Investigating a novel catecholamine-associated cardiomyocyte subpopulation

Alex Lipov

Department of Pharmacology

Linacre College

University of Oxford

A thesis presented for the degree of

Master of Science by Research

Trinity Term 2021

Abstract

Excessive sympathetic stimulation of the heart, via catecholamines, is involved in a range of cardiac disease, from takotsubo cardiomyopathy to heart failure. Canonically, there are two main sources of catecholamines for sympathoexcitation. 1) Adrenaline, secreted by chromaffin cells in the adrenal medulla and carried by the bloodstream to the heart. 2) Noradrenaline, released by sympathetic nerve terminals within the heart. However, there is a growing body of literature highlighting a non-neuronal cellular source of catecholamines within the heart. This project aimed to characterise a cellular cardiac subpopulation associated with the gene expression of a catecholamine-synthesising enzyme. Mouse embryos and hearts from embryonic day 8.5 to neonates were single-cell RNA-sequenced. RNA velocity, a form of *in silico* lineage inference, in combination with differential gene expression analysis and pseudotime inference suggested first heart field splanchnic mesoderm as the developmental origin for this subpopulation. An optogenetic lineage tracing mouse line for the gene encoding this catecholamine-synthesising enzyme was generated. Langendorff-perfused *ex-vivo* hearts from this mouse line were subjected to pulsed light stimulation. In addition, fluorescent microscopy of immunostained heart sections and genetic fluorescence was performed. These methods showed that the subpopulation was spatially distributed in the ventricular endocardium, subendocardium, atria and interventricular septum. The refractory periods of the cells were not significantly heterogeneous across the heart. The cells of the subpopulation had the size, morphology, striations and protein expression of cardiomyocytes. The subpopulation was shown to be proximal to sympathetic nerve terminals. Characterising subpopulations associated with intrinsic cardiac catecholaminergic signalling could lead to a better understanding of, and potential therapeutics for, cardiac pathologies where excessive catecholaminergic signalling plays a key role.

Keywords— Heart - Optogenetics - Electrophysiology - Sympathoexcitation - Catecholamines - RNA sequencing - RNA velocity - Lineage inference

Acknowledgements

I would like to thank my supervisor Dr Ming Lei for giving me the opportunity to do this project and for guidance throughout. I would also like to thank my co-supervisor Professor David Paterson.

Thank you to the entirety of the Lei Group for being very welcoming, open and helpful.

I am grateful to Tianyi Sun, for her tuition in immunohistochemistry and microscopy techniques. Yue Ren (Rina) for assistance in sectioning. Yu He (Hayley) for useful advice and feedback.

Alexander Grassam-Rowe for the constant guidance, advice and troubleshooting during the computational aspect of the project, and during the Langendorff-perfused electrophysiology.

Thank you to Rhea Suribhatla and Sophie Nield for useful feedback during lab meetings.

Table of Contents

1 Introduction **1**

1.1 Motivation 3

1.2 Canonical sources of cardiac catecholaminergic stimulation 4

1.3 Prior work: A third source of cardiac catecholaminergic stimulation? 5

1.4 Aims and thesis outline 10

2 Origins: Lineage inference using scRNA-seq **12**

2.1 Introduction 12

2.2 Methods 14

2.2.1 RNA velocity 20

2.2.2 Clustering 25

2.2.3 Embedding and downstream analysis 29

2.3 Results 32

2.4 Discussion 57

3 Electrophysiological properties: *Ex-vivo* whole-heart optogenetic pacing **63**

3.1 Introduction 63

3.2 Materials and methods 63

3.2.1 Mouse line 63

3.2.2 Reagents 64

3.2.3 Langendorff-perfused *ex-vivo* whole heart electrophysiology 65

3.2.3.1 *Ex-vivo* isolation of murine Dbh-Cre-ChR2/tdTomato hearts 65

3.2.3.2 Langendorff-perfusion experimental setup . . . 66

3.2.3.3 Light stimulation protocol 67
3.2.3.4 Administration of isoprenaline 69
3.2.3.5 ECG and statistical analysis methodology . . . 69
3.3 Results 72
3.4 Discussion 84

4 Spatial and morphological characteristics: Immunohistochemistry 88
4.1 Introduction 88
4.2 Methods 88
4.2.1 Fixation and sectioning 88
4.2.2 Immunostaining 89
4.3 Results 91
4.4 Discussion 96

5 Conclusions 97

Bibliography 98

Appendix A Supplementary figures 108

List of Figures

1.1 Pathways of cardiac stimulation 6
1.2 Catecholamine biosynthesis 7
1.3 Spatial distribution of Pnmt+ derived cells 9

2.1 Single-cell RNA sequencing experiment methodology 15
2.2 Characteristics of the scRNA-seq dataset 19
2.3 RNA velocity schematic 20
2.4 RNA velocity models 23
2.5 Modularities of example partitions 26
2.6 Louvain algorithm 27
2.7 RNA velocity phase portrait 31
2.8 Number of detected genes versus RNA count for E8.5 34
2.9 Percentage of RNA counts from ribosomal genes for E8.5 and neonate 35
2.10 Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E8.5 . . . 36
2.11 RNA velocity streamline plot 38
2.12 RNA velocity streamline plot by embryonic stage 39
2.13 RNA velocity inferred pseudotime, velocity length, velocity confidence and PAGA 40
2.14 RNA count inferred top differentially expressed genes 42
2.15 RNA count inferred top differentially expressed genes (continued) 43
2.16 RNA velocity inferred top differentially expressed genes 44
2.17 Developmental marker gene expressions by cluster 45
2.18 Gene expression heatmap for cluster nine 46
2.19 Cell lineage tracing using RNA velocity 53

2.20 RNA velocity and expression profiles for Dbh and Myl2 genes with a Myl2 phase portrait 54
2.21 Gene expression heatmap for the speculated Dbh expressing subpopulation origin cluster 56
2.22 Transcriptomic topology recapitulates spatial heart field topology . 61

3.1 Dbh-Cre-ChR2/tdTomato mouse line schematic 64
3.2 Langendorff-perfusion experimental setup 66
3.3 Electrophysiology of the mouse compared to the human 70
3.4 ECG measurement methodology 71
3.5 Optogenetic light pacing of the right ventricle 73
3.6 Failure to optogenetically light pace the left ventricle 74
3.7 Partial optogenetic pacing capture of the left ventricle 75
3.8 ECG intervals of a control mouse versus an optogenetic mouse for the right atrium 77
3.9 ECG intervals before, during and after light stimulation for each chamber of the heart using optogenetic mouse hearts 78
3.10 Threshold currents, refractory periods and threshold pulse durations for each chamber under light stimulation 80
3.11 Left ventricular light stimulation with and without isoprenaline . 82
3.12 Threshold currents for complete light pacing with and without isoprenaline 83

4.1 Micrographs of tdTomato fluorescence (marking Dbh+ derived cells) in the right and left ventricular walls 91
4.2 Whole heart spatial distribution of Dbh+ derived cells 92
4.3 Alpha-actinin immunostaining and tdTomato genetic fluorescence 94

4.4 High magnification micrograph (40x) of tdTomato genetic fluorescence (Dbh+ derived cells) in the left ventricular wall 95
4.5 Micrograph of the interventricular septum with tdTomato fluorescence and Th immunostaining 95

A.1 Number of detected genes vs RNA count for E8.5 109
A.2 Number of detected genes vs RNA count for E10.5 110
A.3 Number of detected genes vs RNA count for E12.5 111
A.4 Number of detected genes vs RNA count for E14.5 112
A.5 Number of detected genes vs RNA count for E16.5 113
A.6 Number of detected genes vs RNA count for the neonate 114
A.7 Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E8.5 . . . 115
A.8 Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E10.5 . . 116
A.9 Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E12.5 . . 117
A.10 Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E14.5 . . 118
A.11 Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E16.5 . . 119
A.12 Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for the neonate 120
A.13 Ventricular and atrial marker gene expressions by cluster 121

List of Tables

2.1 Marker genes expressed in each cluster as determined by each of the four differential expression methods used 47

3.1 Genotypes of each mouse 64

3.2 Optogenetic responses of each mouse heart by chamber 76

1 | Introduction

The heart is composed of different cell types.

Fibroblasts make up the fibrous skeleton of the heart, including the fibrotendinous ring that forms the base of the heart, the annulus fibrosus. This provides electrical insulation between the atria and ventricles, and defines the apertures of the atrioventricular and semilunar valves (Herring & Paterson, 2018). The skeleton also separates the four chambers with the interatrial, atrioventricular and interventricular septa.

Endothelial cells, pericytes and smooth muscle cells make up the coronary vasculature of the heart, allowing blood to supply oxygen and nutrients to the heart tissue and remove metabolic waste. Endothelial cells form the inside lining of the coronary veins and arteries, and line the inside of the heart, the endocardium. Pericytes wrap around endothelial cells and are important in controlling blood flow, homeostasis and angiogenesis. Smooth muscle cells provide the contractile power for vasoconstriction and vasodilation.

Myeloid and lymphoid cells make up the cardiac immune system, which acts to protect the heart from pathogens and maintain cardiac homeostasis (Swirski & Nahrendorf, 2018; Litviňuková et al., 2020).

Mesothelial cells contribute to a double-walled sac around the heart, the pericardium, which protects the heart from mechanical stresses and infection, as well as functioning to lubricate the heart.

Adipocytes form the epicardial adipose tissue that surrounds the heart. This provides the heart with fuel in the form of free fatty acids, mechanical protection, immunological support and thermoregulation (Antonopoulos & Antoniades, 2017).

Neuronal cells make up the fibres of the sympathetic and parasympathetic branches of the autonomic nervous system that innervate the heart (Boron & Boulpaep, 2017). The sympathetic fibres innervate the atria, sinoatrial node (SAN), atrioventricular node (AVN) and ventricles; whereas the vagus nerve innervates the atria, SAN, AVN and Purkinje fibres. They function to modulate the heart rate, AVN conduction velocity and the force of ventricular contraction.

Cardiomyocytes form the myocardium, the middle layer of the wall of the heart, between the epicardium and endocardium, which contracts in order pump blood around the body. Cardiomyocytes also make up the cardiac conduction system which efficiently generates and passes a wave of depolarisation in a controlled manner across the heart. The heart's population of cardiomyocytes differ in their phenotype by function and anatomical localisation, and can be grouped into subpopulations. Cardiomyocytes can be divided into those specialised for contraction, working cardiomyocytes, and those specialised for other functions, non-working cardiomyocytes (Grassam-Rowe, Ou, & Lei, 2020).

Within the subpopulation of working cardiomyocytes, there are atrial and ventricular cardiomyocytes. Ventricular cardiomyocytes are longer, wider and have a denser T-tubule network compared to atrial cardiomyocytes. Ventricular cardiomyocytes have larger and longer calcium transients, and longer action potentials and effective refractory periods than atrial cardiomyocytes (Walden, Dibb, & Trafford, 2009). Ventricular and atrial cardiomyocytes also express different isoforms of key contractile proteins. For example, the ventricular isoform of myosin regulatory light chain 2 (MLC-2v) is encoded by the MYL2 gene, whereas the atrial isoform (MLC-2a) is encoded by the MYL7 gene. In addition to atrial and ventricular differences, there is also heterogeneity in cardiomyocytes along transmural and apico-basal axes.

Non-working cardiomyocytes consist of specialised pacemaker and conductive cardiomyocytes. Pacemaker cells in the sinoatrial and atrioventricular nodes are smaller, contain fewer mitochondria and myofibrils, have fewer gap junctions and have smaller sarcoplasmic reticula (Boyett, Honjo, & Kodama, 2000; Honjo, Boyett, Kodama, & Toyama, 1996). They have the property of automaticity, that is, spontaneous depolarisation of membrane potential in phase four of the cardiac action potential. This allows them to function as pacemakers and distinguishes them from other cardiomyocytes. Purkinje fibres are cardiomyocytes specialised for rapid conduction of action potentials. They are larger, contain more mitochondria and glycogen, fewer myofibrils, and have a greater density of connexin 40 gap junctions (Ideker, Kong, & Pogwizd, 2009).

In this project, a specific subpopulation of cardiac cells were studied, distinguished by their association with a catecholamine-synthesising enzyme.

1.1 Motivation

Cardiovascular disease is the leading cause of death globally. Hyperactive catecholaminergic signalling is involved in a range of cardiac disease, from takotsubo cardiomyopathy to heart failure. Excessive levels of catecholamines overactivate α_1-adrenergic receptors on the cell membranes of smooth muscle cells which leads to intracellular Ca^{2+} overload, causing coronary artery spasms. Myocardial cells are damaged via myocardial ischaemia and hypoxia due to these coronary artery spasms. In addition, excessive levels of catecholamines overactivate the β_1-adrenergic receptors on cardiomyocytes, again leading to intracellular Ca^{2+} overload. This excessively increases contractility, causing hypoxia and results in myocardial cell damage. Furthermore, high levels of circulating catecholamines become oxidized into aminolutins, aminochromes and oxygen free radicals which

cause oxidative stress. This can lead to coronary spasms, arrhythmias, ultrastructural cell damage and rapid declines in contractile force (Adameova, Abdellatif, & Dhalla, 2009). In heart failure, excessive sympathetic signalling is used by the body to compensate for reduced cardiac output in the short term, but when this fails, chronic excessive sympathetic signalling then leads to myocardial damage and an accelerated progression of heart failure (Triposkiadis et al., 2009; Lymperopoulos, Rengo, & Koch, 2013). Excessive strain is put on the myocardium and pathological structural remodelling occurs.

1.2 Canonical sources of cardiac catecholaminergic stimulation

Adrenaline and noradrenaline are catecholamines that mediate the sympathetic stimulation of the heart. There are two canonical pathways of cardiac sympathetic stimulation. In the direct pathway, sympathetic post-ganglionic neurons directly innervate the heart and release noradrenaline, this neurotransmitter binds to β_1-adrenergic receptors on cardiac cell membranes. In the indirect pathway, stimulation by pre-ganglionic sympathetic neurons induces the secretion of adrenaline by chromaffin cells in the adrenal medulla of the adrenal glands, which sit above the kidneys. Acting as a hormone, adrenaline is then carried by the bloodstream to the heart, where it also binds to β_1-adrenergic receptors on cardiac cell membranes. Activation of β_1-adrenergic receptors has a positive chronotropic, inotropic, lusitropic and dromotropic effect, and shortens the cardiac action potential.

1.3 Prior work: A third source of cardiac catecholaminergic stimulation?

In 1968, work by Ignarro & Shideman on embryonic chick hearts showed tritium-labelled noradrenaline uptake even before sympathetic innervation. In addition, uptake persisted upon addition of reserpine and cocaine, which block noradrenaline transporter mediated reuptake and storage in sympathetic neurons (Ignarro & Shideman, 1968b, 1968c; Grassam-Rowe et al., 2020). Fluctuating concentrations of L-DOPA and the catecholamines—dopamine, noradrenaline and adrenaline—were found in the embryonic chick heart pre-innervation, in addition to their respective catecholamine-synthesising enzymes (Ignarro & Shideman, 1968a).

In 1974, work by Papka, using fluorescence and electron microscopy, on fetal and early postnatal rabbit hearts, demonstrated the presence of catecholamine-containing cells with many dense-core vesicles. These cells were clustered near the aorta and pulmonary artery (Papka, 1974). This and the work by Ignarro & Shideman were the start of a growing body of literature highlighting a cardiac non-neuronal cellular source of catecholamines (Ebert & Thompson, 2001; Huang et al., 1996). In effect, a potential "third" source of sympathetic stimulation, in addition to the two canonical sources described above (see Figure 1.1).

In 1996, work by Huang *et al.*, presented evidence for a cardiac cell subpopulation putatively involved in, or responsible for, this third sympathetic pathway intrinsic to the heart. They named this subpopulation—intrinsic cardiac adrenergic (ICA) cells. Northern and Western blots showed that isolated ICA cells contained mRNA and proteins involved in catecholamine synthesis. Electron microscopy revealed an abundance of clear vesicles, suggesting secretory capabilities. *In vitro*, isolated rat ICA cells constitutively released noradrenaline and adrenaline which in-

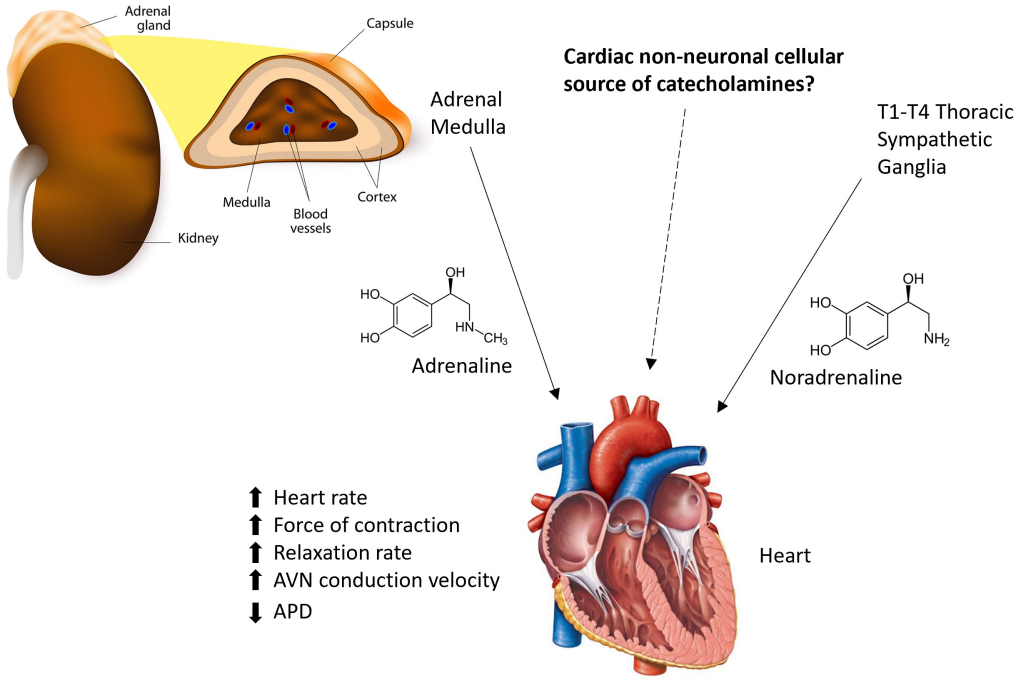


Figure 1.1: Pathways of cardiac stimulation. The two canonical pathways are shown with solid arrows and the putative intrinsic cardiac adrenergic pathway is shown by the dashed arrow. Image sources for the adrenal glands and heart: (Hopkins Medicine, 2021) and (Clip Art Library, 2021), respectively.

duced a positive chronotropic effect on nearby neonatal rat myocytes; the myocyte beating rate decreased by 58% upon addition of timolol, a β-adrenergic receptor antagonist. In addition, these ICA cells were identified in human fetal hearts before sympathetic innervation. Histology showed these cells to be rounded, unlike cardiomyocytes, and proximal to the coronary microvasculature. The ICA cells formed clusters, distributed diffusely across the heart (Huang et al., 1996).

Catecholamine biosynthesis involves a number of enzymes (see Figure 1.2). The enzyme phenylethanoamine N-methyltransferase (Pnmt) catalyses the conversion of noradrenaline into adrenaline.

In 2011, Ebert *et al.* generated a genetic mouse model where Cre-recombinase was inserted into the Pnmt gene locus, under Pnmt promoter regulation. The

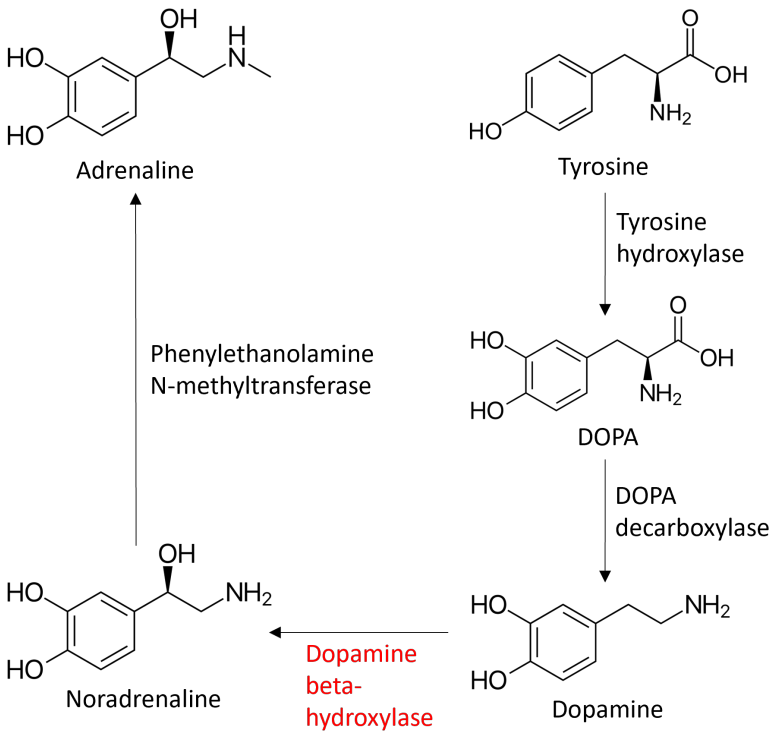


Figure 1.2: Catecholamine biosynthesis. The amino acid L-tyrosine is converted into the amino acid L-DOPA by tyrosine hydroxylase (Th). L-DOPA is converted into the first catecholamine, dopamine, by DOPA decarboxylase. Noradrenaline is synthesised from dopamine by the enzyme dopamine beta-hydroxylase (Dbh)—a focal point for this thesis. Noradrenaline is converted into the final catecholamine, adrenaline, via phenylethanolamine N-methyltransferase (Pnmt).

Pnmt-Cre mice were then crossed with ROSA26 reporter mice which had a stop cassette flanked by LoxP sites upstream of the LacZ gene. In cells that express, or have ever expressed, Pnmt, Cre-recombinase will also be expressed, which excises the stop cassette by recognising the LoxP sites at the ROSA26 locus, thereby allowing transcription of the LacZ gene. This leads to production of the LacZ gene product, β-galactosidase, which is an enzyme that catalyses the conversion of 5-bromo-4-chloro-1H-indol-3-yl β-D-galactopyranoside (X-gal), added previously to the tissue, into a bright blue coloured product. This genetic mouse model therefore labels the cells that express, or have ever expressed, Pnmt (including

progeny) with a bright blue compound. These Pnmt+ derived cells were shown to be distinctly distributed on the left side of heart (left atrium and left ventricle) (Osuala et al., 2011). The Pnmt+ derived cells included pacemaking, conduction and working cardiomyocytes (Ebert et al., 2004). This is contrary to ICA cells which have not been shown to express sarcomeric alpha-actinin (Grassam-Rowe et al., 2020).

In 2017, our group generated an optogenetic mouse model where Pnmt-Cre mice were crossed with mice that had a stop cassette flanked by LoxP sites upstream of a ChR2/tdTomato fusion protein gene (Wang et al., 2017). Channelrhodopsin-2 (ChR2) is a light-gated ion channel found in green algae; tdTomato is a fluorescent protein. Therefore, cells that express, or have ever expressed, Pnmt, will also express Cre-recombinase, which will excise the stop cassette by recognising the LoxP sites. This will permit transcription of the ChR2/tdTomato gene. Hence Pnmt+ derived cells will be fluorescently labelled via tdTomato and will be electrically excitable upon illumination by blue light via ChR2. The left-sided distribution of Pnmt was confirmed (see Figure 1.3). Many of the Pnmt+ derived cells appeared to be working cardiomyocytes. Pulsed illumination with blue light enabled optogenetic pacing of the heart through selective stimulation of the Pnmt+ derived cells.

The project conducted during this MSc by Research was an extension of this previous work, the catecholamine-synthesising enzyme dopamine β-hydroxylase (Dbh) was investigated instead of Pnmt. Dbh catalyses the conversion of dopamine into noradrenaline (Figure 1.2). This project utilised a Dbh-Cre-ChR2/tdTomato optogenetic lineage tracing mouse model, similar to the Pnmt model described above. In this model, cells that express Dbh or are derived from those that have ever expressed the Dbh gene also express channelrhodospin-2 (ChR2) (allowing them to be paced using light) as well as a tdTomato fluorescent marker.

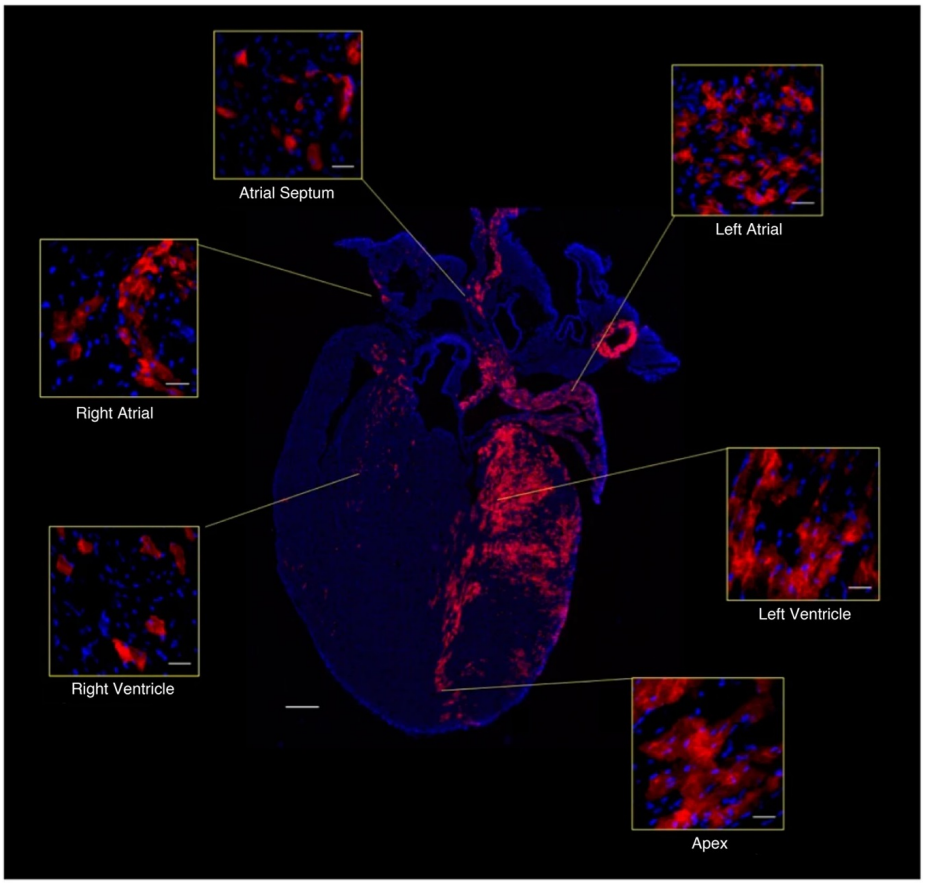


Figure 1.3: Spatial distribution of Pnmt+ derived cells. Coronal section from an adult Pnmt-Cre-ChR2/tdTomato mouse heart. Pnmt+ derived cells are predominantly distributed on the left-side of the heart, as shown by their tdTomato fluorescence (red). DAPI highlights the nuclei in blue. Main scale bar: $500\,\mu\mathrm{m}$; insert scale bars: $100\,\mu\mathrm{m}$. Adapted from (Wang et al., 2017).

Previous work has shown that targeted disruption of the dopamine β-hydroxylase gene leads, in homozygous mice, to death *in utero* (Thomas, Matsumoto, & Palmiter, 1995; Thomas, Marck, Palmiter, & Matsumoto, 1998). Death was suggested to be due to a lack of noradrenaline since uterine lethality could be prevented using a Dbh-independent precursor of noradrenaline, dihydroxyphenylserine. Another notable piece of work similar to that of this project is the Th-Cre-ChR2 optogenetic mouse model by Wengrowski *et al.* in 2015. Here, tyrosine-hydroxylase

(Th) positive cells, mainly sympathetic neurons, selectively expressed ChR2, which allowed for manual photoactivation of sympathetic neurons that then released noradrenaline, with the usual positive chronotropic and inotropic effects (Wengrowski et al., 2015).

1.4 Aims and thesis outline

The goal of this project was to characterise the Dbh expressing/Dbh+ derived cardiac cellular subpopulation. More specifically, the aims of this project were to determine the:

1. developmental cell origin(s);
2. electrophysiological properties and
3. spatial and morphological characteristics

of the Dbh expressing and Dbh+ derived subpopulations. This investigation was carried out using bioinformatical techniques, *ex-vivo* electrophysiology and immunohistochemistry.

The remainder of this thesis is organised as follows:

Chapter 2 — explores the developmental origins of the subpopulation by performing single-cell RNA-sequencing (scRNA-seq) on developing wild-type mouse hearts from embryonic day 8.5 (E8.5) through to neonates. Analysis includes a recently developed form of *in silico* lineage inference —"RNA velocity". From this, further analysis includes pseudotime-inference, RNA velocity inferred differential gene expression tests and an inspection of the transcriptional dynamics of individual genes;

Chapter 3 — examines the electrophysiological properties of the subpop-

ulation through light pacing of *ex vivo* Langendorff-perfused whole Dbh-Cre-ChR2/tdTomato mouse hearts. ECG intervals, ventricular refractory periods, threshold light currents and threshold pulse durations were measured across the four chambers of five hearts;

Chapter 4 — characterises the cellular morphology and spatial distribution of this novel subpopulation using genetic Dbh+ fluorescence combined with immunohistochemistry of the Dbh-Cre-ChR2/tdTomato mouse hearts, staining for markers of cardiomyocytes and sympathetic neurons;

Chapter 5 — summarises the findings and comes to overall conclusions.

2 | Origins: Lineage inference using scRNA-seq

2.1 Introduction

Heart development in the mouse starts from embryonic day 6.5 (E6.5) (i.e. days post-fertilisation) when cardiac progenitor cells either side of the primitive streak of the epiblast ingress through the primitive streak during gastrulation to form the cardiogenic part of the underlying intraembryonic mesoderm (i.e. splanchnic mesoderm—part of the lateral plate mesoderm) (Gilbert et al., 2000). These cardiogenic progenitors delaminate and migrate away from the primitive streak toward the anterior end of the embryo, at around E7.5 (Meilhac, Lescroart, Blanpain, & Buckingham, 2014). The bilateral fields merge cranially to form the first and second heart fields in a horseshoe-shape (Hill, 2021). Cells of the first heart field form bilateral endocardial tubes which then fuse at the midline to form the heart tube. Cells of the second heart field add to the posterior and anterior ends of the heart tube. The first heart field will form the left ventricular myocardium. The anterior second heart field will form the outflow tract at the arterial pole of the heart tube (eventually forming the basal regions of the pulmonary artery and aorta) as well as the right ventricular myocardium. The posterior second heart field will form the left and right atria at the venous pole of the heart (Meilhac et al., 2014).

In addition to splanchnic mesoderm, neural crest cells (originating from neurectoderm) also contribute to the heart. A subpopulation of neural crest cells (cardiac neural crest cells) undergo an epithelial-to-mesenchymal transition before migrating from the neural tube to the heart between E9.0 and E12.5 (Gessert & Kühl,

2010). They form the outflow tract septum separating the aorta and pulmonary artery, and the endocardial cushions (which will differentiate into the aortic and pulmonary valves) (George, Maldonado-Velez, & Firulli, 2020). Cardiac neural crest cells also contribute to the cap of the interventricular septum and both the sympathetic & parasympathetic innervation of the heart. Specifically, truncal neural crest cells form the sympathetic innervation of the heart (Hildreth, Anderson, & Henderson, 2009). Cranial neural crest cells contribute to the parasympathetic innervation of the heart (Owji, Varudkar, & Ebert, 2013).

The origin of the cardiac conduction system (SAN, AVN, Bundle of His and Purkinje fibres) has been controversial (Meilhac et al., 2014). It was originally proposed to be of neural crest origin (Gorza, Schiaffino, & Vitadello, 1988). Though new research suggests the majority of the ventricular cardiac conduction system is of mesodermal origin (Kitajima, Miyagawa-Tomita, Inoue, Kanno, & Saga, 2006). Recent work in our lab has shown that Pnmt+ derived cells are particularly distributed near the SAN and AVN regions of the cardiac conduction system (Wang et al., 2017). In addition, it has been shown that a wave of Pnmt+ cells invade the heart at around E10.5, suggestive of neural crest origins (Owji et al., 2013). Neural crest derived cells appear to contribute to neuroendocrine functions in the heart (Wang et al., 2017). Given all the above, our lab therefore speculated that the Dbh+ derived cells have a neural crest origin.

This chapter tries to test this by single-cell RNA sequencing hearts at different developmental time points. This technique involves extracting RNA from each cell in a sample to determine the gene expression profile of each cell. The data can be analysed to attempt to infer the identities of progenitors of cells of interest.

2.2 Methods

Mouse embryos from embryonic day (E) 8.5 and E10.5, and the hearts of mouse embryos from E12.5, E14.5, E16.5 and neonates, were single-cell RNA-sequenced by Novogene. These were C57 wild type mice from Luzhou, China. The NovaSeq 6000 (Illumina) machine was used. The sequencing was droplet based. Novogene provided raw messenger RNA count matrices, with each row containing, for the given cell, the number of mRNA molecules detected for each gene sequenced. Raw mRNA count matrices were provided for each embryonic stage. For each stage, there were two batches. For each batch, both a spliced mRNA count matrix and an unspliced mRNA count matrix was provided. Both spliced and unspliced abundances are required since a downstream analysis technique, RNA velocity, compares unspliced and spliced abundances. A summary is provided in Figure 2.1. As background, a rough outline of the initial experimental scRNA-seq process conducted by Novogene, from providing the samples to receiving the matrices described above, is as follows:

1. Library construction

    (a) Samples are dissociated into single cells where they are isolated into separate droplets (each assigned a unique "barcode") in a microfluidic device (Klein et al., 2015) (Macosko et al., 2015).

    (b) The cell in each droplet is lysed. The cellular mRNA is broken up into small (200-300 base pair) fragments and captured (e.g. for polyadenylated mRNA, poly(T) primers are used to bind to poly(A) mRNA tails) (Haque, Engel, Teichmann, & Lönnberg, 2017a). This is required since sequencing devices can only sequence short fragments. All mRNA fragments in this droplet are identified from other cells'

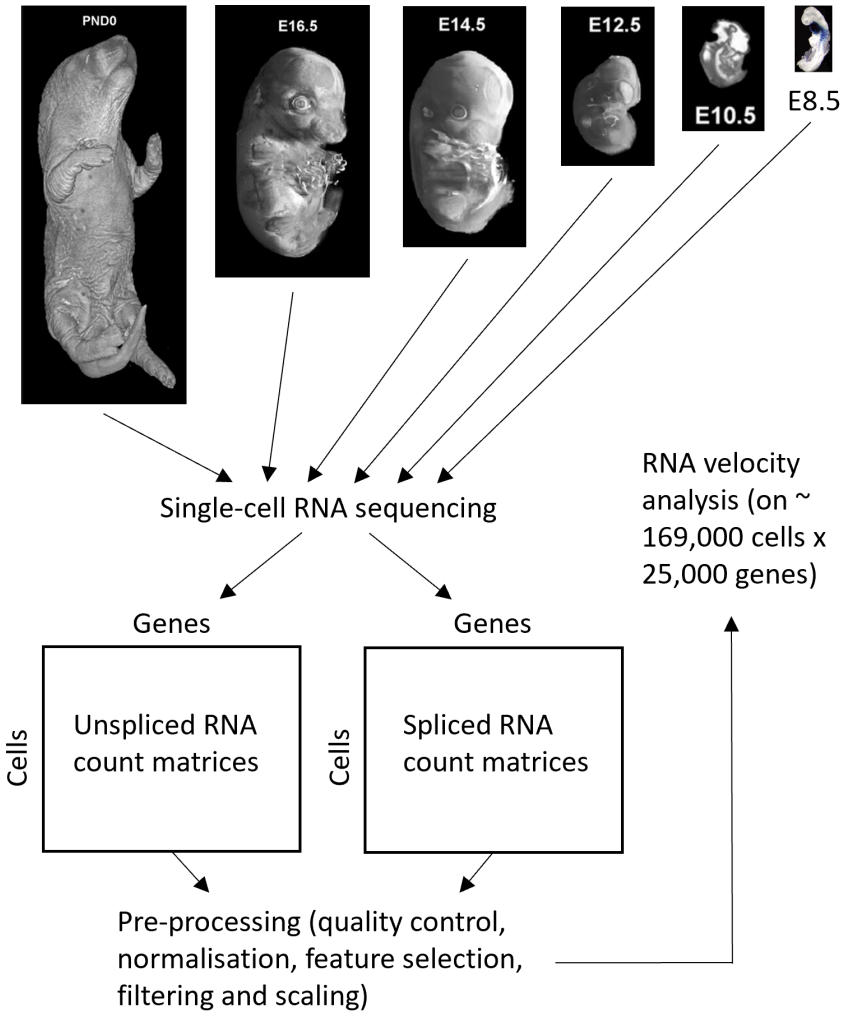


Figure 2.1: Single-cell RNA sequencing experiment methodology. Wild-type mouse embryos from E8.5 to neonates were single-cell RNA sequenced for both spliced and unspliced RNA counts. These count matrices were pre-processed and then RNA velocity analysis was performed using them. The dataset consisted of approximately 169k cells and 25k genes after the pre-processing stage. Mouse embryo images from (Petiet et al., 2008) and (Long et al., 2007).

mRNA by the unique barcode mentioned previously. Each captured mRNA molecule is labelled with a "unique molecular identifier" (UMI) (Luecken & Theis, 2019).

(c) The cellular mRNA molecules are reverse transcribed into complementary DNA (cDNA). cDNA is more stable than RNA and is more easily amplified and modified.

(d) Adapter sequences are added to the cDNA. These allow the next-generation sequencing machine to recognise the fragments and for different cells to be sequenced simultaneously.

(e) The cDNA is amplified by PCR.

(f) Quality control measures are taken. The library concentration of cDNAs and their fragment lengths is verified.

2. Next-generation sequencing

(a) The vertical fragments of cDNA are aligned in a grid called a flow cell. Fluorescent probes, with a unique colour for each nucleotide type (adenine, thymine, cytosine or guanine), are attached to the first base in each sequence (Starmer, 2017).

(b) An image of the flow cell taken by the machine from above identifies the first base of each fragment based on the colour of each probe. The colour is then washed off the probes.

(c) The fluorescent probes are then attached to the second base in each sequence.

(d) Another image of the flow cell is taken by the machine from above, which identifies the second base of each fragment based on the colour

of each probe. The colour is washed off the probes.

(e) Steps (c) and (d) are then repeated for all the remaining bases in each sequence.

(f) The machine outputs the sequence of nucleotides for each cDNA, along with measures of confidence (quality scores) that each nucleotide has been correctly assigned in the sequence. Faded probes and over abundance of single colours (low diversity) can reduce the confidence of the machine in correctly identifying bases.

(g) Low quality reads are filtered out based on the quality scores of the bases and biochemical artefacts (such as adapter sequences bound to each other).

(h) The sequenced reads are split into fragments and aligned to genome fragments, providing locations on the genome (chromosome and locus). By fragmenting the reads and genome, alignment can take place even if the whole read does not exactly match the genome.

(i) The location of each read is compared to the locations of known genes in the genome. An mRNA count matrix is therefore constructed which counts up the number of reads assigned to each gene, for each cell.

The R and Python programming languages were used. The code used for the following bioinformatical processing and analysis can be found at `https://github.com/Alexei-Lipov/rna-velocity-project-2020-21`. The author would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work (Richards, 2015). The script "scVeloPrep13.R" was used first to concatenate all the matrices of each stage and batch for the unspliced mRNA dataset. Then cells

with zero nUMI (which should correspond to cells with no detected RNA in the genes sequenced, and so in this case also corresponds with zero detected active genes, nGene $= 0$) are filtered out. The resulting total matrix (unspliced mRNA counts for all embryonic stages and batches) consisted of 27,933 genes x 448,478 cells sequenced. The spliced dataset had the same dimensions.

A colleague in the lab, Alexander Grassam-Rowe, performed some initial filtering on the spliced dataset resulting in a reduction to 27,477 genes x 169,144 cells. Cells were filtered out according to the following parameters: nUMI < 300, nGene < 270, log10GenesPerUMI < 0.8, mitoRatio > 7.5%; where nUMI (number of unique molecular identifiers) is the number of mRNA molecules detected in the cell, nGene is the number of genes detected as expressed in the cell, log10GenesPerUMI $= [\log_{10}(\text{nGene}) / \log_{10}(\text{nUMI})]$ and where mitoRatio $=$ mtUMI/nUMI (here mtUMI is the number of RNA molecules transcribed from mitochondrial DNA). Quality control plots were generated using the scripts "AllStagesSplicedQC.R", "ColouredQCandHistoPy.py", "e85qc.R", "e105qc.R", "e125qc.R", "e145qc.R", "e165qc.R" and "NRCMqc.R"; the plots are presented in Section 2.3.

This spliced dataset was fed in as an input to "scVeloPrep13.R". The aforementioned unspliced dataset's cells were then filtered according to the spliced filtered dataset's cells. Reducing the unspliced dataset dimensions from 27,933 genes x 448,478 cells to 27,933 genes x 169,144 cells.

Both the filtered unspliced and spliced datasets from "scVeloPrep13.R" were then fed in as inputs to "scVelo21.py", a script primarily using the Scanpy (Wolf, Angerer, & Theis, 2018) and scVelo (Bergen, Lange, Peidli, Wolf, & Theis, 2020) packages in Python 3.6. The unspliced and spliced datasets were merged as layers of Scanpy's adata object, resulting in a dimension of 25,787 genes x 169,144 cells.

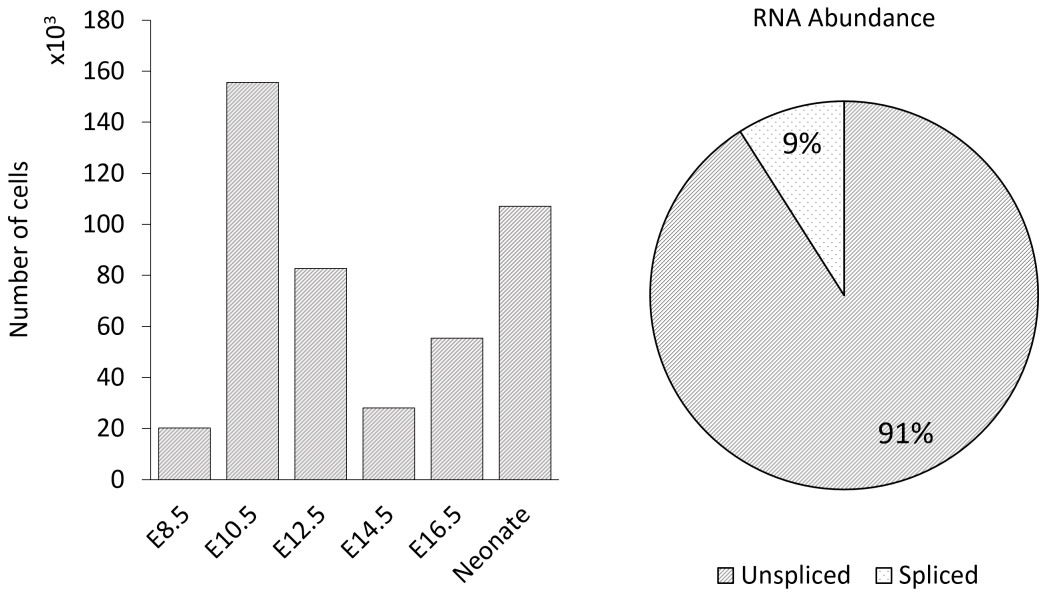


Figure 2.2: Characteristics of the scRNA-seq dataset. The dataset used here is before the pre-processing stage. Left: bar chart showing the number of cells for each embryonic stage. Right: pie chart showing the proportion of RNA counts that were unspliced or spliced for the entire unprocessed dataset (all stages and batches).

By RNA abundance, 91% of the counts were unspliced, while 9% were spliced, see Figure 2.2. The counts of the combined dataset were then normalised per cell such that each cell had a total count equal to the median of the pre-normalisation counts per cell. Normalisation is performed since each cell will have a different total count due to technical variations related to the scRNA-seq process such as varying number of low quality reads or varying flow cell concentrations; this thereby prevents the assignment of an erroneous biological interpretation to the total count per cell variation. Each of the dataset matrices (i.e. both spliced and unspliced layers), X, were then logarithmised, by performing $X = \ln(X + 1)$. Verification of absence of duplicate cells was performed at this stage.

After this, first and second order moments were calculated and RNA velocities were calculated using the stochastic method. A brief explanation of RNA velocity

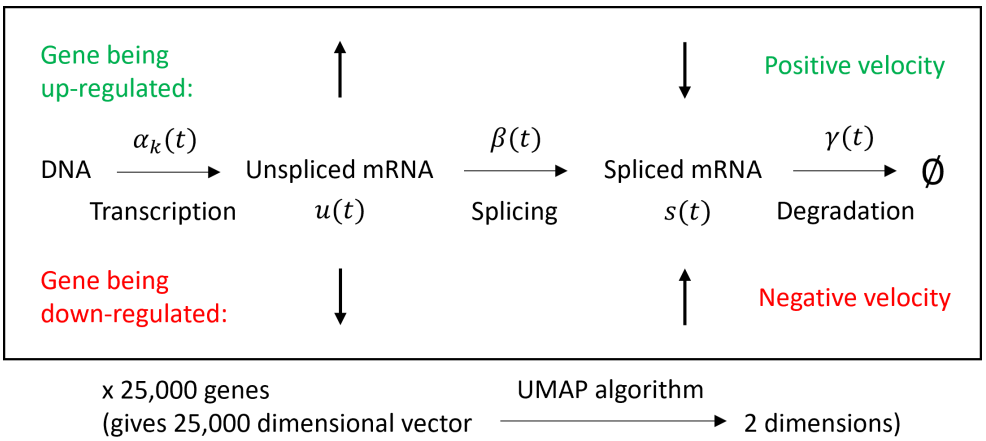


Figure 2.3: RNA velocity schematic. DNA is transcribed into RNA which is spliced and eventually degraded. The transcriptional kinetics of the system describe the rate of these processes. RNA velocity models the up-regulation and down-regulation of genes using these transcriptional kinetics. For each gene, a velocity is inferred (positive or negative). In our case, this gives a ~25,000 dimensional velocity (one for each gene) for each cell (of which there are ~169,000 in our case). A dimensionality reduction algorithm, such as UMAP, collapses this into a 2 dimensional velocity for each cell—allowing humans to interpret the transcriptional trajectories of each cell. Notation is explained in the main text.

will follow.

2.2.1 RNA velocity

RNA velocity (see Figure 2.3) is a way of analysing scRNA-seq data in order to visualise the flow of cellular differentiation across the subpopulations of the sample over a developmental time period. From this, hypotheses can be generated about the origins of certain cellular subpopulations, i.e. *in-silico* lineage inferencing. Unlike other forms of lineage inference, computational RNA velocity analysis is hypothesis-generating. This makes it a useful complement to biological lineage inference (such as the genetic reporter Cre mouse line used in this project) which is often hypothesis-testing. RNA velocity works by making use of both spliced and unspliced mRNA abundances from scRNA-seq. In a cell, for

a particular gene, the concentrations of unspliced pre-mRNA and spliced mature mRNA follow the process as shown in Figure 2.3. DNA is transcribed into unspliced pre-mRNA at a rate $\alpha_k(t)$, where k is the transcriptional state of the cell, i, for that particular gene. The possible transcriptional states of the i-th cell for a particular gene can be "ss_{off}", meaning the cell is not currently transcribing the gene and so there is a steady state of no mRNA for that gene; "on", meaning transcriptional induction of the gene with a growing mRNA concentration; "ss_{on}", meaning the gene is active and is continually being transcribed to give a steady level of mRNA and "off", which indicates transcriptional repression of the gene and decreasing mRNA concentration. That is, $k_i \in \{\text{off}, \text{on}, \text{ss}_{\text{off}}, \text{ss}_{\text{on}}\}$.

Then, the unspliced pre-mRNA, at concentration $u(t)$, is spliced at a splicing rate $\beta(t)$ into spliced mature mRNA, with a concentration $s(t)$. The spliced mature mRNA is then degraded at a rate $\gamma(t)$. The concentration of unspliced mRNA is increased by transcription and decreased by splicing, whereas the concentration of spliced mRNA is increased by splicing and decreased by degradation. We may therefore write the reaction kinetics of the system, describing the rate of increase of unspliced and spliced mRNA concentration, as:

$$\frac{du(t)}{dt} = \alpha_k(t) - \beta(t)u(t), \tag{2.1a}$$

$$\frac{ds(t)}{dt} = \beta(t)u(t) - \gamma(t)s(t). \tag{2.1b}$$

This holds for each gene, separately, independent of other genes, in a given cell. The task is now to solve these equations. The RNA velocity, ν, is given by the time derivative of the spliced mRNA concentration, $ds(t)/dt$. Inspecting the equation for $ds(t)/dt$, we can see that, all else being constant, if the unspliced mRNA abundance, $u(t)$, is sufficiently greater than would be expected in steady state,

then $\nu = ds(t)/dt$ is large and positive, indicating up-regulation of the gene. On the other hand, if the unspliced mRNA abundance, $u(t)$, is sufficiently less than would be expected in steady state, then ν is large and negative, indicating the gene is being down-regulated. Hence, for a given gene in a given cell, the sign of ν indicates whether the gene is being induced or repressed, and the magnitude of ν gives the extent to which the gene is being upregulated or downregulated.

There are numerous approaches to estimating ν. The original model, termed the steady-state model, formulated by the authors of the original RNA velocity paper (La Manno et al., 2018), and the stochastic and dynamical models, proposed later as improvements on the original model and implemented in the Python package scVelo (Bergen et al., 2020). A comparison between the deterministic (steady-state) and dynamical models is given in Figure 2.4.

The steady-state model makes the assumption of time-independence of the rates, $\alpha(t) = \alpha$, $\gamma(t) = \gamma$, and sets the splicing rate, $\beta(t) = 1$, across all genes in a given cell. This is a large assumption given that splicing rates are likely to vary across different genes within a cell. To determine the gene-specific rate γ, a steady-state assumption is used,

$$\frac{ds(t)}{dt} = 0, \tag{2.2}$$

giving, from Equation (2.1b),

$$\gamma = \frac{u}{s}. \tag{2.3}$$

Using this steady-state unspliced to spliced mRNA ratio, the RNA velocity is estimated as the deviation of the observed unspliced to spliced mRNA ratio from this steady-state ratio. The steady-state assumption is realistic for genes in terminally differentiated subpopulations, but not for transiently expressed genes in developing subpopulations.

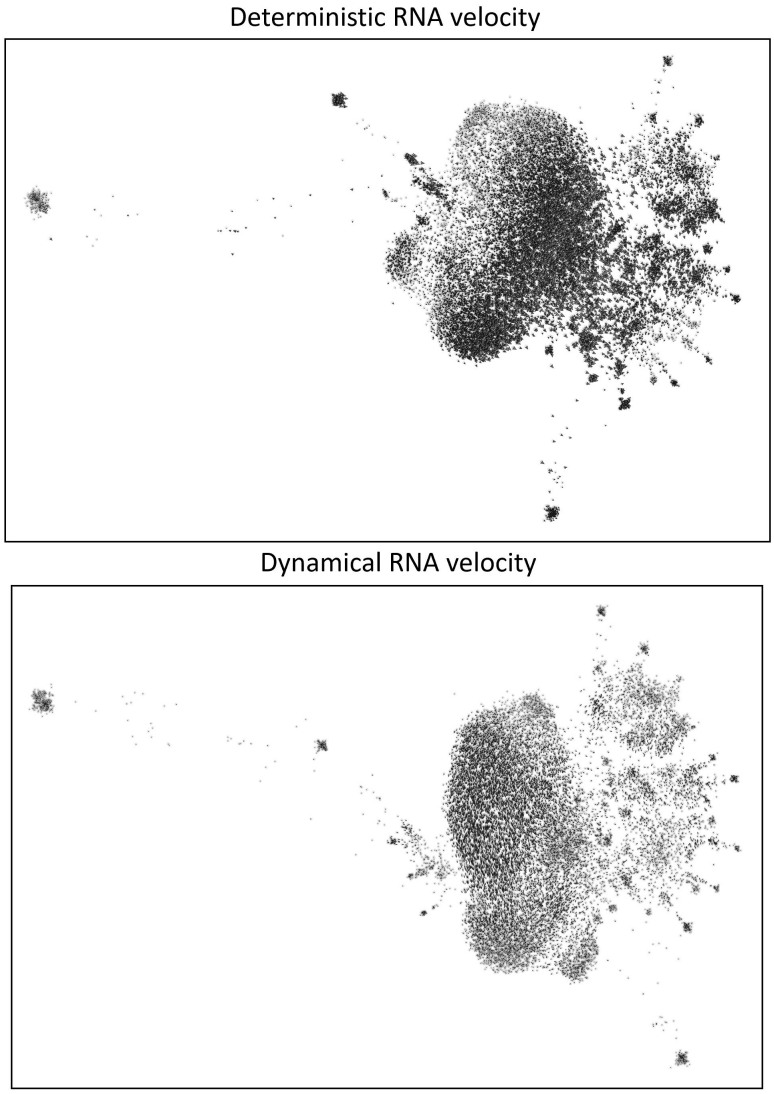


Figure 2.4: RNA velocity models. UMAP embeddings of RNA velocity for a truncated version of the dataset. Top panel used the deterministic RNA velocity model. Bottom panel used the dynamical RNA velocity model. The dynamical model is more powerful but more computationally expensive than the deterministic method.

The later proposed stochastic and dynamical models (Bergen et al., 2020) improve on the steady-state model by resolving the restrictions mentioned above, removing the need for the steady-state and common splicing rate assumptions. The stochastic model treats transcription, splicing and degradation as probabilistic events, and

the Markov jump process is approximated using moment equations. The inclusion of second moments adds the covariance of unspliced and spliced mRNA to the model.

The dynamical model solves Equations (2.1) using a likelihood-based expectation-maximisation framework which iteratively estimates the rates α_k, β and γ, and the i-th cell's transcriptional state and internal time, k_i and t_i.

The dynamical model is the most powerful, but also the most computationally expensive. The stochastic model improves upon the steady-state model and approximates the results of the dynamical model, but is less computationally expensive, on par with the steady-state model. Therefore, in our project, the stochastic model was used, unless otherwise stated.

For a given cell, the models infer an RNA velocity for each gene. This results in a ~25,000 dimensional RNA velocity vector for each cell given that around 25,000 genes were detected. We therefore have around 169,000 RNA velocity vectors, one for each cell, in a 25,000 dimensional "transcriptomic" space. More formally, we have the mRNA abundance matrix $\mathbf{M} \in \mathbb{R}^{N \times D}$, where N is the number of cells and D is the number of genes. The i-th row of $\mathbf{M}$ is the gene expression vector of the i-th cell, $s_i \in \mathbb{R}^D$. The models infer RNA velocity vectors for each cell, $(\nu_i)_{i=1,\dots,N}$, with $\nu_i \in \mathbb{R}^D$. The difference in gene expression between two cells, i and j, is given by $\delta_{ij} = s_j - s_i$. We create a similarity matrix, π_{ij}, to judge, from the i-th cell, which j-th cells produce a gene expression difference closest to that encoded by the RNA velocity ν_i of the i-th cell. We do this by taking the cosine of the angle between δ_{ij} and ν_i,

$$\pi_{ij} = \cos \angle (\delta_{ij}, \nu_i) = \frac{\delta_{ij}^T \nu_i}{|\delta_{ij}||\nu_i|} \tag{2.4}$$

(Bergen et al., 2020). A "velocity graph" is created, showing likely cell-to-cell "transitions". Connections are made between cells when the similarity between the cells in the similarity matrix is high enough. For each cell, only the k nearest neighbours are considered.

The similarity matrix is then converted into a transition matrix describing cell-to-cell transition probabilities, using an exponential kernel:

$$\widetilde{\pi}_{ij} = \frac{\exp\left(\pi_{ij}/\sigma_i^2\right)}{\sum_j \exp\left(\pi_{ij}/\sigma_i^2\right)}, \tag{2.5}$$

where σ_i are the kernel width parameters. A cell's ancestors can be determined by applying the transposition of the transition matrix, this is how the later described cell-tracing plots were generated.

2.2.2 Clustering

The Louvain hierarchical clustering algorithm was used on the velocity graph to identify communities. Here we briefly explain this. A graph consists of a set of nodes (in this case, cells) connected to each other with a set of edges. A community is a subgraph of a network that is locally dense and connected (Barabási, 2016). Therefore, a community is a group of nodes that are preferentially linked to each other, rather than other nodes in the network (locally dense), and is such that there exists a path from any node in the group to any other node in the group (connected). The network's community structure can be derived from the adjacency matrix which describes the connectivity of the network.

To identify the communities within a network, hierarchical clustering algorithms are used. The set of communities that the algorithm divides the network into is called a partition. The clustering algorithm can often create different "resolution" partitions, that is, splitting the network into many small communities (high "reso-

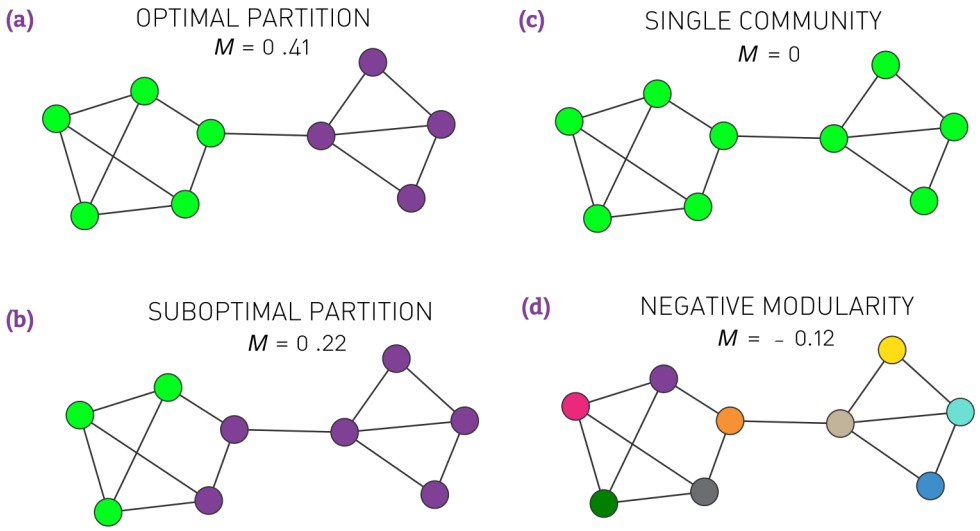


Figure 2.5: Modularities of example partitions. Four different partitions are shown for the same network (a, b, c, and d), with their corresponding modularities. Each node is assigned a community, identified by the colour of the node. Modularity is a metric lying between, and including, -0.5 and 1.0, the closer a partition's modularity is to 1.0, the better the partition. Adapted from (Barabási, 2016).

lution") or into a few large communities (low "resolution"). A dendrogram, often produced by a hierarchical clustering algorithm, shows the different partitions that the algorithm has been able to generate.

Modularity is a metric used to assess how close a partition is to the "true" underlying community structure (see Figure 2.5). It takes values between, and including, -0.5 and 1.0, and is maximised in the partition that most closely approximates the "real" underlying community structure. The modularity, Q, is given by:

$$Q = \frac{1}{2m} \sum_{i,j=1}^{n} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta_{c_i, c_j}. \tag{2.6}$$

Here, δ is the Kronecker delta function, c_i is the i-th community, m is the number of edges, k_i is the degree of the i-th node, A_{ij} is the adjacency matrix and n is the number of nodes (Barabási, 2016). A group of nodes that are preferentially linked

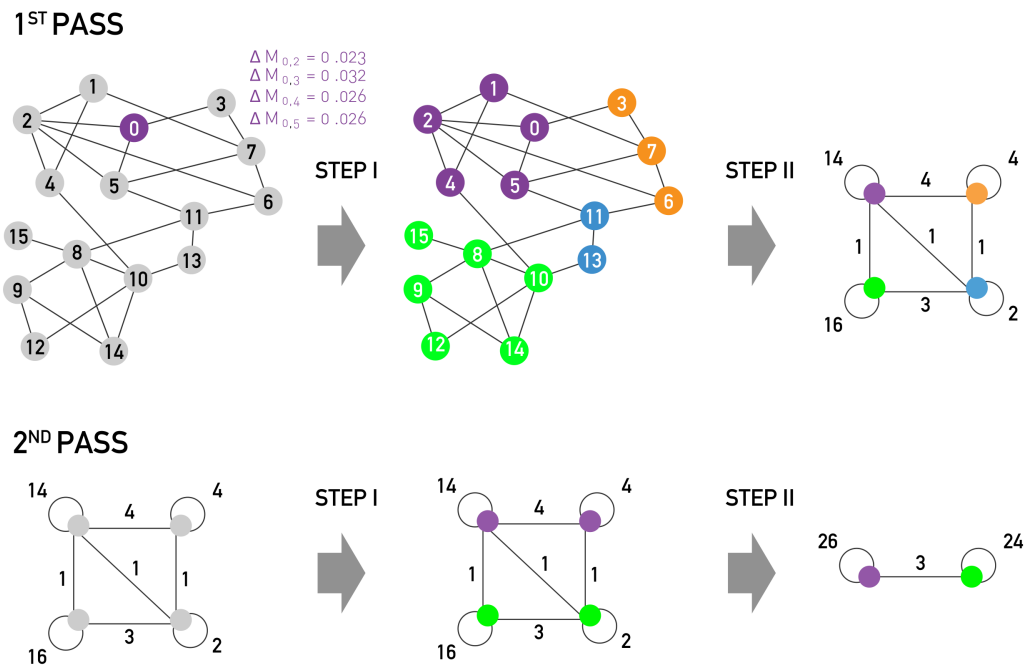


Figure 2.6: Louvain algorithm. This hierarchical clustering algorithm finds the optimal partition of a network through repeated local modularity optimisation and community aggregation. See main text for the steps of the algorithm visualised here. Adapted from (Barabási, 2016).

to each other has more edges within the group than it would otherwise have if the edges in the entire network were assigned randomly. That is, in a community, the nodes link each other more than they would by chance. The modularity metric takes the difference between each community's edge density and the edge density that would be expected in that community in a randomised model of the network. The greater the difference in edge density compared to chance, the greater the difference and so the greater the modularity value, Q, hence hierarchical clustering algorithms often seek to maximise Q.

The Louvain algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) is one of the fastest and most scalable hierarchical clustering algorithms (see Figure 2.6). It has a computational complexity of at most $\mathcal{O}(m)$. The Louvain algorithm con-

sists of "passes", with each pass consisting of two steps (Barabási, 2016). Every node is assigned to its own individual community, then the first step of the first pass begins:

1. The i-th node is put in a community, $\mathcal{C}$, with one of its the neighbouring nodes, j. The change in modularity, ΔQ, as a result of this action is calculated as

$$\begin{aligned}\Delta Q = &\left[\frac{\Sigma_{in} + 2k_{i,in}}{2W} - \left(\frac{\Sigma_{tot} + k_i}{2W}\right)^2\right] \\ &- \left[\frac{\Sigma_{in}}{2W} - \left(\frac{\Sigma_{tot}}{2W}\right)^2 - \left(\frac{k_i}{2W}\right)^2\right],\end{aligned} \tag{2.7}$$

   where Σ_{in} is the sum of the link weights connecting nodes within $\mathcal{C}$, Σ_{tot} is the sum of the link weights of the nodes in $\mathcal{C}$, k_i is the sum of the weights of the links connecting i, $k_{i,in}$ is the sum of the link weights connecting i and any node in $\mathcal{C}$, W is the sum of all the link weights in the network. The i-th node is then instead put into a community with one of its other neighbours, and the ΔQ is calculated. This is repeated for all neighbours of node i. Node i is then put into a community with whichever neighbouring node produced the greatest ΔQ. This process is then repeated for all i nodes in the network.

2. The nodes in each community of the network are aggregated to single nodes. The weight of the link between two nodes is given by the sum of the weights of all the links connecting nodes of each corresponding community. A self loop link is given to each node (corresponding to a community) where the weight is given by the sum of all the weights of the links connecting nodes within the community.

This completes the first pass. Steps (1) and (2) are repeated for the next pass.

Passes are repeated until no change in the network occurs and modularity is maximised. This gives the optimal community structure. In this project, the Louvain algorithm was used to partition the velocity graph into a variety of resolutions, from 9 clusters to 31 clusters.

2.2.3 Embedding and downstream analysis

The 169,000 RNA velocity vectors are in a 25,000 dimensional transcriptomic space. To interpret this high-dimensional data, dimensionality reduction algorithms were used to embed the 25,000 dimensional space into 2 dimensions that we can visualise.

In this project, we used the Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction technique (McInnes, Healy, & Melville, 2018). The 169,000 cells are then represented on a 2D plot with a 2D RNA velocity vector for each cell. By plotting curves that are instantaneously tangent to the velocity vectors, we obtain a streamline RNA velocity plot. This shows the flow of cellular differentiation from cellular subpopulation to cellular subpopulation.

Partition-based graph abstraction (PAGA) was used to infer trajectories between clusters (Wolf et al., 2019). The technique generates a graph where the clusters (obtained by applying a clustering algorithm to the velocity graph) are nodes and edges are assigned between them indicating the connectivity between the clusters. The weight of the edges indicates the confidence in connectivity. Directed edges in RNA-velocity inferred PAGA arise from deducing cluster-to-cluster connection directionality from the RNA velocity vectors. This aims to give a coarse-grained view of the flows of cellular differentiation between the subpopulations.

Differential gene expression analysis was conducted in combination with extracting marker genes from the literature in order to biologically identify the clusters in

the scRNA-seq data. Marker genes are genes which are expressed more in a certain subpopulation compared to other subpopulations, and thereby its expression "marks" that subpopulation. Differential gene expression analysis involves identifying genes that are expressed more in one cluster relative to all the other clusters. In this project, this was done in two ways: (1) using a t-test method for ranking highly differentially expressed genes from the RNA expression matrix; and (2) ranking genes in each cluster by applying a differential expression test (Welch t-test with overestimated variance) on RNA velocities thereby highly ranking genes with differential gene transcriptional dynamics relative to other clusters.

The transcriptional dynamics of individual genes can be visualised using gene "phase portraits". Here, the spliced and unspliced mRNA expression for each cell is plotted for a given gene. The model steady-state spliced/unspliced mRNA ratio line is overlaid on the plot. Cells that express more unspliced than would be expected by the steady-state ratio are likely upregulating the gene and cells that express less unspliced are likely downregulating the gene. Therefore "curves" of gene induction and repression can be seen in the phase portrait, as shown in Figure 2.7, showing the transcriptional dynamics of the gene over latent time. Driver genes, those that are influential in transitioning one cluster to another, can be identified from these phase portraits, since the cells can be coloured by cluster in the phase portrait. Clockwise directionality from cell cluster to cell cluster in the phase portrait suggests that gene may be driving the process of differentiation for the respective clusters.

The RNA-velocity inferred transition matrix allows terminal cells to be predicted (root cells and end cells). A random-walk is then performed from root cells to generate an RNA-velocity graph inferred pseudotime plot. This shows the models' prediction of the "time" point of each cell, in the transcriptomic state it was in when sequenced, within the context of the developmental timeline of the tissue.

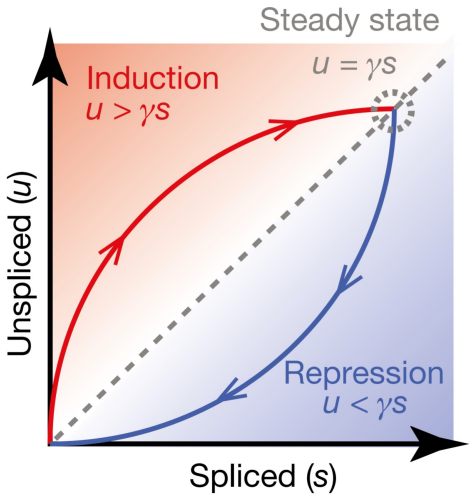


Figure 2.7: RNA velocity phase portrait. The total unspliced mRNA and spliced mRNA counts are plotted on the phase portrait of a specific gene for each cell. The dashed line is the steady-state ratio of unspliced to spliced mRNA count. Cells with above the line are up-regulating the gene, whereas those below the line are down-regulating the gene. Adapted from (La Manno et al., 2018).

This RNA-velocity inferred pseudotime is an advancement over earlier diffusion-based pseudotime methods (Haghverdi, Büttner, Wolf, Buettner, & Theis, 2016).

A measure of the rate of cellular differentiation across the cells of the subpopulations can be visualised by plotting the magnitude of the RNA velocity vector of each cell. A measure of confidence that the RNA velocity vectors in a given area are accurate is given by calculating the coherence of the vector field. That is, an area in which many RNA velocity vectors point in the same direction (high coherence) gives greater confidence in the RNA velocity directionality in that region.

To carry out the biological identification of the scRNA-seq clusters (clustered by the Louvain algorithm), a large number of marker genes was first extracted from the relevant literature. Then, genes were examined in the scRNA-seq data using various plots. The phase portrait (unspliced vs spliced expression), described earlier, for each marker gene was examined to observe the gene dynamics. "Gene

velocity" plots were generated which showed the magnitudes of the RNA velocities for the given gene, across all the cells (in the embedded UMAP transcriptomic space). This gives an indication as to where, and to what extent, the gene is upregulated (or downregulated) in any given area (cellular subpopulation). "RNA expression" plots were generated by plotting the mRNA expression for the given gene in each cell, in the transcriptomic space. Heatmaps were generated that showed the mRNA expression levels (as the heatmap colour) of cells (as rows) for the top 50 genes (as columns) ordered according to differentially expressed RNA velocity gene rankings for each cluster. Stacked violin plots showing the mRNA expression of marker genes for each cluster were generated. These violin plots show the probability density distribution of the mRNA expression, the minimum and maximum values and the median is shown by the colour of the plot. Furthermore, the previously described abundance derived and RNA velocity derived differentially expressed genes were plotted against t-test score for each cluster. By inspecting marker genes in the above described plots and inspecting individual clusters in these plots, biological identities were assigned to each cluster (subpopulation), as will be elaborated upon in Section 2.3.

After the probable biological identities of the clusters were ascertained, inferences could be made as to the directionality of differentiation between clusters using the RNA velocity streamlines. More specifically, speculations on the potential origins of the Dbh expressing subpopulation could be formulated.

2.3 Results

The dataset characteristics were summarised in Figure 2.2. The number of cells per stage ranged from 20,280, for E8.5, to 155,558, for E10.5. Therefore, the E10.5 and neonatal stages were over-represented in the dataset. The majority,

91%, of the mRNA transcripts detected were unspliced. This is large compared to literature values, for example, an unspliced proportion of 15-25% was observed by (La Manno et al., 2018), with previous observations of 14.8% (Gaidatzis, Burger, Florescu, & Stadler, 2015) and ~20% (Picelli et al., 2013).

Quality control plots, shown in Figures 2.8, 2.9 and 2.10, were used in order to manually set filters to remove poor quality cells. For example, for E8.5, Figure 2.8 shows the distribution of cells removed by filtering out cells which have an `nUMI` (`nCount_RNA`) < 300 and `nGene` (`nFeature_RNA`) < 270, shown by the red lines. Each stage contained cells with high percentage mitochondrial counts when the RNA count and number of genes detected was low. Figure 2.9 demonstrates that the percentage of counts from ribosomal genes decreased from ~40% to ~15% when comparing E8.5 to the neonatal stage, suggesting increased protein synthesis during early development relative to later development. A measure of the "novelty" or "complexity" is given by `log10GenesPerUMI`, which is a logarithmised form of the number of genes detected per RNA count for each cell. That is, cells with high "complexity" have mRNA transcripts for a wide variety of genes, whereas cells with low "complexity" have many mRNA transcripts for only a few genes. Red blood cells are an example of a lower complexity cell. All stages had `log10GenesPerUMI` distributions above 0.8, consistent with recommendations in the literature (Mary Piper, 2021). As an example, the distribution for E8.5 is shown in Figure 2.10. This figure also demonstrates that, for E8.5, the average number of genes detected was ~4000, after we have excluded the bottom peak of low-quality cells with very low numbers of genes. A similar story was present with the bimodal distribution of total RNA counts; excluding the low-quality cells, the main peak was ~25,000 total RNA counts per cell. Mitochondrial count percentages were low across all stages, less than ~2.5% for E8.5 as shown. An exception was the neonatal stage which approached 25% (see Ap-

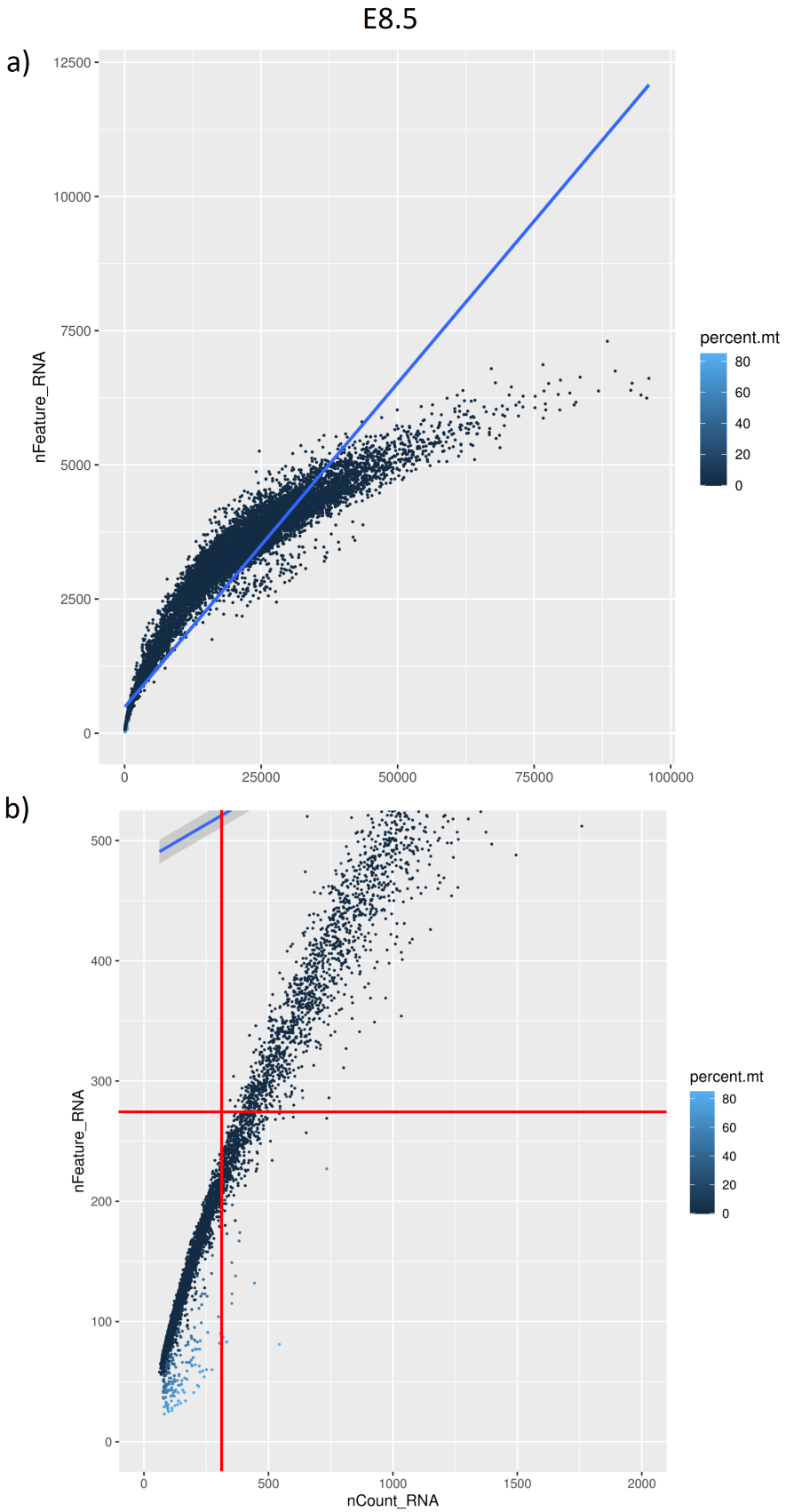


Figure 2.8: Number of detected genes vs RNA count for E8.5. Blue: proportion mitochondrial RNA. Blue line: best fit line. Red lines: filter thresholds. b) A magnified version of the bottom left region of (a).

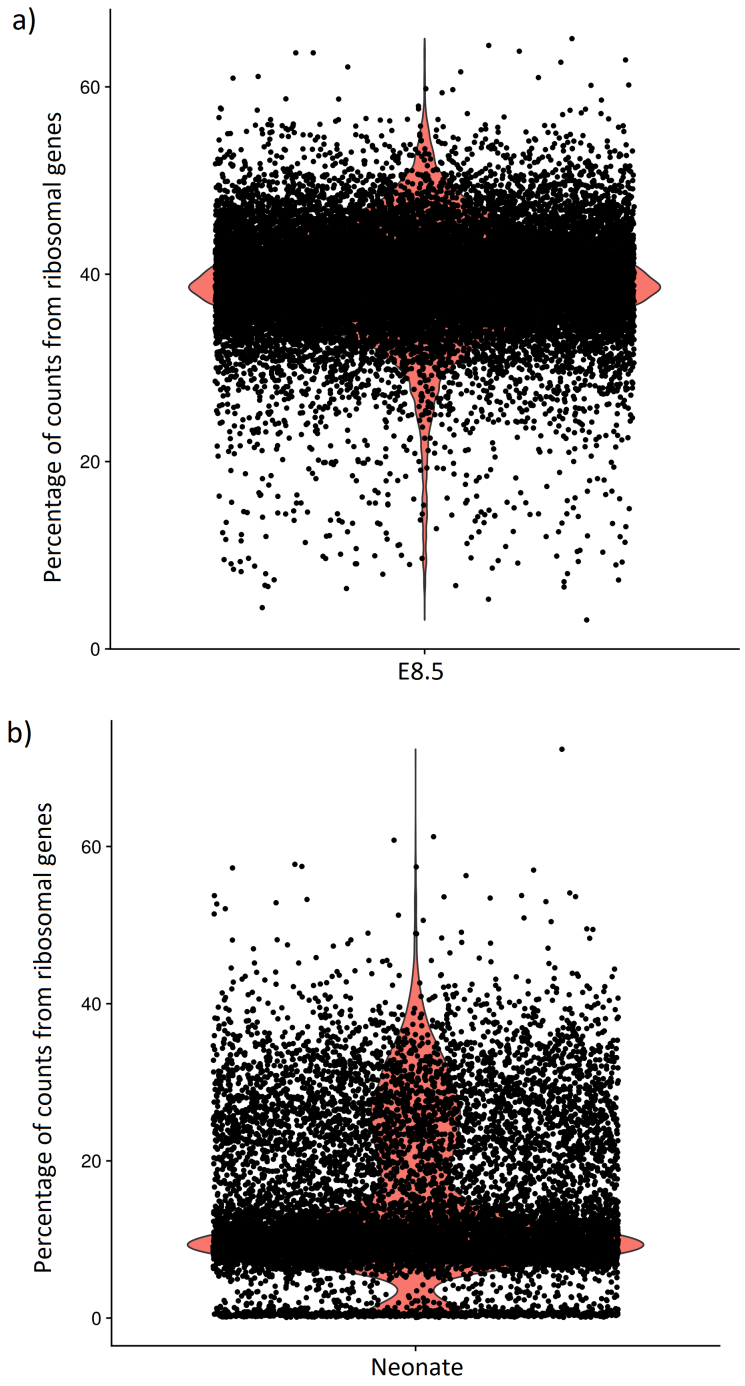


Figure 2.9: Percentage of RNA counts from ribosomal genes for E8.5 and neonate. a) E8.5. b) Neonate. The violin plots show the probability density of the data.

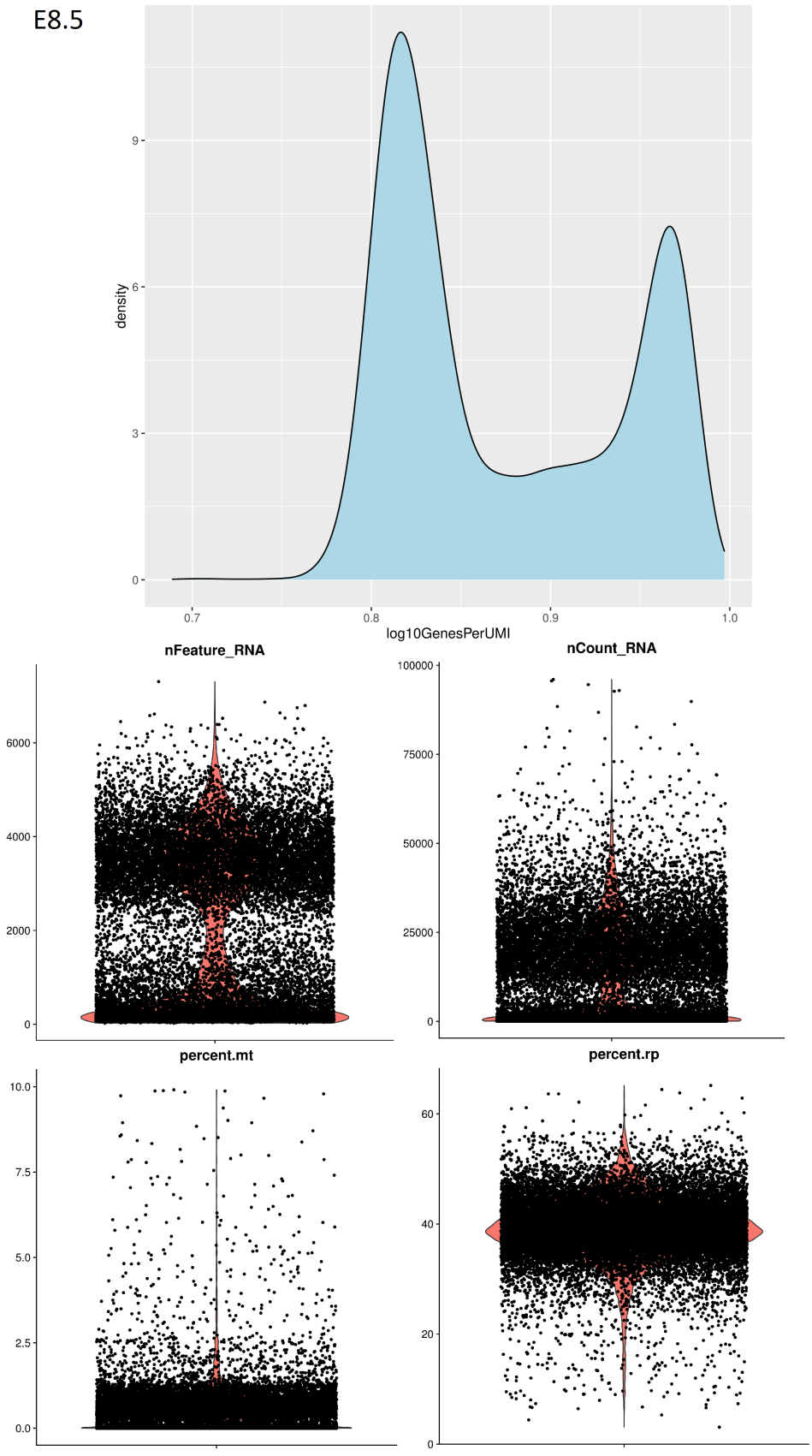


Figure 2.10: Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E8.5. Measure of transcriptomic complexity given by log10GenesPerUMI.

pendix Figure A.12). Ribosomal count percentages gradually decreased across the stages (Appendix Figures A.7 to A.12 and Figure 2.9) with increasing developmental time, starting at ~40% for E8.5 (Figure 2.9). The appendix contains the quality control plots for the remaining embryonic stages.

The RNA velocity streamline plot for all embryonic stages is shown in Figure 2.11, clustered by the Louvain algorithm into 31 clusters. The position, in the UMAP dimensionality reduced transcriptomic space, of each embryonic stage is shown in Figure 2.12. We can see that the right side (e.g. clusters 11, 29 and 15) contains early developmental cells and the left side (e.g. clusters 0, 5 and 1) contains mature cells. Overall, the flow of cellular differentiation is from right to left. This is confirmed by the RNA velocity inferred pseudotime plot (Figure 2.13a). Interestingly, the lower middle region (clusters 8 and 30) is at an earlier developmental stage than the upper middle region (clusters 7, 16 and 27). The RNA velocity length (Figure 2.13b) shows that the upper left arm (cluster 21) is one of the areas with the greatest rate of cellular differentiation. High velocity lengths were observed in the early developmental regions of clusters 15, 22 & 19 and cluster 3. In addition, cluster 30 was an area of rapid cellular differentiation, as well as clusters 1 and 4. The latter was an unexpected finding, given that these clusters were likely to be composed of mature, terminally differentiated cells. The confidence in the RNA velocities was relatively homogeneous across the clusters (Figure 2.13c).

The PAGA graph (Figure 2.13d), described earlier in Section 2.2, corroborates the finding that the overall flow of cellular differentiation is from right to left (e.g. cluster 12 to 22, cluster 15 to 19, cluster 27 to 7 and cluster 21 to 0). It also reveals connections that were not obvious from the earlier RNA velocity streamline plot (Figure 2.11), such as that from cluster 8 to 16, 15 and 3. We would hesitate to place too much importance on fine grain connections and more to overall coarse grain flows, given the many uncertainties and limitations of scRNA-seq and its

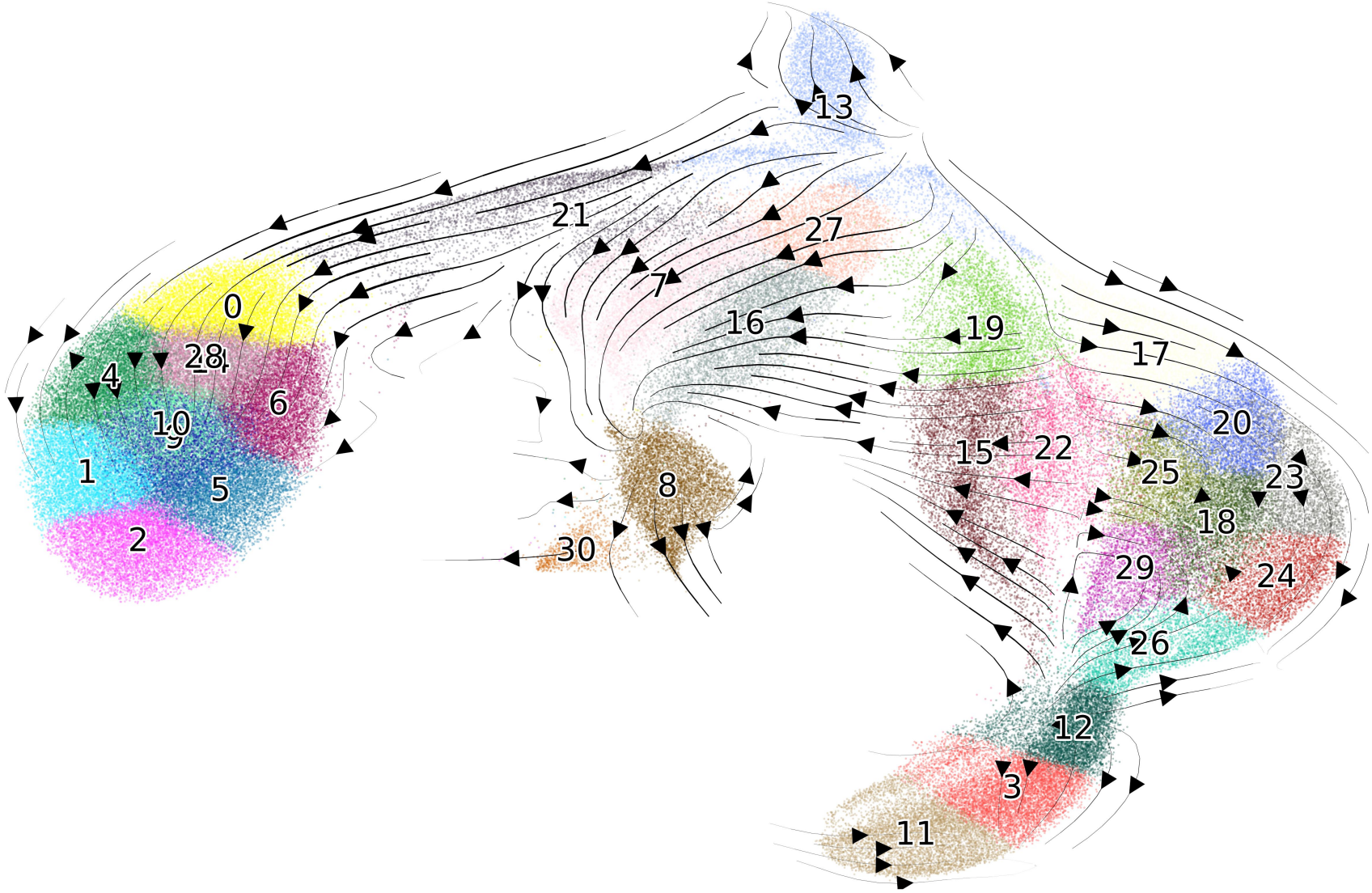


Figure 2.11: RNA velocity streamline plot. UMAP embedded. Louvain clustered into 31 clusters. All embryonic stages.

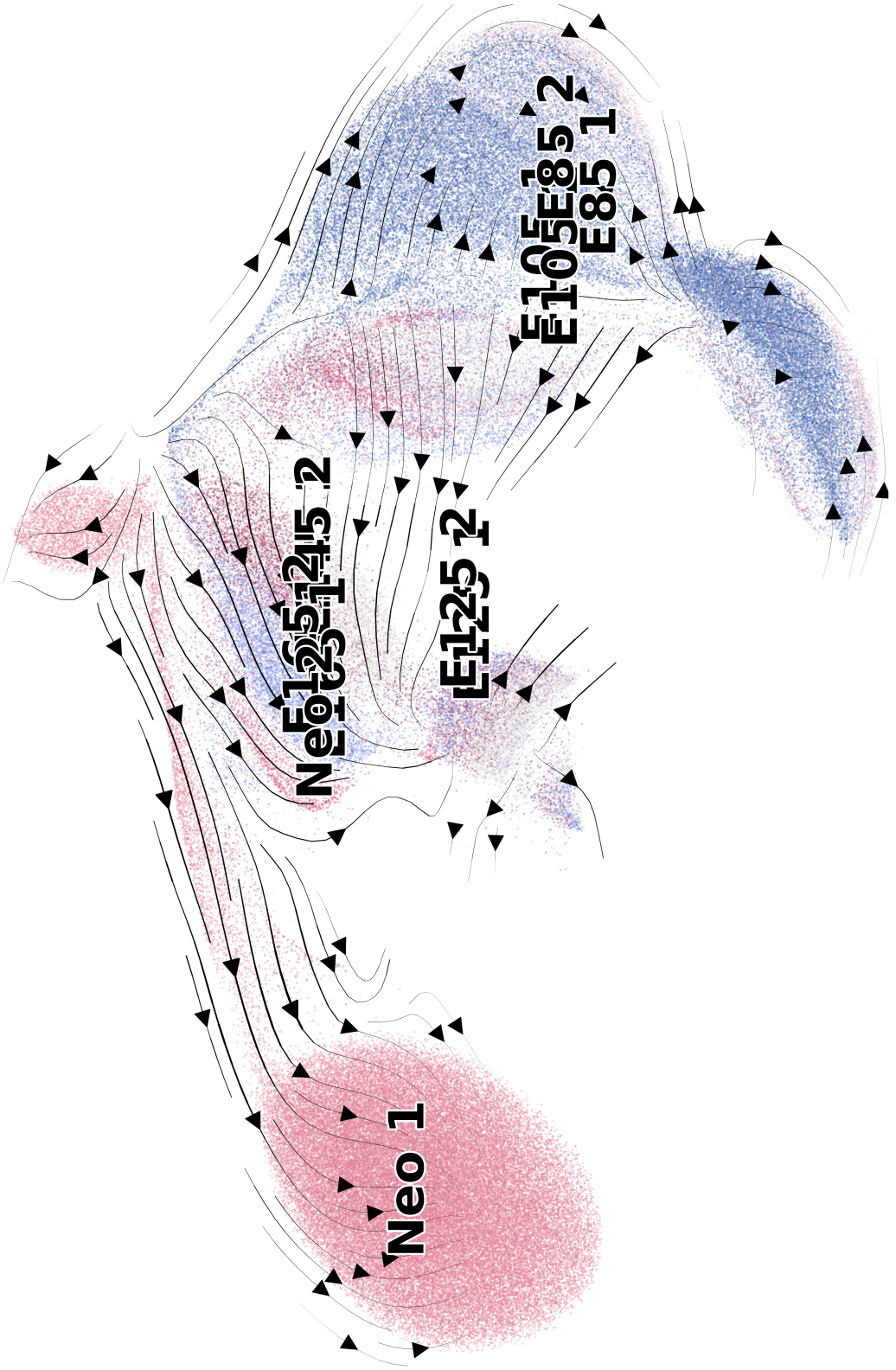


Figure 2.12: RNA velocity streamline plot by embryonic stage. Embryonic stage labels positioned at the centre of mass of each cluster of cells belonging to the given stage. Individual batches are labelled.

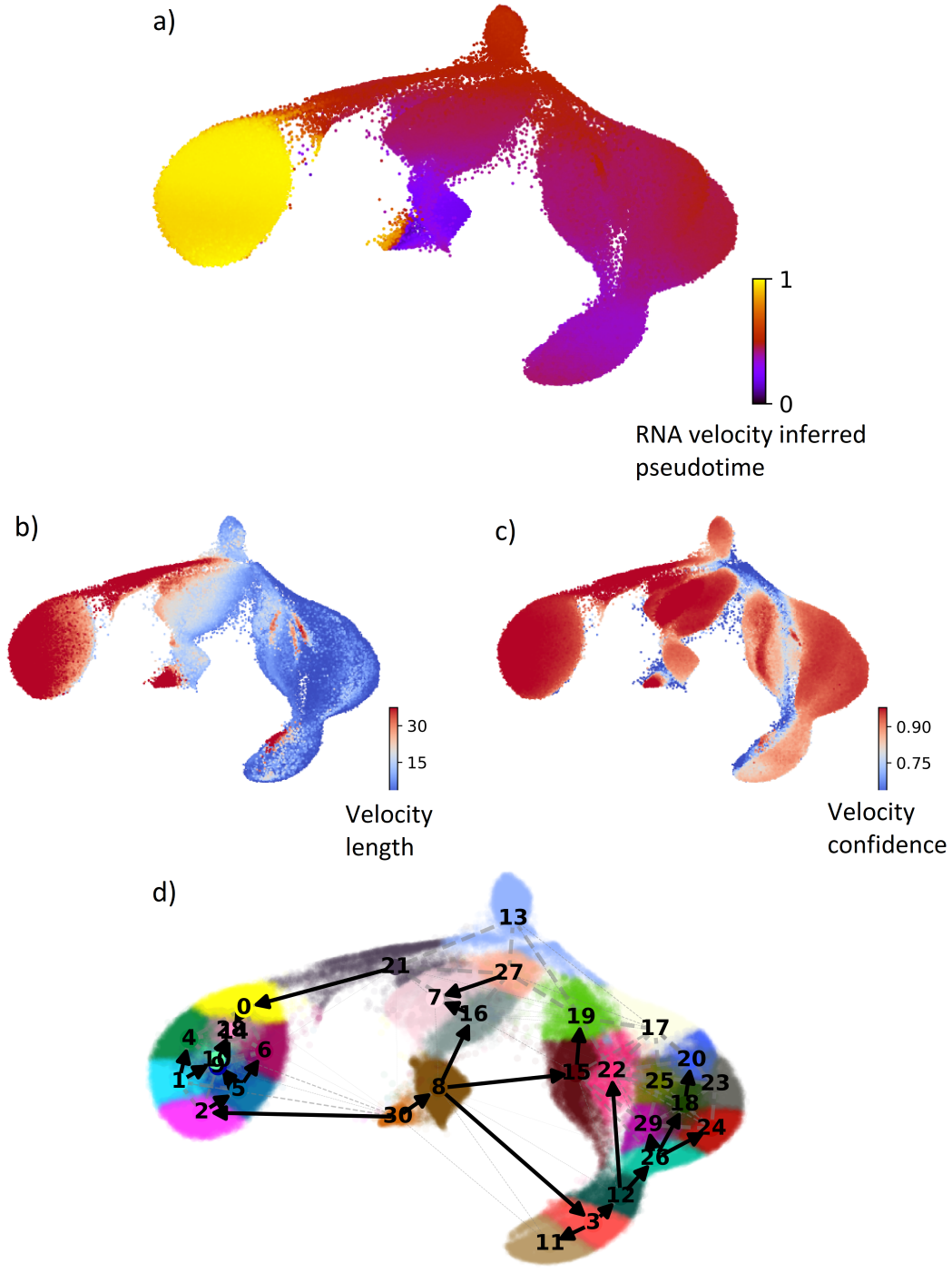


Figure 2.13: RNA velocity inferred pseudotime, velocity length, velocity confidence and PAGA.

post-processing (e.g. dimensionality reduction misinterpretation).

The top 25 differentially expressed genes per cluster, determined using t-tests on RNA abundance, is shown in Figures 2.14 and 2.15. The top 25 differentially expressed genes per cluster, determined using t-tests on RNA velocity, is shown in Figure 2.16; this particular differential expression (DE) analysis was conducted on a more coarse-grain, lower resolution, Louvain clustering, with 10 clusters.

As mentioned in Section 2.2, stacked violin plots of cluster by cluster gene expression were generated and analysed for 47 different lists of marker genes. Figure 2.17 shows one example, with a list of developmental marker genes. (Appendix Figure A.13 shows another stacked violin plot for ventricular and atrial marker genes—showing distinct atrial and ventricular clusters.)

Heatmaps with the top 50 RNA-velocity inferred differentially expressed genes for each of the 31 clusters were generated and analysed. Figure 2.18 shows the heatmap for cluster 9, showing high expression of Myl4 and Myl7, which are markers for atrial cardiomyocytes.

Key differentially expressed marker genes were identified from these various plots for each cluster and a putative biologically identity was assigned. The interpretation of the stacked violin plots, heatmaps and both RNA-count and RNA-velocity DE rankings, along with the putative biological identity of each cluster, are presented in Table 2.1.

We found difficulty in assigning simple biological identities to the early developmental clusters on the right hand side (19, 17, 29, 3, 24 etc.). It seemed that each cluster was closely related, transcriptionally, to each other (relative to, for example, mature atrial cardiomyocytes on the left hand side), and so the clusters appeared to be a mixture of various developmental cell types.

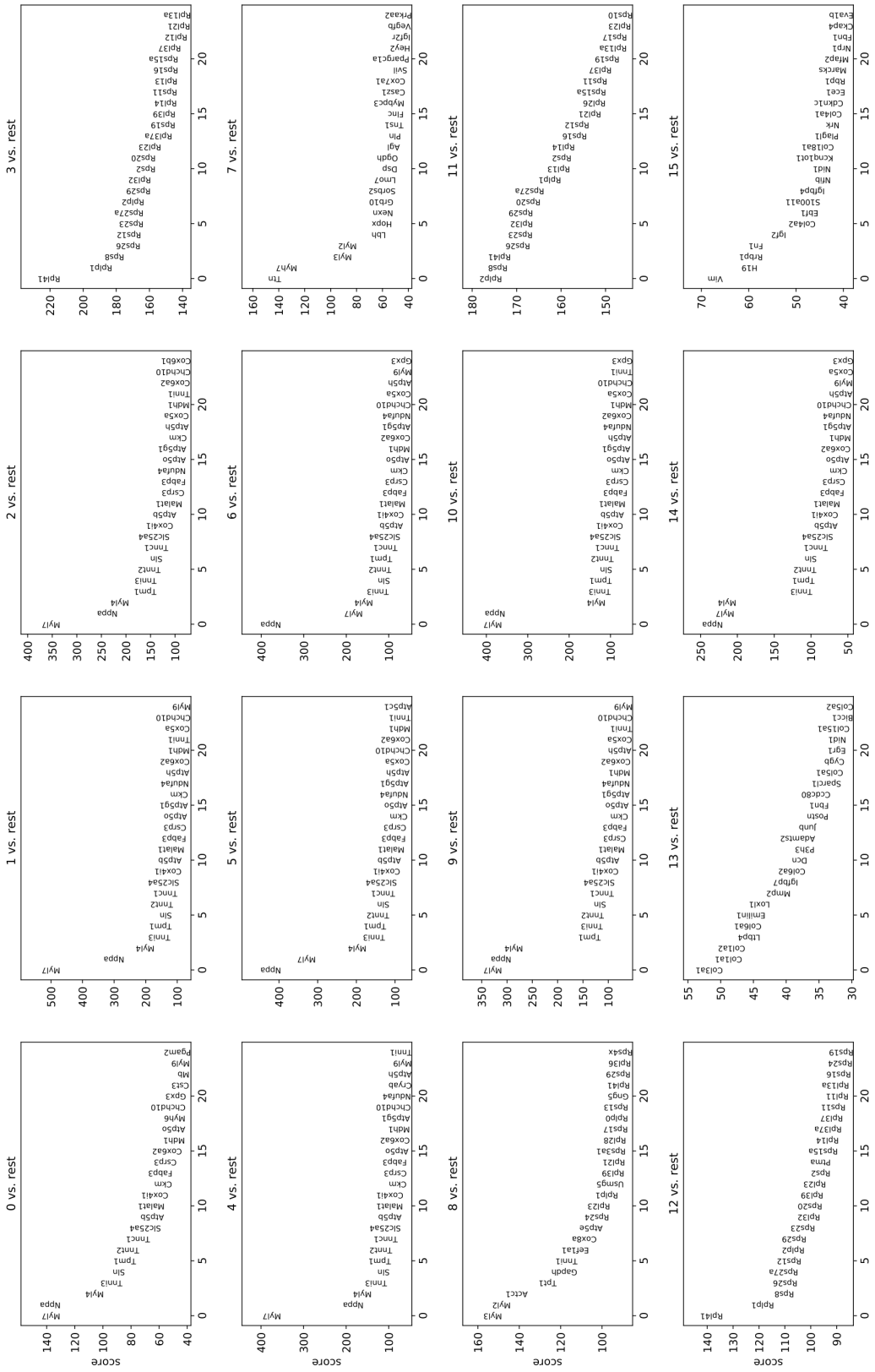


Figure 2.14: RNA count inferred top differentially expressed genes. Top 25 genes are shown for each cluster.

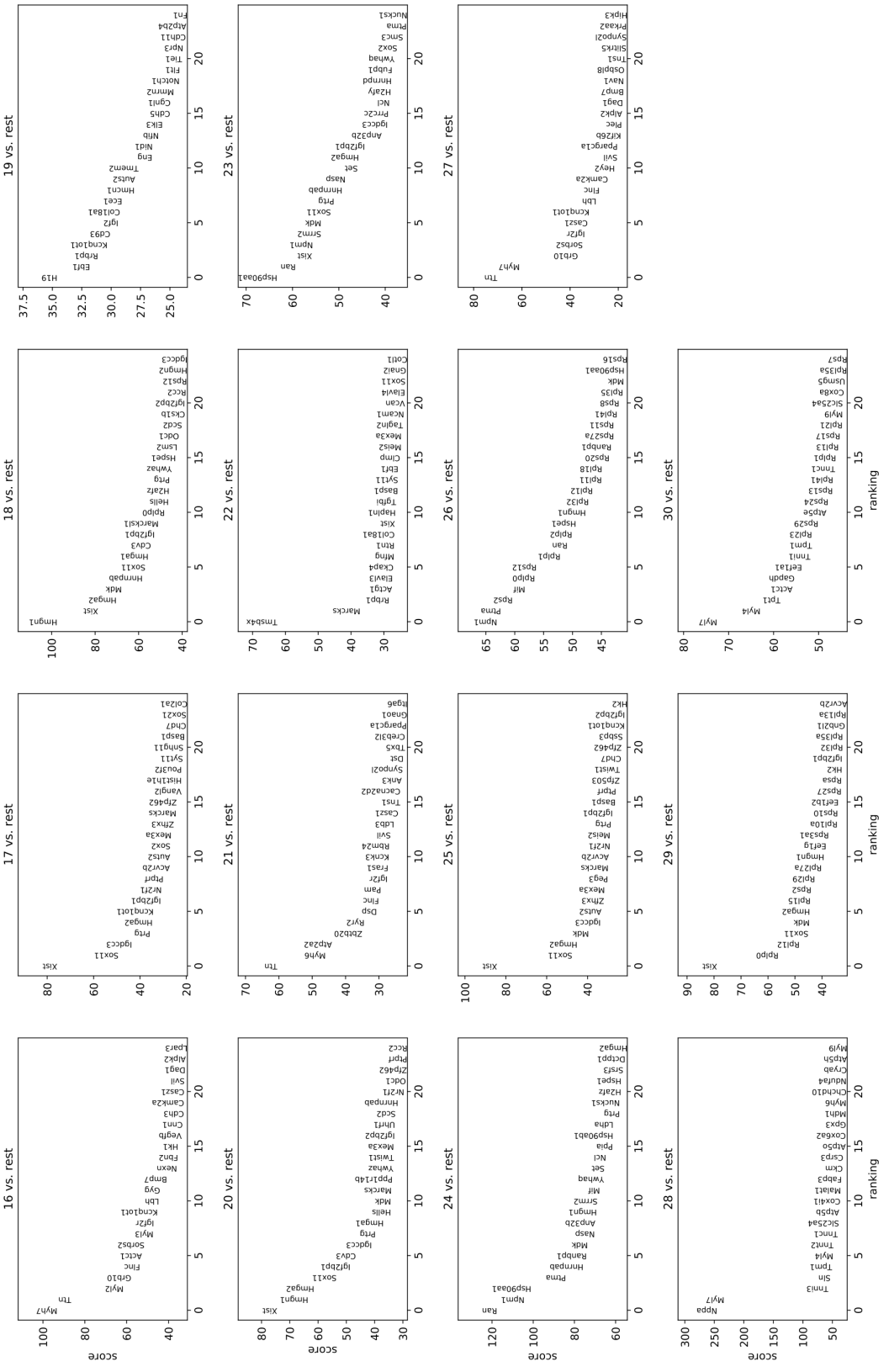
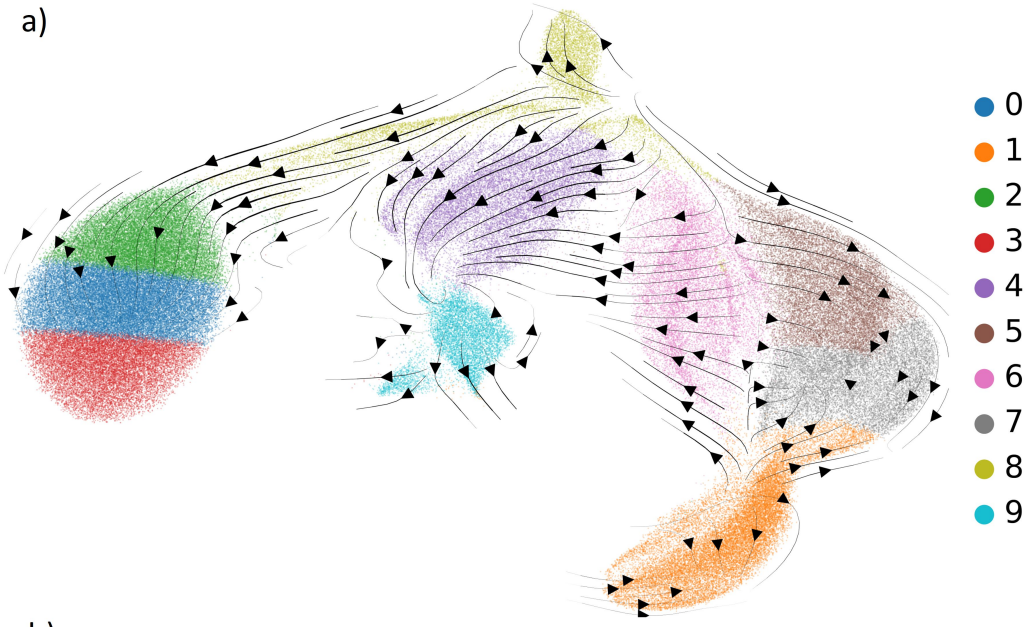


Figure 2.15: RNA count inferred top differentially expressed genes (continued).

a)



b)

Cluster

Gene Ranking	0	1	2	3	4	5	6	7	8	9
0	Myl7	Nnat	Trdn	Myl7	Casz1	Igf2bp1	Col18a1	Hmga2	Sptbn1	Mpped2
1	Myl4	Pantr1	Myl4	Myl4	Grb10	Hmga2	Elk3	Igf2bp1	Nfia	Myl3
2	Trdn	Hoxb5os	Atp5c1	Tnni3	Camk2a	Zfp462	Limch1	Hells	Zeb1	Gtsf1
3	Tnni3	Msi1	Tnni3	Myh6	Bves	Chd7	Nrk	Trim71	Rbms3	Fam178b
4	Atp5c1	Nasp	Myh6	Trdn	Nexn	Ssbp3	Srgap1	Elovl6	Nfkbiz	Snca
5	Myh6	Arg1	Myl7	Atp5c1	Samd4	Elovl6	Rhoj	Lin28b	Col1a2	Ddc
6	Atp2a2	Hmga2	Angpt1	Atp2a2	Lpar3	Igdcc3	Igf1	Greb1	Col5a2	Rapgef5
7	Stard10	Basp1	Atp2a2	Atcayos	Kif26b	Hells	Smad6	Fndc3c1	Ank2	Lpar3
8	Angpt1	Rfx4	Stard10	Angpt1	Myl2	Fndc3c1	Arhgap28	Chd7	Postn	Slc8a1
9	Trim55	Phf6	Ldb3	Kcnj3	Myl3	Zfhx3	Auts2	Plch1	Bicc1	Clvs1
10	Ldb3	Rcc1	Trim55	Rabac1	Lmo7	Zfhx4	Ltbp1	Prtg	Adamts2	Bves
11	Atcayos	Lix1	Atcayos	Postn	Ank3	Greb1	Mecom	Phf6	Fbn2	Trim12a
12	Chrm2	Pax6	Kcnj3	Trim55	Cdh2	Gli3	Rbms3	Cecr2	Sulf1	Sorbs2
13	Kcnj3	Lef1	Rabac1	Stard10	Agl	Phf6	Tspan18	Gli3	Creb5	Cd300lb
14	Rabac1	Slmo1	Chrm2	Pdlim3	Myocd	Trim71	Hapln1	Kif15	Tanc2	Cd74
15	Postn	Lin28b	Postn	Ldb3	Pde4dip	Basp1	Cgnl1	Zfp462	Palld	H19
16	Rabgap1l	Hoxa10	Nexn	Chrm2	Ddc	Ptch1	Tanc2	Nasp	Pcdh7	Robo3
17	Nexn	Crabp1	S100a6	Cdh2	Ank2	Cecr2	Col5a2	Ssbp3	Tgfbr2	Pacrg
18	Cd36	Fndc3c1	Cd36	Nexn	Akap6	Prom1	Fli1	Zfhx4	Pdgfra	Mir142hg
19	S100a6	Gap43	Rabgap1l	Rabgap1l	Sptbn1	Kif15	H19	Rcc1	Ebf2	Gm16867
20	Pdlim3	Kif21a	Nfia	Cd36	Myh7b	Prtg	Tshz2	Nr6a1	Tnfaip6	Slit2
21	Palld	Hoxa9	Etfb	S100a6	Prdm16	Top2a	Celf4	Lix1	Rhoj	Cubn
22	Cdh2	Apba2	Cdh2	Ccdc141	Ldb3	Lin28b	Atp2b4	Igdcc3	Pde1a	Spag16
23	Ccdc141	Pax3	Pdlim3	Nfia	Esrrg	Nhs	Zeb1	Pax3	Arhgap31	Meioc
24	Nfia	Hells	Sorbs2	Palld	Cdh3	Ephb2	Tmem2	Prom1	S100a6	Cacnb2

Figure 2.16: RNA velocity inferred top differentially expressed genes. Top 25 genes are shown for each cluster. The gene rankings (b) are based on the clustering shown in (a).

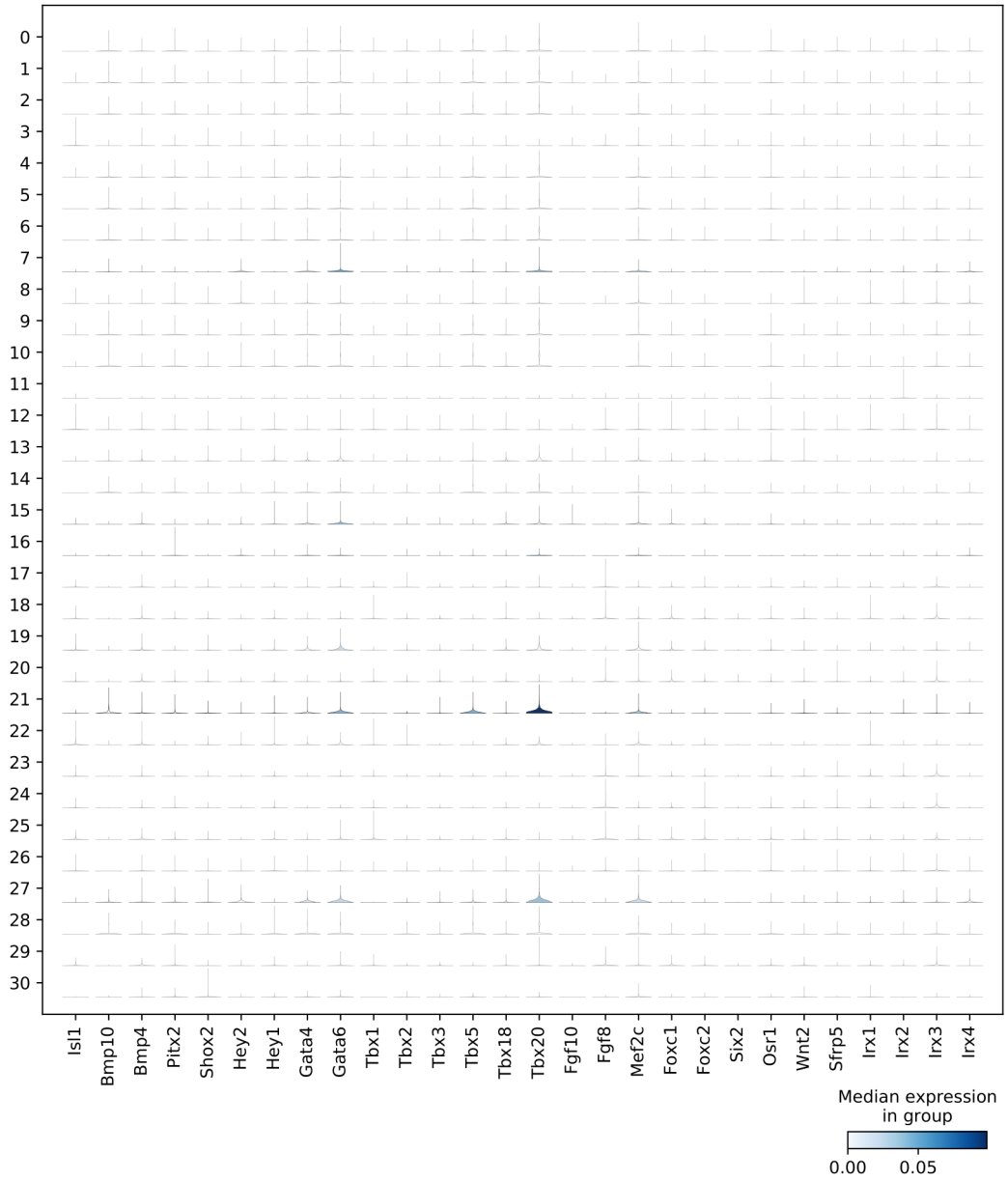


Figure 2.17: Developmental marker gene expressions by cluster. Shown using stacked violin plots—each violin plot shows the probability distribution of the mRNA counts for each cell.

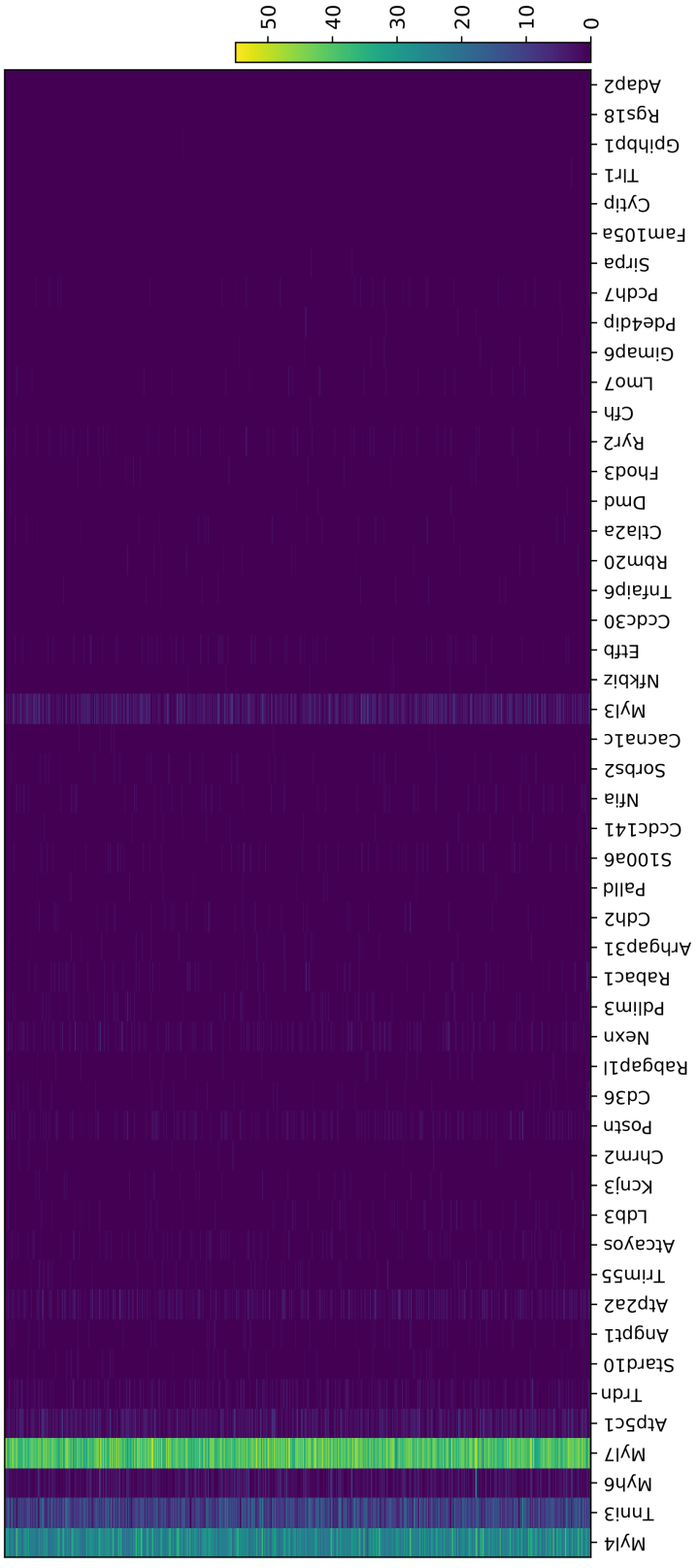


Figure 2.18: Gene expression heatmap for cluster nine. The top-50 RNA-velocity inferred differentially expressed genes are shown along the x-axis for cluster nine.

Table 2.1: Marker genes expressed in each cluster as determined by each of the four differential expression methods used. Cluster numbering corresponds with the numbering in Figure 2.11. The descriptor "medium" indicates lower expression of the marker gene. PGC = Primordial Germ Cell, P-SHF = Posterior Second Heart Field, A-SHF = Anterior Second Heart Field and FHF = First Heart Field.

Cluster	Putative identity	Markers expressed in stacked violin plots	Markers expressed in heatmaps	Markers expressed in RNA count inferred DE rankings	Markers expressed in RNA velocity inferred DE rankings
0	Atrial cardiomyocyte	Nppa (atrial cardiomyocyte), Bmp4 (mesenchyme), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte), Myl7 (cardiomyocyte) and Aldh1a2 (somitic mesoderm)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Myl7 (atrial cardiomyocyte) and Nppa (atrial cardiomyocyte)	Trdn (cardiomyocyte), Myl4 (atrial cardiomyocyte), Tnni3 (cardiomyocyte), Myh6 (atrial cardiomyocyte) and Myl7 (atrial cardiomyocyte)
2	Atrial cardiomyocyte	Mesp1 (cardiac mesoderm; nascent mesoderm), Nppa (pacemaker cell), Bmp4 (mesenchyme), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte), Myl7 (cardiomyocyte) and Aldh1a2 (somitic mesoderm)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Myl7 (atrial cardiomyocyte) and Nppa (atrial cardiomyocyte)	Myl7 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Tnni3 (cardiomyocyte) and Myh6 (atrial cardiomyocyte)
3	Mesoderm (P-SHF tendency) / notochord / neural tube	Mesp1 (cardiac mesoderm; nascent mesoderm), Isl1 (second heart field; pharyngeal mesoderm), Hes1 (neural stem cell), Aldh1a2 (somitic mesoderm), T (notochord), Foxa2 (notochord), Fgf8 (primordial germ cells; medium), Gsc (anterior primitive streak) and Otx2 (anterior primitive streak)	Nnat (neurofilament), Pax6 (neuronal), Crabp1 (neuronal; retinoic acid signalling), Tubb3 (neuronal), Isl1 (neuronal), Hoxa10 (developmental; right side of cardiac crescent), Tfap2a (neuronal) and Ttn (contractile)	Rpl41 (ribosomal) and Rplp1 (ribosomal)	Nnat (neurofilament), Pantr1 (neural progenitor), Hoxb5os (mesenchyme) and Msi1 (neural progenitor)
4	Atrial cardiomyocyte	Nppa (atrial cardiomyocyte), Bmp4 (mesenchymal), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte), Myl7 (cardiomyocyte) and Aldh1a2 (somitic mesoderm)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Myl7 (atrial cardiomyocyte) and Nppa (atrial cardiomyocyte)	Myl7 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte) and Trdn (cardiomyocyte)

5	Atrial cardiomyocyte	Nppa (atrial cardiomyocyte), Bmp4 (mesenchymal), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte), Myl7 (cardiomyocyte) and Aldh1a2 (somitic mesoderm)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Myl7 (atrial cardiomyocyte) and Nppa (atrial cardiomyocyte)	Myl7 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte) and Trdn (cardiomyocyte)
6	Atrial cardiomyocyte	Nppa (atrial cardiomyocyte), Bmp4 (mesenchymal), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte), Myl7 (cardiomyocyte) and Aldh1a2 (somitic mesoderm)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Myl7 (atrial cardiomyocyte) and Nppa (atrial cardiomyocyte)	Myl7 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte) and Trdn (cardiomyocyte)
7	Ventricular cardiomyocyte	Tbx18 (caval vein myocardium), Hopx (conduction), Bves (conduction; epicardium; epithelial cell of coronary vasculature), Gja1 (conduction), Igfbp5 (epicardial), Tbx20 (developmental; first and second heart fields; medium), Irx3 (rostral neurectoderm; medium) and Kdr (endothelium; medium)	Myl3 (ventricular cardiomyocyte), Myh7 (ventricular cardiomyocyte), Myl2 (ventricular cardiomyocyte) and Camk2a (cardiomyocyte)	Ttn (contractile) and Myh7 (ventricular cardiomyocyte)	Casz1 (cardiac development), Myl2 (ventricular cardiomyocyte), Myl3 (ventricular cardiomyocyte) and Nexn (cardiomyocyte)
8	Early ventricular cardiomyocyte	Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte) and Mesp1 (cardiac mesoderm; nascent mesoderm)	Myl3 (ventricular cardiomyocyte), Mpped2 (ventricular cardiomyocyte; less mature cardiomyocyte) and Tnnt1 (cardiomyocyte)	Myl3 (ventricular cardiomyocyte) and Myl2 (ventricular cardiomyocyte)	Mpped2 (early ventricular cardiomyocyte), Myl3 (ventricular cardiomyocyte) and Snca (synaptic terminals)
9	Atrial cardiomyocyte	Hcn4 (first heart field; pacemaker cell), Bmp4 (mesenchyme), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte), Myl7 (cardiomyocyte) and Aldh1a2 (somitic mesoderm)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Myl7 (atrial cardiomyocyte) and Nppa (atrial cardiomyocyte)	Myl7 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte) and Trdn (cardiomyocyte)
10	Atrial cardiomyocyte	Nppa (atrial cardiomyocyte), Bmp4 (mesenchyme), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte), Myl7 (cardiomyocyte) and Aldh1a2 (somitic mesoderm)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Myl7 (atrial cardiomyocyte) and Nppa (atrial cardiomyocyte)	Myl7 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte) and Trdn (cardiomyocyte)

11	Mesoderm (P-SHF tendency) / notochord / neural tube	Isl1 (second heart field; pharyngeal mesoderm), Fgf10 (anterior second heart field), Plat (parietal endoderm; medium), Hes1 (neural stem cell), Aldh1a2 (somitic mesoderm), Osr1 (intermediate mesoderm), T (notochord), Foxa2 (notochord), Fgf8 (primordial germ cell; medium), Gsc (anterior primitive streak) and Otx2 (anterior primitive streak)	Noto (notochord)	Rplp2 (ribosomal) and Rps8 (ribosomal)	Nnat (neurofilament), Pantr1 (neural progenitor), Hoxb5os (mesenchyme) and Msi1 (neural progenitor)
12	Mesoderm (FHF tendency) / notochord / neural tube	Tbx5 (first heart field), Isl1 (second heart field; pharyngeal mesoderm), Tbx1 (anterior second heart field; medium), Mef2c (anterior second heart field; medium), Plat (parietal endoderm), Hes1 (neural stem cell), Osr1 (intermediate mesoderm), T (notochord), Foxa2 (notochord), Fgf8 (primordial germ cell; medium), Gsc (anterior primitive streak), Mixl1 (anterior primitive streak) and Otx2 (anterior primitive streak)	Crabp1 (neuronal)	Rpl41 (ribosomal) and Rplp1 (ribosomal)	Nnat (neurofilament), Pantr1 (neural progenitor), Hoxb5os (mesenchyme) and Msi1 (neural progenitor)
13	Fibroblast	Mesp1 (cardiac mesoderm; nascent mesoderm), Col1a2 (fibroblast), Col3a1 (fibroblast), Col1a1 (fibroblast), Dcn (fibroblast), Cdh5 (endothelium; medium), Postn (myocardial), Apoe (myocardial), Cthrc1 (myocardial), Pdgfra (myocardial), Sox9 (myocardial), Fbln2 (myocardial), Ptn (endocardium), Pecam1 (endothelium), Tcf21 (pharyngeal mesoderm; fibroblast), Gsc (anterior primitive streak) and Otx2 (anterior primitive streak)	Postn (fibroblast), Col5a2 (fibroblast) and Tgfbr2 (fibroblast; required for Postn)	Col3a1 (fibroblast), Col1a1 (fibroblast) and Col1a2 (fibroblast)	Col1a2 (fibroblast), Col5a2 (fibroblast), Fbn2 (fibroblast), Zeb1 (mesenchymal) and Sptbn1 (fibroblast)
14	Atrial cardiomyocyte	Hcn4 (first heart field; pacemaker cell), Bmp4 (mesenchyme; medium), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte), Myl7 (cardiomyocyte) and Aldh1a2 (somitic mesoderm)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Myl7 (atrial cardiomyocyte) and Nppa (atrial cardiomyocyte)	Trdn (cardiomyocyte), Myl4 (atrial cardiomyocyte), Tnni3 (cardiomyocyte), Myh6 (atrial cardiomyocyte) and Myl7 (atrial cardiomyocyte)

15	Endothelial cell	Fgf10 (anterior second heart field), Mef2c (anterior second heart field; medium), Cav1 (endothelium), Gng11 (endothelium), Cdh5 (endothelium), Ecscr (endothelium), Plat (parietal endoderm), Ptn (endocardium), Pecam1 (endothelium) and Kdr (endothelium)	Eng (endothelial)	Vim (mesenchymal), H19 (developmental), Fn1 (developmental; fibroblast) and Col4a2 (fibroblast)	Col18a1 (fibroblast), Elk3 (developing vasculature; cranial mesenchyme; limb buds; neural crest) and Nrk (skeletal muscle)
16	Early ventricular cardiomyocyte	Pax6 (spinal cord; medium), Gypa (erythroid), Gata1 (blood progenitor; erythroid) and Kdr (endothelium)	Tnnt1 (cardiomyocyte) and Myocd (early cardiomyocyte)	Myh7 (ventricular cardiomyocyte), Ttn (contractile) and Myl2 (ventricular cardiomyocyte)	Casz1 (cardiac development), Myl2 (ventricular cardiomyocyte), Myl3 (ventricular cardiomyocyte) and Nexn (cardiomyocyte)
17	Notochord / mesoderm	Isl1 (second heart field; pharyngeal mesoderm; medium), T (notochord), Foxa2 (notochord) and Fgf8 (primordial germ cell; medium)	Nnat (neurofilament)	Sox11 (developing nervous system; neural; mesenchymal) and Prtg (developing nervous system)	Igf2bp1 (neural crest), Hmga2 (developmental) and Chd7 (neurectoderm)
18	Mesoderm (ASHF tendency) / PGC	Isl1 (second heart field; pharyngeal mesoderm; medium), Fgf8 (anterior second heart field; primordial germ cell), Tbx1 (anterior second heart field; paraxial mesoderm), Plat (parietal endoderm), Dkk1 (visceral endoderm), Aldh1a2 (somitic mesoderm), Osr1 (intermediate mesoderm), Mixl1 (anterior primitive streak) and Otx2 (anterior primitive streak)	Hells (developmental) and Trim71 (developmental)	Hmgn1 (developmental), Hmga2 (developmental) and Sox11 (developing nervous system; neural; mesenchymal)	Igf2bp1 (neural crest), Hmga2 (developmental), Hells (developmental) and Trim71 (developmental)
19	Mesenchymal	Dkk1 (visceral endoderm)	Tgfbr2 (fibroblast; mesenchymal), Col5a2 (fibroblast), Col18a1 (fibroblast), Nrk (developmental), Eng (endothelial), Piezo2 (mesenchymal) and Piezo1 (mesenchymal)	H19 (developmental), Ebf1 (lymphopoiesis), Rrbp1 (ribosomal), Kcnq1ot1 (developmental) and Col18a1 (fibroblast)	Col18a1 (fibroblast), Elk3 (developing vasculature; cranial mesenchyme; limb buds; neural crest) and Nrk (skeletal muscle)
20	Mesoderm (ASHF tendency) / neural crest / neural tube / notochord	Fgf8 (anterior second heart field; primordial germ cell), Tbx1 (anterior second heart field; medium), Mef2c (anterior second heart field), Sox10 (neural crest), Foxd3 (neural crest), Sox9 (neural crest), Irx3 (rostral neurectoderm), T (notochord) and Foxa2 (notochord)	Hells (developmental), Trim71 (developmental) and Nnat (neurofilament)	Hmgn1 (developmental), Hmga2 (developmental) and Sox11 (developing nervous system; neural; mesenchymal)	Igf2bp1 (neural crest), Hmga2 (developmental) and Chd7 (neurectoderm)
21	Atrial cardiomyocyte	Mesp1 (cardiac mesoderm; nascent mesoderm), Tbx20 (developmental; first and second heart fields) and Aldh1a2 (somitic mesoderm)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Ttn (contractile), Myh6 (atrial cardiomyocyte) and Atp2a2 (cardiomyocyte)	Col1a2 (fibroblast), Col5a2 (fibroblast), Fbn2 (fibroblast), Zeb1 (mesenchymal) and Sptbn1 (fibroblast)

22	Mesenchymal / endothelial / mesoderm	Isl1 (second heart field; medium) and Kdr (endothelium; medium)	Mecom (mesoderm), Tgfbi (mesenchymal; fibroblast), Ncam1 (mesenchymal), Piezo2 (mesenchymal), Isl1 (second heart field), Vcan (fibroblast) and Kdr (endothelial)	Tmsb4x (development; cardiac endothelial cell; cardiomyocyte) and Marcks (development; developing nervous system)	Col18a1 (fibroblast), Elk3 (developing vasculature; cranial mesenchyme; limb buds; neural crest) and Nrk (skeletal muscle)
23	Mesoderm (A-SHF tendency) / endothelial / notochord / neural tube / PGC	Mesp1 (cardiac mesoderm, nascent mesoderm), Isl1 (second heart field; medium), Fgf8 (anterior second heart field; primordial germ cell), Mef2c (anterior second heart field), Dkk1 (visceral endoderm), Pax6 (spinal cord), Sox2 (forebrain; midbrain; hindbrain), Cdh5 (endothelium), Ptn (endocardium), Pecam1 (endothelium), Aldh1a2 (somitic mesoderm), Osr1 (intermediate mesoderm), T (notochord), Foxa2 (notochord), Mixl1 (anterior primitive streak) and Otx2 (anterior primitive streak)	Map1b (neuronal development), Hells (developmental), Trim71 (developmental) and Nnat (neurofilament)	Hsp90aa1 (neurectoderm), Ran (neurectoderm; neural groove; neural fold; blood island), Sox11 (developing nervous system; neural; mesenchymal) and Prtg (developing nervous system)	Igf2bp1 (neural crest), Hmga2 (developmental), Hells (developmental) and Trim71 (developmental)
24	Mesoderm (A-SHF tendency) / endothelial / notochord / neural tube / PGC	Isl1 (second heart field; medium), Fgf8 (anterior second heart field; primordial germ cell), Dkk1 (visceral endoderm), Pax6 (spinal cord), Sox2 (forebrain; midbrain; hindbrain), Cdh5 (endothelium), Ptn (endocardium), Pecam1 (endothelium), Aldh1a2 (somitic mesoderm), Osr1 (intermediate mesoderm), T (notochord), Foxa2 (notochord) and Otx2 (anterior primitive streak)	Hells (developmental), Trim71 (developmental) and Nnat (neurofilament)	Hsp90aa1 (neurectoderm) and Ran (neurectoderm; neural groove; neural fold; blood island)	Igf2bp1 (neural crest), Hmga2 (developmental), Hells (developmental) and Trim71 (developmental)
25	Mesoderm (A-SHF) / PGC / notochord	Mesp1 (cardiac mesoderm; nascent mesoderm), Isl1 (second heart field; pharyngeal mesoderm; medium), Fgf8 (anterior second heart field; primordial germ cell), Tbx1 (anterior second heart field; paraxial mesoderm), Mef2c (anterior second heart field; medium), Igfbp5 (epicardial; medium), Dkk1 (visceral endoderm), Osr1 (intermediate mesoderm), T (notochord), Foxa2 (notochord) and Otx2 (anterior primitive streak)	Hells (developmental) and Trim71 (developmental)	Hmga2 (developmental), Sox11 (developing nervous system; neural; mesenchymal) and Mdk (neural progenitor)	Igf2bp1 (neural crest), Hmga2 (developmental) and Chd7 (neurectoderm)

26	Mesoderm (FHF tendency)	Tbx5 (first heart field), Isl1 (second heart field; pharyngeal mesoderm), Plat (parietal endoderm), Cdh5 (endothelium), Pecam1 (endothelium; medium), Dll1 (somitic mesoderm), Osr1 (intermediate mesoderm), Fgf8 (primordial germ cell; medium), Gsc (anterior primitive streak) and Otx2 (anterior primitive streak)	Hells (developmental), Trim71 (developmental) and Nnat (neurofilament)	Npm1 (forebrain and blood development), Ran (neurectoderm; neural groove; neural fold; blood island), Hmgn1 (developmental), Ptma (immune system), Rps2 (ribosomal), Mif (immune system; germ layer) and Rplp1 (developing nervous system)	Nnat (neurofilament), Pantr1 (neural progenitor), Hoxb5os (mesenchyme) and Msi1 (neural progenitor)
27	Ventricular cardiomyocyte	Tbx20 (developmental; first and second heart fields; medium) and Irx3 (rostral neurectoderm; medium)	Myh7 (ventricular cardiomyocyte), Tnnt1 (cardiomyocyte), Ttn (contractile) and Myl2 (ventricular cardiomyocyte)	Ttn (contractile) and Myh7 (ventricular cardiomyocyte)	Casz1 (cardiac development), Myl2 (ventricular cardiomyocyte), Myl3 (ventricular cardiomyocyte) and Nexn (cardiomyocyte)
28	Atrial cardiomyocyte	Nppa (atrial cardiomyocyte), Dkk1 (visceral endoderm), Bmp4 (mesenchyme), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte), Myl7 (cardiomyocyte), Aldh1a2 (somitic mesoderm) and Gsc (anterior primitive streak)	Myl7 (atrial cardiomyocyte), Myh6 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte), Kcnj3 (conduction; atrial cardiomyocyte), Stard10 (atrial), Cd36 (fatty acid uptake; maturation), Cacna1c (conduction), Pam (atrial), Atp2a2 (mature cardiomyocyte) and Kcnc2 (conduction)	Myl7 (atrial cardiomyocyte) and Nppa (atrial cardiomyocyte)	Trdn (cardiomyocyte), Myl4 (atrial cardiomyocyte), Tnni3 (cardiomyocyte), Myh6 (atrial cardiomyocyte) and Myl7 (atrial cardiomyocyte)
29	Notochord	Dkk1 (visceral endoderm), T (notochord), Foxa2 (notochord), Fgf8 (primordial germ cell; medium), Gsc (anterior primitive streak) and Otx2 (anterior primitive streak)	Hmga2 (developmental), Dlk1 (developmental), Hells (developmental) and Trim71 (developmental)	Hmga2 (developmental), Sox11 (developing nervous system; neural; mesenchymal) and Mdk (neural progenitor)	Igf2bp1 (neural crest), Hmga2 (developmental), Hells (developmental) and Trim71 (developmental)
30	Mixed atrial and ventricular cardiomyocytes / atrioventricular canal	Smpx (atrioventricular canal), Bmp4 (mesenchyme), Tnnt2 (cardiomyocyte), Myl4 (cardiomyocyte) and Myl7 (cardiomyocyte)	Ryr2 (cardiomyocyte), Kcnq5 (conduction), Rspo3 (atrioventricular canal; outflow tract), Myl2 (ventricular cardiomyocyte), Myl3 (ventricular cardiomyocyte) and Myh6 (atrial cardiomyocyte)	Myl7 (atrial cardiomyocyte), Myl4 (atrial cardiomyocyte) and Actc1 (cardiomyocyte)	Mpped2 (early ventricular cardiomyocyte), Myl3 (ventricular cardiomyocyte) and Snca (synaptic terminals)

Returning to our original aim of this experiment; we are interested in inferring the developmental origin or origins of the Dbh expressing cellular subpopulation. As described in Section 2.2, we can use the transition matrix generated during the analysis to step backwards from a given cell, along high cell-to-cell transition probabilities, to its ancestor cells. Figure 2.19 shows three examples of such "cell-tracing" lineage inference, using the 160,000th, 70,000th and 50,000th cell. Using the probable biological identities identified earlier and summarised in Table 2.1,

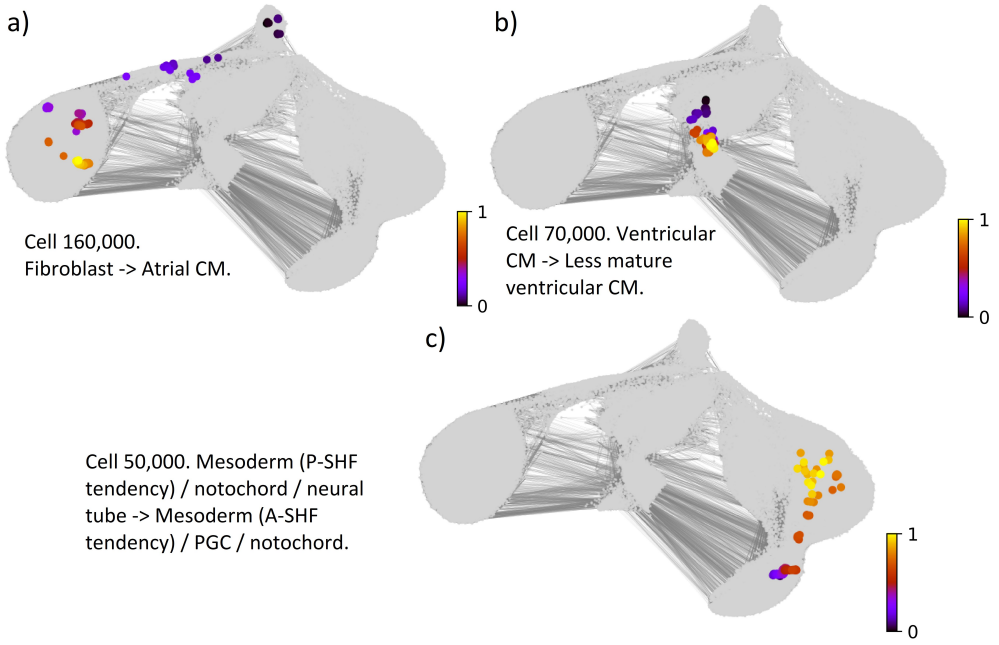


Figure 2.19: Cell lineage tracing using RNA velocity. The ancestors of each of three different starting cells (160,000th, 70,000th and 50,000th, all out of 169,000 total cells) are inferred by stepping backwards through pseudotime (darker: earlier, lighter: later).

the cell type and its ancestor cell type have been annotated.

Figure 2.20 shows the RNA velocity and RNA expression for the Dbh gene, and similarly for the Myl2 gene (a ventricular cardiomyocyte marker) with an additional phase portrait plot. We can see in Figure 2.20d that there is large RNA velocity (for this particular single gene, Myl2) in the "upper" ventricular cardiomyocyte region, indicating upregulation of this myosin protein, an essential component of the sarcomere, allowing contraction by sliding past the actin filaments, as described by the sliding filament theory. This suggests that the cells in this region are actively and rapidly differentiating into ventricular cardiomyocytes.

Figure 2.20e shows large RNA expression of Myl2 in the "lower" ventricular cardiomyocyte (CM) region. However when observing these plots for Myl2, a coher-

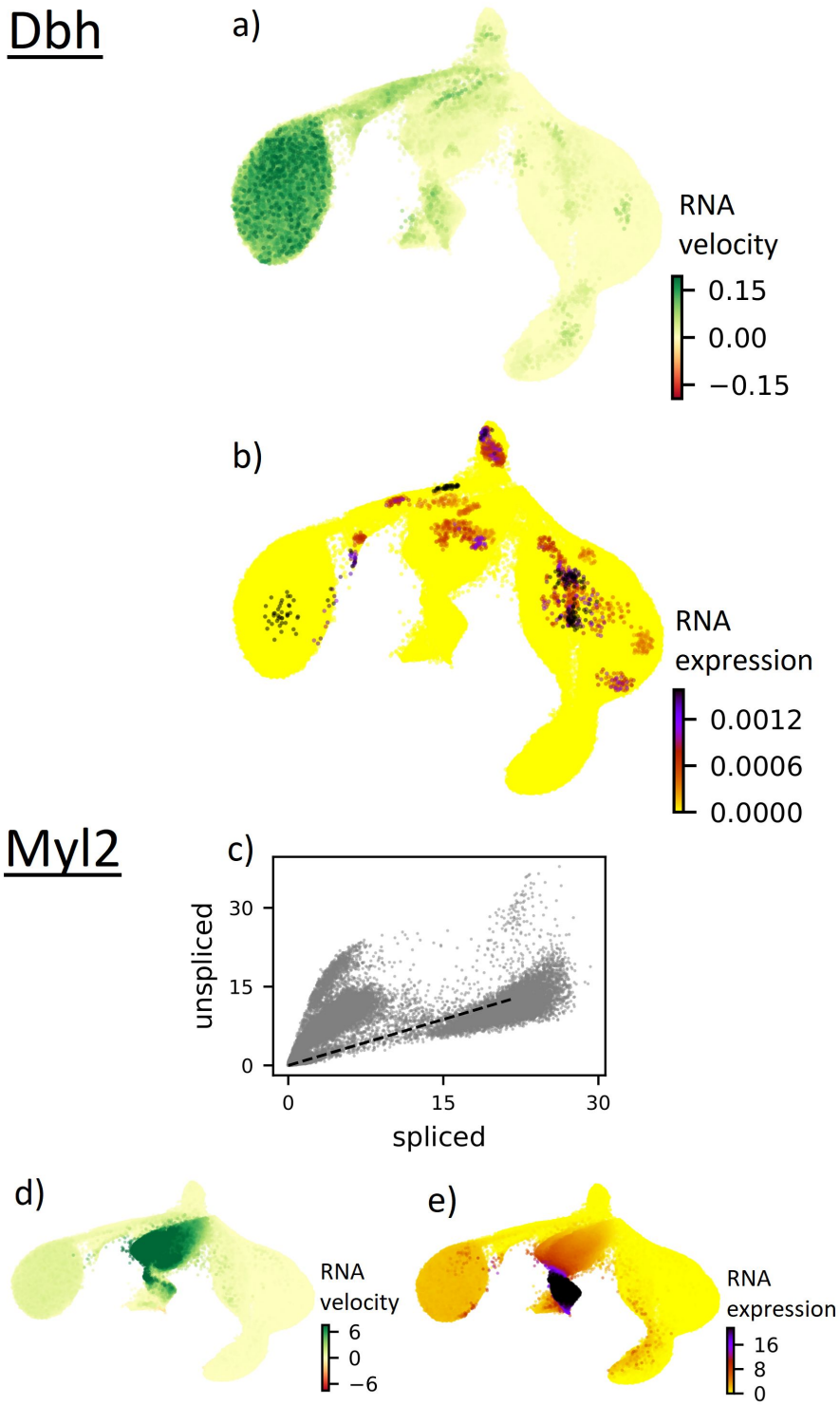


Figure 2.20: RNA velocity and expression profiles for Dbh and Myl2 genes with a Myl2 phase portrait. The phase portrait, (c), shows the unspliced vs spliced RNA counts with the steady state ratio shown as a dashed line.

ent biological interpretation would be that the "upper" ventricular CM region (high RNA velocity) would be early ventricular CMs (Myl2 gene induction and upregulation) and the "lower" ventricular CM region (high RNA expression) would be mature ventricular CMs (Myl2 gene steady-state transcription). This is however inconsistent with the RNA-velocity inferred pseudotime plot (Figure 2.13) and the DE analysis (Table 2.1) which suggest that the "upper" ventricular CM region is more mature than the "lower" ventricular CM region. An example of the unspliced versus spliced "phase portrait" plot, discussed in Section 2.2, is shown in Figure 2.20c for Myl2.

The cells that were upregulating Dbh the most (i.e. had the greatest Dbh RNA velocity) were the mature atrial cardiomyocytes (Figure 2.20a). The cells with the greatest levels of Dbh mRNA are shown in Figure 2.20b. In order of descending pseudotime, the identities of the clusters with greatest Dbh expression are: atrial cardiomyocyte (9, 6, 21), fibroblast (13), (mature) ventricular cardiomyocyte (7, 27), mesenchymal (19), mesoderm (A-SHF tendency) / endothelial / notochord / neural tube / PGC (23), notochord / mesoderm (17), mesoderm (A-SHF) / PGC / notochord (25), mesenchymal / endothelial / mesoderm (22) and mesoderm (FHF tendency) (26).

The earliest cells that expressed significant amounts of Dbh mRNA were in cluster 26. The heatmap with the top 50 RNA velocity inferred differentially expressed genes for cluster 26 is shown in Figure 2.21. From the heatmap genes, it is clear that the cells are of an early developmental type but it is difficult to give a precise developmental cell type. Notable genes expressed are Hells (a general developmental marker), Trim71 (another general developmental marker) and Nnat (a neurofilament marker). These are all non-specific to a particular embryonic cell type. Further interpretation of cluster 26 and conclusions as to the origin of the Dbh expressing subpopulation will be given at the end of Section 2.4.

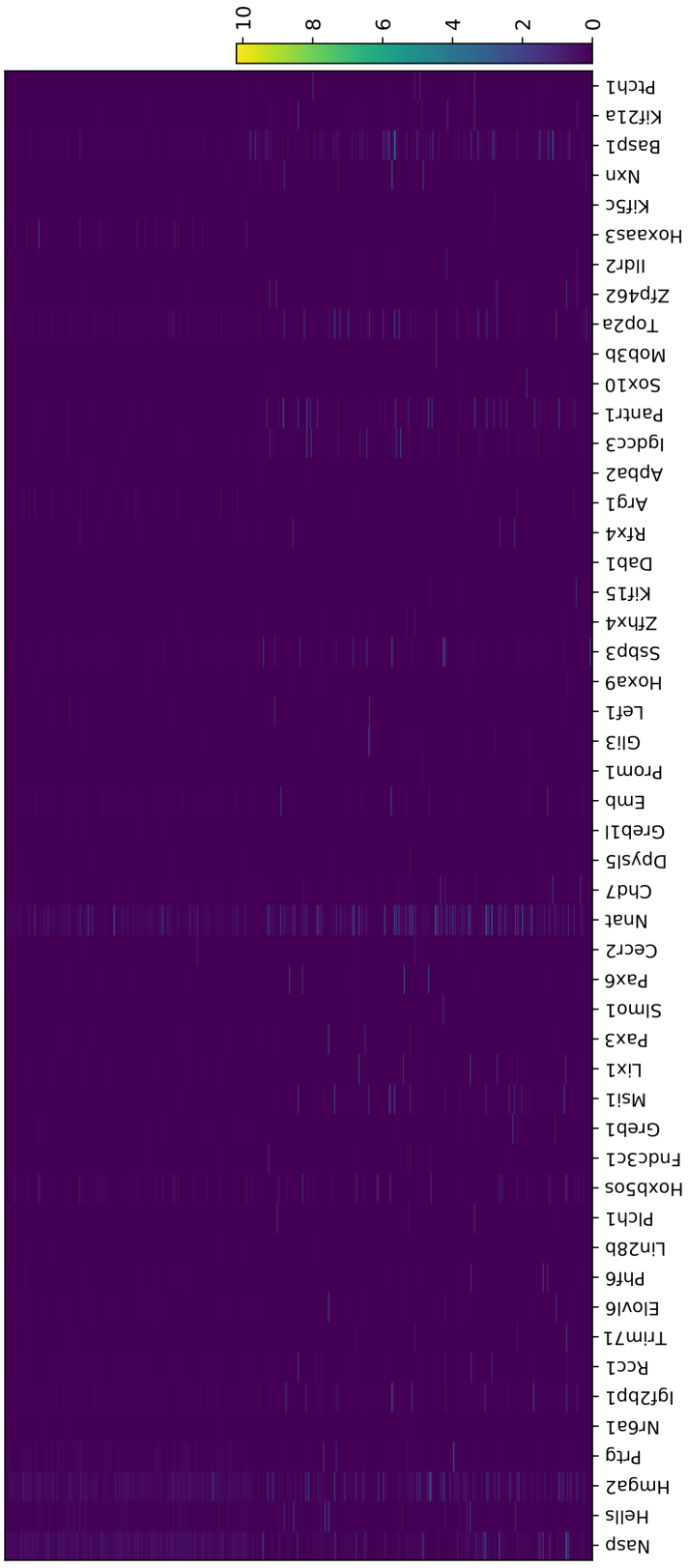


Figure 2.21: Gene expression heatmap for the speculated Dbh expressing subpopulation origin cluster. The top-50 RNA-velocity inferred differentially expressed genes are shown along the x-axis for this cluster.

2.4 Discussion

There were limitations to this investigation. The reason for and impact of the large proportion of unspliced counts relative to spliced counts is unclear. In addition, human judgement was used to set the parameters for filtering out cells deemed to be low quality. This was done by examining the quality control plots in combination with consideration of recommendations in the literature (Luecken & Theis, 2019). Very recent research has shown that it is possible that our filtering parameters, namely the mitochondrial mRNA ratio threshold of 7.5%, could lead to discrimination against pacemaker cells and cardiomyocytes (Galow et al., 2021). This work suggests that higher mitochondrial mRNA ratio thresholds should be used during quality control due to the higher metabolic requirements of cells in the heart relative to other tissues in the body. We took this partially into account—our mitochondrial mRNA ratio threshold is already above the standard in the scRNA-seq literature, but it may not have been high enough, with Galow *et al.* suggesting mitochondrial mRNA ratio thresholds of ~30%.

Quality control plots showed the two batches for each stage to be almost identical in transcriptomic space, hence no batch correction was performed on this data. However, one of the developmental stages, the neonatal stage, did show significant differences between the two batches. For example, in Figure 2.12, it can be seen that the two batches of each stage are localised in the same region of UMAP space, except for the two neonatal stages. This may have increased the chance of erroneous interpretations, particularly in the central region, where E16.5 and E14.5 cells are located.

We used droplet-based RNA sequencing as opposed to well-based sequencing. In droplet-based sequencing, more cells are sequenced (higher cell throughput),

but the number of transcripts detected per cell is lower (lower sequencing depth) than in well-based sequencing (Haque, Engel, Teichmann, & Lönnberg, 2017b). Conversely, in well-based sequencing, the cell throughput is lower but the sequencing depth is higher. It is therefore harder to detect rare genes with low numbers of transcripts with droplet-based sequencing. Missing genes due to this is termed gene dropout, and is a problem universal to all scRNA-seq methods (Qiu, 2020). A single-cell RNA sequencing method comparison study by the Enard lab placed Drop-Seq last place out of five methods for the number of genes detected (Ziegenhain et al., 2017). Thus, it is harder to detect cells based on a single gene such as the Dbh gene as was done in this project. We found that Dbh gene raw mRNA counts were low. Given the low sequencing depth, it is possible that our false negative rate was high, that is, cells that were actually expressing Dbh were not detected as such. It is therefore possible that our analysis has underestimated the number of Dbh expressing cells.

Single-cell RNA sequencing was chosen instead of bulk RNA sequencing. While this provides greater resolution, it also increases the amount of technical noise and biological variation from stochastic transcription (Chen, Ning, & Shi, 2019). An example of a technical error that can occur in droplet-based single-cell RNA sequencing (which was used in this project) is two cells can be captured in one microfluidic droplet. Each microfluidic droplet is assigned one barcode, corresponding to one cell. Therefore, the RNA from two cells is erroneously considered by the method as coming from one cell (doublet error). In addition to doublets and multiplets, empty droplets occur as well (Luecken & Theis, 2019). As a result, the count matrix features "cells" with zero mRNA counts and cells with double the usual amount of mRNA. With the latter, it becomes hard to discern this technical error from biological variation. Moreover, during the RNA sequencing process, some cDNA fragments do not acquire sequencing adaptors and so are not enriched

during the PCR amplifying stage, resulting in loss of mRNA transcripts in the data that were present in the cell (Starmer, 2017).

Another choice that was made was the dimensionality reduction algorithm. Examples of methods that can be used include t-Distributed Stochastic Neighbour Embedding (t-SNE) (van der Maaten & Hinton, 2008), Uniform Manifold Approximation and Projection (UMAP) (McInnes, Healy, & Melville, 2020) and Potential of Heat-diffusion for Affinity-based Trajectory Embedding (PHATE) (Moon et al., 2019). In our project we used UMAP. Both t-SNE and UMAP preserve local non-linear topological structure (distances within clusters) similarly, but UMAP preserves global non-linear topological structure (distances between clusters) better than t-SNE (Becht et al., 2019). UMAP was therefore used, and was closely supported by the RNA velocity package used (scVelo). PHATE is another dimensionality reduction algorithm that aims to improve the visualisation of the intrinsic structure of high-dimensional biological data. PHATE preserves local and global structure with less noise than t-SNE, as well as remaining computationally efficient when scaling to very large datasets (Moon et al., 2019). It also highlights trajectories and branches in the data as opposed to the "shattered" clusters of t-SNE (Moon et al., 2017). In future work, PHATE could be incorporated into scVelo's implementation of RNA velocity; this may have better elucidated the origins of the Dbh expressing subpopulation.

There were multiple choices as to the clustering algorithm. The Markov Cluster algorithm (van Dongen, 2000), based on simulation of stochastic flow in graphs; the Girvan-Newman algorithm (Girvan & Newman, 2002), a divisive form of hierarchical clustering based on link-betweeness and the Louvain algorithm (Blondel et al., 2008) which uses repeated local modularity optimisation and community aggregation, as described in Section 2.2.2. Personal testing suggested that the Markov Cluster algorithm resulted in the best community partitions compared to

Louvain and Girvan-Newman, with a highest community partition modularity of 0.87 compared to 0.81 and 0.80, respectively. However, Louvain was used in this project due to significantly improved runtimes compared to Markov Cluster and Girvan-Newman—Louvain had a runtime four orders of magnitude lower than Markov Cluster. This 10,000x runtime improvement was especially pertinent to this large data project which sequenced ~169,000 cells.

There are limitations to RNA velocity. Transcriptional boosts, where there is a sudden increase in transcription, can confuse the RNA velocity model and has been shown to lead to RNA velocities opposite to that biologically expected in erythroid maturation in mouse gastrulation (Bergen, Soldatov, Kharchenko, & Theis, 2021). Noisy gene dynamics, with no clear induction/repression, in fully mature cells can lead to false positive transition streamlines. This was seen in mature peripheral blood mononuclear cells, where RNA velocity provided an erroneously coherent vector field in one direction, when the biological expectation was a noisy vector field with no clear direction (Bergen et al., 2021). In addition time-variable kinetic parameters can affect gene activation curvature and can cause incorrect RNA velocity direction reversals (Bergen et al., 2021).

Especially in the neonatal stage, some of the mature cardiomyocytes that were being sequenced may have been too large to adequately flow through the microfluidics used. Our quality control measurements noted a large increase in low-quality cells in the neonatal stage (see Figure A.12 in the appendix) judged by the mitochondrial count ratio, in addition there was significant batch inconsistency. We speculate this might be due to overly large cardiomyocytes in the scRNA-seq microfluidic pipeline. Groups such as the Teichmann lab use single-nucleus RNA-sequencing (snRNA-seq) to ensure capture of large cardiomyocytes (Litviňuková et al., 2020). In the future, this could have been used for the neonatal stages, however snRNA-seq only uses nuclear RNA, losing all cytoplasmic RNA information

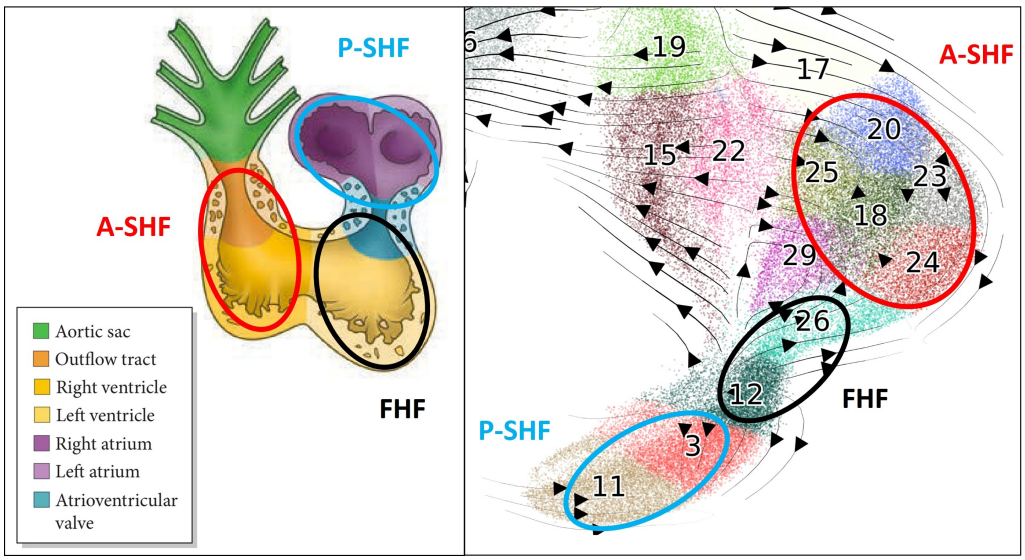


Figure 2.22: Transcriptomic topology recapitulates spatial heart field topology. Left: looping heart at human embryo day 28 (equivalent to the mouse at E10.5). The anatomical regions are coloured and heart fields labelled. Image source: (Gilbert et al., 2000). Right: lower right region of the RNA velocity streamline point with heart field contributions labelled.

and hence this stage of post-transcriptional regulation. Other options include specially designed microfluidic setups for larger molecules.

In terms of the results, we note how the transcriptomic topology of the posterior second heart field (clusters 3 and 11), first heart field (12 and 26) and anterior second heart field (18,20,23,24 and 25) seem to be in correspondence with the biological spatial shape of these embryonic heart fields *in vivo* during heart looping (biologically around E8.5 —which matches the transcriptomics as well). Namely they are contiguous and the first heart field is positioned in-between the second heart fields (see Figure 2.22).

We conclude by answering our initial question as to the origin of the Dbh expressing subpopulation. As mentioned earlier, in Section 2.3, cluster 26 appears to be the strongest candidate from this analysis. This cluster expresses Tbx5, a

marker for the first heart field. Though it does also express Isl1 (a second heart field marker), the other, labelled second, heart field clusters also express numerous SHF markers as well as Isl1, including Fgf8, Tbx1 and Mef2c, unlike cluster 26, and do not express FHF markers, unlike cluster 26. Therefore cluster 26 was judged to be the first heart field. This cluster expresses mesodermal markers such as Osr1 and Dll1. Confusingly, the cluster also expresses neurectoderm and neural progenitor markers such as Ran, Pantr1 and Msi1. The evidence in the literature for these marker associations was weaker though. The cluster identity is further confused by parietal endoderm and endothelial markers such as Plat and Pecam1. These two markers came up often in many of the clusters however, so these were judged to be non-specific markers in this case. In conclusion, the scRNA-seq analysis in this chapter suggests that the Dbh expressing subpopulation originated from first heart field splanchnic mesoderm, not neural crest as originally speculated.

3 | Electrophysiological properties: *Ex-vivo* whole-heart optogenetic pacing

3.1 Introduction

In this chapter we aim to explore the electrophysiological properties and potentially infer the spatial distribution of the Dbh+ derived cell subpopulation in the mouse heart. As mentioned in Chapter 1, we utilised a Dbh-Cre-ChR2/tdTomato mouse line. In this chapter, we describe how the hearts from freshly sacrificed mice were extracted and quickly cannulated and kept alive by connecting them to a perfusion system (i.e. *ex vivo* Langendorff-perfused hearts). The hearts were then subjected to blue light illumination. The selective expression of ChR2 only by Dbh+ derived cells allowed for selective photoactivation of only this subpopulation in the heart.

3.2 Materials and methods

3.2.1 Mouse line

The experiments in Chapter 3 and Chapter 4 utilised five mice from a $Dbh^{Cre/ChR2}$ optogenetic mouse line. This line was obtained by crossing the mouse strain Dbh-e(2A-iCre-WPRE-polyA) with the mouse strain, from The Jackson Laboratory, B6.Cg-Gt(ROSA)26Sor$^{tm27.1(CAG-COP4*H134R/tdTomato)Hze}$/J as shown in Figure 3.1. The mice were bred in the animal facility contained within the University of Oxford's Biomedical Services building. They were bred in individually ventilated cages. The details of each mouse is provided in Table 3.1. Four of the mice expressed ChR2-tdTomato; one mouse, Dbh3.1b, did not express ChR2-tdTomato

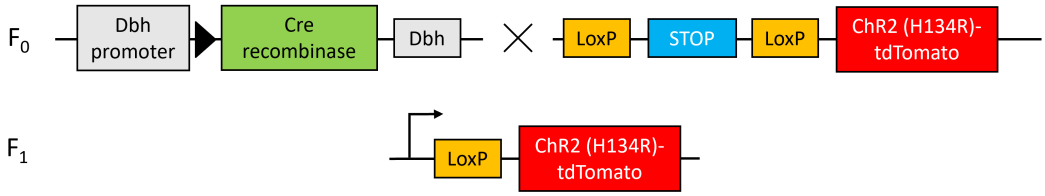


Figure 3.1: Dbh-Cre-ChR2/tdTomato mouse line schematic. In the parental generation, F_0, a Dbh-Cre mouse is crossed with a floxed ChR2/tdTomato mouse. In the resulting first filial generation, F_1, we obtain the genetic construct shown, which expresses ChR2/tdTomato, only in the cells that have ever expressed the Dbh gene (and their progeny).

and so acted as a control.

3.2.2 Reagents

Krebs-Henseleit (Krebs) solution is a physiological buffer used to maintain heart tissue. It was used as the perfusate in the Langendorff-perfusion. Standard Krebs solution consisted of 128 mM NaCl, 4.8 mM KCl, 21.98 mM $NaHCO_3$, 11 mM glucose, 1 mM $MgCl_2$ and 1.8 mM $CaCl_2$.

Ca^{2+}-free Krebs solution was used to block cardiac contraction while cannulating the heart. It is composed of the same solution as standard Krebs solution, as above, but without the 1.8 mM $CaCl_2$.

Avertin (tribromoethanol) is an anaesthetic used before the sacrificing of the mice.

Table 3.1: Genotypes of each mouse.

Mouse Name	Age (weeks)	Sex	iCre	Dbh	ChR2/tdTomato	ROSA
Dbh1.1a	6.6	Male	+	+	+	+
Dbh1.1b	6.6	Male	+	+	+	+
Dbh1.1c	6.6	Male	+	+	+	+
Dbh3.1b	8.6	Male	-	+	+	+
M01614568	26.7	Male	+	+	+	+

1.2% solution was prepared as follows. Avertin stock solution was made by mixing 25 grams of Avertin (tribromoethanol) with 15.5 ml tert-Amyl Alcohol for 12 hours in a dark bottle at room temperature. The compound is photosensitive so it was stored in the dark. Next, 0.5 ml of Averin stock solution was mixed with 39.5 ml 0.9% saline (NaCl) solution. This 1.2% solution was stored in a dark container at 4°C.

Heparin is an anticoagulant used to reduce the likelihood of myocardial infarction during the experiment. 300mg of solid > 180 U/mg heparin was added to 30 ml distilled water giving a > 1800 U/ml heparin solution, stored at 4°C.

Isoprenaline is a beta adrenergic receptor agonist circulated in the Langendorff-perfusion at a certain point in the experiment. Isoprenaline is light sensitive so it was stored in an opaque container in ice. A concentration of 10 mM isoprenaline stock solution was made by adding 1.2 mg isoprenaline hydroxide and 5.4 mg L-ascorbic acid to 0.4844 ml distilled water.

Phosphate-buffered saline (PBS) was made by adding PBS tablets (Sigma-Aldrich) each to 200 ml of distilled water.

3.2.3 Langendorff-perfused *ex-vivo* whole heart electrophysiology

3.2.3.1 *Ex-vivo* isolation of murine Dbh-Cre-ChR2/tdTomato hearts

All of the mice were given 0.3 ml of 1.2% Avertin (tribromoethanol), an anaesthetic, and 0.3 ml of 1800 U/ml heparin, an anticoagulant, by intraperitoneal injection. After 5 minutes, they were sacrificed by cervical dislocation in accordance with the Animals Scientific Procedures Act (1986). Whole hearts from the mice were isolated and placed in 10 ml of Ca^{2+}-free Krebs solution containing 0.1 ml

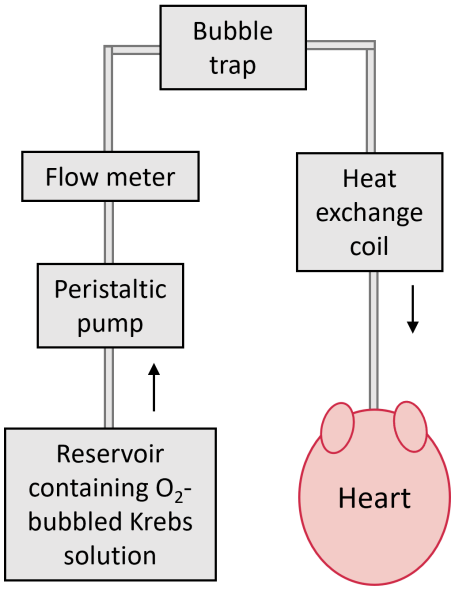


Figure 3.2: Langendorff-perfusion experimental setup. Oxygen bubbled Krebs solution is pumped through a flow meter, bubble trap and heat exchange coil before entering the coronary arteries of the heart.

1800 U/ml heparin. The Ca^{2+}-free Krebs solution arrests heart contraction. Each heart is placed in approximately 100 ml of ice-cold Ca^{2+}-free Krebs solution while the aorta is cannulated under a microscope and held in place with a knot. The cannulation process was done in under approximately 7 minutes to limit ischaemic injury of heart tissue.

3.2.3.2 Langendorff-perfusion experimental setup

The experimental setup of the Langendorff-perfusion is as shown in Figure 3.2. A reservoir containing O_2-bubbled Krebs solution is pumped by a peristaltic pump at a constant flow rate of 3.5–4.0 ml/min as measured by the flow meter. The perfusate is then pumped through a bubble trap to remove air bubbles; air bubbles that enter the coronary arteries of the heart can cause myocardial ischaemia and resulting death of heart tissue. The perfusate is then brought to a physiological temperature of 37°C by pumping it through a heat exchange coil. The

oxygenated, 37°C, Krebs solution then passes through the cannula, into the aorta (retrograde perfusion), through the coronary ostia into the coronary arteries of the heart, maintaining the requirements of the heart tissue in the absence of blood.

Before mouse heart cannulation, the Langendorff-perfusion system was drained and then flushed first with distilled water and then flushed with Krebs solution. The system was checked for bubbles and any that were found were removed. After mouse heart cannulation, each mouse heart was attached to the Langendorff-perfusion setup and began to beat again. Two electrodes were positioned on the left atrium and apex of the heart; this was connected to a computer, running Spike 2 software (Cambridge Electronic Design). This provided an electrocardiogram (ECG) of the mouse heart, approximately corresponding to lead II of the human ECG, allowing the electrophysiological parameters, RR interval, PR interval, QRS duration, QT interval and P-wave duration, to be measured. A blue light (Cairn Research) was directed toward and positioned approximately 2.5 cm away from the right atrium of the mouse heart. Figure 3.5, in Section 3.3, shows this setup. Using a 1401 digitiser and Spike 2 software, the blue light was triggered to deliver 470 nm blue light pulses. The setup allowed the pulse duration (i.e. the time that the light was on during each pulse), current (corresponding with light intensity) and pulse cycle length (time between pulses) to be varied during the experiment. The light intensity was 0.6–$0.8\,\mathrm{mW\,mm^{-2}}$.

3.2.3.3 Light stimulation protocol

The Langendorff-perfused *ex-vivo* hearts were subjected to programmed light stimulation protocols. Two protocols were used: (1) an S1S1 pacing protocol where a train of pulses is delivered at a fixed cycle length, and (2) an S1S2 pacing protocol where a pacing train of eight pulses (S1s) at a fixed cycle length precedes a final (S2) pulse that is delivered at progressively shorter intervals to the train of

S1 pulses. In the S1S2 protocol, the time interval between S1 and S2 that occurs directly before that which fails to produce a ventricular complex on the ECG is recorded as the refractory period.

The experimental protocol for each mouse heart was as follows. The light was positioned over the right atrium. S1S1 pulse trains were delivered with the pulse duration and cycle length fixed at 10 ms and 100 ms respectively, but with the current varied for each pulse train to identify the minimum, threshold, current required to pace the heart with one-to-one correspondence between individual pulses and ventricular complexes (capture), throughout the entire pulse train, as seen on the ECG. Once the threshold current had been established, the S1S1 pulse train was delivered at twice the threshold current, 100 ms cycle length, but with the pulse duration varied. This allowed the threshold pulse duration, i.e. that which allowed capture throughout the stimulus train, to be established. Then, the S1S2 protocol was delivered at 100 ms cycle length, twice the threshold current and twice the threshold pulse duration. This, as described above, allowed the refractory period to be determined. Next, the blue light was shifted to the next chamber of the heart, the right ventricle, and the process above was repeated for this chamber of the heart. This was then repeated for the left atrium and left ventricle.

Videos, including slow-motion capture (0.25x real-time), were taken of the optogenetic light pacing for mouse heart M01614568. They are available at `https://youtu.be/5w3_SZXGrag` or upon request.

Optical fibres were inserted inside the heart by cutting the atria off and inserting the fibres through the opening during Langendorff-perfusion but no successful endocardial optogenetic pacing was achieved using a light current of 1999 mA and 10 ms pulse duration.

3.2.3.4 Administration of isoprenaline

At an earlier point, separately to the experiment, distilled water was flushed through the perfusion system before Krebs solution was flushed through the system. Then food colouring was added to the reservoir flask. The time taken for the food colouring to travel from the reservoir flask to the cannula was measured to be 6 minutes and 30 seconds. This was used as an approximation for the time taken for isoprenaline to reach the heart after adding it to the reservoir flask.

After the electrophysiolgical experiment, described above, was completed under standard Krebs solution conditions, 250 µl of 100 µM isoprenaline was added to the 250 ml Krebs reservoir solution, producing a final concentration of 100 nM isoprenaline in Krebs perfusate. After 6 minutes and 30 seconds, the same experimental protocol, as described earlier, was conducted in the isoprenaline-stimulated condition.

3.2.3.5 ECG and statistical analysis methodology

After concluding the electrophysiological experiments, the ECG data was analysed using Spike 2 software (Cambridge Electronic Design). The RR interval, PR interval, QRS duration, QT interval and P-wave duration were measured before pacing, defined as at least, and approximately, three seconds before the pulse train; during pacing, defined as at least, and approximately two seconds into the pulse train, and after pacing, defined as at least, and approximately, three seconds after the pulse train. The heart-rate corrected QT interval, QTc, was calculated using Bazett's formula (Bazett, 1920) from the RR interval and the QT interval as:

$$QT_c = \frac{QT}{\sqrt{\frac{RR}{1s}}}, \tag{3.1}$$

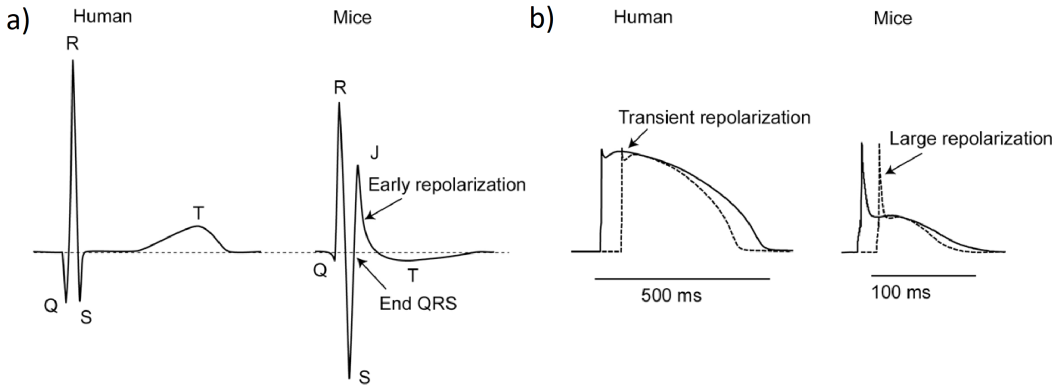


Figure 3.3: Electrophysiology of the mouse compared to the human. a) Differences in the mouse ECG compared to that of the human. b) Differences in the ventricular cardiac action potential in the mouse compared to that of the human. Adapted from (Boukens et al., 2014).

where the intervals, QT and RR, are in seconds. The heart rate (HR), in beats per minute, was calculated as $HR = 60/RR$, with the RR interval in seconds. These parameters were measured for each chamber of the heart and for each of the five mice. The threshold current, threshold pulse duration and refractory period, as described previously, was analysed per chamber of the heart for each of the five mice. This and the ECG parameter measurements were repeated in the 100 nM isoprenaline condition.

The murine ECG differs from that of the human ECG in that there is ST-elevation (J-wave) and an inverted T-wave (Boukens et al., 2014), as shown in Figure 3.3. With this noted, the methodology used to measure the ECG intervals is shown in Figure 3.4 and the ECG interpretation is discussed further in Section 3.4.

All statistical tests were carried out in Microsoft ® Excel ® (Microsoft 365 MSO (16.0.14326.20062) 32-bit version, Excel version 2108, 2021).

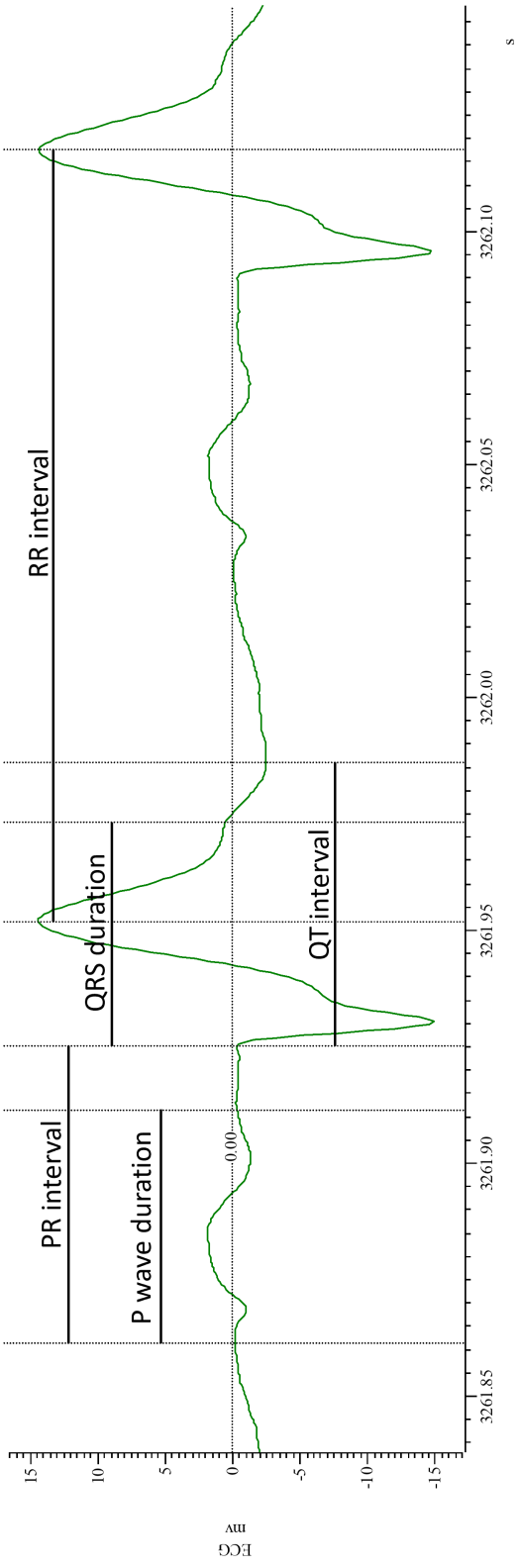


Figure 3.4: ECG measurement methodology. Trace recorded from Langendorff-perfused mouse heart M01614568 without light illumination.

3.3 Results

Figure 3.5 shows the ECG trace during S1S1 light stimulation of the right ventricle (RV) of mouse M01614568 (an optogenetic Cre+ mouse, see Table 3.1) using a cycle length of 100 ms. We can see the light stimulation successfully increased the heart rate from ~325 bpm to ~600 bpm during pacing. The heart was therefore paced one-to-one, since a 100 ms light cycle length corresponds to 600 bpm. After pacing, the heart produced two slow beats before returning to the normal sinus rhythm, as before pacing.

Figure 3.6 shows the ECG trace for the same setup as in Figure 3.5 but for the left ventricle (LV). There was no response to light stimulation. The ECG trace while stimulating the left ventricle for a different heart, that of mouse Dbh1.1b, is shown in Figure 3.7. Here there was a partial response to the S1S1 light stimulation train, with 1-to-1 capture for the first ~500 ms, before a reversion to sinus rhythm.

Table 3.2 summarises the response to light stimulation of each chamber, by mouse. Dbh3.1b was a control mouse, with a Cre negative genotype, and so did not express ChR2. In particular, we note the poor response by the LV to light stimulation.

Figure 3.8 shows the electrophysiological intervals measured, before, during and after pacing, for the right atrium in the control mouse and for the same chamber in the optogenetic mice. For the control mouse, single-factor ANOVA tests showed that there was no significant difference before, during or after pacing for RR ($p \approx 0.808$), PR ($p \approx 0.948$), QRS ($p \approx 0.354$), QT ($p \approx 0.0925$), QTc ($p \approx 0.768$) and P-wave duration ($p \approx 0.594$) where the significance level, α, was taken to be $\alpha = 0.05$ (and for all non-corrected tests henceforth). For the optogenetic mice, single-factor ANOVA tests showed that there was no significant

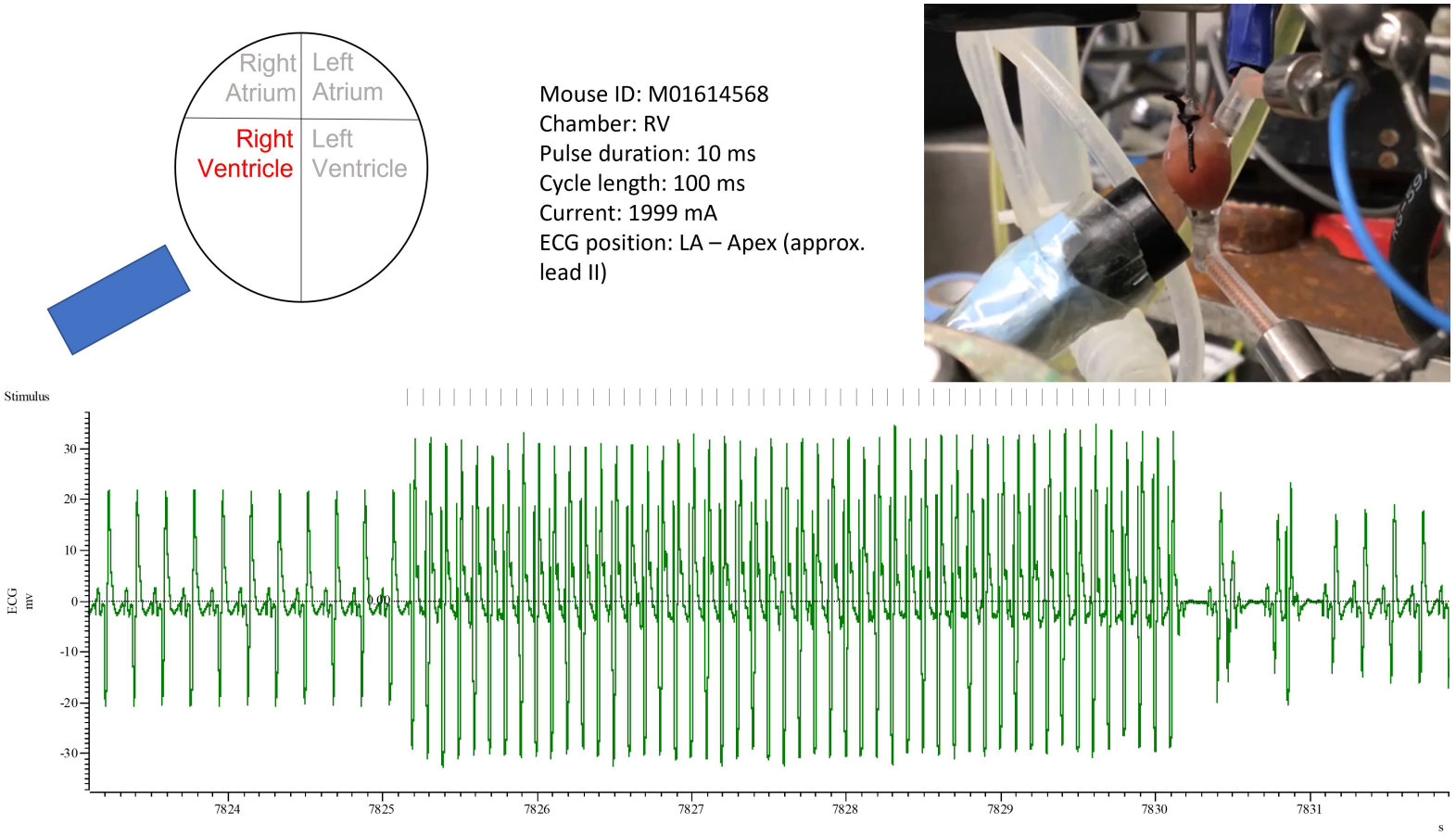


Figure 3.5: Optogenetic light pacing of the right ventricle. Top left: schematic showing position of the light relative to the chambers of the heart. Top middle: details of the experimental setup. Top right: photograph of the experimental setup. Bottom: ECG trace with lines above the trace indicating light pulses. Pulsing light was able to pace the heart from the right ventricle.

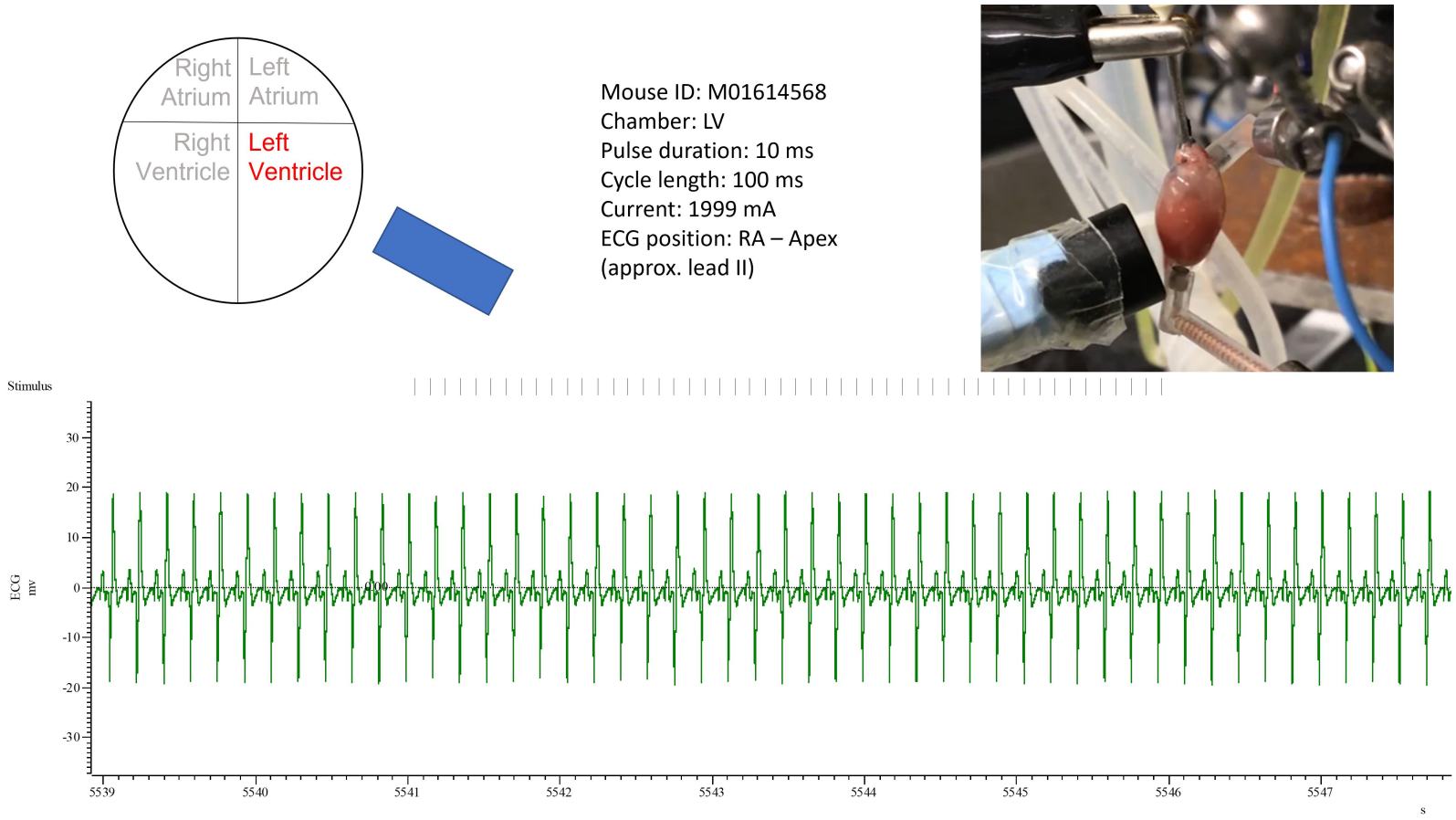


Figure 3.6: Failure to optogenetically light pace the left ventricle. Top left: schematic showing position of the light relative to the chambers of the heart. Top middle: details of the experimental setup. Top right: photograph of the experimental setup. Bottom: ECG trace with lines above the trace indicating light pulses. Pulsing light was not able to pace the heart from the left ventricle.

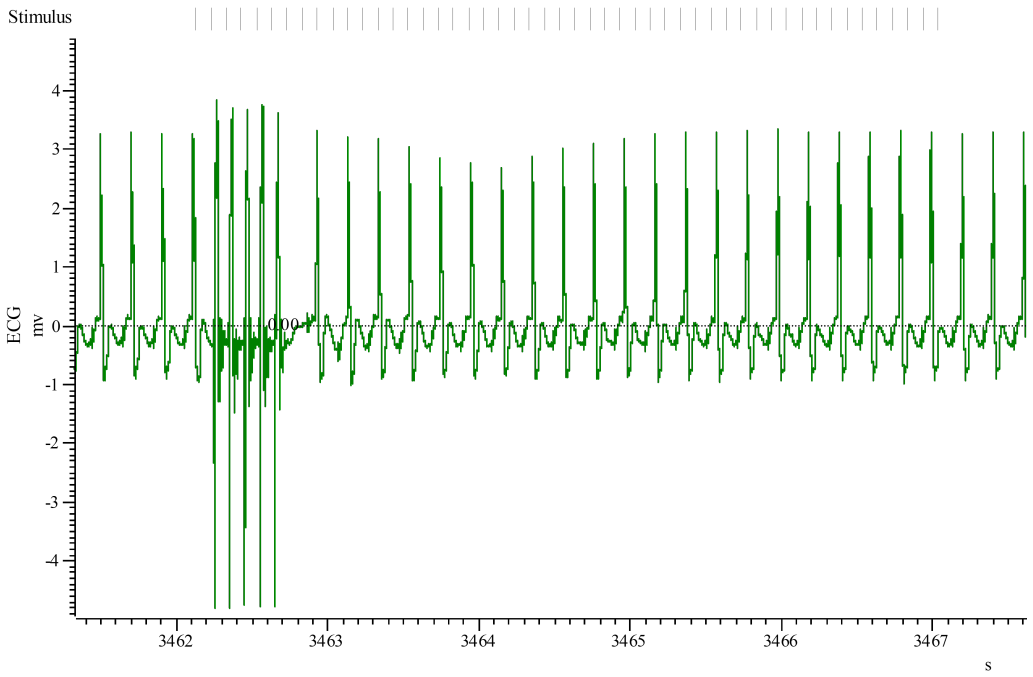


Figure 3.7: Partial optogenetic pacing capture of the left ventricle. Mouse: Dbh1.1b. Current: 1999 mA. Pulse duration: 10 ms. Pulse cycle length: 100 ms.

difference before, during or after pacing for QRS ($p \approx 0.796$), QT ($p \approx 0.683$) and P-wave duration ($p \approx 0.113$). However, single-factor ANOVA tests showed significant difference before, during or after pacing for RR ($p \approx 8.15 \times 10^{-5}$), PR ($p \approx 4.26 \times 10^{-6}$) and QTc ($p \approx 0.0391$). Bonferroni-corrected ($\alpha_c = 0.05/3 \approx 0.0167$) post-hoc two-tailed paired (or unpaired when measurements were missing, e.g. if the P-wave was not visible) t-tests, not assuming equal variances, showed significant differences in RR, PR and QTc before & during pacing ($p \approx \{0.00272, 7.24 \times 10^{-5}, 0.00801\}$, respectively) and during & after pacing ($p \approx \{0.00508, 2.44 \times 10^{-5}, 0.00325\}$, respectively), but not before & after pacing ($p \approx \{0.347, 0.408, 0.951\}$, respectively). This, and specifically the significant reduction in RR to the light cycle length, indicates that the optogenetic mouse model and pacing worked.

Table 3.2: Optogenetic responses of each mouse heart by chamber.

Chamber	Mouse ID	Response to light?
RA	M01614568	Response
	Dbh1.1a	Response
	Dbh1.1c	Response
	Dbh1.1b	Response
	Dbh3.1b	No response
LA	M01614568	Response
	Dbh1.1a	Response
	Dbh1.1c	Response
	Dbh1.1b	Response
	Dbh3.1b	No response
LV	M01614568	No response
	Dbh1.1a	No response
	Dbh1.1c	Response
	Dbh1.1b	Partial response (no full pacing train capture; partial capture at 8ms pulse duration & 1999mA current)
	Dbh3.1b	No response
RV	M01614568	Response
	Dbh1.1a	Response
	Dbh1.1c	Response
	Dbh1.1b	Response
	Dbh3.1b	No response

The ECG intervals before, during and after the pacing stimuli are shown in Figure 3.9 for all four chambers of all the Dbh-Cre-ChR2/tdTomato mouse hearts.

The result for the right atrium was discussed above.

For the left atrium, single-factor ANOVA tests showed that there was no significant difference before, during or after pacing for PR ($p \approx 0.381$), QRS ($p \approx 0.791$), QT ($p \approx 0.888$), QTc ($p \approx 0.211$) and P-wave duration ($p \approx 0.945$). However, single-factor ANOVA tests showed significant difference before, during

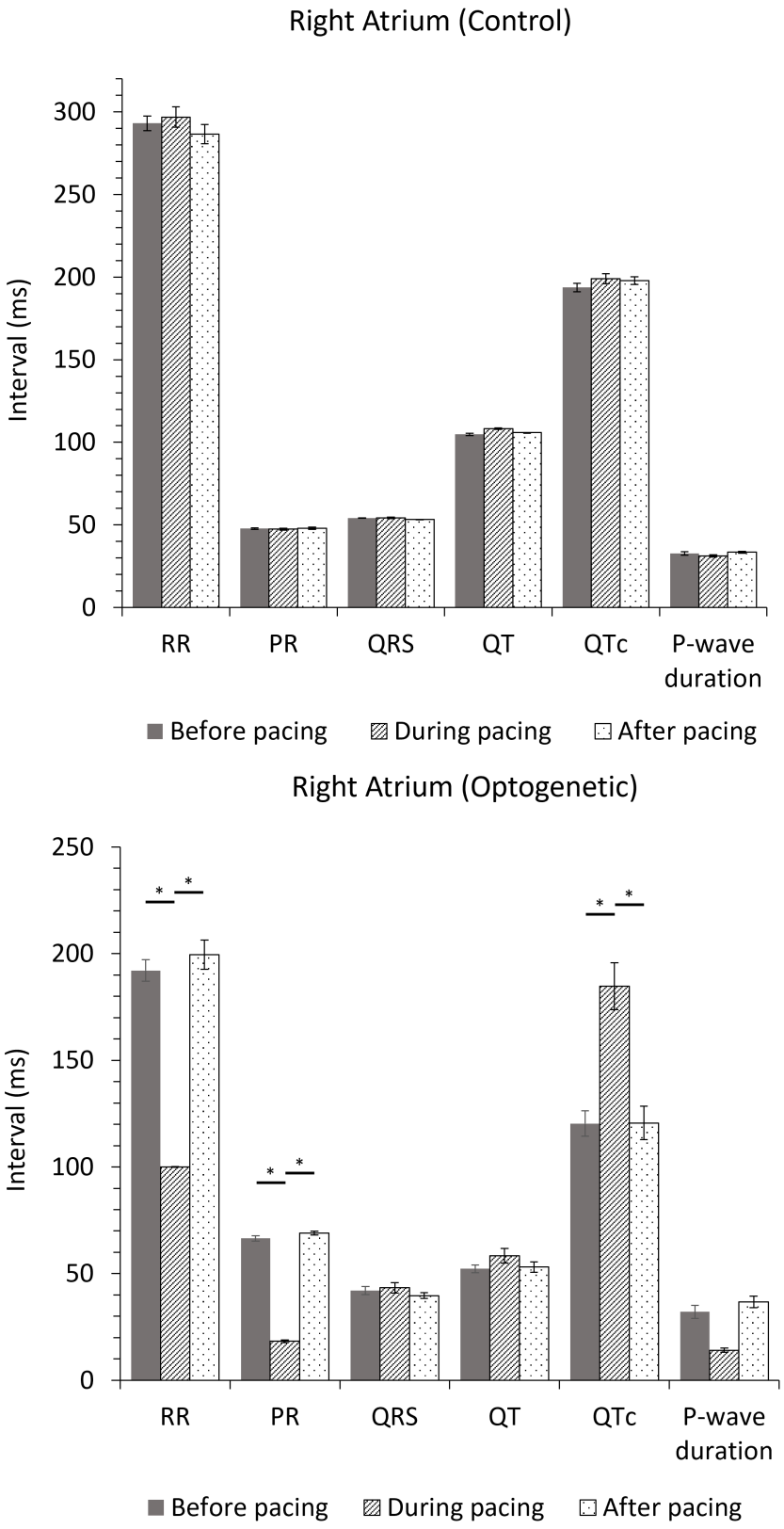


Figure 3.8: ECG intervals of a control mouse versus an optogenetic mouse for the right atrium. Intervals before, during and after light stimulation are shown. Asterisk indicates a statistically significant difference ($p < 0.05$ or $p < 0.0167$ for post-hoc tests). Data points: mean of three beats in the same heart $\pm$ standard error in the mean (SEM).

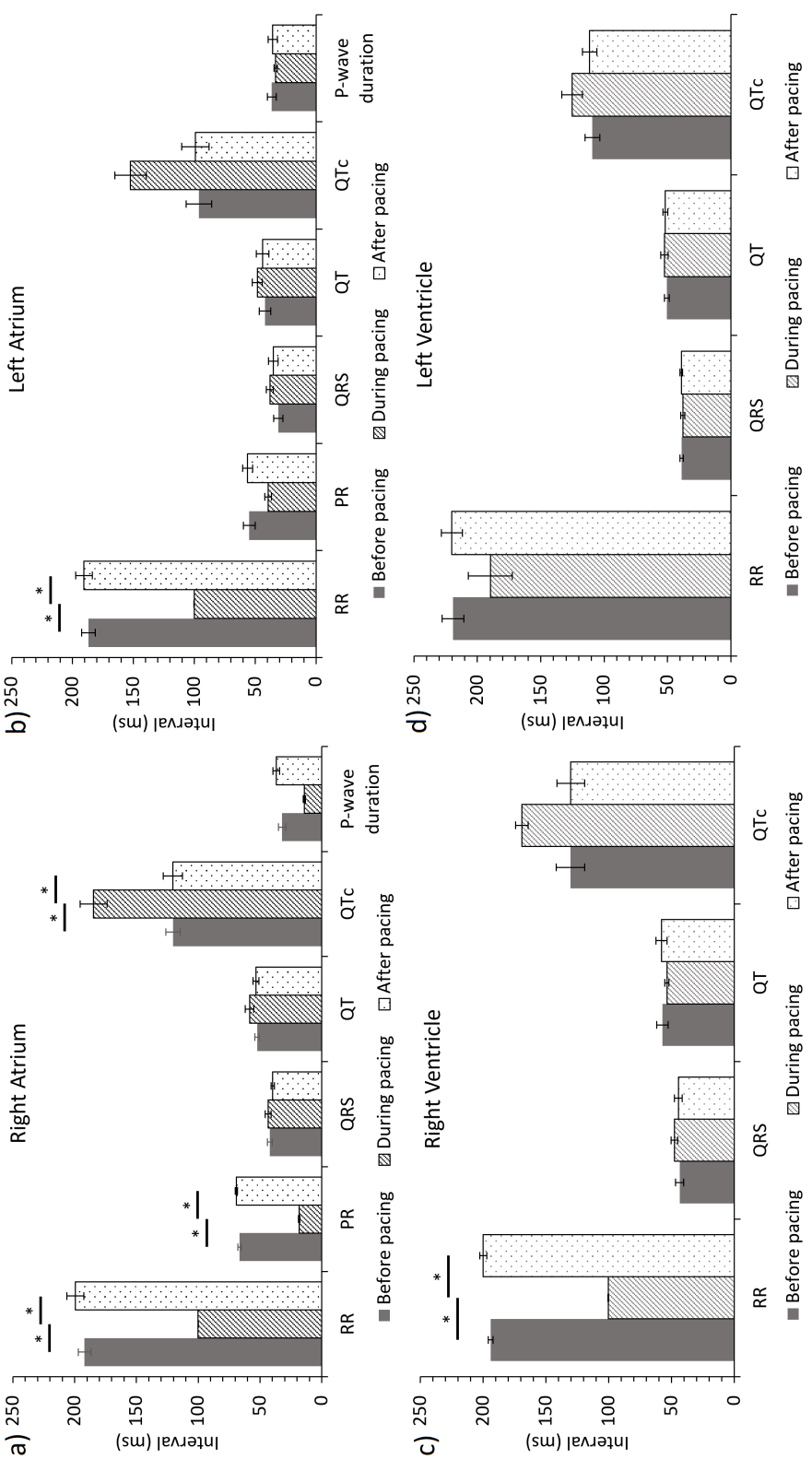


Figure 3.9: ECG intervals before, during and after light stimulation for each chamber of the heart using optogenetic mouse hearts. Asterisk indicates a statistically significant difference ($p < 0.05$ or $p < 0.0167$ for post-hoc tests). Data presented as mean $\pm$ SEM. Means used single beats from $n = 4$ optogenetic mouse hearts.

or after pacing for RR ($p \approx 0.000220$). Bonferroni-corrected ($\alpha_c = 0.05/3 \approx 0.0167$) post-hoc two-tailed paired t-tests, not assuming equal variances, showed significant differences in RR before & during pacing ($p \approx 0.00432$) and during & after pacing ($p \approx 0.00701$), but not before & after pacing ($p \approx 0.267$).

For the right ventricle, single-factor ANOVA tests showed that there was no significant difference before, during or after pacing for QRS ($p \approx 0.878$), QT ($p \approx 0.935$) and QTc ($p \approx 0.391$). However, single-factor ANOVA tests showed significant difference before, during or after pacing for RR ($p \approx 3.91 \times 10^{-8}$). Bonferroni-corrected ($\alpha_c = 0.05/3 \approx 0.0167$) post-hoc two-tailed paired t-tests, not assuming equal variances, showed significant differences in RR before & during pacing ($p \approx 0.000119$) and during & after pacing ($p \approx 0.000434$), but not before & after pacing ($p \approx 0.316$).

For the left ventricle, single-factor ANOVA tests showed that there was no significant difference before, during or after pacing for RR ($p \approx 0.618$), QRS ($p \approx 0.945$), QT ($p \approx 0.958$) and QTc ($p \approx 0.668$).

Figure 3.9 therefore indicates that in the optogenetic mice, the right atrium, left atrium and right ventricle were able to be paced easily, but there was difficulty in pacing the left ventricle.

The threshold current required for full pacing, effective ventricular refractory period and threshold pulse duration required for full pacing for each chamber of the heart for the optogenetic mouse hearts is presented in Figure 3.10. A single-factor ANOVA test showed no significant difference in threshold current between the chambers of the heart ($p \approx 0.315$). Despite this, the mean left ventricular threshold current is considerably higher than the other chambers. We speculate that if there were more data points for the LV, then ANOVA would show statistical significance. Only one data point was taken since all the other optogenetic mouse

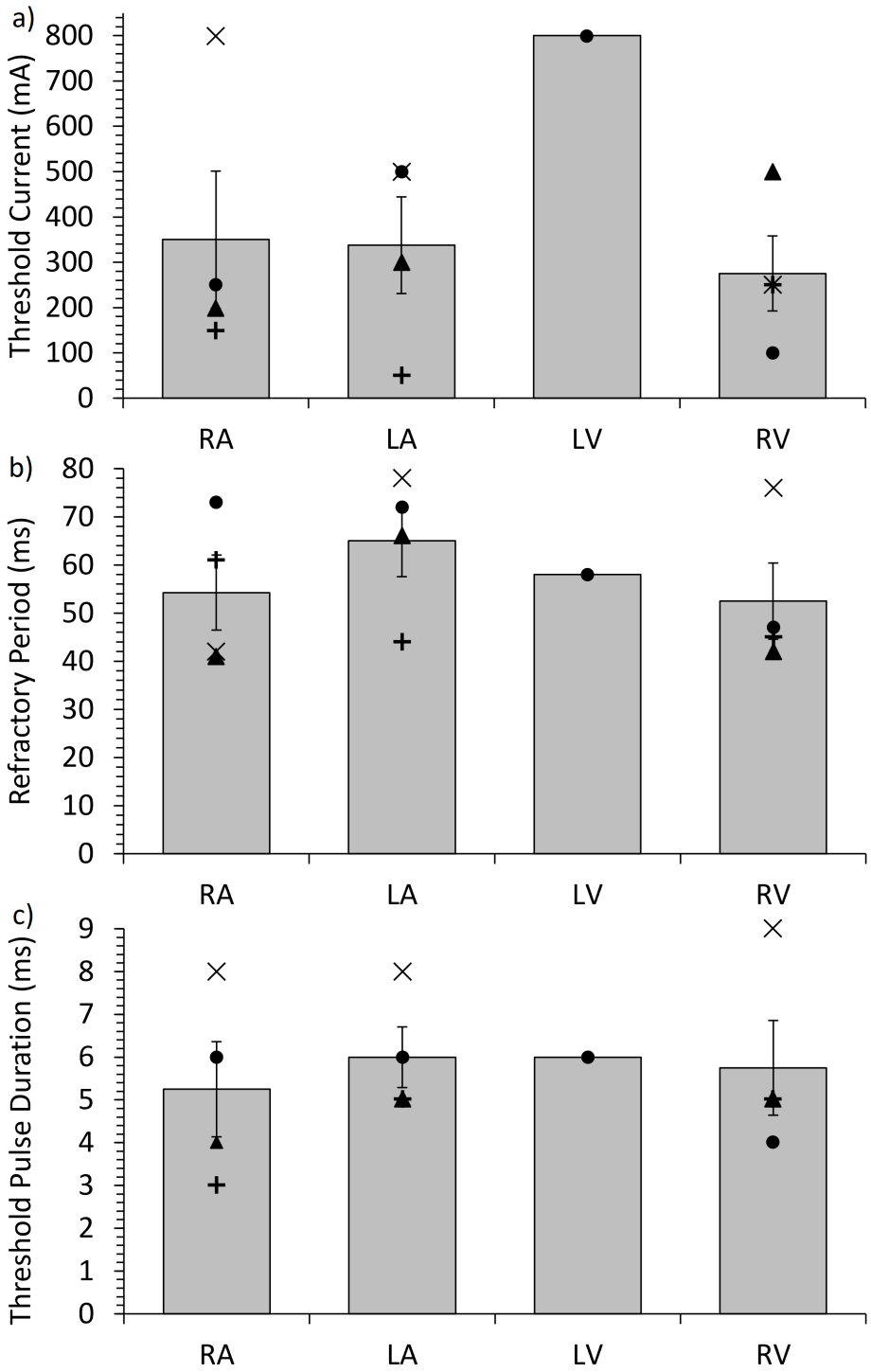


Figure 3.10: Threshold currents, refractory periods and threshold pulse durations for each chamber under light stimulation. Thresholds are for full one-to-one response pacing. Data shown as mean ± SEM. Individual data points shown, with a different symbol for each of $n = 4$ optogenetic mouse hearts (only one optogenetic heart responded to LV light stimulation).

hearts did not respond to light stimulation. It could be argued that the threshold current for these hearts is therefore at least 2000 mA or greater, since the technical maximum current that could be supplied to the setup was 1999 mA.

For the refractory period, a single-factor ANOVA test showed no significant difference between the chambers of the heart ($p \approx 0.681$).

A single-factor ANOVA test showed no significant difference in threshold pulse duration between the chambers of the heart ($p \approx 0.954$).

The isoprenaline experiment was then conducted as per the protocol laid out in Section 3.2.3.4. Isoprenaline is a non-selective beta adrenergic receptor agonist. It has positive chronotropic and inotropic effects through its stimulation of sympathetic signal transduction pathways. It also has a positive bathmotropic effect, that is, it increases the excitability of the cardiac muscle, it does so by raising the resting membrane potential. The aim of the isoprenaline experiment was to see if increased excitability could allow previously unresponsive left ventricles to be paced.

Figure 3.11 shows the results. For before pacing and after pacing, there is a statistically significant decrease in RR without and with isoprenaline (unpaired two-sample one-tail t-test assuming unequal variances; $p \approx 0.0226$ and 0.0285, respectively). For during pacing, there was a decrease in mean RR interval (but not to 100 ms) without and with isoprenaline, though an unpaired two-sample one-tail t-test, assuming unequal variances, did not show significance ($p \approx 0.211$).

This indicates that isoprenaline (100 nM) increases the heart rate before, during (though in this case, not significantly) and after pacing, but the RR interval still fails to match the 100 ms cycle length of the stimulus during the pacing train (i.e. it does not allow complete pacing of the LV).

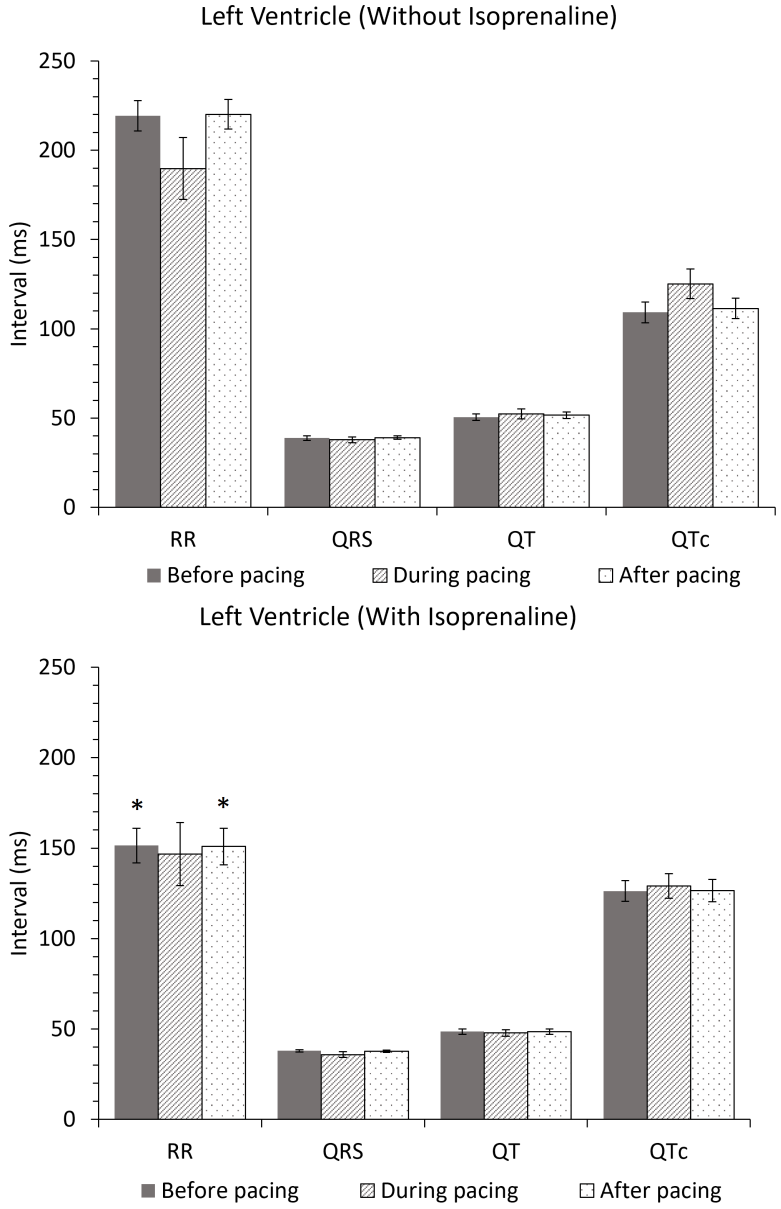


Figure 3.11: Left ventricular light stimulation with and without isoprenaline. Isoprenaline concentration: 100 nM. Asterisk: significant difference relative to without isoprenaline ($p < 0.05$). Data: mean over $n = 3$ hearts $\pm$ SEM.

As a further attempt at measuring whether 100 nM isoprenaline alters Dbh+ derived cellular optogenetic excitability, the threshold currents were measured across each chamber (see Figure 3.12).

There was no statistically significant difference in threshold current with and without isoprenaline for the right ventricle, left atrium and right atrium (unpaired two-sample two-tail t-test, assuming unequal variances; $p \approx \{0.420, 0.446, 0.654\}$ respectively).

There was a decrease in threshold current required to pace the left ventricle with-

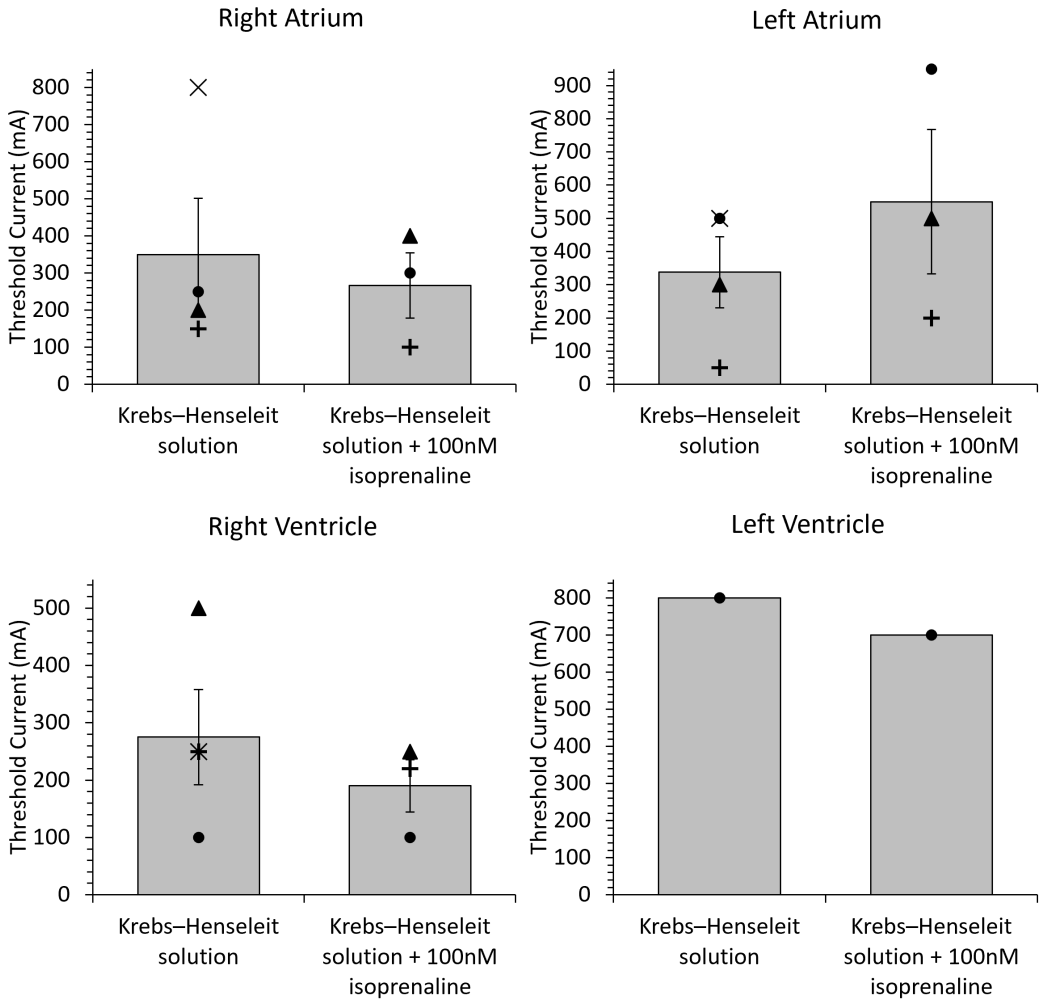


Figure 3.12: Threshold currents for complete light pacing with and without isoprenaline. Data: mean over $n = 3$ hearts $\pm$ SEM. Individual data points shown, with a different symbol for each of $n = 3$ optogenetic mouse hearts (only one optogenetic heart responded to LV light stimulation).

out and with isoprenaline, however a t-test for significance could not be performed since there was only one data point.

These results are inconclusive, it is possible that isoprenaline (100 nM) reduces the threshold current required.

3.4 Discussion

In humans, the QT interval decreases when the heart rate increases (RR interval decreases). Bazett's formula, which was used in this investigation, aims to correct the QT interval for heart rate in order to assess repolarisation time independent of heart rate. However, the formula over-corrects at high heart rates and under-corrects at low heart rates (Salvi, Karnad, Panicker, & Kothari, 2010). A version of Bazett's formula adapted to the murine ECG is often used, $QT_c = QT/\sqrt{RR/100}$, but was not used here (Roussel et al., 2016). In addition, work by Roussel *et al.* in 2016, suggested that, unlike humans, there is no relationship between the RR interval and the QT interval in mice (Roussel et al., 2016). They also showed large intra-individual and inter-individual variability in QT. They concluded that correction formulae for QT should not be used in mice and that it only introduces bias. This controversy does not have a significant impact on our investigation, since QTc was not a major focal point for analysis.

The ECG measuring methodology used in this investigation is shown in Figure 3.4 in Section 3.2.3.5. There are different methods of measuring the QT interval in the mouse. Zhang *et al.* in 2014 compared different QT measurement ECG intervals with a cellular measure of ventricular repolarisation: Q-APR$_{90}$ —the time between the start of the QRS complex and the time point at which the ventricular action potential reaches 90% recovery to baseline. Two measures of QT were compared. Both began at the first deviation from the isoelectric baseline at the

start of the QRS complex. QT1 was then measured to the trough of the negative T-wave whereas QT2 was measured to the return of the negative T-wave trough to the isoelectric baseline. QT1 was found to correlate more with changes in Q-APR$_{90}$ (Zhang, Wu, King, Huang, & Fraser, 2014). In our investigation, we therefore used QT1.

There are numerous differences in cardiac electrophysiology between mice and humans. The average heart rate in a conscious mouse is in the region of 600 bpm *in vivo*. We note this to be considerably higher than the *ex vivo* resting heart rates in our investigation of around 300 bpm, though this is commonly observed in similar experiments (Kass, Hare, & Georgakopoulos, 1998). The average resting heart rate in humans is around 70 bpm, almost 10 times lower. The ventricular action potential duration (to 90% repolarisation) is much shorter in the mouse (~40 ms) than the human (~300 ms) (Speerschneider & Thomsen, 2013). The mouse ventricular action potential lacks the plateau phase seen in humans, instead there is a swift repolarisation almost immediately following depolarisation, providing a more triangular action potential shape. This is a result of species-specific differences in ion channel currents. The difference in action potential morphology leads to a different ECG shape. In mice, there is a prominent J-wave (ST segment elevation) and an inverted T-wave (Boukens et al., 2014).

There was some uncertainty in interpreting some of the ECG traces in our experiment, in particular it was difficult to distinguish notched S waves from J waves. In general, the ECG was interpreted as having a notched S wave instead of an early J wave.

The mice used were all approximately 1 month old except for one which was 6.7 months old. These are all considered to be relatively young mice. There may have been variations in the ECG traces between mice due to age differences.

Merentie *et al.* in 2015 showed that older mice showed longer P wave durations, PQ intervals, QRS durations and smaller R amplitudes (Merentie et al., 2015). In this work, groups of 2-3, 14 and 20-24 month old mice were used. The middle age and old groups were considerably older than our oldest mouse. We therefore speculate that ECG changes due to ageing had little impact on our experiment.

Our unpaced, resting ECG intervals are slightly slower but roughly in line with the literature, for example with the Lei group Pnmt-Cre-ChR2/tdTomato model ECG intervals (Wang et al., 2017). It is notable that our control was not perfect, the control mouse had a slower heart rate than the optogenetic mice (e.g., see Figure 3.8) with an RR of ~300 ms compared to ~200 ms for the optogenetic mice.

We note that in the statistical analysis, the post-hoc Bonferroni correction used likely under powers some of our statistical tests, utilising different post-hoc correction methods could have brought some of the results into significance.

In future work, multiple isoprenaline concentrations could have been used, specifically isoprenaline concetrations in excess of 100 nM, to see whether excitability of the heart could be increased to a sufficient level for successful optogenetic pacing of unresponsive left ventricles.

In addition, we note that the light intensity used in this experiment is higher than some other work, for example in the work by Wengrowski *et al.*, a light intensity of 0.12–$0.3\,\mathrm{mW\,mm^{-2}}$ was used, relative to our range of 0.6–$0.8\,\mathrm{mW\,mm^{-2}}$ (Wengrowski et al., 2015).

In this chapter, we showed that an optogenetically engineered Dbh+ derived subpopulation was difficult to pace in the left ventricle, but was easily photoactivated in the left atrium, right atrium and right ventricle. This was also apparent by an increased threshold light intensity required to photoactivate Dbh+ derived cells in

the left ventricle. On the surface, this suggests the subpopulation spatial distribution is reduced in the left ventricle compared to the other chambers of the heart. There was no statistically significant difference in ventricular refractory period of the Dbh+ derived cells between all four chambers of the heart. Isoprenaline at a concentration of 100 nM did not increase the Dbh+ derived cell excitability enough for pacing in the left ventricle.

4 | Spatial and morphological characteristics: Immunohistochemistry

4.1 Introduction

In this chapter we aim to explore the spatial distribution of the Dbh+ derived cell subpopulation, as well as inferring possible functions. We also aim to characterise the morphology and anatomy of the Dbh+ derived cell. To do this we utilise the same mouse line as in Chapter 3—the Dbh-Cre-ChR2/tdTomato mouse line. The result of this genetic construct is that any cell in the heart that expresses or has ever expressed the Dbh gene, or is the progeny of any cell that at any point expressed the Dbh gene, also expresses the fluorescent marker tdTomato. This is visualised in this chapter with fluorescence microscopy. In addition, we use various antibodies to immunostain key protein markers of cellular phenotype within the heart.

4.2 Methods

4.2.1 Fixation and sectioning

Directly after the Langendorff-perfused electrophysiology experiments (see Section 3.3), the whole heart samples were removed from the cannula in the perfusion setup. Excess fibrous tissue was cut away from the samples. The samples were placed in cold 4% (v/v) paraformaldehyde/phosphate-buffered saline (PFA/PBS) (Sigma-Aldrich/Sigma-Aldrich) for 30 minutes at 4°C. While at 4°C, the samples were transferred into 10% sucrose solution for 1 hour, then transferred into 20% sucrose solution for 1 hour and then transferred into 30% sucrose solution

and kept overnight (for approximately 10 hours). The samples were removed and embedded in a plastic well filled with Optimal Cutting Temperature (OCT) compound. Each embedded sample was frozen by placing the well in a metal container surrounded by dry ice for 10 minutes. The embedded samples were removed and kept in a $-80°\mathrm{C}$ freezer. Coronal sections of the embedded samples were taken with a thickness of 10 $\mu\mathrm{m}$ every 50 $\mu\mathrm{m}$ using a Leica CM3050 S cryostat set at $-20°\mathrm{C}$. The sections were mounted on Polysine microscope slides and kept in a $-80°\mathrm{C}$ freezer.

4.2.2 Immunostaining

The slides were then removed from the $-80°\mathrm{C}$ freezer, placed into a PBS bath and washed in a rocker for 2 minutes to thaw the slides. During the immunostaining process, the samples on the slides were not allowed to dry. The slides were removed and excess PBS wiped off. Hydrophobic barriers were drawn around the samples using an ImmEdge® Hydrophobic Barrier PAP Pen (Vector Laboratories) to keep the reagents pooled in droplets over each sample. At room temperature, droplets of 4% (v/v) PFA in PBS at pH 7.4 were placed on the samples and left for 15 minutes to fix the samples. Excess PFA was removed by tapping the slides on paper towels, then the slides were placed in a PBS bath and washed by rocking for 5 minutes. The slides were washed three times in this way with PBS, each time replacing the bath with fresh PBS. Excess PBS was removed by tapping the slides.

To permeablise the cell membranes, 0.1% (v/v) Triton X-100 (ThermoFisher) in PBS was pipetted onto the samples and incubated at room temperature for 30 minutes. Then the slides were washed in PBS bath by rocking for 5 minutes. The slides were removed and excess PBS was removed. Droplets of 10% (v/v) normal goat serum (NGS) (Abcam) in PBS containing 0.1% (v/v) Tween 20 were

placed on the samples and incubated at room temperature for 1 hour to block non-specific antibody binding. The slides were then put into a PBS bath and taken out. The hydrophobic pen was used again if some of the barrier washed off. The primary antibody in PBS containing 1% (v/v) NGS was pipetted onto the samples. The tray with the slides was covered with foil and incubated in a 4°C refrigerator overnight.

The slides were then removed from the refrigerator and allowed to reach room temperature over 30 minutes. The slides were washed in a PBS bath for 5 minutes on a rocker, twice. Excess PBS was removed from the slides. In a dark room, the secondary antibody in PBS containing 1% (v/v) NGS, covered in foil, was applied as droplets onto the samples. The tray with the slides was covered in foil and placed on a rocker at room temperature for 2 hours. The primary antibodies that were used were anti-tyrosine-hydroxylase (1:200 dilution, rabbit, monoclonal, IgG, Abcam); anti-alpha-actinin (1:200 dilution, mouse, monoclonal, IgG1, Abcam); anti-adrenaline (1:200 dilution, rabbit, polyclonal, IgG, Abcam) and anti-NF-68kDa (1:200 dilution, mouse, monoclonal, IgG1, Abcam). The secondary antibodies that were used were goat anti-mouse (1:200 dilution, Alexa Fluor® 488, Abcam) and goat anti-rabbit (1:200 dilution, Cy5®, Abcam). The slides were then washed in a PBS bath for 5 minutes on a rocker, twice. Excess PBS on the slides was wiped off and droplets of VECTASHIELD® Antifade Mounting Medium with 4',6-diamidino-2-phenylindole (DAPI) (Vector Laboratories) were pipetted onto the samples. The samples were covered with cover slips and sealed with top coat nail polish. The slides were then stored in the dark in a 4°C refrigerator. The slides were imaged using a Zeiss LSM780 confocal fluorescence microscope, for whole-field images, and a Leica TCS SP5 confocal fluorescence microscope. Also, an Olympus BX40 epifluorescence microscope equipped with Olympus UPlanFL 20x, 40x, 60x and 100x objectives was used where the images

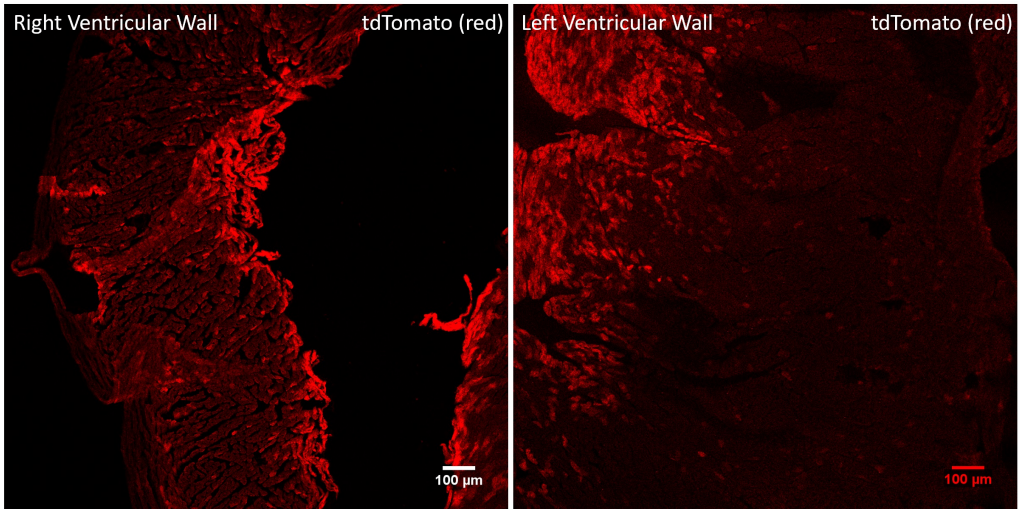


Figure 4.1: Micrographs of tdTomato fluorescence (marking Dbh+ derived cells) in the right and left ventricular walls. Mouse: M01614568. Coronal section.

were acquired using an Orca-ER camera (Hamamatsu) and HCImage software (Hamamatsu). Images were analysed using the Fiji (ImageJ) software (Schindelin et al., 2012).

4.3 Results

Dbh+ derived cells are genetically labelled with tdTomato (excitation wavelength: 554 nm, emission wavelength: 581 nm), coloured red in the figures. Figure 4.1, of the right and left ventricular walls from mouse M01614568, shows that the Dbh+ derived cells are endocardially distributed. All the micrographs in this section are of coronal sections of the heart.

The spatial distribution across the entire heart from mouse Dbh1.1c is shown in Figure 4.2. DAPI (marking nuclei) is coloured in blue (excitation wavelength: 358 nm; emission wavelength: 461 nm) and coloured in red is the tdTomato that labels the Dbh+ derived cells. The Dbh+ derived cells are particularly distributed in the endocardium, subendocardium, interventricular septum and atria.

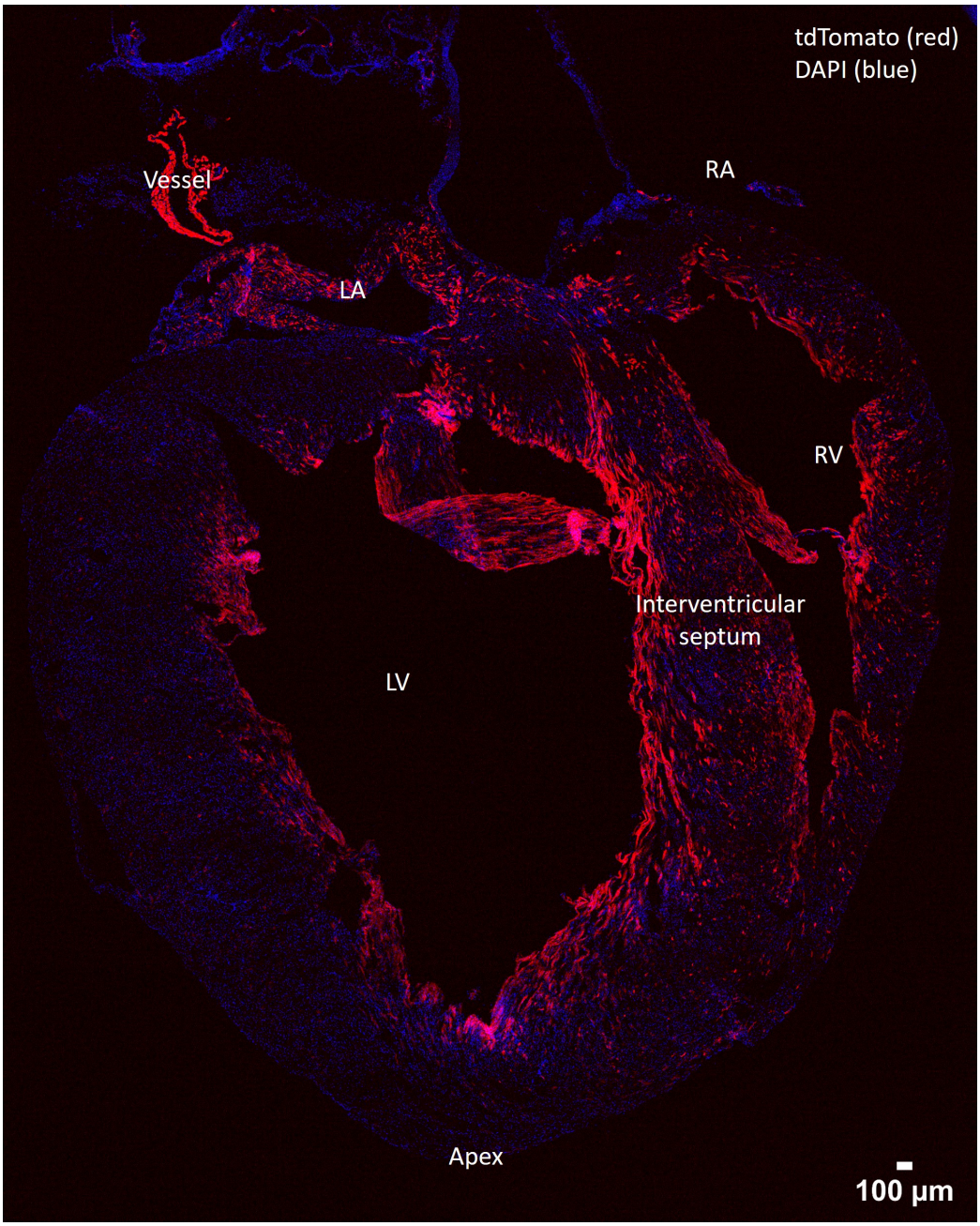


Figure 4.2: Whole heart spatial distribution of Dbh+ derived cells. Mouse ID: Dbh1.1c. Blue: DAPI (nuclei). Red: tdTomato genetic fluorescence (Dbh+ derived cells).

Alpha-actinin 2 is a protein which binds actin to the Z-disk in the sarcomeres of skeletal and cardiac myocytes. It is therefore a marker for cardiomyocytes in our heart sections. There is co-localisation of fluorescence between anti-alpha-actinin (secondary antibody Alexa Fluor® 488; excitation wavelength: 495 nm, emission wavelength: 519 nm) coloured in green and tdTomato in red shown in Figure 4.3. The top panels show the left base of the heart and the bottom panels show the right ventricular wall, both of which originate from the mouse M01614568. This suggests that the Dbh+ derived cells are cardiomyocytes.

A higher magnification micrograph (40x) is shown in Figure 4.4 of the left ventricular wall from the heart of mouse Dbh1.1a. From this, we can observe that the length of each Dbh+ derived cell is approximately $100\,\mu\mathrm{m}$ and the width is approximately $30\,\mu\mathrm{m}$. The morphology is rectangular and elongated. Striations can be seen, indicating the presence of sarcomeric Z-disks. These observations suggest that the Dbh+ derived cells are cardiomyocytes.

Figure 4.5 shows a micrograph of the interventricular septum from the heart of mouse M01614568. Coloured in red is tdTomato, labelling the Dbh+ derived cells. Coloured in green is the fluorescence from anti-Th (secondary antibody Cy5; excitation wavelength: 650 nm, emission wavelength: 670 nm). Tyrosine-hydroxylase is an enzyme that catalyses the conversion of L-tyrosine into L-DOPA. Within the catecholaminergic biosynthetic pathway, L-DOPA is a precursor to noradrenaline, an important neurotransmitter for adrenergic receptors. Tyrosine-hydroxylase is present in the terminals of the sympathetic neurons that innervate the heart. In the figure, we can see the characteristic varicosities (Zaglia & Mongillo, 2017) of the sympathetic innervation. This suggests that Dbh+ derived cells synapse with or are very proximal to sympathetic neurons.

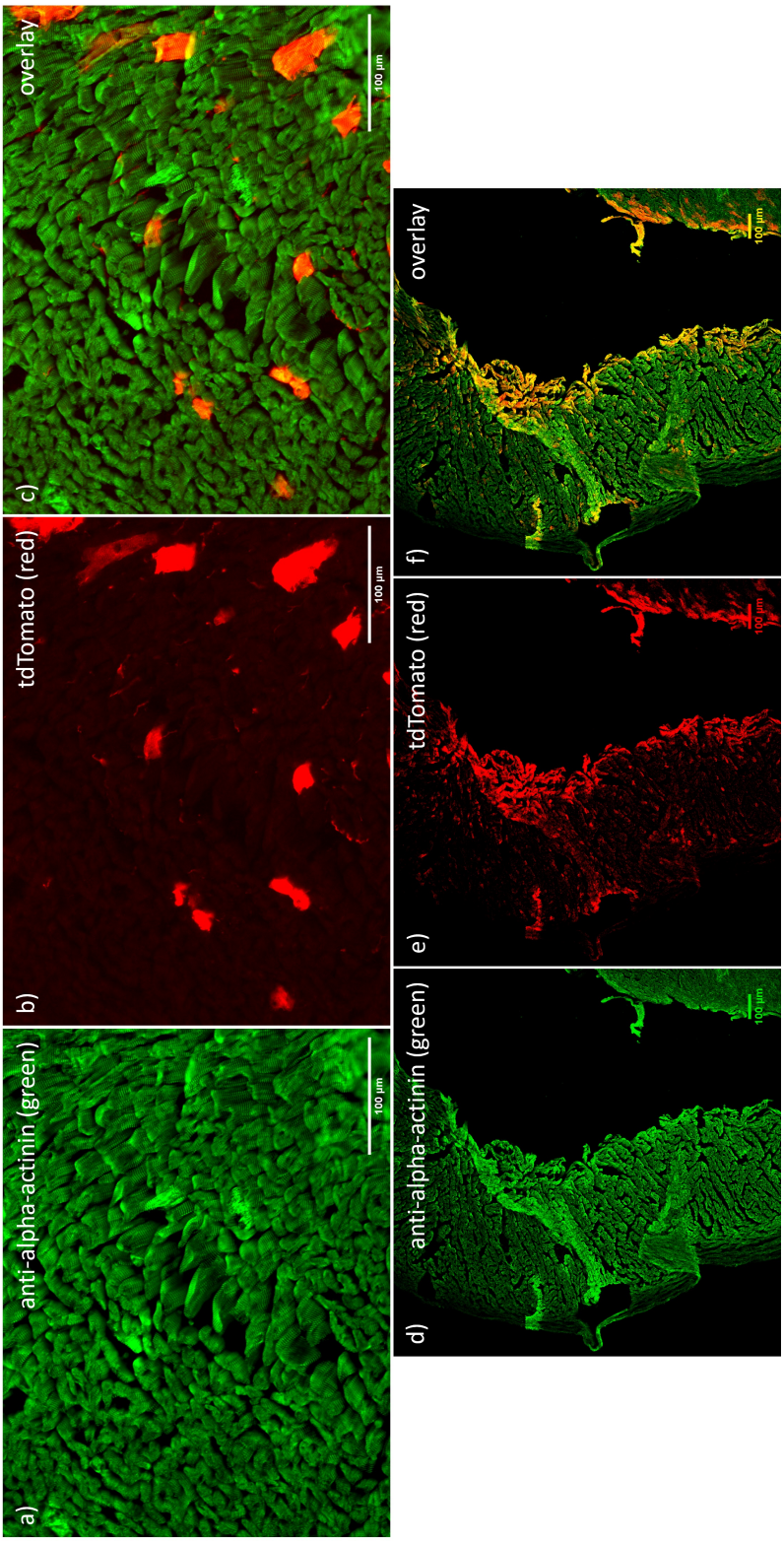


Figure 4.3: Alpha-actinin immunostaining and tdTomato genetic fluorescence. Dbh+ derived cells marked by tdTomato fluorescence (red). Mouse ID: M01614568. Top panels: left base of the heart. Bottom panels: right ventricular wall.

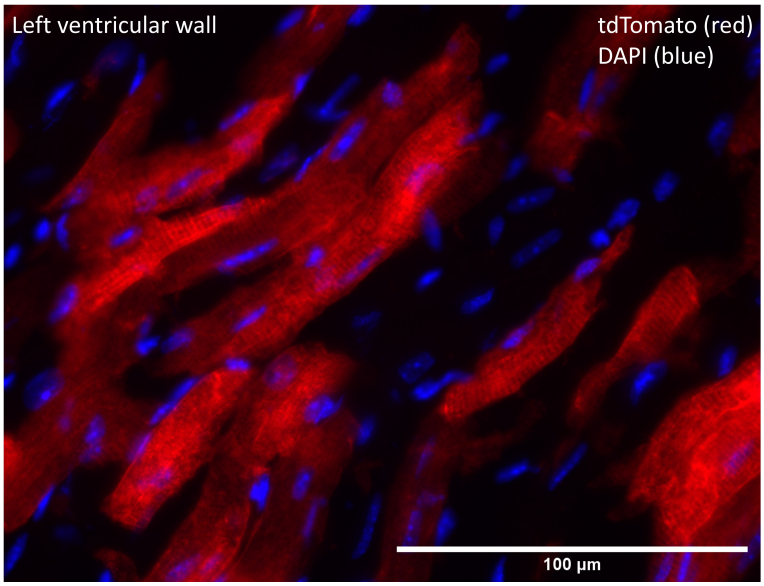


Figure 4.4: High magnification micrograph of tdTomato genetic fluorescence (Dbh+ derived cells) in the left ventricular wall. Blue: DAPI (nuclei). Mouse ID: Dbh1.1a.

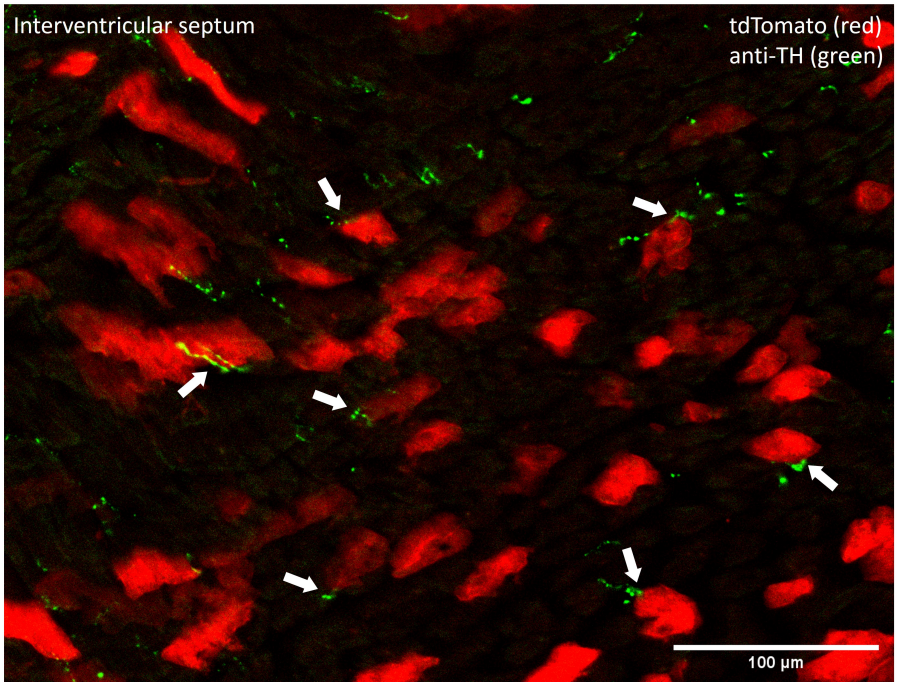


Figure 4.5: Micrograph of the interventricular septum with tdTomato fluorescence and Th immunostaining. Dbh+ derived cells marked by tdTomato (red). Anti-Th fluorescence marks sympathetic nerve terminals. Mouse ID: M01614568.

4.4 Discussion

For future work, it would be interesting to analyse the number of Dbh+ derived cells that are sympathetically innervated and compare it to the total number of cells (both Dbh+ and Dbh-) that are sympathetically innervated. This would provide insight into whether the Dbh+ derived subpopulation is preferentially innervated.

In summary, the genetic tdTomato fluorescence has shown that the Dbh+ derived subpopulation is spatially distributed in the endocardium, subendocardium, interventricular septum and atria. The subpopulation is preferentially distributed endocardially, with a lack of cells in the left ventricular myocardium, as well as a lack of cells in the epicardium and subepicardium. This spatial distribution is close to that of the cardiac conduction system which consists of the sinoatrial node, atrioventricular node, His bundle, left and right bundle branches and Purkinje fibres. This association with the cardiac conduction system would need further investigation.

The immunohistochemistry in combination with genetic fluorescence has shown that the Dbh+ derived cells appear to be cardiomyocytes. They express proteins associated with cardiomyocytes, such as sarcomeric alpha-actinin. The Dbh+ derived cells have the size, morphology and striations of cardiomyocytes. The Dbh+ derived cells appear to be innervated by sympathetic neurons.

5 | Conclusions

Notably, the Dbh+ derived subpopulation is endocardially distributed (Chapter 4) and the left ventricular wall is thicker than the right ventricular wall. Hence, light incident on the left ventricle, during the whole-heart electrophysiology (Chapter 4), was not able to penetrate deep enough to stimulate the endocardially distributed Dbh+ derived cells, whereas it was able to penetrate the thinner right ventricular and atrial walls. This explains the failure to pace the LV.

The original aim of this project was to characterise the Dbh expressing/Dbh+ derived cardiac cellular subpopulation. We end by summarising our findings. The Dbh+ derived cells:

- are spatially distributed in the endocardium, subendocardium, atria and interventricular septum, as evidenced by tdTomato fluorescence and optogenetic photoactivation in a Dbh-Cre-ChR2/tdTomato mouse line;
- have similar refractory periods regardless of heart region, evidenced by optogenetic S1S2 stimulation protocols in an optogenetic mouse line;
- are cardiomyocytes, evidenced by having the size, morphology, striations and protein expression of cardiomyocytes as revealed by genetic fluorescence and immunohistochemistry; and
- appear to be sympathetically innervated, evidenced by proximity to sympathetic terminals shown by genetic fluorescence and immunohistochemistry.

Finally, Dbh expressing cells appear to developmentally originate from first heart field splanchnic mesoderm, as evidenced by the RNA expression profile of the earliest RNA-velocity inferred pseudotime cluster analysed using single-cell RNA sequencing.

Bibliography

Adameova, A., Abdellatif, Y., & Dhalla, N. S. (2009, July). Role of the excessive amounts of circulating catecholamines and glucocorticoids in stress-induced heart disease. *Canadian Journal of Physiology and Pharmacology*, *87*(7), 493–514. doi: 10.1139/Y09-042

Antonopoulos, A. S., & Antoniades, C. (2017, June). The role of epicardial adipose tissue in cardiac biology: Classic concepts and emerging roles. *The Journal of Physiology*, *595*(12), 3907–3917. doi: 10.1113/JP273049

Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.

Bazett, H. (1920). An analysis of the time-relations of electrocardiograms. *Heart: A Journal for the Study of the Circulation...*, *7*, 353.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., ... Newell, E. W. (2019, January). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*, *37*(1), 38–44. doi: 10.1038/nbt.4314

Bergen, V., Lange, M., Peidli, S., Wolf, F. A., & Theis, F. J. (2020). Generalizing rna velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, *38*(12), 1408–1414.

Bergen, V., Soldatov, R. A., Kharchenko, P. V., & Theis, F. J. (2021, August). RNA velocity—current challenges and future perspectives. *Molecular Systems Biology*, *17*(8). doi: 10.15252/msb.202110282

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Boron, W. F., & Boulpaep, E. L. (Eds.). (2017). *Medical Physiology* (Third edition ed.). Philadelphia, PA: Elsevier.

Boukens, B. J., Rivaud, M. R., Rentschler, S., & Coronel, R. (2014, November). Misinterpretation of the mouse ECG: 'musing the waves of Mus musculus'. *The Journal of Physiology*, *592*(Pt 21), 4613–4626. doi: 10.1113/jphysiol.2014.279380

Boyett, M., Honjo, H., & Kodama, I. (2000, September). The sinoatrial node, a heterogeneous pacemaker structure. *Cardiovascular Research*, *47*(4), 658–687. doi: 10.1016/S0008-6363(00)00135-8

Chen, G., Ning, B., & Shi, T. (2019). Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Frontiers in Genetics*, *10*, 317. doi: 10.3389/fgene.2019.00317

Clip Art Library. (2021). *Unlabelled diagram of the heart.* `http://clipart-library.com/clipart/di45ebgLT.htm`. (Online; accessed 17 June 2021)

Ebert, S. N., Rong, Q., Boe, S., Thompson, R. P., Grinberg, A., & Pfeifer, K. (2004). Targeted insertion of the Cre-recombinase gene at the phenylethanolamine n-methyltransferase locus: A new model for studying the developmental distribution of adrenergic cells. *Developmental Dynamics*, *231*(4), 849–858. doi: 10.1002/dvdy.20188

Ebert, S. N., & Thompson, R. P. (2001, January). Embryonic Epinephrine Synthesis in the Rat Heart Before Innervation. *Circulation Research*, *88*(1), 117–124. doi: 10.1161/01.RES.88.1.117

Gaidatzis, D., Burger, L., Florescu, M., & Stadler, M. B. (2015). Analysis of intronic and exonic reads in rna-seq data characterizes transcriptional and post-transcriptional regulation. *Nature biotechnology*, *33*(7), 722–729.

Galow, A.-M., Kussauer, S., Wolfien, M., Brunner, R. M., Goldammer, T., David, R., & Hoeflich, A. (2021, August). Quality control in scRNA-Seq can discriminate pacemaker cells: The mtRNA bias. *Cellular and molecular*

life sciences: CMLS. doi: 10.1007/s00018-021-03916-5

George, R. M., Maldonado-Velez, G., & Firulli, A. B. (2020, October). The heart of the neural crest: Cardiac neural crest cells in development and regeneration. *Development*, *147*(20), dev188706. doi: 10.1242/dev.188706

Gessert, S., & Kühl, M. (2010, July). The Multiple Phases and Faces of Wnt Signaling During Cardiac Differentiation and Development. *Circulation Research*, *107*(2), 186–199. doi: 10.1161/CIRCRESAHA.110.221531

Gilbert, S. F., et al. (2000). *Developmental Biology.* Sinauer associates Sunderland, MA.

Girvan, M., & Newman, M. E. J. (2002, June). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826. doi: 10.1073/pnas.122653799

Gorza, L., Schiaffino, S., & Vitadello, M. (1988, August). Heart conduction system: A neural crest derivative? *Brain Research*, *457*(2), 360–366. doi: 10.1016/0006-8993(88)90707-X

Grassam-Rowe, A., Ou, X., & Lei, M. (2020, August). Novel cardiac cell subpopulations: Pnmt-derived cardiomyocytes. *Open Biology*, *10*(8), 200095. doi: 10.1098/rsob.200095

Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F., & Theis, F. J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods*, *13*(10), 845–848.

Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. (2017a). A practical guide to single-cell rna-sequencing for biomedical research and clinical applications. *Genome Medicine*, *9*(1), 1–12.

Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. (2017b, August). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine*, *9*(1), 75. doi: 10.1186/s13073-017

-0467-4

Herring, N., & Paterson, D. J. (2018). Levick's Introduction to Cardiovascular Physiology, Sixth Edition. , 449.

Hildreth, V., Anderson, R. H., & Henderson, D. J. (2009). Autonomic innervation of the developing heart: Origins and function. *Clinical Anatomy*, *22*(1), 36–46. doi: 10.1002/ca.20695

Hill, M. (2021). *Intermediate Cardiac Embryology.* `https://embryology.med.unsw.edu.au/embryology/index.php/Intermediate_Cardiac_Embryology`. (Online; accessed 15 September 2021)

Honjo, H., Boyett, M. R., Kodama, I., & Toyama, J. (1996, November). Correlation between electrical activity and the size of rabbit sino-atrial node cells. *The Journal of Physiology*, *496*(Pt 3), 795–808.

Hopkins Medicine. (2021). *Adrenal glands.* `https://www.hopkinsmedicine.org/health/conditions-and-diseases/adrenal-glands`. (Online; accessed 17 June 2021)

Huang, M. H., Friend, D. S., Sunday, M. E., Singh, K., Haley, K., Austen, K. F., ... Smith, T. W. (1996, September). An intrinsic adrenergic system in mammalian heart. *The Journal of Clinical Investigation*, *98*(6), 1298–1303. doi: 10.1172/JCI118916

Ideker, R. E., Kong, W., & Pogwizd, S. (2009, March). Purkinje Fibers and Arrhythmias. *Pacing and clinical electrophysiology : PACE*, *32*(3), 283–285. doi: 10.1111/j.1540-8159.2008.02232.x

Ignarro, L. J., & Shideman, F. E. (1968a, January). Appearance and Concentrations of Catecholamines and Their Biosynthesis in the Embryonic and Developing Chick. *Journal of Pharmacology and Experimental Therapeutics*, *159*(1), 38–48.

Ignarro, L. J., & Shideman, F. E. (1968b, January). Norepinephrine and Epinephrine in the Embryo and Embryonic Heart of the Chick: Uptake and Subcellular Distribution. *Journal of Pharmacology and Experimental Therapeutics*, *159*(1), 49–58.

Ignarro, L. J., & Shideman, F. E. (1968c, January). The Requirement of Sympathetic Innervation for the Active Transport of Norepinephrine by the Heart. *Journal of Pharmacology and Experimental Therapeutics*, *159*(1), 59–65.

Kass, D. A., Hare, J. M., & Georgakopoulos, D. (1998, March). Murine Cardiac Function: A Cautionary Tail. *Circulation Research*, *82*(4), 519–522. doi: 10.1161/01.RES.82.4.519

Kitajima, S., Miyagawa-Tomita, S., Inoue, T., Kanno, J., & Saga, Y. (2006). Mesp1-nonexpressing cells contribute to the ventricular cardiac conduction system. *Developmental Dynamics*, *235*(2), 395–402. doi: 10.1002/dvdy.20640

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., … Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, *161*(5), 1187–1201.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., … others (2018). Rna velocity of single cells. *Nature*, *560*(7719), 494–498.

Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C. L., Lindberg, E. L., … Teichmann, S. A. (2020, December). Cells of the adult human heart. *Nature*, *588*(7838), 466–472. doi: 10.1038/s41586-020-2797-4

Long, X., Creemers, E. E., Wang, D.-Z., Olson, E. N., & Miano, J. M. (2007, October). Myocardin is a bifunctional switch for smooth versus skeletal muscle differentiation. *Proceedings of the National Academy of Sciences*, *104*(42), 16570–16575. doi: 10.1073/pnas.0708253104

Luecken, M. D., & Theis, F. J. (2019, June). Current best practices in single-cell RNA-seq analysis: A tutorial. *Molecular Systems Biology*, *15*(6). doi: 10.15252/msb.20188746

Lymperopoulos, A., Rengo, G., & Koch, W. J. (2013, August). The Adrenergic Nervous System in Heart Failure: Pathophysiology and Therapy. *Circulation research*, *113*(6), 10.1161/CIRCRESAHA.113.300308. doi: 10.1161/CIRCRESAHA.113.300308

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., ... others (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, *161*(5), 1202–1214.

Mary Piper, R. K. (2021). *Single-cell RNA-seq: Quality Control Analysis.* Retrieved from `https://hbctraining.github.io/In-depth-NGS-Data-Analysis-Course/sessionIV/lessons/SC_quality_control_analysis.html`

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

McInnes, L., Healy, J., & Melville, J. (2020, September). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*.

Meilhac, S. M., Lescroart, F., Blanpain, C., & Buckingham, M. E. (2014, September). Cardiac Cell Lineages that Form the Heart. *Cold Spring Harbor Perspectives in Medicine*, *4*(9), a013888-a013888. doi: 10.1101/cshperspect.a013888

Merentie, M., Lipponen, J. A., Hedman, M., Hedman, A., Hartikainen, J., Husko, J., ... Ylä-Herttuala, S. (2015, December). Mouse ECG findings in aging, with conduction system affecting drugs and in cardiac pathologies:

Development and validation of ECG analysis algorithm in mice. *Physiological Reports*, *3*(12). doi: 10.14814/phy2.12639

Moon, K. R., van Dijk, D., Wang, Z., Chen, W., Hirn, M. J., Coifman, R. R., … Krishnaswamy, S. (2017). PHATE: A Dimensionality Reduction Method for Visualizing Trajectory Structures in High-Dimensional Biological Data. , 48.

Moon, K. R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D. B., Chen, W. S., … Krishnaswamy, S. (2019, December). Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, *37*(12), 1482–1492. doi: 10.1038/s41587-019-0336-3

Osuala, K., Telusma, K., Khan, S. M., Wu, S., Shah, M., Baker, C., … Ebert, S. N. (2011, July). Distinctive Left-Sided Distribution of Adrenergic-Derived Cells in the Adult Mouse Heart. *PLoS ONE*, *6*(7), e22811. doi: 10.1371/journal.pone.0022811

Owji, A., Varudkar, N., & Ebert, S. N. (2013, December). Therapeutic potential of Pnmt+ primer cells for neuro/myocardial regeneration. *American Journal of Stem Cells*, *2*(3), 137–154.

Papka, R. (1974, December). A study of catecholamine-containing cells in the hearts of fetal and postnatal rabbits by fluorescence and electron microscopy. *Cell and Tissue Research*, *154*(4). doi: 10.1007/BF00219668

Petiet, A. E., Kaufman, M. H., Goddeeris, M. M., Brandenburg, J., Elmore, S. A., & Johnson, G. A. (2008, August). High-resolution magnetic resonance histology of the embryonic and neonatal mouse: A 4D atlas and morphologic database. *Proceedings of the National Academy of Sciences*, *105*(34), 12331–12336. doi: 10.1073/pnas.0805747105

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling

in single cells. *Nature methods*, *10*(11), 1096–1098.

Qiu, P. (2020, March). Embracing the dropouts in single-cell RNA-seq analysis. *Nature Communications*, *11*(1), 1169. doi: 10.1038/s41467-020-14976-9

Richards, A. (2015, August). University of Oxford Advanced Research Computing. *Zenodo*. doi: 10.5281/zenodo.22558

Roussel, J., Champeroux, P., Roy, J., Richard, S., Fauconnier, J., Le Guennec, J.-Y., & Thireau, J. (2016, May). The Complex QT/RR Relationship in Mice. *Scientific Reports*, *6*(1), 25388. doi: 10.1038/srep25388

Salvi, V., Karnad, D. R., Panicker, G. K., & Kothari, S. (2010). Update on the evaluation of a new drug for effects on cardiac repolarization in humans: issues in early drug development. *British Journal of Pharmacology*, *159*(1), 34–48.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... Cardona, A. (2012, Jun). Fiji: an open-source platform for biological-image analysis. *Nat Methods*, *9*(7), 676–682.

Speerschneider, T., & Thomsen, M. B. (2013, December). Physiology and analysis of the electrocardiographic T wave in mice. *Acta Physiologica (Oxford, England)*, *209*(4), 262–271. doi: 10.1111/apha.12172

Starmer, J. (2017). *StatQuest: A gentle introduction to RNA-seq*. `https://www.youtube.com/watch?v=tlf6wYJrwKY`. (Online; accessed 13 July 2021)

Swirski, F. K., & Nahrendorf, M. (2018, December). Cardioimmunology: The immune system in cardiac homeostasis and disease. *Nature Reviews Immunology*, *18*(12), 733–744. doi: 10.1038/s41577-018-0065-8

Thomas, S. A., Marck, B. T., Palmiter, R. D., & Matsumoto, A. M. (1998). Restoration of Norepinephrine and Reversal of Phenotypes in Mice Lacking Dopamine β-Hydroxylase. *Journal of Neurochemistry*, *70*(6), 2468–2476.

doi: 10.1046/j.1471-4159.1998.70062468.x

Thomas, S. A., Matsumoto, A. M., & Palmiter, R. D. (1995, April). Noradrenaline is essential for mouse fetal development. *Nature*, *374*(6523), 643–646. doi: 10.1038/374643a0

Triposkiadis, F., Karayannis, G., Giamouzis, G., Skoularigis, J., Louridas, G., & Butler, J. (2009, November). The Sympathetic Nervous System in Heart Failure. *Journal of the American College of Cardiology*, *54*(19), 1747–1762. doi: 10.1016/j.jacc.2009.05.015

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(86), 2579–2605.

van Dongen, S. (2000). *Graph Clustering by Flow Simulation*.

Walden, A. P., Dibb, K. M., & Trafford, A. W. (2009, April). Differences in intracellular calcium homeostasis between atrial and ventricular myocytes. *Journal of Molecular and Cellular Cardiology*, *46*(4), 463–473. doi: 10.1016/j.yjmcc.2008.11.003

Wang, Y., Lin, W. K., Crawford, W., Ni, H., Bolton, E. L., Khan, H., ... Lei, M. (2017, February). Optogenetic Control of Heart Rhythm by Selective Stimulation of Cardiomyocytes Derived from Pnmt+ Cells in Murine Heart. *Scientific Reports*, *7*(1), 40687. doi: 10.1038/srep40687

Wengrowski, A. M., Wang, X., Tapa, S., Posnack, N. G., Mendelowitz, D., & Kay, M. W. (2015, February). Optogenetic release of norepinephrine from cardiac sympathetic neurons alters mechanical and electrical function. *Cardiovascular Research*, *105*(2), 143–150. doi: 10.1093/cvr/cvu258

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). Scanpy: large-scale single-cell gene expression data analysis. *Genome Biology*, *19*(1), 1–5.

Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., ... Theis, F. J. (2019). Paga: graph abstraction reconciles clustering

with trajectory inference through a topology preserving map of single cells. *Genome biology*, *20*(1), 1–9.

Zaglia, T., & Mongillo, M. (2017, June). Cardiac sympathetic innervation, from a different point of (re)view: Neuro-cardiac communication. *The Journal of Physiology*, *595*(12), 3919–3930. doi: 10.1113/JP273120

Zhang, Y., Wu, J., King, J. H., Huang, C. L.-H., & Fraser, J. A. (2014, June). Measurement and interpretation of electrocardiographic QT intervals in murine hearts. *American Journal of Physiology-Heart and Circulatory Physiology*, *306*(11), H1553-H1557. doi: 10.1152/ajpheart.00459.2013

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., ... Enard, W. (2017, February). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular Cell*, *65*(4), 631-643.e4. doi: 10 .1016/j.molcel.2017.01.023

A | Supplementary figures

Supplementary figures begin on the next page.

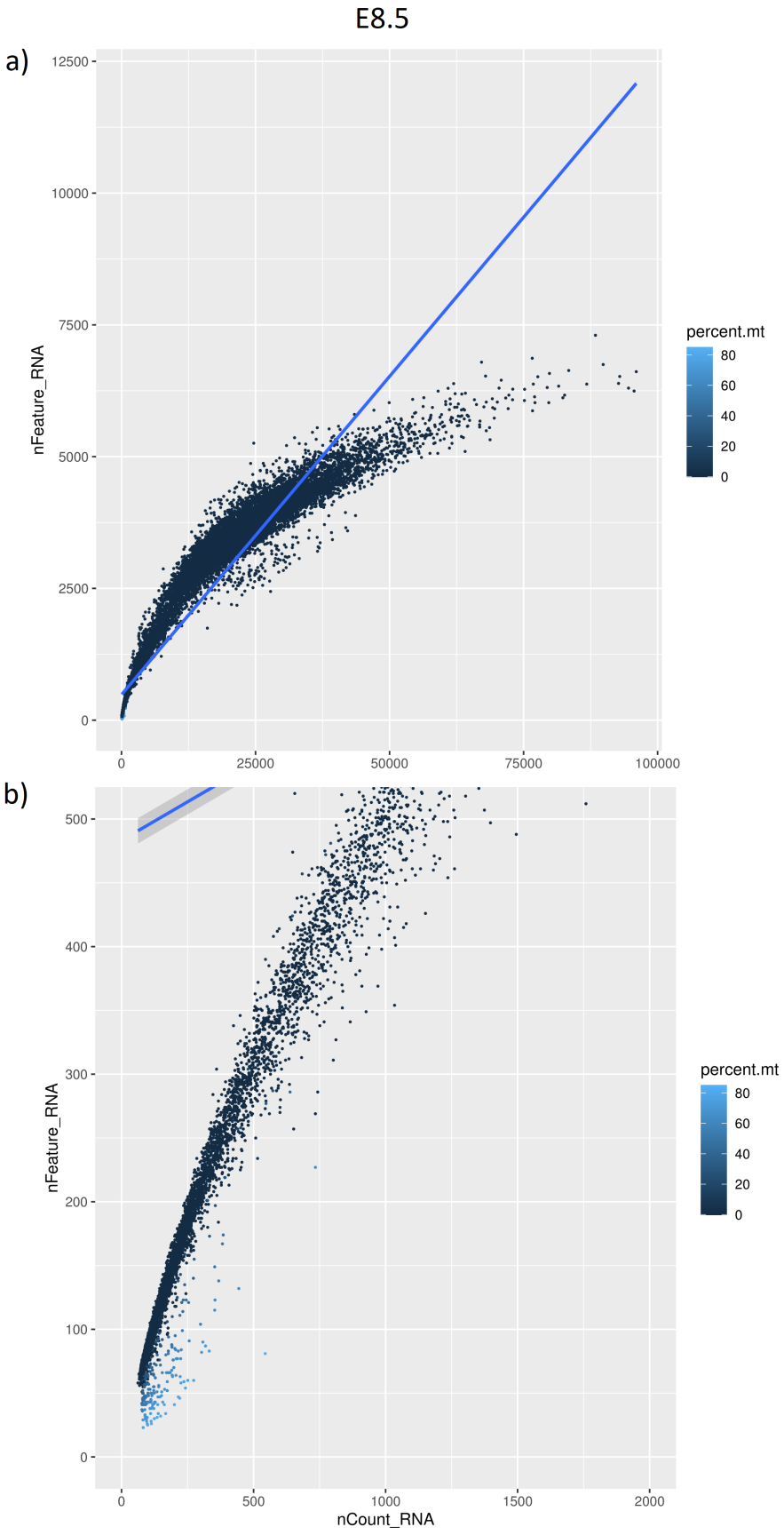


Figure A.1: Number of detected genes vs RNA count for E8.5. Blue: proportion mitochondrial RNA. Blue line: best fit line. b) A magnified version of the bottom left region of (a).

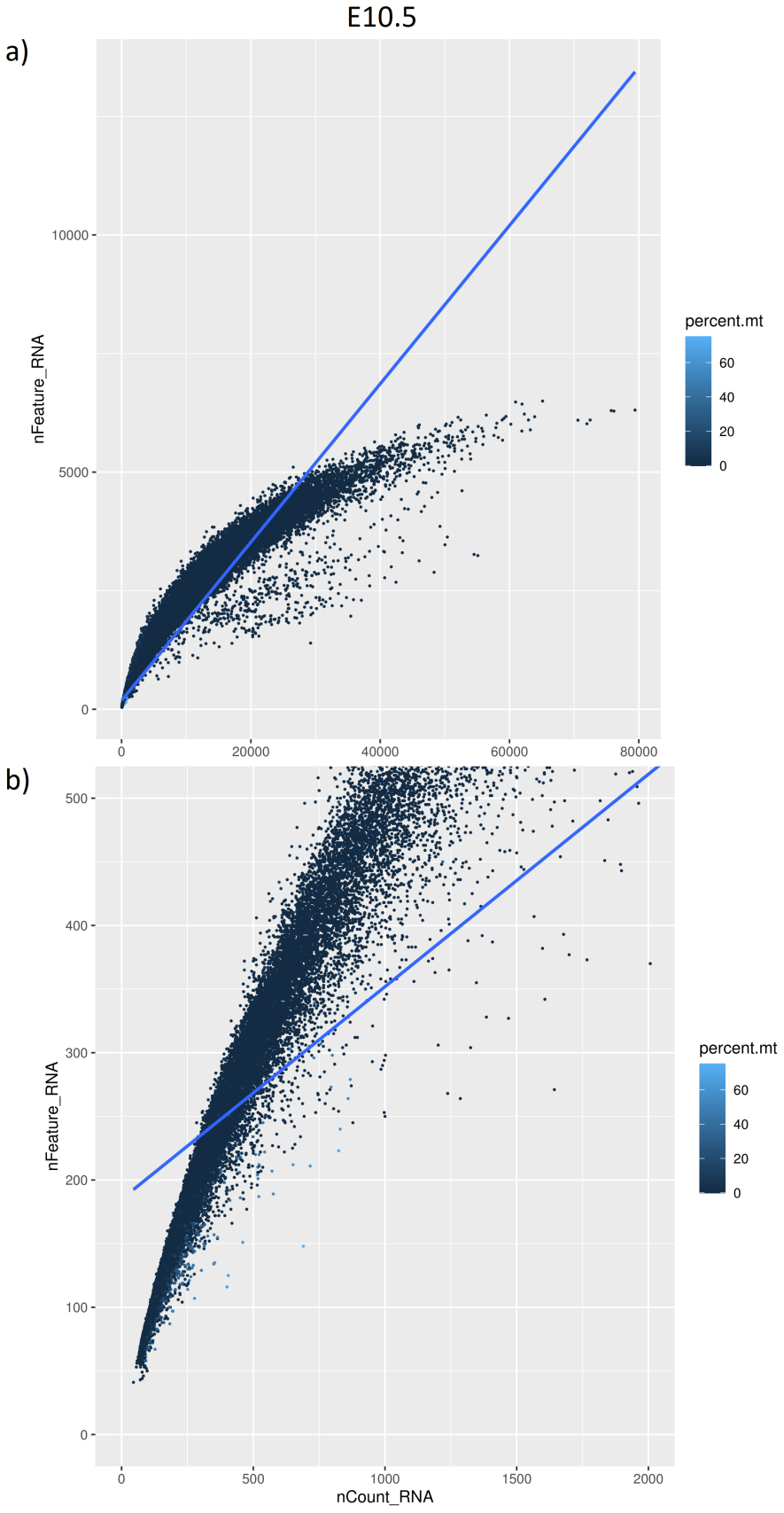


Figure A.2: Number of detected genes vs RNA count for E10.5. Blue: proportion mitochondrial RNA. Blue line: best fit line. b) A magnified version of the bottom left region of (a).

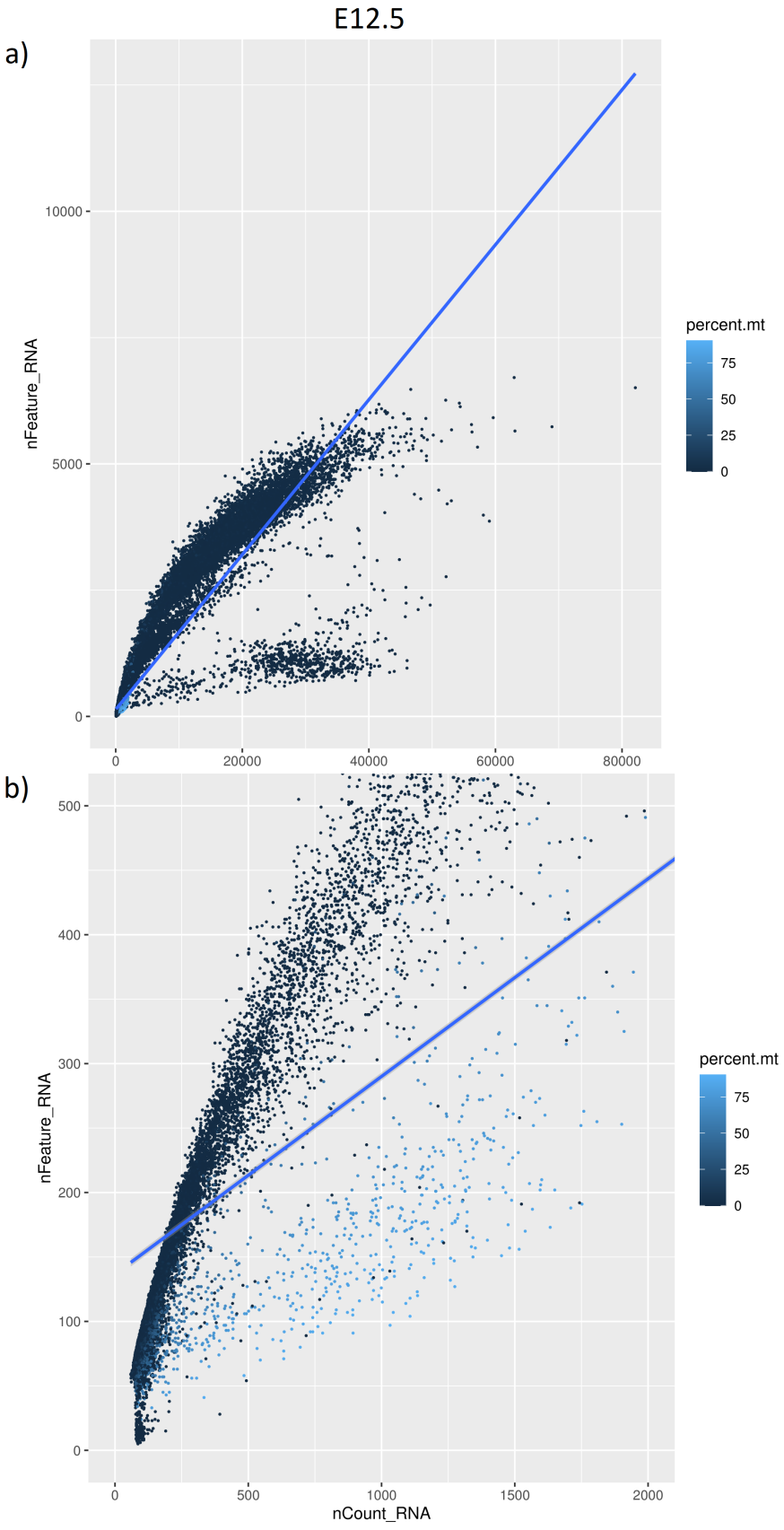


Figure A.3: Number of detected genes vs RNA count for E12.5. Blue: proportion mitochondrial RNA. Blue line: best fit line. b) A magnified version of the bottom left region of (a).

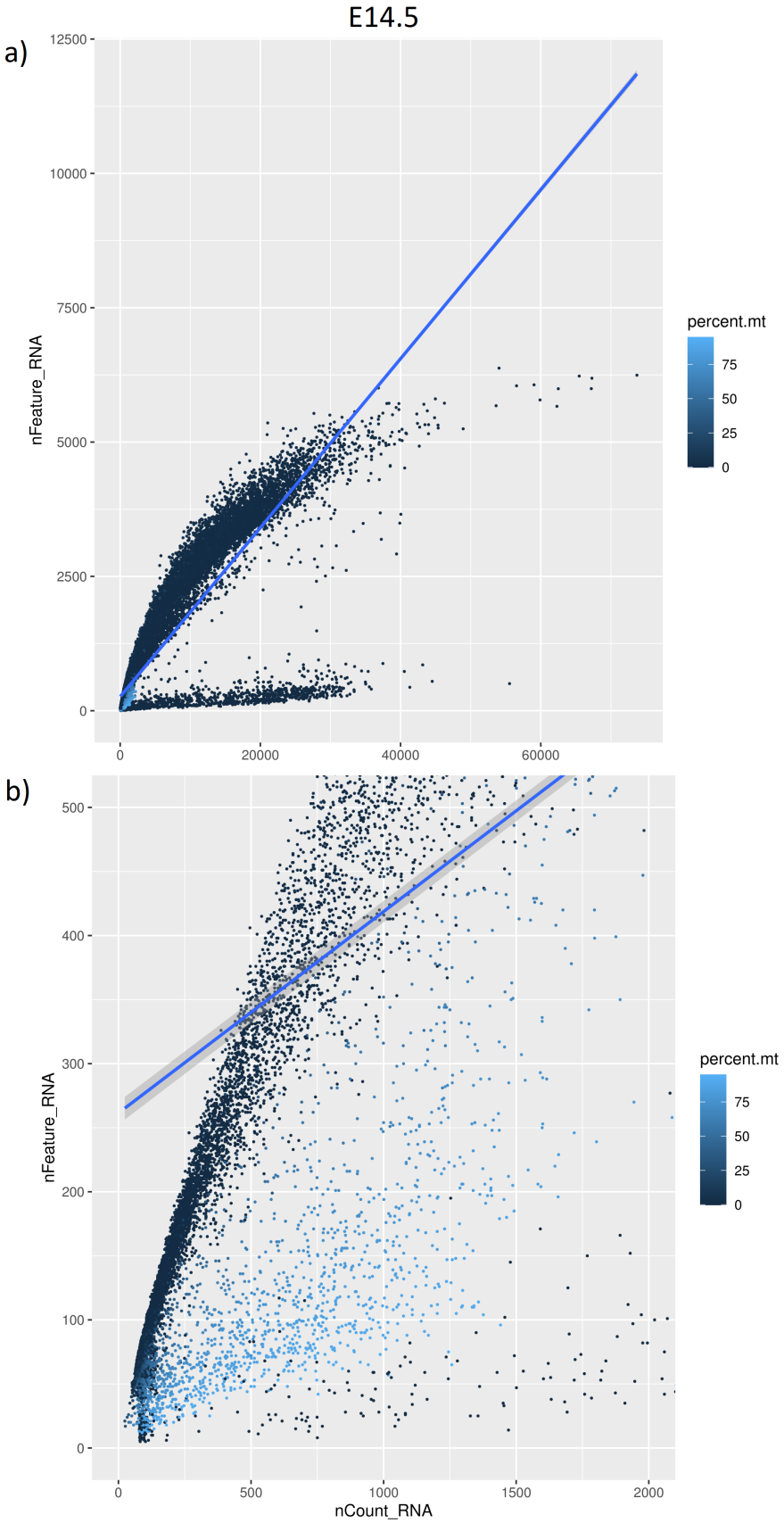


Figure A.4: Number of detected genes vs RNA count for E14.5. Blue: proportion mitochondrial RNA. Blue line: best fit line. b) A magnified version of the bottom left region of (a).

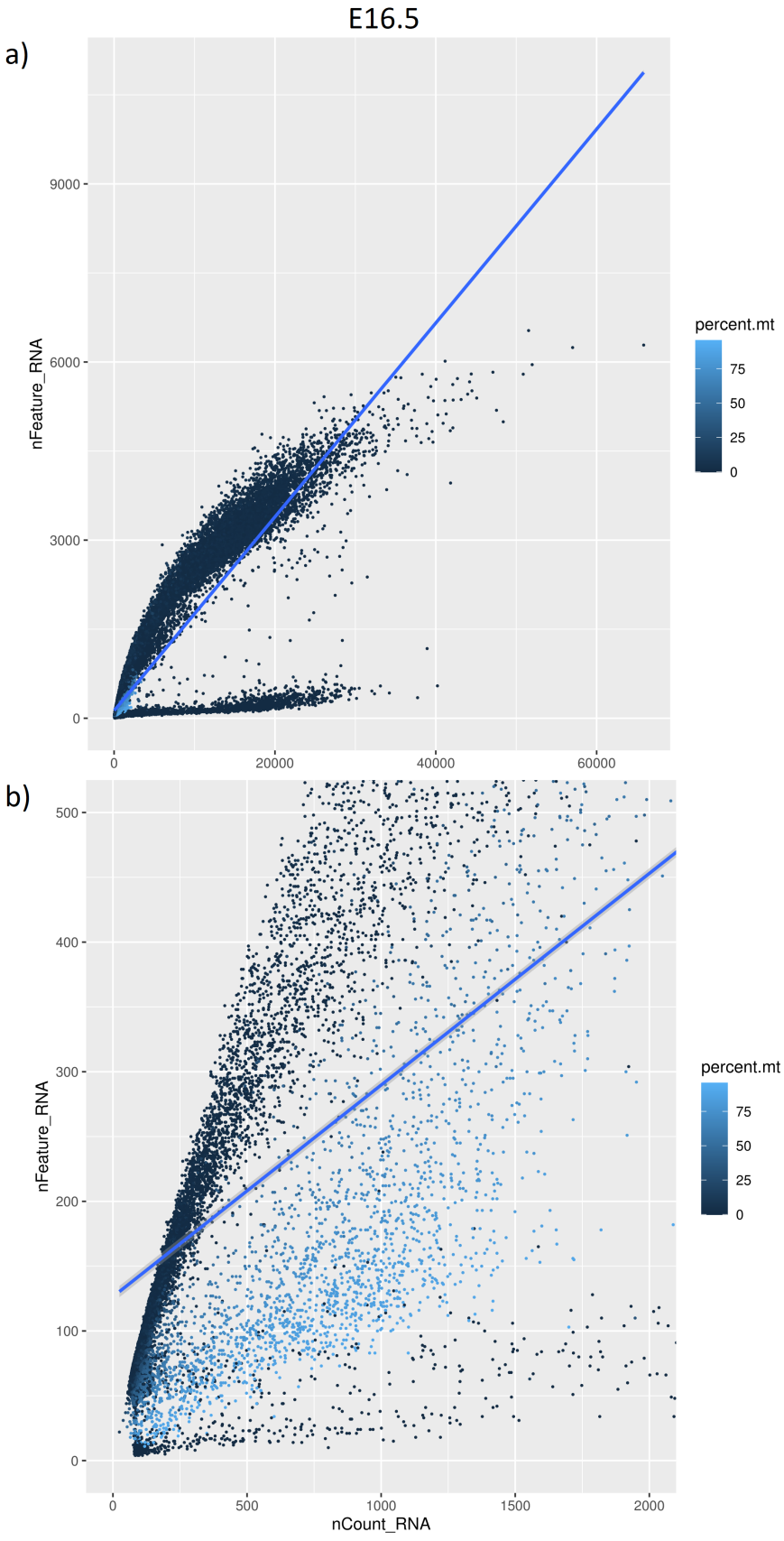


Figure A.5: Number of detected genes vs RNA count for E16.5. Blue: proportion mitochondrial RNA. Blue line: best fit line. b) A magnified version of the bottom left region of (a).

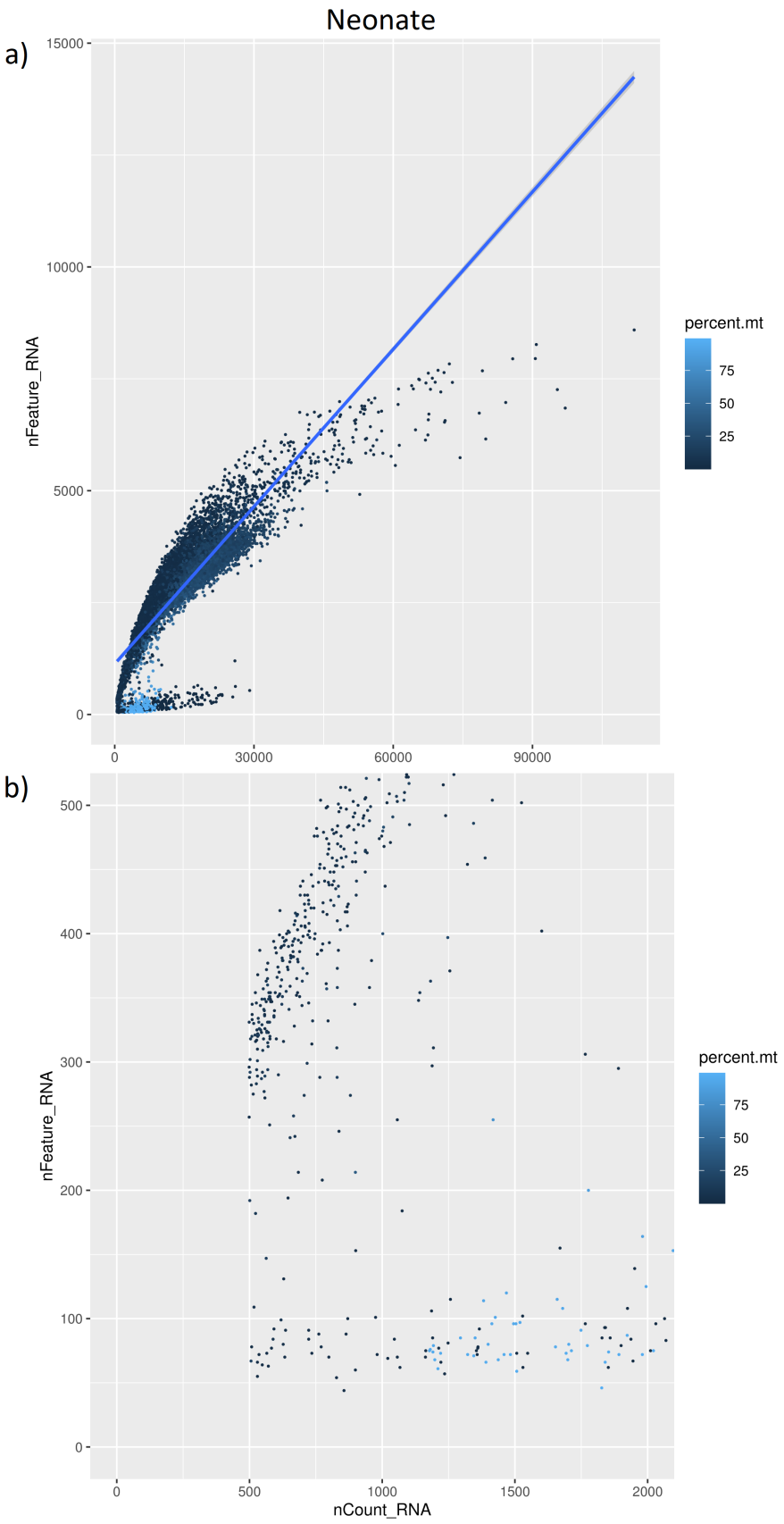


Figure A.6: Number of detected genes vs RNA count for the neonate. Blue: proportion mitochondrial RNA. Blue line: best fit line. b) A magnified version of the bottom left region of (a).

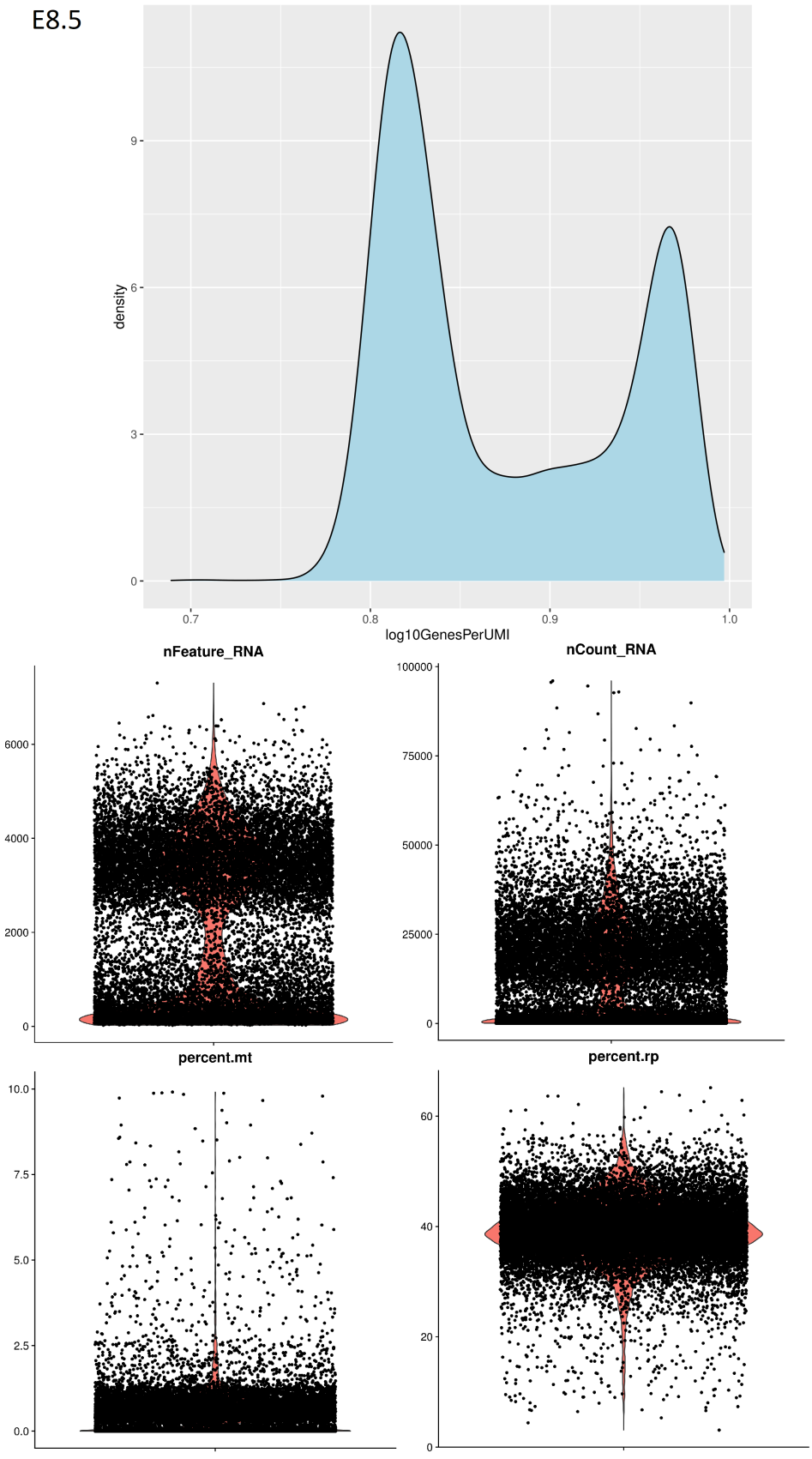


Figure A.7: Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E8.5. Measure of transcriptomic complexity given by log10GenesPerUMI.

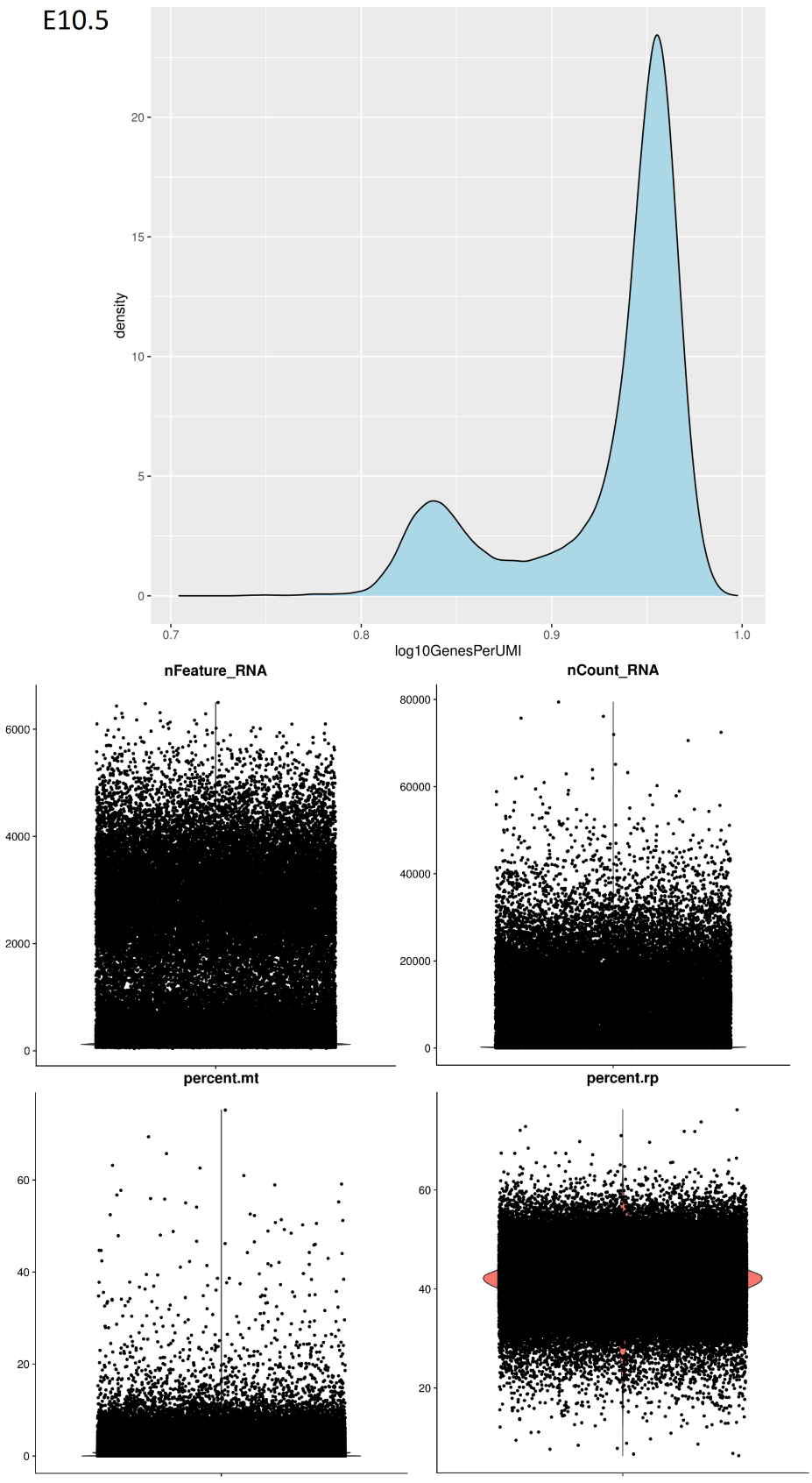


Figure A.8: Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E10.5. Measure of transcriptomic complexity given by log10GenesPerUMI.

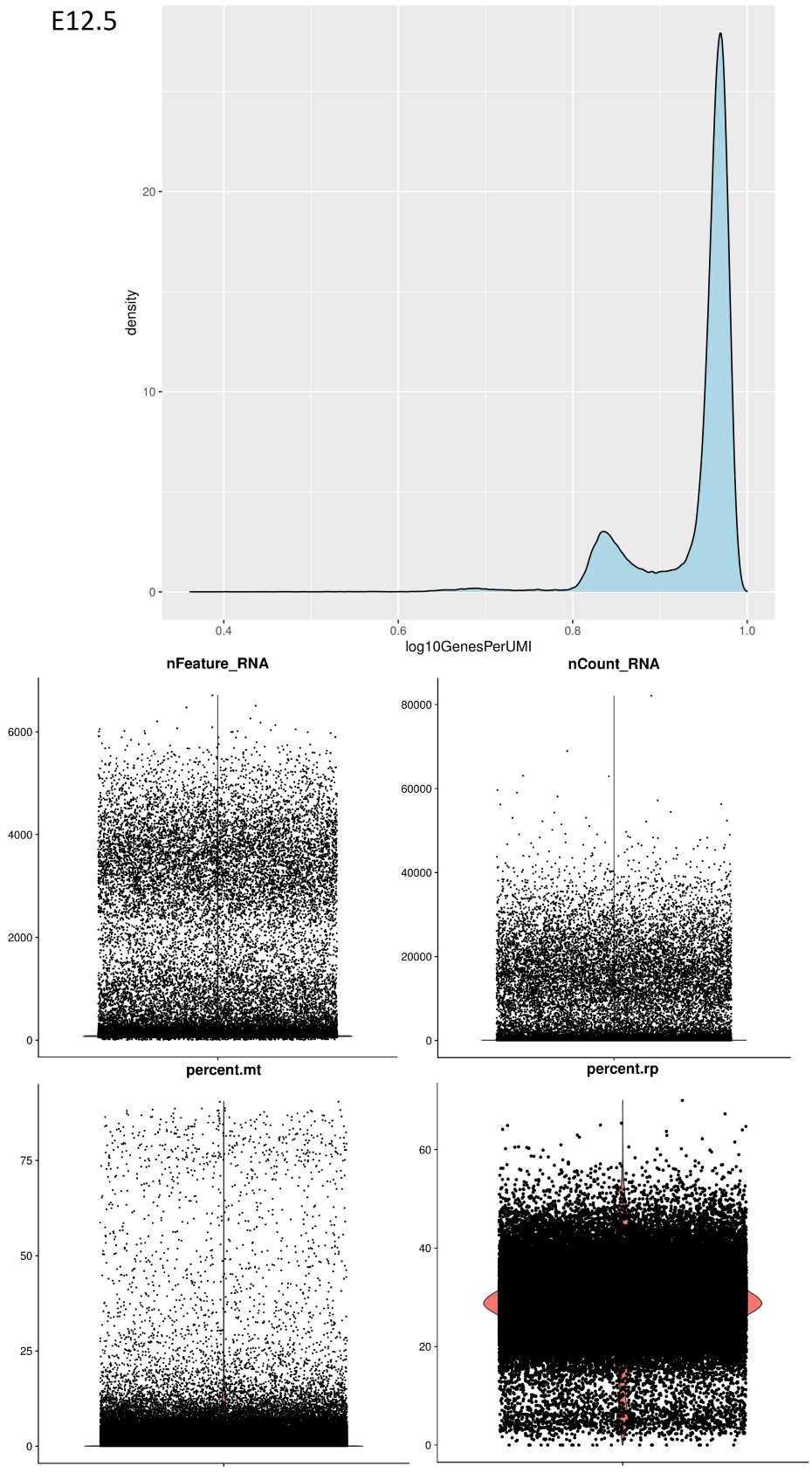


Figure A.9: Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E12.5. Measure of transcriptomic complexity given by log10GenesPerUMI.

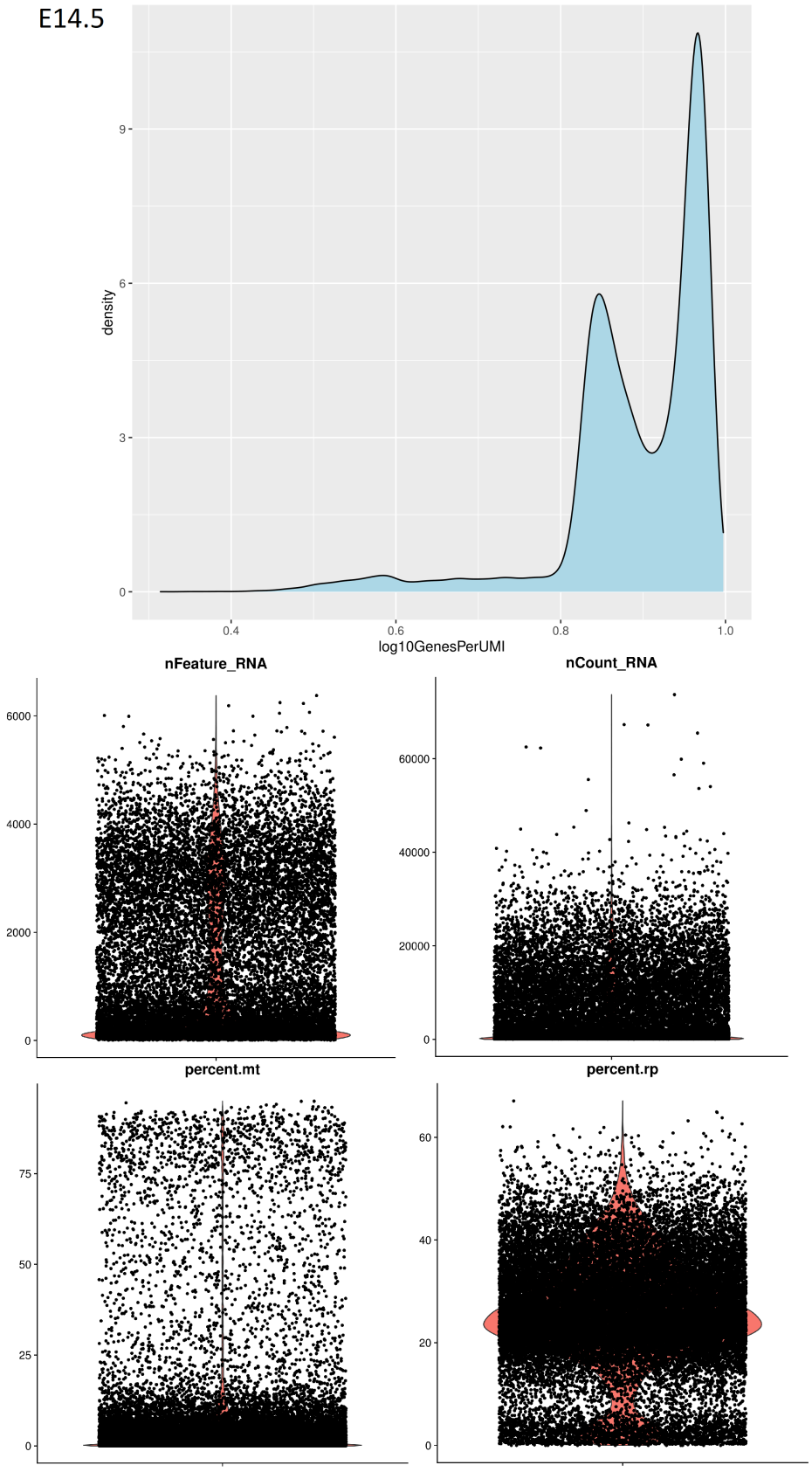


Figure A.10: Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E14.5. Measure of transcriptomic complexity given by log10GenesPerUMI.

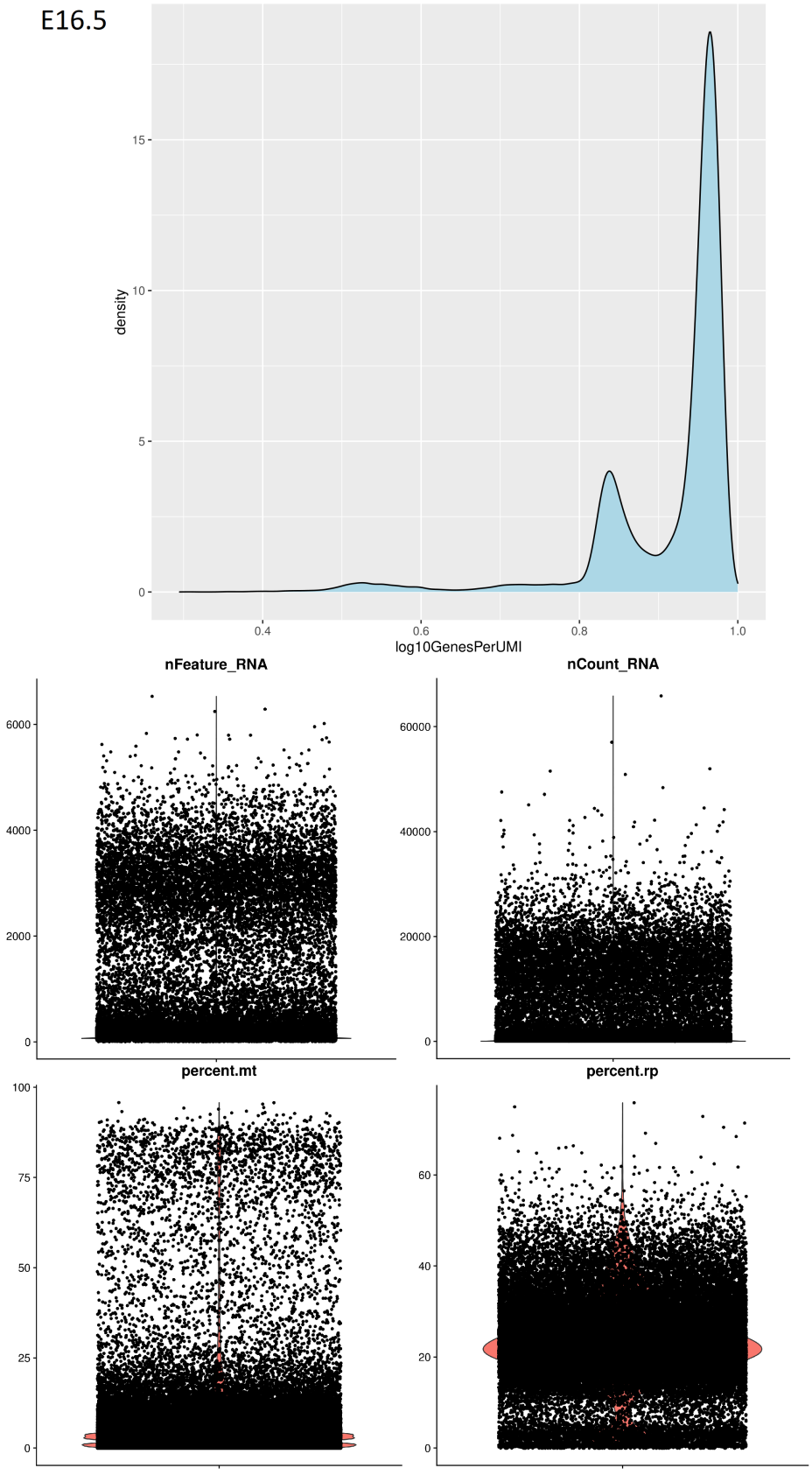


Figure A.11: Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for E16.5. Measure of transcriptomic complexity given by log10GenesPerUMI.

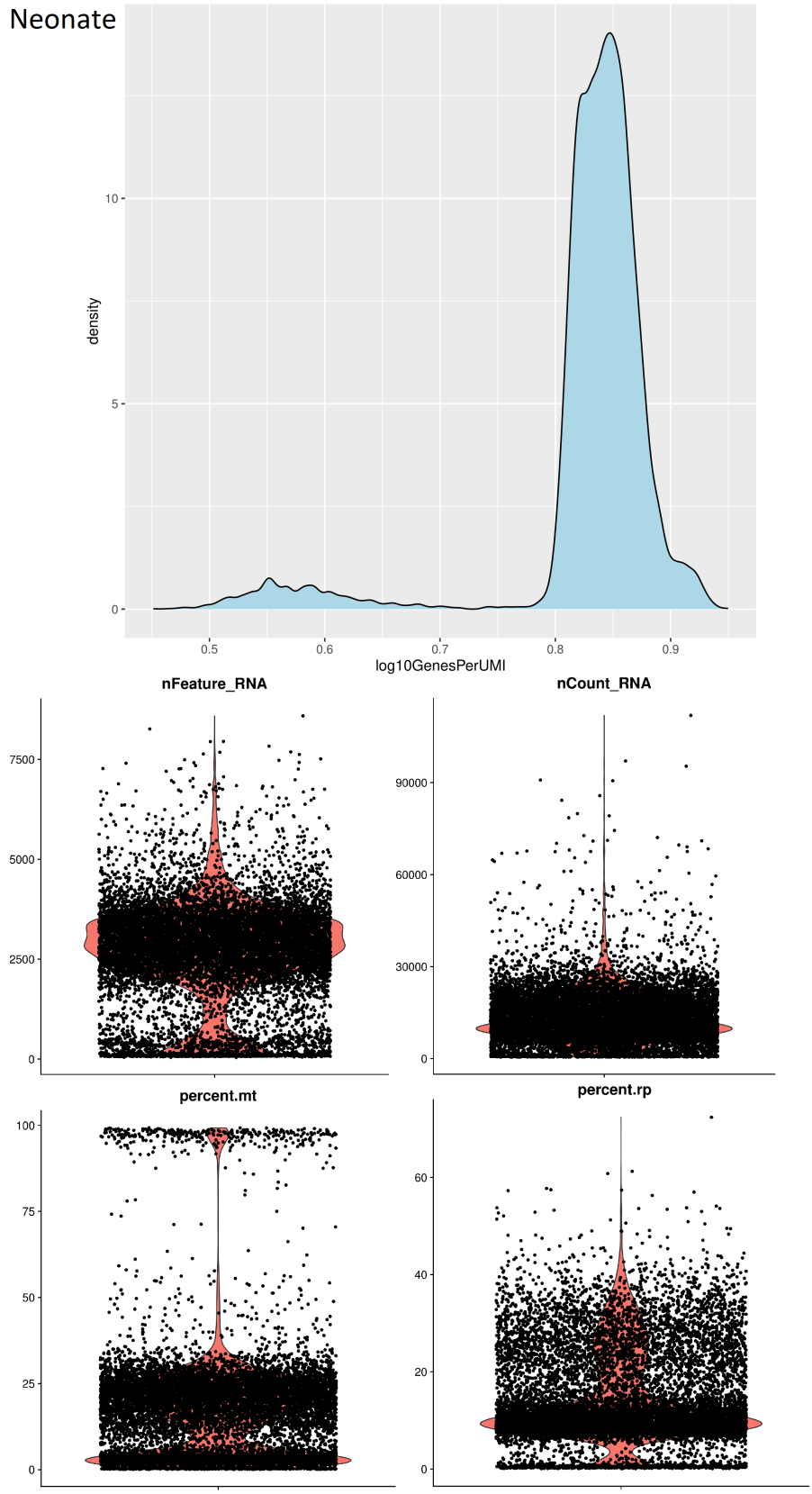


Figure A.12: Transcriptomic complexity, RNA count, proportion ribosomal gene RNA, proportion mitchondrial RNA and gene count for the neonate. Measure of transcriptomic complexity given by log10GenesPerUMI.

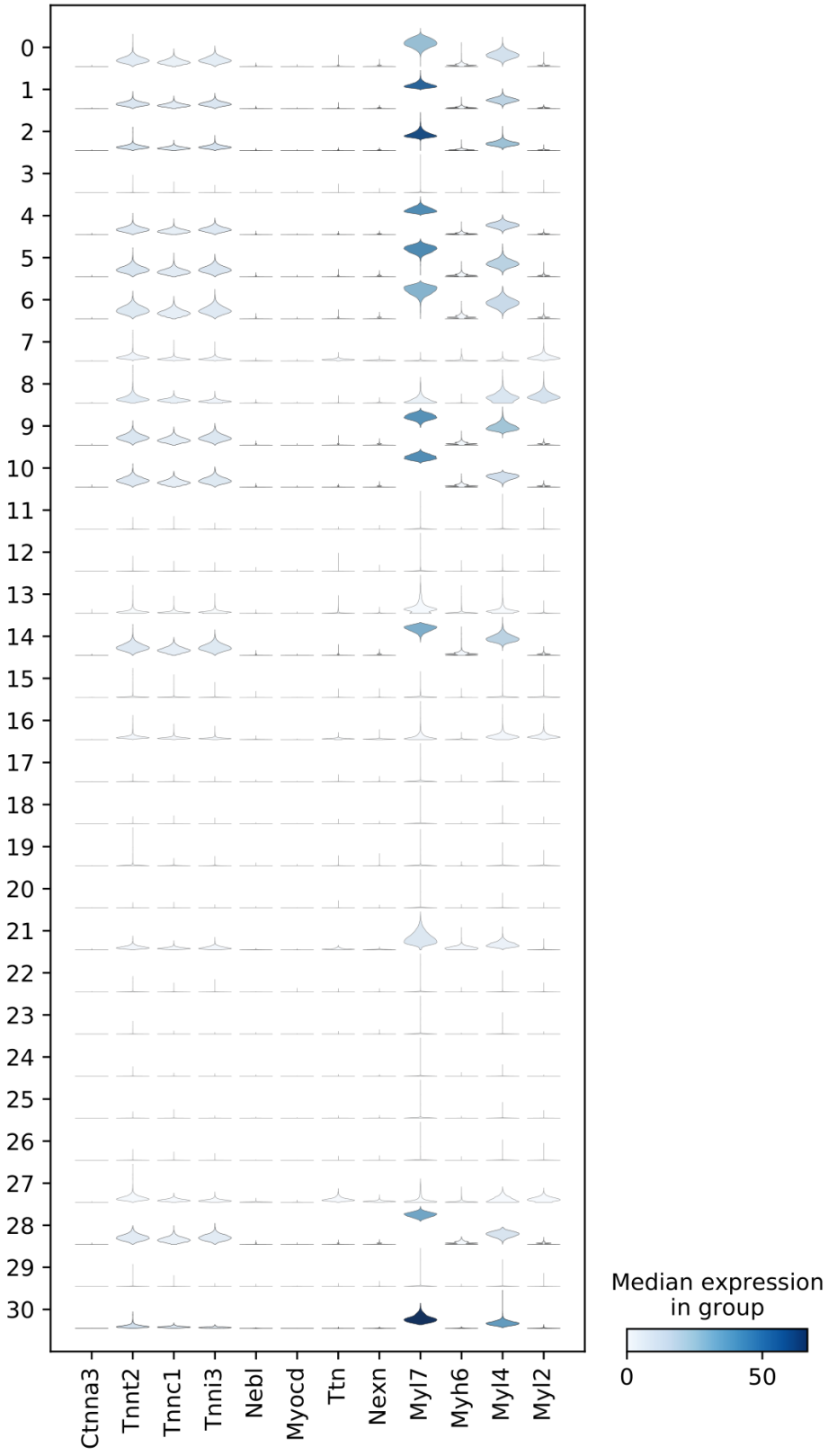


Figure A.13: Ventricular and atrial marker gene expressions by cluster. Shown using stacked violin plots—each violin plot shows the probability distribution of the mRNA counts for each cell. Myl7, Myh6 and Myl4 show high expression in the atrial clusters, whereas Myl2 shows high expression in the ventricular clusters.