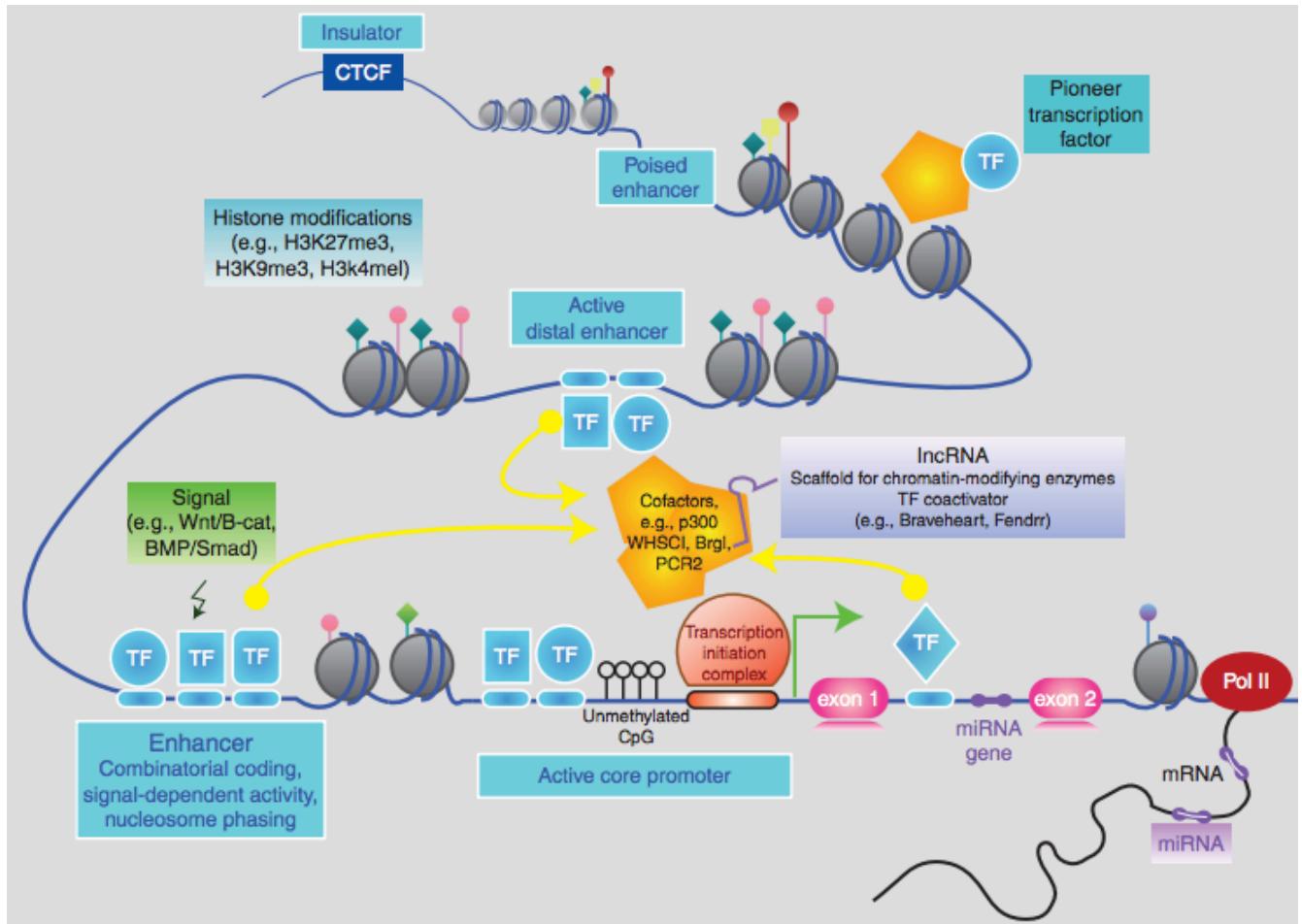


Figure 1.2. An overview of gene regulation, including histone modifications, regulatory elements (enhancers, promoters) bound by transcription factors (TFs), DNA methylation, and non-coding RNA (Waardenberg et al. 2014).

Within the nucleus of a cell, chromatin is tightly wound around proteins called histones to aid in compaction of chromatin. Chromatin compaction is a major factor in gene regulation, as it allows or prevents gene expression, depending on what state it is in (Martin & Cardoso 2010): if the histones are chemically modified to allow the chromatin to be opened, they allow regions of

chromatin to be exposed. These open regions include regulatory elements, which transcription factors bind to in order to induce gene expression or to repress gene expression. This sequence of events eventuating into transcriptional gene expression regulation is presented in figure 1.2. Gene expression regulation is controlled on a number of different levels. The first level of regulation is on the transcriptional level of the gene. Once the gene is transcribed, post-transcriptional modifications can be made to the mRNA. Upon translation into a protein, other translational modification can be made, and even further downstream, post-translational modifications are possible. Protein stability and degradation affect gene regulation. In this section, there will be a focus on gene regulatory mechanisms that act only on the transcriptional level.

1.1.1. Histone modifications

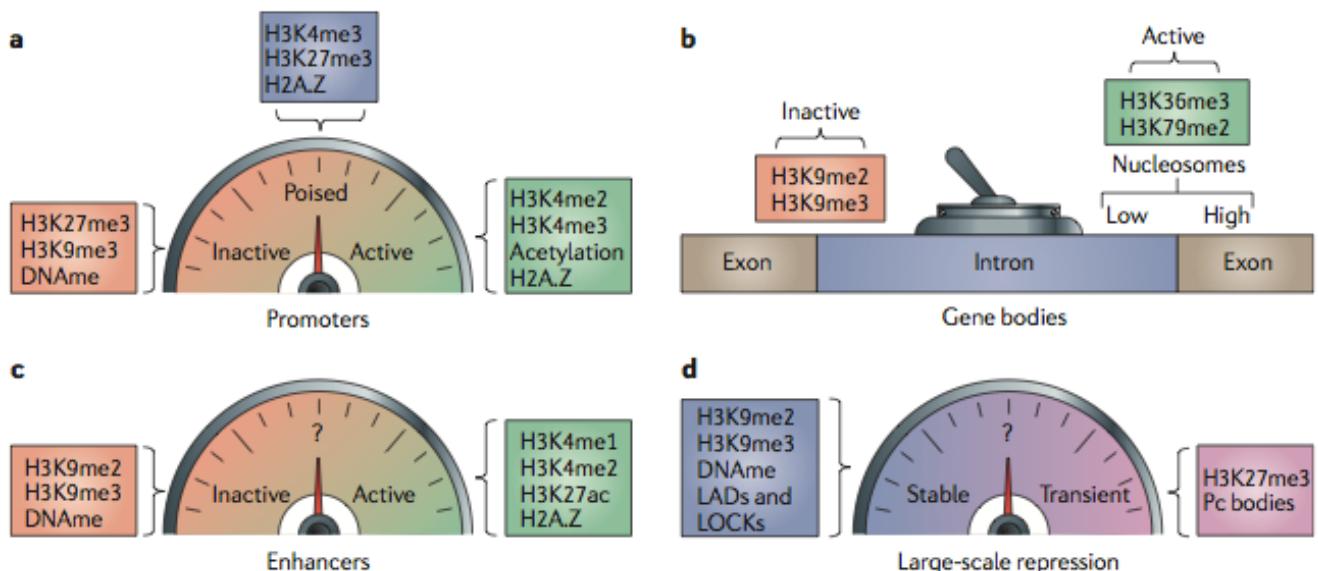


Figure 1.3. Histone modifications and their effects on promoters, enhancers, gene bodies, and larger structures (Zhou et al. 2011).

A number of histone modifications exist which act to affect regulatory elements such as enhancers and promoters. Depending on which histone modification is applied, genes can be induced to be either expressed or repressed (figure 1.3a, c) due to the change in accessibility of their regulatory elements to be bound by transcription factors. Histone acetylation is an epigenetic change that causes condensed chromatin to relax and open, associated with higher rates of

transcription and hence gene expression (Bannister & Kouzarides 2011). Conversely, histone deacetylation and histone methylation instigate tightly packed chromatin, or heterochromatin. Analysing the combinations of histone marks in genomic areas give clues as to what kind of regulatory regions are being observed in the chromatin, for example: enhancer (figure 1.3c), promoter (figure 1.3a), or gene bodies (figure 1.3b, d).

Improper control of epigenetic mechanisms can eventuate in defects and disease. It has been shown that H3K9 histone hypermethylation results in cardiac hypertrophy (Tingare et al. 2013).

1.1.2. Regulatory elements

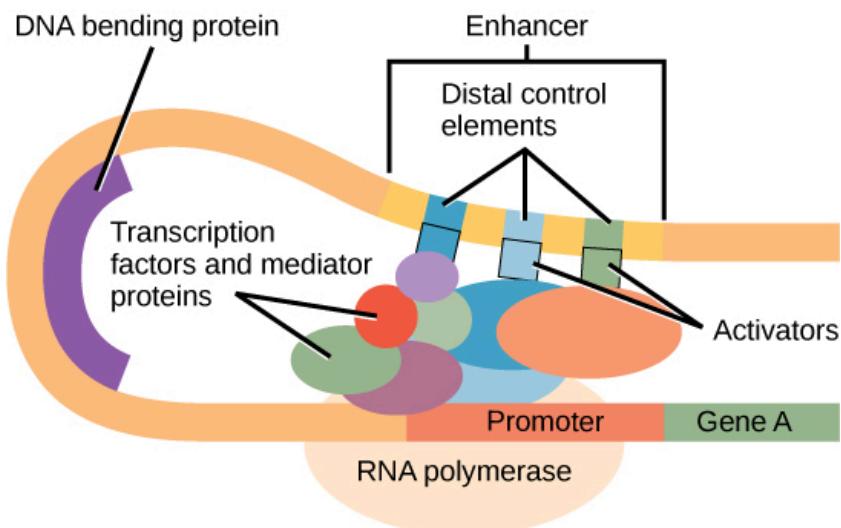


Figure 1.4. How enhancers and promoters interact to enhance gene expression (Boundless Biology 2014).

Regulatory elements are those parts of the chromatin that do not encode proteins or any products, but are bound to by transcription factors to regulate a gene located downstream. Gene expression is dependent on the transcription and eventual translation of the gene into a protein or product which act as effectors in a pathway for a corresponding phenotype. The expression of genes cannot occur unless the relevant sections of chromatin are open and primed for translation. Flanking either sides of a regulatory element are usually specific histone modifications, and so

locating these modifications are good indicators of a nearby regulatory element. Regulatory elements can include enhancers and promoters (figure 1.4), often upstream of their associated gene. However, it has also been found that these elements can be located upstream, downstream, or even in introns (Ecker et al. 2012). Occasionally, enhancers can be located some million or more base pairs away from the target gene (Pennacchio et al. 2013). Enhancers and promoters can affect gene expression depending on which type of transcription factor is bound, and also in what combination. The binding of specific transcription factors in combinations provides the specificity of expression levels. These regulatory elements, associated with specific genes, are bound by specific transcription factors (either activators or repressors) that are responsible for allowing or preventing transcription, respectively.

1.1.3. DNA methylation

DNA methylation is the addition of a methyl group onto cytosine nucleotides. In mammals, up to 70% of CpG islands in the promoter region of a gene are affected (Meehan et al. 1992) (figure 1.2). Genes that are heavily methylated are unable to be transcribed due to the physical restriction of the methyl groups preventing RNA polymerase to pass over the DNA (Meehan et al. 1992). DNA methylation is a necessary element of development and is involved in processes such as embryonic development and gene inactivation (Smith & Meissner 2013). Given this, disregulation of DNA methylation can result in birth defects and disease.

1.1.4. Non-coding RNA

Non-coding RNA molecules, or ncRNAs, are RNA molecules that do not encode for a protein. ncRNAs can act to regulate gene expression, often by acting upon coding RNA molecules (mRNA molecules). MicroRNAs, or ‘miRNAs’, can have either perfect or partial complementarity to mRNA molecules and can bind to ‘silence’ or reduce gene expression, shown in figure 1.2 (Fabian et al. 2010).

In addition to ncRNAs, long non-coding RNA molecules (lncRNAs) act to regulate gene expression in development. It was found that a lncRNA, named Braveheart (Bvht) is a necessity for the correct development of mesoderm towards the cardiac lineage (Klattenhoff et al. 2013).

miRNAs act in a number of ways to regulate gene expression. Most noticeably, a partial complementarity of miRNA to its target can result in the miRNAs hindering translation of its target mRNA into a protein product, as the physical bond prevents ribosomes to translate the molecule (Bartel 2009). Secondly, if the complementarity of the miRNA to its target is perfect, it is possible for miRNA to cleave its target strand. Finally, miRNA can also act to shorten the polyA tail, destabilising the molecule (Fabian et al. 2010). Figure 1.2 shows miRNA acting on the mRNA transcript to regulate gene expression through translational regulation, or cleavage of target mRNA.

There are numerous ways for cells to regulate gene expression, and this results in the formation of a large network of genes, gene products, regulatory elements acting together to achieve gene regulation, termed gene regulatory networks.

1.2. Gene regulatory networks

It has been proposed that diseases and disorders are related to the disruption of particular genes in gene regulatory networks, or GRNs. As defined by Davidson in 2010: gene regulatory networks (GRNs) provide system level explanations of developmental and physiological functions in the terms of the genomic regulatory code (Davidson 2010). GRNs tie together a multitude of information, including gene expression and regulation data. Genes and their products interact with one another frequently for certain developmental or physiological events to occur. Particular genes that interact with a plethora of other genes are termed 'hub' genes due to their comparatively large number of interactions. These hub genes, when disrupted, have been shown to result in heart disease (Vidal & Cusick 2011). For this reason, it is interesting to uncover novel significant hub genes and to investigate the effects of the disruption of these important players in gene networks.

1.3. Investigating gene expression and regulation

Acquiring information on gene networks, including gene expression and regulation, will uncover which genes are important for a specific biological function. Both gene expression and epigenetic changes can be detected using various methods, some qualitative, and others

quantitative. For investigating gene expression, *in situ* hybridisation allows visualisation of gene expression in the form of a coloured marker binding to mRNA, and RNA-seq provides quantitative data on all species of RNA in a sample. For investigating epigenetics, ChIP-seq and ATAC-seq allow researchers to gather information on regions of chromatin that are interact with DNA-binding proteins. A number of different techniques exist to investigate gene expression and regulation. Below details a selection of the most common techniques.

1.3.1. *In situ* hybridisation

In situ hybridisation (ISH) is one of the most common methods used to visualise spatial distribution of gene expression. A probe, labeled by a specific antigen, is bound to a specific gene or gene product. When the antibody, conjugated with an enzyme, specific to the antigen labeling the probe is introduced to the sample, the antibody binds with the antigen. The complementary substrate to the enzyme attached to the antibody can then be added to the sample and a coloured precipitate forms. The organ or embryo can then be inspected using a light microscope (Woodruff 1998). Any areas that are coloured express the gene of interest. This provides fantastic spatial gene expression visualisation within a physical sample and it can be performed on whole-mount embryos, tissues or organs. However, *in situ* experiments may be run only on few genes at a time, and the expression information is purely qualitative.

1.3.2. RNA-seq

RNA-seq involves reducing a sample to its RNA components, and sequencing these to find out levels of gene expression across the entire genome (Wang et al. 2009). RNA species of every type within the cells are incubated with adaptors that attach onto the ends of the RNA, and the sample is sequenced. A dictionary of 'reads' is returned, each row of reads indicating the level of transcripts produced by that corresponding gene.

Information on the transcriptome means that gene expression levels can be investigated genome-wide without bias, and quantitatively. RNA-seq experiments produce a large data set that can be analysed quantitatively using bioinformatics, and statistical significance can be tested. RNA-seq can even be performed on a single cell level, which paves the way for extremely high

resolution sequencing (Wu et al. 2014). Although RNA-seq is powerful, it does not have the benefit of retaining spatial expression information as *in situ* hybridisation does.

1.3.3. ChIP-seq

Chromatin immunoprecipitation sequencing, or ChIP-seq, allows researchers to locate specific places on chromatin where specific histone modifications have occurred, or where specific transcription factors have bound to regulatory elements (Barski et al. 2007). ChIP-seq has previously been used to identify heart enhancers in mice (Blow et al. 2011).

A typical ChIP-seq process begins with the cross-linking of proteins to DNA. The cells are then sonicated to shear the DNA. Loci with DNA-binding proteins are unaffected by the sonication. Once the cells are lysed, the sample containing protein-bound DNA is conjugated with bead-attached antibodies specific to the protein of interest in preparation for immunoprecipitation (IP). IP is used to filter the DNA with proteins attached from the rest of the content, including unbound DNA, in the sample. The precipitate will consist only of the DNA and proteins of interest due to the weight of the bead-attached antibodies attached to the proteins. The DNA is then unlinked from their proteins, sequenced, and mapped to the genome.

ChIP-seq provides valuable information on histone marks genome-wide, allowing researchers to locate areas of chromatin that are either accessible or closed. ChIP-seq can even be utilised to investigate DNA methylation. Unlike RNA-seq, this technology is not readily available on a single cell resolution, but it gives insight into gene regulation.

1.3.4. ATAC-seq

Assay for Transposase-Accessible Chromatic using sequencing, or ATAC-seq, is a novel technique used to gain information on gene regulation in terms of chromatic accessibility and DNA binding factors (Buenrostro et al. 2013). This technique has superseded DNA sequencing due to its comparably low requirement for input cells. Additionally, ATAC-seq allows for the unbiased interrogation of DNA binding molecules, unlike ChIP-seq, meaning that robust results may be generated from ATAC-seq experiments.

ATAC-seq is based on the principle of using the Tn5 transposase enzyme to insert adapters into accessible regions on chromatin. Areas that have not had the adapter sequence inserted are cleaved, leaving only the fragments with Tn5 inserted into them. These chromatin fragments are amplified and sequenced to give information on chromatin accessibility.

Due to its fast generation, ATAC-seq may be used soon in the future for epigenomic profiling in a clinical setting. As well as this, ATAC-seq can be used in conjunction with other techniques, including RNA-seq and flow cytometry. Use of bioinformatics will be useful to derive valuable clues into regulatory networks.

1.4. 3D modeling and gene expression

Intuitively, viewing a three-dimensional representation of a physical three-dimensional object makes for higher quality understanding of positional information. Techniques that allow visual representations of gene expression such as *in situ* hybridisation and histological samples show spatial information, but are limited in their uses. Samples that are processed using histology are output into two-dimensional images. Although these images are useful, it can be difficult to recapitulate structures in context of the entire organ or embryo. For this reason, as technology progresses, there are an increasing number of 3D models of samples for easier viewing.

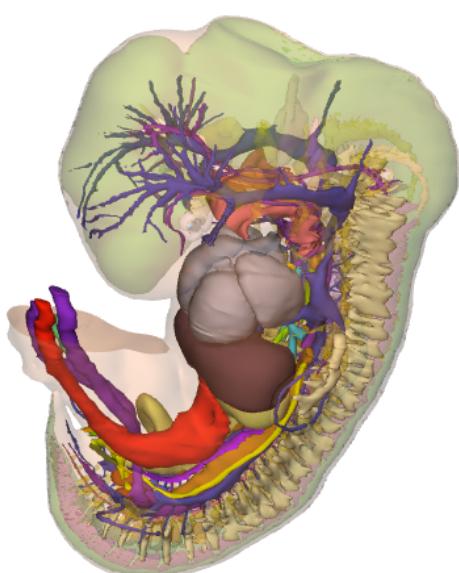


Figure 1.5. 3D model of the human embryo at Stage 16, E39, 10.5mm (de Bakker et al. 2012).

Numerous organs, tissues, and even whole embryos have been modeled in 3D for various uses and clinical applications. With the use of graphic processing software such as Amira by FEI software (FEI Visualization Sciences Group 2012), 3D modeling has advanced to the point of selectively viewing layers, compartments, and anatomical segments. A striking example, shown in figure 1.5, is the 3D model of a human embryo, in which over a dozen different organs and systems can be visualised, all in context of the rest of the organism (de Bakker et al. 2012). The wrist and hand of a human has also been modeled in 3D, providing all of the previous benefits, but additionally allowing a 3D printout to be produced (McMenamin et al. 2014). This printout can then be used for teaching purposes without the use of biological cadavers, overcoming cultural and ethical hindrances.

Given the success of these projects, it then stands to reason that the 3D modelling of the heart will aid greatly in visualisation of various aspects, including gene expression. It is possible to quantitatively sequence the entirety of an organ or embryo and visualise expression in 3D, as with a technique called RNA tomography sequencing.

1.4.1. Tomo-seq

The value of 3D imaging is recognised widely, and some researchers have recreated 3D structures digitally. Recently, Junker et al combined the concepts qualitative measure of gene expression using *in situ* hybridisation with the quantitative measure using RNA-seq in a process termed RNA tomography sequencing, or ‘tomo-seq’ (Junker et al. 2014).

Tomo-seq is performed by slicing each of three separate samples along the three body axes and sequencing each ‘slice’ using RNA-seq. With this data, it is then possible to reconstruct genome-wide gene expression of the sample in three dimensions using a mathematical function.

In addition, tomo-seq greatly reduces the number of samples required for RNA-seq, and hence also the cost. For example, a sample whose resolution of $5 \times 5 \times 5$ is desired, using RNA-seq would require $5^3 = 125$ samples for sequencing, but using tomo-seq, three samples each cut into 5 slices would only require $5 \text{ slices} \times 3 \text{ samples} = 15$ samples for sequencing.

Although tomo-seq provides very high resolution genome-wide transcriptomic data, no studies have been conducted using tomo-seq to give data on epigenetic state of the samples.

Furthermore, another downside to tomo-seq is that the technique is conducted on three separate samples. Each sample would have intrinsic biological and physical variance, making dissection difficult to perform accurately between samples. Adding to that, the technique would require that the slicing be performed in perfect orthogonal relation to the other samples in order for the triangulation algorithm to produce correct output. Any slight deviation from the ideal orthogonal angles will result in inaccurate results.

Even with shortcomings considered, tomo-seq may eventually replace *in situ* hybridisation as it is quantitative, genome-wide, gives spatial information, and is comparatively less expensive to run in terms of resolution. As there have not yet been any studies undertaken on gene regulation in 3D, it may then be possible to use this tomo-seq in conjunction with a tool that examines gene regulation, such as ChIP-seq or ATAC-seq.

1.4.2. Drop-seq

Recently, a technique was devised which allows transcriptomic characterisation of a sample at the single-cell resolution (Achim et al. 2015). The technique is termed ‘drop-seq’ due to its utilisation of a microfluidics system, designed specifically to encapsulate single cells in oil droplets (Thorsen et al. 2001). Drop-seq allows the mRNA transcripts of a sample to be sequenced by passing a single cell suspension through the microfluidics system in which the cells are each met with a barcoded bead primer and the pair is captured by a nanolitre droplet of oil. Within each droplet, the cell is lysed to purify mRNA, which can then be reverse-transcribed into cDNA. From cDNA, libraries can be generated for high-throughput sequencing. This enables genome-wide analysis to be performed on thousands of cells at once. Although quantitative and high-throughput, drop-seq does not provide information on the origin of the cell’s location in the tissue.

1.4.3. *In situ* hybridisation on 3D models

A technique has been devised to map gene expression patterns from *in situ* hybridisation (ISH) experiments onto 3D models of the developing heart (Soufan et al. 2003). This meant that the two-dimensional images of ISH were no longer restricted to a 2D image, but instead, a true-to-life 3D recapitulation of expression patterns. The expression patterns were derived from single

scans of ISH histology slides and mapped onto the model computationally, using a novel algorithm. While this method provides an accurate depiction of spatial location of gene expression, the technique utilises ISH data, which intrinsically can be conducted on one target gene at a time. The absence of current next-generation sequencing techniques proves to be a bottleneck in this technique.

Tomo-seq, drop-seq, and ISH on 3D models analysis all utilise various tools to show gene expression patterns in samples, but no one option allows for the retention of spatial, transcriptomic, and epigenomic state information.

1.5. Murine hearts as a model for investigating gene regulatory networks

The heart makes a fantastic model for gene expression control and epigenetics due to its physical complexity and known disease phenotypes. In terms of using animal models, a mouse is ideal as its heart is physically, developmentally, and physiologically similar to human hearts.

In terms of using the mouse model for investigating gene expression in hearts, it has many similarities to human hearts and hence good translatability to medical and clinical applications. Species such as insects and fish have a simple heart with two or three chambers, but mice closely mimic the developmental and physiological processes as well as the anatomical structures of the human heart. Mouse hearts are compartmentalised in the same way human hearts are: they have four chambers, two of which are atria, and the other two of which are ventricles. de Boer et al have developed a 3D atlas of a mouse heart which has high enough resolution to show these features (de Boer et al. 2012). Additionally, over 99% of the 30,000 genes in mice have homologs in humans, meaning that numerous genes are shared between the species (Gunter & Dhand 2002). A multitude of genes that have been identified as significant in mouse models for heart diseases have been found to have human homologs, such as Nkx2-5 and Wnt11, two major drivers in heart development (Costa et al. 2013; Terami et al. 2004).

It has been previously shown by that neonatal mice retain the ability to regenerate cardiac tissue after injury (Porrello et al. 2011). This is a feature that was previously thought to be exclusive to amphibians and fish; however, this study showed that neonatal mice younger than 7 days of age will repair damage caused to the heart after resection of the ventricular portion of the heart. Although only up to one fifth of the ventricular tissue could be resected, this transient ability to regenerate heart tissue is striking. The study showed through the use of fate mapping that the cells produced to replace the missing tissue were from existing cardiomyocytes; however, it was not determined which genes were responsible for the regeneration of the heart.

Disrupted or ablated expression of particular genes in the heart results in heart disease. A plethora of genes and their associated defects have been identified (Waardenberg et al. 2014). There are a multitude of heart diseases that are localised to specific sub-anatomical regions within the heart, such as septal defects, myocardial defects, or defects restricted to atria or ventricles only. It was recently found that transgenic mice with mutations in the GATA4 gene has a higher incidence of atrial septal defects (Han et al. 2015). Similarly, mutations in the promoter region for GATA6 result in ventricular septal defects (Li et al. 2014). A number of defects within compartments within the heart can lead to various forms of congenital heart disease (CHD) (Bruneau 2008). A selection of CHD can persist into adulthood, causing a decreased quality of life for those affected. Approximately 80% of CHD are genetically inherited, but have causes currently unknown (Van Der Linde et al. 2011). Given the spatial specificity of the defected areas in the heart from these mutations, it may be interesting to attempt to uncover significant changes in differential gene expression in the different compartments of the heart.

1.6. Conclusion

The spatial and temporal control of the expression of genes is undisputedly one of the key factors in the highly specified development of organisms. Epigenetics is another contributing facet to development, responsible for the physical control of transcription of DNA. Both gene expression and epigenetics can be investigated in a spectrum of ways, each with their own advantages and disadvantages. *In situ* hybridisation, RNA-seq, ChIP-seq, and ATAC-seq are the most common

techniques that are utilised, but represent a mere handful of techniques that exist today. As technology climbs and advances, the next logical step is for the manipulation and visualisation of this data to be made as easily interpretable as possible. 3D modelling offers a life-like and intuitive reconstruction of organs and whole organisms. Three-dimensional imaging provides an invaluable resource for a wealth of uses, including better contextual understanding of spatial gene expression, and learning resources. Murine hearts make a good model for investigating gene expression and regulation as they are structurally and developmentally complex and have well-known disease phenotypes. It is hoped that through the 3D reconstruction of genome-wide gene expression and regulation of the adult mouse heart, an understanding of spatial and temporal gene function can be revealed, aiding in understanding fundamental processes underlying disease, development and regeneration.

2. Project outline

2.1. Background

During development, mammalian hearts undergo several morphological changes under strict spatial and temporal control of numerous genes. In order to produce the highly specific segmentation and anatomical structures in the heart, specific genes that control the formation of these individual features must be expressed at the right time and right place (Vincent & Buckingham 2010). When genes that are crucial for heart development are disrupted, heart defects and disease follow (Richards & Garg 2010; Ren et al. 2014).

Although numerous key genes involved in heart development (and hence heart disease) have been identified, only 20% of congenital heart defects (CHD) are attributable to known genetic defects (Blue et al. 2012). CHDs account for almost a third of fetal deaths (Bruneau 2008) and so represent a significant burden on families and hospitals.

Previously, hearts have not been sequenced in high resolution to show quantitative spatial genome-wide gene networks. It has been shown that gene expression patterns during cardiac development could be reconstructed in 3D (Soufan et al. 2003); however, this study showed the expression of only five genes in a qualitative measure using *in situ* hybridisation and is not scalable to the hundreds or thousands of genes involved in heart development. For a quantitative method, RNA-seq performed on a sample, such as a whole embryo or organ, provides precise information on genome-wide gene expression (Wang et al. 2009), but spatial information is lost after sequencing. To counter this, it is possible to use multiple fragments of a sample for RNA sequencing to retain spatial information on gene expression (Combs & Eisen 2013). Furthermore, ATAC-seq (Buenrostro et al. 2013) provides information on protein interactions with DNA and hence offer further clues on gene regulation.

From this, we hypothesise that finding the expression and regulation patterns of every gene within specific sub-compartments in an organ will reveal the specific subset of genes that play an essential role in that anatomical structure.

2.2. Aims

1. To adapt a 3D model of the adult murine heart to our study.
2. To section an adult murine heart in preparation for sequencing techniques.
3. To derive and analyse the transcriptomic and chromatin accessibility signature of sub-anatomical compartments of the adult murine heart.
4. To integrate transcriptional and regulatory information onto a 3D model.

2.3. Experimental design and methodology

Aim 1: To adapt a 3D model of the adult murine heart to our study. An established three-dimensional model of the adult murine heart (Soufan et al. 2003) will be adapted for use in our study. 3D modeling software will be used to computationally ‘dissect’ the model of the heart into virtual sections or “volumes”, reflecting closely the biological samples that will be used for sequencing.

Aim 2: To section an adult murine heart in preparation for sequencing techniques.

Adult mice will be culled and their hearts will be extracted. Hearts will be sectioned along the dorsal-ventral axis to produce five transverse slices of approximately 2mm. Each slice will then be manually dissected by anatomical structures; that is, the atria and ventricles, and left and right structures will be separated from each other.

Aim 3: To derive and analyse the transcriptomic and chromatin accessibility signature of sub-anatomical compartments of the adult murine heart. Each piece will be sequenced and analysed using RNA-seq and ATAC-seq to acquire genome-wide gene expression and gene regulation of the heart. The resulting data will be analysed by supervised and unsupervised methods in order to reveal spatial signatures used to attempt to reconstruct this information into a transcriptome and epigenome map to be mapped onto the 3D model from Aim 1, and to reconstruct cardiac gene networks that could be involved in heart development and disease at a sub-anatomical resolution. Validation will be done by comparing our 3D model and transcriptome map with existing *in situ* hybridisation data on known genes.

Aim 4: To integrate transcriptional and regulatory information onto a 3D model. An algorithmic function will be designed to assign gene expression values onto volumes on the 3D model. Existing RNA-seq and ATAC-seq data will be mapped onto the model to validate the function, and *in situ* hybridisation data will be used to validate the expression patterns.

2.4. Anticipated outcomes

A versatile 3D model of the adult murine heart will be developed that allows RNA-seq and ATAC-seq data to be mapped onto it using a mathematical function.

To prepare for sequencing techniques, adult murine hearts will be sectioned according to sub-anatomical features. Anatomical structures will be separated from each other and prepared for sequencing techniques to be applied.

Heart segments, once sequenced using RNA-seq and ATAC-seq, will be translated into transcriptome and epigenome maps. These maps will subsequently be reconstituted onto the 3D model of the heart to visually depict the spatial expression and regulation of genes.

Through cluster analysis, we will reveal the 3D distribution of gene networks within the heart which will uncover, for the first time, interactions between sub-networks in 3D and thereby identify novel sub-anatomically restricted genes related to heart development and disease.

3. Materials and Methods

3.1. Heart extraction and microdissection

3.1.1. Extraction

Wildtype C57/BL6 female adults mice of approximately 6-10 weeks of age were culled using cervical dislocation method, then sprayed with 80% v/v ethanol and immediately dissected through the abdomen, under the sternum. Once the diaphragm was dissected away to access the upper abdominal cavity, the ribcage was cut and lifted to expose heart and lungs.

To clear blood from the heart, the heart was perfused. Small incisions were made on each lobe of the liver to aid in bloodletting. Perfusion was performed over one minute with 10mL of 1x Hanks' Balanced Salt Solution (HBSS, supplied by Life Technologies) through the left ventricle. HBSS was delivered into the left ventricle using a 10mL syringe with a 26-gauge needle.

Hearts were removed by grasping heart by the root and cutting through major vessels and surrounding connective tissue. Once removed, heart was placed in 310mO (mouse osmolarity) Phosphate Buffered Solution (PBS, prepared by the Australian Regenerative Medicine Institute Internal Store according to receipt in Appendix A) to clean up excess fat, connective tissue, lungs, thymus, and trachea.

3.1.2. Dissection into slices

To prevent the jamming of the mouse heart slicer with the more delicate tissue at the most superior portion of the heart, the isolated heart was first dissected by removing the top fifth of the heart using Vannas scissors.

A Stainless Steel Mouse Heart Slicer Matrix with 0.5mm coronal section slice intervals (Zivic Instruments) was used to ensure slices of even and consistent size. Grooves on the mouse heart slicer are 0.5mm apart. The bottom 4/5 of the heart was placed in the mouse heart slicer with the apex pointed towards the outer edge of the mouse heart slicer (figure 3.1b, right).

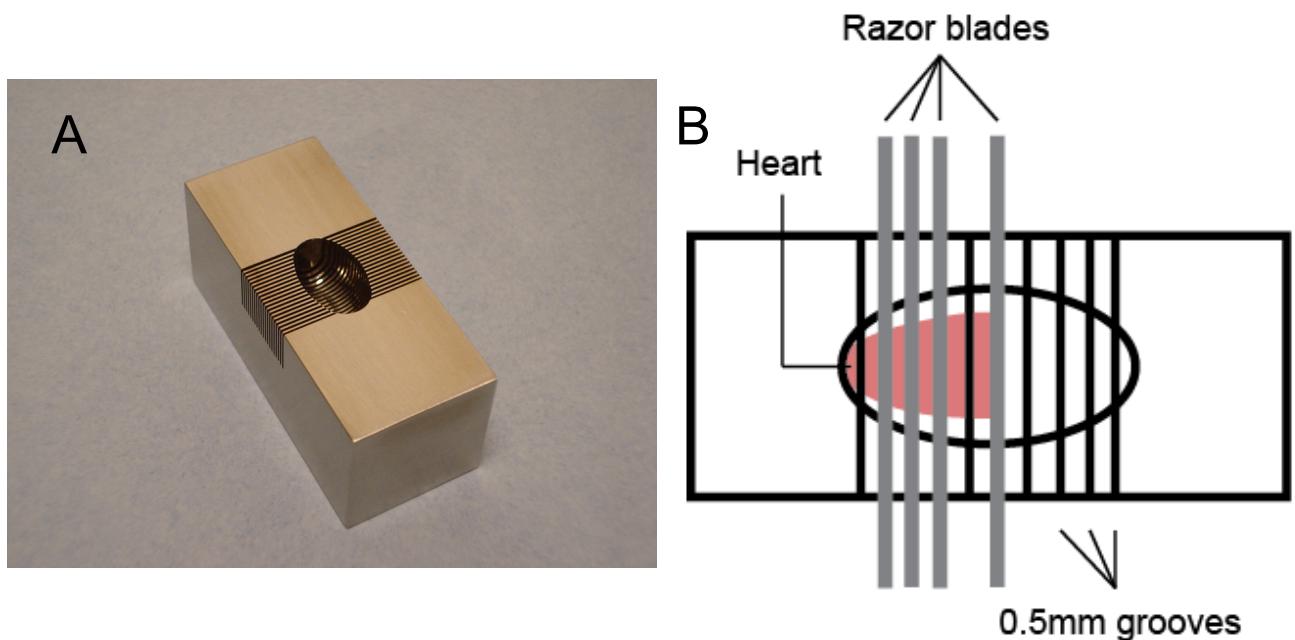


Figure 3.1.a) Layout of heart in mouse heart slicer. b) A diagram of how the ventricular portion of the heart was placed and sliced in the mouse heart slicer.

To ensure the heart does not move when slicing, a razor blade was inserted at a point where the heart is firmly held in place (the 'anchor' razor blade, far right razor blade in figure 3.1b), between the 1/3 – 1/2 point. Because the top fifth of the heart has already been removed, the remaining mass of the heart needs to be sliced into only four more slices. So, three razor blades are placed equidistantly along the heart, which will result in four heart slices being made. Once in place, the three razor blades are pressed firmly down simultaneously to cut through the heart evenly.

To remove the slices, first, the anchor razor blade is removed to reveal the second topmost slice (Slice B). This slice can be taken from the slicer and placed in PBS on a Petri dish. The next razor blade can then be removed for the next slice to be revealed and placed in PBS, and so on, until the final razor blade is removed and the fifth and final slice (Slice E) is taken and placed in PBS.

3.1.3. Dissection into pieces

Each of the five slices was individually dissected into a total of 18 pieces of the heart using Vannas scissors. Figure 3.2 shows the dissection plan used for the heart samples.

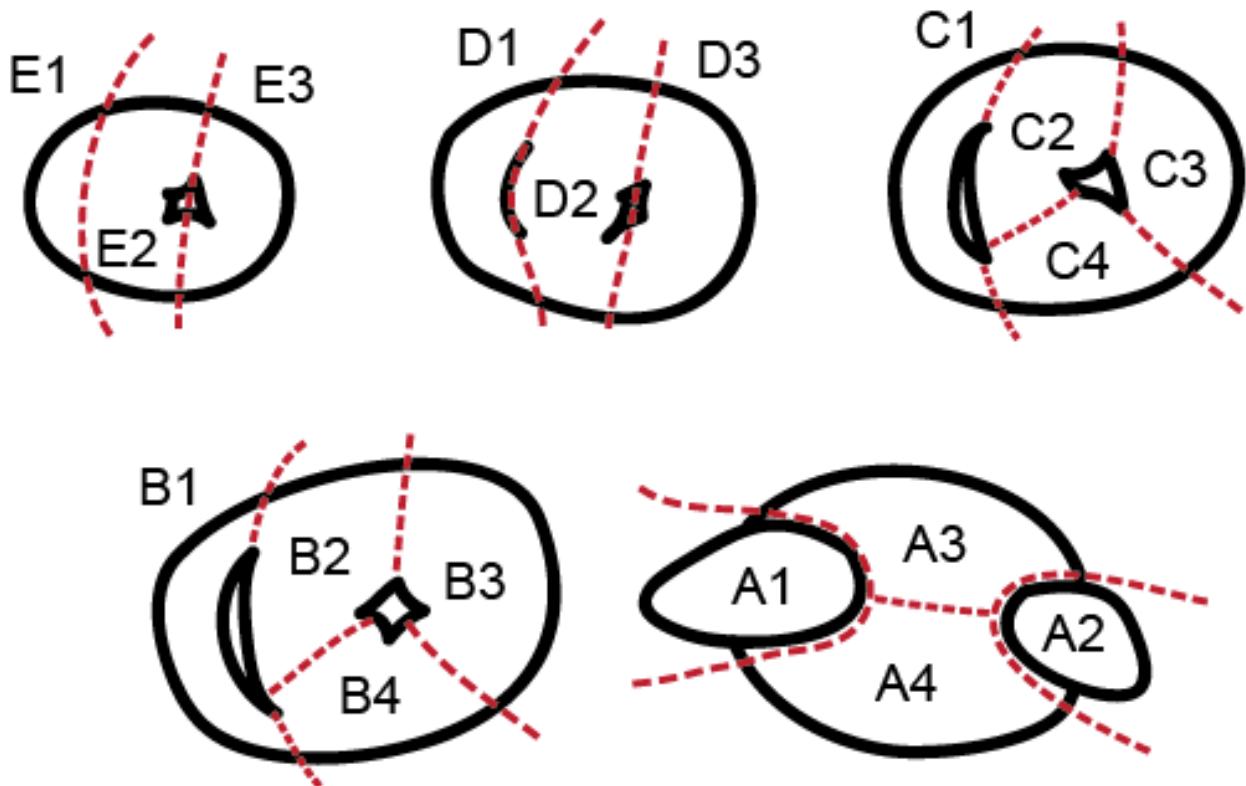


Figure 3.2. Dissection map of the adult murine heart, comprising 18 pieces.

Slice E: Starting from the most apical end of the heart, slice E was cut to isolate the rightmost side of the apex where the right ventricle begins. The remaining left ventricle is bisected, resulting in a total of three pieces for the bottom slice.

Slice D: This slice contains the most inferior part of the lumen of the right ventricle, therefore leaving a small space that allows the free wall of the right ventricle to be easily separated from the rest of the slice. Once the free wall of the right ventricle is dissected away, the remaining left ventricle is bisected to divide left and right sides. This resulted in a total of three pieces for slice D.

Slice C: Slice C includes the middle portion of the right ventricle lumen and is easily dissected away from the thick wall of the left ventricle. The remaining left ventricle, much larger

than the previous slices, can now be dissected into three pieces. This resulted in a total of four pieces for slice C.

Slice B: Slice B is very similar to slice C in terms of overall structure and is dissected in the same way. This resulted in a total of four pieces for slice B.

Slice A: At the most superior part of the heart, a number of major vessels aggregate. Slice A reaches from the bottom of the sinus venosus to the remainder of the heart, including both atria. First, the external portions of the atria are separated. The remaining ‘cap’ of the heart is bisected to give anterior and posterior pieces. This resulted in a total of four pieces for slice A.

3.2. Tissue preparation

3.2.1. Enzymatic digestion

To gather both transcriptomic and epigenomic data from the heart pieces, enzymatic digestion was first attempted. After the method described in Section 3.1.3, ‘Dissection into pieces’, heart tissue pieces were placed into 4mL Fluorescence-Activated Cell Sorting (FACS) tubes and minced using Vannas scissors into fine pieces. Each sample had 500ul collagenase I (Worthington) added to it, then incubated for 20 minutes on a shaker at 37°C. Every five minutes, the samples were triturated for 5-10 seconds each time using needle and syringe, as outlined in table 3.1.

Stage	Minutes spent incubating	Needle gauge
1	5	18
2	10	18
3	15	21
4	20	21

Table 3.1. Summary of stages of trituration in collagenase I digestion.

After the final trituration, each sample was strained through a 40um mesh into a 4mL FACS tube and washed twice using 3mL of 310mO (mouse osmolarity) PBS. 10µl of the samples were

taken for FACS analysis as described in Section 3.3. The remaining volume of the samples was centrifuged at 400g for 5 minutes and excess PBS was aspirated.

The cell suspensions had individual cell concentrations calculated using a haemocytometer. Once cell concentration for each sample was known, 50,000 cells were isolated from the samples and used immediately for ATAC-seq, as described in Section 3.5. The remaining volumes of the samples were frozen at -80°C for RNA purification, as described in Section 3.4.

3.2.2. Tissue homogeniser

Subsequent quantification and qualification of samples generated from the enzymatic digestion method as described in Section 3.2.1 showed low yields and quality. As a result, new heart tissue pieces were derived using methods described in Sections 3.1.1 – 3.1.3.

Heart tissue pieces were placed in 250ul Buffer RLT as supplied by QIAGEN RNeasy Micro Kit. A total of 18 pieces were prepared. Due to the large number of samples, 12 pieces were stored at -80°C while the other pieces were homogenised to optimise sample viability. All samples were homogenised using a Bio-Gen PRO 200 Homogeniser (PRO Scientific, Inc.). Each piece was homogenised for 30 seconds on setting 2 (medium speed). After every homogenisation, the homogeniser tip was cleaned using 80% v/v ethanol for 15 seconds, then deionised water for 15 seconds.

Homogenised tissue samples were immediately used in RNA extraction and purification as described in section 3.4.

3.3. FACS analysis

For viability analysis, 10µl from each of the single cell suspensions from Section 3.2.1 was incubated with propidium iodide (1:500 dilution, Sigma-Aldrich). FACS analysis was done using a BD Biosciences LSR II Flow Cytometer. Gating was set firstly to exclude any debris, and secondly to exclude cells positive for PI staining.

3.4. RNA purification and RNA-seq

RNA purification: RNA extraction and purification of all samples was done using a QIAGEN RNeasy Micro Kit according to manufacturers' instructions.

Library preparation: An Illumina TruSeq Stranded Total RNA kit was used to prepare the libraries for poly-A enrichment for mRNA. For sample amplification, 15 PCR cycles were used, standard to the Illumina kit protocol.

RNA sequencer: For RNA sequencing, an Illumina NextSeq 500 was used on high output for 75bp paired end reads, with 20 million reads per sample.

3.5. ATAC-seq

Cell preparation, transposition reaction and purification, PCR amplification, and qPCR were all performed according to the protocol from Buenrostro et al (Buenrostro et al. 2015).

In brief, the cells are prepared by isolating 50,000 cells and washing cells with PBS and spun. For the transposition reaction and purification of chromatin, a Nextera Tn5 Transposase kit is used in conjunction with a Qiagen MinElute PCR Purification kit. The transposed DNA is eluted in 10µl of elution buffer, again from the MinElute kit. The purified DNA is stored at -20°C overnight.

The DNA fragments are amplified using PCR. Library quality control was done using gel electrophoresis as well as with a Bioanalyser.

3.6. Computational model

The 3D graphical model of the adult mouse heart was provided by J. M. Ruijter and A. F. Moorman from the Heart Failure Research Centre in the Netherlands. The original files were in the Amira software file format (.am and .surf). Amira was used to export these files into Wavefront .obj files, which could then subsequently be used with 3D modeling software (Maya).

3.6.1. Software used

For adapting the heart model for the project, a number of options for software were available. Careful consideration was put into each software, taking into consideration each of their advantages and disadvantages, summarised in table 3.2.

Stage	File/software	Comments
Heart model	Generic model	<ul style="list-style-type: none"> Several anatomically accurate models of the heart available Only human hearts, but close enough to be used as mouse hearts Artists' renditions, i.e.: not based off biological samples of hearts
	Real model (Jan Ruijter's)	<ul style="list-style-type: none"> Generated using 80 histological sections of an adult murine heart
Conversion of model file	Amira	<ul style="list-style-type: none"> Original model was made using Amira No other options available for conversion
3D modeling	AutoCAD	<ul style="list-style-type: none"> Difficult to use
	3DFree	<ul style="list-style-type: none"> Difficult to use
	Maya	<ul style="list-style-type: none"> Easy to use Large user base, good support available
Programming	Java	<ul style="list-style-type: none"> Requires expertise in graphical manipulation
	Python	<ul style="list-style-type: none"> Requires expertise in graphical manipulation Popular language for simulations
	Unity	<ul style="list-style-type: none"> Large user base, good support available Has intrinsic capability to manipulate 3D objects Can use C# for programming (very similar to Java) Can import .fbx files

Table 3.2. Summary of features of software

3.6.2. *In silico* slicing and sectioning

Computational slicing and sectioning of the 3D model of the heart was done using Maya 3D modeling software (2015). The heart was first sliced into 5 transverse pieces as with the biological samples. Slicing of the 3D model was done using the “Slice” tool in Maya, which allows a straight line to be drawn to cut the object completely through. Each slice was then ‘sealed’ to give the appearance of solid tissue. The ‘sealing’ of the slices was done by adding faces individually to the model until the slice was completely covered. Once sealed, each slice was then sectioned into 18 pieces, again using the Slice tool. Similar to the slicing process, each piece was subsequently sealed by adding faces.

3.6.3. Programming

A program was written to integrate RNA-seq datasets onto computational pieces of the heart using the C# programming language. The general algorithm of the program is detailed in box 3.1.

```
for each piece
    load dataset for that piece
    store dataset into hashtable
    upon clicking
        search the hashtable for that gene name
        if (the gene name is found)
            paint that piece with the appropriate intensity
        else if (the gene name is not found)
            return error message
            suggest a similar gene name
done
```

Box 3.1. General algorithm for mapping datasets onto the 3D model of the heart.

3.7. RNA seq pipeline

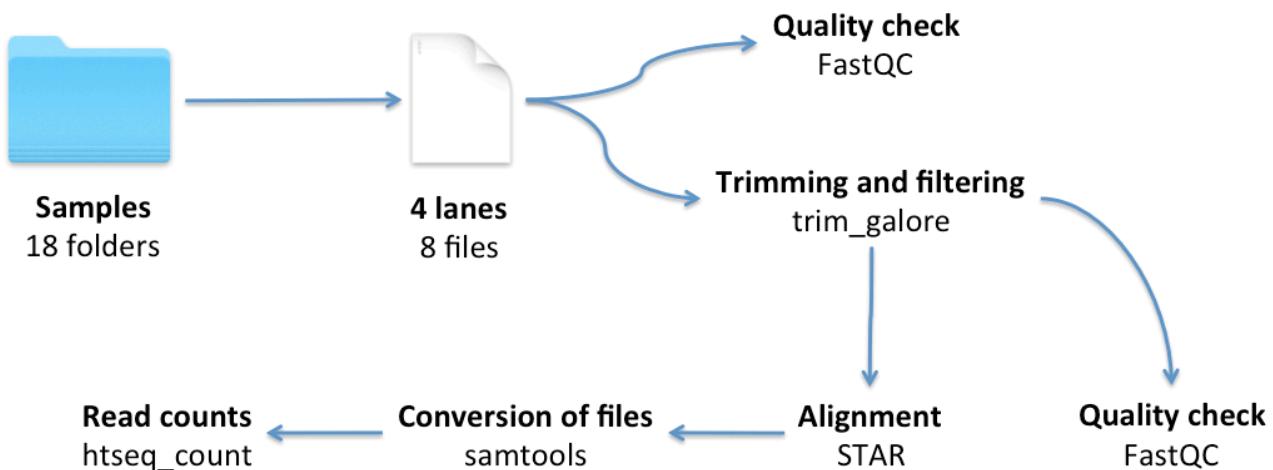


Figure 3.3. RNA-seq data analysis pipeline, showing key software involved in the processing of the data.

The raw data received from the sequencing facility was processed bioinformatically to deduce read counts. A number of software was run in succession, automated by a Bash script (provided in Appendix B). This script requires that there is adequate harddisk space available.

A simple overview of the pipeline is visualised in figure 3.3. The pipeline begins with the quality checking of the raw data. The raw data is then trimmed and filtered to remove redundant and low quality reads. After this, a quality check is run on the trimmed and filtered datasets for comparison with the raw data quality check. The trimmed files can then be aligned (or ‘mapped’) onto a reference genome, in this case, *Mus musculus* 10. These mapped reads can then be visualised, or read counts can be generated.

3.7.1. Batch analysis setup

The datasets for the RNA samples comprises of one folder for each of the 18 samples sent for sequencing. Each folder contains datasets from each of the four lanes the samples were run on. Because the samples sequenced were for paired end reads, there are two files per lane: R1 and R2. R1 contains the forward reads, and R2 contains the reverse reads. Although this effectively halves the number of reads produced per sample, paired end sequencing provides for more accurate alignment further down the pipeline compared to single reads.

3.7.2. Quality check I

To check the quality of the datasets, each file was run through the software FastQC. FastQC takes .fastq or .fastq.gz (compressed) files as input and produces an HTML file as output. The HTML file details a summary report of the quality of the datasets, including general statistics, information on the number of uniquely mapped reads, and per base sequence quality. Once assessed, the files can be either trimmed or filtered as appropriate. The command to run FastQC is shown in box 3.2.

```
fastqc -f fastq $file
```

Box 3.2. Command to be entered on the bash shell to run FastQC.

3.7.3. Trimming and filtering

Trimming and filtering was done using trim_galore software. trim_galore removes low quality reads from the datasets to ‘clean’ up the data and leave only the higher quality reads.

The parameters used for the trim_galore were:

- q-cutoff of 28
- Paired end reads
- Files are gzipped
- phred33 for Illumina .fastq files
- Adapter overlap/stringency of 3 nucleotides

The command to run trim_galore is shown in box 3.3.

```
trim_galore -q 28 --paired --gzip --phred33 --stringency 3 $file1 $file2
```

Box 3.3. Command to be entered on the bash shell to run trim_galore.

3.7.4. Quality check II

After trimming and filtering, the quality of these files are checked again using FastQC. This second batch of FastQC reports should be checked to ensure that the quality is higher or at least

equal to the reports generated prior to trimming and filtering. The command to run FastQC is shown in box 3.4.

```
fastqc -f fastq $file
```

Box 3.4. Command to be entered on the bash shell to run FastQC.

3.7.5. Alignment/mapping

The trimmed files containing raw reads are then mapped to a reference genome using STAR. For paired end sequencing, STAR takes as input two .fastq or .fastq.gz files (the forward and reverse read files), and outputs one .sam file. The resulting command to run STAR is shown in box 3.5.

```
STAR --runThreadN 8 --outSAMattributes All --genomeLoad NoSharedMemory -  
-readFilesCommand zcat -- genomeDir  
/persistent/reference/mouse_mm10/star/  
--readFilesIn file1.fq.gz file2.fq.gz --outFileNamePrefix mapped_
```

Box 3.5. Command to be entered on the bash shell to run STAR.

After alignment, the pipeline follows two endpoints:

1. Visualisation of aligned reads, and
2. Generating read counts.

3.7.6. Conversion into BAM

.sam files input into htseq-count failed to run. Consequently, the resulting .sam files that were generated from alignment using STAR needed to be converted into .bam files. samtools was used for the conversion of .sam files into .bam files. The execution of samtools occurs in three steps: view, sort, and index.

The first step, viewing, is the conversion of the .sam file into the .bam file.

```
samtools view -bSo mapped_Aligned.out.bam mapped_Aligned.out.sam
```

Box 3.6. Command to be entered on the bash shell to run samtools.

Secondly, the .bam file is sorted using the command in Box 3.7.

```
samtools sort mapped_Aligned.out.bam mapped_Aligned.out.sorted
```

Box 3.7. Command to be entered on the bash shell to run samtools.

Finally, the sorted .bam file is then indexed.

```
samtools index mapped_Aligned.out.sorted.bam
```

Box 3.8. Command to be entered on the bash shell to run samtools.

3.7.7. Read counts

Read counts were generated using htseq-count. htseq-count can use either .bam or .sam files, but due to errors caused by attempting to use .sam files, the same .bam files generated from 3.7.6, for the purpose of visualisation using IGV, were used. The resulting command to run htseq-count was:

```
htseq-count -f bam $file /path/to/mm10/gtf/Mus_musculus_GRCm38.gtf >
"$newFileName""_htseq_out.txt"
```

Box 3.9. Command to be entered on the bash shell to run htseq-count.

3.8. Analysis

3.8.1. Cluster analysis

MeV was used for cluster analysis of the RNA-seq data. MeV supplies an easy-to-use graphical user interface in conjunction with a large array of visualisation tools that employ strong statistical analyses. The statistical analysis technique used for the hierarchical cluster analysis was the Spearman correlation method in combination with complete linkage. It has been shown that the Spearman correlation coefficient produces the most robust and accurate results when working with gene expression data (Jaskowiak et al. 2014).

3.8.2. MDS analysis

Degust, a web application made for the purpose of analysing and visualizing RNA-seq data, was used to generate a multi-dimensional scaling (MDS) plot. Degust utilises the R programming language which is a powerful tool used by countless bioinformaticians for data processing and analysis. Degust also makes use of the limma/voom package within R, which allows the creation of simple graphs and plots using statistically sound analysis.

4. Results

4.1. Adaptation of 3D model of adult murine heart

4.1.1. Sectioning

In preparation for integration with RNA-seq datasets, a 3D model of the adult murine heart was adapted for the study. Aanhaanen et al from the Heart Failure Research Centre in the Netherlands provided a model of the whole adult murine heart. Using Maya (3D modeling software, as described in Section 3.6.1), the computational model of the heart was digitally cut first into 5 slices and then sealed, as shown in figure 4.1.

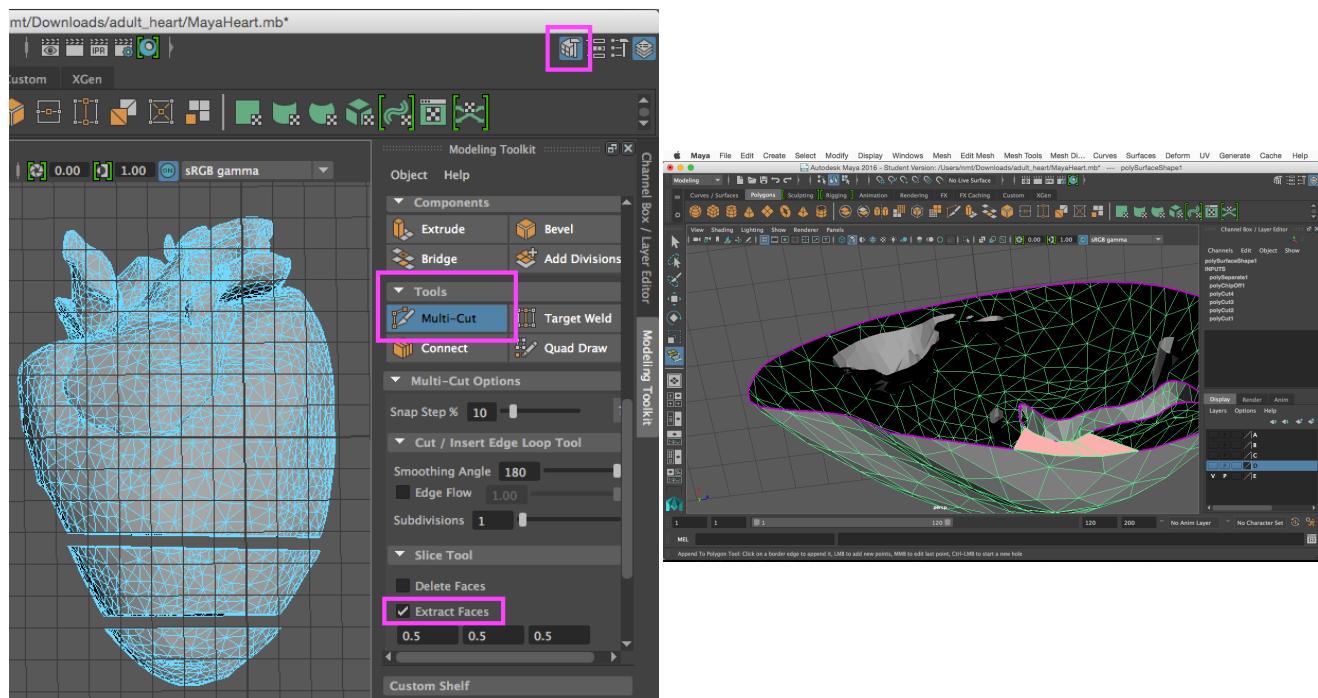


Figure 4.1. Screen captures of the Maya interface and 3D model, showing part of the process of generating and sealing heart slices from the original model.

After the slices were generated, each slice was further computationally dissected into pieces. This process is shown in screen captures in figure 4.2.

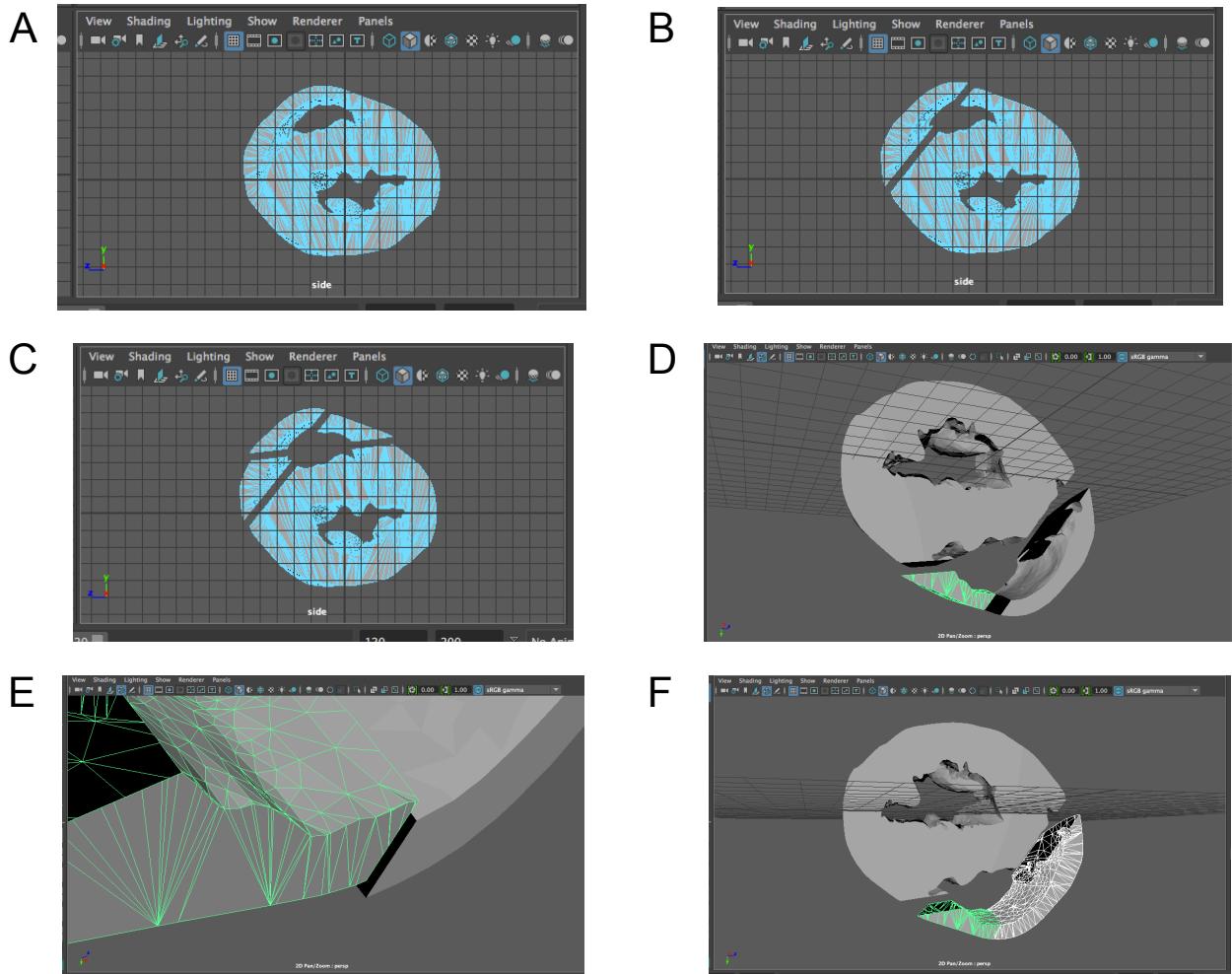


Figure 4.2. a) A sealed slice of heart. b) The first cut into the slice to separate the right ventricle free wall from the piece. c) The second slice into the same slice. d) The same slice, flipped 180 degrees. The cut-up pieces of the right ventricle free wall need to be joined together to form one continuous piece. e) A progress screen capture, showing the matching up of two components from the same piece. f) The final result of the right ventricle free wall piece, joined together into one object.

Once each slice had been dissected further into pieces (as described in Section 3.6.2), a total of 18 distinct pieces were created, shown in figure 4.3. These pieces closely replicate the biological dissections conducted to obtain the transcriptomic datasets, as described in section 3.1.3.

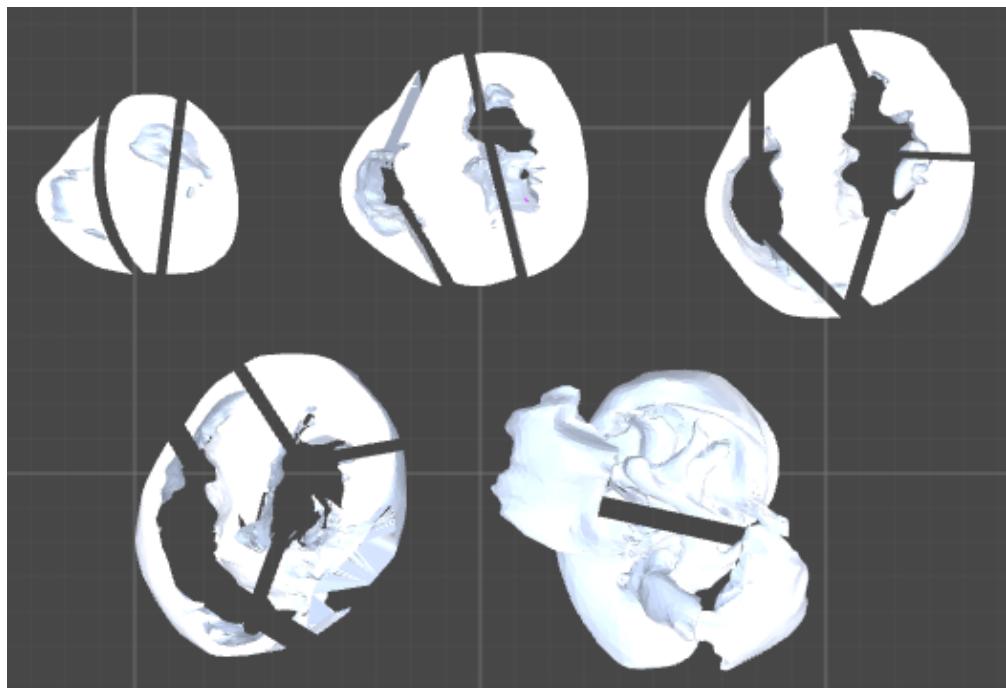


Figure 4.3. Computationally dissected 3D model of the heart, showing 18 pieces.

When the pieces are reassembled to reflect their anatomically correct positions in the heart, the 3D nature of the model allows numerous views of the dissected heart, as shown in figure 4.4.

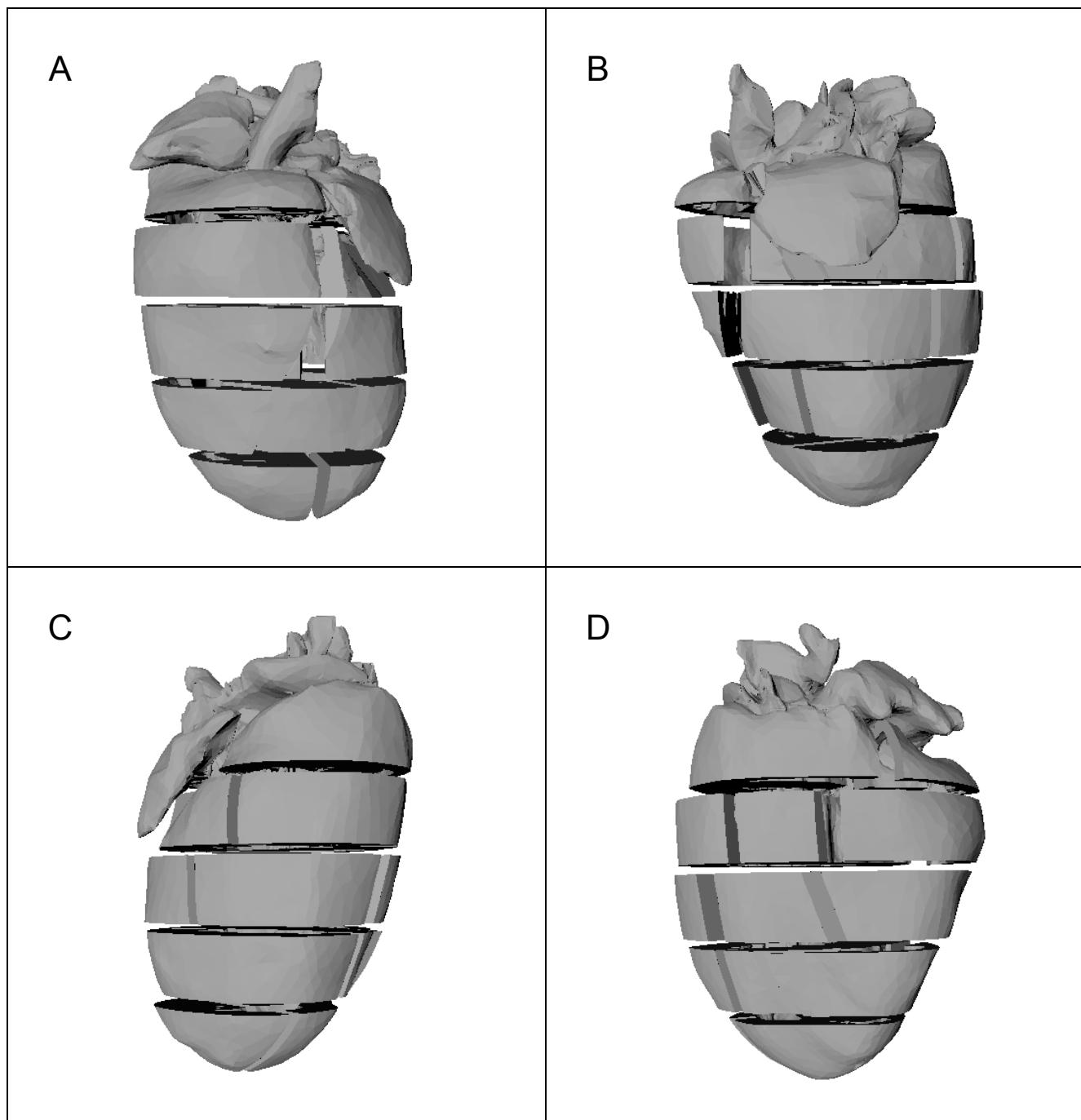


Figure 4.4. Different views of the sectioned 3D model of the adult murine heart. a) Ventral view. b) Lateral right view. c) Dorsal view. d) Lateral left view.

4.1.2. Preparation of 3D model to accept quantitative datasets

The model and the subsequently generated pieces were not able to accept any datasets to display gene expression. For this study, software code was written to allow the mapping of transcriptomic datasets onto the model of the heart algorithmically. This code is provided in Appendix C.

Currently, the software requires that the hexadecimal values of the heatmap colours for each of the pieces are manually entered. This process is tedious but it is a future goal to automate this process.

An existing computational model of the adult murine heart was adapted to our study, first by digital dissected using 3D modeling software, and second by associating software code to the model to allow it to accept datasets.

4.2. Sectioning of adult murine heart *ex vivo*

A protocol was devised to consistently section and dissect an adult murine heart into 18 distinct pieces according to anatomical structure. As described in sections 3.1.2 and 3.1.3, mouse hearts were dissected consistently into portions separating atria and ventricles, and left structures from right. The first cut of the dissection process is shown in figure 4.5a. After placing the ventricular section of the heart into the mouse heart slicer, slices were made out of the remaining portion of the heart (figure 4.5b). Each piece represents an anatomically distinct compartment within the adult murine heart, potentially relating to a section that has a specific set of genes that regulate its development and maintenance. These pieces were then used in RNA extraction, as described in section 3.4.

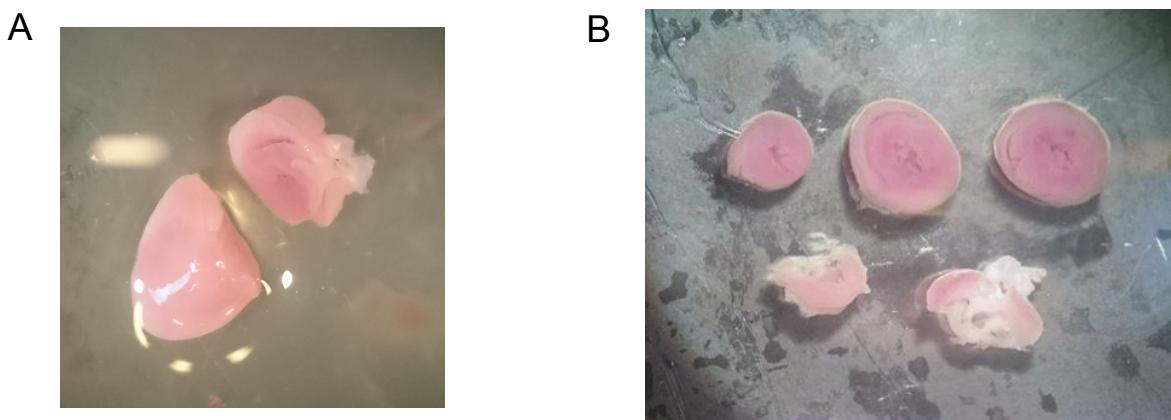


Figure 4.5. a) The first cut into the mouse heart to separate the atrial section from the ventricular section. b) Dissected slices of the heart according to the novel dissection protocol.

4.3. Tissue dissociation

Subsequent to the biological sectioning of the heart, the next step was to acquire a cellularised sample from each of the tissue pieces. In order to do this, various techniques of cellularisation using enzymatic digestion were attempted. These protocols are summarised in table 4.1.

Protocol	Preparation time (minutes)	Incubation time (minutes)	Hands-off?	Viability	
				Countess/EVE	FACS
Tondl				40-50%	0-3% (data not shown)
Galkina				40-60%	N/A
Nilsson				40-50%	90-99%

Table 4.1. Summary of different tissue dissociation protocols used.

Tondl protocol (unpublished, see Appendix D). This protocol is fairly aggressive and uses both mechanical and enzymatic digestion. Overall, this protocol takes approximately 3 hours to complete on 18 pieces of mouse heart tissue. When FACS analysing, cell viability was extremely low (0-3%), indicating that this method of tissue dissociation was not appropriate for this specific study.

Galkina protocol (Galkina et al. 2006). This protocol takes approximately 30 minutes to set up and 1 hour of incubation. Incubation is entirely hands-off. Samples processed using the Pinto protocol were not analysed using FACS, but automatic cell counter readings produced similar values to that of the Tondl protocol, likely indicating similarly low viability.

Nilsson protocol (unpublished). This protocol takes approximately 1 hour to complete. Although the protocol is highly time-sensitive in terms of timing of enzymatic digestion times and hence requires more personnel to complete, the protocol is simple and faster than the other protocols. Additionally, the samples processed using this protocol showed very high viability when

analysed using FACS (figure 4.6). Out of the 10,000 events in the first plot, 29.9% were identified to be single cells. Out of this population, 98.1% of the single cells were PI-, indicating live cells. This protocol was subsequently used.

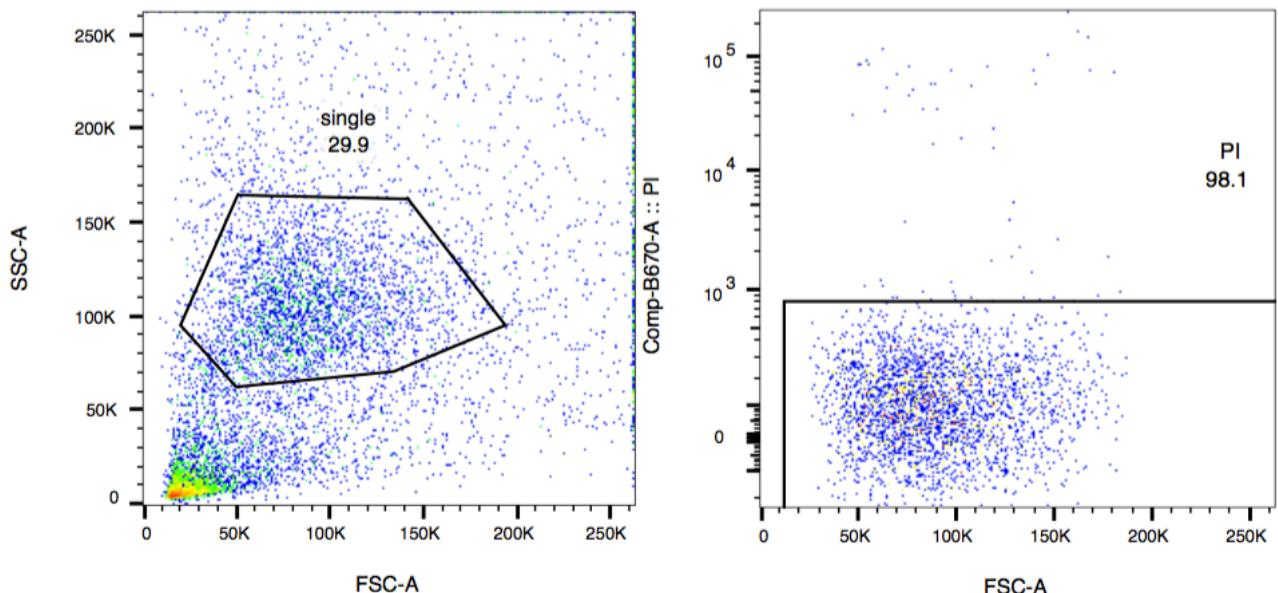


Figure 4.6. FACS plots showing gating for single cells and viability of 98.1%.

However, the samples processed using the Nilsson protocol showed low viability when analysed using an automatic cell counter with values similar or lower to those of the Tondl and Pinto protocols. This inconsistency between analysis techniques may indicate that one is less reliable than the other. Because FACS analysis uses fluorescent staining and is highly sensitive, the results generated from this method of analysis are likely more accurate than those of the automatic cell counter. The automatic cell counter identifies and counts cells based on event shape and size, and it is likely that debris was detected as cells, or true cells were classified as debris.

In order to overcome the problem of low yields of RNA and chromatin from the samples that had undergone enzymatic digestion, a tissue homogeniser was used in an attempt to extract RNA directly from the tissue, bypassing the cellularisation stage.

Tissue homogeniser. The concentration of RNA and chromatin yielded from the enzymatic digestion methods were very low (data not shown). Consequently, the tissue homogeniser was used to ensure a higher yield of RNA as recommended by the RNA purification kit. RNA concentrations from samples processed by the homogeniser were plentiful, as shown in Section 4.4.1.

4.4. Transcriptome of sub-anatomical compartments of the heart

From an adult murine heart, transcriptomic data was gathered from each of the 18 pieces to show gene expression levels sub-anatomical compartment. The RNA purified from the samples (described in Section 3.4) was quantified and quality checked prior to sequencing.

4.4.1. Quantification of RNA

To quantify the concentration of DNA, Qubit assays were done in addition to Quality Control (QC) from the Bioanalyser. It should be noted that prior to running the samples through the Bioanalyser, the samples were diluted from 10 μ l to 35 μ l. The quantification of the RNA was key in confirming that there was sufficient material to send the samples for sequencing. Concentrations were as follows:

Sample name	RNA concentration (ng/ μ l)	
	Qubit	BioAnalyser (after dilution)
A1	103	24.18
A2	41.5	9.12
A3	214	42.23
A4	241	42.12
B1	59	9.18
B2	177	41.29
B3	178	46.17
B4	266	46.54
C1	74.3	18.5
C2	123	22.71
C3	161	41.68
C4	226	40.55
D1	58.2	9.24
D2	148	25.02
D3	105	18.8
E1	64.8	9.86
E2	196	32.53
E3	321	55.26

Table 4.2. Concentrations for each sample.

4.4.2. RNA quality assurance

Prior to sequencing, each sample was run through a LabChip GX. From this, quality assurance information was gathered in the form of RNA Integrity Number (RIN) as well as 260/280 and 260/230 absorbance ratios (Table 4.3). RIN values are used to aid in quantifying the integrity of RNA samples. The purity of the RNA samples was assessed by the 260/280 readings, shown in the third column in Table 4.3. Similarly, nucleic acid purity was assessed by 260/230 readings, shown in the fourth column in Table 4.3.

Sample name	RNA Integrity Number	260/280	260/230
A1	9.1	1.91	1.27
A2	8.8	1.85	0.64
A3	8.4	1.98	1.13
A4	8.6	1.96	1.31
B1	8.5	2.12	0.05
B2	8.6	1.93	1.21
B3	8.7	2.02	1.45
B4	9.1	2.01	1.74
C1	8.2	1.92	1.33
C2	8.3	1.94	1.25
C3	9.3	1.95	1.7
C4	8.9	2.09	0.59
D1	8.8	2.22	0.92
D2	9.1	2.01	1.04
D3	8.5	1.89	0.59
E1	9	2.12	0.78
E2	8.8	1.9	1.26
E3	8.9	1.95	1.04

Table 4.3. Values relating to quality assurance of RNA samples.

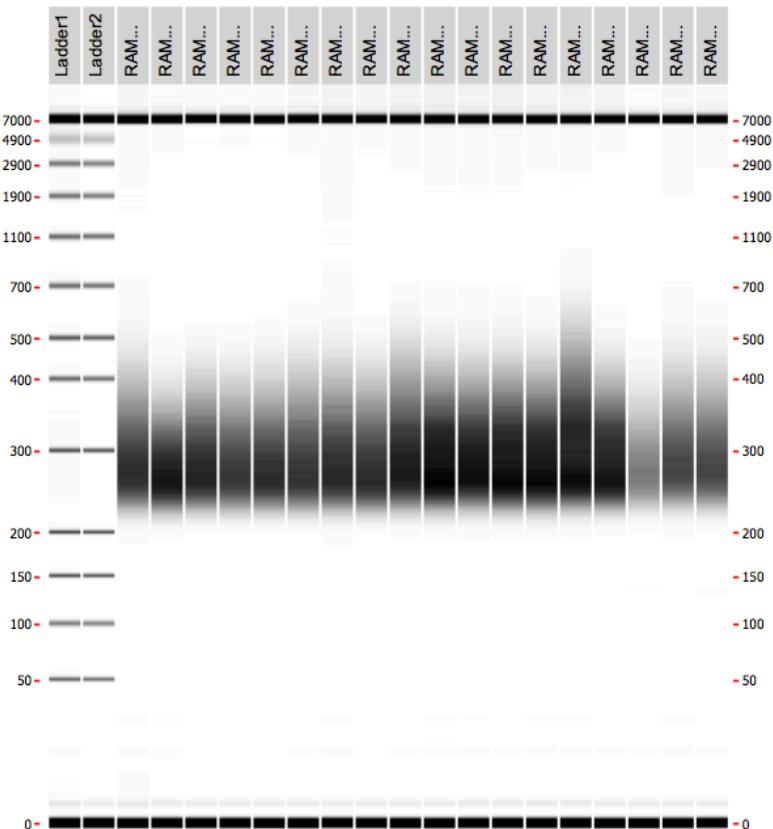


Figure 4.7. Digitised gel from LabChip GX.

From a digitised gel generated from the LabChip GX's output, it can be seen that RNA strands of various sizes are present within each sample, with a peak concentration of strand size at approximately 275bp (figure 4.7).

Given the high RIN values of the samples, the samples were sequenced and reads were obtained for each of the samples.

4.5. Attempt at acquiring chromatin accessibility of sub-anatomical compartments

Chromatin from each of the 18 different pieces was obtained following Section 3.5. A number of quantitative and qualitative tests were run to gauge the quality of the samples, prior to sequencing.

4.5.1. Quality assurance of ATAC-seq samples

To observe the range of components comprising the samples prepared for ATAC-seq, the samples were loaded onto and run on an agarose gel.

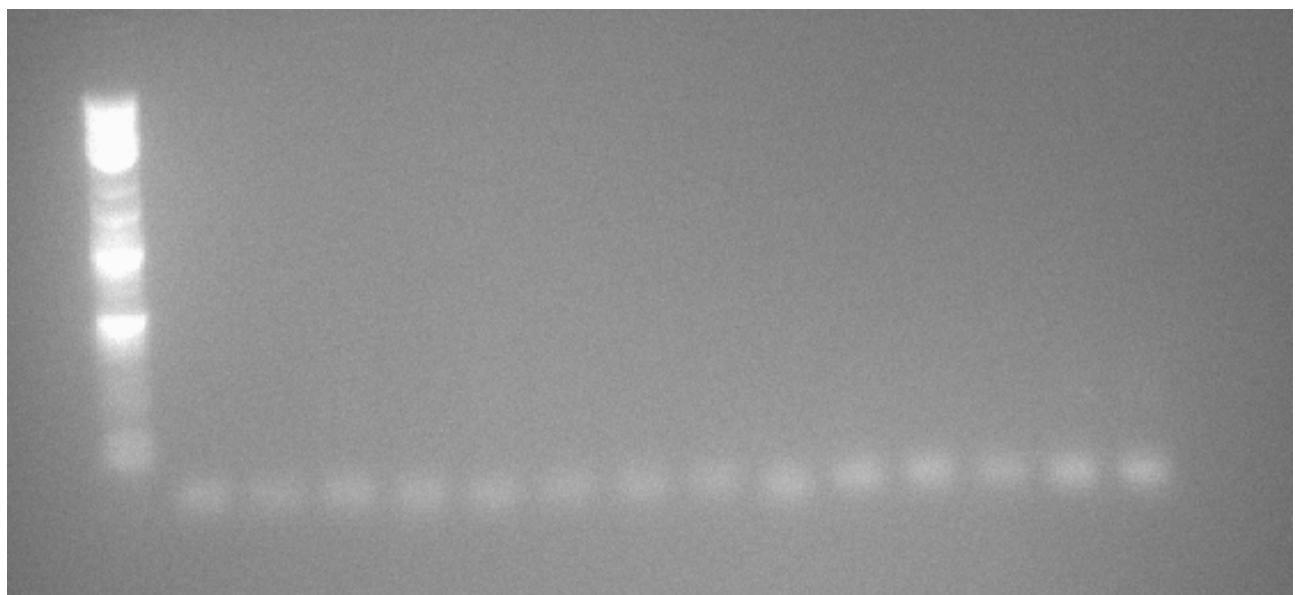


Figure 4.8. Agarose gel to check quality of chromatin in ATAC-seq samples.

The ladder lane ran successfully, showing distinct bands along the gel, demonstrating that the gel had run successfully. For the samples, besides the loading dye at the bottom, no clearly visible bands can be seen on the gel, indicating that the transposase reaction had not taken place.

4.5.2. Quantification of chromatin

DNA concentrations of samples were quantified using a Qubit assay machine.

Concentrations were as follows:

Sample name	DNA concentration (ng/µl)
Elution Buffer only	< 0.005 (non-detectable)
A1	1.05
A3	1.49
B1	2.02
B2	1.70
B3	1.70
C1	1.33
C2	1.34
C3	1.31
D1	1.37
D2	1.53
D3	1.16
E1	1.14
E2	1.57
E3	1.79

Table 4.4. Concentrations for each sample (including negative control of Elution Buffer).

Elution buffer from the QIAGEN MinElute kit was used as a negative control in the Qubit assays to ensure that no readings were positive as a result of proteins in the solution. This buffer was used as a control as it was the buffer used in the experiment in which the final sample was eluted in.

4.5.3. Quality assurance of chromatin

After quantification, 2 μ l of 11 randomly selected samples were analysed using a BioAnalyser.

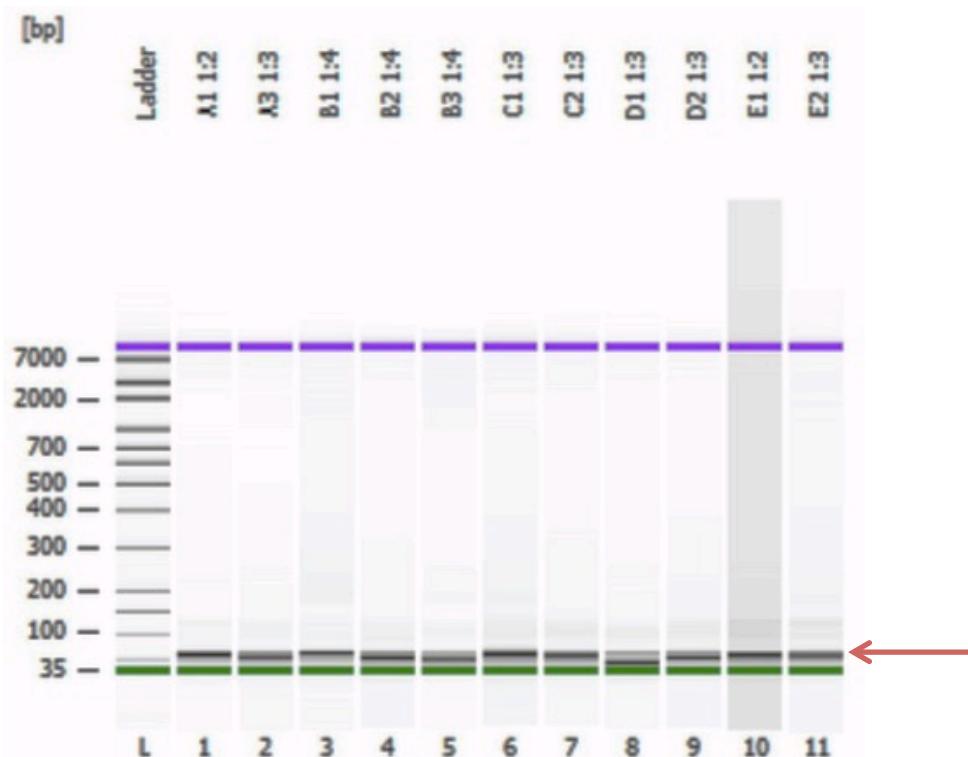


Figure 4.9. BioAnalyser run for ATAC samples, showing a band between 35bp to 100bp.

Very faint bands can be seen on the samples between the 35bp and 100bp marks, as indicated by the red arrow in Figure 4.9. Beside these, the lanes remain fairly clear.

For a successful ATAC-seq experiment, nuclei needs to be isolated from a single-cell suspension. After numerous attempts at cellularisation of tissue samples with none of them yielding sufficient amounts of chromatin, this experiment could not be completed within the timeframe of the study. This is discussed in Section 5.1.6.

4.6. Unsupervised cluster analysis

4.6.1. Hierarchical clustering

Hierarchical clustering of the samples produced a hierarchical cluster plot based on genes and samples. The expression values were normalised to the median intensities of all genes.

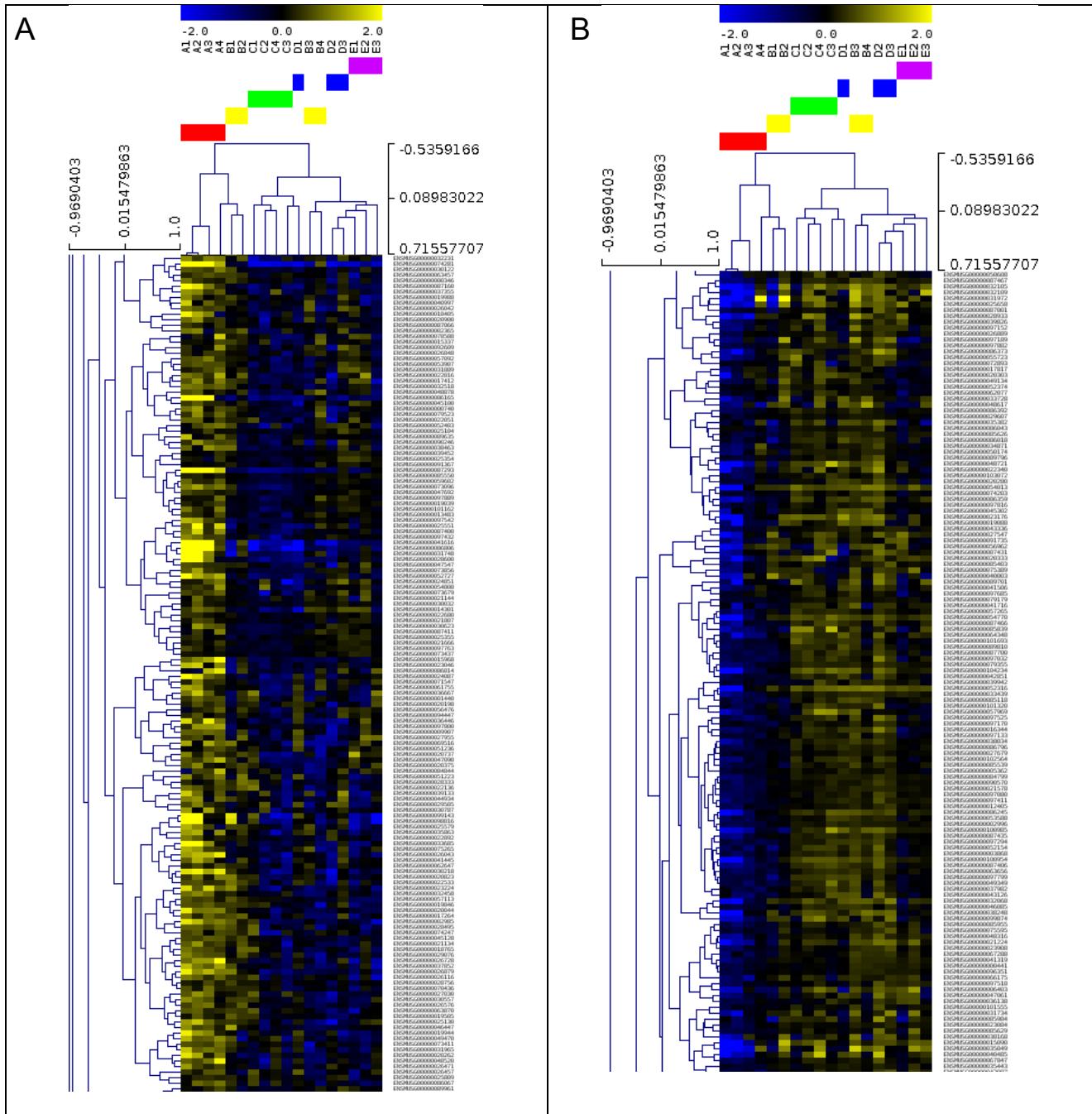


Figure 4.10. Hierarchical clustering of differentially expressed genes. a) Genes highly expressed in the atrial and superior pieces. b) Genes highly expressed in the ventricular pieces.

Figure 4.10 shows sections taken from the full hierarchical cluster tree. These sections exemplify the distinct higher expression and lower expression of genes relative to the median expression of all genes, displayed sample. Figure 4.10a shows genes that are highly expressed in the atrial and superior pieces but expressed less in the ventricular pieces, and figure 4.10b shows the opposite results.

Clustering based on samples indicate that the samples are grouped by heart sections (slice A in red, slice B in yellow, slice C in green, slice D in blue, slice E in purple), indicating similarity within slices.

4.6.2. MDS plots

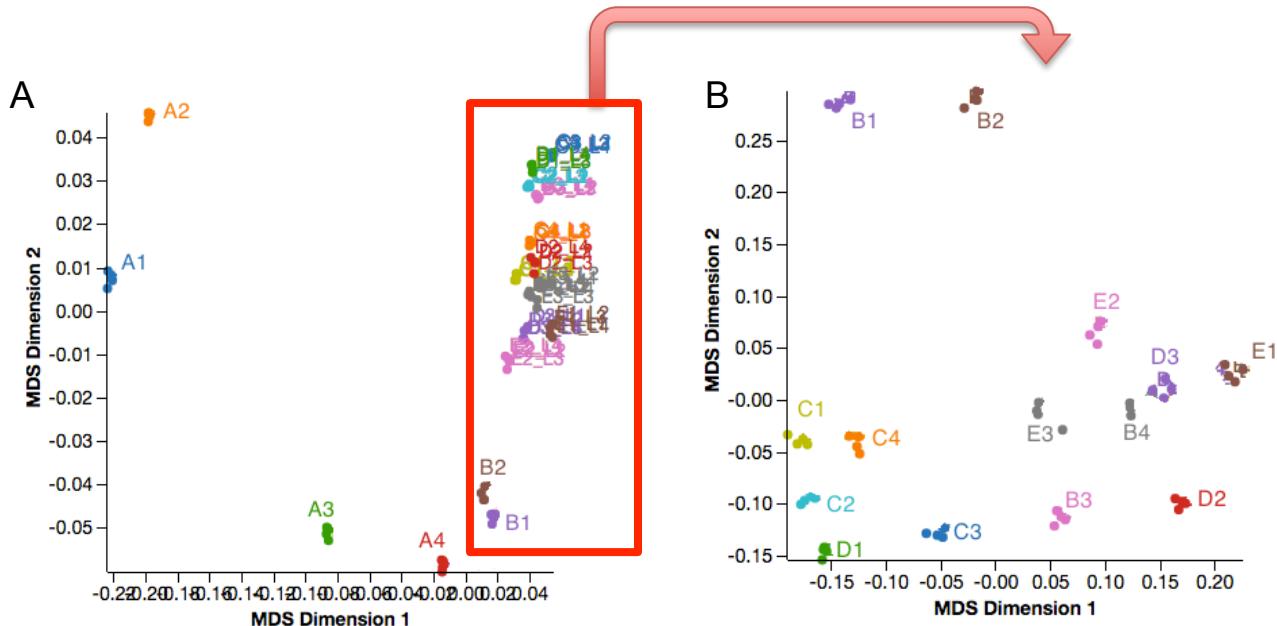


Figure 4.11. MDS plot showing clustering of samples by similarity. a) All pieces. b) MDS plot excluding pieces A1, A2, A3, A4 (atrial pieces).

MDS plots showed strong distinction between atrial and ventricular pieces, as shown in figure 4.11. It can be seen in figure 4.11a that the atrial pieces are distinct from the other pieces by the first dimension.

Further, within each slice, the pieces are distinct in position (figure 4.12).

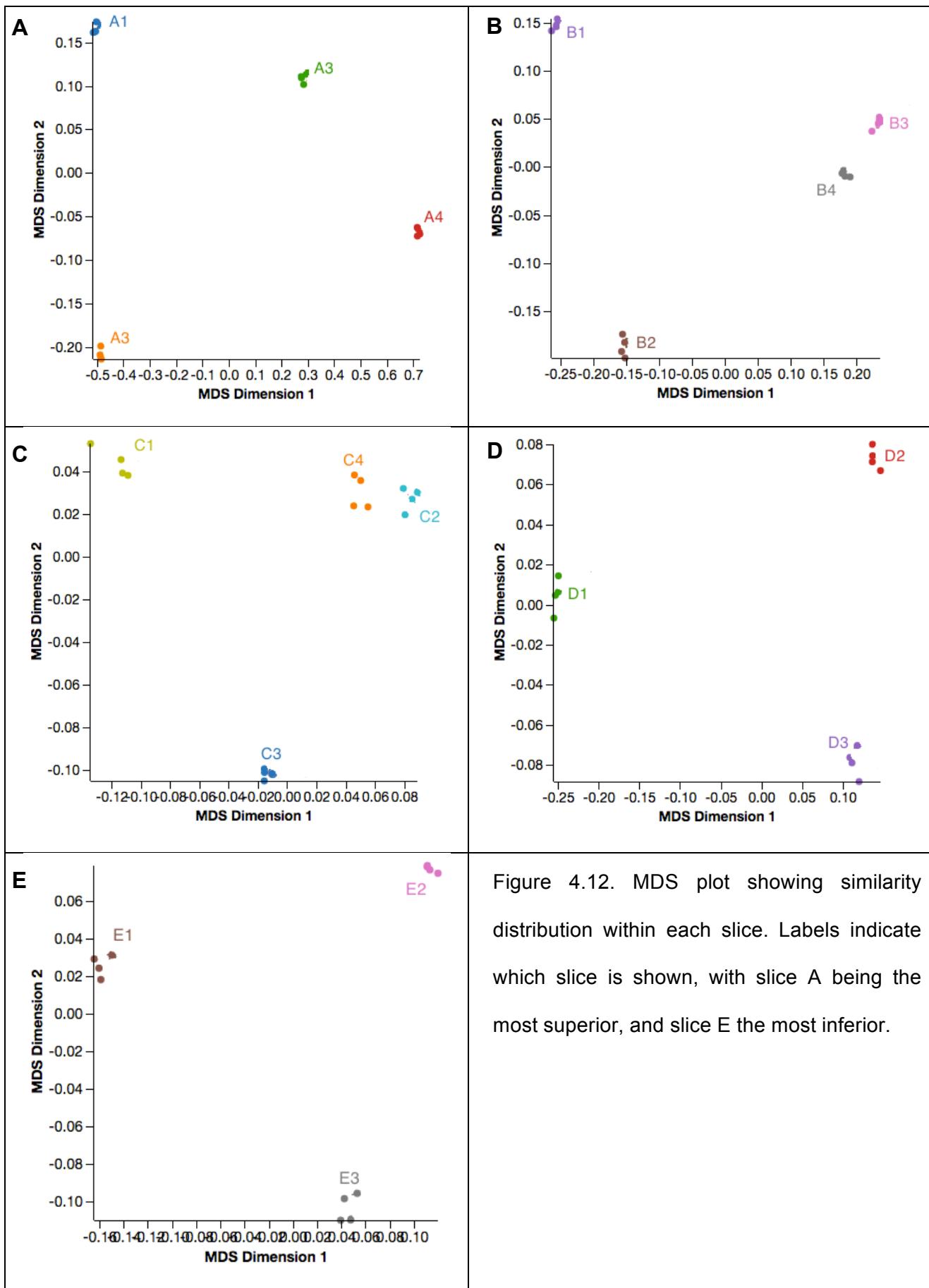
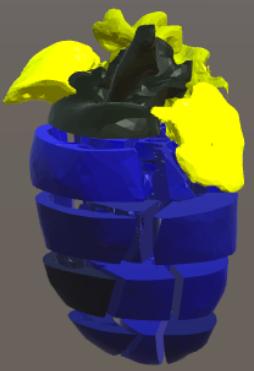
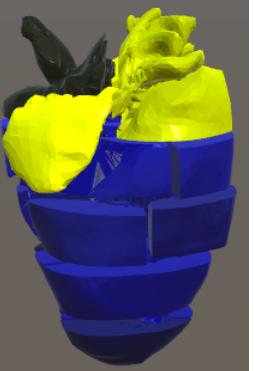
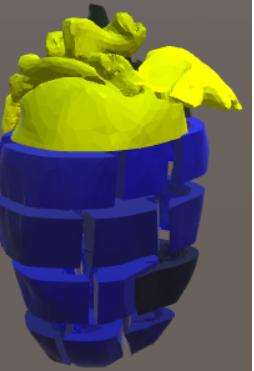
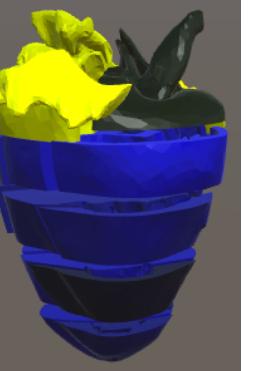
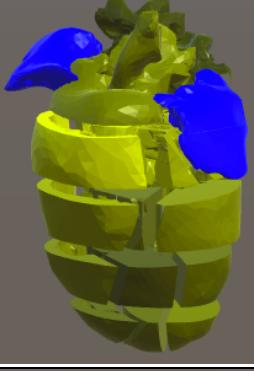
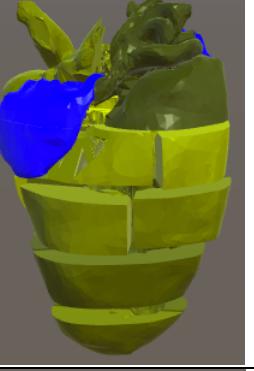
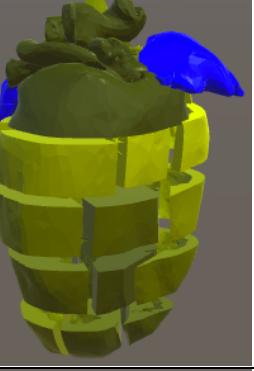
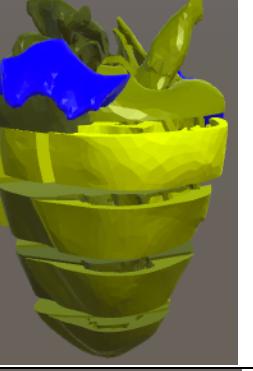
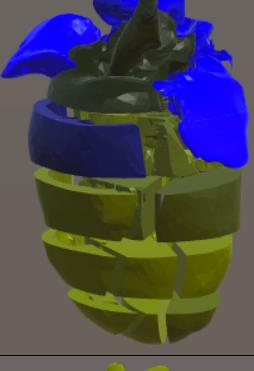
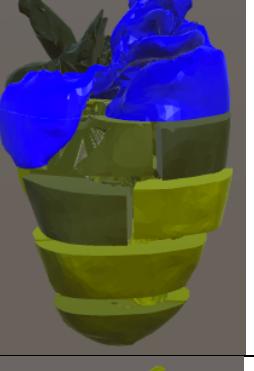
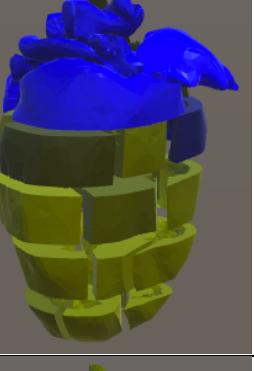
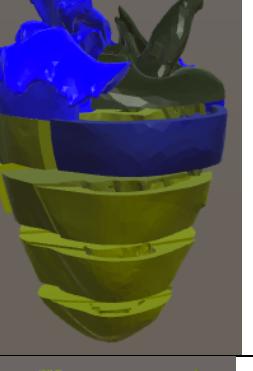
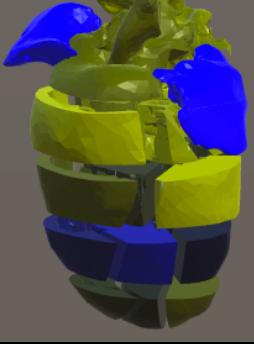
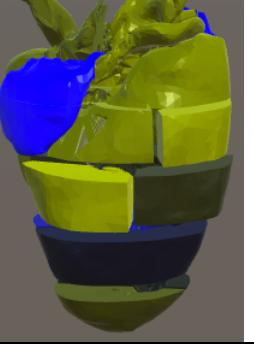
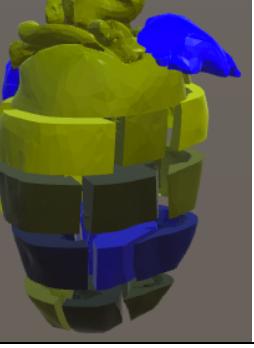
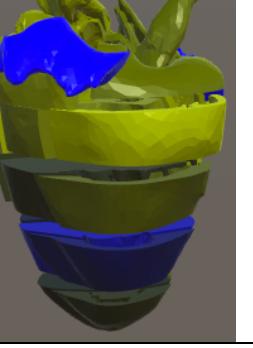


Figure 4.12. MDS plot showing similarity distribution within each slice. Labels indicate which slice is shown, with slice A being the most superior, and slice E the most inferior.

4.7. Integration of RNA-seq datasets with 3D model

As explained in 4.2, the computationally sectioned model of the mouse heart was adapted to the study and displays gene expression patterns of genes specified by user input into the software. The model is able to take files containing read counts and translate the read counts for each gene into a heatmap, visualised on the 3D model.

Standalone software was created which displays the 18 pieces of the sectioned 3D model of the heart and a text field. The text field allows users to enter any gene name as input and search for the expression patterns of that gene. The program, if it finds the gene name, will display the gene expression levels for each piece using the RNA seq datasets generated from the above experiments. The gene expression levels are shown in the form of a two-colour heatmap, that is, yellow toned pieces show where the gene is highly expressed, and blue toned pieces show where the gene is expressed less.

	Ventral view	Lateral left view	Dorsal view	Lateral right view
D630045J12Rik RIKEN				
Myl2				
Ccdc63				
Lrtm1				

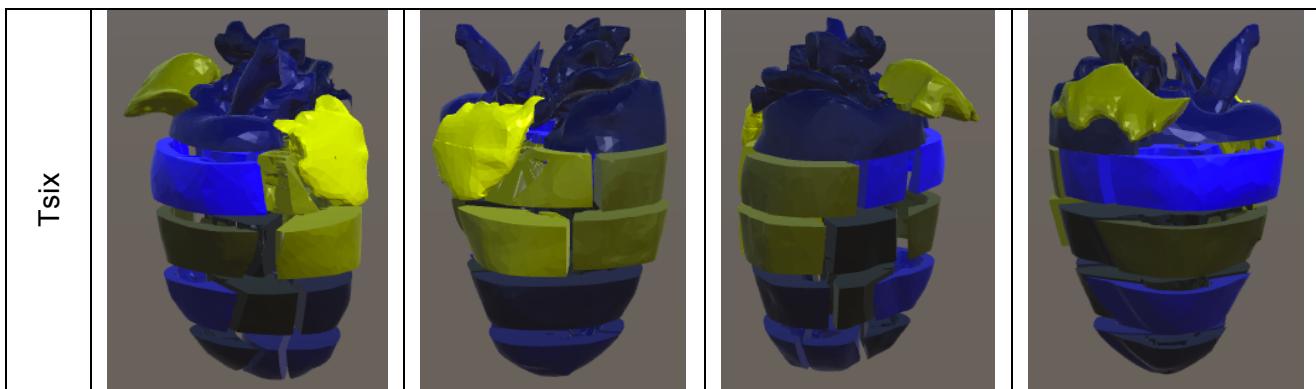


Figure 4.13. The 3D heart model, showing different views of different gene expression patterns of genes across the heart, showing specific expression for respective pieces of the heart.

The model has shown that Myl2, myosin polypeptide 2, is most highly overrepresented in the ventricular area and expressed less in the atrial parts of the heart. This is confirmed by comparing *in situ* hybridisation images from the Mouse Genomic Database (MGI).

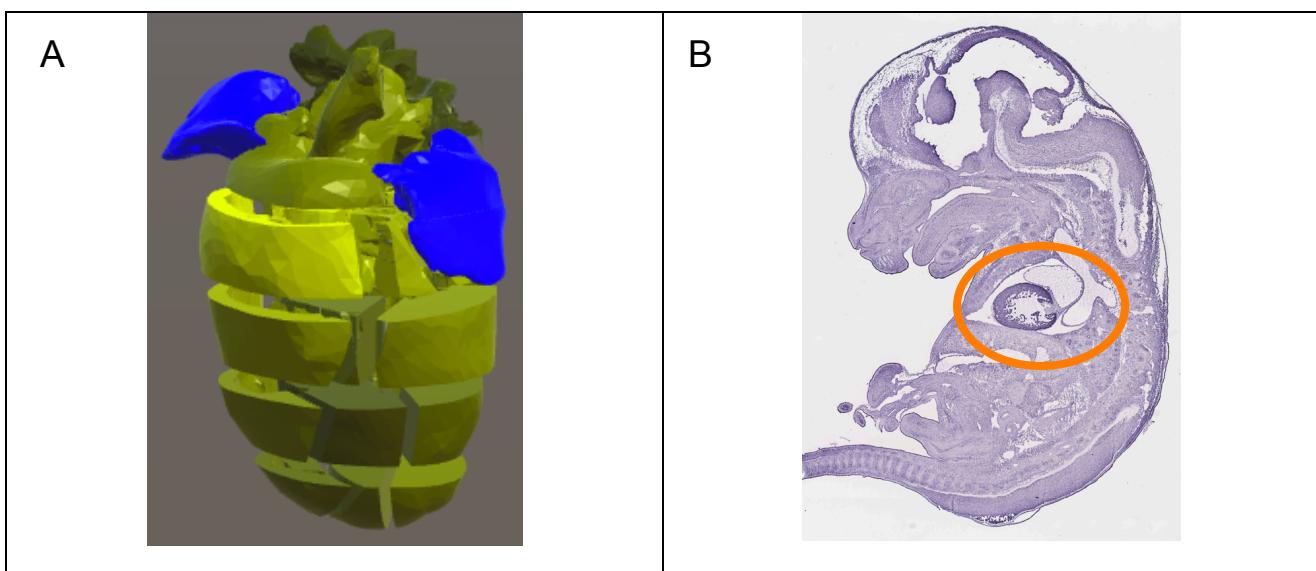


Figure 4.14. a) 3D heart model showing gene expression of Myl2. b) ISH image of gene expression of Myl2 from MGI, showing similarity between gene expression patterns on the 3D model, and real experiments.

5. Discussion

5.1. Technical issues

5.1.1. Computational model of the heart

As with any model, it can be difficult to accurately recapitulate a state that is entirely true to life. The model was generated from histological sections of an adult mouse heart, and so, the model is biologically relevant to the study. However, due to intrinsic variation of organs between animals, the accuracy of the model may not always be high. Technology permitting, photographs could be taken of the biological samples during dissection to aid in digital dissection. Photographs may be taken from perhaps six different angles (top, bottom, four sides at 90°) and scrutinised closely while dissecting the 3D model for a more accurate digital dissection.

Additionally, to achieve the most true-to-life representation of the biological samples, the 3D model was dissected manually. This manual process was time consuming and is estimated to have taken approximately 40 hours. However, this process is necessary and cannot be automated as it requires the ability to discern between any ambiguity within the model.

5.1.2. Mice used for experiments

The strain used were C57BL6/MARP wildtype mice to ensure minimal heart phenotypes. The use of a specific strain of mice ensures that there is fewer genetic variation between replicates for more consistent acquisition of data. However, this strain is highly inbred, meaning that there are very few genetic variations within the strain. It may be more broadly applicable to the general populace if the mice used were truly wildtype and of numerous mixed genetic backgrounds.

Mice used for the experiments were all female. It has been shown that there are significant differences in heart phenotypes between sex in humans (Qin et al. 2015). In the future, it may be interesting to compare differences in gene expression and regulation between the sexes.

Additionally, all mice were adults between 6-10 weeks old. This age was chosen as this is when the animals have reached sexual maturity and are considered adults. It may be possible that this timeframe is wide, and a more specific age should be chosen. For example, some mice may have been 6 weeks old and compared with a mouse of 10 weeks, producing discrepancies in data.

5.1.3. Heart extraction

Mode of mouse culling. The method that is used to cull the mice may affect the results of the experiments. Two of the most common methods for culling mice in a fast and humane way include asphyxiation by carbon dioxide inhalation, and cervical dislocation. It has been shown that asphyxiation can be a time-consuming process which might affect the integrity of the tissue and eventually impair RNA integrity or change the chromatic landmark. Cervical dislocation, on the other hand, results in immediate death of the animal. This allows for more immediate perfusion of the heart, resulting in faster and more efficient clearing of blood cells for a more pure sample of heart tissue only. Moreover, the faster processing of heart tissue would entail fresher organ and tissue isolation, perhaps accounting for the higher cell viability found in the samples

Accuracy of slicing. The slices that are produced from the experiment are relative rather than absolute. That is, the dissection is performed by attempting to separate anatomical structure. Consequently, it may be the case that the pieces are not perfectly reproducible between heart samples.

To achieve more consistently sliced pieces, techniques alternative to the mouse heart slicer could be used. A vibratome allows fresh tissue to be used. Similarly, a cryostat could be used with frozen tissue; however, the freezing of the tissue may affect cell viability. Moreover, ATAC-seq requires fresh cells to be used, that is, cells that have not been frozen, so the cryostat would not be appropriate if ATAC-seq is to be used. It should be noted, however, that these techniques would not allow close scrutiny of slice size by anatomy. Each slice would be of the same size. Additionally, as the project was limited to five slices of the heart, each slice was between 1.5 – 2mm in thickness, outside of the largest setting on both vibratomes and cryostats.

Heart tissue heterogeneity. Additionally, this also begs the question of the biological relevance of using the entire piece of heart that is derived somewhat crudely from the heart. Mammalian hearts consist of three distinct layers of tissue: epicardium, myocardium, and endocardium. Each of these layers has a specific subset of cells that comprise it, which would presumably entail different gene expression patterns between the layers. It follows then that

another way to dissect the heart could have been through manually separating the layers; however, this would prove to be very difficult given the small scale of the heart as well as the challenge of identifying the layers using a dissection microscope.

Another potential method to separate the heart tissue by layer could be to use FACS sorting after dissociating each piece. Cell surface markers specific to each layer of the heart would be useful in separating and collecting samples specific to the layer.

5.1.4. Tissue dissociation

Tondl protocol. This protocol was originally designed for the dissociation of zebrafish embryos, and so, it may be that the protocol needs first to be optimised for mice.

As the time spent in strong enzymes is quite lengthy, it may be possible to reduce the time spent in these enzymes in order to increase cell viability, for example, halving the incubation time and assessing the viability of the cells.

Galkina protocol. The protocol described by Galkina et al specified that DNase 1 was used. However, this was excluded from our revision of the protocol as it was thought that the DNase may potentially degrade the chromatin that was required for the ATAC-seq experiment. Dissociation results may have been improved if DNase was used, however, it is not known whether the enzyme would degrade the chromatin at this stage in the experiment.

5.1.5. RNA-seq

For quality assurance, it is recommended to use RNA samples of RNA Integrity Number (RIN) ≥ 8 (Schroeder et al. 2006). As shown in Table 1, the RINs of all samples were above 8, indicating that the samples were of high enough quality to gain reliable and accurate reads once sequenced. A 260/280 ratio of ~ 2.0 is considered pure for RNA. Only some of the samples were assessed to be above this threshold, but most of the samples were near enough to the 2.0, with the lowest value being 1.85. Additionally, the purity of the RNA samples was assessed by 260/280 absorbance ratios, shown in the third column in Table 1. Nucleic acid purity was assessed by 260/230 absorbance ratios, shown in the fourth column in Table 1. 260/230 ratios should often fall between 2.0 – 2.2. Samples with ratios lower than these values may be due to contaminants. None

of the samples were in this range, indicating potentially large numbers of contaminants in all samples.

5.1.6. Technical replicates

In this study, one biological replicate was produced. However, it should be noted that upon sequencing, the samples were conjugated with barcodes by polymerase chain reaction (PCR) amplification to allow unique identification of each sample. Subsequently, all samples were mixed together and the combination of all 18 samples was run on four different lanes as technical replicates on a sequencing level. The multiplexing of samples across four lanes also ensures that any discrepancies found on a malfunctioning lane would be detected as an outlier.

Next, the datasets generated from the sequencing run comprised of four lanes per sample. At this point, it was possible to either concatenate all of the files together into two files (one for the forward reads, one for the reverse reads), or concatenate two of the lanes together (for example, lanes 1 and 2 and lanes 3 and 4, or lanes 1 and 4, and lanes 2 and 3, etc.), or keep all of the lanes separate for four technical replicates.

5.1.7. ATAC-seq

As seen in the FACS analysis of the samples dissociated using the Nilsson protocol, the majority of the single cell population were live and viable. However, upon chromatin quantification and qualification (Section 4.4.1 and 4.4.2), it appeared that the transposase reaction had failed. This is potentially due to the disruption of the delicate stoichiometry between transposase enzyme and the number of starting cells. It is specified in the protocol (Buenrostro 2015) that the number of cells is vital to the success of the experiment.

In the future, this experiment may be optimised by FACS sorting for viable cells. The current method involved FACS analysis to check for viability, but the cell suspension was used as-is. So, the suspension still contained debris, dead cells, and other miscellaneous artifacts, which would have decreased the concentration of cells in the suspension. Even though the cells were counted manually by microscopy and cytometer, the concentration of cells in solution would have been lower than if first run through a cell sorter. As shown in figure 4.6, less than a third of the

sample was considered to be single cells. Even though the majority of this population was viable as suggested by PI staining, the low cell concentration would have affected the purity of the samples, and hence, the success of the ATAC-seq experiment. FACS sorting the samples would increase the concentration of cells and perhaps aid in an increased yield of RNA and DNA purification.

5.2. Discussion of results

5.2.1. Computational model

The three-dimensional aspect of the model aids greatly in providing an intuitive and true-to-life representation of gene expression patterns in the adult murine heart. Collecting transcriptomic data and displaying it on a 3D model of the organ allows identification of key genes involved in the formation and maintenance of the heart. The model can be used not only as a mode of confirmation of existing knowledge and experiments, but additionally as a tool for discovery of novel genes. With better understanding on the spatial distribution of genes, genes that are likely involved in disease relating to those areas can be identified and tested.

It can be seen from Section 4.6 that the model readily portrays gene expression patterns of any gene of interest in such a way that allows the user to visualise where the gene is being expressed spatially. Pieces of similar anatomical location often had gene expression patterns similar to those adjacent pieces; that is, pieces on the most superior part of the heart had similar gene expression compared to those pieces on the most inferior parts of the heart, and vice versa. This is the first model available to display RNA gene expression in the heart at this resolution.

5.2.2. Transcriptomic data of the heart

Hierarchical clustering showed a strong distinction between atrial and ventricular gene expression (figure 4.10). From the full hierarchical cluster tree, it is frequently seen that genes upregulated in the atria are strongly expressed in the atria and weakly in the ventricles, and vice versa. It would be interesting to investigate each piece for specific clusters of co-expressed genes to perhaps reveal novel analogous genes for that function. Additionally, exploring genes that are

expressed exclusively in specific ventricular compartments may yield interesting results. Identifying genes expressed in specific locations within the heart may provide the basis for future experiments to mutate or knockout those genes, perhaps leading to the discovery of new genes involved in congenital heart defects.

From the MDS plots, while all pieces within the topmost slice, slice A, seemed to frequently group together, it also seemed that two pieces from the next slice down, slice B, followed pieces from slice A quite closely. Pieces B1 and B3 often tended towards pieces from slice A, specifically A1 and A2. A likely cause for this is that pieces B1 and B3 are immediately inferior to pieces A1 and A2. A1 and A2 are the right and left atria respectively, and B1 and B3 were in the area where the atrial compartments persist into the main body of the heart. So, it is likely the B1 and B3 have similar gene expression patterns as A1 and A2 due to this continuity of the atrial sections.

Following on from this, each slice seemed to have separation between the pieces, indicating distinct gene expression patterns according to anatomical location in the heart.

5.2.3. Comparisons between 3D model and ISH

Although there were a number of genes that were overrepresented in the atria but not in the ventricles, a number of ISH images did not reflect the expression patterns expected from the heatmap derived from cluster analysis. Additionally, some ISH images had not yet been created and uploaded to MGI. This may be due to the fact that the ISH data available are from embryonic mice rather than adult mice, creating discrepancy between gene expression patterns. As this study was conducted on adult mice, it may be beneficial to collect more data on adult ISH experiments for comparison with the model.

It then follows that there is a multitude of genes that have not yet had their function defined but are important in heart development and causing disease phenotypes. Candidates for such events include gene x, y, and z, given their distinct expression patterns.

6. Conclusions and future directions

Heart disease is one of the most prevalent causes of death worldwide. Major genes responsible for some conditions have been discovered, such as gene x, gene y, and gene z.

From this study, it has been shown that gene expression patterns are distinct in different spatial locations in the adult mouse heart.

A 3D model of the adult murine heart was adapted to the study through means of computational sectioning using 3D modeling software. Software was specially programmed to use the sectioned model in conjunction with transcriptomic data sets.

A method of slicing and dissecting mouse hearts was devised, allowing for 18 distinct portions of the heart to be isolated and processed in preparation for RNA purification and subsequently RNA sequencing to gather transcriptomic data for each piece.

Using the transcriptomic data, unsupervised cluster analysis showed that atrial and ventricular sections of the heart have distinct gene expression patterns. Further to that, each slice of the heart had distinct gene expression patterns, supporting the hypothesis that finding the expression and regulation patterns of every gene within specific sub-compartments in an organ will reveal the specific subset of genes that play an essential role in that anatomical structure.

Epigenomic information was not gathered within this study, but the opportunity for optimisation and changes in experimental plans has arisen from it. The quantification and quality assurance aided in identifying flaws in the study, and it is hoped that later experiments will push this research so that not only will transcriptomic data be gathered, but epigenomic data too, and all from the one piece of tissue.

Once the transcriptomic data was integrated with the 3D model of the adult murine heart, visualisation of gene expression patterns for each piece within the heart could be easily viewed and scrutinised. Subsequently, the clustering algorithm completed using MeV followed by gene ontology enrichment analysis may aid in finding genes that are potentially homologous in function, or genes that work in conjunction with other genes in the same pathway to regulate a specific sub compartment.

In the future, it may be interesting to extend the study into disease, regeneration, development, and other systems. This study was a proof of concept to show the full pipeline from start to finish using murine hearts, but it is entirely possible to adapt the study for other organs and organisms.

The heart in various states of injury or disease phenotypes would be interesting to investigate. Myocardial infarct and septal defects would be of particular interest.

In terms of regeneration, as in the study conducted by Porrello et al, the transcriptomic and epigenomic state of sub-anatomical pieces within the heart would be interesting to investigate during stages of cardiac regeneration in the neonate. Novel genes relating to heart regeneration may be identified.

Paramount to the correct development of organisms is perfectly orchestrated spatial and temporal gene expression. Our project details a spatially specific transcriptomic map of the murine heart at the adult stage; however, the study could be extended to include key stages in heart development as well. Developmental stages of interest may be E7.5, E8, E9.5, E12, and E14.5 (Rosenthal & Harvey 2010). Analysing datasets from these developmental stages would not only confirm known key developmental genes, but also reveal key genes that are central to heart development within stages as well as between.

Other organs that display interesting morphology, development, and physiology, such as the brain and lungs, may be used in a similar study to identify genes and epigenomic states key to those organs.

A number of 3D models exist for various organisms, including mice, zebrafish, and even humans embryos. The generation of a library of developmental stages and disease states in a variety of organisms would prove to be a wealth of information for researchers in a wide variety of fields.

7. Appendices

Appendix A: Phosphate Buffered Solution recipe

For 1L of 10x Phosphate Buffered Solution:

Ingredient	Amount
Potassium chloride (KCl)	2g
Sodium chloride (NaCl)	80g
Sodium phosphate dibasic ($\text{Na}_2\text{HPO}_4 \cdot 12 \text{ H}_2\text{O}$)	36.33g
Potassium phosphate monobasic (KH_2PO_4)	2.4g
MQ-water	To 1L

Appendix B: RNA-seq pipeline

```
#!/bin/sh

# FastQC I
cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASEq/RAM815_20150903-
25109093"

for sample in $( ls )
do
    # Open the sample folder (this should run 18 times)
    echo $sample
    cd $sample

    # Open the files in the folder (8 times)
    for file in $( ls )
    do
        if [[ $file == *"fastq.gz" ]]
        then
            echo "Running FastQC on $file:"
            fastqc -f fastq $file
        fi
    done

    cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASEq/RAM815_20150903-
25109093"
done


# trim_galore
cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASEq/RAM815_20150903-
25109093"

for sample in $( ls )
do
    # Open the sample folder (this should run 18 times)
    echo $sample
    cd $sample

    # Open the files in the folder (8 times)
    for file in $( ls )
    do
        if [[ $file == *"fastq.gz" ]]
        then
            #echo "This is the fastq.gz file: $file"

            if [[ $file == *"R1"* ]]
            then
                file1=$file
                #echo "File 1: $file1"
            fi
        fi
    done
done
```

```

        if [[ $file == *"R2"* ]]
            then
                file2=$file
                #echo "File 2: $file2"
            fi

            lane1=$(echo $file1 | cut -c11-11)
            lane2=$(echo $file2 | cut -c11-11)

            if [[ $lane1 == $lane2 ]]
                then
                    echo "Running trim_galore on $file1 and $file2:"
                    trim_galore -q 28 --paired --gzip --phred33 --
stringency 3 $file1 $file2
                fi
            fi
        done

        cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASeq/RAM815_20150903-
25109093"
done

# FastQC II
cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASeq/RAM815_20150903-
25109093"

for sample in $( ls )
do
    # Open the sample folder (this should run 18 times)
    echo $sample
    cd $sample

    # Open the files in the folder (8 times)
    for file in $( ls )
    do
        if [[ $file == *"fq.gz" ]]
        then
            echo "Running FastQC II on $file:"
            fastqc -f fastq $file
        fi
    done

    cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASeq/RAM815_20150903-
25109093"
done

# STAR
cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASeq/RAM815_20150903-
25109093"

```

```

for sample in $( ls )
do
    # Open the sample folder (this should run 18 times)
    echo $sample
    cd $sample

    # Open the files in the folder (8 times)
    for file in $( ls )
    do
        if [[ $file == *"fq.gz" ]]
        then
            #echo "This is the fastq.gz file: $file"

            if [[ $file == *"val_1.fq.gz" ]]
            then
                file1=$file
                #echo "File 1: $file1"
            fi

            if [[ $file == *"val_2.fq.gz" ]]
            then
                file2=$file
                #echo "File 2: $file2"
            fi

            fileName=$(echo $file | cut -c1-2)
            lane1=$(echo $file1 | cut -c11-11)
            lane2=$(echo $file2 | cut -c11-11)

            if [[ $lane1 == $lane2 ]]
            then
                echo "Running STAR on $file1 and $file2:"
                STAR --runThreadN 8 --outSAMattributes All --
genomeLoad NoSharedMemory --readFilesCommand zcat --genomeDir
/persistent/reference/mouse_mm10/star/ --readFilesIn $file1 $file2 --
outFileNamePrefix "$fileName"_L$lane1"_mapped_
            fi
        fi
    done

    cd "/home/nmtan5/Desktop/analysis/3DHeartRNASeq/RAM815_20150903-
25109093"
done

# samtools
cd "/home/nmtan5/Desktop/analysis/3DHeartRNASeq/RAM815_20150903-
25109093"

for sample in $( ls )
do

```

```

# Open the sample folder (this should run 18 times)
echo $sample
cd $sample

# Open the files in the folder (8 times)
for file in $( ls )
do
    if [[ $file == *"mapped"*.sam" ]]
    then
        echo "Running samtools view on $file:"
        newFileName=$(echo $file | cut -c1-24)
        samtools view -bSo $newFileName.bam $file
    fi
done

cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASEq/RAM815_20150903-
25109093"
done

# htseq_counts
cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASEq/RAM815_20150903-
25109093"

for sample in $( ls )
do
    # Open the sample folder (this should run 18 times)
    echo $sample
    cd $sample

    # Open the files in the folder (8 times)
    for file in $( ls )
    do
        if [[ $file == *"mapped"*.bam" ]]
        then
            echo "Running htseq-count on $file:"
            newFileName=$(echo $file | cut -c1-5)
            #echo "$newFileName" "_htseq_out.txt"

            htseq-count          -f           bam           $file
/home/nmtan5/reference/mouse_mm10/gtf/Mus_musculus_GRCm38_78.gtf      >
"$newFileName" "_htseq_out.txt"
        fi
    done

    cd      "/home/nmtan5/Desktop/analysis/3DHeartRNASEq/RAM815_20150903-
25109093"
done

```

Appendix C: C# code for mapping transcriptomic data onto 3D model

```
using UnityEngine;
using UnityEngine.UI;
using System;
using System.Collections;
using System.IO;

public class colour : MonoBehaviour {

    string myLine;
    Hashtable ht = new Hashtable ();

    void OnGUI () {

    }

    // Use this for initialisation
    public void Start(){

    }

    // Update is called once per frame
    void Update () {

        // Sanity check - Press R for Red, G for Green
        /*
        if(Input.GetKeyDown(KeyCode.R)){
            gameObject.GetComponent<Renderer>().material.color = Color.red;
        }
        if(Input.GetKeyDown(KeyCode.G)){
            gameObject.GetComponent<Renderer>().material.color = Color.green;
        }*/
    }

    private void SubmitName(string arg0){
        Debug.Log(arg0);
    }

    //public void ColourFromText(String input){
    public void ColourFromText(){
        var input = GameObject.Find ("InputFieldNew").GetComponent<InputField> ();
        var se = new InputField.SubmitEvent ();
        se.AddListener (SubmitName);
        input.onEndEdit = se;

        // Open the text file containing the dataset for that piece
        StreamReader reader = new StreamReader (@"/Users/nmt/Documents/U
```

```

nity/TestColour/Assets/testDataA1.txt");
    myLine = "";
    char delim = '\t';
    string[] cols;
    string[,] vals = new string[40000, 19];

    int rowCount = 0;

    // Load dataset into a hashtable
    while (!reader.EndOfStream) {
        myLine = reader.ReadLine();
        cols = myLine.Split(delim);

        // Populate the gene ID column
        vals[rowCount,0] = cols[0];
        Debug.Log(vals[rowCount,0]);

        // Populate the array with values
        for (int i = 1; i < cols.Length-1; i++){
            vals[rowCount,i] = cols[i];
            Debug.Log(vals[rowCount,i]);
        }
        rowCount++;
        //Debug.Log(myLine);
        //Debug.Log(cols[0]);
        //Debug.Log(cols[1]);
        //ht.Add(cols [0], cols [1]);
    }

    //if (ht.ContainsKey("gene1")) {
    // Get the row number of the gene
    //if (ht.ContainsKey("gene1") > 0) {

        string testStr = "0.421 + "F";
        //float testFloat = float.Parse (testStr);

        // Parse the string array into floats
        GameObject.Find("A_1").GetComponent<Renderer>().material.color =
new Color(1.0F, 1.0F, 0.0F);
        GameObject.Find("A_2").GetComponent<Renderer>().material.color =
new Color(1.0F, 1.0F, 0.0F);
        GameObject.Find("A_3").GetComponent<Renderer>().material.color =
new Color(0.7058824F, 0.7254902F, 0.0F);
        GameObject.Find("A_4").GetComponent<Renderer>().material.color =
new Color(0.05882353F, 0.0627451F, 0.0F);

        GameObject.Find("B_1").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.5019608F);
        GameObject.Find("B_2").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.47843137F);
        GameObject.Find("B_3").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.36078432F);
    }
}

```

```

        GameObject.Find("B_4").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.38431373F);

        GameObject.Find("C_1").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.31764707F);
        GameObject.Find("C_2").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.25490198F);
        GameObject.Find("C_3").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.3372549F);
        GameObject.Find("C_4").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.27450982F);

        GameObject.Find("D_1").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.05490196F);
        GameObject.Find("D_2").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.38039216F);
        GameObject.Find("D_3").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.3882353F);

        GameObject.Find("E_1").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.23137255F);
        GameObject.Find("E_2").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.36862746F);
        GameObject.Find("E_3").GetComponent<Renderer>().material.color =
new Color(0.0F, 0.0F, 0.20784314F);

    }

// Resets all heart pieces to white
public void resetColour(){
    GameObject.Find("A_1").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("A_2").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("A_3").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("A_4").GetComponent<Renderer>().material.color =
Color.white;

    GameObject.Find("B_1").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("B_2").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("B_3").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("B_4").GetComponent<Renderer>().material.color =
Color.white;

    GameObject.Find("C_1").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("C_2").GetComponent<Renderer>().material.color =
Color.white;

```

```
    GameObject.Find("C_3").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("C_4").GetComponent<Renderer>().material.color =
Color.white;

    GameObject.Find("D_1").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("D_2").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("D_3").GetComponent<Renderer>().material.color =
Color.white;

    GameObject.Find("E_1").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("E_2").GetComponent<Renderer>().material.color =
Color.white;
    GameObject.Find("E_3").GetComponent<Renderer>().material.color =
Color.white;
}
}
```

Appendix D: Tondl protocol for tissue dissociation

1. Pre-chill large and small centrifuge to 4deg
2. Euthanise fish with 0.4% tricaine (kept in 4°C fridge)
3. Collect hearts in PBS on ice
4. Rinse hearts twice with ice-cold HBSS
5. Fix hearts for 7min in 1.8% PFA/HBSS on RT
6. Quench with 0.125M glycine (f.c.) on ice for 15min
7. Wash with ice-cold HBSS
8. Discard supernatant and add 500µl collagenase
(100mg/ml use 1:100 in 0.1M TrisHCl pH 7.5)
9. Incubate shaking at 200rpm at 37deg for 30 min, pipette up and down in between using 1000µl pipette with cut-off tip
10. Check at microscope if embryos are almost single cells
11. Centrifuge at 5600RPM in table top centrifuge at 4°C for 5min
12. Discard supernatant
13. Resuspend pellet in 500µL 0.25% Trypsin/EDTA
14. Incubate shaking at 200rpm at 37deg for 30min, pipette up and down using 200µl pipette
15. Terminate by adding 200µl FCS (10% final concentration) and 1.3mL HBSS
16. Incubate 2min on ice
17. Equilibrate 100um (yellow) mesh with 3mL HBSS over clean 50ml Falcon tube
18. Run sample through mesh, rinse tube with 3mL HBSS
19. Centrifuge 4600RPM 4°C 5min
20. Discard supernatant
21. Resuspend pellet in 3mL HBSS
22. Equilibrate 40µm mesh (blue) with 3mL HBSS over clean 50ml Falcon tube
23. Run sample through mesh, wash and rinse tube with HBSS
24. Centrifuge 4600RPM 4°C 5min
25. Discard supernatant
26. Resuspend pellet in 1mL HBSS

27. Take 10µl aliquot for cell counting
28. Mix 10µl aliquot with 10µl Trypan Blue Stain, load 10µl on Countess slide
29. Centrifuge 4600RPM 4°C 5min, discard supernatant
30. Snap freeze pellet with liquid nitrogen and store at -80°C

8. References

- Achim K et al. 2015. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. . (April).
- de Bakker BS et al. 2012. Towards a 3-dimensional atlas of the developing human embryo: The Amsterdam experience. *Reproductive Toxicology*. 34(2), pp.225–236.
- Bannister AJ & Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell research*. 21(3), pp.381–395. Available at: <http://dx.doi.org/10.1038/cr.2011.22>.
- Barski A et al. 2007. High-Resolution Profiling of Histone Methylation in the Human Genome. *Cell*. 129(4), pp.823–837.
- Bartel DP. 2009. MicroRNA Target Recognition and Regulatory Functions. *Cell*. 136(2), pp.215–233.
- Blow MJ et al. 2011. ChIP-seq Identification of Weakly Conserved Heart Enhancers. . 42(9), pp.806–810.
- Blue GM et al. 2012. Congenital heart disease: current knowledge about causes and inheritance. *The Medical Journal of Australia*. 197(August), pp.155–159.
- de Boer B a. et al. 2012. Growth of the developing mouse heart: An interactive qualitative and quantitative 3D atlas. *Developmental Biology*. 368(2), pp.203–213. Available at: <http://dx.doi.org/10.1016/j.ydbio.2012.05.001>.
- Boundless Biology. 2014. Transcriptional Enhancers and Repressors. *Boundless*. Available at: <https://www.boundless.com/biology/textbooks/boundless-biology-textbook/gene-expression-16/eukaryotic-transcription-gene-regulation-113/transcriptional-enhancers-and-repressors-458-11684/> [Accessed April 16, 2015].
- Bruneau BG. 2008. The developmental genetics of congenital heart disease. *Nature*. 451, pp.943–948.
- Buenrostro JD et al. 2015. Current Protocols in Molecular Biology. . (January), p.23. Available at: <http://10e8b09b-2029-4a89-baf0-b90acb9a6c30/Paper/p3768\nhttp://10e8b09b-2029-4a89-baf0-b90acb9a6c30/Paper/p3810>.
- Buenrostro JD et al. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*. 10(12), pp.1213–8. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3830333/>
- Combs P a. & Eisen MB. 2013. Sequencing mRNA from Cryo-Sliced Drosophila Embryos to Determine Genome-Wide Spatial Patterns of Gene Expression. *PLoS ONE*. 8(8), pp.2–8.
- Copp a. 1995. Death before birth: Clues from gene knockouts and mutations. *Trends in Genetics*. 11(3), pp.87–93.
- Costa MW et al. 2013. Functional characterization of a novel mutation in NKX2-5 associated with congenital heart disease and adult-onset cardiomyopathy. *Circulation: Cardiovascular Genetics*. 6(3), pp.238–247.
- Davidson EH. 2010. Emerging properties of animal gene regulatory networks. *Nature*. 468(7326), pp.911–920. Available at: <http://dx.doi.org/10.1038/nature09645>.
- Ecker JR et al. 2012. Genomics: ENCODE explained. *Nature*. 489(7414), pp.52–55.
- Eisenberg E & Levanon EY. 2013. Human housekeeping genes, revisited. *Trends in Genetics*. 29(10), pp.569–574. Available at: <http://dx.doi.org/10.1016/j.tig.2013.05.010>.
- Fabian MR, Sonenberg N & Filipowicz W. 2010. Regulation of mRNA translation and stability by microRNAs. *Annual review of biochemistry*. 79, pp.351–379.

- FEI Visualization Sciences Group. 2012. Amira 3D Software for Life Sciences. Available at: <http://www.fei.com/software/amira-3d-for-life-sciences/>.
- Friedel RH et al. 2011. Generating conditional knockout mice. *Methods in molecular biology* (Clifton, N.J.). 693(1), pp.205–231.
- Galkina E et al. 2006. Lymphocyte recruitment into the aortic wall before and during development of atherosclerosis is partially L-selectin dependent. *The Journal of experimental medicine*. 203(5), pp.1273–1282.
- Gunter C & Dhand R. 2002. Human biology by proxy. *Nature*. 420(6915), pp.509–509.
- Han HUA et al. 2015. GATA4 transgenic mice as an in vivo model of congenital heart disease. . 2007(9).
- Jaskowiak PA, Campello RJ & Costa IG. 2014. On the selection of appropriate distances for gene expression data clustering. *BMC Bioinformatics*. 15(Suppl 2), p.S2. Available at: <http://www.biomedcentral.com/1471-2105/15/S2/S2>.
- Junker JP et al. 2014. Genome-wide RNA Tomography in the Zebrafish Embryo.
- Klattenhoff CA et al. 2013. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*. 152(3), pp.570–83. Available at: <http://www.ncbi.nlm.nih.gov/article/fcgi?artid=3563769&tool=pmcentrez&rendertype=abstract>.
- Li C et al. 2014. Novel and Functional DNA Sequence Variants within the GATA6 Gene Promoter in Ventricular Septal Defects. *International Journal of Molecular Sciences*. 15(7), pp.12677–12687. Available at: <http://www.mdpi.com/1422-0067/15/7/12677>.
- Van Der Linde D et al. 2011. Birth prevalence of congenital heart disease worldwide: A systematic review and meta-analysis. *Journal of the American College of Cardiology*. 58(21), pp.2241–2247. Available at: <http://dx.doi.org/10.1016/j.jacc.2011.08.025>.
- Martin RM & Cardoso MC. 2010. Chromatin condensation modulates access and binding of nuclear proteins. *The FASEB journal: official publication of the Federation of American Societies for Experimental Biology*. 24(4), pp.1066–1072.
- McMenamin PG et al. 2014. The production of anatomical teaching resources using three-dimensional (3D) printing technology. *Anatomical Sciences Education*. 00.
- Meehan R et al. 1992. Transcriptional repression by methylation of CpG. *Journal of cell science. Supplement*. 16, pp.9–14.
- Merrill VK et al. 1989. A genetic and developmental analysis of mutations in labial, a gene necessary for proper head formation in *Drosophila melanogaster*. *Developmental biology*. 135(2), pp.376–391.
- Pennacchio L a et al. 2013. Enhancers: five essential questions. *Nature reviews. Genetics*. 14(4), pp.288–95. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23503198>.
- Porrello ER et al. 2011. Transient Regenerative Potential of the Neonatal Mouse Heart. *Science*. 331(6020), pp.1078–1080.
- Qin J et al. 2015. Sexually dimorphic phenotype of arteriolar responsiveness to shear stress in soluble epoxide hydrolase-knockout mice. *American Journal of Physiology - Heart and Circulatory Physiology*, p.ajpheart.00568.2015. Available at: <http://ajpheart.physiology.org/lookup/doi/10.1152/ajpheart.00568.2015>.
- Ren K et al. 2014. KCTD10 Is Involved in the Cardiovascular System and Notch Signaling during Early Embryonic Development. *PLoS ONE*. 9(11), p.e112275. Available at: <http://dx.plos.org/10.1371/journal.pone.0112275>.

- Richards A a & Garg V. 2010. Genetics of congenital heart disease. *Current cardiology reviews*. 6, pp.91–97.
- Rosenthal N & Harvey RP. 2010. *Heart Development and Regeneration, Volume 1*.
- Schroeder A et al. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology*. 7(1), p.3. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/1413964/>&tool=pmcentrez&rendertype=abstract.
- Smith ZD & Meissner A. 2013. DNA methylation: roles in mammalian development. *Nature reviews. Genetics*. 14(3), pp.204–20. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/23400093/>.
- Soufan AT et al. 2003. Three-dimensional reconstruction of gene expression patterns during cardiac development. *Physiological genomics*. 13(3), pp.187–195.
- Terami H et al. 2004. Wnt11 facilitates embryonic stem cell differentiation to Nkx2.5-positive cardiomyocytes. *Biochemical and Biophysical Research Communications*. 325(3), pp.968–975.
- Thorsen T et al. 2001. Dynamic pattern formation in a vesicle-generating microfluidic device. *Physical Review Letters*. 86(18), pp.4163–4166.
- Tingare A, Thienpont B & Roderick HL. 2013. Epigenetics in the heart: the role of histone modifications in cardiac remodelling. *Biochemical Society transactions*. 41(3), pp.789–96. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/2369793/>.
- Vecoli C et al. 2014. Congenital Heart Disease : The Crossroads of Genetics , Epigenetics and Environment. , pp.390–399.
- Vidal M & Cusick M. 2011. Interactome Networks and Human Disease. *Cell*. 144(6), pp.986–998. Available at: <http://www.sciencedirect.com/science/article/pii/S0092867411001309>.
- Vincent SD & Buckingham ME. 2010. How to make a heart. The origin and regulation of cardiac progenitor cells. *Current Topics in Developmental Biology*. 90(10), pp.1–41.
- Waardenberg AJ et al. 2014. Development. , pp.1–24.
- Wang Z, Gerstein M & Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*. 10(1), pp.57–63.
- Woodruff T. 1998. Cellular localization of mRNA and protein: in situ hybridization histochemistry and in situ ligand binding. *Methods Cell Biol.*. 57, pp.333–351. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/9648114/>.
- Wu AR et al. 2014. Quantitative assessment of single-cell RNA-sequencing methods. *Nature methods*. 11(1), pp.41–46.
- Zhou VW, Goren A & Bernstein BE. 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics*. 12(1), pp.7–18. Available at: <http://dx.doi.org/10.1038/nrg2905>.